Measures of degeneracy and redundancy in biological networks

GIULIO TONONI[†], OLAF SPORNS, AND GERALD M. EDELMAN

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

ABSTRACT   Degeneracy, the ability of elements that are structurally different to perform the same function, is a prominent property of many biological systems ranging from genes to neural networks to evolution itself. Because structurally different elements may produce different outputs in different contexts, degeneracy should be distinguished from redundancy, which occurs when the same function is performed by identical elements. However, because of ambiguities in the distinction between structure and function and because of the lack of a theoretical treatment, these two notions often are conflated. By using information theoretical concepts, we develop here functional measures of the degeneracy and redundancy of a system with respect to a set of outputs. These measures help to distinguish the concept of degeneracy from that of redundancy and make it operationally useful. Through computer simulations of neural systems differing in connectivity, we show that degeneracy is low both for systems in which each element affects the output independently and for redundant systems in which many elements can affect the output in a similar way but do not have independent effects. By contrast, degeneracy is high for systems in which many different elements can affect the output in a similar way and at the same time can have independent effects. We demonstrate that networks that have been selected for degeneracy have high values of complexity, a measure of the average mutual information between the subsets of a system. These measures promise to be useful in characterizing and understanding the functional robustness and adaptability of biological networks.

There are many examples of biological systems composed of elements that are structurally different but which, under certain conditions, can perform similar functions. Classic illustrations include nucleotide triplets degenerate in the third position that code for the same amino acid, different proteins that catalyze the same enzymatic reaction, and different genotypes that produce the same phenotype. In a neurobiological context, the capability of different, nonisomorphic structures to yield isofunctional outputs, effects, or consequences has been called degeneracy (1, 2). This term previously was used in immunology to refer to the ability of different antibodies to bind to the same antigen (3). In physics, the term degeneracy is applied to systems taking on several discrete or distinct energy values or states.

Examples of degeneracy in neurobiology abound (2). The convergent-divergent connectivity of the brain suggests that large numbers of neuronal groups are able to affect the output of any chosen subset of neurons in a similar way. For example, a large number of different brain structures can influence, in series or in parallel, the same motor outputs, and after localized brain lesions, alternative pathways capable of generating functionally equivalent behaviors frequently emerge (4). Degeneracy also is encountered at the neuronal level. Neural signaling mechanisms typically use parallel as well as converging pathways of transmitters, receptors, kinases, phosphatases, and second messengers. A constrained set of outputs such as specific changes in the levels of second messengers or in gene expression thus can be brought about by a large number of different input combinations.

Although many similar examples exist in all fields and levels of biology, a specific notion of degeneracy has yet to be firmly incorporated into biological thinking, largely because of the lack of a formal theoretical framework. As a consequence, instances of degeneracy often are not treated as exemplifications of a biological principle, but are discounted as redundancy. In technical usage, redundancy refers to duplication or repetition of elements within electronic or mechanical components to provide alternative functional channels in case of failure. In information theory, it refers to the repetition of parts or all of a message to circumvent transmission error. Although the structural definition of redundancy applies readily in engineering, the requirement of identity among elements in a population or in biological networks is rarely satisfied and in any case is difficult to assess. The biological examples mentioned above, for instance, do not in general involve structurally identical parts. This fact has important consequences, because unlike redundant elements, degenerate elements may produce different outputs in different contexts.

In the present paper, we propose to clarify these issues by moving explicitly to the functional level, defining both degeneracy and redundancy in information theoretical terms. This approach allows us to distinguish between degeneracy and redundancy within a unified framework and, through the use of well-defined measures, to make both concepts operationally useful. Although these measures are applied here to neural examples, in principle they may be extended to any biological network or complex system.

THEORY

As in previous papers, we consider neural systems consisting of a number of elementary components, which we take to represent neuronal groups sharing anatomical connections (5, 6). We then study the dynamic interactions between these elements and an output sheet, assuming that the global statistical properties of a system do not change with time (stationarity). With these assumptions in place, we characterize degeneracy and redundancy in terms of the average mutual information between subsets of elements within a neural system and the output sheet.

