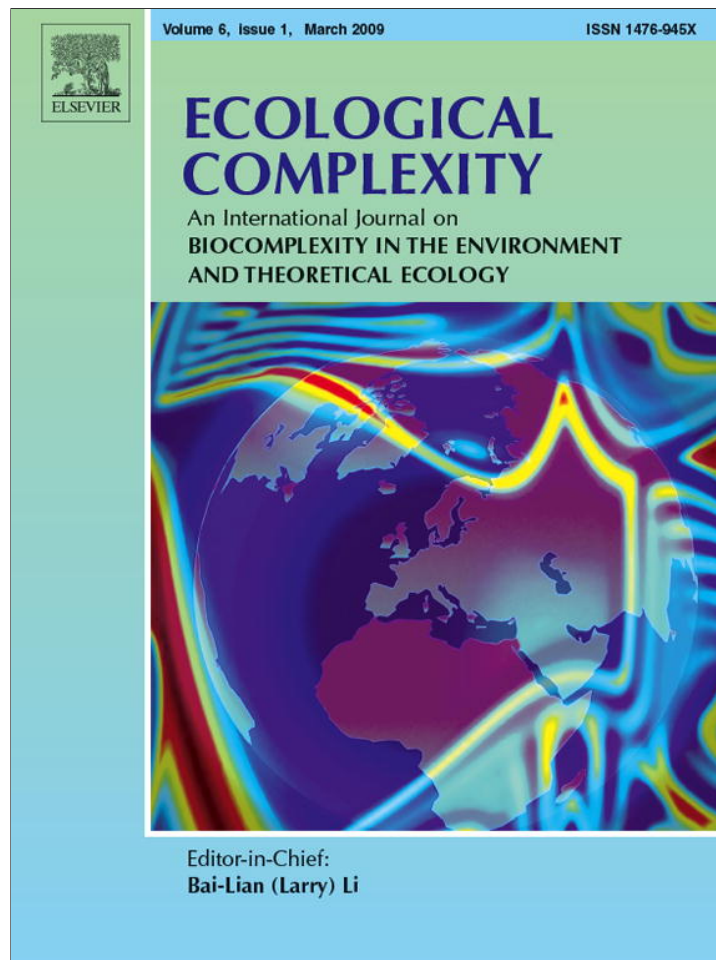


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>

available at www.sciencedirect.comjournal homepage: <http://www.elsevier.com/locate/ecocom>

Quantifying sustainability: Resilience, efficiency and the return of information theory

Robert E. Ulanowicz^{a,*}, Sally J. Goerner^b, Bernard Lietaer^c, Rocio Gomez^d

^a University of Maryland Center for Environmental Science, Chesapeake Biological Laboratory, Solomons, MD 20688-0038, USA

^b Integral Science Institute, 374 Wesley Ct, Chapel Hill, NC 27516, USA

^c Center for Sustainable Resources, University of California, 101 Giannini Hall, Berkeley, CA 94720-3100, USA

^d School of Computer Science, University of Birmingham, Birmingham B15 2TT, United Kingdom

ARTICLE INFO

Article history:

Received 8 November 2007

Received in revised form

26 May 2008

Accepted 1 October 2008

Published on line 29 November 2008

Keywords:

Apophysis

Ascendency

Biodiversity

Fitness

Information theory

Reserve capacity

Resilience

Stability

Sustainability

Window of vitality

ABSTRACT

Contemporary science is preoccupied with that which exists; it rarely accounts for what is missing. But often the key to a system's persistence lies with information concerning lacunae. Information theory (IT), predicated as it is on the indeterminacies of existence, constitutes a natural tool for quantifying the beneficial reserves that lacunae can afford a system in its response to disturbance. In the format of IT, unutilized reserve capacity is complementary to the effective performance of the system, and too little of either attribute can render a system unsustainable. The fundamental calculus of IT provides a uniform way to quantify both essential attributes – effective performance and reserve capacity – and results in a single metric that gauges system sustainability (robustness) in terms of the tradeoff allotment of each. Furthermore, the same mathematics allows one to identify the domain of robust balance as delimited to a “window of vitality” that circumscribes sustainable behavior in ecosystems. Sensitivity analysis on this robustness function with respect to each individual component process quantifies the value of that link “at the margin”, i.e., how much each unit of that process contributes to moving the system towards its most sustainable configuration. The analysis provides heretofore missing theoretical justification for efforts to preserve biodiversity whenever systems have become too streamlined and efficient. Similar considerations should apply as well to economic systems, where fostering diversity among economic processes and currencies appears warranted in the face of over-development.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction: the importance of being absent

The late Bateson (1972) observed that science deals overwhelmingly with things that are present, like matter and energy. One has to dig deeply for exceptions in physics that address the absence of something (like the Pauli Exclusion Principle, or Heisenberg's uncertainty). Yet any biologist can

readily point to examples of how the absence of something can make a critical difference in the survival of a living system. Nonetheless, because biology aspires to becoming more like physics, very little in quantitative biology currently addresses the important roles that lacunae play in the dynamics of living systems.

One might object that the use of information theory (IT) in genomics does indeed address matters like missing alleles, but

* Corresponding author. Tel.: +1 410 326 7266; fax: +1 410 326 7378.

E-mail addresses: ulan@cbl.umces.edu (R.E. Ulanowicz), sgoerner@mindspring.com (S.J. Goerner), blietaer@earthlink.net (B. Lietaer), Rocio@rociogomez.com (R. Gomez).

1476-945X/\$ – see front matter © 2008 Elsevier B.V. All rights reserved.

doi:10.1016/j.ecocom.2008.10.005

the emphasis in bioinformatics remains on information as a positive essence—as something that is transferred between a sender and a receiver. Biology at large has yet to reckon with Bateson’s insight that IT addresses apophysis directly as something even more fundamental than communication theory. In particular, IT is a means for apprehending and quantifying that which is missing. At the same time Bateson made his penetrating observation, he also defined information as any “difference that makes a difference”, and such difference almost always involves the absence of something.

All too many investigators, and even some theoreticians of information, remain unaware that IT is predicated primarily upon the notion of the negative. That this was true from the very beginning can be seen in Boltzmann’s famous definition of surprisal,

$$s = -k \log(p), \tag{1}$$

where s is one’s surprisal at seeing an event that occurs with probability p , and k is an appropriate (positive) scalar constant. Because the probability, p , is normalized to a fraction between zero and one, most offhandedly conclude that the negative sign is a mathematical convenience to make s work out positive (and that may have been Boltzmann’s motivation). But from the perspective of logic one can only read this equation as defining s to gauge what p is not. That is, if p is the weight we give to the presence of something, then s becomes a measure of its absence.¹ If p is very small, then the ensuing large magnitude of s reflects the circumstance that most of the time we do not see the event in question.

