

Complexity and coherency: integrating information in the brain

Giulio Tononi, Gerald M. Edelman and Olaf Sporns

The brains of higher mammals are extraordinary integrative devices. Signals from large numbers of functionally specialized groups of neurons distributed over many brain regions are integrated to generate a coherent, multimodal scene. Signals from the environment are integrated with ongoing, patterned neural activity that provides them with a meaningful context. We review recent advances in neurophysiology and neuroimaging that are beginning to reveal the neural mechanisms of integration. In addition, we discuss concepts and measures derived from information theory that lend a theoretical basis to the notion of complexity as integration of information and suggest new experimental tests of these concepts.

The brain's capacity rapidly to integrate information from many different sources lies at the root of our cognitive abilities. Two general questions are particularly relevant to understanding this capacity. First, how is the specialized information conveyed by the activity of functionally segregated areas and neuronal groups integrated into a unified, coherent scene? This question underlies one of the oldest controversies in neuroscience – that between localizationist and holistic views of brain function. Second, how is the information that is conveyed by incoming stimuli integrated with information present in memory? It is often assumed that the brain controls behavior by processing incoming stimuli in the form of neural activity patterns. On the other hand, it is also assumed that the brain lays down memories of previously encountered stimuli in the form of patterns of connectivity among neurons. Consequently, information processing and information storage are often studied separately and with different methodologies.

In this review, we consider experimental evidence as well as recent theoretical studies suggesting that the question of how the brain integrates information can be addressed within a unified conceptual framework. We first discuss known neural mechanisms underlying cognitive and behavioral integration. We then examine a set of statistical measures derived from information theory that can be used to characterize the integration of information among functionally segregated groups of neurons. Finally, after considering the key role of spontaneous activity in brain function, we examine how these statistical measures can be used to evaluate the integration of incoming stimuli with ongoing neural interactions.

Functional segregation and integration

Phrenologists had imagined that different cognitive functions were allocated to different parts of the brain well before any neurobiological evidence was available. Advances in neuroscience appear to have fulfilled the phrenologists' dream and have conclusively demonstrated that functional specialization at multiple spatial scales is a fundamental principle of brain organization. In the visual system, for example, different brain areas are functionally specialized for visual attributes such as shape, motion, and color¹⁻³, and parcellation of function has been discovered within other sensory modalities and in the motor domain⁴⁻⁶. Functional segregation extends to the level of columns or groups of neurons⁷. In primary visual cortex neuronal groups are specialized for different stimulus orientations^{8,9}, direction of motion¹⁰ and spatial frequency¹¹. A similar specialization along different stimulus dimensions has been discovered in essentially every brain area that has been studied in sufficient detail^{12,13}. Most recently, it has been demonstrated that different brain regions can be activated by specific cognitive tasks or by specific stimulus attributes whether these are perceived, imagined, or remembered¹⁴.

While the evidence for regional specialization in the brain is overwhelming, it is clear that the information conveyed by the activity of specialized groups of neurons must be functionally integrated in order to guide adaptive behavior – just consider how many different signals must be rapidly evaluated and coherently integrated to navigate safely in a busy city. Like functional specialization, functional integration occurs at multiple spatial and temporal scales. In vision, for example, individual elements (dots, edges) are

G. Tononi, G.M. Edelman and O. Sporns are at The Neurosciences Institute, 10640 John J. Hopkins Drive, San Diego, CA 92121, USA.

tel: +1 619 626 2000
fax: +1 619 626 2099
e-mail: tononi@nsi.edu

grouped together to yield shapes according to Gestalt laws, such as those of common motion, colinearity, continuity, and proximity^{15,16}. Different attributes (shape, color, location, size, etc.) must be bound together to form objects, and multiple objects coexist within a single visual image¹⁷. Images themselves are integrated with auditory, somatosensory and proprioceptive inputs to yield a coherent, unified conscious scene¹⁸. Functional integration also occurs whether or not the brain is responding to extrinsic stimuli: we experience a unified, integrated conscious scene whether we are awake and responding to stimuli, whether we are imagining or thinking, or whether we are asleep and dreaming.

Mechanisms of neural integration and binding

How does the brain 'bind' together the attributes of objects and events in space and time to construct a unified conscious scene? Undoubtedly, neurons can integrate frequently co-occurring constellations of features by convergent connectivity^{19,20}. However, convergence is unlikely to be the predominant mechanism for integration. First, no single ('master') brain area has been identified, the activity of which represents entire perceptual or mental states. Second, the vast number of possible perceptual stimuli occurring in ever changing contexts greatly exceeds the number of available neuronal groups (or even single neurons), thus causing a combinatorial explosion²¹. Third, convergence does not allow for dynamic ('on-the-fly') conjunctions in response to novel, previously unencountered stimuli.

A crucial hint about a more general mechanism of neural integration is given by a key feature of the anatomical organization of the brain: reciprocal and parallel connectivity among functionally segregated groups of neurons is the rule rather than the exception. Based on these considerations, a theoretical solution of the problem of integration has been proposed that takes into account the cooperative interactions within and among functionally segregated brain areas as mediated by a process of ongoing, parallel signaling – called 're-entry'^{22,23}. Detailed computer simulations of visual cortical areas have shown that re-entrant interactions can synchronize the activity of functionally specialized groups of neurons²⁴, thus providing a neural basis for several Gestalt laws²⁵. Furthermore, simulations of the visual system have shown that re-entrant interactions among functionally specialized cortical areas generate short-term correlations in their firing that lead to coherent perceptual performance and behavior in the absence of a master area²⁶, thus providing a parsimonious solution to the binding problem¹⁷. It should be emphasized that a key feature of these models was that backward connections as well as horizontal intra-areal connections were voltage-dependent; that is, their efficacy was controlled by postsynaptic depolarization on a short timescale. Such connections do not by themselves drive target cells but they modulate the amount and precise timing of their firing.

Experimentally, the pervasive occurrence of synchronous firing among cortical and thalamic structures during various cognitive tasks has been demonstrated by a large number of studies using different methodologies^{27,28}. Several recent experiments suggest that the synchronization of firing of distinct groups of neurons in the gamma fre-

quency range may be a bona fide indicator of early levels of stimulus integration according to Gestalt principles²⁹. Furthermore, if re-entrant interactions are prevented from occurring; for example, by transection of callosal connections, both perceptual integration and synchronous firing are abolished³⁰. In addition, there have been numerous demonstrations of broad-band synchronization during various sensorimotor and cognitive tasks both within and among multiple brain areas, including areas outside the visual system^{31–35}. Taken together, these findings suggest that the rapid integration of information within the thalamo-cortical system does not occur in a particular location but rather in terms of a unified neural process. How can one establish the presence and the extent of such a process, and what kind of evidence should one look for in neurophysiological and cognitive experiments?

Functional clustering

Imaging studies are typically used to establish whether certain brain regions are more or less active than other brain regions during a cognitive task³⁶. However, techniques such as PET and fMRI can also be employed to explore functional interactions between different brain areas, for example, by using multivariate statistical methods. These include multidimensional scaling, path analysis and cluster analysis^{37–43}.

One of the first questions one might ask about functional interactions is whether, during a cognitive task, there is any evidence of functional clustering. A functional cluster can be defined as a set of brain regions that interact much more strongly with each other than with the rest of the brain⁴³. In order to determine the presence of functional clustering, it is important to employ multivariate measures of statistical dependence and independence that permit the simultaneous characterization of the interactions among many elements (see Box 1). In particular, a measure taken from the statistical foundations of information theory, called 'mutual information', can be used to measure the statistical dependence between two subsets of neural elements. A quantity derived from mutual information, called 'integration', can be used to measure the total statistical dependence among the elements of a neural system. A functional cluster (Fig. 1) can then be defined as a subset of brain regions for which the statistical dependence within the subset (integration) is much larger than that between the subset and rest of the brain (mutual information).

