Patterns of information storage and senescence.

Jonathan D.H. Smith

Department of Mathematics
Iowa State University
Ames, Iowa 50011-2064, U.S.A.
E-mail: jdhsmith@math.iastate.edu
URL: http://orion.math.iastate.edu/jdhsmith/homepage.html

Abstract: Against the background of increasing entropy ordained by the Second Law, we discuss the abstract options for information storage and order. General patterns of growth, homeostasis, senescence and demise emerge. We propose simple formal models of these behaviors, and examine some examples. The central idea is the so-called Grouping Principle, identifying the grouping of microstates into macrostates as the key determinant of long-term behavior. In particular, a model of abstract oncogenesis suggests that a small change in the grouping may convert homeostasis into unconstrained growth.

---

1 Summary

A medium for the storage and transmission of information is an essential prerequisite for organization. Although the existence of such a medium does not guarantee that meaningful information and order are present, the absence of such a medium would certainly preclude any possibility of them. The medium represents a necessary (but not sufficient) capacity for order. Collapse of the capacity by the kind of senescence and demise described below is one of the inherent limitations on the lifetime of an information system. Our primary concern is how such a capacity can evolve in the first place, and, once established, how will it last?

Information tokens are modelled here as the outcomes of an "experiment". For example, the numbers from 1 to 6 might arise from rolling a die. The significance of an information token depends on its probability as an outcome of such an experiment. The minimum storage required to record the outcomes of the experiment is measured, in terms of the probabilities of its outcomes, by the entropy of the experiment. A brief, direct approach to the concepts of entropy, probability and randomness, as they pertain to the analysis of complex systems, is given in Sections 2 and 3 (compare [8]).

Systems complex enough to embody self-organization require energy for their maintenance, and are thus regarded as dynamical systems. Section 4 outlines the mathematical formulation of a dynamical system, by a probability-preserving operator $T : X \to X$ on a phase space X. The map T represents the passage of a discrete time unit, and the elements of the phase space represent the possible histories of an individual particle. As time passes, entropy increases (14), albeit at a decreasing rate (15). The capacity for organization arises when a system's microstates are grouped into macrostates. The micro- and macro-entropies are plotted against time in the **Two Curves Graph**. Both curves are non-decreasing (14) and concave (15). However, the macro-entropy is never greater than the micro-entropy (16). The difference between the two represents the capacity for information storage and organization.

Section 7 presents a basic class of models exhibiting the Two Curves Graph, namely macrostates grouping microstates that are the states of a finite, stationary Markov chain. One of the simplest possible models has four states, visited cyclically in sequence as time passes (18). There are two non-trivial groupings of microstates into macrostates, grouping either adjacent or opposite pairs of microstates on the cycle into macrostates. In the opposite grouping, the storage capacity grows to 1 bit after 1 time unit, and is then maintained. In the adjacent grouping, the capacity again grows to 1 bit after 1 time unit, but then drops back to zero after the next time unit. The distinction between these two behaviors is the simplest illustration of the **Grouping Principle** that the capacity for information storage and organization depends critically on the grouping of microstates into macrostates. Subsequent sections illustrate some typical behaviors exhibited by this capacity: **stasis** as in the opposite grouping and Section 11, **senescence** as in Section 12, **demise** as in the adjacent grouping and Section 13, and **growth** as in Section 14. The picture of stasis provided by Section 11 stands in sharp contrast to Newton's Law, "Every body continues in its state of rest or uniform motion unless acted on by an external force." Newtonian stasis is passive — the absence of action. The stasis represented by Section 11 is rather an actively maintained homeostasis in the biological sense.

The concluding section offers some brief observations on biological interpretation and application of the models that have been presented, including mechanisms for senescence and neoplasia.

2 Entropy, probability, and randomness

Consider a scientific experiment which may have any one of N possible outcomes. For example, $N = 6$ for the experiment of rolling a die. By convention of modern science, a "scientific" experiment has to be repeatable and reproducible. The experiment is said to be *random* if no statistical test available to the experimenter can detect any pattern in repeated outcomes. It is very important to note that this definition is contingent on the power of the apparatus available. Rolling a die under casino conditions should be random. On the other hand, if equipment such as a high speed precision camera was available, it would be possible to predict the outcome of each roll from the initial motion of the die, and the experiment would no longer be random. A subtler distinction arises when rolling a very slightly loaded die under casino conditions. If the number of rolls required for even a sophisticated statistical test to reveal the bias is in excess of the number of rolls sufficient to wear the spots off the die, then the experiment would still qualify as random. The basic concepts of entropy and probability follow from the concept of randomness. For a random experiment with N outcomes, the (*natural*) *entropy* is

$$H = \log N \tag{1}$$

(using natural logarithms) or the (*binary*) *entropy*

$$H = \log_2 N \quad \text{bits} \tag{2}$$

using logarithms to base 2 and bits as units. (Occasionally logarithms to base 10 are used, with Hartleys as units.) The *probability* $\pi(x)$ of any one particular outcome x of the random experiment is

$$\pi(x) = N^{-1}. \tag{3}$$

The probability $\pi(x)$ and natural entropy H are connected by the mutually inverse relationships

$$\text{(a)} \qquad H = -\log \pi(x)\,; \qquad \text{(b)} \qquad \pi(x) = \exp(-H)\,.$$

