Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 3422–3427, April 1996
Neurobiology

A complexity measure for selective matching of signals by the brain

(correlation/entropy/mutual information/neuronal group selection/reentry)

GIULIO TONONI, OLAF SPORNS, AND GERALD M. EDELMAN

The Neurosciences Institute, 10640 John J. Hopkins Drive, San Diego, CA 92121

Contributed by Gerald M. Edelman, December 28, 1995

ABSTRACT **We have previously derived a theoretical measure of neural complexity (C_N) in an attempt to characterize functional connectivity in the brain. C_N measures the amount and heterogeneity of statistical correlations within a neural system in terms of the mutual information between subsets of its units. C_N was initially used to characterize the functional connectivity of a neural system isolated from the environment. In the present paper, we introduce a related statistical measure, matching complexity (C_M), which reflects the change in C_N that occurs after a neural system receives signals from the environment. C_M measures how well the ensemble of intrinsic correlations within a neural system fits the statistical structure of the sensory input. We show that C_M is low when the intrinsic connectivity of a simulated cortical area is randomly organized. Conversely, C_M is high when the intrinsic connectivity is modified so as to differentially amplify those intrinsic correlations that happen to be enhanced by sensory input. When the input is represented by an individual stimulus, a positive value of C_M indicates that the limited mutual information between sensory sheets sampling the stimulus and the rest of the brain triggers a large increase in the mutual information between many functionally specialized subsets within the brain. In this way, a complex brain can deal with context and go "beyond the information given."**

The intricate connectivity that links functionally specialized groups of neurons within and among brain areas is an outstanding characteristic of mammalian brains. Through an ongoing, recursive, and parallel process of signaling called reentry (1), the anatomical connectivity of the brain supports a functional connectivity, a complex pattern of correlations among sets of neuronal groups (2). In a previous paper, we introduced a measure called neural complexity (C_N) that characterizes the functional connectivity of a neural system in terms of the set of statistical correlations among its units (3). It was shown that C_N is low when these units are either completely uncorrelated or completely correlated. Conversely, C_N is high when a neural system displays both functional segregation and integration among its units or, equivalently, when their correlations are both strong and heterogeneous.

In deriving C_N, we considered a neural system in isolation from environmental stimuli and at a given developmental time. A more complete characterization of the functional connectivity of the brain must relate it to the statistical structure of the signals sampled from the environment. Such signals activate specific neuronal populations and, as a result, synaptic connections between them are strengthened or weakened. In the course of development and experience, the fit or matching between the functional connectivity of the brain and the statistical structure of signals sampled from the environment tends to increase progressively through processes of variation and selection mediated at the level of the synapses (1). These processes are particularly well demonstrated by the organization of primary visual areas. Within a visual area, the connectivity is initially organized in a uniform way. During development and experience, it undergoes a selection process such that groups of neurons responding to similar orientations become preferentially connected (4, 5). The resulting functional connectivity, which constitutes a basis for various Gestalt criteria (6, 7), matches the prevalence of extended colinear edges in the retinal image.

In order to characterize the fit or matching to the statistical structure of environmental signals from a more general perspective, we introduce here a statistical measure, called matching complexity (C_M), which reflects the change in C_N observed when a neural system is receiving sensory input. Through computer simulations, we show that when the synaptic connectivity of a simplified cortical area is randomly organized, C_M is low and the functional connectivity does not fit the statistical structure of the sensory input. If, however, the synaptic connectivity is modified and the functional connectivity is altered so that many intrinsic correlations are strongly activated by the input, C_M increases. We also demonstrate that, once a repertoire of intrinsic correlations has been selected that adaptively matches the statistical structure of the sensory input, that repertoire becomes critical to the way in which the brain categorizes individual stimuli. After developing and illustrating the properties of C_M, we consider the contrast between this selectionist approach and standard information processing views, and we discuss its applicability to various experimental paradigms.

Theory

Following a previous paper (3), we consider a neural system X with n units which are taken to represent neuronal groups. We assume that its activity is described by a Gaussian stationary multidimensional stochastic process (8). The joint probability density function describing such a multivariate process, corresponding here to its functional connectivity, can be characterized in terms of entropy and mutual information (*MI*; refs. 8 and 9). *MI* is a general measure of the deviation from independence, or correlation, among many variables. For instance, consider a bipartition of the system X into a jth subset X_j^k composed of k units and its complement $X - X_j^k$. The *MI* between X_j^k and $X - X_j^k$ is given by

$$MI(X_j^k; X - X_j^k) = H(X_j^k + H(X - X_j^k) - H(X)\,, \qquad [1]$$

where $H(X_j^k)$ and $H(X - X_j^k)$ are the entropies of X_j^k and $X - X_j^k$ considered independently, and $H(X)$ is the entropy of the system considered as a whole (joint entropy). *MI* is zero if X_j^k and $X - X_j^k$ are statistically independent, and it is positive otherwise.