This approach has been applied to PET data obtained from normal and schizophrenic subjects performing a set of cognitive tasks. Analysis of the data in terms of integration and mutual information provided evidence that subsets of distributed brain regions activated during the task constituted a functional cluster. A comparative evaluation of functional interactions within the cluster revealed distinct differences in functional integration between schizophrenics and controls despite the absence of differences in activation⁴³. Clearly, the search for functional clustering during cognitive activity has just begun, and it needs to be expanded by imaging methodologies that offer better spatial and temporal resolution. For example, can one find evidence for subsets of integrated brain regions that are functionally insulated as clusters at a timescale of hundreds of

Box 1. Entropy, mutual information, and integration

Entropy and mutual information are basic statistical concepts that were originally introduced in information theory^a but that have gained much wider applications^b. Entropy is a measure of uncertainty or variability, while mutual information is a measure of how much uncertainty is shared (statistical dependence).

Let X be a system composed of a set of elements $\{x_i\}$, and let us assume that the system can assume a number $m = 1 \dots M$ of discrete states. Each of these states is associated with a probability p_m , such that the sum of their probabilities equals 1. For the system X , the entropy is given by:

$$H(X) = - \sum_{m=1}^M p_m \log_2(p_m) \quad [1]$$

$H(X)$ is large if the system has many equally likely states (high uncertainty). On the other hand, $H(X)$ is zero if and only if the system attains only a single state with $p = 1$. In that case, there is no uncertainty about the state of the system.

The uncertainty about the state of a subset X_i of the system X , which is accounted for by the state of the rest of the system ($X - X_i$) is called mutual information^c, given by:

$$MI(X_i; X - X_i) = H(X_i) + H(X - X_i) - H(X) \quad [2]$$

Note that MI is positive and symmetric. Mutual information is a measure of statistical dependence that expresses how much information is provided about the state of a subset by 'knowing' the state of the rest of the system (and vice versa). Mutual information is zero in the case of statistical independence. Compared to other measures of statistical dependence such as, for example, correlation coefficients, mutual information has the advantage that it is multivariate and that it captures both linear and non-linear dependencies (for elegant applications of information theory to neurophysiology see Refs d–g). Note that mutual information reflects a statistical dependence among subsets of a system irrespective of its source. The presence and direction of causal interactions between two subsets of a system can be evaluated, at least in principle, by measuring the change in mutual information obtained by perturbing each subset in turn.

One can also define the overall deviation from statistical independence among all $\{x_i\}$ individual elements of a system X (Ref. h). This quantity, which measures the total entropy loss, is called integration $I(X)$ and it is defined as the difference between the sum of the entropies of the individual elements and the entropy of the entire system:

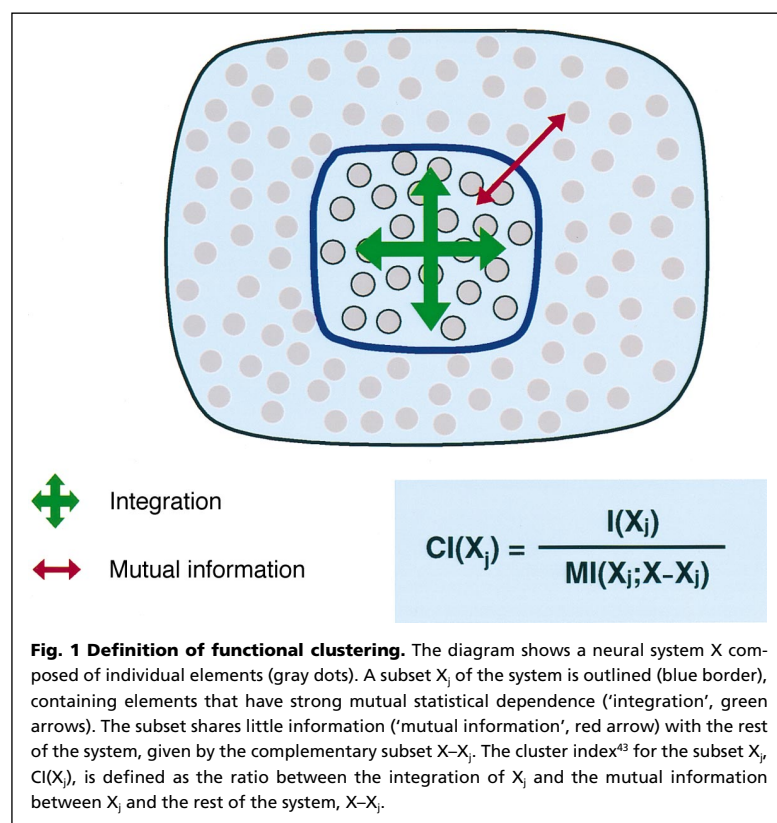
$$I(X) = \sum H(x_i) - H(X) \quad [3]$$

As the term 'integration' implies, $I(X)$ is zero if the elements of a system are statistically independent, and is positive whenever there are statistical dependencies among the elements.

Entropy, mutual information, and integration can also be defined for continuum systems. Standard procedures exist for calculating entropy and thus mutual information values from covariance or correlation matrices^c.

References

- a Shannon, C.E. and Weaver, W. (1963) *The Mathematical Theory of Communication*, University of Illinois Press
- b Zurek, W.H., ed. (1990) *Complexity, Entropy, and the Physics of Information*, Addison-Wesley
- c Jones, D.S. (1979) *Elementary Information Theory*, Clarendon Press
- d Dan, Y., Atick, J.J. and Reid, R.C. (1996) Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory *J. Neurosci.* 16, 3351–3362
- e Richmond, B.J., Gawne, T.J. and Jin, G.X. (1996) Neuronal codes: reading them and learning how their structure influences network organization *Biosystems* 40, 149–157
- f Phillips, W.A. and Singer, W. (1997) In search of common foundations for cortical computation *Behav. Brain Sci.* 20, 657–722
- g Rieke, F., Warland, D., de Ruyter van Steveninck, B. and Bialek, W. (1997) *Spikes: Exploring the Neural Code*, MIT Press
- h Tononi, G., Sporns, O. and Edelman, G.M. (1994) A measure for brain complexity: relating functional segregation and integration in the nervous system *Proc. Natl. Acad. Sci. U. S. A.* 91, 5033–5037



milliseconds and that vary with a cognitive task? Such a demonstration would have important implications for our understanding of cognitive processes, especially of conscious experience⁴⁴, because, by definition, at a given timescale, only signals exchanged within a functional cluster can be integrated.

Reconciling functional segregation and integration: neural complexity

How much information is integrated in a functional cluster of strongly interacting brain regions? From a theoretical point of view, high information and high integration present opposing requirements: the former requires the mutually independent firing of specialized groups of neurons, the latter requires that their joint activity be highly coherent. This paradox can be addressed by considering functional segregation and integration within a unified theoretical framework provided by the statistical foundations of information theory. The key idea is to estimate the average integration for subsets of the neural system of increasing size; that is, at multiple spatial scales⁴⁵. When this is done (Fig. 2A), the degree of specialization and integration within the system can be simultaneously evaluated. For example, if a system is composed of functionally segregated elements, the average integration for small subsets is low. This indicates that, taken a few at a time, such elements have independent