Using binary entropy H, these take the form

$$\text{(a)} \qquad H = -\log_2 \pi(x)\,; \qquad \text{(b)} \qquad \pi(x) = 2^{-H}\,. \tag{4}$$

The probability $\pi(x)$ represents a fair stake to buy into the following game: win one unit if the outcome of the experiment is x. Randomness of the experiment means that there are no winning strategies in this game. The entropies (1) and (2) measure one's ignorance about the outcome of the experiment. If you let someone else run the experiment, and instead question them afterwards as to what the outcome was, then the minimum number of yes/no questions required to elicit the outcome would be given by (2).

3   Non-random experiments

The formulae of Section 2 are too narrow to be of general use, where one wishes to deal with non-random experiments. These may be modelled using an underlying random experiment whose set of outcomes, called the *phase space*, has N elements. The phase space is completely partitioned into a set $\xi = \{C_1, \dots, C_r\}$ of mutually exclusive subsets called *states*. The partition ξ represents the non-random experiment (also denoted ξ) of sampling an outcome x from the phase space and locating the state C_i in which it lies. If the state C_i contains n_i outcomes of the underlying random experiment, each of whose outcomes has probability N^{-1} according to (3), then the *probability* $p(C_i)$ of the state C_i is given as

$$p(C_i) = n_i N^{-1}.$$

The state C_i may be regarded as a random experiment in its own right: select an outcome from C_i. The entropy of this random experiment, according to (1), is

$$H(C_i) = \log n_i.$$

If you perform experiment ξ and obtain the result C_i, then your ignorance will have been reduced by $\log N - \log n_i = -\log p(C_i)$. This happens with probability $p(C_i)$. Thus the average loss of ignorance or gain in knowledge obtained on performing experiment ξ is its *entropy*

$$H(\xi) = -\sum_{i=1}^{r} p(C_i) \log p(C_i). \tag{5}$$

(Of course, one may take logarithms to base 2 and quote $H(\xi)$ in bits.) The mathematical discipline of measure theory extends the definitions of probability and entropy to appropriate infinite phase spaces, where "counting outcomes" may be replaced by "measuring volumes". The entropy $H(\xi)$ satisfies the inequality

$$0 \le H(\xi) \le \log r. \tag{6}$$

Equality obtains on the left in (6) if and only if $p(C_i) = 1$ for some i: if you already know in advance that ξ will come up with state C_i, then you gain no knowledge by performing the experiment. Equality obtains on the right in (6) if and only if $p(C_i) = r^{-1}$ for each i: the most informative experiments are those designed so that all their different outcomes are equally likely. In particular, randomness of an experiment is characterized by its entropy. Moreover, the three concepts of **entropy**, **probability**, and **randomness** turn out to be equivalent. Each leads to the other two. In complexity theory, it is an entropy measure, namely the self-delimiting algorithmic complexity, which is usually taken as basic. Probability is then obtained via (4) [10]. The complexity class of the algorithms invoked corresponds to the power of the apparatus used in the statistical tests for randomness discussed in Section 2.

4 Conditional entropy and dynamical sytems

Let X be a phase space. Two experiment $\xi = \{C_1, \ldots, C_r\}$ and $\eta = \{D_1, \ldots, D_s\}$ on X have a common *refinement*

$$\xi \vee \eta = \{C_i \cap D_j | 1 \le i \le r, \ 1 \le j \le s\}.$$

The *conditional probability* of C_i given D_j is defined to be

$$p(C_i|D_j) = p(C_i \cap D_j)/p(D_j) \tag{7}$$

for $p(D_j) \ne 0$ and 0 otherwise. If the experiment η has already yielded state D_j, then (7) represents the probability that ξ will yield state C_i. The *conditional entropy* of ξ given η is defined as

$$H(\xi|\eta) = -\sum_{j=1}^{s} p(D_j) \sum_{i=1}^{r} p(C_i|D_j) \log p(C_i|D_j). \tag{8}$$

