

Version 4.1

Roman roads: The hierarchical endosymbiosis of cognitive modules

Rodrick Wallace, Ph.D.
Division of Epidemiology
The New York State Psychiatric Institute*

October 22, 2009

Abstract

Serial endosymbiosis theory provides a unifying paradigm for examining the interaction of cognitive modules at vastly different scales of biological, social, and cultural organization. A trivial but not unimportant model associates a dual information source with a broad class of cognitive processes, and punctuated phenomena akin to phase transitions in physical systems, and associated coevolutionary processes, emerge as consequences of the homology between information source uncertainty and free energy density. The dynamics, including patterns of punctuation similar to ecosystem resilience transitions, are largely dominated by the availability of ‘Roman roads’ constituting channels for the transmission of information between modules.

Key Words: cognition, endosymbiosis, free energy, information theory, phase transition, rate distortion

1 Introduction

The rise and management of something like the Roman empire, decision-making in smaller social groups like hunting teams and families, the interaction of unconscious cognitive neural and other modules to produce consciousness in higher animals, the crosstalk between cognitive submodules in modern computing machines and networks, and the serial endosymbiosis leading to the eukaryotic transition that preceded all these by a billion years, seem absurdly disconnected across scales of time, energy, and levels of organization. Here we will argue that serial endosymbiosis theory (SET) provides a fundamental model for the hierarchical ‘fusion’ of cognitive processes, once biological cognition emerged from deep evolutionary time. Indeed, cognition, in the sense we will define it, seems a necessary condition for life, and appears a sufficient marker of its existence.

Witzany (2006) describes SET, often associated with the work of Lynn Margulis (e.g., 1966, 1996, 2004), as explaining the origin of nucleated eukaryotic cells by a merging of archaeobacterial and eubacterial cells in anaerobic symbiosis, followed by acquisition of mitochondria or plastids. In contrast to former evolutionary theories that consider ramification as a driving force for evolution, Margulis uses the language of mechanistic biology: ‘merging’, ‘fusion’, ‘incorporation’, and ‘amalgamation’. By contrast, Witzany’s biosemiotics approach argues that all cell-cell interactions are communication processes. Witzany nonetheless claims SET plausibly demonstrates that chance mutations were not responsible for initiating the evolutionary processes from prokaryotes to eukaryotes. Rather, merging processes by various prokaryotes – via endosymbioses and the fusion of different organisms – ultimately led to the eukaryotic cell.

Wallace and Wallace (2008), rather than taking Witzany’s biosemiotics perspective on biological communication, use the asymptotic limit theorems of information theory to derive a class of necessary conditions statistical models of the interaction between embedding ecosystem, cognitive gene expression, and Darwinian genetic inheritance. These interactions, via an analog to the ‘no free lunch’ theorem of optimization theory, lead to a ‘mesoscale resonance’ by which punctuated ecosystem resilience transitions at appropriate scale, in the sense of Holling (1973, 1992), can entrain both more rapid processes of gene expression, and slower Darwinian genetic selection, to produce punctuated evolutionary transitions on geologic time scales.

R. Wallace and R.G. Wallace (2008) have extended that work to prebiotic circumstances, examining how vesicles engaging in an inherently coevolutionary crosstalk – exchanging information – can, given sufficient time and energy, overcome the constraints of Eigen’s paradox to engage in low error rate replication.

The basis of that work was a reformulation of Eigen’s error catastrophe for replication in terms of an average distortion measure, and formal application of the information theory Rate Distortion Theorem, understanding that the rate distortion function has a homology with free energy density via the relation between channel capacity and information source

*Affiliation for identification only. Address correspondence to Rodrick Wallace, 549 W. 123 St., Apt. 16F, New York, NY, 10027 USA, email wallace@pi.cpmc.columbia.edu.

uncertainty. The error catastrophe then emerges as an analog to a phase transition in a physical system.

Wallace and Wallace (2009) extend the analysis to examine an ‘epigenetic catalysis’ under a cognitive paradigm for gene expression by which embedding information sources can direct the ontology of enveloped structures through mutual information interaction. Here we will extend that argument to more general circumstances.

Wallace (2009) examines the implications of metabolic energy density limitations for the program of Serial Endosymbiosis Theory, as applied to the eukaryotic transition. The essence of the argument is that mutualism, obligate mutualism, symbiosis, and ‘fusion’ represent a progression of increasing fidelity in real-time communication between (at least two) vesicles. That is, a progression of decreasing average distortion in chemical communication between interacting parts of a biological system. This, he shows, requires a progressive increase in available metabolic energy. Thus the transition from anaerobic to aerobic metabolism – to an oxygen-based ecology – may have been necessary to provide sufficient metabolic energy density to permit the ‘fusion’ of biological vesicles operating in real time. That is, low error rate real-time communication between biological structures, as opposed to slower processes of replication that can store energy until needed, requires a high energy metabolism unavailable to anaerobic ecosystems. Thus the (relatively) sudden transition to an aerobic ecosystem entrained the evolutionary transition resulting in eukaryotic cells. Early eukaryotes were apparently not anaerobes (e.g., Embley, 2006; Cavalier-Smith, 2006), and Wallace (2009) shows that a very simple model suggests the transition required the preexistence of a relatively high energy density metabolic cycle.

An essential, and recurring, point is the question of just what scale measures ‘real time’ for a particular cellular, neural, social, institutional, machine, or other process.

‘Real time’ metabolic processes of eukaryotic cells can occur on characteristic time scales of a few milliseconds to a few minutes. The governance of the systems of distributed cognition that constitute human organizations of any size (e.g., Wallace and Fullilove, 2008) typically takes place over periods of hours to years. As with the metabolism of cells, however, an essential limit is the maximum rate at which information can be sent across the organization with a particular level of distortion. Too rapid transmission guarantees high distortion in the messages sent, while, paradoxically, too slow transmission guarantees institutional failure. Hence the reference to Roman roads in the title of this paper.