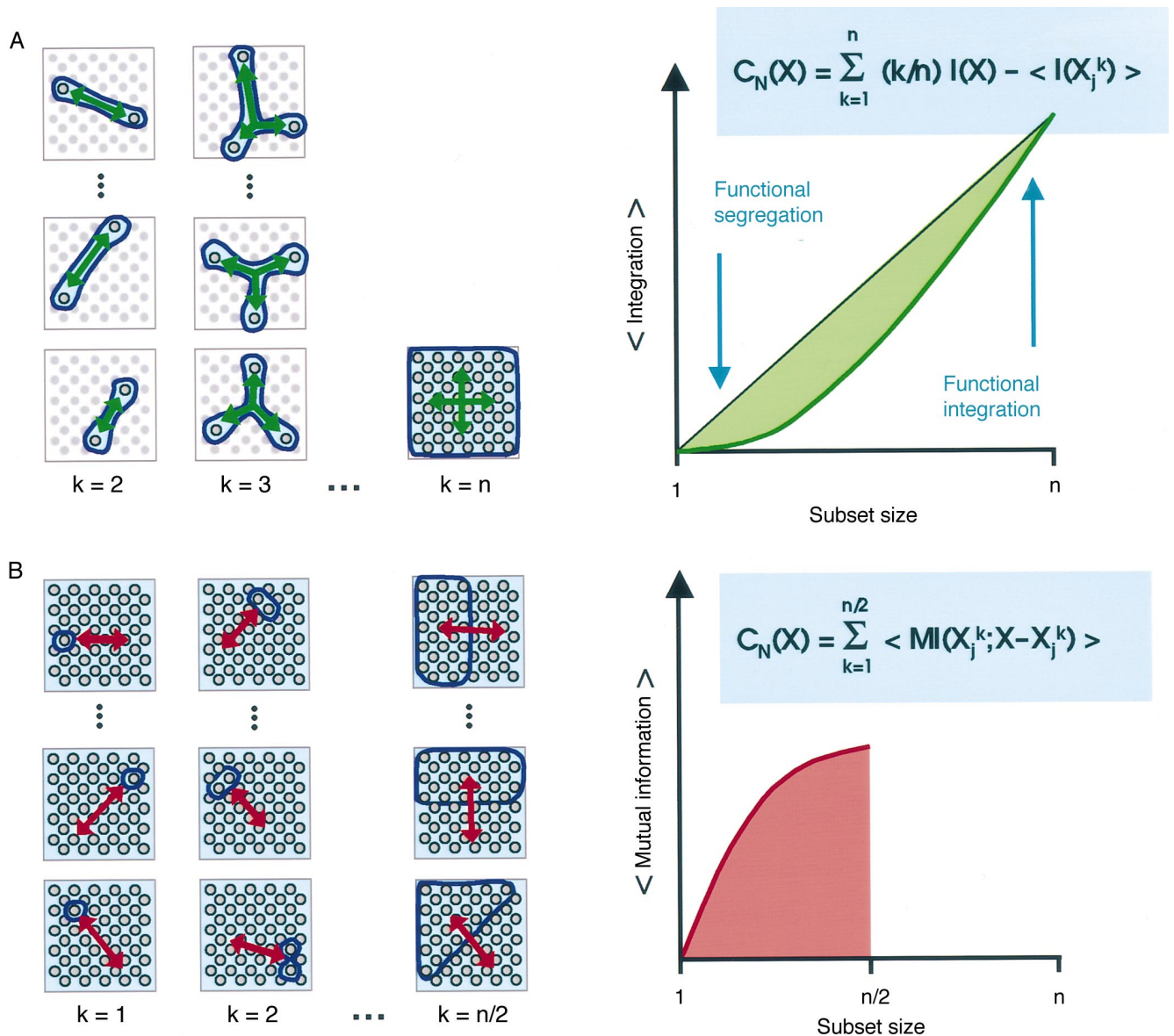


Fig. 2 Two equivalent ways of defining neural complexity. (A) Neural complexity defined in terms of integration, utilizing the ensemble average ($\langle \rangle$) of integration values for subsets composed of increasing numbers of neural elements (subset size). On the left, we show examples of subsets (indicated in blue) of size $k=2$, $k=3$ and $k=n$, with green arrows indicating statistical dependence (integration). The diagram on the right shows the distribution of the mean integration for subset sizes $k=1 \dots n$ (green curve; note that $I(1) = 0$). Neural complexity is given by the area (light green) between the linear increase of integration and the actual average integration over all subset sizes. (B) Neural complexity defined in terms of mutual information, utilizing the ensemble average of mutual integration values between subsets of a given size and their complement (corresponding to bipartitions of the neural system), summed over all subset sizes. On the left, individual examples for subsets of sizes $k=1$, $k=2$, and $k=n/2$ are shown, with red arrows indicating mutual information. The diagram on the right shows the average mutual information for subset sizes $k=1 \dots n/2$ (red curve). Neural complexity is given by the area (light red) under the curve. A complexity measure that does not involve the calculation of average values of integration and mutual information can also be defined. This quantity, which is called 'interaction complexity' or CI and is related but not identical to neural complexity, measures the amount of the entropy of a system that is accounted for by the interactions among its elements. It is given by:

$$H(X) - \sum H(X_j^1 | X - X_j^1) = \sum MI(X_j^1; X - X_j^1) - I(X),$$

where $H(X)$ is the system entropy and

$$H(X_j^1 | X - X_j^1)$$

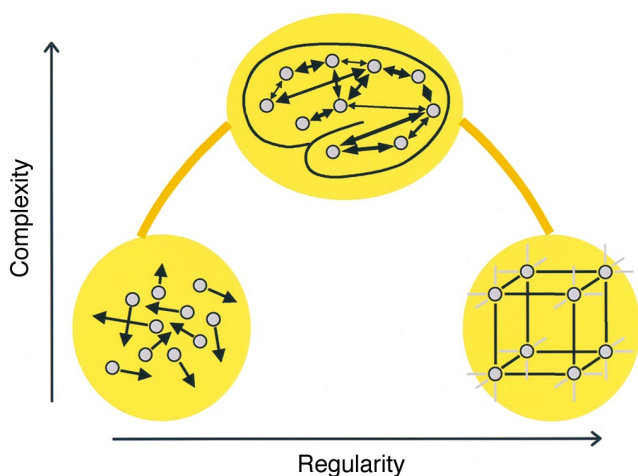
is the entropy of each element conditioned by the entropy of the rest of the system. Note that complexity measures should be applied to a single system (i.e. to a functional cluster) and not to a collection of independent or nearly independent subsystems.

specialized functions: they provide the system with different sources of information. On the other hand, if the same system shows cooperative behavior at the global level, the average integration for large subsets is high. This indicates that different sources of information are being integrated into a coherent whole. Based on the curves shown in Fig. 2A, it becomes feasible to determine quantitatively the coexistence

of functional specialization and integration by defining a measure called neural complexity. It can be seen from the figure that complexity (corresponding to the shaded area in the figure) is high only if a system is both highly integrated and specialized. Figure 2B illustrates also that complexity is mathematically equivalent to the average information exchanged between subsets of a neural system and the rest of

Box 2. Different kinds of complexity

According to the Oxford English Dictionary, something is complex when it constitutes 'a whole... comprehending various parts united or connected together', especially 'parts or elements not simply coordinated, but involved in various degrees of subordination'. While we think that we recognize complexity when we see it, complexity is an attribute that is often employed generically without any attempt at conceptual clarity or, even less, quantification. Recently, scientific approaches to complexity have attempted to retain the intuitive, common sense notion of complexity by emphasizing the idea that complex systems are neither completely regular nor completely random. For example, neither a random string nor a periodically repeating string of letters is complex, while a string of English text certainly is. More generally, any system of elements arranged at random (e.g. gas molecules) or in a completely regular or homogeneous way (molecules in a crystal lattice) is not complex. By contrast, the arrangement and interactions of neurons in a brain or of molecules in a cell is obviously extremely complex (see Fig.).



A number of complexity measures have been proposed, but only a few satisfy the requirement of attaining small values for both completely random and completely regular systems. In neurobiology, for example, one often encounters the term 'dimensional complexity' or just 'complexity' referring to the so-called correlation dimension of EEG signals^a. Its value appears to increase, for instance, from sleep to waking states, or with brain maturation^{b,c}. The correlation dimension is a measure developed in the context of nonlinear dynamics, which should be proportional, roughly speaking, to the number of independent neuronal populations giving rise to an EEG signal^d. But because the correlation dimension would be higher for complete independence than for the mixture of functional segregation and integration that characterizes brain dynamics, it violates the criterion for complexity mentioned above.

Complexity measures have been proposed in the context of algorithmic information theory, which deals with the information necessary to generate individual bit strings. For example, the well-known algorithmic (or

Kolmogorov) complexity is defined as the length of the shortest computer program that generates a particular bit string^e. While this measure is appropriately low for completely regular strings, it is highest for random strings, and thus it too does not satisfy the above criterion for complexity.