It satisfies the equality

$$H(\xi|\eta) = H(\xi \vee \eta) - H(\eta), \tag{9}$$

and thus represents the additional information gained as a result of experiment ξ if the result of experiment η is already known. Given a third experiment $\zeta = \{E_1, \ldots, E_t\}$, the fundamental inequality

$$H(\xi|\eta \vee \zeta) \le H(\xi|\eta) \tag{10}$$

holds. It may be interpreted as saying: the more you know in advance, the less new information any subsequent experiment can bring you. It is sometimes useful to be able to consider the entropy $H(\xi)$ of (5) as a conditional entropy $H(\xi|\eta)$ as in (8). This may be achieved by taking η to be the singleton partition $\{X\}$. The sum (8) then reduces to (5), so that

$$H(\xi|\{X\}) = H(\xi).$$

A *dynamical system* is an invertible, probability-preserving map $T : X \to X$ of a phase space X. Intuitively, the space X may be viewed as the set of all possible histories of particles. The map T represents the passage of (discrete) time: if an element x of X describes the history of a certain particle, then the element xT of X describes the history of a second particle which is always one time unit ahead of the first. Thus if the membership $x \in C_i$ means that the first particle is in state C_i at time 0, the equivalent memberships $xT \in C_i$ or $x \in C_i T^{-1}$ mean that the second particle is in state C_i at time 0, or in other words that the first particle is in state C_i at time 1. The *history*

$$\xi^k = \xi \vee \xi T^{-1} \vee \cdots \vee \xi T^{1-k} \tag{11}$$

represents performing the experiment ξ over a period of $k-1$ time units and recording the evolution of the system. Preservation of probability by T means that, for any state C_i, one has

$$p(C_i) = p(C_i T). \tag{12}$$

This is the expression of Liouville's Theorem: in classical physics it corresponds to a Hamiltonian system, and in quantum physics to a unitary evolution operator. As a consequence of (12), one obtains the preservation

$$H(\xi) = H(\xi T) \tag{13}$$

of the entropy of an experiment ξ.

5 The Two Curves Graph

A *hierarchical dynamical system* is a dynamical system $T : X \to X$ with two experiments ξ and η such that ξ is a refinement of η, i.e. $\xi \vee \eta = \xi$. The states of ξ are known as *microstates* and the states of η are known as *macrostates*. Thus each macrostate is a grouping of microstates. A history ξ^k as in (11) is called a *microhistory*. It is a refinement of the corresponding *macrohistory* η^k. The non-negative real-valued functions $H(\xi^k)$ and $H(\eta^k)$ of the discrete time parameter k are respectively described as the *micro-entropy* and the *macro-entropy*. The major characteristic of hierarchical dynamical systems is the **Two Curves Graph**, on which the micro- and macro-entropies are plotted together against time. Several examples are displayed in Sections 9 and 11–14. The key features of the Two Curves Graph are as follows:

1. Both entropy curves should be **non-decreasing**;

2. Both curves should be **concave**;

3. The difference between the micro- and macro-entropy should always be **non-negative**.

These features and their interpretation are discussed in detail in References [1]–[6]. Briefly summarizing that discussion, one may characterize the monotonicity (1) of the entropy curves as an expression of the Second Law of Thermodynamics. The concavity (2) is interpreted as the effect of historical constraints retarding the rate of entropy increase (systems with memory are less random or more predictable than memoryless systems, since potential new states inconsistent with the history are inaccessible). The non-negative difference (3) between the micro- and macro-entropies is interpreted as a measure of the **capacity for organization or information storage** within the hierarchical dynamical system. Brooks and Wiley wrote that

> Information is defined ... as the difference between the maximum possible entropy and the entropy of the observed state of a system

[1]. As discussed at the beginning, it is more precise in this context to refer to the *information capacity*, necessary rather than sufficient for meaningful information.

6 Hierarchical dynamical systems

All the features of the Two Curves Graph are demonstrated mathematically within the hierarchical dynamical system model.

1. Using (9) and the basic inequality (10), one has

$$H(\xi^{k+1}) - H(\xi^{k+1}) = H(\xi^{k+1}|\xi^{k+1}) = H(\xi^{k+1}|\xi^k \vee \xi T^{-k}) \leq H(\xi^{k+1}|\xi^k) = H(\xi^{k+1}) - H(\xi^k).$$

Subtracting $H(\xi^{k+1})$ from each side and multiplying by -1 gives the inequality

$$H(\xi^k) \leq H(\xi^{k+1}) \tag{14}$$

showing that the **micro-entropy is non-decreasing**. A similar calculation shows that the **macro-entropy is non-decreasing**.