Consider a neural system X with n elements whose activity is described by a Gaussian stationary multidimensional stochastic process (7) that produces a set of outputs by means of an output sheet O and a fixed connectivity matrix CON(X;O) between O and a subset of the system units (Fig. 1*A*). The joint probability density function describing such a multivariate process, corresponding here to its functional connectivity, can be characterized (7, 8) in terms of entropy (H) and mutual information (MI). Entropy and mutual information are used here in their statistical connotation; they can be thought of as

---

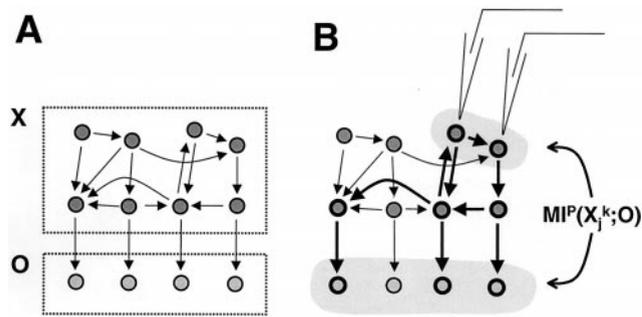[†]To whom reprint requests should be addressed. e-mail: tononi@nsi.edu.

FIG. 1.   Schematic diagram illustrating bases for the proposed measure of degeneracy. (*A*) We consider a system X, composed of individual elements (gray circles, *n* = 8) that are interconnected (arrows) among each other [with a connection matrix CON(X)] and that also have connections to a set of output units O [with a connection matrix CON(X;O)]. (*B*) A subset (shaded in gray) of the system X is perturbed by injecting (large Vs at the top) a fixed amount of variance (uncorrelated noise) into each of its constituent units. This perturbation activates numerous connections within the system (thick arrows) and produces changes in the variance of a number of units within X and O. The resulting mutual information under perturbation $MI^P(X^k_j;O)$ is computed (see Eq. **1**). This procedure is repeated for all subsets of sizes $1 \le k \le n$ of the system X.

multivariate generalizations of variance and covariance in univariate statistics that are sensitive both to linear and nonlinear interactions. For instance, consider a *j*th subset X^k_j of the system X composed of *k* units, and the output sheet O. The MI between X^k_j and O is given by:

$$MI(X^k_j;O) = H(X^k_j) + H(O) - H(X^k_j,O),   [1]$$

where $H(X^k_j)$ and $H(O)$ are the entropies of X^k_j and O considered independently, and $H(X^k_j, O)$ is the joint entropy of subset X^k_j and output O. Thus, $MI(X^k_j;O)$ measures the portion of entropy shared by the system subset X^k_j and the output O. It is high if both X^k_j and O have high entropy (high variance) and share a large fraction of it (high covariance); it is low or zero if X^k_j and O have low entropy or are statistically independent.

Degeneracy and Redundancy. By itself, mutual information is a measure of the deviation from statistical independence between two subsets and is insensitive to the direction of the interaction: a positive value of mutual information between X^k_j and O could be the result of a causal effect of X^k_j on O through forward connections, of O on X^k_j through back connections, or of another subset simply providing statistically correlated input to both X^k_j and O. Here we are interested only in measuring the causal effects on the output of changes in the state of subsets of the system. To evaluate these causal effects, we consider the change in mutual information between each subset X^k_j and O when that subset is injected with a fixed amount of variance (uncorrelated random noise, Fig. 1*B*). We call the mutual information value obtained by such perturbation of each subset $MI^P(X^k_j;O)$. For simplicity, we shall assume that before injecting the variance the variance of the system is zero, so that the initial value of mutual information between the system and the output is also zero.

We define the degeneracy $D_N(X;O)$ of X with respect to O as:

$$D_N(X;O) = \sum_{k=1}^{n}[<MI^P(X^k_j;O)> - (k/n)MI^P(X;O)].   [2a]$$

According to Eq. **2a**, $D_N(X)$ is high when the mutual information between the whole system (*k* = *n*) and the output is high and at the same time the average mutual information between small subsets of the system (small values of *k*) and the

output is higher than would be expected from a linear increase over increasing subset size (Fig. 2*A*).

$D_N(X;O)$ can be expressed in mathematically equivalent ways. In particular, $D_N(X;O)$ corresponds to the average mutual information shared between bipartitions of X and the output O, summed over all bipartition sizes:

$$D_N(X;O) = 1/2\sum_{k=1}^{n}<MI^P(X^k_j;X - X^k_j;O)>,   [2b]$$

where the mutual information that is shared between X^k_j, X-X^k_j, and O is $MI^P(X^k_j;X-X^k_j;O) = MI^P(X^k_j;O) + MI^P(X-X^k_j;O) - MI^P(X;O)$. Thus, according to Eq. **2b**, $D_N(X;O)$ is high when, on average, the mutual information shared between any bipartition of the system and the output is high (Fig. 2*B*).