Attempts at modifying the notion of algorithmic complexity in order to capture 'true complexity' have recently been proposed^{f,g}. The key idea is to discount pure randomness or noise and measure complexity by the shortest computer program capable of describing the remaining regularities. By definition, such a measure would be satisfactorily low both for random and trivially regular strings, but would be high for systems incorporating a large number of regularities that cannot be further reduced. Of course, insofar as the notion remains algorithmic, it requires that the observer can distinguish between what represents genuine organization and what is instead randomness or noise, and this may be difficult or even impossible. The length of the description of the regularities is also highly dependent on the understanding of the observer. Thus, a system that appears highly complex or random might turn out to be considerably simpler once the organizing principles are understood. Low-dimensional chaotic systems, for example, might appear random, yet their behavior can be fully determined by as few as three equations.

The definitions of complexity considered in this review (see Fig. 2, main article) are statistical measures that capture regularities based on the deviation from independence (mutual information) among subsets of a system^h. In this way, noise can be distinguished from genuine regularities in a way that is relatively independent of an observer's understanding of the system's organization. A degree of subjectivity remains, of course, in deciding which variables to measure and in choosing the appropriate level of coarse-graining for averaging. However, it is easy to show that these measures satisfy the requirement of being low both for completely random and for trivially regular (homogeneous) systems^h.

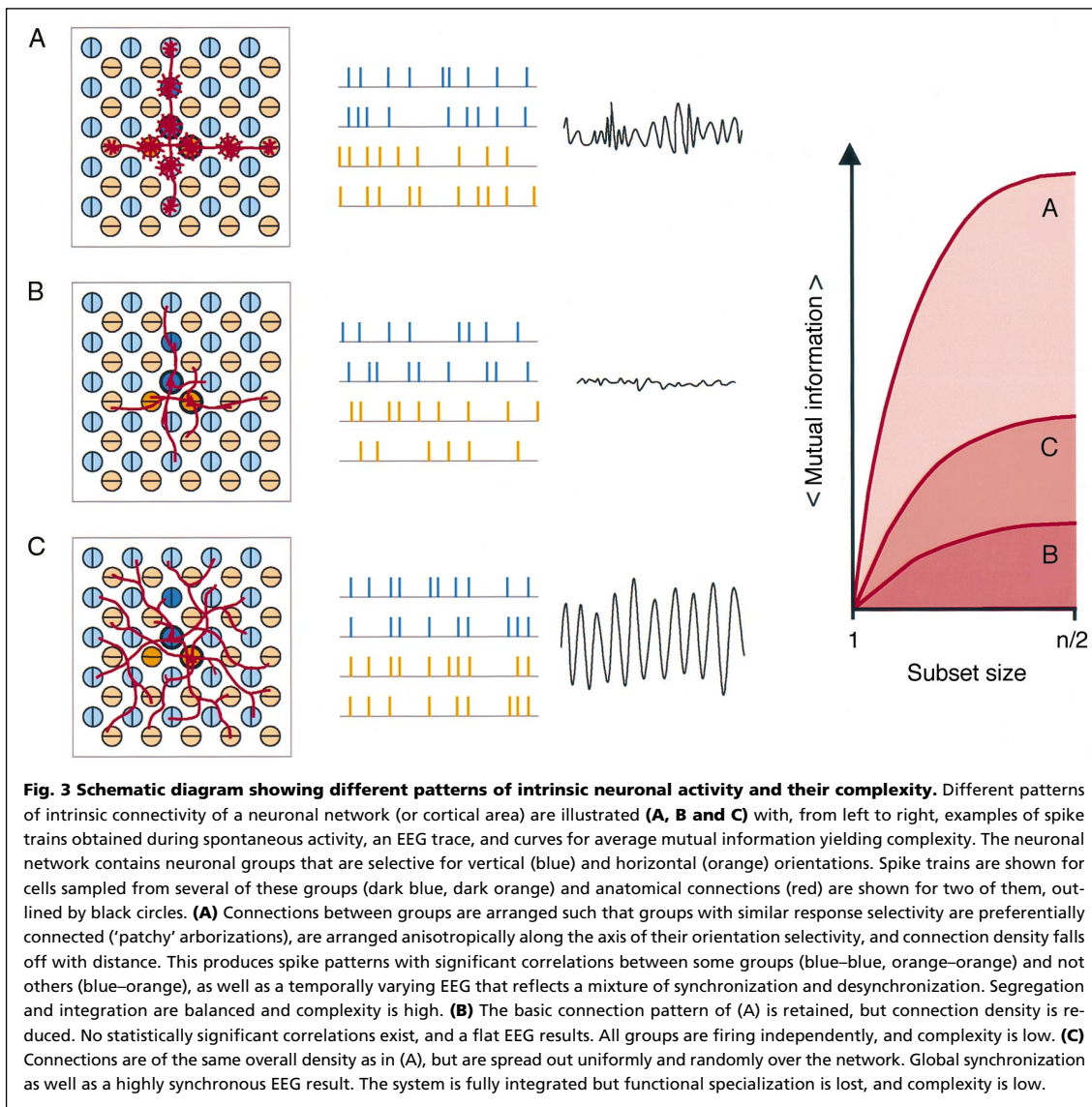
References

- a Babloyantz, A., Salazar, J.M. and Nicolis, C. (1985) Evidence of chaotic dynamics of brain activity during the sleep cycle *Phys. Lett. (A)* 111, 152–156
- b Anokhin, A.P. et al. (1996) Age increases brain complexity *Electroencephalogr. Clin. Neurophysiol.* 99, 63–68
- c Meyer-Lindenberg, A. (1996) The evolution of complexity in human brain development: an EEG study *Electroencephalogr. Clin. Neurophysiol.* 99, 405–411
- d Lutzenberger, W., Preissl, H. and Pulvermüller, F. (1995) Fractal dimension of electroencephalographic time series and underlying brain processes *Biol. Cybern.* 73, 477–482
- e Kolmogorov, A.N. (1965) Three approaches to the quantitative definition of information *Inf. Trans.* 1, 3–11
- f Crutchfield, J.P. and Young, K. (1989) Inferring statistical complexity *Phys. Rev. Lett.* 63, 105–108
- g Gell-Mann, M. and Lloyd, S. (1996) Information measures, effective complexity, and total information *Complexity* 2, 44
- h Tononi, G., Sporns, O. and Edelman, G.M. (1994) A measure for brain complexity: relating functional segregation and integration in the nervous system *Proc. Natl. Acad. Sci. U. S. A.* 91, 5033–5037

the system, summed over all subset sizes. Thus, complexity provides a measure for the amount of information that is integrated within a neural system (for a discussion of complexity measures, see Box 2).

A schematic illustration of the notion of complexity, based on the results of large-scale computer simulations^{45,46} is given in Fig. 3 (for details see Fig. 3 legend). Complexity is evaluated for the spontaneous activity of three simulated examples of a cortical area that differ in the anatomical pattern of their intra-areal (voltage-dependent) re-entrant connections. Figure 3A shows the activity patterns that emerge when neuronal groups are connected to each other accord-

ing to anatomical rules (i.e. specificity^{47–49}, anisotropy⁵⁰, and fall-off with distance) derived from the actual organization of primary visual cortex. The spontaneous dynamic behavior of this system is complex: neighboring neurons of similar orientation preference tend to fire synchronously more often than neurons belonging to functionally unrelated groups, in agreement with the Gestalt laws of similarity and continuity. From one moment to the next, however, the particular subsets of neuronal groups that are firing together changes, so that a large number of coherent patterns is continuously generated. This results in a calculated electroencephalogram (EEG) that shows waxing and waning



and ever-changing episodes of transient synchrony, resembling the one seen during waking or rapid-eye-movement (REM) sleep. If one measures the average integration for increasing subset sizes, this system turns out to be functionally integrated and segregated at the same time. Its complexity is, accordingly, quite high. If we decrease the density of intra-areal connections drastically (Fig. 3B), individual groups of neurons fire more independently, resulting in a desynchronized EEG. The system is not integrated and its complexity is low. Finally, if every neuronal group is connected to many other neuronal groups in a uniform way (Fig. 3C), most of them tend to fire synchronously, irrespective of their response properties. The calculated EEG is hypersynchronous, resembling the high voltage waves of slow-wave sleep or of generalized epilepsy. The system is thus highly integrated but, because all groups of neurons behave in essentially the same way, functional specialization is completely lost. Correspondingly, the integrated information is low and so is complexity.

