SELF-ORGANIZATION IN BIOLOGY AND ECONOMICS

Robert U. Ayres
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

and

International Institute for Applied Systems Analysis
Laxenburg, Austria

RR-88-1
January 1988

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
Laxenburg, Austria

International Standard Book Number 3-7045-0091-7

Research Reports, which record research conducted at IIASA, are independently reviewed before publication. However, the views and opinions they express are not necessarily those of the Institute or the National Member Organizations that support it.

Cover design by Anka James

Printed by Novographic, Vienna, Austria

Contents

Summary

This report reviews some of the basic relationships between energy, entropy, order, and information, especially in the context of self-organizing dissipative systems. It has two parts. Four general ideas are discussed in the first part: (1) Information (negative entropy) can be captured and stored in dissipative structures, including living organisms. (2) "Evolutionary level" can be usefully defined as the ability of living organisms to capture and store information in structures. This is a variant of Lotka's principle. It is further suggested (3) that intelligence can be defined as the ability to modify or create *external* (nonliving) structures capable of storing information. It is also suggested (4) that information may be stored in two forms: (a) as "free energy" and (b) as structure (morphological differentiation) *per se*.

The report then focuses on the economic system as a self-organizing dissipative system in which intelligent activity (accumulation) of information-storing structures is more and more consciously controlled and managed. The main agent of negentropic accumulation is technology, generated endogenously by the economic system or adapted by it. The fundamental role of technological change as a driver of economic growth is emphasized, as is the increasing degree to which change and growth are intentionally managed. This trend also creates new vulnerabilities.

Foreword

The author is deputy leader of IIASA's Technology–Economy–Society Program. This constitutes the third (in chronological order) of a series of his Research Reports exploring information-theoretic perspectives on economic issues. Whereas the first two papers in this set focused on specific problems, this one essentially proposes a new paradigm for economics.

In brief, the author suggests that the economic system should be viewed as a kind of "living system" that is never in a true static equilibrium, but rather in a stationary (or quasi-stationary) state that must be maintained against dissipative forces by a continuous input of free energy from outside the system. The report discusses some implications of this paradigm, and it provides a particularly lucid integration of the concepts of economics in the context of living systems.

ROBERT H. PRY
Director
International Institute for Applied Systems Analysis

Preface and acknowledgments

The first version of this paper was originally written for the Fourteenth International Conference on the Unity of the Sciences (ICUS), held in November 1985. Eventual publication both in a book and as a journal article is expected, but has been delayed. It is being disseminated now (in a revised version) as an IIASA Research Report because of its relevance to two other Research Reports (RR-87-11 and RR-87-19) that explore more detailed aspects of the same broad theme, viz, applications of information concepts in resource allocation, manufacturing, and ergonomics. The three papers together constitute a natural set, of which this is the broadest in perspective and should therefore be regarded as a primary reference for the others.

I am indebted to Professor Georgescu-Roegen for pointing out several errors in my earlier draft. I am also grateful to Marcelo Alonso, Charles Berg, Harvey Brooks, Bill Clark, and James Grier Miller for inspiration and/or constructive comments during more recent discussions. None of them are responsible for any conceptual or other errors I may have made.

ROBERT U. AYRES
Deputy Program Leader
Technology–Economy–Society
Project Leader
Computer-Integrated Manufacturing

SELF-ORGANIZATION IN BIOLOGY AND ECONOMICS

1. Introduction

The second law of thermodynamics has been causing confusion and consternation since it was first formulated by Clausius in 1847. Once it was clearly recognized that processes are of two distinct kinds – reversible and irreversible – and that entropy is unchanged in the former but increases in the latter, an evolutionary principle or law of sorts was suggested for the physical universe. Clausius claimed in 1865 that for any closed system not in equilibrium, entropy must increase continuously. This is an extrapolation from the known fact that isolated thermodynamic systems not in equilibrium (but near it) must approach equilibrium irreversibly. Thus, thermodynamic equilibrium is characterized – in fact, defined – as a state of maximum entropy. Since entropy in an isolated system never decreases (according to the second law), no such system can stray further from equilibrium than its original position. Many irreversible processes tend to push any system in the direction of thermodynamic equilibrium, unless the system is open and acted upon by external forces.

The discouraging aspect of the situation is that for a closed system thermodynamic equilibrium is a changeless state in which all matter is uniformly distributed and there are no temperature or concentration differences – hence, no gradients and no structure or order. If the universe continues to expand physically without limit, such an asymptotic limit might be envisioned: Nernst called it the "heat death" (*Wärmetod*) of the universe. It is a state of absolute uniformity and maximum disorder or mixing.

Order is unavoidably a somewhat anthropomorphic concept. As the term is normally used, it implies sorting or separation, as opposed to mixing. A more general definition relates order to long-range spatial (or momentum) correlations. It is important to emphasize that the term "order" is often used carelessly as a synonym for structure or organization (e.g., by Schrödinger, 1945). These terms are discussed further below. Disorder, in the sense of mixing or randomness, is not necessarily equivalent to entropy, as suggested by Boltzmann (1896). It corresponds with entropy only in three very special (and unrepresentative) cases: an ideal perfect gas, a crystal at low temperature, and isotopic mixing (McGlashan, 1966).

At almost the same time the theory of thermodynamics was being developed, Charles Darwin formulated his theory of biological evolution by the mechanism of natural selection. Notwithstanding some severe challenges from physicists (notably, by William Thompson, Lord Kelvin) during the nineteenth century [1], Darwin's work in biology presented a clear and convincing story of evolution of living systems toward ever-increasing complexity and structure or organization, at least on the Earth's surface, culminating in the development of the human brain (*Figure 1*).

Structure and organization are used here as nearly interchangeable words meaning morphological and/or functional specialization (or "division of labor"). Given the fossil record and other evidence, this trend toward increasing organization appears indisputable. Here it must be pointed out that *order*, in the sense of separation, is not necessarily the same as *organization*, in the sense of specialization, though the concepts are undoubtedly related. Structure definitely implies some sort of order, if not the converse; but for a long time it was very hard to reconcile Darwin's theory – or evolution itself – with the second law of thermodynamics. On the other hand, there has never been any convincing evidence of a definite contradiction between biological evolution and thermodynamics.

New developments in mathematics and in nonequilibrium, nonlinear thermodynamics in the past three decades have created the basis for a positive reconciliation. In particular, Glansdorff and Prigogine (1971) and Nicolis and Prigogine (1977) have constructed models of simple chemical systems that exhibit stable, coherent, "self-organizing" behavior, yet are far from thermodynamic equilibrium. Eigen (1971) has even shown convincingly how complex macromolecules can not only reproduce themselves and form simple chemical building blocks via stable biochemical hypercycles, but can also evolve by random mutation (copying errors) and selection according to an optimization principle that is definable in molecular terms.

Four decades ago, Alfred Lotka (1945) suggested that the direction of biological evolution could be explained in terms of ability to capture and utilize energy from the environment. He should have specified *free energy*, since only the component of energy that is available to do work matters. Hereafter, this modification is assumed: free energy is used in the general sense of availability for useful work. (The terms "essergy" and "exergy" have been proposed, but neither has achieved wide acceptance.) Intuitively, Lotka's hypothesis helps explain at least two observable aspects of biological evolution.

(1) Organisms that use free energy (food) most efficiently will tend to compete most effectively, *ceteris paribus*, within a given niche.
(2) Organisms have evolved to utilize any and all available sources of free energy, including wastes. This implies ever-increasing diversity. It explains the development of specialized scavengers and parasites, as well as the hierarchy of predators.

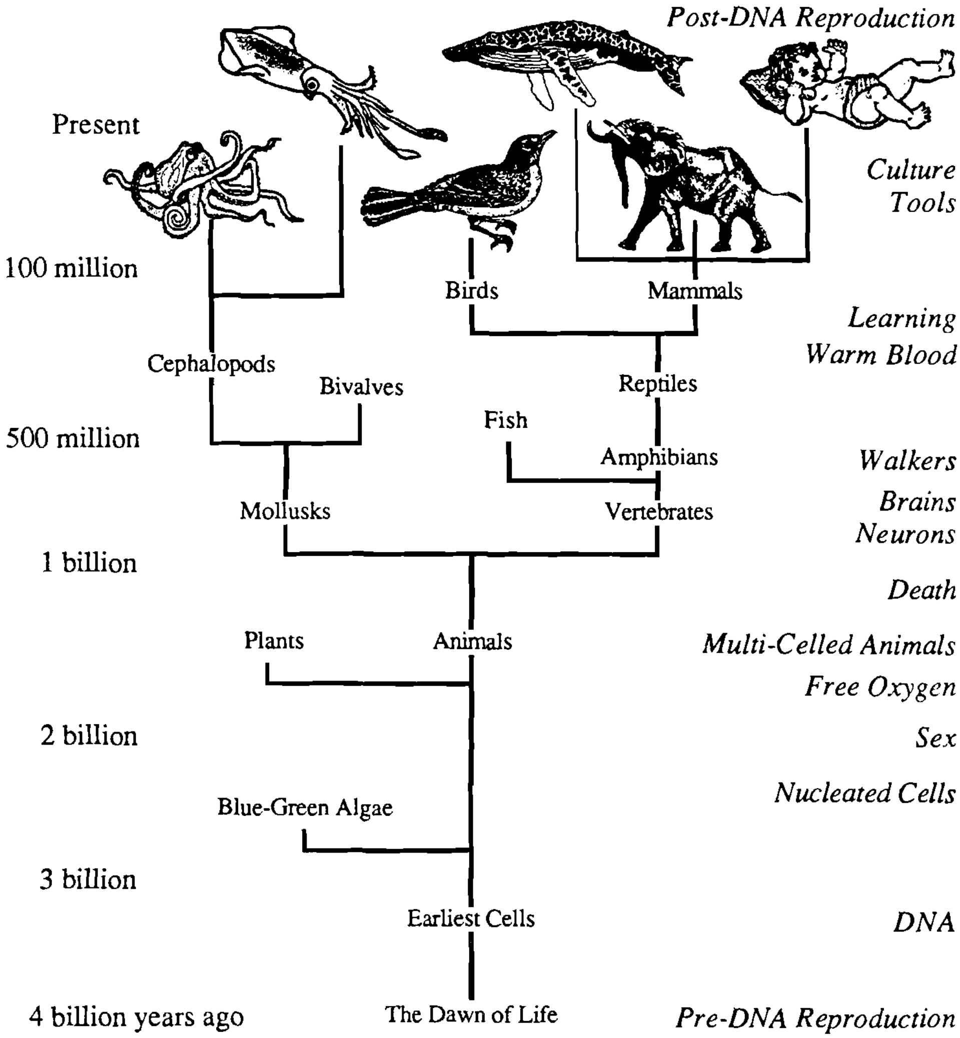

Figure 1. Intelligence on earth. The diagram gives timing, family relationships, and significant innovations in the development of terrestrial intelligence. It is likely that very early evolution occurred in an information carrier other than DNA. With the advent of learned behavior in mammals and birds, DNA lost a significant part of its job. More than half of what makes a modern human being is passed culturally. Source: Moravec (1985).