Boltzmann’s gift to science – the feasibility of quantifying what is not – remains virtually unappreciated. It is akin to the contribution of the Arabian mathematicians who invented the number zero. (Anyone who doubts the value of that device should try doing long division using Roman numerals.) In particular, we shall attempt to build upon Boltzmann’s invention and to demonstrate that IT literally opens new vistas to which classical physics remains blind. More importantly, the interplay between presence and absence plays a crucial role in whether a system survives or disappears. As we shall see, it is the very absence of order (in the form of a diversity of processes) that makes it possible for a system to persist (sustain itself) over the long run.

2. Evolution and indeterminacy

That Boltzmann’s definition is actually a quantification of the negative lends an insight into IT that few appreciate—namely, that the product of the measure of the presence of an event, i , (p_i) by a magnitude of its absence (s_i) yields a quantity that represents the *indeterminacy* (h_i) of that event,

$$h_i = -k p_i \log(p_i) \tag{2}$$

When $p_i \approx 1$, the event is almost certain, and $h_i \approx 0$; then when $p_i \approx 0$, the event is almost surely absent, so that again $h_i \approx 0$. It is only for intermediate, less determinate values of

p_i that h_i remains appreciable, achieving its maximum at $p_i = (1/e)$.

It is helpful to reinterpret (2) in terms germane to evolutionary change and sustainability. When $p_i \approx 1$, the event in question is a virtual constant in its context and unlikely to change ($h_i \approx 0$). Conversely, whenever $p_i \approx 0$, the event exhibits great potential to change matters ($s_i \gg 1$), but it hardly ever appears as a player in the system dynamics (so that, again, $h_i \approx 0$). It is only when p_i is intermediate that the event is both present frequently enough and has sufficient potential for change. In this way, h_i represents the capacity for event i to be a significant player in system change or evolution.

Seeking a perspective on the entire ensemble of events motivates us to calculate the aggregate systems indeterminacy, H , as

$$H \sum_i h_i = -k \sum_i p_i \log(p_i), \tag{3}$$

which we can now regard as a metric of the total capacity of the ensemble to undergo change. Whether such change will be coordinated or wholly stochastic depends upon whether or not the various events i are related to each other and by how much. In order for any change to be meaningful and directional, constraints must exist among the possible events (Atlan, 1974).

In order better to treat relationships between events, it is helpful to consider bilateral combinations of events, which for clarity requires two indices. Accordingly, we will define p_{ij} as the joint probability that events i and j co-occur. Boltzmann’s measure of the non-occurrence of this particular combination of events (1) thus becomes,

$$s_{ij} = -k \log(p_{ij}). \tag{4}$$

If events i and j are entirely independent of each other, the joint probability, p_{ij} , that they co-occur becomes the product of the marginal probabilities that i and j each occur independently anywhere. Now, the marginal probability that i occurs for any possible j is $p_i = \sum_j p_{ij}$, while the likelihood that j occurs regardless of i is $p_j = \sum_i p_{ij}$.² Hence, whenever i and j are totally independent, $p_{ij} = p_i p_j$. Here the assumption is made that the indeterminacy s_{ij} is maximal when i and j are totally independent. We call that maximum s_{ij}^* . The difference by which s_{ij}^* exceeds s_{ij} in any instance then becomes a measure of the constraint that i exerts on j , call it x_{ij} , where,

$$x_{ij} = s_{ij}^* - s_{ij} = -k \log(p_i p_j) - [-k \log(p_{ij})] = k \log \left(\frac{p_{ij}}{p_i p_j} \right) = x_{ji}. \tag{4}$$

The symmetry in (4) implies that the measure also describes the constraint that j exerts upon i . In other words (4) captures the mutual constraint that i and j exert upon each other (an analog of Newton’s Third Law of motion).

In order to calculate the average mutual constraint (X) extant in the whole system, one weights each x_{ij} by the joint

¹ Here the reader might ask why the lack of i is not represented more directly by $(1 - p_i)$? The advantage and necessity of using the logarithm will become apparent presently.

² For the remainder of this essay a dot in the place of an index will represent summation over that index.

probability that i and j co-occur and sums over all combinations of i and j :

$$X = \sum_{i,j} p_{ij} x_{ij} = k \sum_{i,j} p_{ij} \log\left(\frac{p_{ij}}{p_i p_j}\right) \quad (5)$$

Here is where the advantage of (1) as the formal estimate of lacunae becomes apparent, because the convexity of the logarithmic function guarantees (Abramson, 1963) that:

$$H \geq X \geq 0 \quad (6)$$

In words, (6) says that the aggregate indeterminacy is an upper bound on how much constraint (order) can appear in a system. Most of the time, $H > X$, so that the difference

$$\psi = (H - X) = -k \sum_{i,j} p_{ij} \log\left(\frac{p_{ij}^2}{p_i p_j}\right) \geq 0 \quad (7)$$

as well. In the jargon of IT ψ is called the “conditional entropy”.

Relationship (7) can be rewritten as

$$H = X + \psi, \quad (8)$$

and it makes a very valuable statement. It says that the capacity for evolution or self-organization (H) can be decomposed into two components. The first (X) quantifies all that is regular, orderly, coherent and efficient. It encompasses all the concerns of conventional science. By contrast, ψ represents the lack of those same attributes, or the irregular, disorderly, incoherent and inefficient behaviors. It quantifies and brings into the scientific narrative a host of behaviors that heretofore had remained external to scientific discourse. Furthermore, it does so in a way that is wholly commensurate with X , the usual object of interest. The key point is that, if one is to address the issues of persistence and sustainability, ψ becomes the indispensable focus of discussion, because it represents the reserve that allows the system to persist (Conrad, 1983). To help one see this, it is useful to demonstrate how one might attach numbers to these quantities.

3. Measuring the missing

Up to this point we have spoken only vaguely about events i and j . Without loss of generality, we now narrow our discussion to consider only transfers or transformations. That is, event i will signify that some quantum of medium leaves or disappears from component i . Correspondingly, event j will signify that a quantum enters or appears in component j . We now identify the aggregation of all quanta both leaving i and entering j during a unit of time – or, alternatively, the flow from i to j (or the transformation of i into j) – as T_{ij} . Thus, T_{ij} might represent the flow of electrons from point i to point j in an electrical circuit; the flow of biomass from prey i to predator j in an ecosystem; or the transfer of money from sector i to sector j in an economic community.

