Degeneracy, Redundancy & Complexity in Biological Systems & Their Measures

Qing-jun Wang

When I just came to experimental biology from theoretical physics five years ago, Dr. Ladislav Nedbal, my mentor, gave me an advice I remember since then, "you must treat them as living things". With more time spent in biological fields, I understand and appreciate it more and more. Biology deals with living things, which can never be satisfactorily explained in a mechanical and deterministic way.

A good portion of biological complexity comes from degeneracy. Degeneracy is used to designate different wave function satisfying the same energy state in quantum mechanics. In biology, degeneracy is "the ability of elements that are structurally different to perform the same function". However, for many years, the concept of degeneracy is lacking and confused with redundancy, which occurs when the same function is performed by identical elements. Unlike redundant elements, degenerate elements can produce different outputs under different conditions.

In engineering, redundancy refers to duplication of elements within electric or mechanical components to provide additional power for protection from failure or the repetition of messages sending to decrease transmission error. Despite the commonness of redundancy in engineering, true redundancy in biological system is rarely seen due to the rareness of the presence of identical elements.

On the contrast, examples of degeneracy exist in all fields and all levels of biology, which were discounted as redundancy (Figure 1). In immunology, different antibodies can bind to the same antigen. In neurobiology, a large number of different brain structures can influence the same motor output and a constrained set of signaling events can be brought about by a large number of different combinations of stimuli. In a cellular level, genetic code is degenerate with different codons coding for the same amino acid; transcription of a gene is degenerate with different 5' start site, 3' termination site, and degenerate transcription machinery; translation is degenerate with alternative splicing; protein folding is degenerate with different primary sequences lead to similar protein structures and functions; enzyme activity is degenerate with different proteins catalyzing the same enzyme reaction; metabolism is degenerate with the existence of multiple parallel anabolic and catabolic pathways. In a multicellular level, many different patterns of muscle contraction yield equivalent outcome; neural connectivity is highly degenerate in that although no two neural cells within an individual are identical. In the level concerning the interaction of an individual with other individuals and the environment, enormous kinds of diets are equivalent; although no two "equivalent" neurons taken form two individuals have exactly the same morphology, the two individuals generally behave the same; there are a large number different ways to transmit messages among individuals. A list of degeneracy at different levels of biological organization is shown in Figure 1.

With the widespread of degeneracy in biological system, a question has to be asked is why. It is argued that "degeneracy is a necessary accompaniment of natural selection". However, it is "not a property simply selected by evolution, but rather is a prerequisite for and an inescapable product of the process of natural selection" because in an evolution system there is no *a priori* design on how to survive and any change, such as

mutation or interaction with the environment, is possible to lead to strong selection. In the absence of degeneracy, there are not many chances for changes to be beneficial since there is only one correct way which not many may be lucky enough to find.

To better understand biological complexity, to clarify the concepts of degeneracy and redundancy, and to make these concepts more easily to operate, several measures based on information theory are introduced. These measures are applied to neural examples in the papers. However, they are equally applicable to any (biological) complex system.

Consider an neural system X with n elementary components. Assume that its activity is described by a Gaussian stationary multidimensional stochastic process, i.e. the dynamic interactions between these elements do not change with time. No assumption is made about messages, codes, or noisy channels. The joint probability density function of this system can be characterized in terms of entropy (H) and mutual information (MI). If the components of the system are independent, entropy is maximal. Intrinsic constraints can make the components deviate from statistical independence and reduce entropy. Mutual information is used to measure the deviation. For a bipartition of the system X into a j-th subset X_j^k composed of k components and its complement $X- X_j^k$, the mutual information be X_j^k and $X-X_j^k$ is

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

where $H(X_j^k)$ and $H(X- X_j^k)$ are the entropies of the two subsets X_j^k and $X- X_j^k$, and H(X) is the entropy of the whole system. MI=0 if the two subsets X_j^k and $X- X_j^k$ are statistically independent and MI>0 if they aren't, i.e. MI≥0.

The concept of mutual information can be generalized to express the deviation from independence among the n components of a system X by means of a single measure — integration I(X).

$$I(X) = \Sigma_{i=1}^{n} H(x_i)-H(X), \quad (2.1)$$

where x_i is the i-th individual component. For a bipartition,

$$I(X) = I(X_j^k)+I(X- X_j^k)+MI(X_j^k;X- X_j^k). \quad (2.2)$$

The average integration for subsets of size k is denoted as $< I(X_j^k)>$.

$$<I(X_j^n)> = I(X),$$
$$< I(X_j^1)> = 0,$$
$$I(X) \geq I(X_j^k)+I(X- X_j^k) \geq 0,$$
$$<I(X_j^{k+1})> \geq <I(X_j^k)>.$$

Define complexity $C_N(X)$ of a system X without receiving signal from the environment as

$$C_N(X) = \Sigma_{k=1}^{n} [(k/n)I(X)- <I(X_j^k)>], \quad (3.1)$$
$$C_N(X) \geq 0.$$

$C_N(X)$ can also be expressed as

$$C_N(X) = \Sigma_{k=1}^{n} [<H(X_j^k)>-(k/n)H(X)] \quad (3.2)$$
$$C_N(X) = \Sigma_{k=1}^{n/2} <MI(X_j^k;X- X_j^k)>. \quad (3.3)$$

$C_N(X)$ is high when the integration of the system is high and at the same time the average integration for small subsets is lower than would be expected from a linear increase over increasing subset size. Or $C_N(X)$ is high when the average mutual information between any subset of the system and its complement is high.