2. Using (9), (10) and (13), one has

$$\begin{aligned}
H(\xi^{k+1}) - H(\xi^k) &= H(\xi T^{-k}|\xi^k) = H(\xi T^{-k}|\xi \vee \xi^{k-1}T) \\
&\leq H(\xi T^{-k}|\xi^{k-1}T) = H(\xi^{k-1}T \vee \xi T^{-k}) - H(\xi^{k-1}T) \\
&= H(\xi^{k-1} \vee \xi T^{1-k}) - H(\xi^{k-1}) = H(\xi^k) - H(\xi^{k-1}),
\end{aligned}$$

i.e. the inequality

$$H(\xi^{k+1}) - H(\xi^k) \leq H(\xi^k) - H(\xi^{k-1}) \tag{15}$$

showing that the **micro-entropy is a concave function of time**. A similar calculation shows that the **macro-entropy is concave**.

3. Using (9) and the non-negativity of conditional entropies that is immediate from their definition (8), one has

$$H(\xi^k) = H(\xi^k \vee \eta^k) = H(\eta^k) + H(\xi^k|\eta^k) \geq H(\eta^k),$$

so that

$$H(\eta^k) \leq H(\xi^k). \tag{16}$$

The **non-negative difference** $H(\xi^k) - H(\eta^k)$ measures the capacity for organization.

7 Markov chains

Some of the simplest hierarchical dynamical systems illustrating the Two Curves Graph as presented above come from finite stationary Markov chains. The microstates $C_1, \ldots, C_r$ are the states of the chain. Stationarity corresponds to equation (12). The microhistory

$$\xi^2 = \xi \vee \xi T^{-1} = \{C_i \cap C_j T^{-1} | 1 \leq i, j \leq r\}$$

determines the transition matrix as follows. The (i, j)-th coefficient of the transition matrix, giving the probability of passing to state C_j in one time unit conditional on starting from state C_i, is specified according to (7) as

$$p(C_j T^{-1} | C_i) = p(C_i \cap C_j T^{-1}) \Big/ \sum_{k=1}^{r} p(C_i \cap C_k T^{-1}) \,.$$

Conversely, if the transition matrix is the stochastic matrix

$$\Pi = (\pi_{ij})_{r \times r} \,,$$

so that

$$\sum_{j=1}^{r} \pi_{ij} = 1$$

for $1 \leq i \leq r$, and if there is a non-negative eigenvector

$$\pi = [\pi_1 \pi_2 \ldots \pi_r] \tag{17}$$

of Π such that

$$\pi \Pi = \pi \,,$$

then the probabilities of the microhistories are determined. Namely

$$p(C_i) = \pi_i \,,$$

while

$$p(C_{i_1} \vee C_{i_2} T^{-1} \vee \cdots \vee C_{i_k} T^{1-k}) = \pi_{i_1} \pi_{i_1 i_2} \ldots \pi_{i_{k-1} i_k} \,.$$

The micro-entropy values may then be calculated by (5).

8   The Grouping Principle

This principle summarizes the most important characteristic of hierarchical dynamical systems:

> **The self-organization inherent in a hierarchical dynamical system depends critically on the grouping of microstates into macrostates.**

The simplest illustration of the Grouping Principle is obtained from a finite stationary Markov chain whose transition matrix is the permutation matrix

$$\Pi = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{18}$$

of a cycle of length 4. The eigenvector of (17) is

$$\pi = \begin{bmatrix} \dfrac{1}{4} & \dfrac{1}{4} & \dfrac{1}{4} & \dfrac{1}{4} \end{bmatrix}.$$

There will be two macrostates, denoted by $\triangle$ and $\square$, each comprising two microstates. The *opposite grouping* groups diametrically opposite microstates on the cycle,
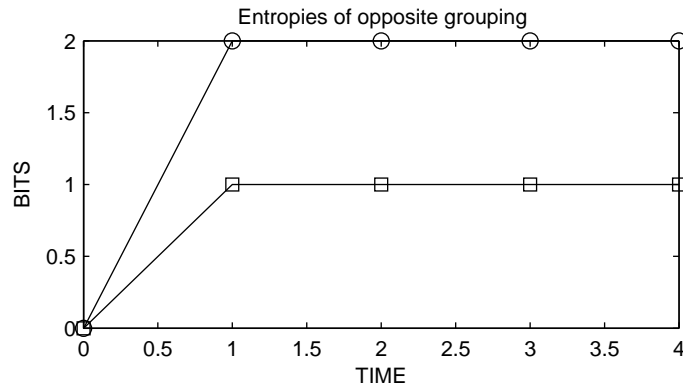
$$
\begin{array}{ccc}
\square & \longrightarrow & \triangle \\
\uparrow & & \downarrow \\
\triangle & \longleftarrow & \square
\end{array}
$$
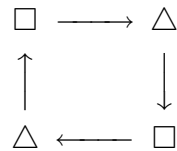
while the *adjacent grouping* groups adjacent microstates on the cycle:

$$
\begin{array}{ccc}
\triangle & \longrightarrow & \triangle \\
\uparrow & & \downarrow \\
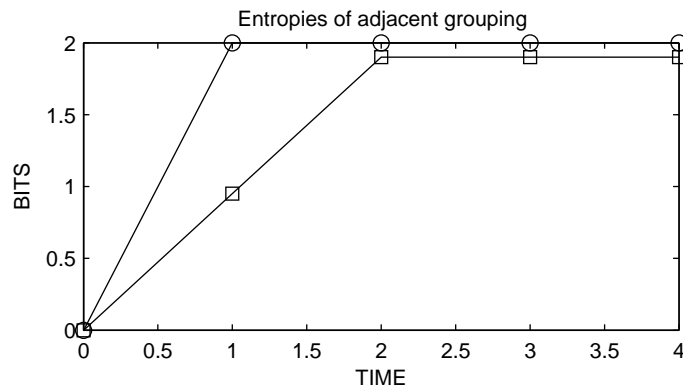\square & \longleftarrow & \square
\end{array}.
$$
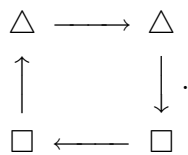
9 Adjacent and opposite groupings

The corresponding macro- and micro-entropies for the groupings of Section 8 are plotted below. Binary entropies have been used to give integer values. In the **opposite grouping**,

the capacity for organization grows to 1 bit after 1 time unit, and is then maintained at that value. In the **adjacent grouping**,

the capacity for organization again grows to 1 bit after 1 time unit, but then dies back down to zero after the next time unit.

In the language of the following sections, the adjacent grouping displays **demise**, while the opposite grouping displays **stasis**.

10   Model behaviors

Consider a long sequence of symbols from a finite set. Such a sequence . . .

- . . . may be the result of discretizing a time series;

- . . . may describe the behavior of an automaton running through different states, the n-th symbol in the sequence locating the state of the automaton at time n;

- . . . may arise from symbolic dynamics in chaos [7], or by the related indexing procedure used in the subsequent sections;

- . . . may be a DNA sequence, or a codon sequence;

- . . . may come from a finite Markov chain.

For example, the sequence

$$0\ 1\ 2\ 3\ 0\ 1\ 2\ 3\ 0\ 1\ 2\ 3\ 0\ 1\ \ldots$$

corresponds to the Markov chain (18). The symbols of the finite set determine microstates. A microhistory as in (11) is then determined by k-tuplets of successive symbols in the sequence. Probabilities may be simulated by relative frequencies within the sequence, and a micro-entropy may then be calculated on the basis of these relative frequencies. (In certain cases it may be appropriate to start with r-tuplets from the original sequence as the basic symbols or microstates. For instance, codons may be more appropriate than base pairs from a DNA sequence. To achieve subsequent non-trivial groupings of microstates into macrostates with a sequence of zeroes and ones from symbolic dynamics in chaos, one will have to start at least with doublets.)

Now suppose that the symbols or microstates are grouped into macrostates. Using relative frequencies to simulate probabilities again, a macro-entropy can be calculated. The micro- and macro-entropies may then be plotted on a Two Curves Graph. Of course, in concrete cases the relevance of the interpretation of the difference between the micro- and macro-entropies as a measure of the capacity for organization will depend on the significance of the grouping used. The kinds of behavior to be expected are illustrated by the examples in the following sections. These examples are the result of computer experimentation conducted by Dan Ashlock.

