Version 7

Dysfunctions of highly parallel real-time machines as 'developmental disorders': Security concerns and a *Caveat Emptor*

Rodrick Wallace
Division of Epidemiology
The New York State Psychiatric Institute*

September 15, 2009

Abstract

A cognitive paradigm for gene expression in developmental biology that is based on rigorous application of the asymptotic limit theorems of information theory can be adapted to highly parallel real-time computing. The coming Brave New World of massively parallel 'autonomic' and 'Self-X' machines driven by the explosion of multiple core and molecular computing technologies will not be spared patterns of canonical and idiosyncratic failure analogous to the developmental disorders affecting organisms that have had the relentless benefit of a billion years of evolutionary pruning. This paper provides a warning both to potential users of these machines and, given that many such disorders can be induced by external agents, to those concerned with larger scale matters of homeland security.

Key Words: developmental disorders, cognition, highly parallel computation, homeland security, information theory, real time.

1 Introduction

A cognitive paradigm for gene expression has emerged in which contextual factors determine the behavior of what Cohen calls a 'reactive system', not at all a deterministic, or even stochastic, mechanical process (e.g., Cohen, 2006; Cohen and Harel, 2007; Wallace and Wallace, 2008, 2009). The different highly formal approaches are much in the spirit of the pioneering efforts of Maturana and Varela (1980, 1992) who foresaw the essential role that cognitive process must play across a broad spectrum of biological phenomena.

A relatively simple model of cognitive process as an information source – a generalized 'language' – permits use of Dretske's (1994) insight that any cognitive phenomenon must be constrained by the limit theorems of information theory, in the same sense that sums of stochastic variables are constrained by the Central Limit Theorem. The viewpoint permits a new formal approach to gene expression and its dysfunctions (e.g., Wallace and Wallace, 2009, 2010). We have already applied something like this to highly parallel computation (Wallace 2006, 2008a, 2009, 2010), and now extend that work to illustrate the broad range of pathological behaviors such machines may express, taking a perspective much like that of developmental biology.

Machine failure becomes, from this viewpoint, analogous to developmental dysfunction, in the sense of Wallace (2008b), who argues that ecosystem resilience theory permits novel exploration of comorbidity in developmental psychiatric and chronic physical disorders. Structured psychosocial stress, chemical agents, and similar noxious exposures, can write distorted images of themselves onto child growth, and, if sufficiently powerful, adult development as well, inducing a punctuated life course trajectory to characteristic forms of comorbid mind/body dysfunction. For an individual, within the linked network of broadly cognitive physiological and mental subsystems, this occurs in a manner recognizably similar to resilience domain shifts affecting a stressed ecosystem, suggesting that reversal or palliation may often be exceedingly difficult.

We will suggest similar problems must necessarily afflict highly parallel 'autonomic' or 'Self-X' machines, particularly those acting in real time. The current stampede by industry toward using such devices to govern our transportation, power, and communication systems, to control individual vehicles under fly-by-wire protocols, run nuclear reactors, chemical plants, oil refineries, and so on, raises important issues of homeland security, given the relative ease of inducing developmental disorders in the highly parallel systems that determine gene expression.

We begin by recapitulating recent models of development, following Wallace and Wallace (2008, 2009). Subsequent sections significantly expand that work.

*Address correspondence to: R. Wallace, 549 W. 123 St., 16F, New York, NY, 10027. rdwall@ix.netcom.com. Affiliation for identification only.

2 Models of development

The popular spinglass model of development (e.g., Ciliberti et al., 2007a, b) assumes that N transcriptional regulators, are represented by their expression patterns $\mathbf{S}(t) = [S_1(t), ..., S_N(t)]$ at some time t during a developmental or cell-biological process and in one cell or domain of an embryo. The transcriptional regulators influence each other's expression through cross-regulatory and autoregulatory interactions described by a matrix $w = (w_{ij})$. For nonzero elements, if $w_{ij} > 0$ the interaction is activating, if $w_{ij} < 0$ it is repressing. w represents, in this model, the regulatory genotype of the system, while the expression state $\mathbf{S}(t)$ is the phenotype. These regulatory interactions change the expression of the network $\mathbf{S}(t)$ as time progresses according to a difference equation

$$S_i(t + \Delta t) = \sigma[\sum_{j=1}^{N} w_{ij} S_j(t)],$$

(1)

where Δt is a constant and σ a sigmodial function whose value lies in the interval $(-1, 1)$. In the spinglass limit σ is the sign function, taking only the values ± 1.

The regulatory networks of interest here are those whose expression state begins from a prespecified initial state $\mathbf{S}(0)$ at time $t = 0$ and converge to a prespecified stable equilibrium state $\mathbf{S}_\infty$. Such networks are termed *viable* and must necessarily be a very small fraction of the total possible number of networks, since most do not begin and end on the specified states. This 'simple' observation is not at all simple in our reformulation, although other results become far more accessible, as we can then invoke certain asymptotic limit theorems of applied probability.

This treatment of development is formally similar to spinglass neural network models of learning by selection, e.g., as proposed by Toulouse et al. (1986) nearly a generation ago. Much subsequent work, summarized by Dehaene and Naccache (2001), suggests that such models are simply not sufficient to the task of understanding high level cognitive function, and these have been largely supplanted by complicated 'global workspace' concepts whose mathematical characterization is highly nontrivial (Atmanspacher, 2006).

Wallace and Wallace (2008, 2009) shift the perspective on development by invoking a cognitive paradigm for gene expression, following the example of the Atlan/Cohen model of immune cognition.

Atlan and Cohen (1998), in the context of a study of the immune system, argue that the essence of cognition is the comparison of a perceived signal with an internal, learned picture of the world, and then choice of a single response from a large repertoire of possible responses.

Such choice inherently involves information and information transmission since it always generates a reduction in uncertainty, as explained by Ash (1990, p. 21).

More formally, a pattern of incoming input – like the $\mathbf{S}(t)$ above – is mixed in a systematic algorithmic manner with a pattern of internal ongoing activity – like the (w_{ij}) above – to create a path of combined signals $x = (a_0, a_1, ..., a_n, ...)$ – analogous to the sequence of $\mathbf{S}(t + \Delta t)$ above, with, say, $n = t/\Delta t$. Each a_k thus represents some functional composition of internal and external signals.