Admittedly, it was never clear whether Lotka's maximum principle (expressed in terms of free energy) would provide an explanation of evolution compatible with Darwin's principle of natural selection on the basis of "survival of the fittest". Unfortunately, there is a tendency to circularity in Darwin's theory, since "fitness" can probably only be defined in terms of ability to survive. In any case, the thermodynamic implications of evolution have been explored far more deeply in recent decades by Prigogine and his colleagues (1972), among others.

2. Order, entropy and structure

It is useful to distinguish two kinds of structure. One kind of orderly structure is exemplified by a crystalline solid, such as a snowflake. The second kind of structure occurs in systems that are stationary, but not necessarily in equilibrium. They may, in fact, be far away from thermodynamic equilibrium and maintained so by a continuous renewable flow of free energy from outside the system. Examples range from relatively simple chemical systems, such as the "Brusselator" (Nicolis and Prigogine, 1977) or the "Oregonator" (Field, 1985), to genetic material (Schrödinger, 1945) and living cells. A rotating gyroscope maintaining its dynamic equilibrium against the force of gravity offers an intriguing mechanical example of the same kind of phenomenon.

For most chemical systems, and some physical systems, the concept of thermodynamic equilibrium is well defined and characterized by a minimum value of G, the Gibbs free energy. By definition $G = U + PV - TS$. (The Gibbs free energy is the correct measure of departure from equilibrium only for processes that take place at constant temperature and pressure, whence $\Delta G = \Delta U + P\Delta V - T\Delta S$.) The "force" driving a chemical system toward equilibrium are also well defined. In simple cases, this can be expressed quantitatively as a function of G. In general, the governing relationships are nonlinear. Nonlinear equations can have multiple discrete solutions (or even continua of solutions), only one of which reduces to the solution of the linearized equation. This roughly describes the situation far from thermodynamic equilibrium, where there are solutions corresponding to situations where G is at a local – but not global – minimum (Glansdorff and Prigogine, 1971; Nicolis and Prigogine, 1977).

Such a local minimum may persist at finite temperatures and entropy (because of the term TS in the free energy function), but it depends on a continuous in-flow of free energy from the environment. This cannot happen in an isolated system. Such persistent stable states tend to exhibit "coherent" behavior, which is equivalent to long-range order in phase space. Physical examples include lasers (Graham and Haken, 1970) and thermal convection patterns (Velarde and Normand, 1980). The term "dissipative structure" has been applied by Nicolis and Prigogine (1977). Transitions from the disordered state to a coherent (ordered) state of a system can occur suddenly, triggered by random fluctuations. Such transitions between two solutions of a nonlinear equation have been classified topologically and termed "catastrophes" by Thom (1972).

A more recent development initiated by the work of May (1976), and since elaborated by a number of mathematicians (e.g., Jensen, 1987), essentially challenges the distinction between deterministic and random behavior. It is now known that even quite simple nonlinear dynamic systems can exhibit both chaotic and unpredictable behavior and regular behavior for certain ranges of parametric values. Chaotic behavior has been defined as the behavior of deterministic dynamic systems whose solutions are extremely sensitive to initial conditions, such that two trajectories started at nearby initial conditions diverge at an exponential rate called the "average Liaponov exponent" or the Kolmogorov-Sinai entropy (*ibid.*).

As the parameters of the nonlinear dynamic system vary, the transition between regularity and chaos is characterized by regions of chaotic behavior gradually expanding until they occupy most of the phase space of possible solutions. Nevertheless, under certain conditions even a chaotic dynamic system will spend most of its time in one (or more) particular region(s) in phase space. Such regions have been given the name "strange attractors" (Ott, 1981). The relevance of all this to the problem of self-organization is simply that standard notions of "stability" or persistence are too limiting: they will have to be modified so as to be applicable in certain circumstances, even in dynamic systems exhibiting chaotic (i.e., apparently random) behavior.

Of course, in most complex open systems - including living systems - thermodynamic equilibrium conditions cannot be specified in explicit terms. In biological systems, for instance, structure and organization are reflected in morphological differentiation and functional specialization, not in simple regularities of the kind found in a snowflake. In such systems, thermodynamic variables may not be definable at all. It is simply assumed that in such cases, forces driving the system toward equilibrium are zero at equilibrium and increase, in general, with "distance", where the latter term is itself difficult to define precisely. Again the (unknown) equation governing reaction rates is likely to be nonlinear, with multiple solutions, only one of which corresponds to the linear near-equilibrium case. However, detailed quantitative examples of such behavior in chemistry or biology are still quite scarce.

As noted above, the classical notions of order in near-equilibrium vis-à-vis far-from-equilibrium situations are very different. Boltzmann's famous "order principle" (1896), relating order to entropy, is not generally valid and, in any case, it would apply only in the near-equilibrium case (termed the "thermodynamic branch" by Glansdorff and Prigogine). Orderliness must be defined independently of entropy for a self-organizing dissipative structure. Intuitively, the orderliness of a structure, such as DNA molecule, is a function of its ability to contain (i.e., embody) information and to preserve it from destruction or contamination. Information is a well defined concept, applicable in principle to any system, including physical, chemical, biological, social, and even technological systems. Indeed, it offers a bridge between the various disciplines.

3. What is information?

Information is commonly used in three ways: (a) in the semantic sense as data; (b) in the pragmatic sense of knowledge; and (c) in a formal technical sense as the resolution of doubt or uncertainty. It must be emphasized that, while knowledge is inherently an anthropocentric concept, information is technically defined in terms of an abstract observer. For a good recent discussion, see Cherry (1978). A classic reference is Shannon and Weaver (1949).

Information in the third (Shannonian) sense of the word is a function of the *a priori* probability of selecting a given state or outcome from the universe of physically possible states. The more physically possible states there are, the more information is embodied in a given selection or set of equivalent selections. The first formal definition of information H (Hartley, 1928) was given in the context of telegraphic communications. Let

$$H = -\mathrm{K} \log_2 f \tag{1}$$

where f is the frequency with which a given code element – for example, a letter of the alphabet expressed in Morse code – appears in a *given* message. Here the frequency is defined naturally as

$$f = Z_1/Z_0 \tag{2}$$

where Z_1 is the number of times the code element (e.g., letter) occurs in the message and Z_0 is the total number of elements in the message.

More generally, in a system with a large number of possible states, the information content is defined by Z_0, the number of possible end-states of the system (material), and Z_1 is the number of states of particular interest, assuming each state to be equally probable *a priori*. Anticipating later generalizations, the numbers Z_0, Z_1 can be estimated at any convenient level of detail (atomic, molecular, cellular, microcrystalline, etc.), since only the ratio matters.

The constant K in (1) is a scale factor that is fixed as soon as one selects a unit of measurement for information. We can permanently fix $\mathrm{K} = 1$ by defining a "bit" of information as the amount provided by the elimination of doubt between two equally probable outcomes. The number of bits of information required to solve a problem (i.e., to decode a message) corresponds exactly to the minimum number of distinct binary decisions necessary to make the correct final choice. This is a function only of the number of possible binary choices and the *a priori* probability of each outcome.

Imagine the possibilities are "cells" (for example, in a dungeon) and that one seeks to find a particular prisoner, known to be in one of them. Suppose the searcher has no other information, whence all cells can be assumed to be equally probable *a priori*. (This is, of course, a good illustration of the pragmatic use of the word information as knowledge.) Assume W cells in the dungeon. There are many alternative search strategies. The simplest, in some sense, is sequential interrogation of the jailer: "Is the prisoner in cell #1? Cell #2? Etc." The searcher might be lucky and hit on the correct choice immediately or, conversely, he might be unlucky and not find the prisoner until the last (Wth) question. These outcomes are equally probable, as noted. If this sequential strategy were adopted, the average (or expected) number of questions required would be $W/2$.

On reflection, a much more efficient strategy is possible. The searcher could ask questions that reduce the number of remaining possibilities by a factor of 2 each time. For instance, if the cells are numbered $1 \ldots W$, one could ask: "Is the prisoner located in the subset of cells from 1 to $W/2$ inclusive?" Whether the answer is yes or no, the uncertainty is reduced by a factor of 2. If the answer is yes the prisoner is known to be in the first group of addresses $(1,2,\ldots W/2)$; if the answer is no, the prisoner must be in the second half $(W/2 + 1, \ldots W]$. In either case, the same procedure is repeated until only two final possibilities remain. The information gained by each of these questions is exactly 1 bit. The smallest number of such questions (or decisions) is $\log_2 W$. (Obviously, if W is not an even number, $W/2$ might not be an integer; but for large values of W, this complication can be ignored.)