11 Stasis

Stasis occurs if an approximately constant difference between the entropies is maintained for a considerable time. The following graph displays entropy counts for microstates obtained from cyclic repeats of the sequence 0000000001211212212333333333, while the two corresponding macrostates are $\{0, 1, 3\}$ and $\{2\}$. The homeostatic state has a storage capacity of just under 1.5 bits
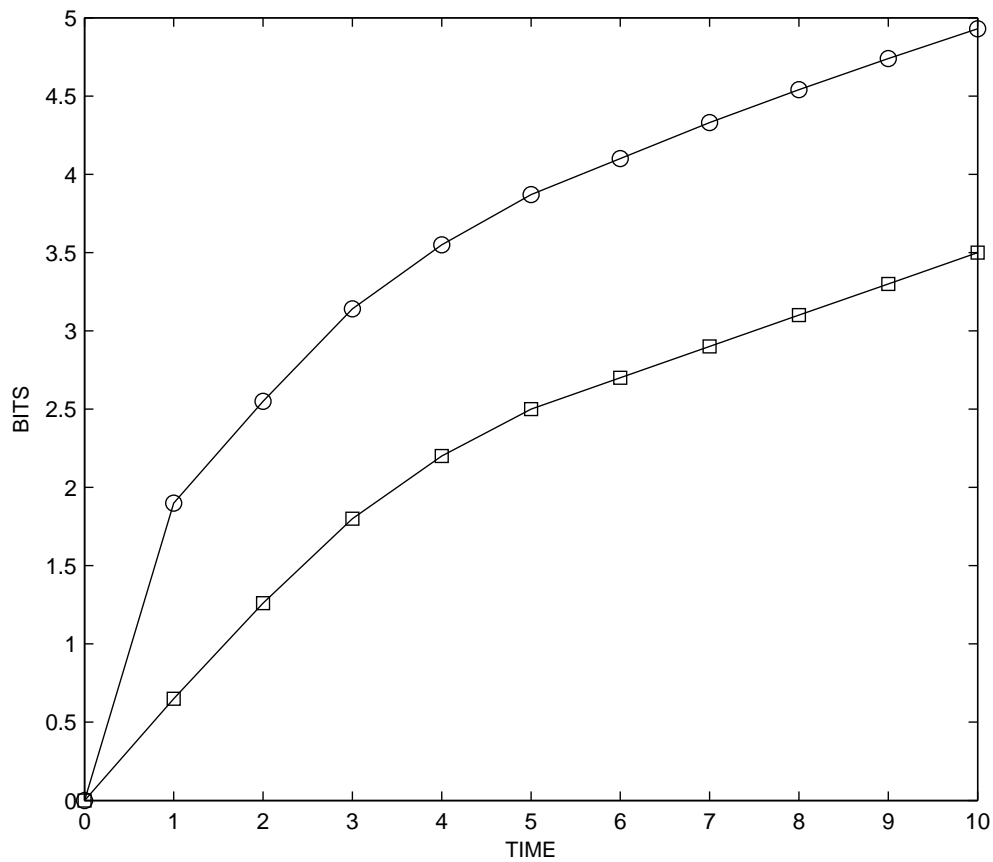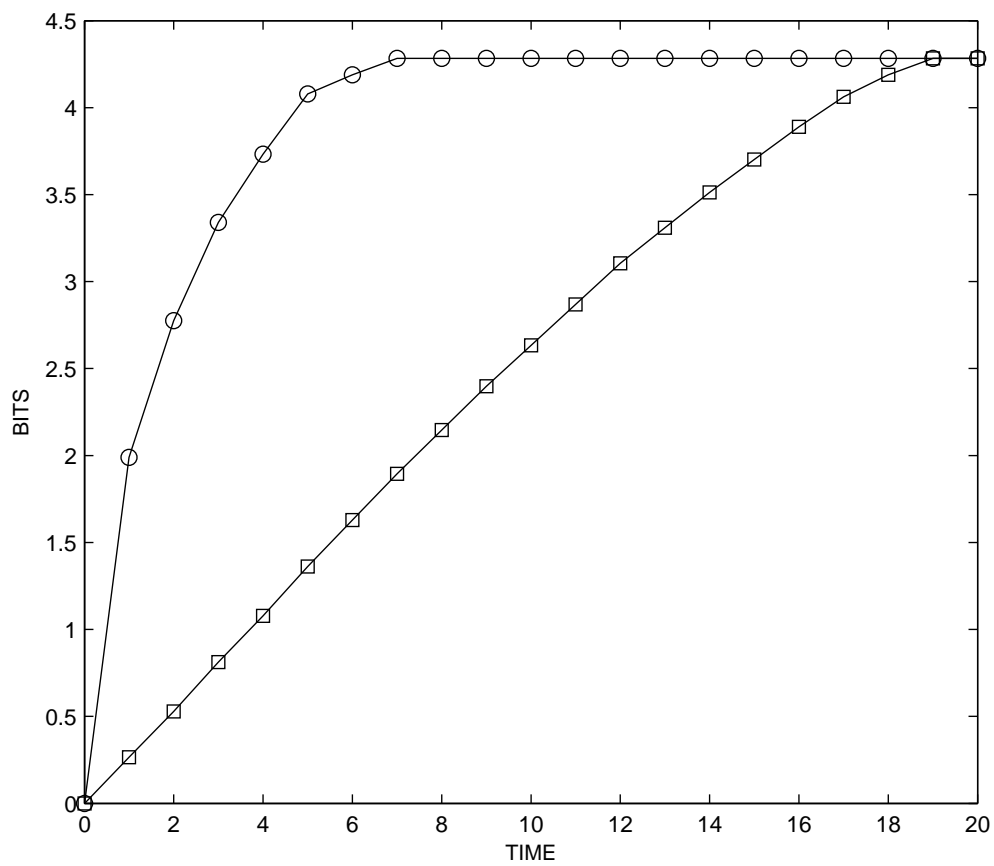
12 Senescence