Previously (3), we considered a neural system X, characterized by a synaptic connectivity $CON(X)$, that is isolated from the environment (Fig. 1*A*). To thoroughly characterize its functional connectivity, all correlations among the units of the

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

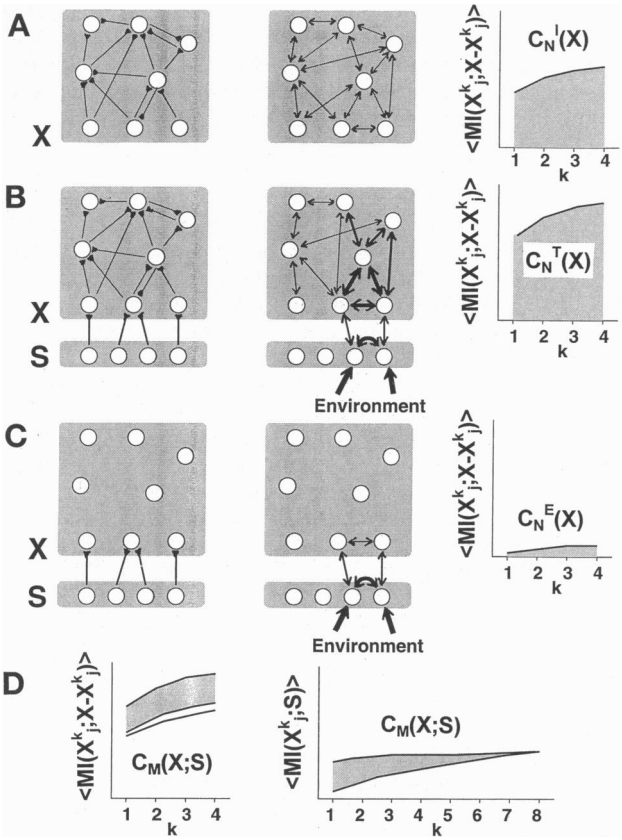

FIG. 1. Schematic illustration of neural complexity (C_N) and matching complexity (C_M). In *A–C* a neural system X is composed of units linked by a matrix of anatomical connections $CON(X)$ (*Left*). This anatomical connectivity gives rise to a functional connectivity $COV(X)$ (*Center*). Correlations are indicated by arrows between the units (thick = strong, thin = weak). (*A*) Isolated system. Intrinsic neuronal activity gives rise to $COV^I(X)$, from which one obtains the system's intrinsic complexity, $C_N^I(X)$ (1). (*Right*) The average mutual information $\langle MI(X_j^k; X - X_j^k)\rangle$ for all bipartitions of the system is plotted versus subset sizes (Eq. **2**); $C_N^I(X)$ is the area under the curve. (*B*) (*Left*) A sensory sheet S is connected to the system X by anatomical connections $CON(S, X)$. (*Center*) Events in the environment cause coactivation of units within S. The functional connectivity of the system is modified by the input: some correlations are enhanced, others are diminished. In this case, the *MI* between some of the subsets increases and there is an overall increase in complexity (*Right*). (*C*) (*Left*) All connections within X are set to 0. Remaining correlations within the system (*Center*) are due to the stimulus *per se* and correspond to the extrinsic complexity $C_N^E(X)$ (*Right*). (*D*) Complexity matching C_M is obtained by subtracting intrinsic and extrinsic complexity from the total complexity (Eq. **3**). The two plots illustrate graphically two equivalent formulations of C_M, in terms of *MI* values between subsets within the system (*Left*; Eq. **4**) and in terms of *MI* values between subsets of the system and the sensory sheet (*Right*; Eq. **6**). Shaded areas under the curves are equal.

system (one-to-many, many-to-many) must be considered. This was done by taking into account the set of all possible bipartitions of the system. The complexity $C_N(X)$ of a neural system is then defined as

$$C_N(X) = \sum_{k=1}^{n/2} \langle MI(X_j^k; X - X_j^k)\rangle\,, \qquad [2]$$

where the ensemble average is taken over all possible bipartitions of size k. According to Eq. **2**, $C_N(X)$ measures the average *MI* for bipartitions of the system over all bipartition sizes. For $C_N(X)$ to be high, two conditions have to be met: the average *MI* between individual units and the rest of the brain must be high, indicating that there are many strong correlations in the system (i.e., the system is integrated). Moreover, the average *MI* must be higher for larger subsets, indicating that such correlations are heterogeneous (i.e., if considering more units adds to *MI*, then individual units must have different, specialized functions).