Before proceeding to a detailed analysis, we present some simple models and facts from information theory that will be the basis of the argument.

2 Cognition as an information source

According to Atlan and Cohen (1998), the essence of cognition is comparison of a perceived external signal with an internal,

learned picture of the world, and then, upon that comparison, the choice of one response from a much larger repertoire of possible responses. Such reduction in uncertainty inherently carries information, and, following, e.g., Wallace (2005), it is possible to make a very general model of this process as an information source in a manner that one observer has called ‘trivial but not unimportant’.

A pattern of ‘sensory’ input, say an ordered sequence y_0, y_1, \dots , is mixed in a systematic (but unspecified) algorithmic manner with internal ‘ongoing’ activity, the sequence w_0, w_1, \dots , to create a path of composite signals $x = a_0, a_1, \dots, a_n, \dots$, where $a_j = f(y_j, w_j)$ for some function f . This path is then fed into a highly nonlinear, but otherwise similarly unspecified, decision oscillator which generates an output $h(x)$ that is an element of one of two disjoint sets B_0 and B_1 .

Take $B_0 \equiv \{b_0, \dots, b_k\}$ and $B_1 \equiv \{b_{k+1}, \dots, b_m\}$.

Now permit a graded response, supposing that if $h(x) \in B_0$ the pattern is not recognized, and if $h(x) \in B_1$ the pattern is recognized and some action $b_j, k+1 \leq j \leq m$ takes place.

The focus is on those composite paths x that trigger pattern recognition-and-response. That is, given a fixed initial state a_0 , such that $h(a_0) \in B_0$, examine all possible subsequent paths x beginning with a_0 and leading to the event $h(x) \in B_1$. Thus $h(a_0, \dots, a_j) \in B_0$ for all $0 \leq j < m$, but $h(a_0, \dots, a_m) \in B_1$.

For each positive integer n , let $N(n)$ be the number of grammatical and syntactic high probability paths of length n beginning with some particular a_0 having $h(a_0) \in B_0$ and leading to the condition $h(x) \in B_1$. We shall call such paths meaningful and take $N(n)$ to be considerably less than the number of all possible paths of length n – pattern recognition-and-response is comparatively rare.

The particular assumption is that the longitudinal finite limit

$$H \equiv \lim_{n \rightarrow \infty} \log[N(n)]/n$$

both exists and is independent of the path x , and – not surprisingly – call such a cognitive process *ergodic*.

It thus becomes possible to define an ergodic information source \mathbf{X} associated with stochastic variates X_j having joint and conditional probabilities $P(a_0, \dots, a_n)$ and $P(a_n|a_0, \dots, a_{n-1})$ such that appropriate joint and conditional Shannon uncertainties may be defined which satisfy the standard relations defining information sources, as in equation (7) below (e.g., Ash, 1990).

This information source is taken as *dual* to the ergodic cognitive process.

Dividing the full set of possible responses into the sets B_0 and B_1 may itself require higher order cognitive decisions by another module or modules, suggesting the necessity of choice within a more or less broad set of possible quasi-languages. This would directly reflect the need to shift gears according to the different challenges faced by the organism, machine, or social group.

‘Meaningful’ paths – creating an inherent grammar and syntax – have been defined entirely in terms of system re-

sponse, as Atlan and Cohen propose. This formalism can easily be applied to the stochastic neuron in a neural network (Wallace, 2005).

A formal equivalence class algebra can now be constructed for a cognitive process characterized by a dual information source by choosing different origin points a_0 , in the sense above, and defining equivalence of two states by the existence of a high-probability meaningful path connecting them with the same origin. Disjoint partition by equivalence class, analogous to orbit equivalence classes for dynamical systems, defines the vertices of a network of cognitive dual languages. Each vertex then represents a different information source dual to a cognitive process. This is not a direct representation as in a neural network, or of some circuit in silicon. It is, rather, an abstract set of ‘languages’ dual to the cognitive processes instantiated by biological structures, machines, social process, or their hybrids. Our particular interest, however, is in an interacting network of cognitive processes.

This structure generates a groupoid (Wallace, 2008): states a_j, a_k in a set A are related by the groupoid morphism if and only if there exists a high-probability grammatical path connecting them to the same base point, and tuning across the various possible ways in which that can happen – the different cognitive languages – parameterizes the set of equivalence relations and creates the groupoid representing a network of dual information sources, a groupoid formed by the disjoint union of the underlying transitive groupoids that emerge when the base point changes. This construction leads to topological descriptions of phase transition dynamics in terms of holonomy groupoids and the like that we will not pursue here.

3 The Rate Distortion Theorem

Real time problems are inherently rate distortion problems. The interaction between two communicating biological vesicles or other cognitive structures can be restated in communication theory terms. Suppose a sequence of signals is generated by a biological or institutional information source – a vesicle or a subgroup – Y having output $y^n = y_1, y_2, \dots$. This is ‘digitized’ in terms of the observed behavior of the system – vesicle or cognitive institutional module – with which it communicates, say a sequence of observed behaviors $b^n = b_1, b_2, \dots$. The b_i happen in real time. Assume each b^n is then deterministically retranslated back into a reproduction of the original biological signal,

$$b^n \rightarrow \hat{y}^n = \hat{y}_1, \hat{y}_2, \dots$$

Define a distortion measure $d(y, \hat{y})$ which compares the original to the retranslated path. Many distortion measures are possible. The Hamming distortion is defined simply as

$$d(y, \hat{y}) = 1, y \neq \hat{y}$$

$$d(y, \hat{y}) = 0, y = \hat{y}$$

For continuous variates the squared error distortion is just (3)

$$d(y, \hat{y}) = (y - \hat{y})^2.$$

There are many such possibilities. The distortion between paths y^n and \hat{y}^n is defined as

$$d(y^n, \hat{y}^n) \equiv \frac{1}{n} \sum_{j=1}^n d(y_j, \hat{y}_j).$$

A remarkable fact of the Rate Distortion Theorem is that *the basic result is independent of the exact distortion measure chosen* (Cover and Thomas, 1991; Dembo and Zeitouni, 1998).

