

Version 10

# Epigenetic catalysis in an evolutionary context

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

January 2, 2009

## Abstract

Recent work by Ciliberti et al. finds the spinglass model of regulatory gene networks adapted from neural network studies to have a single giant connected component in a metanetwork space of interaction matrices, permitting only gradual evolutionary transition, in conflict with empirical studies of development in sea urchin species that found evidence for punctuated equilibrium evolutionary transition. Shifting perspective from the highly parallel matrix space to the grammar/syntax of the time series of expressed phenotypes via a recently introduced cognitive paradigm permits import of other techniques from statistical physics to the study of gene expression, in particular application of Landau's spontaneous symmetry breaking arguments. This produces a straightforward multi-component model generating punctuated equilibrium in the evolution of development. Analogs to Bennett and Hacker's mereological fallacy and to Krebs' sufficiency failure that haunt neural network models of high order cognition also severely constrain the usefulness of spinglass models in the study of gene expression dynamics. Our methods, by contrast, permit incorporation of epigenetic effects in a highly natural manner, finding them analogous to a tunable enzyme catalyst in a process also subject to phase transition analogs.

**Key Words** cognitive paradigm, development, evolution, groupoid, information theory, mereological fallacy, phase transition, statistical mechanics

## 1 Introduction

One of the more remarkable observations in contemporary biology is that, while gene sequence and organization are often highly conserved in evolutionary process, the regulatory machinery controlling their expression may not be (e.g., Chouard, 2008; Britten and Davidson, 1969; Waddington, 1942, 1957). Just as strikingly, conserved function does not necessarily imply conserved genetics. Chouard puts it thus:

A given function can result from diverse combinations of the same genes, or different genes, even in

very closely related species... there are many combinations of contributing factors that can reach the same outcome... one of evo-devo's key questions is how the genes and their interactions can change while keeping the output of the system as good as, or better than, it was before.

A recent computational exercise by Andreas Wagner and colleagues (Ciliberti et al., 2007a, b), based on the classic spinglass analysis of Reinitz adapted from neural network studies (e.g., Mjolsness et al., 1991; Toulouse et al., 1986), appears to prove that only gradual evolutionary transitions are possible in development. Recent empirical observations by Raff et al. (2003) contradict that result. They examined hybrids of related developing sea urchin species having essentially identical adult phenotypes, finding

The concept of punctuated equilibrium has played a major role in palentological and evolutionary thought... This hypothesis proposes that species may endure for geologically long intervals of time without substantial change but can undergo speciation and morphological change in geologically short time intervals... It has been predicted that developmental processes and gene regulatory systems might change rapidly during periods of speciation and accompanying morphological change... [The sea urchin] *Heliocidaris tuberculata* is an indirect developer and produces a typical feeding pluteus larva. *Heliocidaris erythrogramma* is a direct developer that develops into a highly modified nonfeeding larva that generates the juvenile adult sea urchin in a matter of days. Molecular phylogenetic studies show that the two... species are separated by as little as 4 million years... For the remarkable change observed in developmental pathways during this short divergence time, a rapid evolution of the regulatory mechanisms that guide development must have occurred. However, the punctuated equilibrium model predicts that regulatory evolution would not only have been rapid in *H. eryth.*, but regulatory evolution in *H. tuber.* and other indirect-developers should have been slow, with major control of pluteus development remaining similar over large spans of time... These

\*Address correspondence to: Rodrick Wallace, 549 W. 123 St., Suite 16F, New York, NY, 10027 USA, 212-865-4766, wallace@pi.cpmc.columbia.edu.

hypotheses formed the basis for the test crosses we made between *Heliocidaris* species and *P. maculata*. Our results are consistent with deep conservation of indirect development and with the punctuated equilibrium model in the switch to direct development in *H. eryth...*

Here we will briefly review the work of Ciliberti et al., explore some of its less obvious topological implications – in particular the mapping of disjoint directed homotopy classes of phenotype paths into interaction matrix space – and then extend the approach by applying a cognitive paradigm for gene expression developed by Wallace and Wallace (2008). Analogs to phase transition arguments in physical systems generate punctuated equilibrium evolutionary transitions in a ‘highly natural’ manner, even for the spinglass treatment, and a hierarchical extension permits incorporation of epigenetic effects as a kind of tunable catalysis.

## 2 The spinglass model

Following closely Ciliberti et al. (2007a, b), the spinglass model assumes that  $N$  transcriptional regulators, are represented by their expression patterns

$$\mathbf{S}(t) = [S_1(t), \dots, S_N(t)]$$

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

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

(1)

where  $\tau$  is a constant and  $\sigma$  a sigmoidal function whose value lies in the interval  $(-1, 1)$ . In the spinglass limit  $\sigma$  is the sign function, taking only the values  $\pm 1$ .