We now define the redundancy R(X;O) of X with respect to O as the difference between the summed mutual information upon perturbation between *n* subsets of size *k* = 1 and O and the mutual information upon perturbation between the entire system (*k* = *n*) and O:

$$R(X;O) = \sum_{j=1}^{n}[MI^P(X^1_j;O)] - MI^P(X;O).   [3]$$

According to this definition, redundancy is high if the sum of the mutual information between each element and the output is much larger than the mutual information between the entire system and the output. This means that each of the elements of the system contributes similar information with respect to the output. Redundancy is zero if all elements of the system contribute to the output independently and the mutual information between the entire system and O is equal to the sum of the mutual information between each element of the system and O.

Based on this definition, we also can express $D_N(X;O)$ in terms of average redundancy values with respect to O for increasing subset sizes:

$$D_N(X;O) = \sum_{k=1}^{n}[(k/n)R(X;O) - <R(X^k_j;O)>].   [4]$$

According to Eq. **4**, $D_N(X;O)$ is high when the redundancy of the system (i.e., for *k* = *n*) with respect to the output is high and at the same time the average redundancy for small subsets (small values of *k*) is lower than would be expected from a linear increase over increasing subset size (Fig. 2*C*).

A measure of degeneracy that does not require averaging among different subsets is also usefully introduced. This measure, which we shall designate by italics as *D*(X;O), and
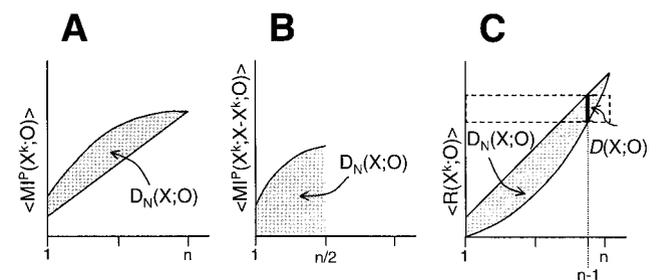

FIG. 2.   Graphical representation of different expressions for degeneracy. (*A*) Degeneracy expressed in terms of the average mutual information between subsets of X and O under perturbation (see Eq. **2a**). (*B*) Degeneracy expressed in terms of the average mutual information shared between bipartitions of X and O (see Eq. **2b**). (*C*) Degeneracy expressed in terms of the average redundancy (see Eq. **4**). A graphical interpretation for the degeneracy *D*(X;O) (see Eq. **5**) is indicated as a dotted rectangular area with height corresponding to that of bar at *n* − 1.

Neurobiology: Tononi *et al.*

Proc. Natl. Acad. Sci. USA 96 (1999)   3259

which is related to degeneracy $D_N(X;O)$, measures the portion of the entropy of the output that is jointly accounted for by different elements of the system. It is given by:

$$D(X;O) = \Sigma\ MI^P(X^1_j;X - X^1_j;O) - R(X;O)$$

$$= MI^P(X,O) - \Sigma\ MI^P(X^1_j;O|X-X^1_j;O), \quad [5]$$

where the sum is over the n elements and $MI^P(X^1_j;O|X-X^1_j;O)$ is the conditional mutual information between each element and O given the mutual information between the rest of the system and O. The relationship between $D_N(X;O)$ and $D(X;O)$ is displayed graphically in Fig. 2*C*.

Relationship to Complexity and Integration. In previous work (5), we introduced a measure called neural complexity $C_N(X)$, which measures the degree to which an isolated system is both functionally integrated (different elements of the system are integrated and behave coherently) and functionally segregated or specialized (different elements are relatively independent and the system is highly differentiated). We also introduced the system integration I(X), a measure of the total reduction of entropy because of the interactions within the system (5). The expressions for complexity and integration of a system bear a striking formal resemblance to the expressions for the degeneracy and redundancy of a system with respect to an output (Fig. 3). If, in the expression for $D_N(X;O)$, one replaces $MI^P(X^k_j;O)$—the mutual information between each subset and the output—by $H(X^k_j)$—the entropy of each subset—one obtains $C_N(X)$ and the equations defining the two measures are formally identical. As defined, $I(X) = \Sigma\ H(x_i) - H(X)$, i.e., the difference between the sum of the entropies of the n individual components considered independently and the entropy of X considered as a whole [compare Eq. **3**; compare also the original information theoretical expression for redundancy (7) as the ratio $(\Sigma\ H(x_i) - H(X))/\Sigma\ H(x_i)$]. The neural complexity $C_N(X)$ of system X is defined in three mathematically equivalent ways (Fig. 3 *A–C*): $C_N(X) = \Sigma\ [<H(X^k_j)> - (k/n)H(X)] = 1/2\Sigma\ <MI(X^k_j;X-X^k_j)> = \Sigma\ [(k/n)I(X) - <I(X^k_j)>]$ (compare with Eqs. **2a**, **2b**, and **4**).