Suppose that with each path y^n and b^n -path retranslation into the y -language, denoted \hat{y}^n , there are associated individual, joint, and conditional probability distributions

$$p(y^n), p(\hat{y}^n), p(y^n, \hat{y}^n), p(y^n | \hat{y}^n).$$

The average distortion is defined as

$$D \equiv \sum_{y^n} p(y^n) d(y^n, \hat{y}^n).$$

(1)

It is possible, using the distributions given above, to define the information transmitted from the Y to the \hat{Y} process using the Shannon source uncertainty of the strings:

$$I(Y, \hat{Y}) \equiv H(Y) - H(Y | \hat{Y}) = H(Y) + H(\hat{Y}) - H(Y, \hat{Y}),$$

(2)

where $H(\dots, \dots)$ is the joint and $H(\dots | \dots)$ the conditional uncertainty (Cover and Thomas, 1991; Ash, 1990).

If there is no uncertainty in Y given the retranslation \hat{Y} , then no information is lost, and the systems are in perfect synchrony.

In general, of course, this will not be true.

The *rate distortion function* $R(D)$ for a source Y with a distortion measure $d(y, \hat{y})$ is defined as

$$R(D) = \min_{p(y, \hat{y}); \sum_{(y, \hat{y})} p(y) p(\hat{y}) d(y, \hat{y}) \leq D} I(Y, \hat{Y}).$$

(3)

where $P(X)$ is chosen so as to maximize the rate of information transmission along a channel Y .

Finally, recall the analogous definition of the rate distortion function from equation (3), again an extremum over a probability distribution.

4 Information and free energy

Our own work (e.g., Wallace and Wallace, 2008, 2009) focuses on the homology between information source uncertainty and free energy density. More formally, if $N(n)$ is the number of high probability ‘meaningful’ – that is, grammatical and syntactical – sequences of length n emitted by an information source X , then, according to the Shannon-McMillan Theorem, the zero-error limit of the Rate Distortion Theorem (Ash, 1990; Cover and Thomas, 1991; Khinchin, 1957),

$$\begin{aligned} H[X] &= \lim_{n \rightarrow \infty} \frac{\log[N(n)]}{n} \\ &= \lim_{n \rightarrow \infty} H(X_n | X_0, \dots, X_{n-1}) \\ &= \lim_{n \rightarrow \infty} \frac{H(X_0, \dots, X_n)}{n+1}, \end{aligned} \tag{7}$$

where, again, $H(\dots|\dots)$ is the conditional and $H(\dots, \dots)$ is the joint uncertainty.

In the limit of large n , $H[X]$ becomes homologous to the free energy density of a physical system at the thermodynamic limit of infinite volume. More explicitly, the free energy density of a physical system having volume V and partition function $Z(\beta)$ derived from the system’s Hamiltonian – the energy function – at inverse temperature β is (e.g., Landau and Lifshitz 2007)

$$\begin{aligned} F[K] &= \lim_{V \rightarrow \infty} -\frac{1}{\beta} \frac{\log[Z(\beta, V)]}{V} \equiv \\ &= \lim_{V \rightarrow \infty} \frac{\log[\hat{Z}(\beta, V)]}{V}, \end{aligned}$$

where $\hat{Z} = Z^{-1/\beta}$. The latter expression is formally similar to the first part of equation (7), a circumstance having deep implications: Feynman (2000) describes in great detail how information and free energy have an inherent duality. Feynman, in fact, defines information precisely as the free energy needed to erase a message. The argument is surprisingly direct (e.g., Bennett, 1988), and for very simple systems it is easy to design a small (idealized) machine that turns the information within a message directly into usable work – free energy. Thus the persistent intellectual myth that information is somehow a mystical ‘negative entropy’ is quite

The minimization is over all conditional distributions $p(y|\hat{y})$ for which the joint distribution $p(y, \hat{y}) = p(y)p(y|\hat{y})$ satisfies the average distortion constraint (i.e., average distortion $\leq D$).

The *Rate Distortion Theorem* states that $R(D)$ is the minimum necessary rate of information transmission which ensures the communication between the modules does not exceed average distortion D . Thus $R(D)$ defines a minimum necessary channel capacity. Cover and Thomas (1991) or Dembo and Zeitouni (1998) provide details. The rate distortion function has been calculated for a number of systems.

There is an absolutely central fact characterizing the rate distortion function: Cover and Thomas (1991, Lemma 13.4.1) show that $R(D)$ is necessarily a decreasing convex function of D for any reasonable definition of distortion.

That is, $R(D)$ is always a reverse J-shaped curve. This will prove crucial for the overall argument. Indeed, convexity is an exceedingly powerful mathematical condition, and permits deep inference (e.g., Rockafellar, 1970). Ellis (1985, Ch. VI) applies convexity theory to conventional statistical mechanics.

For a Gaussian channel having noise with zero mean and variance σ^2 (Cover and Thomas, 1991),

$$R(D) = 1/2 \log[\sigma^2/D], 0 \leq D \leq \sigma^2$$

$$R(D) = 0, D > \sigma^2.$$

(4)

Recall, now, the relation between information source uncertainty and channel capacity (e.g., Ash, 1990):

$$H[\mathbf{X}] \leq C,$$

(5)

where H is the uncertainty of the source X and C the channel capacity, defined according to the relation (Ash, 1990)

$$C \equiv \max_{P(X)} I(X|Y),$$

(6)

wrong. Information is a form of free energy and the construction and transmission of information within living things consumes metabolic free energy, as it does within organizational structures, with inevitable losses via the second law of thermodynamics. If there are limits on available free energy density there will necessarily be limits on the ability to process information in real time across metabolic and social, domains, and in particular across the complex cultural structures of distributed cognition that mark human life in community.