A related measure, matching complexity (C_M) is defined to reflect the change in complexity (C_N) upon receiving signals from the environment. It is a measure of how well the connectivity of a system distributes the mutual information between the input set and the system to all subsets of the system.

$$C_M(X;S_i) = C_N^T(X) - C_N^I(X) - C_N^E(X), \quad (4.1)$$

where $C_N^T(X)$ is total complexity when the system samples a stimulus, $C_N^I(X)$ is intrinsic complexity when the system is isolated, $C_N^E(X)$ is extrinsic complexity due to the stimulus *per se*, and S_i is the i-th stimulus. Alternatively,

$$C_M(X;S_i) = \Sigma_{k=1}^{n} <MI^T(X_j^k;S_i)> - <MI^E(X_j^k;S_i)>. \quad (4.2)$$

Degeneracy (D_N) is defined when the whole system is partitioned into output O and other subsets X_j^k and only the causal effects of X_j^k on O through forward connection is considered, regardless of the effect of O on X_j^k through backward connections or the effect of another subset providing statistically correlated inputs to both X_j^k and O.

$$D_N(X;O) = \Sigma_{k=1}^{n} [<MI^P(X_j^k;O)> - (k/n)MI^P(X;O)], \quad (5.1)$$

where mutual information is obtained when uncorrelated random perturbations are applied.

$D_N(X)$ is high when the mutual information between the whole system and the output is low and at the same time the average mutual information between small subsets of the system and the output is higher than would be expected from a linear increase over increasing subset size. Alternatively, for bipartitioned system X,

$$D_N(X;O) = 0.5 \times \Sigma_{k=1}^{n} <MI^P(X_j^k;X - X_j^k;O)>. \quad (5.2)$$

Redundancy is defined as

$$R(X;O) = \Sigma_{j=1}^{n} [<MI^P(X_j^1;O)>] - MI^P(X;O). \quad (6)$$

Redundancy is zero when all the elements of the system contribute to the output independently.

The relation between degeneracy and redundancy can be expressed as

$$D_N(X;O) = \Sigma_{k=1}^{n} [(k/n)R(X;O) - <R(X_j^k;O)>] \quad (7)$$

Comparing equations (3.2) and (5.1), the definitions for complexity and degeneracy are formally identical if the mutual information between each subset and the output in degeneracy is substituted with entropy. Moreover, comparing equations (2.1) and (6), the definitions of integration and redundancy are formally similar. In addition, the relationship between complexity and integration (equation 3.1) and the relationship between degeneracy and redundancy (equation 7) are formally identical.

$C_N(X)$ is high only if a system is both functional integrated and functional segregated. Similarly, $D_N(X;O)$ is high only if the elements of a system are both functionally redundant and functionally independent with respect to the output set.

Results from computer simulation of the effects of system connectivity on redundancy and degeneracy (Figure 2) show that a total independent system (top panel) does not zero degeneracy (indicated in lanes C-E as a shaded area) and zero redundancy; a fully connected network (bottom panel) has very high redundancy and relatively low degeneracy because different combinations of the elements have similar effect on the output; and a degenerate system (middle panel) have the highest degeneracy because different combination of the elements can affect the output in a similar way and at the same time have independent effects.

Results from computer simulation of the effects of system connectivity on degeneracy and complexity (Figure 3) show that random network (top panel) has low degeneracy and low complexity. It also gives uniform patterns of correlation (lane B). On the contrary, high connectivity (bottom panel) yields high degeneracy and high complexity. It also gives strong patterns of correlation, with connections that support higher degeneracy strengthened while others weakened.

In summary, it is a great discovery to distinguish degeneracy from redundancy in biological system, It is a great idea to quantify biological redundancy, degeneracy and complexity. However, how to obtain the knowledge of joint probability distribution of the elements of a biological system and the changes in mutual information upon perturbation is a great challenge.

REFERENCES

1. Edelman GM & Gally JA. (2001) Degeneracy and complexity in biological systems. PNAS. 98:13763-13768.
2. Tononi G., Sporns O., & Edelman G. (1999) Measures of degeneracy and redundancy in biological networks. PNAS. 96:3257-3262.
3. Tononi G., Sporns O., & Edelman G. (1996) A complexity measure for selective matching of signals by the brain. PNAS. 93:3422-3427.
4. Tononi G., Sporns O., & Edelman G. (1994) A measures for brain complexity: relating functional segregation and integration in the nervous system. PNAS. 91:5033-5037.

Figure 1.