A related expression for complexity (9) that does not require averaging among different subsets measures the portion of the entropy of a system that is accounted for by the interactions among its elements (Fig. 3*C*). This measure, which we shall designate by italics as *C*(X), is given by $\Sigma MI(X^1_j;X-X^1_j) - I(X) = H(X) - \Sigma H(X^1_j|X-X^1_j)$, where $(X^1_j|X-X^1_j)$ is the conditional entropy of each element given the entropy of the rest of the system (compare Eq. **5**). Note that both $C_N(X)$ and *C*(X) have zero values for systems composed of disconnected elements, low values for systems composed of elements that are integrated and homogeneous (undifferentiated) and high values for systems that are both integrated and differentiated.
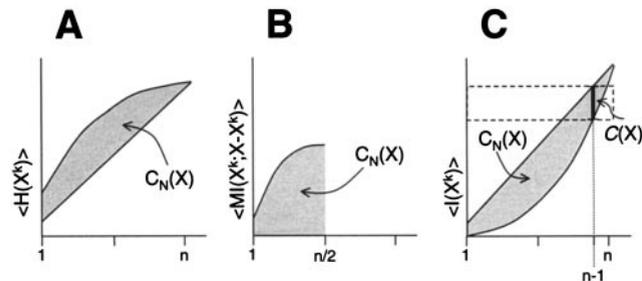

FIG. 3.   Graphical representation of different expressions for complexity. Note the homology between these expressions and those illustrated in Fig. 2. (*A*) Complexity, $C_N(X)$, expressed in terms of the average entropy. (*B*) Complexity expressed in terms of the average mutual information. (*C*) Complexity expressed in terms of the average integration (see ref. 9). A graphical interpretation for the complexity *C*(X) is indicated as a dotted rectangular area with height corresponding to that of bar at $n - 1$.

Implementation. To evaluate $D_N(X;O)$ for systems with many different connectivity patterns, we implemented several model systems as linear realizations. As described (5, 6), this implementation allows us to derive covariance matrices analytically. Each linear system X consisted of n neural elements connected to each other according to a connection matrix CON(X) with no self-connections. 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 $w < 1$. A subset of m output elements ($m<n$) was connected (CON(X;O), Fig. 1*A*) one by one to elements of the output sheet O with connections of unit weight.

We consider the vector **A** of random variables that represents the activity of the elements of X after perturbation, i.e., after injecting uncorrelated Gaussian noise **R** of unit magnitude into a subset X^k_j of elements. For example, if the subset receiving the injection of uncorrelated noise corresponds to the entire system, under stationary conditions we have that $\mathbf{A} = \mathbf{A}*CON(X) + \mathbf{R}$. By defining $Q = [1 - CON(X)]^{-1}$ and averaging over the states produced by successive values of **R**, we obtain the covariance matrix under perturbation $COV^P(X) = <\mathbf{A}^t*\mathbf{A}> = <Q^t*\mathbf{R}^t*\mathbf{R}*Q> = Q^t*Q$, where the superscript t refers to the transpose. The covariance matrix then is normalized to a correlation matrix to ensure that all output elements have unit variance, and the correlation matrix is multiplied through the output connection matrix CON(X;O) to obtain the correlation matrix between the system and the output sheet. Under Gaussian assumptions, all deviations from independence among the units are expressed by their covariances; from these, values of H(X) and therefore of $MI^P(X;O)$ can be derived according to standard formulae (8). This procedure is repeated by applying the uncorrelated Gaussian noise in turn to all possible subsets of the system, in such a way that **R** has unit magnitude for the elements of each subset and is set to zero for the other elements. If the resulting covariance matrix has some elements with zero variance, the corresponding rows and columns are eliminated.

ILLUSTRATION

We first illustrate the measures of degeneracy defined above by applying them to several simple examples. Then, we demonstrate that an increase in degeneracy with respect to an output pattern tends to be accompanied by an increase in complexity.