In the real world, one cannot always proceed quite so efficiently, because *possible* states are seldom, in fact, equally *probable*. For instance, consider the game of "Twenty Questions". Suppose I am thinking of a famous author, and you have to identify that author by asking a few questions. Possible questions might include

- Is the author male?
- Does his/her name begin with letters A–M (the first half of the alphabet)?
- Is the author living?
- Does (did) the author live in England?

In all cases, the *a priori* probability of a yes answer is likely to be different from the probability of a no. For instance, considerably more than half of all famous authors are male and well over half have names beginning with letters in the first half of the alphabet. On the other hand, probably fewer than half of all famous authors are living. Thus, the values of yes versus no answers are unequal in terms of reducing uncertainty, and the information elicited by each question is normally either more than 1 bit or less than 1 bit, depending on the case.

Let P_i be the *a priori* probability of the ith state (or event). The information provided by an actual realization of that event (i.e., the answer to the question is yes) is therefore

$$H_i = \log_2 (1/P_i) = -\log_2 P_i \tag{3}$$

For example, if 3/4 of all "famous writers" are male, then a yes is worth *less* than 1 bit, because the number of possibilities was only cut by a factor of 1/4 rather than by a factor of 1/2. In fact, the value of a yes in this case is $\log_2 (4/3) = 2 - \log_2 3 = 0.4147$ bits. On the other hand, the value of a no is $\log_2 4 = 2$ bits! The information provided by a series of affirmative answers is the simple sum of their individual information values, provided the probabilities are truly independent of each other. For nonindependent states i, j, one must adjust by adding or subtracting a term representing the information value corresponding to the deviation of joint probability from a simple product of probabilities. However, the assumption of independence is generally adequate for purposes of the discussion that follows.

Returning for a moment to the "Twenty Questions" example above, it was pointed out that the value of a yes versus a no is generally unequal. It is therefore natural to ask: what is the probable or expected amount of information elicited by each question? Evidently, the information value of each outcome must be weighted by its probability, P_i. It is helpful to note that probabilities add up to unity, viz.

$$\Sigma P_i = 1, \qquad i = 1,\ldots,N \tag{4}$$

$$<H> = -<\log_2 P_i> = -\frac{\Sigma_i P_i \log_2 P_i}{\Sigma_i P_i} = -\Sigma_i P_i \log_2 P_i \tag{5}$$

This expression for expected information was first introduced by Shannon (1948), also in the context of communications.

As an illustration of this formula, consider the question posed above: "Is the famous author (I am thinking of) male?" If, in fact, 3/4 of all famous authors are male, then the answer yes on my part adds 0.4147 bits of information to your stock, but the answer no is worth 2 bits. But yes has a probability of 3/4, while no has a probability of 1/4. Thus, the expected amount of information to be gained from this question is

$$(3/4) \times 0.4147 \times (1/4) \times 2 = 0.811 \text{ bits}$$

which is, as already pointed out, somewhat less than 1 bit.

This way of looking at information opens a link with physics, via statistical mechanics. One may consider the case of a physical system (of atoms or molecules) that is changing or evolving over time. At any moment it is in a definite state, but at another moment it will be in another state. Let P_i be a set of state probabilities ($P_i \geq 0$), satisfying (4). When a definite observation is made, confirming that event i (among all possible events) has occurred (i.e., the system is in the ith state), the information gained by that observation is given by (3). But the original probability of that event (state) was only P_i. Another measurement might find the system in quite a different state. The *expected* amount of information provided by an observation of the system is the average value of the expression (3) over *many observations*, which is exactly (5).

In reality, we have no means of observing large systems (e.g, molecules) so precisely that the exact state of each molecule can be determined at a moment in time. One can only say how much information would be provided by such an observation, on the assumption that all microstates are equally probable and that the total number of such microstates is W. Then the information value of the (hypothetical) observation would be [from equation (3)]

$$H = -\log_2 (1/W) = \log_2 W \tag{6}$$

However, in the more realistic case of unequal probabilities, Shannon's formula (5) must be used.

It is of some interest to note that multi-component *systems* embody information in their structure. Assume a closed system of n "compartments". One of the compartments can always be defined, for convenience, as the external environment. Assume each compartment contains a "stock" of some universal medium (it could be railroad cars, mass-energy, or money) and there are "flows" between compartments. The medium is conserved in the system as a whole. Let T be the total throughput of the system, and T can be composed into inputs to the various compartments, or outputs from them.

$$T = \sum_{i=1}^{n} T_i = \sum_{j=1}^{n} T_j \tag{7}$$

One can further decompose the flows into each compartment into a sum of flows *from* the other compartments and, similarly, the flow out of each compartment can be decomposed into flows *to* each of the others. Thus, if T_{ij} represents the flow i to j.

$$T_i = \sum_{j=1}^{n} T_{ij} \tag{8}$$

$$T_j = \sum_{i=1}^{n} T_{ij} \tag{9}$$

The ratio T_i/T is the probability per unit time that a quantum of the medium will enter (or leave) the ith compartment. The joint probability that a quantum will flow from the ith compartment to the jth compartment is T_{ij}/T, and so on. It can be shown that the average total information needed to specify the n-compartment system is given by

$$H(n) = -\frac{1}{T}\sum_{i=1}^{n} T_i \log T_i - \frac{1}{T}\sum_{j=1}^{n} T_j \log T_j + \frac{1}{T}\sum_{i=1}^{n}\sum_{j=1}^{n} T_{ij}(\log T + \log T_{ij}) \tag{10}$$

The first two terms can be interpreted as "input" information and "output" information, respectively. The third term can be interpreted as informtion embodied in "structure". This formalism is applicable to any multi-component system that can be characterized as a network of flows.

4. Information and entropy

In point of historical fact, Shannon actually used the term entropy, rather than (expected) information, as his measure of uncertainty. His choice of terminology was apparently whimsically suggested by the mathematician John von Neumann, who defined information gain as the difference between the uncertainty prevailing before an event (e.g., receipt of a message) and the uncertainty remaining afterward. Using the standard notation for entropy (S)

$$\Delta H = -\Delta S \tag{11}$$

But (11) is merely an assertion of identity, not a proof. The relationship between Shannonian entropy and classical thermodynamic entropy (from Clausius) has been the center of an extended discussion, with significant contributions by Brillouin (1951 and 1962), Jaynes (1957a and 1957b), Khinchin (1957), and Tribus (1961a and 1961b). Although there are still doubters, most physicists now accept that information and (negative) entropy are, essentially, the same thing – not merely analogs.

Boltzmann's famous statistical definition of entropy– it is inscribed on his tomb in Vienna – for a system of many identical particles was

$$S = \mathrm{k} \ln W \tag{12}$$

where k is the Boltzmann constant (1.38×10^{-23} joules/°K) and W was the number of possible distinct states (or complexions) of the system. For an ideal gas of distinguishable but quantized particles, for example, it is known that

$$\ln W = N\left\{5/2 + \ln\left[(2j+1)\left(\frac{V}{N}\right)\left(\frac{4\pi mE}{3h^2N}\right)^{3/2}\right]\right\} \tag{13}$$

where N is the number of particles, V is the volume, E is the total energy of the system, m is the mass of a particle, j is the spin, and h is Planck's constant. For the macroscale, N is typically of the order of the number of molecules in a gram-molecular weight (or mole) of any gas at standard pressure and temperature (STP). This number is called Avogadro's number, A, which is equal to 6.02 $\times 10^{23}$.) Taking helium gas at STP as an example, the numerical value of absolute entropy S is 28.7 cal/°K or 7.2 cal/mole°K = 30.1 joules/mole°K.

In information terms, this is the total amount of micro-uncertainty in the gas volume as a whole, with respect to the states of all the molecules in it. It is therefore the amount of information that would be gained if all micro-uncertainty were removed by some hypothetical observation. By comparison, the potential amount of uncertainty removed (or information gained) with respect to a single molecule that could be in either one of two isotopic states, is exactly 1 bit. Using Boltzmann's formula (8), the entropy of such a system is kln 2.

By this logic, Brillouin determined the thermodynamic equivalent of 1 bit to be:

$$\begin{aligned} 1 \text{ bit} = k\ln 2 &= 9.52 \times 10^{-24} \text{ joules/°K} \\ &= 2.27 \times 10^{-24} \text{ cal/°K} \end{aligned}$$

Thus, thermodynamic "information" has units of energy divided by temperature. A flow of 1 bit of information corresponds to a very small flow of energy or a very small change of temperature or both. Tribus and McIrvine (1971), who provide a clearer explanation than Brillouin, have particularly emphasized the very small absolute amounts of energy required to process and deliver information in most communication and information-processing activities. For example, a TV broadcast requires about 6 joules of energy to deliver 300,000 bits of information, but that information has an available energy equivalent of only 1.3 $\times 10^{-15}$ joules. The rest of the 6 joules of energy is just lost to the environment. (Note that 1 cal = 4.186 joules.)

Several physicists have called attention to the enormous amplification factor implicit in the fact that small amounts of energy in the form of information suffice to control very large flows of energy. Allred (1977) has calculated several commonplace examples with power gain factors ranging from 10^8 for an automobile to 10^{13} for a large aircraft homing on a radio beacon. For purposes of this paper, the interesting implications of the large amplifier effect will not be pursued further.

The physical identification of information as negentropy suggests the possibility that stocks of information embodied in structure/organization can be regarded in some sense as reserves or storehouses of negative entropy. These can be utilized to increase the ability of dissipative open systems to capture both negative entropy and materials from the environment and, thus, to grow. This perspective will be elaborated further hereafter.