Table 1. Degeneracy at different levels of biological organization

1. Genetic code (many different nucleotide sequences encode a polypeptide)
2. Protein fold (different polypeptides can fold to be structurally and functionally equivalent)
3. Units of transcription (degenerate initiation, termination, and splicing sites give rise to functionally equivalent mRNA molecules)
4. Genes (functionally equivalent alleles, duplications, paralogs, etc., all exist)
5. Gene regulatory sequences (there are degenerate gene elements in promoters, enhancers, silencers, etc.)
6. Gene control elements (degenerate sets of transcription factors can generate similar patterns of gene expression)
7. Posttranscriptional processing (degenerate mechanisms occur in mRNA processing, translocation, translation, and degradation)
8. Protein functions (overlapping binding functions and similar catalytic specificities are seen, and ''moonlighting'' occurs)
9. Metabolism (multiple, parallel biosynthetic and catabolic pathways exist)
10. Food sources and end products (an enormous variety of diets are nutritionally equivalent)
11. Subcellular localization (degenerate mechanisms transport cell constituents and anchor them to appropriate compartments)
12. Subcellular organelles (there is a heterogeneous population of mitochondria, ribosomes, and other organelles in every cell)
13. Cells within tissues (no individual differentiated cell is uniquely indispensable)
14. Intra- and intercellular signaling (parallel and converging pathways of various hormones, growth factors, second messengers, etc., transmit degenerate signals)
15. Pathways of organismal development (development often can occur normally in the absence of usual cells, substrates, or signaling molecules)
16. Immune responses (populations of antibodies and other antigen-recognition molecules are degenerate)
17. Connectivity in neural networks (there is enormous degeneracy in local circuitry, long-range connections, and neural dynamics)
18. Mechanisms of synaptic plasticity (changes in anatomy, presynaptic, or postsynaptic properties, etc., are all degenerate)
19. Sensory modalities (information obtained by any one modality often overlaps that obtained by others)
20. Body movements (many different patterns of muscle contraction yield equivalent outcomes)
21. Behavioral repertoires (many steps in stereotypic feeding, mating, or other social behaviors are either dispensable or substitutable)
22. Interanimal communication (there are large and sometimes nearly infinite numbers of ways to transmit the same message, a situation most obvious in language)

Figure 2.

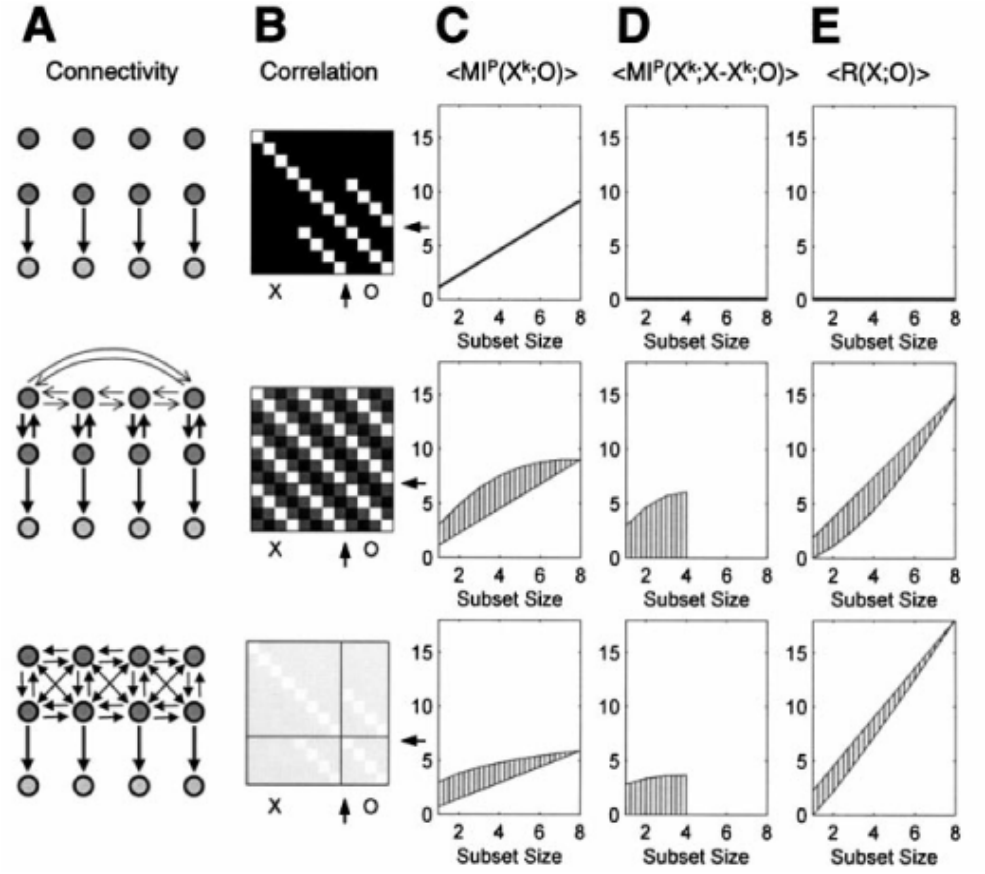

Figure 3.