This path is fed into a highly nonlinear decision oscillator, h, a 'sudden threshold machine', in a sense, that generates an output $h(x)$ that is an element of one of two disjoint sets B_0 and B_1 of possible system responses. Let us define the sets B_k as

$$B_0 \equiv \{b_0, ..., b_k\},$$

$$B_1 \equiv \{b_{k+1}, ..., b_m\}.$$

Assume 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 has been recognized, and some action $b_j, k + 1 \leq j \leq m$ takes place.

The principal objects of formal interest are paths x triggering pattern recognition-and-response. That is, given a fixed initial state a_0, examine all possible subsequent paths x beginning with a_0 and leading to the event $h(x) \in B_1$. Thus

$$h(a_0, ..., a_j) \in B_0$$

for all $0 < j < m$,

but

$$h(a_0, ..., a_m) \in B_1.$$

For each positive integer n, let $N(n)$ be the number of high probability grammatical and syntactical paths of length n which begin with some particular a_0 and lead to the condition $h(x) \in B_1$. Call such paths 'meaningful', assuming, not unreasonably, that $N(n)$ will be considerably less than the number of all possible paths of length n leading from a_0 to the condition $h(x) \in B_1$.

While the combining algorithm, the form of the nonlinear oscillator, and the details of grammar and syntax are all unspecified in this model, the critical assumption permitting inference of the necessary conditions constrained by the asymptotic limit theorems of information theory is that the finite limit

$$H \equiv \lim_{n \to \infty} \frac{\log[N(n)]}{n}$$

both exists and is independent of the path x.

Define such a pattern recognition-and-response cognitive process as *ergodic*. Not all cognitive processes are likely to be ergodic in this sense, implying that H, if it indeed exists at all, is path dependent, although extension to nearly ergodic processes seems possible (Wallace and Fullilove, 2008).

Invoking the spirit of the Shannon-McMillan Theorem, as choice involves an inherent reduction in uncertainty, it is then possible to define an adiabatically, piecewise stationary, ergodic (APSE) information source $\mathbf{X}$ associated with stochastic variates X_j having joint and conditional probabilities $P(a_0, ..., a_n)$ and $P(a_n|a_0, ..., a_{n-1})$ such that appropriate conditional and joint Shannon uncertainties satisfy the classic relations

$$H[\mathbf{X}] = \lim_{n \to \infty} \frac{\log[N(n)]}{n} =$$

$$\lim_{n \to \infty} H(X_n|X_0, ..., X_{n-1}) =$$

$$\lim_{n \to \infty} \frac{H(X_0, ..., X_n)}{n+1}.$$

(2)

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

Adiabatic means that the source has been parametrized according to some scheme, and that, over a certain range, along a particular piece, as the parameters vary, the source remains as close to stationary and ergodic as needed for information theory's central theorems to apply. *Stationary* means that the system's probabilities do not change in time, and *ergodic*, roughly, that the cross sectional means approximate long-time averages. Between pieces it is necessary to invoke various kinds of phase transition formalisms, as described more fully in Wallace (2005) or Wallace and Wallace (2008).

In the developmental vernacular of Ciliberti et al., we now examine paths in phenotype space that begins at some $\mathbf{S}_0$ and converge $n = t/\Delta t \to \infty$ to some other $\mathbf{S}_\infty$. Suppose the system is conceived at $\mathbf{S}_0$, and h represents (for example) reproduction when phenotype $\mathbf{S}_\infty$ is reached. Thus $h(x)$ can have two values, i.e., B_0 not able to reproduce, and B_1, mature enough to reproduce. Then $x = (\mathbf{S}_0, \mathbf{S}_{\Delta t}, ..., \mathbf{S}_{n\Delta t}, ...)$ until $h(x) = B_1$.

Structure is now subsumed *within the sequential grammar and syntax of the dual information source* rather than within the cross sectional internals of (w_{ij})-space, a simplifying shift in perspective.

This transformation carries computational burdens, as well as providing mathematical insight.

First, the fact that viable networks comprise a tiny fraction of all those possible emerges easily from the spinglass formulation simply because of the 'mechanical' limit that the number of paths from $\mathbf{S}_0$ to $\mathbf{S}_\infty$ will always be far smaller than the total number of possible paths, most of which simply do not end on the target configuration.

From the information source perspective, which inherently subsumes a far larger set of dynamical structures than possible in a spinglass model – not simply those of symbolic dynamics – the result is what Khinchin (1957) characterizes as the 'E-property' of a stationary, ergodic information source. This property allows, in the limit of infinitely long output, the classification of output strings into two sets:

[1] a very large collection of gibberish which does not conform to underlying (sequential) rules of grammar and syntax, in a large sense, and which has near-zero probability, and

[2] a relatively small 'meaningful' set, in conformity with underlying structural rules, having very high probability.

The essential content of the Shannon-McMillan Theorem is that, if $N(n)$ is the number of meaningful strings of length n, then the uncertainty of an information source X can be defined as $H[X] = \lim_{n\to\infty} \log[N(n)]/n$, that can be expressed in terms of joint and conditional probabilities. Proving these results for general stationary, ergodic information sources requires considerable mathematical machinery (e.g., Khinchin, 1957; Cover and Thomas, 1991; Dembo and Zeitouni, 1998).

Second, according to Ash (1990) information source uncertainty has an important heuristic interpretation in that we may regard a portion of text in a particular language as being produced by an information source. 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.

Third, information source uncertainty is homologous with free energy density in a physical system, a matter having implications across a broad class of dynamical behaviors.

The free energy density of a physical system having volume V and partition function $Z(K)$ derived from the system's Hamiltonian – the energy function – at inverse temperature K is (e.g., Landau and Lifshitz 2007)

$$F[K] = \lim_{V \to \infty} -\frac{1}{K} \frac{\log[Z(K, V)]}{V} =$$

$$\lim_{V \to \infty} \frac{\log[\hat{Z}(K, V)]}{V},$$

(3)

where $\hat{Z} = Z^{-1/K}$.

The partition function for a physical system is the normalizing sum in an equation having the form

$$P[E_i] = \frac{\exp[-E_i/kT]}{\sum_j \exp[-E_j/kT]},$$

(4)

where E_i is the energy of state i, k a constant, and T the system temperature.