Examples of Independence, Degeneracy, and Redundancy. Three basic examples can be constructed to illustrate key properties of degeneracy. Fig. 4 (*Top*) shows an example of a network for which different elements produce totally independent effects or do not affect the output at all (independence). Fig. 4 (*Bottom*) shows an example of a fully connected network; all of its units can affect the output but their effects on the output are essentially identical, i.e., the units show functional redundancy. Fig. 4 (*Middle*) shows an example of a degenerate network with structured connections; many of the elements have effects on the output that are both functionally redundant and functionally independent (degeneracy). Graphs of connectivities are given in Fig. 4*A*; resulting correlation matrices are presented in Fig. 4*B*. Plots in Fig. 4 *C–E* display, respectively, the distributions of average mutual information between subsets of X and O (Eq. **2a**), average mutual information shared between bipartitions of X and O (Eq. **2b**) and average redundancy (Eq. **4**), with the resulting degeneracy $D_N(X;O)$ displayed as the shaded area. $D_N(X;O)$ is zero for a system (Fig. 4 *C–E, Top*) for which all elements affect the output independently (zero redundancy). A system (Fig. 4 *C–E, Bottom*) that has very high functional redundancy has relatively low values of degeneracy, as different combinations of the system's units have similar effects on the output. A system (Fig. 4 *C–E, Middle*) in which many combinations of
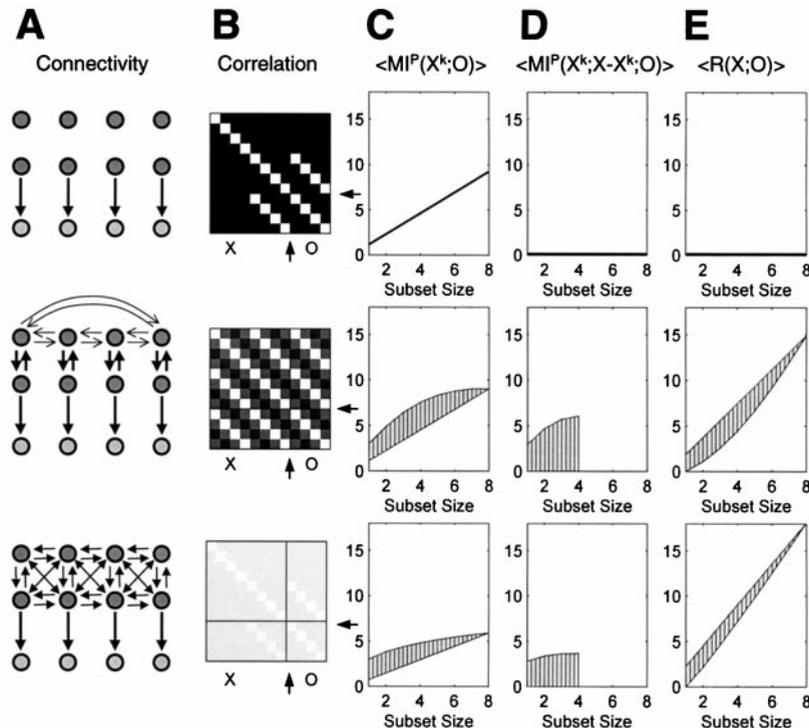
FIG. 4.   Results from simulations of three examples of systems of eight units, each connected to an output sheet of four units. The three examples are: a system that lacks all intrinsic connectivity (*Top*, independent case), a system that is characterized by four modules of two strongly interconnected units that are weakly interconnected among themselves (*Middle*, degenerate case), and a system whose units are fully interconnected (*Bottom*, redundant case). (*A*) Graphs of connections among units of the system and among system units and the output. Arrow width indicates connection strengths (not all connections shown). (*B*) Correlation matrices, displaying correlations intrinsic to X, intrinsic to O, as well as cross-correlations between X and O. The two arrows demarcate the output portion of the correlation matrix. (*C*) Distribution of the average mutual information between subsets of X and O over all subset sizes (see Eq. **2a**). (*D*) Distribution of the average mutual information shared between bipartitions of X and O (see Eq. **2b**). (*E*) Distribution of the average redundancy over all subset sizes (see Eq. **4**). Degeneracy is indicated in *C–E* as a shaded area (compare Fig. 2). Expressions for the ordinates of graphs in *C–E* are at the top of each column.

units have both functionally redundant and functionally independent effects on the output has the highest degeneracy. This structure is accompanied by relatively high redundancy for large subsets, and by relatively low redundancy for smaller subsets of the system. Note that for systems larger than those shown the degeneracy values for degenerate systems would be much larger than those for fully redundant systems. The examples shown are representative; tests for variation in all critical parameters, in particular the total amount of connectivity (w), the amount of variance injected, and the ratio of system to output units produced similar results.

Optimizing Degeneracy with Respect to a Set of Outputs. Fig. 5 summarizes results obtained from simulations of neuronal networks whose intrinsic connectivity gradually was changed to produce high degeneracy with respect to a particular set of outputs—given here by a particular correlation matrix among the units of the output sheet. To achieve the appropriate synaptic changes we used a constrained nonlinear optimization algorithm (Matlab Optimization Toolbox, Natick, MA). During each iteration, the algorithm produced sets of correlation matrices and allowed calculation of the degeneracy with respect to the desired output as given in Eq. **5** (similar results were obtained when optimizing using Eq. **2a**). Connection strengths of the intrinsic connections within the system then were incrementally modified so that, on average, connections that supported higher overall degeneracy were strengthened while others were weakened. No self-connections were allowed. For the results shown in Fig. 5, networks were initialized to have positive connections only, and all connections remained positive for the entire simulation. As in previous studies (6), the total amount of connectivity per network unit (w, computed as the sum of afferent connection

strengths) was kept constant for all units throughout the simulation (here $w = 0.9$).