This approach is formally similar to spinglass neural network models of learning by selection, e.g., as proposed by Toulouse et al. (1986). Subsequent work by Baars (1988; 2005) on consciousness in higher animals suggests that such models are simply not sufficient to the task, a matter we will explore further below.

The networks of interest in this model 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 a network is termed *viable*, for obvious reasons.

After an elaborate and very difficult simulation exercise, a particular series of results emerges (Ciliberti et al., 2007a):

...[V]iable networks comprise a tiny fraction of possible ones. They could be widely scattered in the space of all possible networks and occupy disconnected islands in this space. However, our analysis indicates precisely the opposite. The metagraph of viable networks has one ‘giant’ connected component that comprises most or all viable networks. Any two networks in this component can be reached from one another through gradual changes of one regulatory interaction at a time, changes that never leave the space of viable networks.

In general, within the giant component, randomly chosen pairs of networks with the same phenotype will have vastly different organization, in terms of the matrix  $(w_{ij})$ .

Define  $d$  as the the Hamming distance between two  $\mathbf{S}_0$  and  $\mathbf{S}_\infty$ . A typical result is that for  $N = 5$  genes,  $6 \leq M \leq 7$  total regulatory interactions, and  $d = 0.4$ , full enumeration finds a total of only 37,338 viable networks out of  $6.3 \times 10^7$  possible ones (Ciliberti et al., 2007a). Long random walks through the space of viable networks, however, visit all but a very small fraction of the nodes of the metagraph, and this missing fraction decreases as  $N$  increases. Large  $N$  require elaborate Monte Carlo sampling for simulation, a difficult and computationally intensive enterprise.

In  $w$ -space Ciliberti et al. define a metric characterizing the distance between two network topologies as

$$D(w, w') = \frac{1}{2M_+} \sum_{i,j} |\text{sign}(w_{ij}) - \text{sign}(w'_{ij})|,$$

where  $M_+$  is the maximum number of regulatory interactions, and  $\text{sign}(x) = \pm 1$  depends on the sign of  $x$ , and is 0 for  $x = 0$ .

Several observations emerge directly.

[1] What Ciliberti et al. (2007a, b) observe, in another idiom, is that in phenotype space, in  $\mathbf{S}$ -space, characterized by the Hamming distance  $d$ , the set of all paths associated with viable networks forms an equivalence class, closely analogous to the directed homotopy equivalence classes in the sense of Goubault (2003) and Goubault and Raussen (2002). Directed homotopy differs from simple homotopy (e.g., Lee, 2000) in that one uses paths from one point to another rather than loops, and seeks continuous deformations between them. See Wallace and Wallace (2008) for discussion in a biological context. Thus there is, in this spinglass model, a mapping from  $\mathbf{S}$ -space into  $(w_{ij})$  space, characterized by the metric  $D$ , that associates a unique simply connected component with each dihomotopy-like equivalence class of paths connecting

two particular phenotype points. Indeed, the  $w$ -space component can clearly be treated according to standard homotopy arguments, i.e., using loops.

[2] What one does with homotopically simply connected components is patch them together build larger, and more interesting, topological structures, using the Seifert-Van Kampen Theorem (SVKT) (e.g., Lee, 2000, Ch. 10): If paths within  $S$ -space are not continuously transformable into one another, using the d-metric, (if there are ‘holes’), then several distinct dihomotopy classes will exist, e.g., as in figures 1 and 2 of Wallace and Wallace (2008). The obvious conjecture is that, under such a circumstance, very complex topological objects may lurk in  $w$ -space, not just the simply connected component discovered by Ciliberti et al. These may, according to the SVKT, intersect as well as exist as isolated and disconnected sets.

In particular, if there are dihomotopy ‘holes’ in  $S$ -space, consequently reflected in disconnected patches in  $w$ -space, then punctuated transition events of various sorts may well become an evolutionary norm, as in Gould (2002), even for the spinglass model.

[3] A large and increasing body of work surrounding coupled cell networks invokes groupoids, a natural generalization of symmetry groups. As Dias and Stewart (2004) comment,

Until recently the abstract theory of coupled cell systems has mainly focused on the effects of symmetry in the network... and the consequent formation of spatial and spatiotemporal patterns. The formal setting for this theory centers upon the symmetry group of the network.

The analysis of robust patterns of synchrony in general coupled cell systems – that is, dynamics in which sets of cells behave identically as a consequence of the network topology – has led to the fruitful notion of the ‘symmetry groupoid’ of a coupled cell network... A groupoid is a generalization of a group, in which products of elements are not always defined... The symmetry groupoid of a coupled cell network is a natural algebraic formalization of the ‘local symmetries’ that relate subsets of the network to each other. In particular ‘admissible’ vector fields – those specified by the network topology – are precisely those that are equivariant under the action of the symmetry groupoid...