We maintain the convention introduced earlier that a dot in the place of a subscript indicates summation over that index. Thus $T_i (= \sum_j T_{ij})$ will represent everything leaving i during the unit time interval, and T_j will gauge everything entering j during

the same duration. In particular, $T_{..} (= \sum_{i,j} T_{ij})$ represents the total activity of the system and is given the name “total system throughput”.

These definitions allow us to estimate all the probabilities defined above in terms of their measured frequencies of occurrence. That is,

$$p_{ij} \sim \frac{T_{ij}}{T_{..}}, \quad p_i \sim \frac{T_i}{T_{..}}, \quad \text{and} \quad p_j \sim \frac{T_j}{T_{..}} \quad (9)$$

Substituting these estimators in Eqs. (3), (5) and (7), yields

$$H = -k \sum_{i,j} \frac{T_{ij}}{T_{..}} \log\left(\frac{T_{ij}}{T_{..}}\right), \quad X = k \sum_{i,j} \frac{T_{ij}}{T_{..}} \log\left(\frac{T_{ij} T_{..}}{T_i T_j}\right), \quad \text{and} \\ \psi = -k \sum_{i,j} \frac{T_{ij}}{T_{..}} \log\left(\frac{T_{ij}^2}{T_i T_j}\right), \quad (10)$$

respectively.

The dimensions in the definitions (10) remain problematic, however. All of the ratios that occur there are dimensionless (as required of probabilities), so that the only dimensions that the variables H , X and ψ carry are those of the base of the logarithm used in their calculation. For example, if the base of the logarithm is 2, the variables are all measured in bits. (A “bit” is the amount of information required to resolve one binary decision.) Unfortunately, bits do not convey any sense of the physical magnitude of the systems to which they pertain. For example, a network of flows among the populations of microbes in a Petri Dish could conceivably yield an H of the same order of magnitude as a network of trophic exchanges among the mammalian species on the Serengeti Plain.

Tribus and McIrvine (1971) spoke to this inadequacy of information indices and suggested that the scalar constant, k , which appears in each definition, be used to impart physical dimensions to the measures. Accordingly, we elect to scale each index by the total system throughput, $T_{..}$, which conveys the overall activity of the system. In order to emphasize the new nature of the results, we give them all new identities. We call

$$C = T_{..} \cdot H = - \sum_{i,j} T_{ij} \log\left(\frac{T_{ij}}{T_{..}}\right) \quad (11)$$

the “capacity” for system development (Ulanowicz and Norden, 1990). The scaled mutual constraint,

$$A = T_{..} \cdot X = \sum_{i,j} T_{ij} \log\left(\frac{T_{ij} T_{..}}{T_i T_j}\right), \quad (12)$$

we call the system “ascendency” (Ulanowicz, 1980). The scaled conditional entropy,

$$\phi = T_{..} \cdot \psi = - \sum_{i,j} T_{ij} \log\left(\frac{T_{ij}^2}{T_i T_j}\right), \quad (13)$$

we rename the system “reserve”,³ for reasons that soon should become apparent.

³ From here on “reserve” will apply to what heretofore has been called “reserve capacity”.

4. A two-tendency world

Of course, this uniform scaling does not affect the decomposition (8), which now appears as

$$C = A + \Phi. \tag{14}$$

In other words, (14) says that the capacity for a system to undergo evolutionary change or self-organization consists of two aspects: It must be capable of exercising sufficient directed power (ascendency) to maintain its integrity over time. Simultaneously, it must possess a reserve of flexible actions that can be used to meet the exigencies of novel disturbances. According to (14) these two aspects are literally complementary. That they are conceptually complementary as well is suggested by the following example:

Fig. 1 depicts three pathways of carbon flow ($\text{mg C m}^{-2} \text{y}^{-1}$) in the ecosystem of the cypress wetlands of S. Florida that lead from freshwater shrimp (prawns) to the American alligator via the intermediate predator categories—turtles, large fish and snakes (Ulanowicz et al., 1996).

Of course, these species are entwined in a myriad of relationships with other populations, but for the purposes of illustrating a point, this sub-network will be considered as if it existed in isolation. $T_{..}$ for this system is $102.6 \text{ mg C m}^{-2} \text{y}^{-1}$; the ascendency, A , works out to $53.9 \text{ mg C-bits m}^{-2} \text{y}^{-1}$ and the reserve, Φ , is $121.3 \text{ mg C-bits m}^{-2} \text{y}^{-1}$.

Inspection of the pathways reveals that the most efficient pathway between prawns and alligators is via the large fishes. If efficiency were the sole criterion for development, the route via the fish would grow at the expense of the less efficient pathways until it dominated the transfer, as in Fig. 2.

The total system throughput of the simplified system rises to $121.8 \text{ mg C m}^{-2} \text{y}^{-1}$, as a result of the increase in overall efficiency, but the greatest jump is seen in the ascendency, A , which almost doubles to $100.3 \text{ mg C-bits m}^{-2} \text{y}^{-1}$. Meanwhile, the reserve has vanished completely ($\Phi = 0$). To use a cliché,

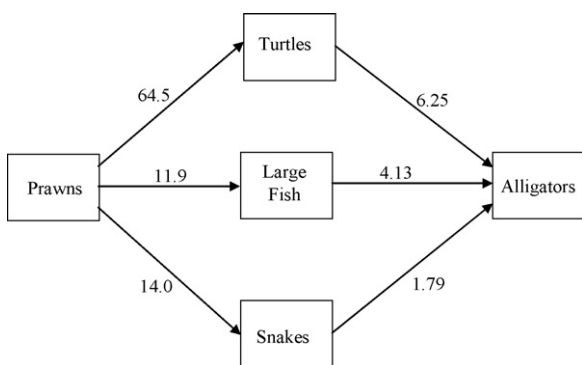


Fig. 1 – Three pathways of carbon transfer ($\text{mg C m}^{-2} \text{y}^{-1}$) between prawns and alligators in the cypress wetland ecosystem of S. Florida (Ulanowicz et al., 1996).

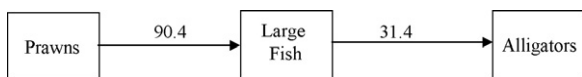


Fig. 2 – The most efficient pathway in Fig. 1 after it had eliminated parallel competing pathways.

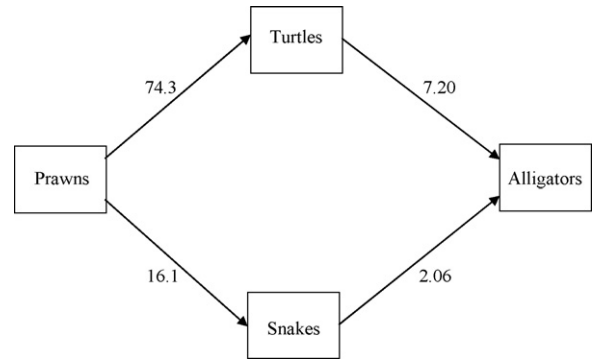


Fig. 3 – Possible accommodation by turtles and snakes to the disappearance of fish as intermediaries between prawns and alligators.

the system has put all its eggs in one basket (efficiency). Should some catastrophe, like a virus affecting fish, devastate the fish population, all transfer from prawns to alligators in this rudimentary example would suffer in direct proportion.