Other simulations⁴⁵ have shown that complexity is associated with high density of connections, strong local connectivity helping to link cells into neuronal groups, patchiness in the connectivity among neuronal groups, and large numbers of short re-entrant circuits, all of which are funda-

mental organizational principles of the cerebral cortex. In general, systems that are composed of functionally specialized but highly interactive elements (e.g. the brain) will attain high values of complexity. By contrast, systems composed of elements that either lack integration (e.g. a gas) or specialization (e.g. a homogeneous crystal) will have minimal complexity (see Box 2 Fig.).

Once complexity is defined in information-theoretical terms, it is possible to apply it to data obtained from neurophysiological experiments⁵¹ and to test several experimental predictions. For example, based on large-scale computer simulations of the thalamocortical system, one would predict that, even though the anatomical connectivity remains unchanged, major changes in complexity should accompany changes in firing mode brought about by neuromodulators such as acetylcholine. A simple test of this prediction would be a comparison of complexity values obtained during waking or REM sleep, when acetylcholine release is high, versus slow-wave sleep, when it is low⁵², in line with the observation that during slow-wave sleep the integration of cognitive information is considerably reduced⁵³. One would also predict that disorders characterized by a diffuse impairment of connectivity should be associated with reduced complexity.

The role of spontaneous activity

The simulations just discussed demonstrate that, in neural systems incorporating connectivity rules found in cortical areas, spontaneous activity leads to intrinsic patterns of correlations that are far from random. Experimentally, patterned spontaneous activity is indeed encountered in a great variety of neural systems from the earliest stages of development through to adulthood. For example, neurons in the developing retina as well as in the cortex undergo spontaneous patterned discharges which play an important role in shaping adult connectivity^{54,55}. Most remarkably, while still *in utero* the brain spends many hours a day in a form of sleep, called active sleep, which is characterized by self-generated, spontaneous activity in most brain structures⁵⁶.

Spontaneous activity continues to be a fundamental feature of the adult brain. When an animal is not engaged in any particular task, most neurons are found to fire at a spontaneous level that ranges from 0.5 to 10–15 Hz (Ref. 57). Even neurons in isolated cortical 'islands' continue to fire in the absence of any extrinsic input⁵⁸. The EEG recorded when a subject is resting with the eyes closed and in the absence of stimulation also reveals ongoing brain activity that is spontaneously changing. Studies using real-time optical imaging in visual cortex of cats indicate that the great variability of evoked responses to stimuli results largely from the dynamics of ongoing activity. Indeed, the effect of stimuli has been likened to 'ripples caused by tossing a stone in a wavy sea'⁵⁹. Finally, imaging studies in humans have revealed that stimulus-induced changes in blood flow, which are assumed to reflect changes in synaptic activity, represent only a small percentage of resting blood flow⁶⁰.

Perhaps the most striking demonstration of internally generated neural and cognitive activity is given by REM sleep. It is well known that, during REM sleep, the overall mode of activity in the thalamocortical system, both at the EEG and at the single unit level, resembles that of waking^{61,62}. Despite certain cognitive peculiarities of dreams, such as suspension of disbelief, singlemindedness and loss of self-reflectiveness⁶³, which may be related to recently demonstrated differences in regional activation⁶¹, the structural similarity between dreaming and waking mentation is remarkable. For example, visual objects and scenes are usually recognizable, Gestalt laws apply as much in waking as they apply in dreaming, integrated multimodal scenes are structurally indistinguishable from those experienced during waking, language is intelligible, and even the narrative structure of dreams is highly coherent⁶⁴. Longitudinal studies have also shown that the development of the cognitive structure of dreams precisely parallels the development of cognitive abilities of the waking infant⁵³. Indeed, it appears that the structural characteristics that dreams share with waking cognition reflect those regularities in the environment that have been incorporated in the brain's anatomical connectivity. By contrast, the seemingly arbitrary associations that often occur in dreams may indicate which aspects of knowledge can be flexibly recombined during spontaneous activity.

These observations suggest that a spontaneous exchange of re-entrant signals among functionally segregated groups of neurons takes place all the time, whether the brain is responding to extrinsic inputs or not⁶⁵. They also suggest that a con-

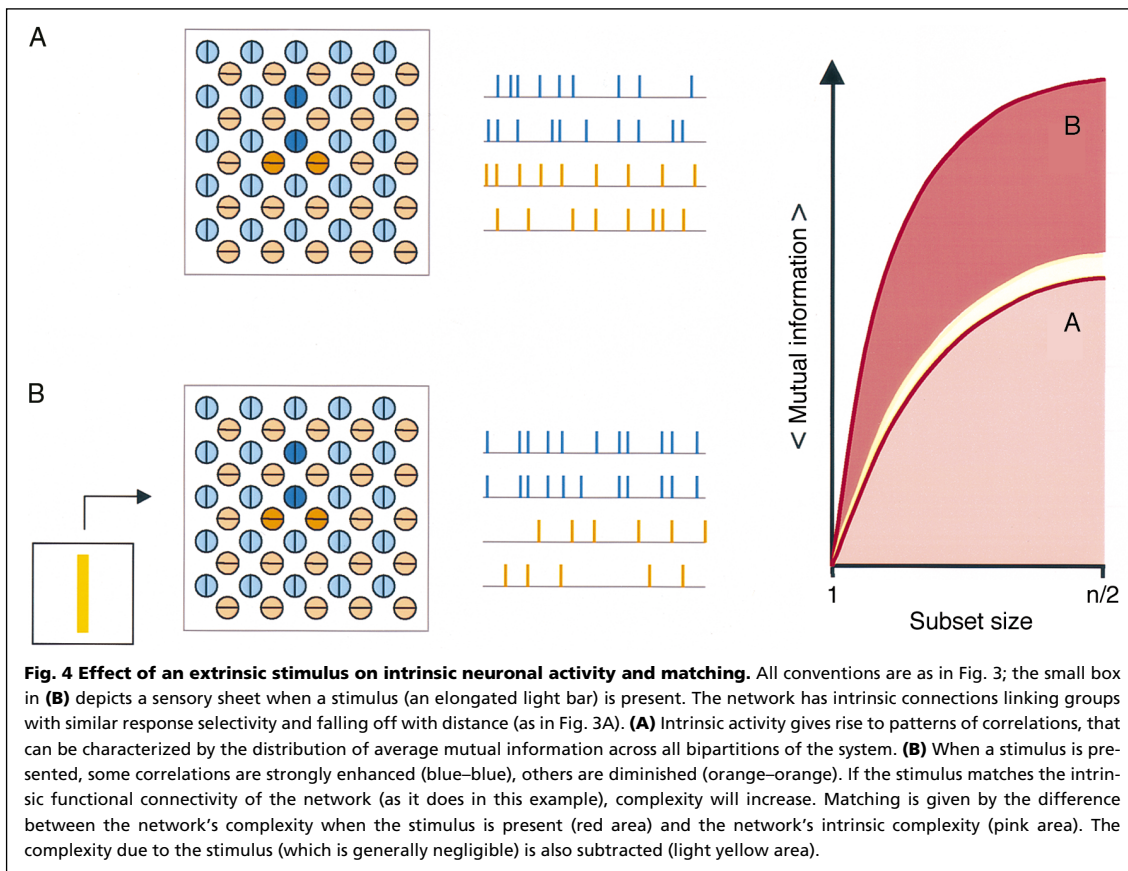
siderable fraction of intrinsic signals do not constitute noise, but rather express a set of functional relationships among neuronal groups that have been selected over evolution, development, and individual experience to reflect statistical regularities in the environment⁶⁶. Based on the simulations mentioned in the previous section, for instance, one would predict that spontaneous firing patterns in visual cortex should reflect the Gestalt principles of similarity, continuity and proximity. Although direct experimental confirmation for this prediction is not yet available, supporting evidence in a different context has recently been obtained, through multi-unit studies of hippocampal place cells in the rat⁶⁷. When an awake rat navigates in its environment, hippocampal cells that have corresponding place fields are activated together. Later, the same cells begin to fire together spontaneously during sleep even if the animal has been removed from that environment. These studies provide experimental support for the notion that the brain incorporates statistical relationships sampled in the environment in terms of its functional connectivity – the pattern of correlations among different brain regions – and that this functional connectivity is spontaneously reactivated during sleep (see Stickgold, this issue pp. 484–492).