The Appendix provides a summary of standard material on groupoids that will be of later use.

[4] Both of these – analogous – approaches can apparently be coarse-grained into a symbolic dynamics associated with (simple) information sources having particular grammar and syntax. The method is straightforward (e.g., Beck and Schlogl, 1995; McCauley, 1994). One could, thus, probably translate the spinglass results of Ciliberti et al. into symbolic dynamics, using groupoid methods to study the underlying topological objects.

[5] The model used by Ciliberti et al. is abstracted from spinglass treatments of neural networks, as is made clear in

the seminal papers by the Reinitz group (e.g., Mjolsness et al. 1991; Reinitz and Sharp, 1995; Sharp and Reinitz, 1998; Jaeger et al., 2004). Thus and consequently, Ciliberti et al. are invoking an implicit cognitive paradigm for gene expression (e.g., Cohen, 2006; Cohen and Harel, 2007; Wallace and Wallace, 2008), and cognitive process, as the philosopher Fred Dretske eloquently argues (e.g., Dretske, 1994), is constrained by the necessary conditions imposed by the asymptotic limit theorems of information theory. A little work produces a very general cognitive gene expression metanetwork structure recognizably similar to that found by Ciliberti et al. and the Reinitz group, but one capable of highly punctuated dynamics. The massively parallel computations are hidden, somewhat, in the required empirical fitting of regression model analogs based on the asymptotic limit theorems of information theory rather than on the central limit theorem.

[6] A salient characteristic of high level cognitive process is precisely its inherent punctuation (e.g., Baars, 1988, 2005; Wallace, 2005), and this easily emerges using an information theory approach via the famous homology between information and free energy (e.g., Feynman, 2000). ‘Simple’ neural network analogs will inevitably have more difficulty replicating such behavior, but as discussed, the mapping of disjoint dihomotopy equivalence classes from phenotype sequence space to disjoint sets in interaction matrix space provides a straightforward example for spinglass models.

The next sections use information theory methods to make the transition from crosssectional  $w$ -space into that of serially correlated sequences of phenotypes, expanding on the results of Wallace and Wallace (2008).

### 3 Cognition as an information source

Atlan and Cohen (1998) argue, in the context of immune cognition, that the essence of cognitive function involves comparison of a perceived signal with an internal, learned picture of the world, and then, upon comparison, choosing a response from a much larger repertoire of possible responses.

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

Cognitive pattern recognition-and-response proceeds by an algorithmic combination of an incoming ‘external sensory signal’ with an ‘internal ongoing activity’ – incorporating the learned picture of the world. The combination triggers an appropriate action on the condition that the pattern of sensory activity requires a response.

More formally, a pattern of incoming input – like the  $\mathbf{S}(t)$  of equation (1) – is mixed in a systematic algorithmic manner with a pattern of internal ongoing activity – like the  $(w_{ij})$  according to equation (1) – to create a path of combined signals  $x = (a_0, a_1, \dots, a_n, \dots)$  – analogous to the sequence of  $\mathbf{S}(t + \tau)$  of equation (1), with, say,  $n = t/\tau$ . Each  $a_k$  thus represents some functional composition of internal and external signals.

This path is fed into a highly nonlinear decision oscillator,

$h$ , which 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

$$B_0 \equiv b_0, \dots, b_k,$$

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

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

The principal objects of formal interest are paths  $x$  which trigger 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  $h(a_0, \dots, a_j) \in B_0$  for all  $0 < j < m$ , but  $h(a_0, \dots, a_m) \in B_1$ .

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 which permits inference of the necessary conditions constrained by the asymptotic limit theorems of information theory is that the finite limit

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

(2)

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  $P(a_0, \dots, a_n)$  and  $P(a_n|a_0, \dots, a_{n-1})$  such that appropriate

conditional and joint Shannon uncertainties satisfy the classic relations

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

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

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

(3)

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).

Wallace (2005, pp. 34-36) applies this formalism to a standard neural network model.

In the developmental vernacular of Ciliberti et al., we now examine paths in phenotype space that begins at some  $\mathbf{S}_0$  and converges  $n = t/\tau \rightarrow \infty$  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  $x = (\mathbf{S}_0, \mathbf{S}_\tau, \dots, \mathbf{S}_{n\tau}, \dots)$  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.

## 4 Consequences of the perspective change

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 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  $H[X] = \lim_{n \rightarrow \infty} \log[N(n)]/n$ , that can be expressed in terms of joint and conditional probabilities as in equation (3) above. 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, information source uncertainty has an important heuristic interpretation. Ash (1990) puts it this way:

...[W]e 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.

This will prove important below.

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)

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

(4)

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

Feynman (2000), following the classic work by Bennett (1988), concludes that the information contained in a message is simply 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  $H = \lim_{n \rightarrow \infty} \log[N(n)]/n$ . Ash’s comment above then has a corollary: If, for a biological system,  $H_1 > H_2$ , source 1 will require more metabolic free energy than source 2.