If healthy populations of turtles and snakes had been present when the fish population was incapacitated, it is possible that the pathways they provide might have buffered the loss, as in Fig. 3.

Rather than total system collapse, $T_{..}$ drops modestly to $99.7 \text{ mg C m}^{-2} \text{y}^{-1}$, and the ascendency falls back only to $44.5 \text{ mg C-bits m}^{-2} \text{y}^{-1}$. The chief casualty of the disappearance of the fishes is the reserve, which falls by almost half to $68.2 \text{ mg C-bits m}^{-2} \text{y}^{-1}$. In other words, in the alternative scenario the system adapts in homeostatic fashion to buffer performance (A) by expending reserves (Φ) (Odum, 1953). The reserve in this case is not some palpable storage, like a cache of some material resource. Rather, it is a characteristic of the system structure that reflects the absence of effective performance.

The hypothetical changes in Figs. 1-3 were deliberately chosen as extremes to make a didactic point. In reality, one might expect some intermediate accommodation between A and Φ as the outcome. Identifying where such accommodation might lie is the crux of this essay, for it becomes obvious that the patterns we see in living systems are the outcomes of two antagonistic tendencies (Ulanowicz, 1986, 1997). On one hand are those processes that contribute to the increase in order and constraint in living systems. Paramount among them seems to be autocatalysis, which is capable of exerting selection pressure upon its constituents and of exerting a centripetal pull upon materials and energy, drawing resources into its orbit (Ulanowicz, 1986, 1997). In exactly the opposite direction is the slope into dissipation that is demanded by the second law of thermodynamics. At the focal level, these trends are antagonistic. At higher levels, however, the attributes become mutually obligate: A requisite for the increase in effective orderly performance (ascendency) is the existence of flexibility (reserve) within the system. Conversely, systems that are highly constrained and at peak performance (in the second law sense of the word) dissipate external gradients at ever higher gross rates (Schneider and Kay, 1994; Ulanowicz, 2009).

5. The survival of the most robust

While the dynamics of this dialectic interaction can be quite subtle and highly complex, one thing is boldly clear—systems with either vanishingly small ascendancy or insignificant reserves are destined to perish before long. A system lacking ascendancy has neither the extent of activity nor the internal organization needed to survive. By contrast, systems that are so tightly constrained and honed to a particular environment appear “brittle” in the sense of [Holling \(1986\)](#) or “senescent” in the sense of [Salthe \(1993\)](#) and are prone to collapse in the face of even minor novel disturbances. Systems that endure – that is, are sustainable – lie somewhere between these extremes. But, where?

Recognizing the importance of achieving an intermediate balance, [Wilhelm \(2003\)](#) suggested that the product $A\phi$ might serve as an appropriate metric for robustness. It becomes zero whenever either A or ϕ is zero and it takes on a maximum when $A = \phi$. This is an interesting suggestion, and strongly parallels the treatment of power production by [Odum and Pinkerton \(1955\)](#); but, like the latter analysis, it remains problematic. For example, there is no obvious reason why the optimal balance should fall precisely at $A = \phi$ (other than mathematical convenience). Secondly, with regard to this analysis, the simple product $A\phi$ does not accord with the formulaic nature of the IT used to construct each of its factors.

This second shortcoming is rather easy to rectify once we recall that (2) behaves in much the same manner as $A\phi$. It is zero at the extremes of $p_i = 0$ and $p_i = 1$ and achieves a single maximum in between. Accordingly, we define $a = A/C$ and notice that $1 > a > 0$. Here a is a relative measure of the organized power flowing within the system. In lieu of $\sim a$, or $(1 - a)$, we choose the Boltzmann formulation, $-k \log(a)$, so that the product of a and $\sim a$, or what we shall call the system’s “fitness for evolution”,

$$F = -ka \log(a), \tag{15}$$

becomes our measure of the system’s potential to evolve or self-organize. It is 0 for $a = 1$ and approaches the limit of 0 as $a \rightarrow 0$. One can normalize this function by choosing $k = e \log(e)$ (where “ e ” is the base of natural logarithms), such that $1 > F > 0$.

This does not solve our second problem, however, as F is still constrained to peak at $a = (1/e)$. There is no more reason to force the balance between A and ϕ to occur at $[A/(A + \phi)] = (1/e)$ than it was to mandate that it happen when $A = \phi$. Clearly, the location of the optimum could be the consequence of (as yet) unknown dynamical factors, rather than one of mathematical convenience. One way to permit the maximum to occur at an arbitrary value of a is to introduce an adjustable parameter, call it β , and to allow the available data to indicate the most likely value of β . Accordingly, we set $F = -ka^\beta \log(a^\beta)$. This function can be normalized by choosing $k = e/\log(e)$, so that $F_{\max} = 1$ at $a = e^{-1/\beta}$, where β can be any positive real number. Whence, our measure for evolutionary fitness becomes

$$F = -\left[\frac{e}{\log(e)}\right] a^\beta \log(a^\beta) \tag{16}$$

The function F varies between 0 and 1 and is entirely without dimensions. It describes the fraction of activity that is effective in creating a sustainable balance between A and ϕ . That is, the total activity (e.g., the GDP in economics, or $T_{..}$ here) will no longer be an accurate assessment of the robustness of the system. Our measure, $T_{..}$, must be discounted by the fraction $(1 - F)$. Equivalently, the robustness, R , of the system becomes

$$R = T_{..} \cdot F. \tag{17}$$

The focus of attention now turns to identifying the most propitious value for β . This is a very crucial point, because the value of β fixes the optimal value of a against which the status of any existing network will be reckoned. There is no apriori reason to assume that the value of β is universal. There might be one value of β most germane to ecosystem networks, another for economic communities, and still another for networks of genetic switching. Since the data most familiar to the authors of this work pertain to ecosystem networks of trophic exchanges, ecology seems a reasonable domain in which to begin our search.