5. Information and organization in biology

Three related notions are suggested by the identification of information stocks with negative entropy. First, a dissipative structure far from equilibrium may itself capture and store negative entropy from its environment for future use in the form of *more complex organization or structure.*

Second, *evolutionary level* can be defined, tentatively, as the ability of a living system to capture and store negentropy (or information) in genetic structures or in brains. Genetic information predominates in lower forms of life and brain information predominates in higher forms, as indicated in *Figure 2.* All living organisms have the ability to store information in lesser or greater degree.

Third, *intelligence* can be defined as the ability to learn, to modify behavior, and/or to modify the external environment, i.e., to create external structures (cultures and cultural artifacts) for information storage. The ability to modify the external environment is apparently possessed to some extent by all higher animals, but the ability to create external structures explicitly for information storage is possessed only by humans. Carl Sagan (1977) has called this "extra-somatic" information. Whereas Lotka's principle suggests that evolution seeks to maximize the ability to process free energy, it is suggested below that evolution seeks to maximize the ability to capture free energy and, most important, to convert some of it to *morphological information embodied as structure or organization.* To the extent that intelligence enhances this ability, evolution seeks to maximize intelligence. Yet the new perspective does not seriously conflict with Lotka's, inasmuch as capturing free energy also implies processing (i.e., metabolizing) it.

That organisms capture disembodied information from the environment and store it in physical structures has been pointed out in various ways many times. The underlying notion that living organisms seem to retard the global increase of entropy has been frequently been expressed in the biological literature, e.g., by Johnstone (1921), Breder (1942), Needham (1943) and Blum (1955) [2]. The idea that advancing evolutionary level corresponds to increasing structural complexity goes back at least to Herbert Spencer and was reiterated by Huxley (1956), although this notion is still questioned by some biologists. All of this literature was well summarized by Polgar (1961).

The gene itself is nothing more nor less than a packet of information stored compactly in molecular form (Schrödinger, 1945). It contains both morphological and functional information needed by the organism. The information embod-

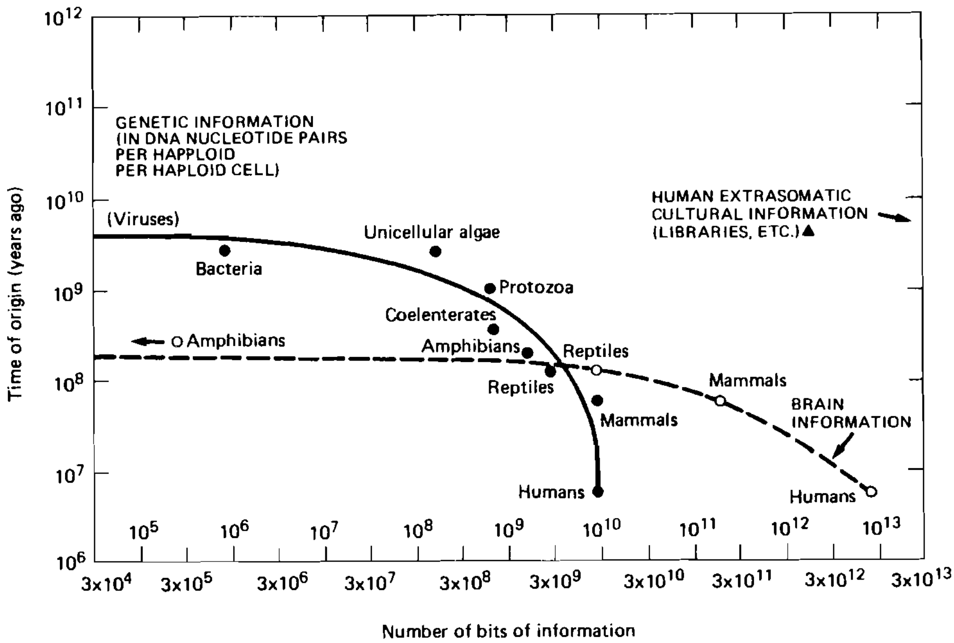

Figure 2. Genetic information. Source: Sagan (1977).

ied in genes tells cells how and when to divide, how and when to differentiate, how to manufacture various enzymes, hormones, etc. It specifies the many coupled biochemical pathways (hypercycles) that carry out the metabolic processes. It also tells the organism as a whole how to react to various stimuli, what food to eat, when and how to mate, where to lay eggs, etc. This information storehouse is the result of long evolutionary learning process, described by Darwin as natural selection and by Eigen (1971) as "value maximization". The cumulative nature of the process is evident from the fact that the higher organisms, arriving later on the evolutionary scene, carry far more genetic information than the simplest, earliest organisms (*Figure 2*). The ability of higher organisms to accumulate and reproduce this information by extragenetic means can also be regarded as evidence of increasing intelligence in the sense defined above.

Returning to the accumulation of "stored negentropy-as-structure", Polgar (1961) has identified four key mechanisms, viz. *persistence, replication, environmental modification*, and *social evolution.* On deeper reflection, Polgar's classification seems incomplete. Since self-replication and environmental modification are both mechanisms for information storage and transfer, it would seem that *self-modification* (i.e., learning) might be considered as another category of mechanism.

At the molecular (i.e., genetic) level, Schrödinger (1945) showed that persistence, as reflected by the low rate of natural mutation, can be explained in terms of quantum mechanics. The critical difference between classical and quantized systems is that the former are characterized by a continuum of possible states, whereas the latter are characterized by discrete states with discrete transitions between them. At the level of chemical system, Glansdorff and Prigogine (1971) and Nicolis and Prigogine (1977) have given an explanation of the persistence of structure in terms of nonequilibrium thermodynamics.

Replication is, of course, one of nature's basic long-term survival (and growth) mechanisms. A detailed understanding of replication at the cellular level may still be years away, but the famous discovery of the double-helix structure of DNA and the so-called genetic code provided a very useful starting point (Crick, 1962). The other important natural means whereby living systems ensure their own long-term survival are by self-modification and environmental modification. The species modification process may be unconscious, as in Darwin's theory of natural selection, or take place at the microlevel, as postulated by Eigen (1971) and Dyson (1982). Or self-modification can be conscious, as in human learning processes.

By comparison, environmental modification has received less attention, although it is immensely important. Living organisms have massively modified the atmosphere and the ocean over the past two billion years (Oparin, 1953). The earth's primordial atmosphere was mostly carbon dioxide (CO_2) and water (H_2O) plus methane (CH_4) and ammonia (NH_3) with no free oxygen or nitrogen, while the primitive oceans were much smaller and saltier. Life as we know it today could not survive in such an environment. By the same token, the present oxygen-rich environment is quite unsuitable to the spontaneous formation of organic molecules by known mechanisms (e.g., Miller and Orgell, 1974). The driver of these environmental changes was the evolutionary development of increasingly efficient energy conversion mechanisms (*Table 1*). Whereas the earliest known organisms obtained energy by inefficient anaerobic processes akin to fermentation, later aerobic organisms utilized the more efficient photosynthetic processes (Wald, 1955; Gaffron, 1965).

Environmental modification in some form is still practiced today at the microlevel by viruses (upon their host cells), by wasps that paralyze their prey and lay eggs in them, by every nest-building species, and on the mesolevel by some animals such as beavers. On a larger scale, forests and savannahs create their own macroenvironments. (Deforestation in the tropics can lead to irreversible desertification). Coral reefs are the undersea analog. Of course, human activities such as agriculture and fossil fuel consumption have enormous environmental impacts, which we need not consider here at length.

Table 1. Energy conversion processes during various evolutionary eras.

Form of evolution	*Era*	*Environment*	*Energy source*	*Structure and process outcomes*
Chemical	I	Anaerobic; methane, ammonia, hydrogen	Ultraviolet light; heat	Acetate, glycine, uracil, adenine, other organic molecules in aqueous medium
		Loss of most free hydrogen		
	II	Anaerobic; traces of gasenous oxygen	Ultraviolet light; heat; visible light	Polyphosphates, porphyrins, peptides; porphyrin catalysis of photoreduction and oxidation
		Loss of most ultraviolet light		
	III	Anaerobic; traces of gaseous oxogen and carbon dioxide	Visible light	Replicating organic molecules; photochemical reactions
		Loss of many free organic molecules		
	IV	Anaerobic; gaseous carbon dioxide and traces of gaseous oxygen	Photo-reduction; fermentation	Organelles; cells; two-step light–energy conversion process
Biological		Loss of most anaerobic environmental regions		
	V	Aerobic; anaerobic pockets	Photosynthesis; respiration	Free-living cells; organs and organisms

Source: Gaffron (1965).

Another way in which negentropy-as-structure in the biosphere operates to cut down on the free energy flux per unit of biomass (and, thus, retard the rate of global entropy increase) is by means of species diversity and specialization.

The key is that there are relationships between function and structure. Each species of plant or animal having a different function (i.e., a different niche) will have a different structure – hence, a different microcomposition. Each species can be expected to require a somewhat different combination of inputs, as noted by Leibig in 1876 (Lotka, 1950), so that its limiting factors will be different. Thus the total biomass of two species exploiting the same resource base can generally be larger than that of any one alone. A third species, with a different specialization, can add still more to the total biomass without exhausting the available resources, and so on.

The lowest-level trophic animals obtain their food mainly as carbohydrates or cellulose, directly from photosynthetic plants. Higher animals, in turn, are able to exploit lower animals and obtain their food mainly as proteins and fats, which are more useful and easier to digest. The top predators are therefore able to eat and digest much less bulk than they would otherwise need, by consuming foods more similar to their own tissues and more easily broken down into sugars and amino acids for reassembly. This enables active carnivores to consume more food (free energy), but spend much less time eating and digesting than the herbivores they prey on.