Reconciling information processing and information storage: matching complexity

If the brain is spontaneously active, and if such spontaneous activity consists in considerable measure of organized patterns of correlations among distributed groups of neurons, it is worth considering how ongoing signaling in the brain is affected when a stimulus is encountered. How is the external stimulus integrated with the intrinsic functional states of the nervous system?

To answer this question, it is useful to resort again to the results of some computer simulations, which are presented in schematic form in Fig. 4. As was mentioned above, spontaneous activity in a simulated cortical area produces intrinsic correlation patterns that largely reflect its internal connectivity, which incorporates statistical regularities in the environment (Fig. 4A). For example, neurons sharing the same orientation preference tend to emit coincident spikes. When a stimulus consistent with that environment is presented, such as an elongated light bar, the number of coincident spikes between neighboring neurons of matching orientation increases significantly (Fig. 4B). The increase is fast and effective because groups of neurons linked by cortico-cortical and thalamocortical re-entrant loops are kept in a ready-to-fire mode by the opening of voltage-dependent channels through spontaneous activity^{46,68}. In the absence of spontaneous activity, the triggering of coincident spikes is much slower and less effective. Most importantly, changes in the strength of correlated firing occur not only for a particular couple of neurons, but for many subsets of neurons distributed throughout the simulated neural system.

How can such a widespread change in the spontaneously occurring functional connectivity of a neural system be understood in information-theoretical terms? As we have seen, the amount of information exchanged between all subsets of a neural system (their average mutual information) is measured by neural complexity. It was recently shown that



the change in complexity when such a system encounters extrinsic stimuli reflects the degree to which intrinsic statistical relationships between its constituent neuronal groups match statistical relationships present in the environment⁶⁹. For this reason, this increase in complexity has been called matching complexity (see Fig. 5A). It is important to note that high values of matching do not imply any simple isomorphism between the functional connectivity of a neural system and the statistical regularities of the input stimuli. On the other hand, matching is zero if a system has random connectivity (which implies low complexity) or when the input contains no statistical regularities (i.e. it is just noise). A low value of matching between a structured stimulus and a system of high complexity signals statistical ‘novelty’⁶⁹.

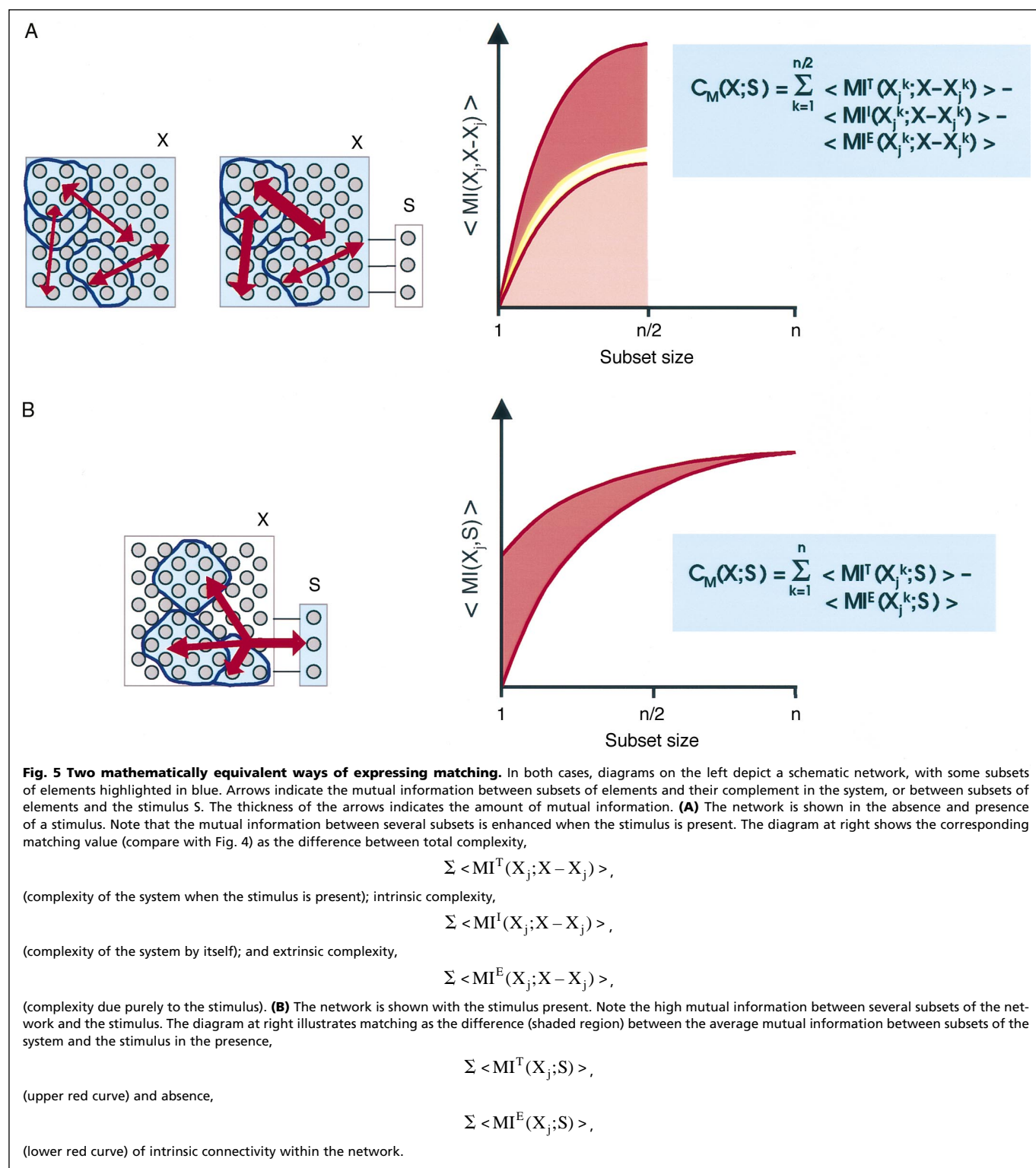
A mathematically equivalent way of expressing matching is as a measure of how well the connectivity distributes information from the stimulus to all subsets of a neural system (Fig. 5B). This equivalence is important because it relates the mutual information between the input and the neural system to changes in the mutual information within the system. Thus, just as complexity captures both functional segregation and integration, matching accommodates extrinsic and intrinsic sources of information within the same theoretical framework.