## 5 Symmetry arguments

A formal equivalence class algebra, in the sense of the Appendix, 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*, in the sense of Ciliberti et al. 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). 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 – parametrizes the set of equivalence relations and creates the (very large) groupoid.

There is a hierarchy of structure here. 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 know what that structure may be, but we will 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.

### 5.1 The first level

The spinglass model of Ciliberti et al. produced a simply connected, but otherwise undifferentiated, metanetwork of gene expression dynamics that could be traversed continuously by single-gene transitions in the highly parallel  $w$ -space. 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 actually 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 an available metabolic energy density available at a rate  $M$ , and an allowed development time  $T$ , let  $K = 1/\kappa MT$  for some appropriate scaling constant  $\kappa$ , so that  $MT$  is total developmental free energy. Then the probability of a particular  $H_\alpha$  will be determined by the standard relation (e.g., Landau and Lifshitz, 2007),

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

(5)

where the sum may, in fact, be a complicated abstract integral. The basic requirement is that the sum/integral always converges.  $K$  is the inverse product of a scaling factor, a metabolic energy density rate term, and a characteristic development time. The developmental energy might be raised to some power, e.g.,  $K = 1/(\kappa(MT)^b)$ , suggesting the possibility of allometric scaling.

Thus, in this formulation, there must be structure *within* a (cross sectional) connected component in the  $w$ -space of Ciliberti et al., determined in no small measure by available energy. Some dual information sources will be ‘richer’/smarter than others, but, conversely, must use more metabolic energy for their completion. Evidently this would set up a particular tension under selection pressures.

The next generalization is crucial:

While we might simply impose an equivalence class structure based on equal levels of energy/source uncertainty, producing a groupoid in the sense of the Appendix (and possibly allowing a Morse Theory approach in the sense of Matsumoto, 2002 or Pettini, 2007), 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.

## 5.2 The second level

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, so that  $K = 1/\kappa MT$ , we propose that the metabolic-energy-constrained probability of a reproductive information source representing equivalence class  $D_i$ ,  $H_{D_i}$ , will again be given by the classic relation

$$P[H_{D_i}] = \frac{\exp[-H_{D_i} K]}{\sum_j \exp[-H_{D_j} K]},$$

(6)

where the sum/integral 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

$$Z_D \equiv \sum_j \exp[-H_{D_j} K].$$

(7)

We now define the *Groupoid free energy* of the system,  $F_D$ , at inverse normalized metabolic energy density  $K$ , as

$$F_D[K] \equiv -\frac{1}{K} \log[Z_D[K]],$$

(8)

following the standard arguments from statistical physics (again, Landau and Lifshitz, 2007, or Feynman, 2000).

The groupoid free energy construct permits introduction of deep results from statistical physics.

## 5.3 Spontaneous symmetry breaking

We have expressed the probability of a reproductive information source in terms of its relation to a fixed, scaled, available (inverse) metabolic free energy density, seen as a kind of equivalent (inverse) system temperature. This gives a statistical thermodynamic path leading to definition of a ‘higher’

free energy construct  $-F_D[K]$  – 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.

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 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 requires invocation of groupoid representations, no small matter (e.g., Buneci, 2003; Bos 2006).

Somewhat more rigorously, the biological renormalization schemes of the Appendix to Wallace and Wallace (2008) may now be imposed on  $F_D[K]$  itself, leading to a spectrum of highly punctuated transitions in the overall system of developmental information sources.

Most deeply, however, an extended version of 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  $F_D[K]$ . Necessity, but not sufficiency, is important, as it, in theory, allows mixed groupoid symmetries.

The essential insight is that the single simply connected giant component of Ciliberti et al. is unlikely to be the full story, and that more complete models will likely be plagued – or graced – by highly punctuated evolutionary dynamics.

## 5.4 Tunable epigenetic catalysis

Incorporating the influence of embedding contexts – epigenetic effects – is most elegantly done by invoking the Joint Asymptotic Equipartition Theorem (JAEPT) and the extensions of Network Information Theory in equations (6-8) (Cover and Thomas, 1991). For example, given an embedding contextual information source, say  $Z$ , that affects development, then the dual cognitive source uncertainty  $H_{D_i}$  is replaced by a joint uncertainty  $H(X_{D_i}, Z)$ . 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, while there are approx-

imately  $\exp[nH(X)]$  typical  $X$  sequences, and  $\exp[nH(Z)]$  typical  $Z$  sequences, there are only about  $\exp[nH(X, Z)]$  jointly typical sequences, so that the effect of the embedding context, in this model, is to greatly lower the ‘developmental free energy’ at a given metabolic energy  $MT$ . Thus, for a given  $MT$ , the effect of epigenetic regulation is to make possible developmental pathways otherwise inhibited by their high values of uncertainty/free energy. 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 distinguishes our approach, as it permits hierarchical models similar to those of higher order cognitive neural function that incorporate Baars’ contexts in a natural way (e.g., Wallace and Wallace, 2008; Wallace and Fullilove, 2008).