In this paper, we set out to evaluate how the set of correlations measured by $C_N(X)$ changes when the system samples stimuli from the environment (Fig. 1*B*). We assume that the system X samples a stimulus i through a sensory sheet S and a fixed extrinsic connectivity represented by a matrix $CON(S; X)$ between S and a subset of the system units. S and $CON(S; X)$ are not considered to be part of the system, so that we can express the *MI* between the system and the sensory sheet when it samples the ith stimulus as $MI(X; S_i)$. The complexity $C_N(X)$ observed when the system is isolated is called *intrinsic complexity* $C_N^I(X)$. When the system samples a stimulus through the sensory sheet we observe a *total complexity* $C_N^T(X)$. To evaluate the response of the system, the contribution to $C_N^T(X)$ due to the stimulus *per se* should be discounted. This *extrinsic complexity* $C_N^E(X)$ can be obtained by setting $CON(X) = 0$ (Fig. 1*C*). We can now define the matching complexity $C_M(X; S_i)$ between X and S_i as

$$C_M(X; S_i) = C_N^T(X) - C_N^I(X) - C_N^E(X). \qquad [3]$$

Thus, $C_M(X; S_i)$ is the change in the complexity of the system beyond that accounted for by its intrinsic complexity $C_N^I(X)$ and extrinsic complexity $C_N^E(X)$ (Fig. 1*D Left*); $C_M(X; S_i)$ can be positive, negative, or zero. Eq. **3** refers to a given stimulus S_i; in general, since the statistical structure of the signals sampled from the environment will be characterized by many different stimuli, one should calculate the average matching $\langle C_M(X; S_i)\rangle$ for a set of n stimuli, where $i = 1 \ldots n$. According to Eq. **2**, it follows that

$$C_M(X; S_i) = \sum_{k=1}^{n/2} \langle MI^T(X_j^k; X - X_j^k)\rangle - \langle MI^I(X_j^k; X - X_j^k)\rangle - \langle MI^E(X_j^k; X - X_j^k)\rangle. \qquad [4]$$

Given that the *MI* over the set of all bipartitions comprises all correlations within the system, C_M measures the extent to which the set of correlations intrinsic to the system are enhanced or reduced, on average, by the signals sampled from the environment.

This definition of C_M in terms of changes in *MI* within X is precisely related to the distribution of the *MI* between X and S_i. Since extrinsic input and intrinsic noise in X are on average uncorrelated, the *MI* between any subset X_j^k and S_i is equal to the total entropy of X_j^k minus the intrinsic entropy of X_j^k (8):

$$MI^T(X_j^k; S_i) = H^T(X_j^k) - H^I(X_j^k). \qquad [5]$$

By considering Eq. **4** and substituting, using first Eq. **1** and then Eq. **5**, it can be shown that $C_M(X; S_i) = \sum_{k=1}^{n/2} \langle MI^T(X_j^k; S_i) + MI^T(X - X_j^k; S_i)\rangle - \langle MI^E(X_j^k; S_i) + MI^E(X - X_j^k; S_i)\rangle$. By extending the sum over subset sizes to $k = n$, one obtains that

$$C_M(X; S_i) = \sum_{k=1}^{n} \langle MI^T(X_j^k; S_i)\rangle - \langle MI^E(X_j^k; S_i)\rangle. \qquad [6]$$

Thus, C_M also measures the change in the average *MI* between S_i and all subsets of X, summed over all subset sizes, which is due to the connectivity of the system. Given a fixed value of *MI* between X and S_i, matching measures how well the *MI* between X and S_i is distributed to subsets of units of X (Fig. 1*D Right*).