Senescence occurs if the difference grows for a time, but then decays. The following graph displays entropy counts for microstates that have been obtained from cyclic repeats of the sequence 0000000000011111111122223333344, while the three corresponding macrostates are $\{0, 2, 4\}$, $\{1\}$ and $\{3\}$.
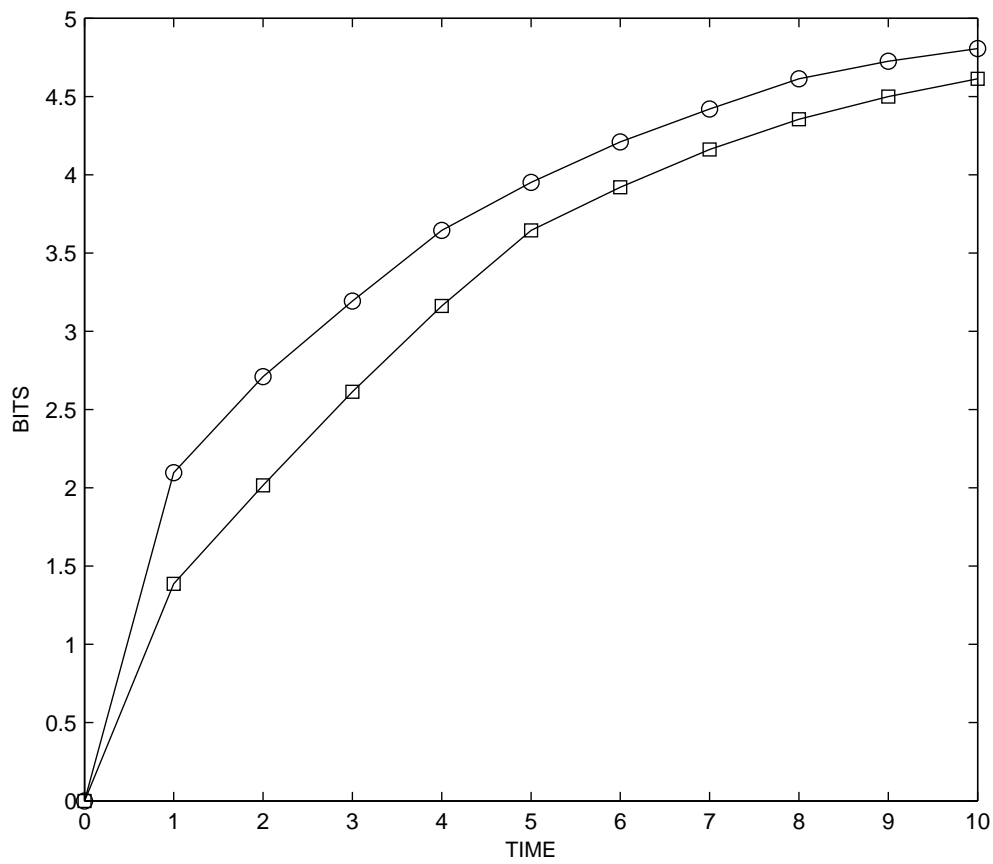
13   Demise

Demise occurs if decay brings the difference down to zero. The following graph displays entropy counts for microstates obtained from cyclic repeats of the sequence 01001000100001234555, while the two corresponding macrostates are $\{0, 1, 2, 3, 5\}$ and $\{4\}$.