Another possible extension is to allow multiple, simultaneous, gene network dynamics, so that a single gene may serve more than one purpose. This is similar to the multiple workspace institutional distributed cognition models of Wallace and Fullilove (2008). There, an individual actor can participate in multiple, ongoing simultaneous (or rapidly sequential) ‘projects’. A principal complication then becomes the necessity of crosstalk (i.e., coordination) between those projects, a matter governed by the rate distortion theorem and its possible pathologies.

Parenthetically, cognitive gene expression must likely conserve resources by temporal or process focus, leading to the potential for the inattentive blindness that afflicts most cognitive process (e.g., Wallace, 2007).

These elaborations allow 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. West-Eberhard (2005) puts the matter as follows:

Any new input, whether it comes from the genome, like a mutation, or from the external environment, like a temperature change, a pathogen, or a parental opinion, has a developmental effect only if the preexisting phenotype is responsive to it... A new input... causes a reorganization of the phenotype, or ‘developmental recombination.’...In developmental recombination, phenotypic traits are expressed in new or distinctive combinations during ontogeny, or undergo correlated quantitative change in dimensions...Developmental recombination can result in evolutionary divergence... at all levels of organization.

Individual development can be visualized as a series of branching pathways. Each branch point is a developmental decision, or switch point, governed by some regulatory apparatus, and each switch point defines a modular trait. Developmental recombination implies the origin or deletion of a branch and a new or lost modular trait. It is important to realize that the novel regulatory response and the novel trait originate simultaneously. Their origins are, in

fact, inseparable events: you cannot have a change in the phenotype, a novel phenotypic state, without an altered developmental pathway...

This is accomplished in our formulation by allowing the set  $B_1$  in section 3 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  $H(X_{D_i}, Z)$ , so that the epigenetic information source  $Z$  serves to direct as well the possible final states of  $X_{D_i}$ .

The mechanics of such channeling can be made more precise as follows.

Real time problems, like the crosstalk between epigenetic and genetic structures, are inherently rate distortion problems, and the interaction between biological structures can be restated in communication theory terms. Suppose a sequence of signals is generated by a biological information source  $Y$  having output  $y^n = y_1, y_2, \dots$ . This is ‘digitized’ in terms of the observed behavior of the system 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,

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

Here the information source  $Y$  is the epigenetic  $Z$ , and  $B$  is  $X_{D_i}$ , but the terminology used here is more standard (e.g., Cover and Thomas, 1991).

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

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

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

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

(9)

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:

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

(10)

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

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

(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 communication between the biological vesicles 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 explicitly calculated for a number of simple systems.

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

$$H[\mathbf{X}] \leq C, \quad (12)$$

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

$$C \equiv \max_{P(X)} I(X|Y). \quad (13)$$

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

Finally, recall the analogous definition of the rate distortion function above, again an extremum over a probability distribution.

Recall, again, equations (4-8), i.e., that the free energy of a physical system at a normalized inverse temperature-analog  $K = 1/\kappa T$  is defined as  $F(K) = -\frac{1}{K} \log[Z(K)]$  where  $Z(K)$  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  $K$ , the probability of a state  $E_i$  is determined by the relation  $P[E_i] = \exp[-E_i K] / \sum_j \exp[-E_j K]$ .

The partition function is simply the normalizing factor.

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 free metabolic energy,  $K = 1/\kappa MT$ .

The key is to take the  $R(D)$  as representing energy as a function of the average distortion. Assume a fixed  $K$ , so that the probability density function of an average distortion  $D$ , given a fixed  $K$ , is then

$$P[D, K] = \frac{\exp[-R(D)K]}{\int_{D_{min}}^{D_{max}} \exp[-R(D)K] dD}. \quad (14)$$

Thus lowering  $K$  in this model rapidly raises the possibility of low distortion communication between linked systems.

We define the *rate distortion partition function* as just the normalizing factor in this equation:

$$Z_R[K] \equiv \int_{D_{min}}^{D_{max}} \exp[-R(D)K] dD, \quad (15)$$

again taking  $K = 1/\kappa MT$ .

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

$$F_R[K] \equiv -\frac{1}{K} \log[Z_R[K]], \quad (16)$$

and apply Landau's spontaneous symmetry breaking argument to generate punctuated changes in the linkage between the genetic information source  $X_{D_i}$  and the embedding epigenetic information source  $Z$ . Recall that Landau's insight was that certain phase transitions were usually in the context of a significant symmetry change in the physical states of a system.