Feynman (2000), following the classic approach by Bennett (1988), who examined idealized machines using information to do work, concludes that *the information contained in a message is most simply measured by the free energy needed to erase it.*

Thus, according to this argument, source uncertainty is homologous to free energy density as defined above, i.e., from the similarity with the relation $H = \lim_{n\to\infty} \log[N(n)]/n$.

Ash's perspective then has an important corollary: If, for a biological system, $H_1 > H_2$, source 1 will require more metabolic or other free energy than source 2.

For highly parallel cognitive machines, we envision the system as beginning a calculation at a 'machine phenotype' $\mathbf{S}_0$ and converging on an 'answer' $\mathbf{S}_\infty$ that may be some fixed point or, in the sense of Wallace (2008a, 2009, 2010), may be a dynamic behavior represented by a particular information source.

2.1 Tunable epigenetic catalysis

Following the direction of Wallace and Wallace (2009), incorporating the influence of embedding contexts – epigenetic effects or the desires of the programmer – is most elegantly done by invoking the Joint Asymptotic Equipartition Theorem (JAEPT) (Cover and Thomas, 1991). For example, given an embedding contextual information source, say Z, that affects development represented by a dual information source X, then the dual cognitive source uncertainty $H[X]$ is replaced by a joint uncertainty $H(X, Z)$. The objects of interest then become the jointly typical dual sequences $y^n = (x^n, z^n)$, where x is associated with cognitive gene expression and z with the embedding context. Restricting consideration of x and z to those sequences that are in fact jointly typical allows use of the information transmitted from Z to X as the splitting criterion.

One important inference is that, from the information theory 'chain rule' (Cover and Thomas, 1991),

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

while there are approximately $\exp[nH(X)]$ typical X sequences of length n, and $\exp[nH(Z)]$ typical Z sequences, and hence $\exp[n(H(x) + H(Z))]$ independent joint sequences, there are only

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

(5)

jointly typical sequences, so that the effect of the embedding context Z is to lower the *relative* free energy of a particular developmental channel for the system, (while, of course, raising the total free energy needed to sustain both the main and the regulatory processes).

Thus the effect of epigenetic regulation – programming – is to channel development into pathways that might otherwise be inhibited. Hence the epigenetic information source Z acts as a *tunable catalyst*, a kind of second order cognitive enzyme, to enable and direct developmental pathways. 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).

This elaboration allows a spectrum of possible 'final' phenotypes, what Gilbert (2001) calls developmental or phenotype plasticity. Thus gene expression is seen as, in part, responding to environmental or other, internal, developmental signals.

Wallace and Wallace (2009) thus provide a formal basis for the important arguments of West-Eberhard (2005) who finds that individual development can be visualized as a series of branching pathways, each a developmental decision, or switch point, governed by some regulatory apparatus, and each switch point defines a modular trait. There cannot, West-Eberhard concludes, be a change in the phenotype, a novel phenotypic state, without an altered developmental pathway.

Machine analogs seem obvious.

In what follows we will present three distinct portraits of this process in highly parallel systems, all formally analogous to ecosystem resilience shifts.

3 Phase transitions in cognitive systems

3.1 Spontaneous symmetry breaking

A formal equivalence class algebra can now be constructed by choosing different origin and end points $\mathbf{S}_0, \mathbf{S}_\infty$ and defining equivalence of two states by the existence of a high probability meaningful path connecting them with the same origin and end. Disjoint partition by equivalence class, analogous to orbit equivalence classes for dynamical systems, defines the vertices of the proposed network of cognitive dual languages, much enlarged beyond the spinglass example. We thus envision a network of metanetworks. Each vertex then represents a different equivalence class of information sources dual to a cognitive process. This is an abstract set of metanetwork 'languages' dual to the cognitive processes of gene expression and development.

This structure generates a groupoid, in the sense of Weinstein (1996). Wallace (2008a) provides a summary.

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 and end points, 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 (very large) groupoid.

There is a hierarchy in groupoid structures. First, there is structure *within the system having the same base and end points*, as in Ciliberti et al. Second, there is a complicated groupoid structure defined by sets of dual information sources surrounding the variation of base and end points. We do not need to know what that structure is in any detail, but can show that its existence has profound implications.

First we examine the simple case, the set of dual information sources associated with a fixed pair of beginning and end states.

Taking the serial grammar/syntax model above, we find that not all high probability meaningful paths from $\mathbf{S}_0$ to $\mathbf{S}_\infty$ are the same. They are structured by the uncertainty of the associated dual information source, and that has a homological relation with free energy density.

Let us index possible dual information sources connecting base and end points by some set $A = \cup\alpha$. Argument by abduction from statistical physics is direct: Given, for a biological system, some metabolic energy density available at a rate M, and an allowed development time τ, let $K = 1/\kappa M\tau$ for some appropriate scaling constant κ, so that $M\tau$ is total developmental free energy. Then the probability of a particular H_α will be determined by the standard expression (e.g., Landau and Lifshitz, 2007),

$$P[H_\beta] = \frac{\exp[-H_\beta K]}{\sum_\alpha \exp[-H_\alpha K]},$$

where the sum may, in fact, be a complicated abstract integral. This is just a version of the fundamental probability relation from statistical mechanics, as above. The sum in the denominator, the partition function in statistical physics, is a crucial normalizing factor that allows the definition of of $P[H_\beta]$ as a probability. A basic requirement, then, is that the sum/integral always converges. K, in a biological system, is the inverse product of a scaling factor, a metabolic energy density rate term, and a characteristic development time τ. The developmental energy might be raised to some power, e.g., $K = 1/(\kappa(M\tau)^b)$, suggesting the possibility of allometric scaling.

Some dual information sources will be 'richer'/smarter than others, but, conversely, must use more metabolic energy for their completion.

Wallace (2010), as summarized in the Appendix, argues that the behavior of highly parallel cognitive machines acting in real time is driven, among other factors, by the average distortion, D, between machine intent and machine impact, measured in terms of a *Rate Distortion Function* $R(D)$ and a characteristic machine response time τ. $R(D)$ is necessarily a convex function in D (Cover and Thomas, 1991), and this, since it defines a channel capacity that can be measured as a free energy analog, drives much of the dynamics of machine/environment interaction. Wallace (2010), again summarized in the Appendix, identifies a rate distortion free energy for the system, in terms of a particular groupoid associated with dual information sources within the machine.