Data on existing flow networks of ecosystems do not appear sufficient to determine a precise value for β . They do, however, indicate rather clearly those configurations of flows that are not sustainable. [Zorach and Ulanowicz \(2003\)](#), for example, compare how a collection of estimated flow structures differs from networks that have been created at random. For their demonstration, they plotted the networks, not on the axes A vs. ϕ , but rather on the transformed axes $c = 2^{\phi/2}$ and $n = 2^A$. As they explain in the course of their analysis, c measures the effective connectivity of the system in links per node, or how many nodes on (logarithmic) average enter or leave each compartment. The variable n gauges the effective number of trophic levels in the system, or how many transfers, on (logarithmic) average, a typical quantum of medium makes before leaving the system. Their results are displayed in [Fig. 4](#).

It is immediately obvious that the empirical networks all cluster within a rectangle that is bounded roughly in the vertical direction by $c = 1$ and $c \approx 3.01$ and horizontally by $n = 2$ and $n \approx 4.5$. It happens that three of the four sides of this “window of vitality” can be explained heuristically. The fact that $c \geq 1$ says simply that the networks being considered are all fully connected. Any value $c < 1$ would imply that the graph

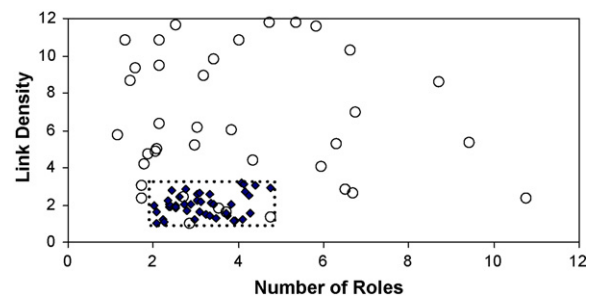


Fig. 4 – Combinations of link-density (c) plotted against number of effective roles (n) in a set of randomly assembled networks (circles) and empirically estimated ecosystem networks (dark squares).

is separated into non-communicating sub-networks. Similarly, $n > 2$ for all ecosystem networks, because it is in the very definition of an ecosystem that it encompass complementary processes, such as oxidation/reduction reactions or autotrophy/heterotrophy interactions (Fiscus, 2001).

The boundary delimiting maximal link-density, $c \approx 3.01$, is the result of applying the May–Wigner stability hypothesis in its information-theoretic homolog (Ulanowicz, 2002). The precise theoretical value of the boundary, as derived by Ulanowicz, is $c = e^{(3/e)}$. In essence, this says that systems can be either strongly connected across a few links or weakly connected across many links, but configurations of strong connections across many links and weak connections across a few links tend to break up or fall apart, respectively (May, 1972). The “magic number” 3 in association with maximal connectivity has been cited by Pimm (1982) and by Wagensberg et al. (1990) for ecosystems, and by Kauffman (1991) for genetic networks.

Only the fourth boundary remains largely unexplained. Pimm and Lawton (1977) commented on how one rarely encounters trophic pathways greater than 5 in nature. Efforts to relate this limit to thermodynamic efficiencies have (thus far) proved unsuccessful (Pimm and Lawton, 1977). The available data reveal no values for n close to 5, and so an upper limit of 4.5 has been chosen arbitrarily.

The emerging picture seems to be that sustainable ecosystems all plot within the window of vitality. It has yet to be investigated whether any sub-regions of the window might be preferred over others, and the scatter appears to be without statistically discernible pattern. It might be surmised, however, that systems plotting too close to any of the four boundaries could be approaching their limits of stability for one reason or another. Under such consideration, the most conservative assumption would be that those systems most distant from the boundaries are those most likely to remain sustainable. We therefore choose the geometric center of the window ($c = 1.25$ and $n = 3.25$) as the best possible configuration for sustainability under the information currently available. These values translate into $a = 0.4596$, from which we calculate a most propitious value of $\beta = 1.288$.

6. Vectors to sustainability

Systems can risk unsustainability in relation to this “optimum” on two accounts. When $a < 0.4596$, the system likely requires more coherence and cohesion. There may be insufficient or under-developed autocatalytic pathways that could impart additional robustness to the system. Conversely, when $a > 0.4596$, the system might be over-developed or too tightly constrained. Some autocatalytic pathways may have arrogated too many resources into their orbit, leaving the system with insufficient reserves to persist in the face of novel exigencies.

Should it survive further scrutiny, this threshold in a provides an extremely useful guide towards achieving sustainable communities. In fact, the measure of robustness, R , can even be employed to indicate which features of a given configuration deserve most remediation. Once again, the algebra of IT proves most convenient, because the functions C , A and Φ all happen to be homogeneous Euler functions of the

first order. This means that the derivatives with respect to their independent variables are relatively easy to calculate (Courant, 1936).

Starting from our definition of robustness (17), we seek to establish the direction in which this attribute responds to a unit change in any constituent flow. That is we wish to calculate $(\partial R / \partial T_{ij})$. Employing the chain rule of differentiation, we see that

$$\begin{aligned} \frac{\partial R}{\partial T_{ij}} &= F + T_{..} \frac{\partial F}{\partial T_{ij}} \\ \frac{\partial R}{\partial T_{ij}} &= F + T_{..} F' \frac{\partial a}{\partial T_{ij}} \\ \frac{\partial R}{\partial T_{ij}} &= F + \frac{T_{..} F'}{C} \left\{ \log \left[\frac{T_{ij} T_{..}}{T_i T_j} \right] + a \log \left[\frac{T_{ij}^2}{T_i T_j} \right] \right\} \end{aligned} \quad (18)$$

where F' is the derivative of F with respect to a , i.e.,

$$F' = -e\beta a^{\beta-1} \left[\frac{\log(a^\beta)}{\log(e)} + 1 \right] \quad (19)$$

In particular, when the system is at its optimum ($F = 1$ and $F' = 0$) we see from (18) that a unit increment in each and every flow in the system would contribute exactly one unit to system robustness. Once away from the optimum, however, contributions at the margin will depend on which side of the optimum the system lies, and where in the network any particular contribution is situated.

When $a < a_{opt}$, then F' will be positive, so that those flows that dominate the inputs to or output from any compartment will result in a positive sum within the braces, and the contribution of that transfer at the margin will be > 1 . For the relatively smaller flows, the negative second term in braces will dominate, and the contribution of those links at the margin will be < 1 . One observes both situations within the network of energy flows occurring in the Cone Spring ecosystem (Tilly, 1968), one of the most widely used examples of a simple ecosystem flow network (Fig. 5). The value of a for this network (0.418) is $< a_{opt}$ (≈ 0.460), so that the community can still grow and develop without jeopardizing its sustainability. Those flows with highest contributions at the margin (in parentheses) serve to vector the system towards configurations of greater sustainability. One notes in particular that increases in the values of the contributions at the margin along the pathway $\rightarrow 1 \rightarrow 2 \rightarrow 3$ are all favored to move a towards a_{opt} . Conversely, increases in flows that parallel mainstream flows (such as $3 \rightarrow 4$) contribute proportionately much less than towards system robustness, so that there is a disincentive against augmenting those flows.