Conversely, information source uncertainty has an important heuristic interpretation (Ash, 1990) in that we may regard a portion of text in a particular language as being produced by an information source. The probabilities $P[X_n = a_n | X_0 = a_0, \dots, X_{n-1} = a_{n-1}]$ may be estimated from the available data about the language; in this way we can estimate the uncertainty associated with the language. A large uncertainty means, by the Shannon-McMillan Theorem, a large number of ‘meaningful’ sequences. Thus given two languages with uncertainties H_1 and H_2 respectively, if $H_1 > H_2$, then in the absence of noise it is easier to communicate in the first language; more can be said in the same amount of time. On the other hand, it will be easier to reconstruct a scrambled portion of text in the second language, since fewer of the possible sequences of length n are meaningful.

In sum, if a biological or institutional system of cognitive modules characterized by H_1 has a richer and more complicated internal communication structure than one characterized by H_2 , then necessarily $H_1 > H_2$ and system 1 will consume metabolic or other free energy at a greater rate than system 2.

5 Rate Distortion Dynamics

$R(D)$ defines the minimum channel capacity necessary for the system to have average distortion less than or equal D , placing a limit on information source uncertainty. Thus, we suggest distortion measures can drive information system dynamics. That is, the rate distortion function also has a homological relation to free energy density, similar to the relation between free energy density and information source uncertainty.

We are led to propose, as a heuristic, that the dynamics of vesicles or cognitive institutional submodules interacting in real time will be constrained by the system as described in terms of a parameterized rate distortion function. To do this, take R as parameterized, not only by the distortion D , but by some vector of variates $\mathbf{D} = (D_1, \dots, D_k)$, for which the first component is the average distortion. The assumed dynamics are, as in Wallace and Wallace, (2008), then driven by gradients in the rate distortion disorder defined as

$$S_R \equiv R(\mathbf{D}) - \sum_{i=1}^k D_i \partial R / \partial D_i. \quad (8)$$

This leads to the deterministic and stochastic systems of equations analogous to the Onsager relations of nonequilibrium thermodynamics:

$$dD_j/dt = \sum_i L_{j,i} \partial S_R / \partial D_i \quad (9)$$

and

$$dD_t^j = L^j(D_1, \dots, D_k, t) dt + \sum_i \sigma^{j,i}(D_1, \dots, D_k, t) dB_t^i, \quad (10)$$

where the dB_t^i represent added, often highly structured, stochastic ‘noise’ whose properties are characterized by the quadratic variation (e.g., Protter, 1995).

A central focus of this paper is to generalize these equations in the face of richer structure, for example the existence of characteristic time constants within nested processes, and the influence of an embedding source of free energy.

For a simple Gaussian channel with noise having zero mean and variance $\sigma^2 = 1$,

$$S_R(D) = R(D) - D dR(D)/dD = 1/2 \log(1/D) + 1/2. \quad (11)$$

The simplest possible Onsager relation becomes

$$dD/dt \propto -dS_R/dD = \frac{1}{2D}, \quad (12)$$

where $-dS_R/dD$ represents the force of an entropic wind, a kind of internal dissipation inevitably driving the real-time, system of interacting (cognitive) information sources toward greater distortion.

This has the solution

$$D \propto \sqrt{t}. \quad (13)$$

A central observation is that *similar results must necessarily accrue to any of the reverse-J-shaped relations that inevitably characterize $R(D)$* , since the rate distortion function is necessarily a convex decreasing function of the average distortion D , whatever distortion measure is chosen. Again, see Cover and Thomas (1991, Lemma 13.4.1) for details.

The explicit implication is that a system of cognitive modules interacting in real time will inevitably be subject to a relentless entropic force, requiring a constant free energy expenditure for maintenance of some fixed average distortion in the communication between them: The distortion in the communication between two interacting modules will, without free energy input, have time dependence

$$D = f(t), \quad (14)$$

with $f(t)$ monotonic increasing in t .

This necessarily leads to the punctuated failure of the system.

6 Error Catastrophe and the dual energy limit

The mean field model of Wallace and Wallace (2008) explores the spectrum of phase transition dynamics possible to information processes. Actual architectures will involve crosstalking hierarchies and/or networks of subsystems. The phase transition discussions of Wallace and Wallace (2008) can, accordingly, be rephrased in terms of a critical distortion D_C exactly analogous to the Eigen's error catastrophe (Eigen, 1971, 1996). That is, the lowest possible energy state for a highly parallel, real time, structure of interacting cognitive modules is just above the critical distortion. Further, absent a continual metabolic or other free energy input, the dynamics of interaction can, from the homology between $R(D)$ and free energy density, drive the system to that limit.

Beyond D_C the interaction simply collapses in a phase transition. The modules cease to interact, and the 'cell' dies in a parallel to Eigen's famous error catastrophe (R. Wallace and R. G. Wallace, 2008).