While we might simply impose an equivalence class structure based on equal levels of energy/source uncertainty, producing a groupoid, we can do more by now allowing both source and end points to vary, as well as by imposing energy-level equivalence. This produces a far more highly structured groupoid that we now investigate.

To retierate, equivalence classes define groupoids, by standard mechanisms (e.g., Weinstein, 1996; Brown, 1987; Golubitsky and Stewart, 2006). The basic equivalence classes – here involving both information source uncertainty level and the variation of $\mathbf{S}_0$ and $\mathbf{S}_\infty$, will define transitive groupoids, and higher order systems can be constructed by the union of transitive groupoids, having larger alphabets that allow more complicated statements in the sense of Ash above.

Again, given an appropriately scaled, dimensionless, fixed, inverse available metabolic energy density rate and development time, or in machine terms, a fixed $R(D)\tau$, so that $K = 1/\kappa M\tau$, or $K = 1/\kappa R(D)\tau$, we propose that the probability of an information source representing equivalence class G_i, H_{G_i}, will be given by

$$P[H_{G_i}] = \frac{\exp[-H_{G_i} K]}{\sum_j \exp[-H_{G_j} K]},$$

where the sum/integral of the denominator is over all possible elements of the largest available symmetry groupoid. By the arguments of Ash above, compound sources, formed by the union of underlying transitive groupoids, being more complex, generally having richer alphabets, as it were, will all have higher free-energy-density-equivalents than those of the base (transitive) groupoids.

Let

$$Z_G \equiv \sum_j \exp[-H_{G_j} K].$$

We now define the *Groupoid free energy* of the system, F_G, at inverse temperature' K, as

$$F_G[K] \equiv -\frac{1}{K}\log[Z_G[K]].$$

The groupoid free energy construct permits introduction of important ideas from statistical physics.

We have expressed the probability of an information source in terms of its relation to a fixed, scaled, available (inverse) 'temperature' K. This gives a statistical thermodynamic path leading to definition of a 'higher' free energy construct – $F_G[K]$ – to which we now apply Landau's fundamental heuristic phase transition argument (Landau and Lifshitz 2007; Skierski et al. 1989; Pettini 2007). See, in particular, Pettini (2007) for details.

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 metabolic free energy densities – ensures that higher symmetry/energy states – mixed transitive groupoid structures – will then be accessible to the system. Absent high metabolic free energy rates and densities, however, only the simplest transitive groupoid structures can be manifest. A full treatment from this perspective seems to require invocation of groupoid representations, no small matter (e.g., Buneci, 2003; Bos 2006).

Something like Pettini's (2007) Morse-Theory-based topological hypothesis can now be invoked, i.e., that changes in underlying groupoid structure are a necessary (but not sufficient) consequence of phase changes in $F_G[K]$. Necessity, but not sufficiency, is important, as it, in theory, allows mixed groupoid symmetries.

Using this formulation, the mechanisms of epigenetic catalysis are accomplished by allowing the set B_1 above to span a distribution of possible 'final' states $\mathbf{S}_\infty$. Then the groupoid arguments merely expand to permit traverse of both initial states and possible final sets, recognizing that there can now be a possible overlap in the latter, and the epigenetic effects are realized through the joint uncertainties $H(X_{G_i}, Z)$, so that the epigenetic information source Z serves to direct as well the possible final states of X_{G_i}. Scherrer and Jost (2007a, b) use information theory arguments to suggest something similar.

3.2 The developmental holonomy groupoid in phenotype space

There is another, more direct, way to look at phase transitions in cognitive systems, adapting perspectives of homotopy and holonomy directly within 'machine phenotype' space. This is related to the constructions above. We begin with ideas of directed homotopy.

In conventional topology one constructs equivalence classes of loops that can be continuously transformed into one another on a surface. The prospect of interest is to attempt to collapse such a family of loops to a point while remaining within the surface. If this cannot be done, there is a hole. Here we are concerned, as in figure 1, with sets of one-way developmental trajectories, beginning with an initial machine phenotype $\mathbf{S_i}$, and converging on some final phenotype, here characteristic behavioral patterns simplified as 'metabolic' cycles labeled, respectively, $\mathbf{S_n}$ and $\mathbf{S_o}$. The filled triangle represents the effect of some external epigenetic 'farmer/programmer' in the sense of Wallace (2009, 2010) acting at a critical developmental period represented by the initial phenotype $\mathbf{S_i}$.

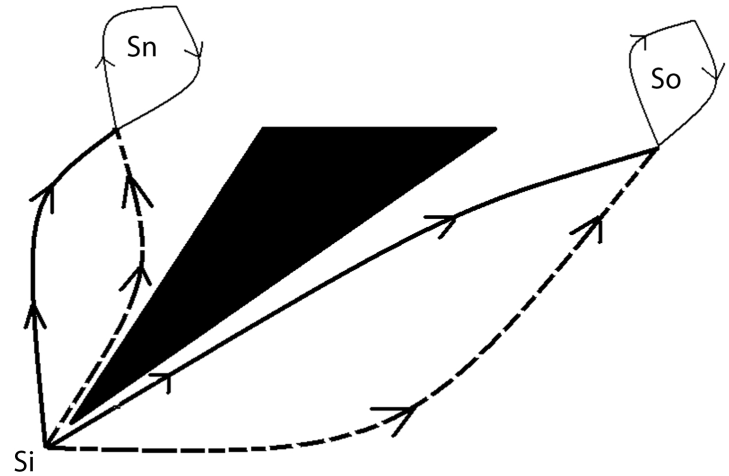

Figure 1: Developmental homotopy equivalence classes in machine phenotype space. The set on one-way paths from $\mathbf{S_i}$ to $\mathbf{S_n}$ represents an equivalence class of developmental trajectories converging on a 'machine phenotype'. In the presence of an external epigenetic catalyst, a 'farming' program, or a noxious interaction by a real-time system with its context, for example an increasing average distortion, D, between machine intent and machine impact, developmental trajectories can converge on another, possibly pathological, phenotype, $\mathbf{S_o}$.