To demonstrate that increases along the pathway $\rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow$ will indeed raise the value of the relative ascendancy, we artificially add, say 8000 kcal $m^{-2} y^{-1}$, to each of those links (Fig. 6). These additions mimic the process of eutrophication, whereby the addition of some additional resource inflates the primary production of the plants [1], which primarily die uneaten to become detritus [2], which in its turn is consumed and dissipated by bacteria [3]. The ensuing value of a is 0.529 ($> a_{opt}$).

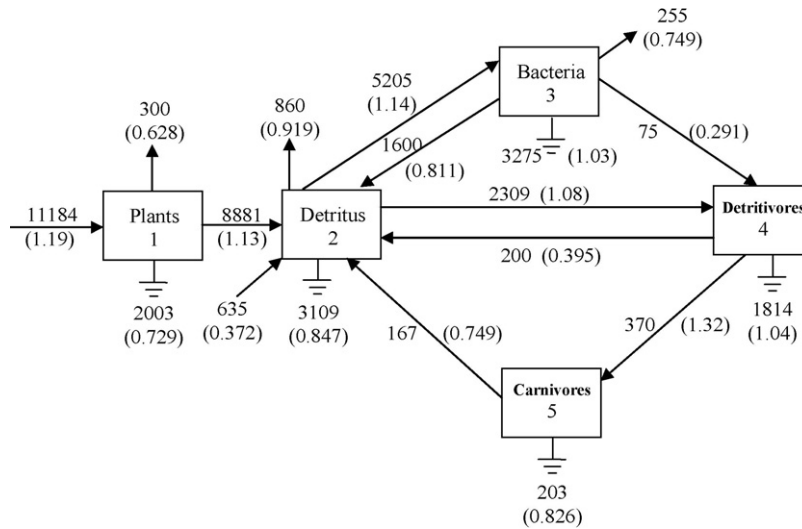


Fig. 5 – The trophic exchanges of energy (kcal m⁻² y⁻¹) in the Cone Spring ecosystem (Tilly, 1968). Arrows not originating from a box represent exogenous inputs. Arrows not terminating in a box portray exogenous outputs. Ground symbols represent dissipations. Numbers in parentheses accompanying each flow magnitude indicate the value of 1 kcal m⁻² y⁻¹ increment at the margin.

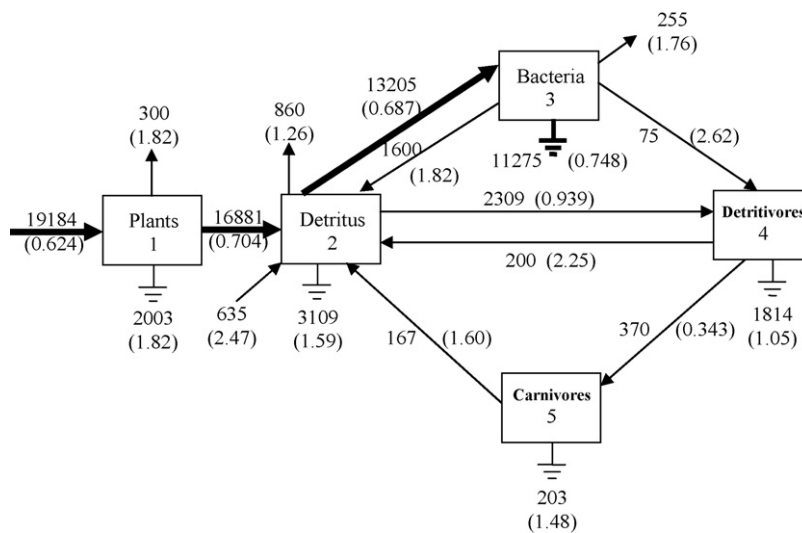


Fig. 6 – The Cone Spring network, as in Fig. 5, except that 8000 kcal m⁻² y⁻¹ has artificially been added over the route, → 1 → 2 → 3 →. Corresponding changes in the increments at the margin are shown in parentheses.

The absolute values of the contributions at the margin in Fig. 6 are pretty much the qualitative inverses of those in Fig. 5. One sees, for example, that the contributions at the margin along the “eutrophic axis” are now all less than one, whereas the corresponding values of the small, parallel transfers (such as 3 → 4) are now significantly greater than unity. One concludes, not surprisingly, that in a system with a surfeit of ascendancy over reserve, system survival is abetted by the addition of small, diverse parallel flows.

7. One-eyed ecology

One may conclude several things from the model developed here, but one in particular stands out: Many ecologists, in their

desire for a science that is derivative of physics, have unnecessarily blinded themselves to much of what transpires in nature. Physics does address matter and energy as they are present in ecosystems, but it tells us almost nothing about that which is lacking. Such latter considerations remain external to the core dynamics and can only be accounted as boundary constraints in ad-hoc terms, such as “rules” (Pattee, 1978) or particular “material laws” (Salthe, 1993), that usually do not conveniently mesh with the formal structure or dimensionality of the primary description.

As we have seen, the notions of both presence and absence are built into the formal structure of IT. Such architecture accounts for relationships like (8) and (14) wherein complementary terms of ‘what is’ and ‘what is not,’ share the same dimensions and almost the same structure. That is, one is

comparing apples with apples. Furthermore, the effects of lacunae no longer remain external to the statement of the dynamics; they become central to it. Most importantly, by incorporating apophasis into the core of the problem, one can avoid the pursuit of ill-fated directions, as will be discussed presently.

In fairness, it should be recalled that ecologists were not always indifferent to IT. In fact, soon after Claude Shannon (1948) had resuscitated Boltzmann's (1872) formulation, Robert MacArthur (1955) used the index to quantify the diversity of flows in an ecosystem and suggested that such diversity enhanced the stability of an ecosystem. Unfortunately, focus soon switched away from flows to the diversity of populations, and the leading aspiration among theoretical ecologists during the decade of the 1960s became how to demonstrate that biodiversity augments system stability (which in the interim had been formulated in terms of linear dynamical theory (Ulanowicz, 2002)). To the chagrin of most who were pursuing this intuition, physicist May (1972) upset the applecart by showing how, under the assumption of random connections, increased diversity is more likely to decrement, rather than bolster, system stability. May's counter-demonstration proved an embarrassment of the first magnitude to ecologists, while at the same time reinforcing their physics envy (Cohen, 1976). As a result, most ecologists came to eschew IT, and retreated to approaches that resembled what remained orthodox in physics.