Figure 1 shows $R(D)$ and $S_R(D)$ vs. D for a Gaussian channel. The limit D_C represents the critical average distortion

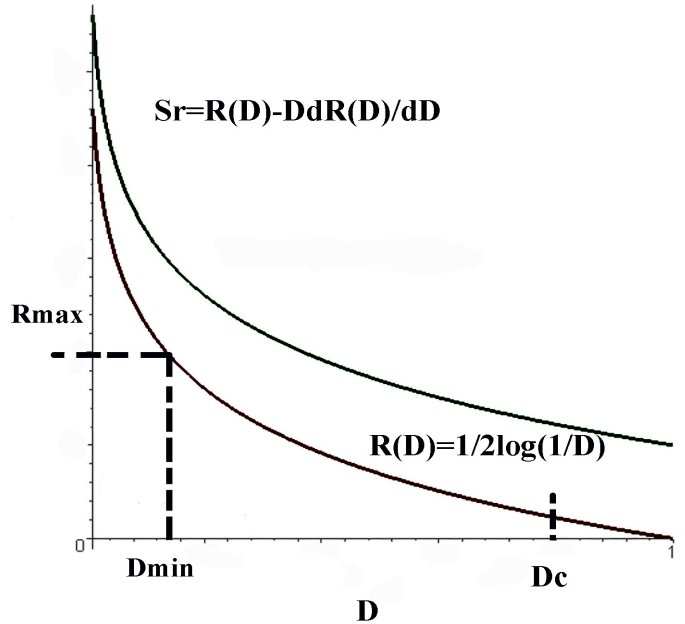


Figure 1: Rate distortion dynamics for a real-time Gaussian channel connecting two cognitive modules. Critical distortion is D_C and the maximum possible rate distortion value is R_{max} . Beyond D_C the two modules break up. D_{min} is the minimum distortion possible under the energy density constraint R_{max} . An extended model shows how the embedding available free energy density acts as a temperature analog to determine a probability distribution for D .

at which a real-time system of two interacting vesicles suffers punctuated failure. This analog to the error catastrophe in viral replication is now applied to a system of two biological vesicles or cognitive institutional submodules interacting in real time.

The results of Wallace and Wallace (2008) suggest that such systems will further be subject to large deviations jets that can be systematically described using variations of the Onsager/Machlup formalism, a matter we do not pursue here.

We now enter a new realm, and propose a second critical limit, determined by enthalpic constraints imposed by availability of metabolic or other sources of free energy.

Information source uncertainty is homologous to free energy density, and the rate distortion function, $R(D)$, is a minimum channel capacity placing a lower limit on information source uncertainty for a given average distortion D , and hence on the minimum free energy density necessary to attain that distortion. Thus, while the maximum average distortion in the real time communication between generalized vesicles is constrained by an analog to Eigen's error catastrophe, the minimum possible is constrained by the available free energy density.

Thus, in figure 1, most simply, in addition to the error catastrophe defined by D_C , there is an energy-determined

limit to the rate distortion function, R_{max} . This, in turn, defines a minimum possible average distortion for interacting generalized vesicles, D_{min} , and is a dual constraint to the error catastrophe.

7 Reconsidering phase transition

It is possible to more precisely characterize a probability distribution of distortion values associated with a given available free energy density by making a natural extension of standard theory. This leads to deep topological considerations via a generalization of Landau's famous theory of spontaneous symmetry breaking for phase transitions.

Recall that the free energy of a physical system at a normalized inverse temperature $\beta = 1/kT$ is defined as

$$F(\beta) = -\frac{1}{\beta} \log[Z(\beta)], \quad (15)$$

where $Z(\beta)$ the partition function defined by the system Hamiltonian. More precisely, if the possible energy states of the system are a set $E_i, i = 1, 2, \dots$ then, at normalized inverse temperature β , the probability of a state E_i is determined by the relation

$$P[E_i] = \frac{\exp[-E_i\beta]}{\sum_j \exp[-E_j\beta]}. \quad (16)$$

The partition function is simply the normalizing factor (Landau and Lifshitz, 2007)

$$Z[\beta] \equiv \sum_j \exp[-E_j\beta]. \quad (17)$$

Applying this formalism, it is possible to extend the rate distortion model by describing a probability distribution for D across an ensemble of possible rate distortion functions in terms of available metabolic energy density, using the obvious generalization of equations (16) and (17). That is, *available free energy density is viewed as an embedding temperature-analog.* (20)

The key is to take the $R(D)$ as representing energy as a function of the average distortion. Assume there to be a fixed and appropriately normalized inverse available free energy density, β , so that the probability density function of an average distortion D , given a fixed β , is then

$$P[D, \beta] = \frac{\exp[-R(D)\beta]}{\int_{D_{min}}^{D_{max}} \exp[-R(D)\beta] dD}. \quad (18)$$

The rate distortion function for a Gaussian channel, having $D_{min} = 0$ and $D_{max} = \sigma^2$, gives the normalizing factor simply as $2\sigma^2/(2 + \beta)$ and probability distribution

$$P[D, \beta, \sigma] = (1/2)\sigma^{-(\beta+2)}(1/D)^{-\beta/2}(\beta + 2). \quad (19)$$

Letting $\sigma = 1, \beta = 2$, the density values for $D = 0.1$ and $D = 0.01$ are 0.2 and 0.02, respectively, a ten-fold decrease, at that value of normalized available free energy density $1/\beta = .5$. Increasing available energy density rapidly lowers this ratio: for $1/\beta = 2$ the density values are, respectively, 0.70 and 0.40, which differ by less than a factor of two.

More generally, the mean value of D for the distribution above is easily calculated to be

$$\langle D \rangle = \frac{\sigma^2(\beta + 2)}{\beta + 4}.$$

As $\beta \rightarrow \infty, 0$ then $\langle D \rangle \rightarrow \sigma^2, \sigma^2/2$. The corresponding limits of the standard deviation of D are, respectively, $0, \sigma^2/2\sqrt{3}$.

Thus increasing available embedding free energy density – lowering β – in this model, rapidly raises the possibility of low distortion communication between linked cognitive sub-modules.

We define the *rate distortion partition function* as just the normalizing factor in equation (18):

$$Z_R[\beta] \equiv \int_{D_{min}}^{D_{max}} \exp[-R(D)\beta] dD, \quad (20)$$

again taking β as proportional to the inverse metabolic energy density.

We now define a new free energy-analog, the *rate distortion free-energy*, as

$$F_R[\beta] \equiv -\frac{1}{\beta} \log[Z_R[\beta]].$$

(21)

One further important question regards what constitutes ‘real time’ for the system of interest. If ‘real time’ has a characteristic time constant τ in which response typically takes place, then the temperature analog β might be better represented by the product $\beta\tau$, i.e., available energy rate \times characteristic system time.