The top row of Fig. 5 displays data obtained from random networks (used as initial conditions in the optimization runs). Connection matrices typically consisted of randomly chosen normally distributed connection strengths (Fig. 5*A*, *Top*) that gave rise to fairly uniform patterns of correlations (Fig. 5*B*, *Top*). Random networks produced low degeneracy with respect to their outputs (mean value of $D(X;O) = 4.37 \pm 0.25$; Fig. 5*C*, *Top*). Their complexity was also low (mean value of $C(X) = 0.18 \pm 0.04$; Fig. 5*D*, *Top*).

After connection strengths were adjusted to yield high degeneracy, networks contained patterned connectivity, arranged in two weakly linked modules that were more strongly connected to pairs of output units (Fig. 5*A*, *Bottom*). As required by simulation constraints, they produced the appropriate correlation patterns in their output units (Fig. 5*B*, *Bottom*). Their degeneracy was significantly higher than that of random networks (mean value of $D(X;O) = 8.10 \pm 0.06$; Fig. 5*C*, *Bottom*). In addition, their intrinsic complexity was high (mean value of $C(X) = 0.84 \pm 0.02$; Fig. 5*D*, *Bottom*). We tested a variety of other output patterns and parameter settings (i.e., varying the amount of connectivity, the relative number of system units versus output units, and inclusion of inhibitory connections). Essentially similar results were obtained in all cases.

To test whether the networks shown in Fig. 5, *Bottom*, correspond to optimally or near-optimally degenerate solutions, we constructed a variety of other networks designed to yield the output pattern used in the simulations. Consistently, the values for degeneracy of these networks were considerably lower than those obtained from the optimization runs. This
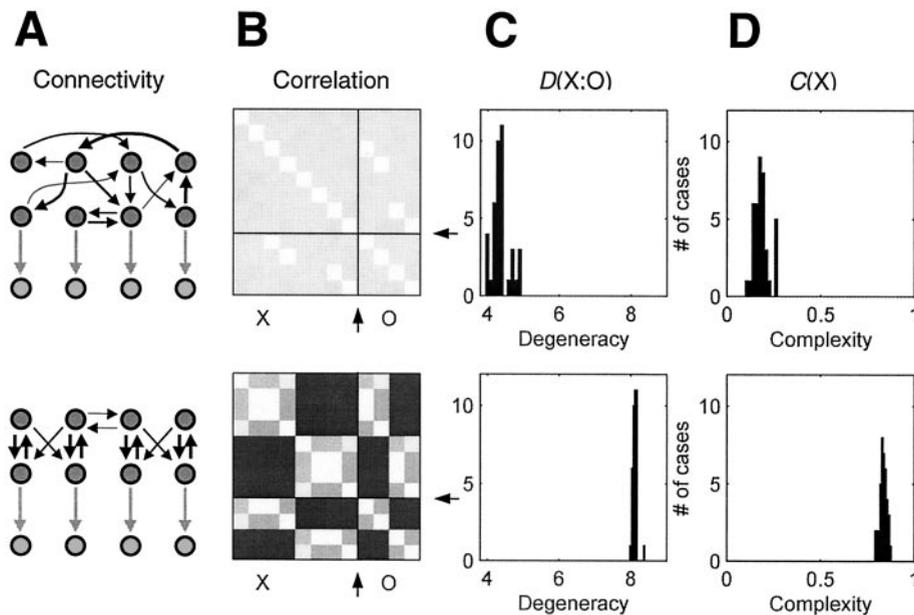
FIG. 5.   Typical results obtained from simulations of 40 systems of eight units connected to an output sheet of four units; during optimization the system's connectivity was modified by gradient ascent to increase degeneracy $D(X;O)$ (Eq. **5**) with respect to a fixed output pattern. This output was given as a correlation matrix of the units of the output sheet. In the results shown, the output consisted of a high cross-correlation (0.75) among pairs of two units of the output and a low cross-correlation (0.25) between them. (*Top*) Data obtained from one example of a randomly connected system before the beginning of the runs. (*Bottom*) Data from one example obtained after gradient ascent resulted in a stable value for degeneracy. (*A*) Schematic representations of graphs showing the interconnections between units of the system and the output. (*B*) Correlation matrices (conventions as in Fig. 4*B*). (*C*) Distribution of values for degeneracy $D(X;O)$ (see Eq. **5**) for 40 randomly chosen (*Top*) and optimized (*Bottom*) systems. (*D*) Distribution of values for complexity $C(X)$ for the corresponding systems.

finding suggests that the network architectures found by gradient ascent are indeed close to optimal with respect to their degeneracy. We found that, whenever a patterned output was required (i.e., an output characterized by both high and low correlations), the architectures resembled groups of units with strong intra-group connections and weaker inter-group connections.