We now discern a larger vision of the diversity/stability issue. It is not that May was wrong in his elegant demonstration. Rather (referring to Fig. 4), May showed how, as a system approached the top frame of the window of vitality, further diversification will indeed accelerate the system across the threshold and into oblivion. Furthermore, May's stability index was pivotal to establishing the location of this upper frame. The system is over-connected (see Allen and Starr, 1982), but we now see that transgressing May's threshold is only one of four different ways that a system can get into trouble. A system could also approach the bottom frame (under-connected), at which time, according to the model just presented, further diversification indeed will reduce the tendency of the system to become unsustainable. This latter argument, however, requires an appropriate measure of the Reserve that will keep the system from approaching that edge too closely. Of course, a system could also exit the window via the end members, but the exact location of the right-hand limit and the reasons for its existence remain poorly understood.

Most fortunately, May's demonstration did little to quench the widely held conviction that biodiversity does have value in maintaining sustainable ecosystems. Major worldwide efforts have been justifiably mounted to conserve ecological diversity. Yet, although some empirical evidence does exist to support such intuition (e.g., Van Voris et al., 1980; Tilman et al., 1996), few convincing theoretical models have emerged to defend such conservation. Few, that is, save for that of Rutledge et al. (1976), who focused undauntedly on the utility of IT and in particular upon the conditional entropy as an index of merit. Unfortunately, the ecologists of Rutledge's time were in no mood to countenance a return to the shambles of IT that lay in the wake of May's deconstruction.

Instead, ecology marched on with one eye kept deliberately shut.

8. Evolution as moderation

The model just discussed highlights the necessary role of reserve capacities in sustaining ecosystems. It contrasts with Darwinian theory, which unfortunately is espoused by many simply as the maximization of efficiency⁴ (e.g., the survival of the fittest). Such emphasis on efficiency is evident as well in a number of approaches to ecology, such as optimal foraging theory.⁵ Our results alert us to the need to exhibit caution when it comes to maximizing efficiencies. Systems can become too efficient for their own good. Autocatalytic configurations can expand to suck away resources from nonparticipating taxa, leaving them to wither and possibly to disappear. In particular, the human population and its attendant agro-ecology is fast displacing reserves of wild biota and possibly driving the global ecosystem beyond a_{opt} . In the face of such monist claims, our model illustrates the pressing need to conserve the diversity of biological processes (which, after all, was MacArthur's original concern).

Although possibly less enamored of physics, economics, too, seems in pursuit of monistic goals and all too willing to sacrifice everything for the betterment of market efficiency. Doubtless, maximizing efficiency is a good strategy to apply to inchoate economic systems that occupy the upper-left-hand corner of the economic window of vitality. Preoccupation with efficiency in today's global theatre could, however, propel into disaster a global economy that is fast approaching the lower-right-hand corner. Economists have long recognized, usually for ethical reasons, the need for "externalities" to brake the pell-mell rush towards increased market "efficiencies." In the model presented here, such brakes appear as necessary internal constraints on the system and point up the need to retain "subsidiarity." At times the brakes appear spontaneously within the system, as when societies adjust to problems by increasing their complexity (Tainter, 1988). Our model suggests that the establishment of complementary currencies can make impressive contributions at the margin towards sustaining the global economic system (Lietaer, 2001).

While this exercise may have illumined the dialectical nature of natural development in consistent and quantitative terms, it unfortunately leaves other issues clouded. The exact location of a_{opt} , for example, is bound to remain controversial. Two major uncertainties further obscure this problem. The first is the pressing need for a larger collection of ecological flow networks with which to explore whether any sub-region within the window of vitality may exist that is favored by the most sustainable systems. In this context it is worth remarking that almost all ecosystem networks plotted in Fig. 4 for which $a > a_{opt}$ constitute early renditions of

⁴ Efficiency is being used here as a synonym for "effectiveness".

⁵ Optimal foraging theory has been criticized elsewhere for focusing on the wrong null hypothesis. Allen et al. (2003) argue that optimal foraging is not the signal of interest, but is rather the null hypothesis, given evolved systems.

ecosystem flows that were cast in terms of only a few compartments. Recent and more fully resolved networks of flows tend to possess lower values of a (Robert Christian personal communication). Christian further notes that the values of a for the stable subset of a large collection of weighted, randomly assembled networks approached an asymptote very close to $1/e$. (See Fig. 7.11a on p. 84 of Morris et al., 2005.) It therefore bears further investigation as to whether β differs significantly from unity. If $\beta \approx 1$, then the a_{opt} chosen here is decidedly an over-estimate. This has practical implications, because an inflated a_{opt} would not sound warning bells soon enough.

In addition, there are the nagging ambiguities concerning the origins of the truncation of the right side of the window. Our choice of $n = 4.5$ as the limit was, to a degree, arbitrary. Unfortunately, little is yet known as to what poses a limit on the effective number of trophic levels in ecosystems. Obviously, thermodynamic losses play a key role, but they do not seem to be the whole story (Pimm and Lawton, 1977). Needed is a theoretical explanation of that limit akin to May's exegesis, which fixes the position of the top member. (Apropos this limit is the observation by Cousins (1990) that it is the top members of the trophic web that often control what transpires at lower levels.)

Of course, there remains the question of how well (if at all) this ecological analysis pertains to economic communities. It seems not unreasonable to assume that many of the same dynamics are at work in economics as structure ecosystems, and that, over "deep time," nature has solved many of the developmental problems for ecosystems that still beset human economies. It is perhaps useful that this model suggests that Adam Smith's "invisible hand" is not alone in sculpting the patterns of economies. Yet another hand would appear necessary to work opposite to Smith's. (No one claps with one hand.) Furthermore, it now appears unfortunate that economists by and large have abandoned the study of "input-output" networks as puerile. Perhaps this exercise will motivate a few to dust off some of the archived, large data sets on input-output networks of cash flows to scope out better the dimensions of the economic counterpart to ecology's window of vitality.

In all likelihood, the dynamics portrayed here pertain to other domains of inquiry as well. Kauffman (1991), for example, has written about limits on the stability of genetic control networks. These systems appear more mechanical than do ecosystems, and their rigidities may narrow the window of vitality sufficiently that it could be characterized more as an "edge". As with economics, a large collection of genetic control networks might help resolve better the domain of ontogenetic stability. Similarly, Vaz and Carvalho (1994) have portrayed the immune system as a network. Might not elucidating its window of vitality provide significant new insights into the health of organisms?

The possibilities of this wider perspective are truly exciting. Their potential helps to vanquish the pessimism implicit in ecologists' physics envy and to give new life to Hutchinson's optimistic view of "ecology as the study of the Universe" (Jolly, 2006). All that is necessary is that ecologists keep both eyes wide open.