Implementation. In order to evaluate $C_M(X; S_i)$ for many sets of stimuli and for systems with many different connectivity

patterns, we implemented model systems as linear realizations. As in ref. 3, this allowed us to derive covariance matrices analytically. Each linear system X consisted of n units which received connections from m other units ($1 \leq m \leq n - 1$, no self-connections), resulting in a connection matrix $CON(X)$. $CON(X)$ was normalized so that the absolute value of the sum of the afferent synaptic weights per unit was set to a constant value $w < 1$. If we consider the vector $\mathbf{A}$ of random variables that represents the activity of the units of X, subject to uncorrelated Gaussian noise $\mathbf{R}$ of unit magnitude, we have that, under stationary conditions, $\mathbf{A} = \mathbf{A}*CON(X) + \mathbf{R}$. By defining $Q = [1 - CON(X)]^{-1}$ and averaging over the states produced by successive values of $\mathbf{R}$, we obtain the intrinsic covariance matrix $COV^{I}(X) = \langle \mathbf{A}^t*\mathbf{A} \rangle = \langle Q^t*\mathbf{R}^t*\mathbf{R}*Q \rangle = Q^t*Q$, where the superscript t refers to the transpose. If we include the contribution of extrinsic input $\mathbf{S}_i$, we have that $\mathbf{A} = \mathbf{A}*CON(X) + \mathbf{S}_i*CON(S; X) + \mathbf{R}$. Substituting $W = CON(S; X)*Q$, we obtain $\mathbf{A} = \mathbf{S}_i*W + \mathbf{R}*Q$. By averaging over the states produced by successive values of $\mathbf{R}$ and $\mathbf{S}_i$, we obtain the total covariance matrix $COV^{T}(X) = \langle \mathbf{A}^t*\mathbf{A} \rangle = \langle W^t*\mathbf{S}_i^t*\mathbf{S}_i*W \rangle + \langle Q^t*\mathbf{R}^t*\mathbf{R}*Q \rangle = W^t*COV(S_i)*W + Q^t*Q$. The extrinsic covariance matrix $COV^{E}(X)$ can be derived by setting $CON(X) = 0$. Under Gaussian assumptions, all deviations from independence among the units are expressed by their covariances; from these values of $H(X)$ and therefore of $C_N(X)$ can be derived according to standard formulae (8).

Results

In order to illustrate the notion of C_M with simple examples, we consider model systems and stimuli that capture some basic aspects of the organization of a primary visual area and of its inputs. The model systems contained four units that responded to vertical segments and four units that responded to horizontal segments. The input to the eight units consisted of a set of 18 individual stimuli representing either vertical or horizontal elongated bars (Fig. 2*A Upper*). Assuming stationarity, each stimulus activated the eight units with a given probability. The overall statistical structure of this set of stimuli was represented by positive correlations between contiguous units with similar orientation selectivity, as well as by negative correlations between units with different orientation selectivity (Fig. 2*A Lower*).

Matching the Statistical Structure of the Input. Fig. 2 *B–F* shows results obtained for 50 model systems using this set of stimuli. For systems with randomly generated connectivities, $\langle C_M(X; S_i) \rangle$ values were near 0 (mean = 0.01 ± 0.03; Fig. 2*C Upper*). The connectivity matrix for one of these systems illustrates that positive and negative synaptic weights were distributed at random among all the elements (Fig. 2*C Upper*) giving rise to an almost flat covariance matrix (Fig. 2*D Upper*). For these systems, $C_N(X)$ was also low (mean = 0.31 ± 0.07; Fig. 2*F Upper*). The broad distribution of the angles between their eigenvectors indicates that $COV^{I}(X)$ differed from the average $COV^{T}(X)$ (Fig. 2*F Upper*).

For any set of stimuli, many different synaptic connectivities can lead to an increase in $\langle C_M(X; S_i) \rangle$. For instance, substantial gains in $\langle C_M(X; S_i) \rangle$ were obtained by implementing paradigms for synaptic change similar to those used in ref. 12 (data not shown). Since the focus here is on the significance of an increase in $\langle C_M(X; S_i) \rangle$ rather than on any particular neural mechanisms, we present the results obtained with a constrained nonlinear optimization algorithm (Matlab Optimization Toolbox, Natick, MA). This is a standard method of gradient ascent on a function [in this case, $\langle C_M(X; S_i) \rangle$, with $w < 0.5$ and no self-connections]. For any given stimulus, this algorithm evaluated the change in MI for all bipartitions of the system. Synaptic weights were incrementally modified such that, on average, intrinsic correlations that supported an increase in $MI(X_j^k; X - X_j^k)$ in response to the stimulus were enhanced. This process was repeated for a fixed number of iterations, after which a different stimulus was presented in a random sequence.