Two corollaries of these observations are worth mentioning. First, if complexity and matching are both high, even for a small value of the extrinsic mutual information between an individual stimulus and the brain there will be a much larger increase in the intrinsic mutual information among subsets of units within the brain. This means that, by relying on its intrinsic functional connectivity, the brain

can literally go ‘beyond the information given’⁷⁰. Second, the same stimulus can convey radically different amounts of information (as reflected in matching complexity) depending upon the functional connectivity of the system that receives it. This theoretical conclusion is consistent with an everyday observation: the same stimulus, say, a Chinese

Outstanding questions

- Can one find solid neurophysiological evidence for functional clustering occurring at the timescale of hundreds of milliseconds? And are such functional clusters limited to the thalamocortical system?
- It is almost impossible to do two things at once. Even simple choices cannot be made simultaneously, but require an interval of at least 150 milliseconds (the so-called psychological refractory period). What are the underlying neural mechanisms of this? Are they related to the mechanisms of conscious integration?
- In the brain, and particularly in the cerebral cortex, any group of neurons is potentially connected to many other groups of neurons. What prevents synchronization from spreading to the entire brain?
- Can the concept of complexity usefully be extended to temporal sequences, including spike patterns? In particular, can one show that a completely random or a completely periodic sequence is not complex, while a sequence that contains many different kinds of regularities is complex?
- What is the relative importance of cortico-cortical and thalamo-cortical mechanisms of integration?
- What accounts for the relative functional disconnection from extrinsic inputs that occurs during dreaming?
- Does the extent of functional integration in the brain change between automatic and controlled tasks? For example, are the neural processes underlying automatic tasks functionally insulated (involving a small set of brain regions)? By contrast, do controlled tasks require the integration of a large set of neuronal groups that are widely distributed?



character, can be meaningful to Chinese speakers and meaningless to English speakers even though the extrinsic information conveyed to the retina must be the same. Unlike the usual information-processing approaches, the concept of matching clarifies this paradox.

These theoretical observations lead to several experimental predictions. For example, meaningful, association-rich stimuli should lead to an increase in the mutual information between many subsets of neuronal groups (i.e. matching should be high), while meaningless stimuli should not (i.e. matching should be low). For example, words and

non-words should be associated with different values of matching. And specifically, there should be a positive change in matching when a hidden figure suddenly emerges from a noisy background, as in random-dot stereograms (see Ref. 71). In neural terms, this increase in mutual information should go along with an increase in correlated activity along specific cortico-cortical and cortico-thalamic re-entrant loops.

Conclusion

Concepts such as functional clustering, complexity, and matching, which are based on the statistical foundations of

information theory, provide a unified framework for conceptualizing the integration of information in the brain as a selectionist system⁷². This includes both the integration of the activity of functionally segregated groups of neurons as well as the integration of incoming stimuli with ongoing, spontaneous brain activity. The availability of measures that can be applied to actual neural processes leads to experimental predictions that, at least in principle, can be tested with neurophysiological and neuroimaging techniques. If these predictions are fulfilled, traditional information-processing approaches that see the brain as a relatively passive device for processing or storing information will have to be augmented by theoretical approaches, of which the one outlined here is an example, that emphasize the intrinsic dynamics of the brain and that view memory and perception as closely related.

Acknowledgements

This work was carried out as part of the theoretical neurobiology program at The Neurosciences Institute, which is supported by Neurosciences Research Foundation. The Foundation receives major support for this program from Novartis Pharmaceutical Corporation.

References

- 1 Zeki, S. (1978) Functional specialization in the visual cortex of the Rhesus monkey *Nature* 274, 423–428
- 2 Colby, C.L. and Duhamel, J.R. (1991) Heterogeneity of extrastriate visual areas and multiple parietal areas in the macaque monkey *Neuropsychologia* 29, 517–537
- 3 Felleman, D.J. and Van Essen, D.C. (1991) Distributed hierarchical processing in the primate cerebral cortex *Cereb. Cortex* 1, 1–47
- 4 Kaas, J.H. and Pons, T.P. (1988) The somatosensory system of primates, in *Comparative Primate Biology, Vol. 5: Neuroscience* (Steklis, H.P., ed.), pp. 421–868, Alan Liss
- 5 Morel, A., Garraghty, P.E. and Kaas, J.H. (1993) Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys *J. Comp. Neurol.* 335, 437–459
- 6 Rizzolatti, G., Luppino, G. and Matelli, M. (1998) The organization of the cortical motor system: new concepts *Electroencephalogr. Clin. Neurophysiol.* 106, 283–296
- 7 Mountcastle, V. (1957) Modality and topographic properties of single neurons of cat's somatic sensory cortex *J. Neurophysiol.* 20, 408–434
- 8 Hubel, D.H. and Wiesel, T.N. (1968) Receptive fields and functional architecture of monkey striate cortex *J. Physiol.* 195, 215–243
- 9 Bartfeld, E. and Grinvald, A. (1992) Relationships between orientation-preference pinwheels, cytochrome oxidase blobs, and ocular-dominance columns in primate striate cortex *Proc. Natl. Acad. Sci. U. S. A.* 89, 11905–11909
- 10 Weliky, M., Bosking, W.H. and Fitzpatrick, D. (1996) A systematic map of direction preference in primary visual cortex *Nature* 379, 725–728
- 11 Shoham, D. et al. (1997) Spatio-temporal frequency domains and their relation to cytochrome oxidase staining in cat visual cortex *Nature* 385, 529–533
- 12 Albright, T.D., Desimone, R. and Gross, C.G. (1984) Columnar organization of directionally selective cells in visual area MT of the macaque *J. Neurophysiol.* 51, 16–31
- 13 Fujita, I., Tanaka, K., Ito, M. and Cheng, K. (1992) Columns for visual features of objects in monkey inferotemporal cortex *Nature* 360, 343–346
- 14 Frackowiak, R.S.J. et al. (1997) *Human Brain Function*, Academic Press
- 15 Koffka, K. (1935) *Principles of Gestalt Psychology*, Harcourt Brace
- 16 Kanizsa, G. (1979) *Organization in Vision*, Praeger
- 17 Treisman, A. (1996) The binding problem *Curr. Opin. Neurobiol.* 6, 171–178
- 18 Tononi, G. and Edelman, G.M. Consciousness and complexity *Science* (in press)
- 19 Barlow, H.B. (1972) Single units and sensation: a neuron doctrine for perceptual psychology? *Perception* 1, 371–394
- 20 Rolls, E.T. (1992) Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas *Philos. Trans. R. Soc. London Ser. B* 335, 11–20
- 21 von der Malsburg, C. (1995) Binding in models of perception and brain function *Curr. Opin. Neurobiol.* 5, 520–526
- 22 Edelman, G.M. (1987) *Neural Darwinism: The Theory of Neuronal Group Selection*, Basic Books
- 23 Edelman, G.M. (1989) *The Remembered Present: A Biological Theory of Consciousness*, Basic Books
- 24 Sporns, O. et al. (1989) Re-entrant signaling among simulated neuronal groups leads to coherency in their oscillatory activity *Proc. Natl. Acad. Sci. U. S. A.* 86, 7265–7269
- 25 Sporns, O., Tononi, G. and Edelman, G.M. (1991) Modeling perceptual grouping and figure-ground segregation by means of active re-entrant connections *Proc. Natl. Acad. Sci. U. S. A.* 88, 129–133
- 26 Tononi, G., Sporns, O. and Edelman, G.M. (1992) Re-entry and the problem of integrating multiple cortical areas: simulation of dynamic integration in the visual system *Cereb. Cortex* 2, 310–335
- 27 Bressler, S.L. (1995) Large-scale cortical networks and cognition *Brain Res. Rev.* 20, 288–304
- 28 Singer, W. and Gray, C.M. (1995) Visual feature integration and the temporal correlation hypothesis *Annu. Rev. Neurosci.* 18, 555–586
- 29 Gray, C.M. et al. (1989) Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties *Nature* 338, 334–337
- 30 Engel, A.K. et al. (1991) Interhemispheric synchronization of oscillatory neuronal responses in cat visual cortex *Science* 252, 1177–1179
- 31 Bressler, S.L., Coppola, R. and Nakamura, R. (1993) Episodic multiregional cortical coherence at multiple frequencies during visual task performance *Nature* 366, 153–156
- 32 Murthy, V.N. and Fetz, E.E. (1996) Synchronization of neurons during local field potential oscillations in sensorimotor cortex of awake monkeys *J. Neurophysiol.* 76, 3968–3982
- 33 Riehle, A. et al. (1997) Spike synchronization and rate modulation differentially involved in motor cortical function *Science* 278, 1950–1953
- 34 Donoghue, J.P. et al. (1998) Neural discharge and local field potential oscillations in primate motor cortex during voluntary movements *J. Neurophysiol.* 79, 159–173
- 35 Singer, W. et al. (1997) Neuronal assemblies: necessity, signature and detectability *Trends Cognit. Sci.* 1, 252–261
- 36 Posner, M.I. and Raichle, M.E. (1994) *Images of Mind*, Scientific American Library
- 37 McLaughlin, T. et al. (1992) Potential language and attentional networks revealed through factor analysis of rCBF data measured with SPECT *J. Cereb. Blood Flow Metab.* 12, 535–545
- 38 Moeller, J.R. et al. (1987) Scaled subprofile model: a statistical approach to the analysis of functional patterns in positron emission tomographic data *J. Cereb. Blood Flow Metab.* 7, 649–658
- 39 McIntosh, A.R. et al. (1994) Network analysis of cortical visual pathways mapped with PET *J. Neurosci.* 14, 655–666
- 40 McIntosh, A.R. et al. (1996) Changes in limbic and prefrontal functional interactions in a working memory task for faces *Cereb. Cortex* 6, 571–584
- 41 Friston, K.J. et al. (1996) Functional topography: multidimensional scaling and functional connectivity in the brain *Cereb. Cortex* 6, 156–164
- 42 Büchel, C. and Friston, K.J. (1997) Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI *Cereb. Cortex* 7, 768–778
- 43 Tononi, G., McIntosh, A.R., Russell, D.P. and Edelman, G.M. (1998) Functional clustering: identifying strongly interactive brain regions in neuroimaging data *Neuroimage* 7, 133–149
- 44 Tononi, G. and Edelman, G.M. (1998) Consciousness and the integration of information in the brain, in *Consciousness* (Jasper, H.E.A., ed.), Plenum Press
- 45 Tononi, G., Sporns, O. and Edelman, G.M. (1994) A measure for brain complexity: relating functional segregation and integration in the nervous system *Proc. Natl. Acad. Sci. U. S. A.* 91, 5033–5037
- 46 Lumer, E.D., Edelman, G.M. and Tononi, G. (1997) Neural dynamics in a model of the thalamocortical system. I. Layers, loops and the emergence of fast synchronous rhythms *Cereb. Cortex* 7, 207–227