Significantly, the amount of genetic information required to reproduce an organism tends to increase with its trophic level in the predator–prey hierarchy. So, in general, does the information-processing capacity of the organism itself: the processing capacities of its brain and central nervous system seem to increase with trophic level, as shown in *Figure 2*. For higher mammals, the storage capacity of the brain significantly exceeds the storage capacity of DNA.

It is worthy of note that Miller (1978 and 1987) has identified 20 subsystems that appear to be characteristic of most living systems, from cells to human societies. Of these, 10 subsystems are primarily information processors. The complete list, together with some examples (given by Miller) is attached as an *Appendix*.

A recent contribution of some importance is the observation that structural complexity is proportional in information-content for interactive biotic communities or ecosystems, as well as individual organisms.

In particular, Ulanowicz (1986) has proposed an information-based formalism [see equations (7)-(10)] for describing biotic communities and helping to explain the countervailing tendencies of such systems towards increasing *efficiency*, on the one hand, and increasing *resiliency* on the other. Ulanowicz introduces a composite measure for systems called "ascendency". A combining total system throughput flow T (of nutrient or energy) and "average mutual information" H in the Shannonian sense. The defining relationship is

$$A = TH \tag{14}$$

6. Extragenetic processes of information storage and transfer

Humans are not the first organisms in nature to transmit nongenetic information from generation to generation. Most mammals and birds teach their young, to some degree at least. Humans are, however, the first species to store nongenetic information systematically *as such* in external repositories, such as libraries that are maintained by the society as a whole. The developments of spoken and, later, written languages were obviously critical steps. Indeed, although humans may have existed as a distinct biological species for several million years, the processes of extragenetic information storage began only about 6000 years ago, as shown in *Table 2*. Moreover, in recent centuries, cultural information accumulation and transfer, recently enhanced by the use of computers, has undoubtedly approached, and possibly surpassed, the genetic transfer process, as suggested in *Figure 3*.

Table 2. Evolution of external means of information transmission.

Ideographs	4000 BC	Sumer.
Pictographs	3500 BC	Sumer.
Government (State)/law	3500 BC	Sumer.
Writing (cuneiform)	3000 BC	Sumer.
Phonograms	3000–2800 BC	Sumer.
Hieroglyphics	3000–2800 BC	Egypt
Dictionaries	2800 BC	Sumer./tablets
Book (papyrus roll)	2800 BC	Sumer./Egypt
Archives	2500 BC	Sumer.
University	2500 BC	Sumer.
Library	2500 BC	Sumer.
Multiplication tables	2500 BC	Sumer.
Code of law	2100 BC	Sumer.
Mail (letters)	2200 BC	Sumer.
Fractions	2000 BC	Babylon
Algebra	1800 BC	Mesopotamia
Alphabet	1500–1400 BC	Palestine (Phoenicians)
Coinage	700 BC	Lydia
Zero	300 BC	Babylon
Encyclopedia	18th century AD	France
Public education	18th century AD	New England
Telegraph	1840s	US, UK, Germany
Telephone	1876	US
Radio-telegraph	1890s	Italy, UK
Radio broadcast	1922	US, UK
TV broadcast	1928	US, UK

For example, the systematic use of extragenetic information to enhance the food gathering and production process began in prehistoric times. Knowledge about plant reproduction, animal behavior, weather, climate, geography, and even astronomy [3] had an enormous payoff in terms of increasing the ability of a human population to support itself and grow in numbers. At first, much of this knowledge was acquired and retained by a process not dissimilar to natural selection. More recently, the application of knowledge to increase production has become increasingly intentional, systematic, and analytic. As I have emphasized elsewhere (Ayres, 1987b), manufacturing can be regarded as the process of implanting "useful" information in materials. A similar notion has also been expressed by Berg (1987): "Production begins in the human nervous system and ends in the human nervous system". The processes of learning about nature, and disseminating and applying that knowledge for the benefit of man, are themselves now institutionalized, as will be discussed later.

The notion of stored information (knowledge) as an effective *substitute* for resource inputs follows immediately from the notion of using knowledge to increase the efficiency of capturing and accumulating "natural" information (free energy) from the environment. In effect, resources are created by human intelligence. Solar energy is, of course, captured by photosynthesis and stored as fossil

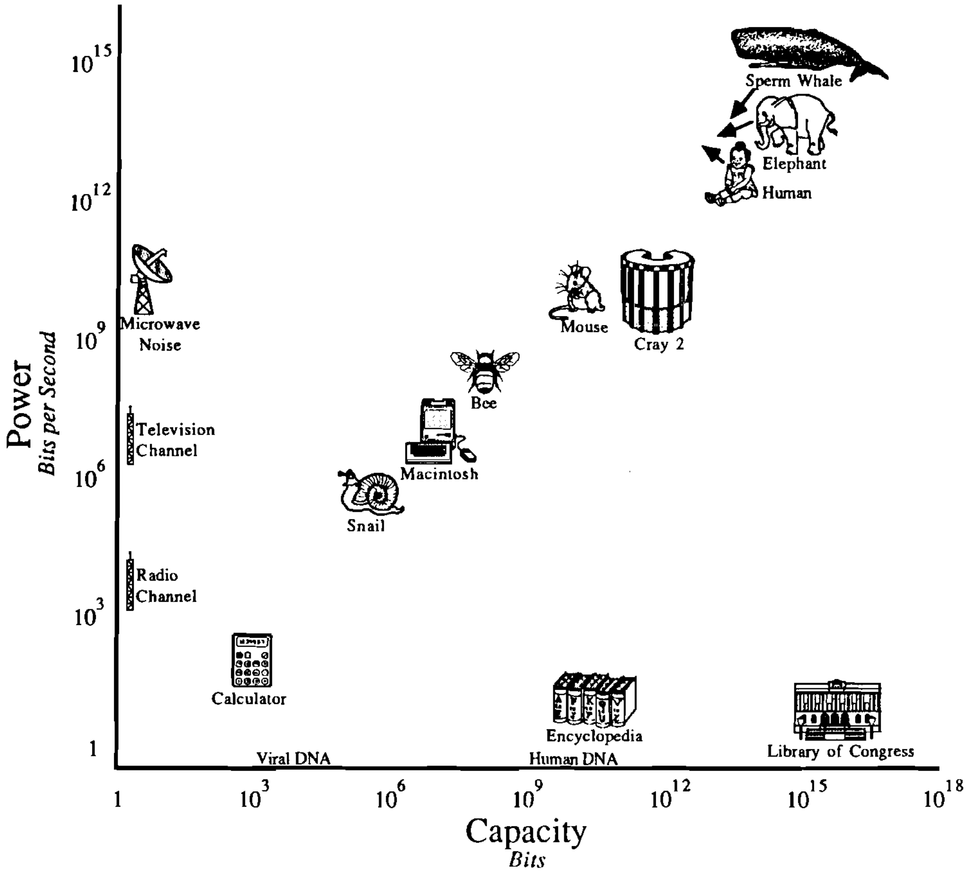

Figure 3. Computers versus living systems. Source: Moravec (1981).

fuels, but to find and use these resources requires knowledge. The essential equivalence of knowledge and resources is neatly summarized by the proverb: "Give a man a fish, and he can feed his family for a day. Teach him how to fish, and he can feed his family forever." Ancient man learned how to make useful objects from copper, silver, iron, and other metals from nearly pure nuggets. As one resource is exhausted, typically, another of lower grade is exploited. Thus, charcoal gave way to coke, and whale oil was replaced by kerosene. To be fair, the innovations that make such substitutions possible are often motivated by other considerations. In particular, the first distillation of petroleum to extract kerosene (and other by-products) was motivated by the desire to find some use for an abundant nuisance. (I am indebted to Charles Berg for this observation).

Knowledge also permits more efficient utilization of any resource. As the demand for motor gasoline grew rapidly in the early twentieth century, means of increasing the gasoline and octane output from each barrel of crude oil were actively sought. Using sophisticated modern refinery technology, and anti-knock additives, the gasoline recovery fraction has risen from 15% to as much as 60% of each barrel, while average octane levels have almost doubled from about 50 (for natural gasoline) to over 90 in the late 1960s. This, in turn, permits higher engine compression ratios and correspondingly higher levels of thermal efficiency. Thus, technological knowledge has enormously increased the amount of useful transportation work that can be extracted from crude oil (Enos, 1962). It is not unreasonable to attribute this enhancement to technological knowledge created by the techno-economic system. This system comprises both embodied knowledge (in capital equipment) and disembodied knowledge in various portable forms, such as books and computer software.

A further observation regarding resources: the vast storehouse of fossil fuels in the earth's crust that humans are currently exploiting (and using up) is an accumulation of surplus natural negentropy left over from incompletely decayed living organisms in earlier geological periods. It is stored (from a molecular perspective) as chemical structure, i.e., energy-rich hydrocarbon molecules that are quite stable at ambient temperatures, but which combine exothermically with oxygen above ignition temperature. Obviously, humans will have to find other energy sources to replace hydrocarbons within a century, more or less. It is already clear that several alternative possibilities exist, including fission, fusion, and photovoltaic cells on the Earth or in space. However, all of them will require large capital (i.e., stored information) investments, not to mention technologies more advanced than are currently available. In any case, some surplus negentropy in the form of capital will have to be set aside from the existing fossil fuel store to "finance" the eventual changeover.

It is worth emphasizing once again that physical capital represents a store of useful information. The process of accumulating capital, which economists traditionally assumed to be the mechanism of economic growth, is evidently a process of accumulating embodied information. The technological progress represented by successive generations of capital equipment is, from this perspective, merely another aspect of the same accumulation process.

7. The economic system as a self-organizing dissipative structure [4]

Economists do not think of economic activities and relationships in thermodynamic terms. When economists talk about equilibrium, they refer to a balance between supply and demand, or (looking at it another way) among prices, wages, and profits. Neoclassical economic models consider labor, capital goods, and services to be abstractions. The exception is, perhaps, resource/environmental economics, in which some physical properties of matter (e.g., mass, toxicity) cannot be neglected.