The histogram in Fig. 2*B Lower* shows $\langle C_M(X; S_i) \rangle$ values obtained for 50 systems of eight units whose connectivity was modified by gradient ascent. As expected, $\langle C_M(X; S_i) \rangle$ was considerably increased (mean = 0.33 ± 0.01). The connectivity matrix shown (Fig. 2*C Lower*) is representative of what was found in all cases. Each unit tended to have positive connections with contiguous units of similar orientation specificity and negative connections with other units. The resulting functional connectivity, as illustrated by the intrinsic covariance matrix (Fig. 2*D Lower*), was also such that units with the

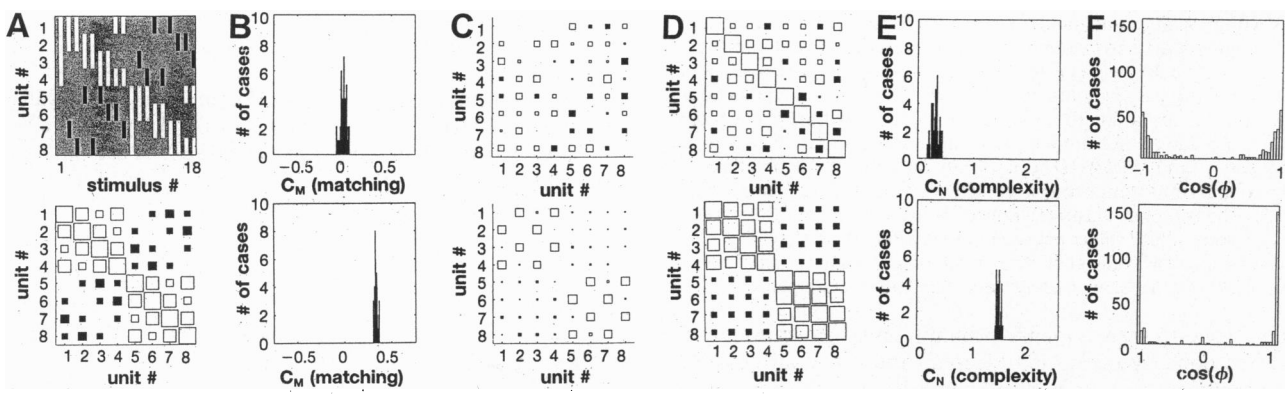

FIG. 2. Results from simulations of networks of eight interconnected units exposed to a series of 18 stimuli presented sequentially. (*B–F*). Results from 50 networks with connectivities that were random (*Upper*) and with connectivities that had been modified by gradient ascent to increase C_M (*Lower*). (*A Upper*) Set of individual stimuli represented as activation vectors impinging upon units 1–8 (white = activation, black = inhibition, gray = no change). Units 1–4 and 5–8 represent orientation detectors selective for vertical and horizontal bars, respectively. Note that units may be coactive within subsets 1–4 and 5–8 but are never coactive between these subsets. (*A Lower*) The corresponding overall statistical structure of the input expressed by an average covariance matrix. Open and filled squares indicate positive and negative correlations, respectively, with size proportional to correlation strength. (*B*) Histogram plot of C_M values for random (*Upper*) and modified (*Lower*) networks. (*C*) Connectivity matrix $CON(X)$ for one example of a random (*Upper*) and of a modified (*Lower*) network. Open and filled squares indicate positive and negative connections, respectively, with size indicating strength. (*D*) Intrinsic covariance matrix $COV^{I}(X)$ for the same networks. (*E*) Histogram plot of values for $C_N^I(X)$. Note that the mean $C_N^I(X)$ after gradient ascent on C_M (*Lower*) is significantly higher than for random networks (*Upper*). (*F*) Cosine of the angle θ between the eigenvectors, sorted by their eigenvalue, of COV^{I} and COV^{T}. For random networks (*Upper*), the wide distribution of cosines indicates that the matrices are structurally different. Clustering of the distribution around 1 and −1 for networks with high values of C_M (*Lower*) indicates that COV^{I} and COV^{T} have, on average, similar structures.

same orientation preference were positively correlated, while units with different orientation preferences were slightly negatively correlated. Furthermore, it was found that, although $C_M(X; S_i)$ and $C_N(X)$ values could occasionally increase or decrease independently, after gradient ascent the functional connectivities associated with increased values of $\langle C_M(X; S_i)\rangle$ showed high values of $C_N(X)$ (mean = 1.45 ± 0.04; Fig. 2*E Lower*). The small angle between the sorted eigenvectors indicated that the average $COV^T(X)$ was a scaled version of $COV^1(X)$ (i.e., it had the same eigenvectors but higher eigenvalues; Fig. 2*F Lower*).