14 Growth

Growth occurs if the difference increases for a considerable time. The following graph displays entropy counts for microstates obtained from cyclic repeats of the sequence 01010101010202020202, while the two corresponding macrostates are $\{0\}$ and $\{1, 2\}$.
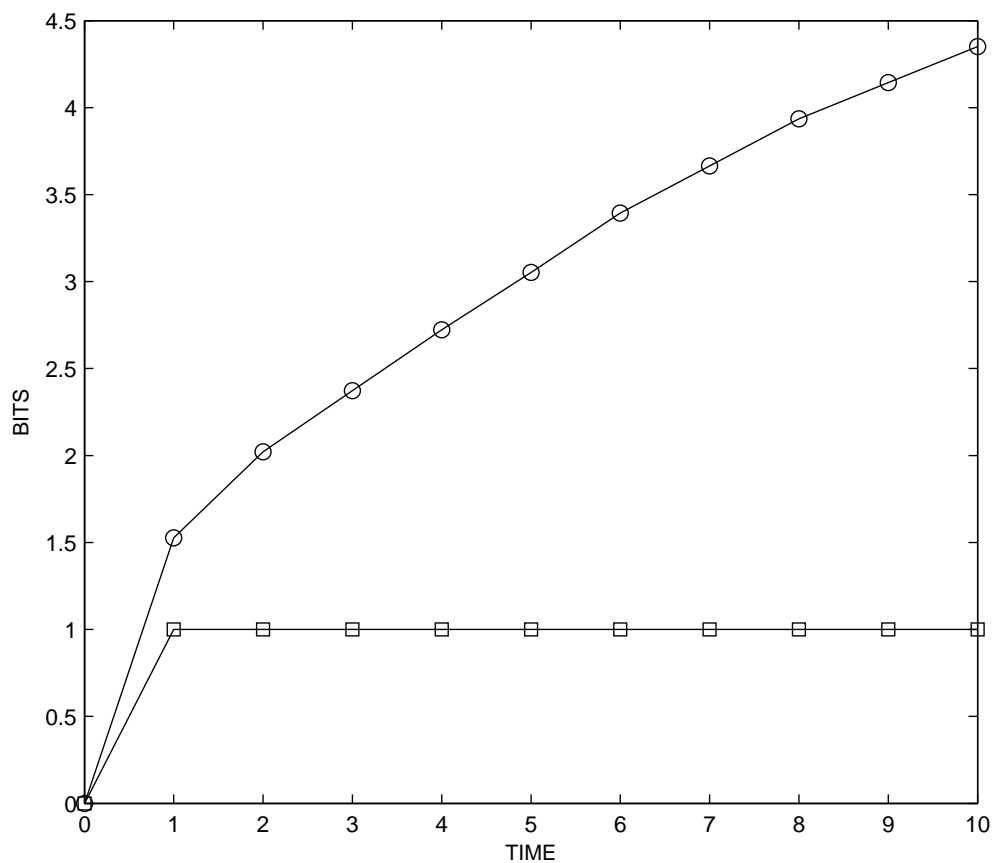
15   Biological applications

The Grouping Principle and the features displayed by the Two Curves Graph invite speculation on their relevance as abstract models of biological phenomena. Here are two examples.

- One might expect to observe grouping leading to senescence or demise (as displayed by Sections 12–13) in **biological clocks** responsible for the aging of cells, populations, etc. In this case the fact that the difference between the micro- and macro-entropies only measures the *capacity* for organization would not matter, since the organization itself would vanish once there was no capacity for it.

- In another direction, the Grouping Principle suggests an underlying **general mechanism for neoplasia**. Beneath a fixed micro-entropy, a change in the grouping of microstates into macrostates may change the behavior from stasis or senescence (as in Sections 11–12) to growth (as in Section 14). The change might occur randomly, or as a result of the action of a carcinogen. In this case the growing difference between the micro- and macro-entropies would not be accompanied by a growth in organization beneficial to the host organism.

In the hierarchical dynamical systems discussed above, the hierarchy has just consisted of two levels: microstates and macrostates. Biological systems are (by definition) much more complex. It will thus become necessary to extend the models to cover such cases, including the possibility of multiple time scales (compare [9], for example).

References

[1] Brooks, D.R.; Wiley, E.O., *Evolution as Entropy: Toward a Unified Theory of Biology*, University of Chicago Press, Chicago, IL, 1986. (2nd. ed. 1988.)

[2] Brooks, D.R.; Collier, J.; Maurer, B.A.; Smith, J.D.H.; Wiley, E.O., *Entropy and information in evolving biological systems*, Biology and Philosophy **4** (1989), p.407 ff.

[3] Frautschi, S., *Entropy in an expanding universe*, Science **217** (1982), p.593 ff.

[4] Landsberg, P.T., *Is equilibrium always an entropy maximization?*, J. Stat. Physics **35** (1984), p.159 ff.

[5] Landsberg, P.T., *Can entropy and order increase together?*, Physics Lett. A **102** (1984), p.171 ff.

[6] Layzer, D., *The arrow of time*, Sci. Amer. **233** (1975), p.56 ff.

[7] Smale, S., *Differentiable dynamical systems*, Bull. Amer. Math. Soc. **73** (1967), p.747 ff.

[8] Smith, J.D.H., *Some observations on the concepts of information-theoretic entropy and randomness*, Entropy **3** (2001), p.1 ff. (electronic journal).

[9] Smith, J.D.H., *A macroscopic approach to demography*, J. Math. Biol. **48** (2004), p.105 ff.

[10] Uspenskii, V.A.; Semenov, A.L.; Shen', A.Kh., *Can an individual sequence of zeros and ones be random?*, Russian Mathematical Surveys **45** (1990), p.121 ff.