The proof of the existence of a static supply–demand equilibrium (conjectured by Walras in 1877; finally proved by Arrow and Debreu in 1951) was one of the great achievements of neoclassical economics because it seems to provide a theoretical explanation of Adam Smith's price-setting "invisible hand". There can be no question that the operation of a money-based free competitive market generates a kind of coherence, or long-range order, in contrast to the unstable price–wage anarchy that prevails in a barter society, for instance. Central planning attempts to introduce order of another kind. The static competitive free market-based economic system described in textbooks does reflect a kind of order very similar to cooperative phenomena in physics. It has also been proved that an idealized market-based system tends toward a so-called Pareto optimum – a situation where nobody can be better off without making somebody else worse off – although it does not necessarily allocate resources equitably. (Equity is, of course, a moral concept.) Finally, the market system is, in theory, self-regulating and capable of recovering from a perturbation in demand, for instance. It therefore displays self-organizing characteristics.

Even the abstract model of the economic system depends on resource inputs, but in a closed Walrasian model *resources are assumed to be generated by labor and capital.* The neo-classical (Walrasian) equilibrium system does *not* qualify as a dissipative structure. The neoclassical system is, in effect, a perpetual motion machine. This fact was emphatically pointed out by the Nobel prize-winning chemist F. Soddy in 1922 (Daly, 1980), but Soddy's work was virtually ignored by economists. The first economists to stress the dissipative nature of the economic system were Boulding (1966) and Georgescu-Roegen (1971). The relevance of mass and energy conservation to environmental–resource economics was first emphasized by Kneese *et al.* (1970).

In reality, resource inputs originate outside the economic system *per se*: they include air, water, sunlight and material substances, fuels, food, and fiber crops, all of which embody free energy or available work. "Outputs" of the economic system, on the other hand, are final goods. They are ultimately discarded or, in rare cases, recycled. Wastes that are not recycled are, again, outside the economic system. Free energy is expended and lost at every state, viz. extraction, refining, manufacturing, construction, and even final consumption (Ayres, 1978). Though total energy is always conserved, free energy is not. Energy inputs (fuels) are rich in free energy, while energy outputs are mostly in the form of low-temperature waste heat, oxidation products, or degraded materials. Thus, the economic system, in reality, is absolutely dependent on a continuing flow of free energy from the environment. In preindustrial times, it was the sun that provided almost all free energy in the form of wood, food crops, animals, water power, or wind power. Today, the major source, by far, is fossil fuels: petroleum, natural gas, and coal from earth's crust. These resources are being exhausted, of course, and will eventually have to be replaced.

Evidently, the real economic system looks very much like a self-organizing dissipative structure in Prigogine's sense: it is dependent on a continuous flow of free energy (the sun or fossil fuels), and it exhibits coherent, orderly behavior. Moreover, like living organisms, it embodies structural information as morphological differentiation and functional specialization [5]. In economic terms, specialization and differentiation of form and function were first recognized as the "division of labor". Of course, production itself is highly differentiated into sectors, products, and services. Similarly, labor skills are increasingly subdivided into occupational classifications.

To carry the analogy with living systems a step further, economic systems are also capable of growth. Economic growth can be of two distinct kinds. First, an economic system can, in principle, expand in size, like a balloon, without accompanying technological or structural change. Output grows but structure complexity does not. It simply gets bigger, as capital and labor inputs increase proportionally. This kind of quasi-static growth can lead to increased final consumption per capita (for a stationary population) by producing more of everything, in exactly fixed ratios. This is only possible, however, if there are no economies or diseconomies of scale, which is an unrealistic but common economic assumption. Technological change is permitted in the quasi-static case, but only in the sense of gradual increases in productivity due to "learning". This kind of change is sometimes called "Usherian", in contrast to more radical "Schumpeterian" changes.

The second kind of growth is dynamic: it involves revolutionary (Schumpeterian) changes in structure, i.e., the creation of new sectors and the obsolescence or destruction of old ones. These changes are driven by radical innovations – new products, new processes – resulting not only in quantitative increases in per capita consumption, but also in qualitative changes in the mix of goods and services generated by the economy. In general, dynamic technology-driven growth results in increased complexity and an increased amount of information stored in the form of structure/organization. Recall the earlier discussion, especially equations (7)–(10).

Quasi-static growth "of the first kind" can be modeled theoretically as an optimal control model with aggregate consumption (or welfare) as the objective function. The control variable is the rate of savings diverted from immediate consumption to replace depreciated capital and add new capital to support a higher level of future consumption [6]. The rate of growth in a simple model of this kind is directly proportional to the rate of savings, which, in turn, depends on the assumed depreciation rate and an assumed temporal discount rate to compare present versus future benefits. Note that assumptions about the operation of the market play almost no role in this type of growth model. Savings, in this model, can be voluntary or enforced by government. Technological change is strictly exogenous.

It is noteworthy (and unfortunate) that many economic development programs in the Third World have been based on the generalized Harrod–Domar type of model, assuming growth of the first kind only. This model suggests a primary role for aggregate capital investment and depending on central planners to maintain balance between the capital needs of various sectors [7]. Yet, empirical research carried out as early as the 1950s established quite clearly that per capita economic growth in the USA and other industrial countries cannot be accounted for primarily in terms of increased capital inputs, e.g., Abramovitz (1956), Fabricant (1954), and Solow (1957). In fact, the linked notion of increasing factor productivity as a reflection of technological progress was introduced into economics at this time by Kendrick (1956). Unfortunately, the relative contribution of Usherian (incremental) vs. Schumpeterian (radical) innovation was not clarified by these early studies. However, the relatively poor performance of most centrally planned economic development programs is probably due in part to their focus on investment *per se*, to the neglect of parallel structural adjustment and radical innovation.

Dynamic growth "of the second kind" is less dependent on savings and/or capital investment and far more dependent on R&D and innovation. However, such growth cannot occur without capital investment since new production technologies, in particular, are largely embodied in capital equipment. Technological innovation drives this kind of dynamic growth, as will be discussed later.

An essential characteristic of self-organizing systems seems to be feedback control, e.g., Lotka (1950), Nicolis and Prigogine (1977), and Odum (1983).

This applies not only to stabilization processes, but also to growth processes. Feedback controls in complex systems involve information transfer. In biological organisms there are a number of specialized physical and chemical subsystems for monitoring the state-of-the-environment and the state-of-the-system and responding to changes in either.

In the economic system the same thing is true. In an idealized free-market economy, the "signals" are market prices. In an idealized centrally and planned economy, the necessary control information must be derived by administrative means (from data on production, shipments, inventory changes, sales, etc.). The processes of economic growth, in particular, also require a flow of information on new products and new means of production. In the distant past, such information was obtained by accident or by trial and error. In recent centuries it has become increasingly more intentional and more organized. Technology nowadays is created on purpose, as an explicit economic activity, although this activity has not yet been treated as an explicit sector of the economy.

8. Technology-creation system

The "technosystem" is, by definition, the creator of new techniques new products and new applications. Its activities can and do enable an economy to grow beyond the limits set by any given level of technology by finding more efficient methods to exploit existing resources, discovering new sources or finding viable substitutes, and discovering new products and processes. The technosystem operates within the larger economic framework, however. In particular, it is the macrosystem that determines both demand for technology and its supply. The impressions of certain social critics to the contrary, technology is not an autonomous or self-acting force outside its economic context.

Similarly, the economic system functions in a social framework, which in turn functions in an ecological–biological framework. The latter functions within climactic, geochemical, and astrophysical frameworks. The fundamental laws of physics (e.g., mechanics and thermodynamics) operate directly or indirectly at all levels of the hierarchy, including the highest. Basic biological laws also govern social behavior, and so on. On the other hand, higher-level laws are irrelevant at lower levels of the hierarchy.

A debate has raged for many decades over the extent to which technology is created in response to exogenously determined demand, vis-à-vis the extent to which supply create its own demand – a variant of Say's law in economics. Extensive empirical work by economists and sociologists of science tends toward the view that perceived demand is by far the dominant factor. That is to say, most successful inventors and innovations, and most industrial R&D establishments, have responded to a clearly articulated need by consumers, government, or industry itself. On the other hand, it could be argued that, occasionally, a spectacular technological opportunity comes along before there is any immediate need for it. Nuclear power is probably an apt example. The laser, invented in the early 1960s, seems to be another. But, in both cases, major future applications were immediately obvious – to the point of stimulating continuing R&D expenditures.

Quite apart from the question of primacy of demand versus supply, however, it is clear that the economic and political frameworks determine the pattern of prices, including wages, and profits that actually govern the existing allocation of societal resources to, and within, the technosystem.

The pattern of prices, wages, and profits constitutes a set of signals, transmitted by society as a whole (i.e., consumers, government, and industry) that guide individual decisions through established market mechanisms. For example, enrollment in engineering schools, competing with liberal arts schools, apparently reflects relative salaries and job prospects in the different fields. Similarly, investors tend to move out of stagnant or unprofitable sectors and into more profitable, growing sectors.

Signals are sometimes confused, as when government interference or private collusion distort the operations of the competitive market. In addition, there are pervasive market imperfections. Some of these can only be compensated for by government action. One of these imperfections is the inherent difficulty of protecting technological information, which makes it relatively easy for imitators and "free riders" to prosper and inhibits the development of an effective market for exchanging technological knowledge in pure form (i.e., not embodied in any product). This, in turn, makes it impossible for those who invest in new knowledge to capture more than a small fraction of the benefits, in most cases. The consequence is to discourage such investment by the private sector.