For the above examples, we chose systems in which all the units (n = 8) received direct input from S and all the connections among them were allowed to change. We found that, under these conditions, an increase in $\langle C_M(X; S_i)\rangle$ values was associated with a nearly one-to-one correspondence between the overall statistical structure of the input and both the synaptic and functional connectivity of the system. These examples were specifically chosen to demonstrate the notion of fit or matching in a straightforward way. We emphasize, however, that such one-to-one correspondence is in general neither present nor necessary. For example, we computed the value of $\langle C_M(X; S_i)\rangle$ obtained by gradient ascent in four systems with 10 units, 2 of which were not directly connected to S. An analysis of the synaptic connectivity indicated that these two additional units became specialized to detect the presence of long vertical or horizontal bars, respectively. Thus, these units discovered an important statistical feature of the input stimuli, although in this case a one-to-one correspondence did not hold. In addition, C_M and C_N values were higher than those obtained with systems of 8 units [$\langle C_M(X; S_i)\rangle$ = 0.65 ± 0.06 and $C_N(X)$ = 2.30 ± 0.15], indicating that larger repertoires of intrinsic correlations provide the potential for higher values of $\langle C_M(X; S_i)\rangle$. By contrast, $\langle C_M(X; S_i)\rangle$ and $C_N(X)$ remained around zero for systems of 10 or more units with random connectivities. Results qualitatively similar to the ones presented here were obtained upon using many different sets of stimuli as well as different numbers of units in the systems (data not shown).

Matching Individual Stimuli: The Role of Context. Once a system has come to match the overall statistical structure of the signals sampled from the environment, what happens when it is presented with individual stimuli? By itself, a given stimulus will inevitably contain only a small subset of the statistical regularities in the input. The following examples show that such stimuli enhanced or reduced the set of intrinsic correlations in the system, depending on whether they were consistent or inconsistent with the overall statistical structure of previously encountered inputs.

We examined the responses of the two networks whose connectivities are shown in Fig. 2*C*. Fig. 3 shows the responses of the two networks to the two different stimuli. The first stimulus was similar to some of those in the original stimulus set (Fig. 2*A*). The second stimulus was novel and unlike any of the stimuli encountered before in that a vertical and a horizontal line segment were present together. The histograms represent the change in the strength of all intrinsic correlations $[MI^T(X_j^k; X - X_j^k) - MI^1(X_j^k; X - X_j^k) - MI^E(X_j^k; X - X_j^k)]$ in response to the given stimulus. The histograms obtained from the random network (Fig. 3 *Upper*) show that it did not "recognize" either stimulus and C_M was close to 0 [$C_M(X; S_1)$ = 0.01; $C_M(X; S_2)$ = 0.02]. Fig. 3 *Lower* illustrates the responses of the system modified by gradient ascent on C_M with the original stimulus set (Fig. 2*A*). Fig. 3*A Lower* shows that, for the first stimulus, there was a marked increase in the strength of many intrinsic correlations. The system "recognized" the stimulus and C_M was positive [$C_M(X; S_1)$ = 0.24]. A detailed analysis indicated that the intrinsic correlations triggered by this stimulus provided a specific context which could be interpreted as filling in missing evidence and implying certain

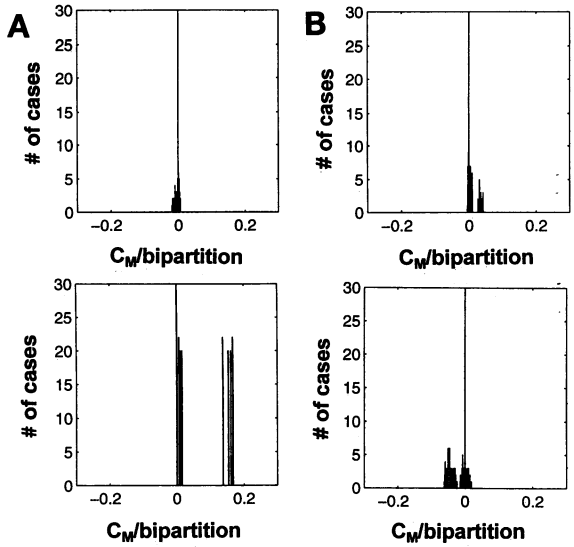

FIG. 3. Responses of a network before (*Upper*) and after (*Lower*) gradient ascent on C_M to known (*A*) and novel (*B*) stimuli. (*A*) Histograms showing the difference between total, intrinsic, and extrinsic *MI* for all bipartitions of the system upon presentation of a stimulus activating units 2 and 3 to the random network (*Upper*) and to the network modified by gradient ascent (*Lower*). In the case of the modified network, this stimulus enhances not only the positive correlation between units 2 and 3 but also their positive correlations with other units responding to vertical bars (1–4) and their negative correlations with units responding to horizontal bars (5–8). (*B*) As in *A*, but with a stimulus that activates units 2 and 6.

predictions—for example, that the two active elements were likely to be part of an elongated vertical bar, and that vertical bars were rarely present simultaneously with horizontal bars. By contrast, when the system was tested with the novel stimulus (Fig. 3 *Lower*), there was a decrease in the strength of intrinsic correlations and C_M was negative [$C_M(X; S_2)$ = −0.09]. Such a response can be considered as an indication of novelty. On the other hand, when this novel stimulus was included in the original set and the process of gradient ascent was allowed to proceed, the stimulus was found to result in positive values for C_M (data not shown).