In the sense of Wallace (2010), however, a real-time system is 'programmed' as much by interaction with embedding context as by the desires of the programmer, and will become subject to noxious epigenetic influence in much the same manner as a developing organism.

We assume phenotype space to be directly measurable and to have a simple 'natural' metric defining the difference between developmental paths.

Developmental paths continuously transformable into each other without crossing the filled triangle define equivalence classes characteristic of different information sources dual to cognitive gene expression, as above.

Given a metric on phenotype space, and given equivalence classes of developmental trajectories having more than one path each, we can *pair one-way developmental trajectories* to make loop structures. In figure 1 the solid and dotted lines above and below the filled triangle can be pasted together to make loops characteristic of the different developmental equivalence classes. Although figure 1 is represented as topologically flat, there is no inherent reason for the phenotype manifold itself to be flat. The existence of a metric in phenotype space permits determining the degree of curvature, using standard methods. Figure 2 shows a loop on some manifold. Using the metric definition it is possible to *parallel transport* a tangent vector starting at point s around the loop, and to measure the angle between the initial and final vectors, as in-

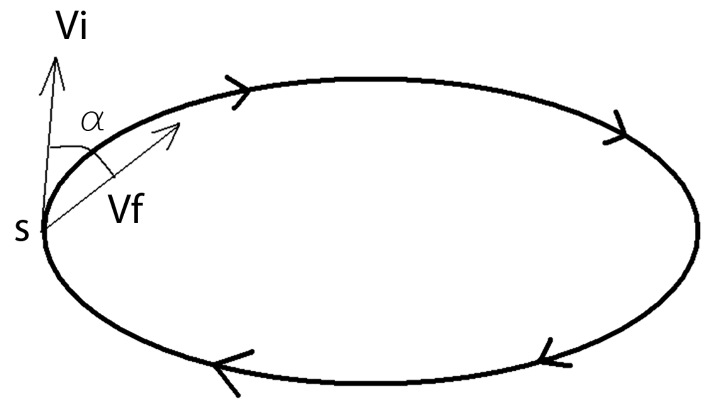

Figure 2: Parallel transport of a tangent vector $V_i \rightarrow V_f$ around a loop on a manifold. Only for a geometrically flat object will the angle between the initial and final vectors be zero. By a fundamental theorem the path integral around the loop by parallel displacement is the surface integral of the curvature over the loop.

dicated. A central result from elementary metric geometry is that the angle α will be given by the integral of the curvature tensor of the metric over the interior of the loop (e.g., Frankel, 2006, Section 9.6).

The *holonomy group* is defined as follows (e.g., Helgason, 1962):

If s is a point on a manifold M having a metric, then the holonomy group of M is the group of all linear transformations of the tangent space M_s obtained by parallel translation along closed curves starting at s.

For figure 1 the *phenotype holonomy groupoid* is the disjoint union of the different holonomy groups corresponding to the different branches separated by 'developmental shadows' induced by epigenetic information sources acting as developmental catalysts/programs.

The relation between the phenotype groupoid as defined here and the phase transitions in $F_G[K]$ as defined above is an open question.

3.3 Higher holonomy: Imposing a metric on the manifold of dual information sources

Glazebrook and Wallace (2009a) examined holonomy groupoid phase transition arguments for networks of interacting information sources dual to cognitive phenomena. A more elementary form of this arises directly through extending holonomy groupoid arguments to a manifold of different information source dual to cognitive phenomena as follows.

Different cognitive phenomena will have different dual information sources, and we are interested in the local properties of the system near a particular reference state. We impose a topology on the system, so that, near a particular 'language' A, dual to an underlying cognitive process, there is an open set U of closely similar languages $\hat{A}$, such that $A, \hat{A} \subset U$. It may be necessary to coarse-grain the system's responses to define these information sources. The problem is to proceed in such a way as to preserve the underlying essential topology, while eliminating 'high frequency noise'. The formal tools for this can be found elsewhere, e.g., in Chapter 8 of Burago et al. (2001).

Since the information sources dual to the cognitive processes are similar, for all pairs of languages $A, \hat{A}$ in U, it is possible to:

[1] Create an embedding alphabet which includes all symbols allowed to both of them.

[2] Define an information-theoretic distortion measure in that extended, joint alphabet between any high probability (grammatical and syntactical) paths in A and $\hat{A}$, which we write as $d(Ax, \hat{A}x)$ (Cover and Thomas, 1991, and the appendix). Note that these languages do not interact, in this approximation.

[3] Define a metric on U, for example,

$$\mathcal{M}(A, \hat{A}) = |\lim \frac{\int_{A,\hat{A}} d(Ax, \hat{A}x)}{\int_{A,A} d(Ax, A\hat{x})} - 1|, \tag{6}$$

using an appropriate integration limit argument over the high probability paths. Note that the integration in the denominator is over different paths within A itself, while in the numerator it is between different paths in A and $\hat{A}$. Other metric constructions on U seem possible. Structures weaker than a conventional metric would be of more general utility, but the mathematical complications are formidable.

Note that these conditions can be used to define equivalence classes of *languages* dual to cognitive processes, where previously we defined equivalence classes of *states* that could be linked by high probability, grammatical and syntactical paths connecting two phenotypes. This led to the characterization of different information sources. Here we construct an entity, formally a topological manifold, *that is an equivalence class of information sources*. This is, provided $\mathcal{M}$ is a conventional metric, a classic differentiable manifold. The set of such equivalence classes generates the *dynamical groupoid*, and questions arise regarding mechanisms, internal or external, which can break that groupoid symmetry.

Since H and $\mathcal{M}$ are both scalars, a 'covariant' derivative can be defined directly as

$$dH/d\mathcal{M} = \lim_{\hat{A} \to A} \frac{H(A) - H(\hat{A})}{\mathcal{M}(A, \hat{A})}, \tag{7}$$

where $H(A)$ is the source uncertainty of language A.

Suppose the system to be set in some reference configuration A_0.

To obtain the unperturbed dynamics of that state, impose a Legendre transform using this derivative, defining another scalar

$$S \equiv H - \mathcal{M} dH/d\mathcal{M}. \quad (8)$$

The simplest possible generalized Onsager relation – here seen as an empirical, fitted, equation like a regression model – is $d\mathcal{M}/dt = LdS/d\mathcal{M}$, where t is the time and $dS/d\mathcal{M}$ represents an analog to the thermodynamic force in a chemical system. This is seen as acting on the reference state A_0.