Again, the biological 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 biological substructures.

Since  $1/K$  is proportional to the embedding metabolic free energy, we assert that

[1] the greatest possible set of symmetries will be realized for high developmental free metabolic energies, and

[2] phase transitions, related to total available developmental metabolic free energy, will be accompanied by fundamental changes in the final topology of the system of interest – phenotype changes – recognizing that evolutionary selection acts on phenotypes, not genotypes.

The relation  $1/K \propto MT$  suggests the possibility of evolutionary tradeoffs between development time and the rate of available metabolic free energy.

## 6 Discussion and conclusions

We have hidden the massively parallel calculations made explicit in the work of Ciliberti et al. and the Reinitz group, burying them as 'fitting regression-model analogs to data', possibly at a second order epigenetic hierarchical level. In the real world such calculations would be quite difficult, particularly given the introduction of punctuated transitions that must be fitted using elaborate renormalization calculations, typically requiring such exotic objects as Lambert W-functions (e.g., Wallace, 2005). From one perspective we are

providing a different formalism for this work in the sense that matrix mechanics and wave mechanics provide alternate pictures of quantum mechanics. The approach here lends itself easily to analysis of punctuated dynamics, although, as argued, the spinglass model itself appears to be discretized by disjoint directed homotopy equivalence classes. Toulouse et al. (1986) provide another standard perspective on multiple quasistability in spinglass models using classic ‘frustration’ arguments.

It is important to recognize that the developmental strategies represented by the  $H(X_{D_j})$  or the  $H(X_{D_i}, Z)$  are not merely passive actors. Quite the contrary, they characterize full-scale Darwinian individuals subject to variation, selection, and chance extirpation, forces acting at both individual and species levels (Gould, 2002). Thus groupoid symmetries and available metabolic free energy are, as a consequence of the Darwinian individuality of reproduction, contexts for, rather than determinants of, evolutionary process, including punctuated equilibrium. They are the banks between which the evolutionary glacier flows – sometimes slowly, and sometimes in sudden advance.

Analogies with neural network studies suggest intractable conceptual difficulties for spinglass-type models of gene expression and development dynamics. In spite of nearly a century of sophisticated neural network model studies – including elegant treatments like Toulouse et al. (1986) – Atmanspacher (2006) felt compelled to state that

To formulate a serious, clear-cut and transparent formal framework for cognitive neuroscience is a challenge comparable to the early stage of physics four centuries ago. Only very few approaches worth mentioning are visible in contemporary literature.

Furthermore, Krebs (2005) has identified what might well be described as the sufficiency failing of neural network models, that is, neural networks can be constructed as Turing machines that can replicate any known dynamic behavior in the same sense that the Ptolemaic Theory of planetary motion, as a Fourier expansion in epicycles, can, to sufficient order, mimic any observed orbit. Keplerian central motion provides an essential reduction.

Likewise, Bennett and Hacker (2003) conclude that neural-centered explanations of high order mental function commit the mereological fallacy, that is, the fundamental logical error of attributing what is in fact a property of an entirety to a limited part of the whole system. ‘The brain’ does not exist in isolation, but as part of a complete biological individual who is most often deeply embedded in social and cultural contexts.

Neural network-like models of gene expression and development applied to complex living things inherently commit both errors, particularly in a social (or cultural) milieu. This suggests a particular necessity for the formal inclusion of the effects of embedding contexts – the epigenetic  $Z$  – in the sense of Baars (1988, 2005). That is, gene expression and development are conditioned by internal and external signals from embedding physiological, social, and for humans, cul-

tural, structures. As described above, our formulation can include such influences in a highly natural manner.

It is worth repeating that this work provides necessary conditions statistical models of development and gene expression analogous to regression models, although based on the Shannon-McMillan and Rate Distortion Theorems rather than the Central Limit Theorem. It seems necessary to iterate the models to second order, in the same spirit as hierarchical regression, before addressing complex observations. The principal scientific utility of such things is typically the empirical comparison of fitted models for similar systems under different, or different systems under similar, conditions, and the discovery of new structure from analysis of the residuals from fitted models. Although most of the messy computational machinery is hidden one layer deeper than in the work Ciliberti et al., considerable challenges confront attempts to use these methods on real systems.

Accepting such subtleties, nonetheless the observations of Ciliberti et al. regarding the simply connected component in  $w$ -space, although not a complete analysis, may well have opened the door to a far deeper understanding of developmental process in an evolutionary context.