Discussion

In mammalian brains, most neurons receive signals from other neurons rather than directly from sensory inputs. Moreover, there is increasing evidence that neural responses to such sensory inputs cannot be characterized in terms of fixed receptive fields, since neurons are sensitive to multiple contextual cues (10). Furthermore, the brain is spontaneously active. After development and experience, such intrinsically generated activity constitutes a highly organized functional connectivity seen even in the absence of sensory inputs, as strikingly demonstrated by dreaming and imagery (11). These observations raise serious questions about attempts to describe the brain as an information processing device.

In this paper, we provide some conceptual and operational tools to address these issues from a different, selectionist perspective (1). Relying on the notion of neural complexity (C_N) (3), we have here introduced a related statistical measure, matching complexity (C_M), which represents the change in complexity in a neural system when it responds to signals from an environment. While C_N measures the amount and heterogeneity of the intrinsic functional connectivity of a neural system, C_M measures how well that functional connectivity fits or matches the statistical structure of its sensory inputs. We showed that as C_M increases the intrinsic functional connec-

tivity becomes progressively more adapted to the statistical structure of the sensory input. Our analysis also indicated that the functional connectivity of the brain constitutes an intrinsic "context" which by necessity dominates its responses to any single stimulus. Previous large-scale computer simulations (7, 12) suggest that reentry, involving ongoing recursive signaling among multiple sets of neuronal groups (1, 13), is the key process by which such intrinsic context is made available in a rapid and parallel way. The constructive and correlative properties of reentry allow the brain to solve the binding problem (12), to recreate the effects of extrinsic signals even in their current absence, to fill in ambiguous signals, and to predict future occurrences. Its associative properties ensure that local changes in synaptic efficacy are dependent on this intrinsic context.

The present analysis sheds a new light on the relationship between the processes of categorization and association. For categorization, the units of a system must become specialized and respond to different inputs; for association, these units must be able to correlate their responses with those of many other units. According to the notions introduced here, a single principle, the increase in *MI* among subsets of units within the brain, can account for the seemingly opposite requirements of categorization and association, provided that there are biological mechanisms to ensure that this increase occurs for both small and large subsets. An increase in the *MI* between individual units and their inputs will reflect an increase in functional integration; an increase in the *MI* between subsets composed of many units and their inputs will reflect the functional specialization of individual units within these larger subsets. Thus, both the functional integration and the specialization of a neural system will increase, translating into an increase in C_N. For C_M to increase, it is additionally required that, *on average*, C_N increases more when extrinsic input is present than when it is absent. This can be achieved if mechanisms mediating synaptic changes have a way to determine the degree to which their responses are modulated by extrinsic inputs. A possibility is that these mechanisms are sensitive to the level of activity of diffuse neuromodulatory systems that signal the alternation between sleep and wakefulness as well as the saliency of stimuli. Another possibility is that connections more likely to relay extrinsic inputs (e.g., forward connections) may differ, for example with respect to voltage dependency, from connections that carry mostly intrinsic inputs [e.g., backward and lateral connections (references in ref. 12)].