- 47 Gilbert, C.D. and Wiesel, T.N. (1989) Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex *J. Neurosci.* 9, 2432–2442
- 48 Malach, R. et al. (1993) Relationship between intrinsic connections and functional architecture revealed by optical imaging and *in vivo* targeted biocytin injections in primate striate cortex *Proc. Natl. Acad. Sci. U. S. A.* 90, 10469–10473
- 49 Weliky, M. and Katz, L.C. (1994) Functional mapping of horizontal connections in developing ferret visual cortex: experiments and modeling *J. Neurosci.* 14, 7291–7305
- 50 Schmidt, K.E. et al. (1997) The perceptual grouping criterion of colinearity is reflected by anisotropies of connections in the primary visual cortex *Eur. J. Neurosci.* 9, 1083–1089
- 51 Friston, K.J. et al. (1995) Characterising the complexity of neuronal interactions *Hum. Brain Mapp.* 3, 302–314
- 52 Marrosu, F. et al. (1995) Microdialysis measurement of cortical and hippocampal acetylcholine release during sleep–wake cycle in freely moving cats *Brain Res.* 671, 329–332
- 53 Foulkes, D. (1985) *Dreaming: A Cognitive–Psychological Analysis*, Lawrence Erlbaum
- 54 Brivanlou, I.H., Warland, D.K. and Meister, M. (1998) Mechanisms of concerted firing among retinal ganglion cells *Neuron* 20, 527–539
- 55 Penn, A.A. et al. (1998) Competition in retinogeniculate patterning driven by spontaneous activity *Science* 279, 2108–2112
- 56 Roffwarg, H.P., Muzio, J.N. and Dement, W.C. (1966) Ontogenic development of the human sleep–dream cycle *Science* 152, 614–619
- 57 Eggermont, J.J. (1990) *The Correlative Brain: Theory and Experiment in Neural Interaction*, Springer-Verlag
- 58 Burns, B.D., Stean, J.P. and Webb, A.C. (1979) The effects of sleep on neurons in isolated cerebral cortex *Proc. R. Soc. London Ser. B* 206, 281–291
- 59 Arieli, A. et al. (1996) Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses *Science* 273, 1868–1871
- 60 Roland, P.E. (1993) *Brain Activation*, Wiley–Liss
- 61 Hobson, J.A., Stickgold, R. and Pace-Schott, E.F. (1998) The neuropsychology of REM sleep dreaming *NeuroReport* 9, R1–14
- 62 Steriade, M. (1997) Synchronized activities of coupled oscillators in the cerebral cortex and thalamus at different levels of vigilance *Cereb. Cortex* 7, 583–604
- 63 Rechtschaffen, A. (1978) The single-mindedness and isolation of dreams *Sleep* 1, 97–109
- 64 Foulkes, D. (1982) *Children's Dreams: Longitudinal Studies*, Wiley
- 65 Llinas, R.R. and Ribary, U. (1994) Perception as an oneiric-like state modulated by the senses, in *Large-scale Neuronal Theories of the Brain* (Koch, C. and Davis, J.L., eds), pp. 111–124, MIT Press
- 66 Lowel, S. and Singer, W. (1992) Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity *Science* 255, 209–212
- 67 Wilson, M.A. and McNaughton, B.L. (1994) Reactivation of hippocampal ensemble memories during sleep *Science* 265, 676–679
- 68 Lumer, E.D., Edelman, G.M. and Tononi, G. (1997) Neural dynamics in a model of the thalamocortical system: II. The role of neural synchrony tested through perturbations of spike timing *Cereb. Cortex* 7, 228–236
- 69 Tononi, G., Sporns, O. and Edelman, G.M. (1996) A complexity measure for selective matching of signals by the brain *Proc. Natl. Acad. Sci. U. S. A.* 93, 3422–3427
- 70 Bruner, J.S. (1973) *Beyond the Information Given: Studies in the Psychology of Knowing*, W.W. Norton
- 71 Gaetz, M., Weinberg, H., Rzepoluck, E. and Jantzen, K.J. (1998) Neural network classifications and correlation analysis of EEG and MEG activity accompanying spontaneous reversals of the Necker cube *Cognit. Brain Res.* 6, 335–346
- 72 Sporns, O. and Tononi, G., eds (1994) *Selectionism and the Brain*, Academic Press

Sleep: off-line memory reprocessing

Robert Stickgold

Behavioral studies of memory and learning in both humans and animals support a role for sleep in the consolidation and integration of memories. Physiological studies of hippocampal and cortical activity as well as of brainstem neuromodulatory systems demonstrate the state-dependence of communication both between and within the neocortex and hippocampus. These findings are consonant with observed cognition during sleep and immediately following awakening.

The study of the relationship between sleep and memory has had a long and complex history. Two hundred years ago, David Hartley proposed that dreaming altered the strength of associative links in memory¹. But the first systematic studies of sleep and memory were only carried out in 1924, when Jenkins and Dallenbach² used sleep studies to test Ebbinghaus' theory of memory decay³. They demonstrated that recall was diminished less by a night of sleep than an equivalent amount of wake time, and concluded that memory loss normally resulted from sensory interference rather than passive decay; no active role of sleep was considered.

With the discovery of rapid eye movement sleep (REM) in 1953 (Ref. 4), the question of sleep and memory re-emerged. Still, twenty years later Greenberg and Pearlman⁵ found no general acceptance of the hypothesis that REM was involved in information processing, despite numerous reports supporting such a role. Since then possible roles of sleep in strengthening, integrating, and even erasing memories^{6,7} have been continually debated. In what follows I will review evidence from (1) behavioral studies of memory and learning, (2) physiological studies of hippocampal activity and brainstem neuromodulatory systems, and (3) neural network models of information storage and

R. Stickgold is at the Laboratory of Neurophysiology, Department of Psychiatry, Harvard Medical School, 74 Fenwood Road, Boston, MA 02115, USA.

tel: +1 617 734 1300
ext. 316
fax: +1 617 734 7851
e-mail: rstickgold@hms.harvard.edu