It is, in any event, now possible to apply Landau’s famous *spontaneous symmetry breaking argument* (Landau and Lifshitz, 2007; Pettini, 2007). The essence of Landau’s insight was that second order phase transitions were usually in the context of a significant symmetry change in the physical states of a system, with one phase being far more symmetric than the other. A symmetry is lost in the transition, a phenomenon called spontaneous symmetry breaking, and symmetry changes are inherently punctuated. The greatest possible set of symmetries in a physical system is that of the Hamiltonian describing its energy states. Usually states accessible at lower temperatures will lack the symmetries available at higher temperatures, so that the lower temperature phase is less symmetric: The randomization of higher temperatures – in this case limited by available free energy densities – ensures that higher symmetry/energy states – vesicles interacting in real time – will then be accessible to the system. Absent high available free energy rates and densities, however, only the simplest structures can be manifest.

Somewhat more rigorously, the generalized renormalization schemes of the Appendix to Wallace and Wallace (2008) may now be imposed on $F_R[K]$ itself, leading to a spectrum of highly punctuated transitions in the overall system of interacting generalized vesicles, neural, or institutional cognitive modules.

Since β (or $\beta\tau$) is proportional to the embedding available free energy density (and/or characteristic or allowed response time),

[1] the greatest possible set of symmetries will be realized for high available energies (and/or characteristic or allowed response time), and

[2] phase transitions, as available energy density rises, will be accompanied by fundamental changes in the underlying topology of the system of interest. These shifts appear to involve the kind of groupoid structures used in Wallace and Wallace (2009).

Wallace (2009) asserts that the sequence of increasingly intimate relations of mutualism, obligate mutualism, symbiosis, and the ‘fusion’ of Serial Endosymbiosis Theory constitute precisely such basic topological transitions. It appears that similar transitions will accrue to any set of interacting cognitive modules, including those of (cognitive) gene expression, internal cellular dynamics, neural structures in higher animals, organized cognitive modules within groups of social animals, and the institutional cognitive modules that have come to so markedly characterize human life.

8 Hierarchical epigenetic catalysis

Following the arguments of Wallace and Wallace (2009), it is possible to model the influence of embedding contexts – what might be called generalized epigenetic effects – on a particular cognitive process rather simply by invoking the Joint Asymptotic Equipartition Theorem (JAEPT) (Cover and Thomas, 1991). For example, given an embedding epigenetic information source, say Y , that affects some cognitive process X having a dual information source uncertainty $H[X]$ is replaced by a joint uncertainty $H(X, Y)$. The objects of interest then become the jointly typical dual sequences $z^n = (x^n, y^n)$, where x is associated with the cognitive process of interest and y with the embedding, and broadly speaking, epigenetic regulatory context. Restricting consideration of x and y to those sequences that are in fact jointly typical allows use of the information transmitted from Y to X as the splitting criterion.

One important inference is that, from the information theory ‘chain rule’ (Cover and Thomas, 1991), $H(X, Y) = H(X) + H(Y|X) \leq H(X) + H(Y)$, while there are approximately $\exp[nH(X)]$ typical X sequences, and $\exp[nH(Y)]$ typical Y sequences, and hence $\exp[n(H(X) + H(Y))]$ independent joint sequences, there are only $\exp[nH(X, Y)] \leq \exp[n(H(X) + H(Y))]$ jointly typical sequences, so that the effect of the embedding context is to lower the relative free energy of a particular developmental channel. Equality occurs only for stochastically independent processes.

Thus the effect of epigenetic regulation is to channel a system’s ontology into pathways that might otherwise be inhibited by an energy barrier. Hence the epigenetic information source Y acts as a *tunable catalyst*, a kind of second order cognitive enzyme, to enable and direct developmental pathways in X . This result permits hierarchical models similar to those of higher order cognitive neural function that incorporate contexts in a natural way (e.g., Wallace and Wallace, 2008; Wallace and Fullilove, 2008). The cost of this ability to channel is the energy necessity for supporting *two* information sources, X and Y , rather than just X itself.

Including hierarchical effects, for example the influence of embedding culture in the development of the human mind, is, according to this formalism, quite straightforward: Consider, for example, culture as another embedding information source, Z , having source uncertainty $H(Z)$. Then the information chain rule becomes

$$H(X, Y, Z) \leq H(X) + H(Y) + H(Z)$$

(22)

and

$$\exp[nH(X, Y, Z)] \leq \exp[n(H(X) + H(Y) + H(Z))],$$

(23)

where, again, equality occurs only under stochastic independence.

Adiabatic changes in these nested information source may occur at significantly different rates, that is, have different time constants, so that the slower might well be considered, in a first approximation, as boundary conditions for the more rapid. More specifically, the embedding information sources program, in a sense, the inner ones by a synergistic regulatory catalysis that makes certain ontological paths more probable than others.

9 Coevolution and coevolutionarily stable states

9.1 The basic model

A somewhat more complicated model can be developed for multiple interacting cognitive processes acting at different rates. The essential point is that the different cognitive phenomena, each characterized by a dual information sources $H_m, m = 1 \dots s$, interact through various signals and that *they become each other's principal environments*, a broadly coevolutionary phenomenon parameterized as well by the inherent time constants of each process, represented by $\tau_m, m = 1 \dots s$, and by the overall available free energy, F .

We write

$$H_m = H_m(\tau_1 \dots \tau_s, F, \dots H_j \dots), j \neq m.$$

(24)

The dynamics of such a system is assumed to be driven by a recursive network of stochastic differential equations, similar to those used to study many other highly parallel dynamic structures (e.g., Wymer, 1997).