It follows (*ceteris paribus*) that the private sector tends to underinvest in R&D. The fact that R&D expenditures are nevertheless very high in some industries (e.g., electronics and pharmaceuticals) does not contradict this result. Quite possibly, the social optimum would be still higher. On the other hand, there may be other explanations for the observed facts, since organizational behavior in the real world is more complex than the neoclassical model. The public sector must make up the gap, particularly in those areas where the private incentives are most lacking. One cogent example of such an area is the development of technical means for controlling pollutant emissions. Pollution is, in itself, a market imperfection, and there is very little profit motive in this field. Other areas of minimal private incentive are the mature public sector monopolies, e.g., defense, public health, and public safety.

It is axiomatic that technological progress depends on the knowledge base, and that the knowledge base can be increased, at the margin, by deliberate investment in R&D. One question of fundamental interest is why individuals or enterprises should invest in R&D. This is tantamount to asking: how and why does an R&D investment pay off in economic terms? The general outlines of an answer to this question have been clear for some time. Many detailed issues, however, still remain to be cleared up. However, in contrast to the case of quasi-static economic growth, which is driven by savings and does not depend on market structure, there is reason to believe that market structure plays a significant role in the process of technology creation. Joseph Schumpeter (1912, 1961) first pinpointed the driving force underlying dynamic economic growth as technological innovation by entrepreneurs seeking "supernormal" profits. Such profits arise from a temporary monopoly position conferred by each innovation until successful imitators are able to enter the market. Many questions raised by Schumpeter's theory remain to be answered, but economists today are increasingly inclined to accept his basic hypothesis.

9. Thermodynamic constraints on economic growth

Several themes from the prior discussion can now be summarized in terms of their implications for economic growth:

(1) Since the economy is, by assumption, a dissipative structure, it depends on a continuous flow of free energy and materials from and to the environment. Such links are precluded by closed neoclassical general equilibrium models, either static or quasi-static.

(2) The energy and physical materials inputs to the economy have shifted over the past two centuries from mainly renewable to mainly nonrenewable sources.

(3) Dynamic economic growth is driven by technological change (generated, in turn, by economic forces), which also results in continuous structural change in the economic system. For instance, so-called Leontief input–output coefficients do not remain constant.

(4) It follows, incidentally, that a long-term survival path must sooner or later reverse the historical shift away from renewable resources. This will only be feasible if human technological capabilities continue to rise to levels much higher than current ones [8]. But, since technological capability is itself an output of the economic system, it will continue to increase if, and only if, deliberate investment in R&D is continued or even increased.

In short, the role of knowledge-generating activity in retarding global entropy seems to be growing in importance.

Looking at it in another way, external resource constraints in themselves may not constitute an ultimate limit to growth, since technological improvements and substitutions appear to offer a possible way out. This has always been the basis on which most scientists and economists have criticized the "limits" thesis of the Club of Rome and others. But the critique itself has tended to assume that new technology always appears (essentially) costlessly in response to any perceived scarcity or need. This is not the case in reality. Large-scale future substitutions, such as the eventual replacement of motor gasoline (perhaps by methanol or ethanol) will necessarily entail massive R&D investments for scaling up from pilot plants to full-scale production, not to mention even more massive capital outlays for construction. But because of market failures, the existing private incentives to invest in this kind of research may be inadequate while governments may neglect it for short-term political reasons (e.g., "industry should do it"). The economic system is not necessarily stable against all perturbations, and the more it is intentionally managed to optimize growth, the more it becomes vulnerable to the consequences of human error.

APPENDIX*

A.1. The 20 critical subsystems of a living system.

Subsystems that process both matter-energy and information

1. *Reproducer* carries out instructions in the genetic information or charter of a system and mobilizes matter and energy to produce one or more similar systems.
2. *Boundary* at the perimeter of a system holds together the components that make up the system, protects them from environmental stresses, and excludes or permits entry to various sorts of matter-energy and information.

Subsystems that process matter-energy

3. *Ingestor* brings matter-energy across the system boundary from the environment.
4. *Distributor* carries inputs from outside the system or outputs from its subsystems around the system to each component.

Subsystems that process information

11. *Input transducer*, a sensory subsystem, brings markers bearing information into the system, changing them to other matter-energy forms for transmission within it.
12. *Internal transducer*, another sensory subsystem, receives, from subsystems or components within the system, markers bearing information about significant alterations in those subsystems or components, changing them to other matter-energy forms that can be transmitted within it.
13. *Channel and net* are composed of a single route in physical space or multiple interconnected routes over which markers bearing information are transmitted to all parts of the system.
14. *Timer*, the clock, set by information from the input transducer about states of the environment, uses information about processes in the system to measure the passage of time, and transmits to the decider signals that facilitate coordination of the system's processes in time.

5. *Converter* changes certain inputs to the system into forms more useful for the special processes of that particular system.

6. *Producer* forms stable associations that endure for significant periods among matter-energy inputs to the system or outputs from its converter, the materials synthesized being for growth, damage repair, or replacement of components of the system, or for providing energy for moving or constituting the system's outputs of products or information markers to its suprasystem.
7. *Matter-energy storage* places matter or energy at some location in the system, retains it over time, and retrieves it.

8. *Extruder* transmits matter-energy out of the system in the forms of products or wastes.
9. *Motor* moves the system or parts of it in relation to part or all of its environment or moves components of its environment in relation to each other.
10. *Supporter* maintains the proper spatial relationships among components of the system, so that they can interact without weighting each other down or crowding each other.

15. *Decoder* alters the code of information input to it through the input transducer or internal transducer into a "private" code that can be used internally by the system.

17. *Memory* carries out the second stage of the learning process, storing information in the system for different periods of time, and then retrieving it.
18. *Decider*, an executive subsystem, receives information inputs from all other subsystems and transmits to them outputs for guidance, coordination, and control of the system.
19. *Encoder* alters the code of information input to it from other information-processing subsystems, from a "private" code used internally by the system into a "public" code that can be interpreted by other systems in its environment.

20. *Output transducer* emits markers bearing information from the system, changing markers within the system into other matter-energy forms that can be transmitted over channels in the system's environment.

*Both sections adapted from Miller (1987).

A.2. Selected major components of the 20 critical subsystems at the three levels of living systems.

	Level		
Subsystem	*Cell*	*Organ*	*Organism*
1. *Reproducer*	DNA and RNA molecules	Upwardly dispersed to organism	Testes, ovaries, uterus, genitalia
2. *Boundary*	*Matter-energy and information*: outer membrane	*Matter-energy and information*: capsule or outer layer	*Matter-energy and information*: skin or other outer covering
3. *Ingestor*	Transport molecules	Input artery	Mouth, nose, skin in some species
4. *Distributor*	Endoplasmic reticulum	Intercellular fluid	Vascular system of higher animals
5. *Converter*	Enzyme in mitochondrion	Gastric mucosa cell	Upper gastrointestinal tract
6. *Producer*	Chloroplast in green plant	Islets of Langerhans of pancreas	Organs that synthesize materials for metabolism and repair
7. *Matter-energy storage*	Adenosine triphosphate	Central lumen of glands	Fatty tissues
8. *Extruder*	Contractile vacuoles	Output vein	Sweat glands of animal skin
9. *Motor*	Cilia, flagellae, pseudopodia	Smooth muscle, cardiac muscle	Skeletal muscle of higher animals
10. *Supporter*	Cytoskeleton	Stroma	Skeleton
11. *Input transducer*	Receptor sites on membrane for activation of cyclic AMP	Receptor cell of sense organ	Sense organs
12. *Internal transducer*	Repressor molecules	Specialized cell of sinoatrial node of heart	Proprioceptors
13. *Channel and net*	Pathways of mRNA, second messengers	Nerve net or organ	Hormonal pathways, central and peripheral nerve nets
14. *Timer*	Components not known	Upwardly dispersed to organism	Supraoptic nuclei of thalamus
15. *Decoder*	Molecular binding sites	Second echelon cell of sense organ	Sensory nuclei

Subsystem	Level: Cell	Level: Organ	Level: Organism
16. *Association*	Unknown	None found; upwardly dispersed to organism	Unknown neural components
17. *Memory*	Unknown	None found; upwardly dispersed to organism	Unknown neural components
18. *Decider*	Regulator genes	Sympathetic fibers of sinoatrial node of heart	Components at several echelons of nervous system
19. *Encoder*	Structure that synthesizes hormones	Presynaptic region of output neuron	Temporoparietal area of dominant hemisphere of human cortex
20. *Output transducer*	Presynaptic membrane of neuron	Presynaptic region of output neuron	Larynx; other components that put out signals

Notes

[1] The objection raised by Thompson had to do with the geological age of the earth. Lacking any inkling of the true source of the sun's heat or the earth's interior heat (nuclear reactions), physicists calculated a maximum age for the earth to be of the order of 25 million years – far too brief to accommodate Darwin's theory of slow natural selection.

[2] Some scientists, including Brillouin (1949), even suggested that the second law of thermodynamics might not be applicable in open systems. However, this sort of speculation appears unfounded, in view of empirical evidence that biological growth processes actually dissipate much more energy than ordinary metabolic (maintenance) processes. Prigogine *et al.* (1972) raise the question of "how living systems have acquired the ability to dissipate intensely".

[3] Astronomy evolved historically as means of predicting dates of annual spring runoff in the Nile Valley, where seasonal changes at lower latitudes are very slight.

[4] Since this paper was written I have become aware that others are thinking along similar lines, notably Silverberg (1987a, b) and Dosi *et al* (1986).

[5] In fact, the economic system includes at least 15 of the 20 characteristic subsystems of living systems (*Appendix*), lacking only the motor (9), the input transducer (11), the encoder (19), and the output transducer (20). The last three subsystems are absent because the ecosystem is unique: it does not interact with competitors, predators, prey, etc.