DISCUSSION

In this paper we have defined the concepts of degeneracy and redundancy in functional terms and provided a principled way to formulate these concepts by using information theoretical measures. By using simple linear models we have illustrated the conditions under which degeneracy and redundancy are each either high or low and have analyzed the relationship between degeneracy and redundancy. Finally, we have shown in linear model systems that procedures leading to the selection of systems having high degeneracy with respect to a set of outputs are accompanied by an increase in the complexity of these systems, i.e., an increase in the degree to which these systems are both functionally integrated and locally segregated.

Degeneracy and Redundancy. We originally suggested that nonisomorphic structures that are isofunctional in a biological context under some criterion should be considered degenerate (1–3). It is implicit in this definition that as long as some biological elements are structurally different yet may often produce a similar output there may arise situations in which these elements lead to different outputs. Identical elements, i.e., structurally redundant ones, cannot do so. A degenerate system, unlike a fully redundant one, is thus extremely adaptable to unpredictable changes in circumstances and output requirements (2). It is no accident that selectional processes such as those that underlie evolution and the immune system show widespread evidence of degeneracy.

The approach taken here stresses a functional definition of both degeneracy and redundancy. Although this functional

approach is in general consistent with structural definitions, it presents certain theoretical advantages. First, by providing a definition of degeneracy in information theoretical terms, it permits the development of general, quantitative measures. Second, by considering degeneracy and redundancy within the same theoretical framework, it allows a precise distinction between the two concepts. Thus, this approach makes it explicit that, whereas a degenerate system appears to be functionally redundant with respect to particular outputs in a particular context, it may perform differently in different contexts.

According to this functional approach, in a system constituted of completely independent elements, each element accounts for a different portion of the entropy of the output units and thus can be said to have an independent function. However, no element can functionally substitute for any other element and the system is extremely brittle. In a fully redundant system, each element shares the same portion of the output entropy. All elements thus can be said to perform the same function. Although that function may be extremely robust, there is no flexibility to accommodate different functions when circumstances change. In a degenerate system, a large number of elements of the system jointly contribute to portions of the entropy of the output units. The system is thus functionally redundant and fault-tolerant with respect to many output functions. At the same time, however, different elements will contribute to different portions of the entropy of the output units, making the system highly adaptive. The relationship between degeneracy and redundancy is therefore the following: to be degenerate, a system must have a certain degree of functional redundancy. However, a completely redundant system will not be degenerate, because the functional differences between distinct elements and thus their ability to contribute independently to a set of outputs will be lost.

Degeneracy and Complexity. As we have mentioned, if one substitutes the mutual information value between each subset

and the output with the entropy value of each subset, the equations defining the two measures are formally identical. Similarly, integration, which measures the decrease of entropy within a system, is formally similar to the redundancy of the system with respect to the output. The close relationship between degeneracy $D_N(X;O)$ and complexity $C_N(X)$ is otherwise not merely formal. Just as $C_N(X)$ is high only if a system is both functionally integrated and functionally segregated, $D_N(X;O)$ is high only if the elements of a system are both functionally redundant and functionally independent with respect to a set of outputs.

To measure degeneracy, perturbation is applied and the mutual information between subsets of the system and the output then is evaluated. This approach considers only genuine causal effects of a system subset on the output and discounts any effects of the output back on the system or any effects caused by statistically correlated input from other subsets. Although the original definition of complexity was based on mutual information, a definition of complexity that takes into account the causal direction of the interaction between subsets of the system is equally warranted. As demonstrated here for degeneracy, this analysis can be done by considering the change in mutual information upon perturbation of each subset in turn (10). Unlike complexity, however, both degeneracy and redundancy as defined here can take on negative values. This may occur if different elements of the system have no effect on the output taken one by one, but have a cooperative effect taken together; for example, if the state of the output is a logical "exclusive or" of the states of two system elements.