Letting the τ_i, F , and the H_j all be represented as parameters Q_j , (with the caveat that H_m not depend on itself), one can define, according to the generalized Onsager development of Wallace and Wallace (2008),

$$S^m \equiv H_m - \sum_i Q_i \partial H_m / \partial Q_i$$

(25)

to obtain a complicated recursive system of phenomenological ‘Onsager relations’ stochastic differential equations,

$$dQ_t^j = \sum_i [L_{j,i}(t, \dots \partial S^m / \partial Q^i \dots) dt + \sigma_{j,i}(t, \dots \partial S^m / \partial Q^i \dots) dB_t^i],$$

(26)

where, again, for notational simplicity only, we have expressed both the H_j , the time constants τ_j , and the available free energy F , in terms of the same symbols Q_j .

m ranges over the H_m and we could allow different kinds of ‘noise’ dB_t^i , having particular forms of quadratic variation that may, in fact, represent a projection of environmental factors under something like a rate distortion manifold (Wallace and Fullilove, 2008).

As usual for such systems, there will be multiple quasi-stable points within a given system’s H_m , representing a class of generalized resilience modes accessible via punctuation.

Second, however, may well be analogs to fragmentation when the system exceeds the critical values of Q_c according to the approach of Wallace and Wallace (2008). That is, the Q -parameter structure will represent full-scale fragmentation of the entire system, and not just punctuation within it.

We thus infer two classes of punctuation possible for this kind of structure.

There are other possible patterns:

[1] Setting equation (26) equal to zero and solving for stationary points gives a set of ‘coevolutionarily stable’ attractor states since the noise terms preclude unstable equilibria.

[2] This system may also converge to limit cycle or ‘strange attractor’ behaviors in which the system seems to chase its tail endlessly, crudely analogous to the cycle of climate-driven phenotype changes in communities of persistent temperate region plants.

[3] What is converged to in both cases is not a simple state or limit cycle of such states. Rather it is an equivalence class, or set of them, of highly dynamic information sources coupled by mutual interaction through crosstalk. Thus ‘stability’ in this extended model represents particular patterns of ongoing dynamics rather than some identifiable ‘state’, although

such dynamics may be indexed by a ‘stable’ set of generalized phenotypes, as it were.

Here we become enmeshed in a highly recursive phenomenological stochastic differential equations, but at a deeper level than the stochastic reaction model of Zhu et al. (2007), and in a dynamic rather than static manner: the objects of this dynamical system are equivalence classes of information sources and their crosstalk, rather than simple final states of a chemical system.

9.2 Roman roads

For the purposes of this paper – a focus on the importance of ‘Roman roads’ connecting cognitive modules having different characteristic time constants – the model above must invoke the rate distortion functions for mutual crosstalk between the different cognitive modules, using the homology of the rate distortion function itself with free energy.

Given different cognitive processes $1\dots s$, the quantities of special interest thus become the mutual rate distortion functions $R_{i,j}$ characterizing communication (and the distortion $D_{i,j}$) between them, while the essential parameters remain the characteristic time constants of each process, $\tau_j, j = 1\dots s$, and the overall, embedding, available free energy density, F .

Taking the Q^α to run over all the relevant parameters and mutual rate distortion functions (including distortion measures $D_{i,j}$), equation (25) becomes

$$S_R^{i,j} \equiv R_{i,j} - \sum_k Q^k \partial R_{i,j} / \partial Q^k, \quad (27)$$

which generalizes equation (8).

Equation (26) accordingly becomes

$$dQ_t^\alpha = \sum_{\beta=(i,j)} [L_\beta(t, \dots \partial S_R^\beta / \partial Q^\alpha \dots) dt + \sigma_\beta(t, \dots \partial S_R^\beta / \partial Q^\alpha \dots) dB_t^\beta], \quad (28)$$

and this generalizes equation (10).

Here we are explicitly modeling the role of crosstalk, its distortion, the inherent time constants of the different cognitive modules, and the overall available free energy density.

This is a very complicated structure indeed, but its general dynamical behaviors will obviously be analogous to those described in just above. For example, setting equation (28) to zero gives the ‘coevolutionary stable states’ of a system of interacting cognitive modules in terms of channel capacity, average distortion, system time constants, and overall available

free energy density. Again, limit cycles and strange attractors seem possible as well. And again, what is converged to is a dynamic behavior pattern, not some fixed ‘state’. And again, such a system will display highly punctuated dynamics almost exactly akin to resilience domain shifts in ecosystems (e.g., Holling, 1973, 1992; Gunderson, 2000). Indeed, the formalism seems directly applicable to ecosystem studies.

10 Discussion and Conclusions

According to Wallace (2009), mutualism, obligate mutualism, symbiosis, and the fusion of SET represent progressively more rapid and less distorted real-time communication between biological structures instantiating information sources. Such communication requires progressively greater channel capacity that, through the homology between information source uncertainty and free energy density, requires progressively more energetic metabolism. The eukaryotic transition, according to this model, may have been entrained by an ecosystem resilience shift from anaerobic to aerobic metabolism.

Wallace (2009) further speculates that, asymptotically, for multicellular structures, while the capacity of a given highly parallel real time biological system will grow according to some power law of its size, the energy needed to maintain and regulate the system, to oppose the entropic distortion torrent or its sudden large deviations jets, in the sense of Wallace and Wallace, (2008), may grow according to another power law altogether. Thus there might well be a crossover point at which the necessity for stabilization will overcome the increase in overall system capacity, suggesting the need to carefully adapt architecture to system size and demand. Thus architectures ‘economical’ at one scale may not be so at other scales. Such relations represent ‘allometric scaling’ (e.g., White and Seymour, 2005; Speakman, 2005) whose form is largely determined by the question of what constitutes ‘real time’ for the system of interest, and by what energy sources are available.