[6] Aggregative models have been studied by Harrod (1936) and Domar 1956). Sectoral growth models have been studied by von Neumann (1945) and others. The literature is well summarized by Burmeister and Dobell (1970).

[7] Harrod (1936) called this balancing process "walking on the razor's edge". However, it was later shown that the Harrod–Domar models' extreme sensitivity to balancing is an artifact of their particular choice of production function (Solow, 1956).

[8] An optimal growth model incorporating the information perspective discussed in this paper, but focusing on the substitution of renewable resources for exhaustible resources, is described in detail in Ayres (1987a).

References

Abramovitz, M. (1956), Resources and output trends in the United States since 1870, *American Economic Review* **46** (May).

Allred, J.C. (1977), Application of entropy concepts to national energy problems, Entropy Economics Symposium, Houston, TX, November 2. (NOTE: Unpublished.)

Ayres, R.U. (1978), *Resources, Environment and Economics: Applications of the Materials/Energy Balance Principle*, John Wiley, New York, NY.

Ayres, R.U. (1987a), *The Industry-Technology Life Cycle: An Integrating Meta-Model?* Research Report RR-87-3 (March), International Institute for Applied Systems Analysis, Laxenburg, Austria.

Ayres, R.U. (1987b), *Manufacturing and Human Labor as Information Processes*, Research Report RR-87-19 (November), International Institute for Applied Systems Analysis, Laxenburg, Austria.

Berg, C. (1987), Letter to the author, October 22.

Blum, H.F. (1955), *Time's Arrow and Evolution*, Princeton University Press, Princeton, NJ.

Boltzman, L. (1896), *Vorlesungen über Gastheorie*, Barth, Leipzig.

Boulding, K.E. (1966), Environmental quality in a growing economy, in: *Essays from the Sixth RFF Forum*, Johns Hopkins University Press, Baltimore, MD.

Breder, C. (1942), A consideration of evolutionary hypotheses in reference to the origin of life, *Zoologica* **27**: 132–143.

Brillouin, L. (1949), Life, thermodynamics and cybernetics, *American Scientist* **37**: 554–568.

Brillouin, L. (1951), Physical entropy and information, *Journal of Applied Physics* **22** (3): 338–343.

Brillouin, L. (1962), *Science and Information Theory*, 2nd edition, Academic Press, New York, NY.

Burmeister, E. and Dobell, A.R. (1970), *Mathematical Theories of Economic Growth*, Macmillan, London.

Cherry, C. (1978), *On Human Communications*, 3rd edition, MIT Press, Cambridge, MA.

Crick, H.C. (1962), The genetic code, *Scientific American* (Oct).

Daly, H. (1980), The economic thought of Frederick Soddy, in: *History of Political Economy*, **12** (4): 469–488, Duke University Press, Durham, NC.

Domar, E.D. (1956), Capital expansion, rate of growth and employment, *Econometrica* **14**: 137–147.

Dosi, G., Orsenigo, L., and Silverberg, G. (1986), *Innovation, Diversity and Diffusion: A Self-Organization Model* (unpublished).

Dyson, F.J. (1982), Origin of life, *Journal of Molecular Evolution* **18**: 344–355.

Eigen, M. (1971), Self-organization of matter and the evolution of biological macromolecules, *Die Naturwissenschaften* **58** (Oct).

Enos, J.L. (1962), *Petroleum Progress and Profits: A History of Process Innovation*, MIT Press, Cambridge, MA.

Fabricant, S. (1954), Economic progress and economic change, in: National Bureau of Economic Research, *34th Annual Report*, NBER, New York, NY.

Field, R.J. (1985), Chemical organization in time and space, *American Scientist* **73** (Mar–Apr): 142–149.

Gaffron, H. (1965), The role of light in evolution – the transition from a one quantum to a two quantum mechanism, p. 450 in: S.W. Fox (ed), *The Origins of Prebiological Systems and of Their Molecular Matrices*, Academic Press, New York, NY.

Georgescu-Roegen, N. (1971), *The Entropy Law and the Economic Process*, Harvard University Press, Cambridge, MA.

Glansdorff, P. and Prigogine, I. (1971), *Thermodynamic Theory of Structure, Stability and Fluctuations*, Wiley-Interscience, New York, NY.

Graham, R. and Haken, H. (1970), *Zeitschrift für Physik* **237** (31).

Harrod, R.F. (1936), *The Trade Cycle*, Oxford University Press, London.

Hartley, R.V.L. (1928), Transmission of information, *Bell System Technical Journal* **7**.

Huxley, J.S. (1956), Evolution, cultural and biological, pp. 3–25 in: W.I. Thomas (ed), *Current Anthropology*, University of Chicago Press, Chicago, IL.

Jaynes, E.T. (1957a), Information theory and statistical mechanics, I, *Physical Review* **106**: 620.

Jaynes, E.T. (1957b), Information theory and statistical mechanics, II, *Physical Review* **108**: 171.

Jensen, R.V. (1987), Classical chaos, *American Scientist* **75** (Mar–Apr): 168–181.

Johnstone, J. (1921), *The Mechanism of Life in Relation to Modern Physical Theory*, Longmans, London.

Kendrick, J.W. (1956), Productivity trends: capital and labor, *Review of Economics and Statistics* (Aug).

Khinchin, A.I. (1957), *Mathematical Foundations of Information Theory*, Dover, New York, NY.

Kneese, A.V., Ayres, R.U., and d'Arge, R.C. (1970), *Aspects of Environmental Economics: A Materials Balance – General Equilibrium Approach*, Johns Hopkins University Press, Baltimore, MD.

Lotka, A.J. (1945), The law of evolution as a maximal principle, *Human Biology* **17** (3): 167–194.

Lotka, A.J. (1950), *Elements of Physical Biology*, Dover, New York, NY. (NOTE: First published in 1925.)

May, R.M. (1876), Simple mathematical models with very complicated dynamics, *Nature* **261**: 459.

McGlashan, M.L. (1966), The use and misuse of the laws of thermodynamics, *Journal of Chemical Education* **43** (5): 226–232.

Miller, J.G. (1978), *Living Systems* **5**, McGraw-Hill, New York, NY.

Miller, J.G. (1987), Keynote address at IIASA Workshop on "Life Cycles and Long Cycles", Montpellier, France, July 8. (NOTE: Unpublished.)

Miller, S. and Orgell, L. (1974), *The Origins of Life on the Earth*, Prentice-Hall, Englewood Cliffs, NJ.

Moravec, H., *Mind Children: The Future of Robot and Human Intelligence*, to be published by Harvard University Press (forthcoming).

Moravec, H. (1985), *Life, the Universe and Everything*, unpublished, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA.

Needham, J. (1943), *Time, the Refreshing River*, Allen & Unwin, London.

Nicolis, G. and Prigogine, I. (1977), *Self-Organization in Non-Equilibrium Systems*, Wiley-Interscience, New York, NY.

Odum, H.T. (1983), *Systems Ecology: An Introduction*, John Wiley, New York, NY.

Oparin, A.I. (1953), *Origin of Life*, Dover, New York, NY.

Ott, E. (1981), Strange attractors and chaotic motions of dynamical systems, *Review of Modern Physics* **53**: 655.

Polgar, S. (1961), Evolution and the thermodynamic imperative, *Human Biology* **33** (2): 99–109.

Prigogine, I., Nicolis, G., and Babloyantz, A. (1972), Thermodynamics of evolution, *Physics Today* **23** (11): 23–28; (12): 38–44.

Sagan, C. (1977), *Dragons of Eden – Speculations on the Evolution of Human Intelligence*, Random House, New York, NY.

Schrödinger, E. (1945), *What Is Life? The Physical Aspects of the Living Cell*, Cambridge University Press, London.

Schumpeter, J.A. (1912), *Theorie der wirtschaftlichen Entwicklungen*, Duncker & Humboldt, Leipzig.

Schumpeter, J.A. (1961), *The Theory of Economic Development*, Oxford University Press, New York, NY. (NOTE: English translation from the 1911 German edition.)

Shannon, C.E. (1948), A mathematical theory of communication, *Bell System Technical Journal* **27**.

Shannon, C.E. and Weaver, W. (1949), *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, IL.

Silverberg, G. (1987a), Technical progress, capital accumulation and effective demand: A self-organization model, in: D.F. Batten, J. Casti, B. Johansson (eds) *Economic Evolution and Structural Adjustment*, Springer-Verlag, Berlin.

Silverberg, G. (1987b), Modelling economic dynamics and technical change: Mathematical approaches to self-organization and evolution, to appear in: G. Dosi, C. Freeman, R. Nelson, G. Silverberg and L. Soete (eds) *Technical Change and Economic Theory* (in press).

Solow, R.M. (1956), A contribution to the theory of economic growth, *Quarterly Journal of Economics* **70**: 65–94.

Solow, R.M. (1957), Technical change and the aggregate production function, *Review of Economics and Statistics* (Aug).

Thom, R. (1972), *Structural Stability and Morphogenesis. General Theory of Models.* Benjamin-Cummings Pubs., Reading, MA.

Tribus, M. (1961a), Information theory as the basis for thermostatics and thermodynamics, *Journal of Applied Mechanics* **28** (Mar): 1–8.

Tribus, M. (1961b), *Thermostatics and Thermodynamics*, Van Nostrand, New York, NY.

Tribus, M. and McIrvine, E.C. (1971), Energy and information, *Scientific American* **225** (Sept): 179–188.

Ulanowicz, R. (1986), *Growth and Development: Ecosystem Phenomenology*, Springer-Verlag, New York, NY.

Velarde, M. and Normand, C. (1980), Convection, *Scientific American* **243** (July).

von Neumann, J. (1945), A model of general economic equilibrium, *Review of Economic Studies* **13**: 1–9.

Wald, G. (1955), *The Origin of Life in the Physics and Chemistry of Life*, Simon & Schuster, New York, NY.