Explicit parameterization of $\mathcal{M}$ introduces standard – and quite considerable – notational complications (Burago et al., 2001): Imposing a metric for different cognitive dual languages parameterized by $\mathbf{K}$ leads to Riemannian, or even Finsler, geometries, including the usual geodesics (e.g., Wallace and Fullilove, 2008; Glazebrook and Wallace, 2009a, b).

The dynamics, as we have presented them so far, have been noiseless. The simplest generalized Onsager relation in the presence of noise might be rewritten as

$$d\mathcal{M}/dt = LdS/d\mathcal{M} + \sigma W(t),$$

where σ is a constant and $W(t)$ represents white noise. Again, S is seen as a function of the parameter $\mathcal{M}$. This leads directly to a familiar family of stochastic differential equations,

$$d\mathcal{M}_t = L(t, \mathcal{M})dt + \sigma(t, \mathcal{M})dB_t, \quad (9)$$

where L and σ are appropriately regular functions of t and $\mathcal{M}$, and dB_t represents the noise structure, characterized by its quadratic variation. In the sense of Emery (1989), this leads into deep realms of stochastic differential geometry and related topics: Imagine figure 1 as blurred by Brownian or more structured stochastic fuzz.

We have defined a groupoid for the system based on a particular set of equivalence classes of information sources dual to cognitive processes. That groupoid parsimoniously characterizes the available dynamical manifolds, and breaking of the groupoid symmetry by epigenetic crosstalk creates more complex objects of considerable interest. This leads to the possibility, indeed, the necessity of epigenetic *Deus ex Machina* mechanisms – analogous to programming, stochastic resonance, etc. – to force transitions between the different possible modes within and across dynamic manifolds. In one model the external 'programmer' creates the manifold structure, and the system hunts within that structure for the 'solution' to the problem according to equivalence classes of paths on the manifold. Noise, as with random mutation in evolutionary algorithms, precludes simple unstable equilibria, but not other possible structures.

Equivalence classes of *states* gave dual information sources. Equivalence classes of *information sources* give different characteristic dynamical manifolds. Equivalence classes of one-way developmental *paths* produce different directed homotopy topologies characterizing those dynamical manifolds. This introduces the possibility of having different quasi-stable modes *within* individual manifolds, and leads to ideas of holonomy and the holonomy groupoid of the set of quasi-stable developmental modes.

4 Discussion

For highly parallel real time cognitive machines, deterioration in the communication with the embedding environment can be measured by decline of a rate distortion function $R(D)$ consequent on increase of the average distortion D between machine intent and machine impact (Wallace, 2010, and the Appendix). In the context of some fixed response time τ, this acts to lower an internal 'machine cognitive temperature', driving the system to simpler and less-rich behaviors, and ultimately condensing to some 'absolute zero' configuration, possibly in a highly punctuated manner.

This is only one failure mode of many possible.

While we have not yet fully rationalized the three versions of punctuated resilience shifts in machine development presented above – we do not know precisely how the three pictures represent each other, and this is a subject of active research – the implications for highly parallel real time cognitive machines are not encouraging.

Such machines necessarily undergo developmental trajectories, from an initial configuration to a final 'machine phenotype' that is either some fixed 'answer' or else a desired dynamic behavior (Wallace, 2008a, 2009, 2010). These are affected, not only by the desires of the programmer or the limitations of the machine, by also by the distortion between machine intent and machine impact, and by the effects of epigenetic 'noxious exposures' that may not be random, causing the functional equivalence of developmental disorders resulting in pathological machine phenotypes. These might be either 'wrong answers' or 'wrong strategies', and the analogy with ecosystem resilience shifts suggests that prevention or correction will not be easy (Wallace, 2008b).

A comparison with the status of our understanding of human psychopathology is of some interest. In spite of some two hundred years of the scientific study of mental disorders, Johnson-Laird et al. (2006) felt compelled to write that

> Current knowledge about psychological illnesses

> is comparable to the medical understanding of epidemics in the early 19th century. Physicians realized that cholera, for example, was a specific disease, which killed about a third of the people whom it infected. What they disagreed about was the cause, the pathology, and the communication of the disease. Similarly, most medical professionals these days realize that psychological illnesses occur..., but they disagree about their cause and pathology. Notwithstanding [recent official classification schemes for mental disorders], we doubt whether any satisfactory a priori definition of psychological illness can exist... because it is a matter for theory to elucidate.

At any given time, about one percent of a human population will experience debilitating schizophrenoform disorders, and perhaps another one percent crippling levels of analogs to Western depression/anxiety. A smaller proportion will express dangerous degrees of some forms of psychopathy or sociopathy. Indeed, most of us, of whatever cultural background, over the life course will suffer one or more periods of mental distress or disorder in which we simply cannot properly carry out socially expected duties. A billion years of evolutionary selection has not eliminated mental illness any more than it has eliminated infectious disease.

The Brave New World of highly parallel autonomic and Self-X machines that will follow the current explosion of multiple core and molecular computing technologies will not be spared analogous patterns of canonical and idiosyncratic failure. These, we have argued, will resemble developmental disorders of gene expression often caused by exposure to teratogenic chemicals, by social exposures including psychosocial stress, or their synergisms. That is, they can be induced. Technological momentum, driven by industrial economic interests, will see such machines widely deployed to control a great swath of critical systems in real time, ranging from transportation, power, and communication networks, to individual vehicles, various industrial enterprises, and the like.

There is a recent, closely studied, case history of serious, indeed fatal, failure of just such a real-time system. Hawley (2006, 2008) describes how, during combat operations of the second Gulf War, US Army Patriot antimissile units were involved in two fratricide incidents involving coalition aircraft. A British Tornado was misclassified as an anti-radiation missile, and a US Navy F/A-18 misclassified as a tactical ballistic missile. Both were shot down, and their crews lost. Two of 11 US Patriot shots were fratricides, some 18 percent. Detailed analysis found a 'perfect storm' that, among other things, involved a synergism of a fascination with, and blind faith in, technology with unacknowledged system fallibilities, reinforced by an organizational culture of 'React quickly, engage early, and trust the system without question'.