Acknowledgements

The authors are grateful to the Citerra Foundation of Boulder, Colorado for partial support of the travel involved in creating this work and to Edgar and Christine Cahn for hosting our initial meeting in their home. They would also like to extend their gratitude to Timothy F.H. Allen for his enthusiastic support of the ideas contained herein. Stanley N. Salthe and Robert R. Christian also contributed several points to the narrative following their reviews of an early draft.

REFERENCES

- Abramson, N., 1963. *Information Theory and Coding*. McGraw-Hill, New York, p. 201.
- Allen, T.F.H., Starr, T.B., 1982. *Hierarchy*. University of Chicago Press, Chicago, p. 310.
- Allen, T.F.H., Giampietro, M., Little, A., 2003. Distinguishing ecological engineering from environmental engineering. *Ecol. Eng.* 20, 389–407.
- Atlan, H., 1974. On a formal definition of organization. *J. Theor. Biol.* 45, 295–304.
- Bateson, G., 1972. *Steps to an Ecology of Mind*. Ballantine Books, New York.
- Boltzmann, L., 1872. Weitere Studien über das Wärmeleichgewicht unter Gasmolekülen. *Wien. Ber.* 66, 275–370.
- Cohen, J.E., 1976. Irreproducible results and the breeding of pigs. *Bioscience* 26, 391–394.
- Conrad, M., 1983. *Adaptability: The Significance of Variability from Molecule to Ecosystem*. Plenum Press, New York, p. 383.
- Courant, R., 1936. P108 in *Differential and Integral Calculus*, vol. 2. Interscience Publishers, New York, p. 681.
- Cousins, S.H., 1990. Countable ecosystems deriving from a new foodweb entity. *Oikos* 57, 270–275.
- Fiscus, D.A., 2001. The ecosystemic life hypothesis. I: introduction and definitions. *Bull. Ecol. Soc. Am.* 82 (4), 248–250.
- Holling, C.S., 1986. The resilience of terrestrial ecosystems: local surprise and global change. In: Clark, W.C., Munn, R.E. (Eds.), *Sustainable Development of the Biosphere*. Cambridge University Press, Cambridge, UK, pp. 292–317.
- Jolly, A., 2006. A Global Vision. *Nature* 443, 148.
- Kauffman, S.A., 1991. Antichaos and adaptation. *Sci. Am.* 265, 78–84.
- Lietaer, B., 2001. *The Future of Money*. Century, London, p. 371.
- MacArthur, R.H., 1955. Fluctuations of animal populations and a measure of community stability. *Ecology* 36, 533–536.
- May, R.M., 1972. Will a large complex system be stable? *Nature* 238, 413–414.
- Morris, J.T., Christian, R.R., Ulanowicz, R.E., 2005. Analysis of size and complexity of randomly constructed food webs by information theoretic metrics. In: Belgrano, A., Scharler, U.M., Dunne, J., Ulanowicz, R.E. (Eds.), *Aquatic Food Webs*. Oxford University Press, Oxford, p. 262, pp. 73–85.
- Odum, E.P., 1953. *Fundamentals of Ecology*. Saunders, Philadelphia, p. 384.
- Odum, H.T., Pinkerton, R.C., 1955. Time's speed regulator: the optimum efficiency for maximum power output in physical and biological systems. *Am. Sci.* 43, 331–343.
- Pattee, H.H., 1978. The complementarity principle in biological and social structures. *J. Soc. Biol. Struct.* 1, 191–200.
- Pimm, S.L., 1982. *Foodwebs*. Chapman and Hall, London, p. 219.
- Pimm, S.L., Lawton, J.H., 1977. Number of trophic levels in ecological communities. *Nature* 268, 329–331.

- Rutledge, R.W., Basorre, B.L., Mulholland, R.J., 1976. Ecological stability: an information theory viewpoint. *J. Theor. Biol.* 57, 355–371.
- Salthe, S.N., 1993. *Development and Evolution: Complexity and Change in Biology*. MIT Press, Cambridge, MA, p. 257.
- Schneider, E.D., Kay, J.J., 1994. Life as a manifestation of the second law of thermodynamics. *Math. Comput. Model.* 19, 25–48.
- Shannon, C.E., 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423.
- Tainter, J.A., 1988. *The Collapse of Complex Societies*. Cambridge University Press, NY, p. 250.
- Tilly, L.J., 1968. The structure and dynamics of Cone Spring. *Ecol. Monogr.* 38, 169–197.
- Tilman, D., Wedin, D., Knops, J., 1996. Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature* 379, 718–720.
- Tribus, M., McIrvine, E.C., 1971. Energy and information. *Sci. Am.* 225 (3), 179–188.
- Ulanowicz, R.E., 1980. An hypothesis on the development of natural communities. *J. Theor. Biol.* 85, 223–245.
- Ulanowicz, R.E., 1986. *Growth & Development: Ecosystems Phenomenology*. Springer-Verlag, NY, p. 203.
- Ulanowicz, R.E., 1997. *Ecology the Ascendent Perspective*. Columbia University Press, NY, p. 201.
- Ulanowicz, R.E., 2002. Information theory in ecology. *Comput. Chem.* 25, 393–399.
- Ulanowicz, R.E., 2009. *A third window: natural life beyond Newton and Darwin*. Templeton Foundation Press, West Conshohocken, PA.
- Ulanowicz, R.E., Norden, J.S., 1990. Symmetrical overhead in flow networks. *Int. J. Syst. Sci.* 1, 429–437.
- Ulanowicz, R.E., Bondavalli, C., Egnotovitch, M.S., 1996. *Network Analysis of Trophic Dynamics in South Florida Ecosystems, FY 96: The Cypress Wetland Ecosystem*. Annual Report to the United States Geological Service Biological Resources Division University of Miami Coral Gables, FL 33124.
- Van Voris, P., O'Neill, R.V., Emanuel, W.R., Shugart, H.H., 1980. Functional complexity and ecological stability. *Ecology* 61, 1352–1360.
- Vaz, N.M., Carvalho, C.R., 1994. Assimilation, tolerance and the end of innocence. *Ciênc. cult. (São Paulo)* 46 (5/6), 351–357.
- Wagensberg, J., Garcia, A., Sole, R.V., 1990. Connectivity and information transfer in flow networks: two magic numbers in ecology? *Bull. Math. Biol.* 52, 733–740.
- Wilhelm, T., 2003. An elementary dynamic model for non-binary food-webs. *Ecol. Model.* 168, 145–152.
- Zorach, A.C., Ulanowicz, R.E., 2003. Quantifying the complexity of flow networks: how many roles are there? *Complexity* 8 (3), 68–76.