## 7 Mathematical appendix

### 7.1 Groupoids

#### 7.1.1 Basic ideas

Following Weinstein (1996) closely, a groupoid,  $G$ , is defined by a base set  $A$  upon which some mapping – a morphism – can be defined. Note that not all possible pairs of states  $(a_j, a_k)$  in the base set  $A$  can be connected by such a morphism. Those that can define the groupoid element, a morphism  $g = (a_j, a_k)$  having the natural inverse  $g^{-1} = (a_k, a_j)$ . Given such a pairing, it is possible to define ‘natural’ end-point maps  $\alpha(g) = a_j, \beta(g) = a_k$  from the set of morphisms  $G$  into  $A$ , and a formally associative product in the groupoid  $g_1 g_2$  provided  $\alpha(g_1 g_2) = \alpha(g_1), \beta(g_1 g_2) = \beta(g_2)$ , and  $\beta(g_1) = \alpha(g_2)$ . Then the product is defined, and associative,  $(g_1 g_2) g_3 = g_1 (g_2 g_3)$ .

In addition, there are natural left and right identity elements  $\lambda_g, \rho_g$  such that  $\lambda_g g = g = g \rho_g$  (Weinstein, 1996).

An orbit of the groupoid  $G$  over  $A$  is an equivalence class for the relation  $a_j \sim G a_k$  if and only if there is a groupoid element  $g$  with  $\alpha(g) = a_j$  and  $\beta(g) = a_k$ . Following Cannas da Silva and Weinstein (1999), we note that a groupoid is called transitive if it has just one orbit. The transitive groupoids are the building blocks of groupoids in that there is a natural decomposition of the base space of a general groupoid into orbits. Over each orbit there is a transitive groupoid, and the disjoint union of these transitive groupoids is the original groupoid. Conversely, the disjoint union of groupoids is itself a groupoid.

The isotropy group of  $a \in X$  consists of those  $g$  in  $G$  with  $\alpha(g) = a = \beta(g)$ . These groups prove fundamental to classifying groupoids.

If  $G$  is any groupoid over  $A$ , the map  $(\alpha, \beta) : G \rightarrow A \times A$  is

a morphism from  $G$  to the pair groupoid of  $A$ . The image of  $(\alpha, \beta)$  is the orbit equivalence relation  $\sim_G$ , and the functional kernel is the union of the isotropy groups. If  $f : X \rightarrow Y$  is a function, then the kernel of  $f$ ,  $\ker(f) = [(x_1, x_2) \in X \times X : f(x_1) = f(x_2)]$  defines an equivalence relation.

Groupoids may have additional structure. As Weinstein (1996) explains, a groupoid  $G$  is a topological groupoid over a base space  $X$  if  $G$  and  $X$  are topological spaces and  $\alpha, \beta$  and multiplication are continuous maps. A criticism sometimes applied to groupoid theory is that their classification up to isomorphism is nothing other than the classification of equivalence relations via the orbit equivalence relation and groups via the isotropy groups. The imposition of a compatible topological structure produces a nontrivial interaction between the two structures. Below we will introduce a metric structure on manifolds of related information sources, producing such interaction.

In essence, a groupoid is a category in which all morphisms have an inverse, here defined in terms of connection to a base point by a meaningful path of an information source dual to a cognitive process.

As Weinstein (1996) points out, the morphism  $(\alpha, \beta)$  suggests another way of looking at groupoids. A groupoid over  $A$  identifies not only which elements of  $A$  are equivalent to one another (isomorphic), but *it also parametrizes the different ways (isomorphisms) in which two elements can be equivalent*, i.e., all possible information sources dual to some cognitive process. Given the information theoretic characterization of cognition presented above, this produces a full modular cognitive network in a highly natural manner.

Brown (1987) describes the fundamental structure as follows:

A groupoid should be thought of as a group with many objects, or with many identities... A groupoid with one object is essentially just a group. So the notion of groupoid is an extension of that of groups. It gives an additional convenience, flexibility and range of applications...

EXAMPLE 1. A disjoint union [of groups]  $G = \cup_{\lambda} G_{\lambda}$ ,  $\lambda \in \Lambda$ , is a groupoid: the product  $ab$  is defined if and only if  $a, b$  belong to the same  $G_{\lambda}$ , and  $ab$  is then just the product in the group  $G_{\lambda}$ . There is an identity  $1_{\lambda}$  for each  $\lambda \in \Lambda$ . The maps  $\alpha, \beta$  coincide and map  $G_{\lambda}$  to  $\lambda$ ,  $\lambda \in \Lambda$ .

EXAMPLE 2. An equivalence relation  $R$  on [a set]  $X$  becomes a groupoid with  $\alpha, \beta : R \rightarrow X$  the two projections, and product  $(x, y)(y, z) = (x, z)$  whenever  $(x, y), (y, z) \in R$ . There is an identity, namely  $(x, x)$ , for each  $x \in X$ ...

Weinstein (1996) makes the following fundamental point:

Almost every interesting equivalence relation on a space  $B$  arises in a natural way as the orbit equivalence relation of some groupoid  $G$  over  $B$ . Instead of dealing directly with the orbit space  $B/G$  as an object in the category  $S_{map}$  of sets and mappings,

one should consider instead the groupoid  $G$  itself as an object in the category  $G_{htp}$  of groupoids and homotopy classes of morphisms.