Under the stress of actual combat, the distortion of the fog of war in the real world, the Patriot system appears to suffer a condensation to a kind of absolute zero, a grossly simplified behavioral state, unable to properly differentiate friend from foe. More complicated military systems currently under consideration will likely display more convoluted but equally debilitating failure patterns.

Failure modes of existing large scale, massively parallel systems, like power blackouts, cloud computing outages, and internet viruses, will, as multicore technoloy relentlessly advances, become failure modes of physically small devices having similar numbers of components. These will have been tasked with controlling critical real time operations across a broad swath of civilian enterprise, consequent on fascination with, and blind faith in, technology with unacknowledged system fallibilities, reinforced by organizational culture. Such devices will, in addition, be subject to induced failure by unfriendly external agents. *Caveat Emptor* indeed.

5 Appendix: Cognitive systems in real time

Following closely the arguments of Wallace (2010), real time problems are inherently rate distortion problems: The implementation of a complex cognitive structure, say a sequence of control orders generated by some dual information source Y, having output $y^n = y_1, y_2, ...$ is 'digitized' in terms of the observed behavior of the regulated system, say the sequence $b^n = b_1, b_2, ...$ The b_i are thus what happens in real time, the actual impact of the cognitive structure on its embedding environment. Assume each b^n is then deterministically retranslated back into a reproduction of the original control signal,

$$b^n \to \hat{y}^n = \hat{y}_1, \hat{y}_2, ...$$

Define a distortion measure $d(y, \hat{y})$ that compares the original to the retranslated path. 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

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

The average distortion is defined as

$$D \equiv \sum_{y^n} p(y^n) d(y^n, \hat{y}^n). \qquad (10)$$

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

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

If there is no uncertainty in Y given the retranslation $\hat{Y}$, then no information is lost, and the regulated system is perfectly under control.

In general, this will not be true.

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

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

(11)

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 transmission does not exceed average distortion* D. Thus $R(D)$ defines a minimum necessary channel capacity. Cover and Thomas (1991) provide details. The rate distortion function has been calculated for a number of systems. Cover and Thomas (1991, Lemma 13.4.1) show that $R(D)$ is necessarily a decreasing convex function of D, that is, always a reverse J-shaped curve.

Recall, now, the relation between information source uncertainty and channel capacity (e.g., Ash, 1990; Cover and Thomas, 1991):

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

(12)

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

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

(13)

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

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

The disjunction between intent and impact of a cognitive system interacting with an embedding environment can be modeled using a simple extension of the language-of-cognition approach. Recall that cognitive processes can be formally associated with information sources, and how a formal equivalence class algebra can be constructed for a complicated cognitive system by choosing different origin points in a particular abstract 'space' and defining the equivalence of two states by the existence of a high probability meaningful path connecting each of them to some defined origin point within that space. Disjoint partition by equivalence class is analogous to orbit equivalence relations for dynamical systems, and defines the vertices of a network of cognitive dual languages available to the system: Each vertex represents a different information source dual to a cognitive process. The structure creates a large groupoid, with each orbit corresponding to a transitive groupoid whose disjoint union is the full groupoid, and each subgroupoid associated with its own dual information source. Larger groupoids will, in general, have 'richer' dual information sources than smaller. We can apply the spontaneous symmetry breaking argument to increasing disjunction between cognitive intent and system impact as follows:

With each subgroupoid G_i of the (large) cognitive groupoid we can associate a dual information source H_{G_i}. Let $R(D)$ be the rate distortion function between the message sent by the cognitive process and the observed impact. Remember that both H_{G_i} and $R(D)$ are free energy density measures.

The essential argument is that $R(D)$ is an embedding context for the underlying cognitive process. The argument-by-abduction from physical theory is, then, that $R(D)$ constitutes a kind of thermal bath for the processes of cognition. Thus we can write the probability of the dual cognitive information source H_{G_i} as

$$P[H_{G_i}] = \frac{\exp[-H_{G_i}/\kappa R(D)]}{\sum_j \exp[-H_{G_j}/\kappa R(D)]},$$

(14)

where κ is an appropriate dimensionless constant characteristic of the particular system. The sum is over all possible subgroupiods of the largest available symmetry groupoid. Again, compound sources, formed by the union of underlying transitive groupoids, being more complex, will have higher free-energy-density equivalents than those of the base transitive groupoids. This follows directly from the argument of Ash (1990) quoted above.

We can apply the Groupoid Free Energy phase transition arguments from above, remembering that the Rate Distortion Function $R(D)$ is always a decreasing convex function of D

(Cover and Thomas, 1991). For real time cognitive systems, increasing average distortion between cognitive intent and observed impact will 'lower the temperature' so as to drive the cognitive process, possibly in a highly punctuated manner, relentlessly toward simpler and less rich behaviors.

One important question is what constitutes 'real time' for the system of interest. If 'real time' has a characteristic time constant τ in which response takes place, then the temperature analog $\kappa R(D)$ might be better represented by the product $\alpha R(D)\tau$, i.e., energy rate $\times$ time. Given sufficient time to think matters over, as it were, the cognitive system might well converge to an acceptable solution. The canonical example being the increased time taken by autofocus cameras to converge under very low light conditions. The Morse Theory arguments of Wallace (2010), as extended by the Siefert-Van Kampen Theorem, then suggest a possibly complicated topology underpins system dynamics since R and τ (and other possible factors represented by α) then become eigenvalues of an operator on a highly abstract base space.

6 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.

Atmanspacher, H., 2006, Toward an information theoretical implementation of contextual conditions for consciousness, *Acta Biotheoretica*, 54:157-160.

Bak, A., R. Brown, G. Minian, and T. Porter, 2006, Global actions, groupoid atlases and related topics, *Journal of Homotopy and Related Structures*, 1:1-54.

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.

Bos, R., 2007, Continuous representations of groupoids, arXiv:math/0612639.

Brown, R., 1987, From groups to groupoids: a brief survey, *Bulletin of the London Mathematical Society*, 19:113-134.

Buneci, M., 2003, *Representare de Groupoizi*, Editura Mirton, Timisoara.