Matching, Information Processing, and Context. Since C_M is expressed in terms of *MI*, the notion of matching in a selectional system can usefully be related to the statistical foundations of information theory (9). Standard applications of information theory consider the problem of transmitting signals across a channel with limited capacity and noise. Various strategies are employed to deal with problems such as dimensionality reduction (fewer outputs than inputs), redundancy reduction (information compression), and noise. However, several fundamental aspects of brain organization are puzzling when viewed in these terms: (*i*) the number of units in the brain is much larger than that of the units that receive direct sensory inputs (the opposite of dimensionality reduction); (*ii*) there is evidence for extensive correlated neural activity in response to a stimulus (14) (the opposite of redundancy reduction); and (*iii*) the signals coming from the sensory periphery are a minority with respect to reentrant signals from the rest of the brain even in so-called "relay" nuclei (the opposite of noise reduction). Considered from the present perspective, however, the larger the number of units, the vaster the repertoire of subsets that can be selected by interacting with signals from the environment. Moreover, if C_M is high, *MI* from the same stimulus is efficiently distributed to many different subsets in the brain, implying a high degree of correlated neural activity rather than a reduction of redundancy. If C_N is also high, however, each of the subsets responding to a given stimulus will have a specialized relation to the rest of the brain, so that redundant inputs from the stimulus can lead to different functional consequences. The existence of more than one way satisfactorily to recognize a given input is an instance of degeneracy, another fundamental property of selective systems (1). Finally, the present analysis relates changes in *MI* among subsets of units within the brain to changes in the distribution of *MI* between these subsets and the sensory input (compare Eq. **3** with Eq. **6**, and see Fig. 1*D*). This means that, if C_N and C_M are both high, for a small value of the extrinsic *MI* between an individual stimulus and the brain there will be a large increase in the intrinsic *MI* among subsets of units within the brain. If incoming stimuli act largely by modulating intrinsic correlations, the reentrant interactions that support such correlations should not be considered as a source of noise, but rather as providing the context that makes incoming stimuli meaningful. By characterizing context in terms of *MI* within the brain, our analysis shows that the brain can literally go "beyond the information given" (15).

Applications and Predictions. The present analysis leads to several predictions. For example, stimuli that reveal the constructive nature of perception, such as illusory contours, two-dimensional diagrams that lead to three-dimensional percepts (Necker cube), and random-dot sterograms should be associated with positive values of C_M. In particular, there should be a change in C_M when a hidden figure suddenly emerges from a noisy background. This should go along with an increase in correlated activity along specific corticocortical and corticothalamic reentrant loops. More generally, meaningful and meaningless stimuli, such as words vs. nonwords, should be associated with different values of C_M. As for C_N (16), among the most interesting applications of C_M should be those in the field of functional neuroimaging. It will be important to determine the sensitivity of various neuroimaging techniques with respect to changes in these measures. It remains to be seen whether measuring brain responses to repeated presentations of stimuli approximates quasistationary conditions. If that is the case, paradigms measuring C_M should provide, compared with subtraction paradigms based on activity, a more comprehensive assessment of the "meaning" of a given stimulus to a given subject in terms of the range and specificity of the set of associations that the stimulus triggers.

We thank the Fellows of The Neurosciences Institute for extensive discussions. This work was carried out as part of the theoretical neurobiology program at the Institute, which is supported by Neurosciences Research Foundation. The Foundation receives major support from Sandoz Pharmaceutical Corporation. O.S. is a W. M. Keck Foundation Fellow.

1. Edelman, G. M. (1987) *Neural Darwinism: The Theory of Neuronal Group Selection* (Basic, New York).
2. Aertsen, A. & Preissl, H. (1991) in *Nonlinear Dynamics and Neuronal Networks*, ed. Schuster, H. (VCH, Weinheim, Germany), pp. 281–301.
3. Tononi, G., Sporns, O. & Edelman, G. M. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 5033–5037.
4. Callaway, E. M. & Katz, L. C. (1990) *J. Neurosci.* **10,** 1134–1149.
5. Gilbert, C. D. & Wiesel, T. N. (1989) *J. Neurosci.* **9,** 2432–2442.
6. Koffka, K. (1935) *Principles of Gestalt Psychology* (Harcourt, New York).
7. Sporns, O., Tononi, G. & Edelman, G. M. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 129–133.
8. Jones, D. S. (1979) *Elementary Information Theory* (Oxford Univ. Press, Oxford).
9. Shannon, C. E. & Weaver, W. (1949) *The Mathematical Theory of Communication* (Univ. of Illinois Press, Chicago).
10. Nelson, J. I. (1985) in *Models of the Visual Cortex*, eds. Rose, D. & Dobson, V. G. (Wiley, Chichester, U.K.), pp. 108–122.

11. Llinás, R. R. & Paré, D. (1991) *Neuroscience* **44,** 521–535.
12. Tononi, G., Sporns, O. & Edelman, G. M. (1992) *Cereb. Cortex* **2,** 310–335.
13. Edelman, G. M. (1989) *The Remembered Present: A Biological Theory of Consciousness* (Basic, New York).
14. Gawne, T. J. & Richmond, B. J. (1993) *J. Neurosci.* **13,** 2758–2771.
15. Bruner, J. S. (1973) *Beyond the Information Given* (Norton, New York).
16. Friston, K. J., Tononi, G., Sporns, O. & Edelman, G. M. (1996) *Hum. Brain Mapping*, in press.