Interacting cognitive modules are ubiquitous in cellular function, neural process, social interaction within small groups, institutional distributed cognition in larger ones, the function of computing machinery and networks, and the operation of economic and imperial hegemonies at the greatest scale. The availability of free energy within these systems determines the maximum possible channel capacities, and hence the maximum rate at which accurate information may be communicated across them. Failure to provide resources, (or the necessity of responding in less than the characteristic time of the system) triggers rate distortion dynamics of decline until punctuated processes of disintegration become manifest.

Conversely, increasing communication between nested sets of interacting cognitive modules, contingent on availability of free energy in various forms, will almost inevitably be followed by a different phase transition, leading to the punctuated establishment of globally-structured, hierarchically-organized and highly integrated phenomena of various kinds. Roman roads are essential, at all scales of living organization.

11 References

- Ash R., 1990, *Information Theory*, Dover Publications, New York.
- Atlan, H., and I. Cohen, 1998, Immune information, self-organization, and meaning, *International Immunology*, 10:711-717.
- Bennett, C., 1988, Logical depth and physical complexity. In *The Universal Turing Machine: A Half-Century Survey*, R. Herkin (ed.), pp. 227-257, Oxford University Press.
- Canfield, D., M. Rosing, and C. Bjerrum, 2006, Early anaerobic metabolisms, *Philosophical Transactions of the Royal Society, B*, 361:1819-1836.
- Cavalier-Smith, T., 2006, Cell evolution and earth history: stasis and revolution, *Philosophical Transactions of the Royal Society, B*, 351:969-1006.
- Cover, T., and J. Thomas, 1991, *Elements of Information Theory*, John Wiley and Sons, New York.
- Dembo, A., and O. Zeitouni, 1998, *Large Deviations and Applications*, Second Edition, Springer, New York.
- Eigen, M., 1971, Self-organization of matter and the evolution of biological macromolecules, *Naturwissenschaften*, 58:465-523.
- Eigen, M., 1996, *Steps Toward Life: Perspectives on Evolution*, Oxford University Press, New York.
- Ellis, R., 1985, *Entropy, Large Deviations, and Statistical Mechanics*, Springer, New York.
- Embley, T., 2006, Multiple secondary origins of the anaerobic lifestyle in eukaryotes, *Philosophical Transactions of the Royal Society, B*, 361:1055-1067.
- Feynman, R., 2000, *Lectures on Computation*, Westview Press, New York.
- Gunderson, L., 2000, Ecological resilience – in theory and application, *Annual Reviews of Ecological Systematics*, 31:425-439.
- Holling, C., 1973, Resilience and stability of ecological systems, *Annual Reviews of Ecological Systematics*, 4:1-23.
- Holling, C., 1992, Cross-scale morphology, geometry, and dynamics of ecosystems, *Ecological Monographs*, 41:1-50.
- Khinchin, A., 1957, *Mathematical Foundations of Information Theory*, Dover, New York.
- Landau, L., and E. Lifshitz, 2007, *Statistical Physics, 3rd Edition*, Part I, Elsevier, New York.
- Margulis/Sagan, L., 1966, On the origin of mitosing cells, *Journal of Theoretical Biology*, 14:255-274.
- Margulis, L., 1996, Archaeal-eubacterial mergers in the origin of eukarya: phylogenetic classification of life, *Proceedings of the National Academy of Sciences*, 93:1071-1076.
- Margulis, L., 2004, Serial endosymbiotic theory (SET) and composite individuality. Transition from bacterial to eukaryotic genomes, *Microbiology Today*, 31:173-174.
- Pettini, M., 2007, *Geometry and Topology in Hamiltonian Dynamics and Statistical Mechanics*, Springer, New York.
- Protter, P., 1995, *Stochastic Integration and Differential Equations: A New Approach*, Springer, New York.
- Rockafellar, R., 1970, *Convex Analysis*, Princeton University Press, Princeton, NJ.
- Speakman, J., 2005, Body size, energy metabolism, and lifespan, *Journal of Experimental Biology*, 208(Pt 9):1717-1730.
- Wallace, R., 2005, *Consciousness: A Mathematical Treatment of the Global Neuronal Workspace Model*, Springer, New York.
- Wallace, R., 2008, Toward formal models of biologically inspired, highly parallel, machine cognition, *International Journal of Parallel, Emergent, and Distributed Systems*, 23367-408.
- Wallace, R., 2009, Metabolic constraints of the eukaryotic transition, *Origins of Life and Evolution of Biospheres*, 39:165-176.
- Wallace, R., and M. Fullilove, 2008, *Collective Consciousness and its Discontents: Institutional Distributed Cognition, Racial Policy, and Public Health in the United States*, Springer, New York.
- Wallace, R., and R.G. Wallace, 2008, On the spectrum of prebiotic chemical systems: an information-theoretic treatment of Eigen's paradox, *Origins of Life and Evolution of Biospheres*, 38:419-456.
- Wallace, R., and D. Wallace, 2008, Punctuated equilibrium in statistical models of generalized coevolutionary resilience: how sudden ecosystem transitions can entrain both phenotype expression and Darwinian selection, *Transactions on Computational Systems Biology IX*, LNBI 5121:23-85.
- Wallace, R., and D. Wallace, 2009, Code, context, and epigenetic catalysis in gene expression, *Transactions on Computational Systems Biology XI*, LNBI 5750:283-334.
- White, C., and R. Seymour, 2005, Allometric scaling of mammalian metabolism, *Journal of Experimental Biology*, 208(Pt 9):1611-1619.
- Witzany, G., 2006, Serial endosymbiotic theory (SET): the biosemiotic update, *Acta Biotheoretica*, 54:103-117.
- Wymer, C., 1997, Structural nonlinear continuous-time models in econometrics, *Macroeconomic Dynamics*, 1:518-548.
- Zhu, R., A. Rebirio, D. Salahub, and S. Kaufmann, 2007, Studying genetic regulatory networks as the molecular level: delayed reaction stochastic models, *Journal of Theoretical Biology*, 246:725-745.