Burago D., Y. Burago, and S. Ivanov, 2001, *A Course in Metric Geometry*, Graduate Studies in Mathematics 33, American Mathematical Society.

Cannas Da Silva, A., and A. Weinstein, 1999, *Geometric Models for Noncommutative Algebras*, American Mathematical Society, RI.

Ciliberti, S., O. Martin, and A. Wagner, 2007a, Robustness can evolve gradually in complex regulatory networks with varying topology, *PLoS Computational Biology*, 3(2):e15.

Ciliberti, S., O. Martin, and A. Wagner, 2007b, Innovation and robustness in complex regulatory gene networks, *Proceeding of the National Academy of Sciences*, 104:13591-13596.

Cohen, I., 2006, Immune system computation and the immunological homunculus. In Nierstrasz, O., J. Whittle, D. Harel, and G. Reggio (eds.), MoDELS 2006, LNCS, vol. 4199, pp. 499-512, Springer, Heidelberg.

Cohen, I., and D. Harel, 2007, Explaining a complex living system: dynamics, multi-scaling, and emergence. *Journal of the Royal Society: Interface*, 4:175-182.

Cover, T., and J. Thomas, 1991, *Elements of Information Theory*, John Wiley and Sons, New York.

Dehaene, S., and L. Naccache, 2001, Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework, *Cognition*, 79:1-37.

Dembo, A., and O. Zeitouni, 1998, *Large Deviations: Techniques and Applications*, 2nd edition, Springer, New York.

Dretske, F., 1994, The explanatory role of information, *Philosophical Transactions of the Royal Society, A*, 349:59-70.

Emery, M., 1989, *Stochastic Calculus on Manifolds*, Springer, New York.

Feynman, R., 2000, *Lectures on Computation*, Westview Press, New York.

Frankel, T., 2006, The Geometry of Physics: An Introduction, Second Edition, Cambridge University Press.

Gilbert, S., 2001, Mechanisms for the environmental regulation of gene expression: ecological aspects of animal development, *Journal of Bioscience*, 30:65-74.

Glazebrook, J.F., and R. Wallace, 2009a, Small worlds and red queens in the global workspace: an information-theoretic approach, *Cognitive Systems Reserch*, 10:333-365.

Glazebrook, J.F., and R. Wallace, 2009b, Rate distortion manifolds as model spaces for cognitive information. In press, *Informatica* Overview Article.

Golubitsky, M., and I. Stewart, 2006, Nonlinear dynamics and networks: the groupoid formalism, *Bulletin of the American Mathematical Society*, 43:305-364.

Gunderson, L., 2000, Ecological resilience – in theory and application, *Annual Reviews of Ecological Systematics*, 31:425-439.

Hawley, J., 2006, Patriot Fratricides: The human dimension lessions of Operation Iraqui Freedom, *Field Artillery*, January-February.

Hawley, J., 2008, The Patriot Vigilance Project: A case study of Patriot fratricide mitigations after the Second Gulf War, Third System of Systems Conference, December 10, 2008.

Helgason, S., 1962, *Differential Geometry and Symmetric Spaces*, Academic Press, New York.

Holling, C., 1973, Resilience and stability of ecological systems, *Annual Reviews of Ecological Systematics*, 4:1-23.

Johnson-Laird, P., F. Mancini, and A. Gangemi, 2006, A hyper-emotion theory of psychological illness, *Psychological Review*, 113:822-841,

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.

Maturana, H., and F. Varela, 1980, *Autopoiesis and Cognition*, Reidel Publishing Company, Dordrecht, Holland.

Maturana, H., and F. Varela, 1992, *The Tree of Knowledge*, Shambhala Publications, Boston, MA.

Pettini, M., 2007, *Geometry and Topology in Hamiltonian Dynamics and Statistical Mechanics*, Springer, New York.

Sarlio-Lahteenkorva, S., E. Lahelma, 2001, Food insecurity is associated with past and present economic disadvantage and body mass index, *Journal of Nutrition*, 131:2880-2884.

Scherrer, K., and J. Jost, 2007a, The gene and the genon concept: a functional and information-theoretic analysis, *Molecular Systems Biology* 3:87-93.

Scherrer, K., and J. Jost, 2007b, Gene and genon concept: coding versus regulation, *Theory in Bioscience* 126:65-113.

Skierski, M., A. Grundland, and J. Tuszynski, 1989, Analysis of the three-dimensional time-dependent Landau-Ginzburg equation and its solutions, *Journal of Physics A* (Math. Gen.), 22:3789-3808.

Toulouse, G., S. Dehaene, and J. Changeux, 1986, Spin glass model of learning by selection, *Proceedings of the National Academy of Sciences*, 83:1695-1698.

Wallace, R., 2005, *Consciousness: A Mathematical Treatment of the Global Neuronal Workspace Model*, Springer, New York.

Wallace, R., 2006, Pitfalls in biological computing: canonical and idiosyncratic dysfunction of conscious machines, *Mind and Matter*, 4:91-113.

Wallace, R., 2007, Culture and inattentional blindness, *Journal of Theoretical Biology*, 245:378-390.

Wallace, R., 2008a, Toward formal models of biologically inspired, highly parallel machine cognition, *International Journal of Parallel, Emergent and Distributed Systems*, 23:367-408.

Wallace, R., 2008b, Developmental disorders as pathological resilience domains, *Ecology and Society*, 13:29 (online).

Wallace, R., 2009, Programming coevolutionary machines: the emerging conundrum. In press, *International Journal of Parallel, Emergent and Distributed Systems*

Wallace, R., 2010, Tunable epigenetic catalysis: programming real-time cognitive machines. In press, *International Journal of Parallel, Emergent and Distributed Systems*

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 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.

Wallace, R., and D. Wallace, 2010, *Gene Expression and its Discontents: The social production of chronic disease*, Springer, New York.

Wallace R.G., and R. Wallace, 2009, Evolutionary radiation and the spectrum of consciousness, *Consciousness and Cognition*, 18:160-167.

Weinstein, A., 1996, Groupoids: unifying internal and external symmetry, *Notices of the American Mathematical Association*, 43:744-752.

West-Eberhard, M., 2005, Developmental plasticity and the origin of species differences, *Proceedings of the National Academy of Sciences*, 102:6543-6549.