Degeneracy $D_N(X;O)$ also is usefully compared with matching complexity $C_M(S;X)$, a measure of how well the connectivity of a system distributes the mutual information between an input sheet S and the system to all subsets of the system (6). The relationship between degeneracy, matching, and complexity has another interesting aspect. In previous work, it was shown that systems whose connectivity is modified to increase matching with a given set of inputs show a concomitant increase in complexity. As described here, systems selected for high degeneracy with a given set of outputs also show a concomitant increase in complexity. This finding suggests that high complexity values may reflect the statistical structure of a complex environment as well as responses to various selective pressures with multiple alternatives yielding adaptive outputs.

Degeneracy and Biological Networks. As with other statistical measures borrowed from information theory, the measure of degeneracy introduced here requires some knowledge of the joint probability distribution of the elements of the system to which it is applied. Moreover, to obtain an estimate of the changes in mutual information caused by perturbation, the effects of perturbation on elements of the network, singly or in conjunction with other elements, must be known. These are demanding requirements on any application to real biological networks. However, within a limited domain, estimates of the correlations among a small subset of biological variables can be obtained and, from these, approximate measures of degeneracy and redundancy can in principle be derived. Such measures also are readily applicable to computer simulations incorporating detailed knowledge of a particular biological system. Because degeneracy and redundancy are defined here as statistical measures, they are meant to refer to sets of possible outputs rather than to a single output. The evaluation of degeneracy and redundancy with respect to a single output may require the development of algorithmic measures corresponding to the statistical measures presented here (cf. ref. 11).

Despite certain limitations on their practical applicability, the measures and distinctions introduced here can be helpful in analyzing and understanding biological networks. Because evolutionary selective pressure typically is applied to a long series of events involving many interacting elements at multiple temporal and spatial scales, it is unlikely that well-defined functions can be neatly assigned to independent subsets of elements or processes in biological networks. For example, if selection occurs for the ability to locomote, connections within and among many different brain structures, such as the spinal cord, brainstem nuclei, the cerebellum etc., in conjunction with parameters of the muscoloskeletal apparatus, are likely to be modified in a degenerate fashion to contribute to this ability. Locomotion will be affected, but many other functions influenced by these structures also will likely be affected in parallel, resulting in a concomitant increase in the degeneracy of the system.

A manifest consequence of degeneracy in the nervous system is that neurological lesions often may appear to have little effect, at least within a familiar context. For example, for a long time it was thought that the section of the corpus callosum—hundreds of millions of reentrant neural fibers joining the two hemispheres—has little or no consequence on behavior—a striking example of apparent redundancy. However, careful experiments have revealed subtle and not-so-subtle deficits of split brain patients (12), indicating that the reentrant connections between the two hemispheres provide not just redundancy but also degeneracy. Similarly, the deletion of a particular gene in so-called knockout experiments often has no apparent phenotypic consequence. On the other hand, changes in the context may reveal functionally important interactions (e.g., ref. 13). Such examples indicate that degeneracy is not just a likely consequence of mutational and selectional events acting on integrated and multilevel systems, but also that, in turn, it provides a rich substrate for selective events. The ability of natural selection to give rise to a large number of nonidentical structures capable of producing similar functions appears to increase both the robustness of biological networks and their adaptability to unforeseen environments by providing them with a large repertoire of alternative functional interactions.

1. Edelman, G. M. (1978) in *The Mindful Brain*, eds. Edelman, G. M. & Mountcastle, V. B. (MIT Press, Cambridge, MA), pp. 51–100.
2. Edelman, G. M. (1987) *Neural Darwinism: The Theory of Neuronal Group Selection* (Basic Books, New York).
3. Edelman, G. M., ed. (1974) in *Cellular Selection and Regulation in the Immune Response* (Raven, New York), pp. 1–37.
4. Kolb, B. & Whishaw, I. Q. (1996) *Fundamentals of Human Neuropsychology* (Freeman, New York).
5. Tononi, G., Sporns, O. & Edelman, G. M. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 5033–5037.
6. Tononi, G., Sporns, O. & Edelman, G. M. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 3422–3427.
7. Shannon, C. E. & Weaver, W. (1949) *The Mathematical Theory of Communication* (Univ. of Illinois Press, Chicago).
8. Papoulis, A. (1991) *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, New York).
9. Tononi, G., Edelman, G. M. & Sporns, O. (1998) *Trends Cognit. Sci.* **2,** 474–484.
10. Tononi, G. & Edelman, G. M. (1998) *Science* **282,** 1846–1851.
11. Gell-Mann, M. & Lloyd, S. (1996) *Complexity* **2,** 44–52.
12. Gazzaniga, M. S. (1995) *Neuron* **14,** 217–228.
13. Nakao, N., Hiraiwa, N., Yoshiki, A., Ike, F. & Kusakabe, M. (1998) *Am. J. Pathol.* **152,** 1237–1245.