The groupoid approach has become quite popular in the study of networks of coupled dynamical systems which can be defined by differential equation models, (e.g., Golubitsky and Stewart 2006).

### 7.1.2 Global and local symmetry groupoids

Here we follow Weinstein (1996) fairly closely, using his example of a finite tiling.

Consider a tiling of the euclidean plane  $R^2$  by identical 2 by 1 rectangles, specified by the set  $X$  (one dimensional) where the grout between tiles is  $X = H \cup V$ , having  $H = R \times Z$  and  $V = 2Z \times R$ , where  $R$  is the set of real numbers and  $Z$  the integers. Call each connected component of  $R^2 \setminus X$ , that is, the complement of the two dimensional real plane intersecting  $X$ , a tile.

Let  $\Gamma$  be the group of those rigid motions of  $R^2$  which leave  $X$  invariant, i.e., the normal subgroup of translations by elements of the lattice  $\Lambda = H \cap V = 2Z \times Z$  (corresponding to corner points of the tiles), together with reflections through each of the points  $1/2\Lambda = Z \times 1/2Z$ , and across the horizontal and vertical lines through those points. As noted by Weinstein (1996), much is lost in this coarse-graining, in particular the same symmetry group would arise if we replaced  $X$  entirely by the lattice  $\Lambda$  of corner points.  $\Gamma$  retains no information about the local structure of the tiled plane. In the case of a real tiling, restricted to the finite set  $B = [0, 2m] \times [0, n]$  the symmetry group shrinks drastically: The subgroup leaving  $X \cap B$  invariant contains just four elements even though a repetitive pattern is clearly visible. A two-stage groupoid approach recovers the lost structure.

We define the transformation groupoid of the action of  $\Gamma$  on  $R^2$  to be the set

$$G(\Gamma, R^2) = \{(x, \gamma, y) | x \in R^2, y \in R^2, \gamma \in \Gamma, x = \gamma y\},$$

with the partially defined binary operation

$$(x, \gamma, y)(y, \nu, z) = (x, \gamma\nu, z).$$

Here  $\alpha(x, \gamma, y) = x$ , and  $\beta(x, \gamma, y) = y$ , and the inverses are natural.

We can form the restriction of  $G$  to  $B$  (or any other subset of  $R^2$ ) by defining

$$G(\Gamma, R^2)|_B = \{g \in G(\Gamma, R^2) | \alpha(g), \beta(g) \in B\}$$

[1]. An orbit of the groupoid  $G$  over  $B$  is an equivalence class for the relation

$x \sim_G y$  if and only if there is a groupoid element  $g$  with  $\alpha(g) = x$  and  $\beta(g) = y$ .

Two points are in the same orbit if they are similarly placed within their tiles or within the grout pattern.

[2]. The isotropy group of  $x \in B$  consists of those  $g$  in  $G$  with  $\alpha(g) = x = \beta(g)$ . It is trivial for every point except

those in  $1/2\Lambda \cap B$ , for which it is  $Z_2 \times Z_2$ , the direct product of integers modulo two with itself.

By contrast, embedding the tiled structure within a larger context permits definition of a much richer structure, i.e., the identification of local symmetries.

We construct a second groupoid as follows. Consider the plane  $R^2$  as being decomposed as the disjoint union of  $P_1 = B \cap X$  (the grout),  $P_2 = B \setminus P_1$  (the complement of  $P_1$  in  $B$ , which is the tiles), and  $P_3 = R^2 \setminus B$  (the exterior of the tiled room). Let  $E$  be the group of all euclidean motions of the plane, and define the local symmetry groupoid  $G_{loc}$  as the set of triples  $(x, \gamma, y)$  in  $B \times E \times B$  for which  $x = \gamma y$ , and for which  $y$  has a neighborhood  $\mathcal{U}$  in  $R^2$  such that  $\gamma(\mathcal{U} \cap P_i) \subseteq P_i$  for  $i = 1, 2, 3$ . The composition is given by the same formula as for  $G(\Gamma, R^2)$ .

For this groupoid-in-context there are only a finite number of orbits:

$\mathcal{O}_1$  = interior points of the tiles.

$\mathcal{O}_2$  = interior edges of the tiles.

$\mathcal{O}_3$  = interior crossing points of the grout.

$\mathcal{O}_4$  = exterior boundary edge points of the tile grout.

$\mathcal{O}_5$  = boundary ‘T’ points.

$\mathcal{O}_6$  = boundary corner points.

The isotropy group structure is, however, now very rich indeed:

The isotropy group of a point in  $\mathcal{O}_1$  is now isomorphic to the entire rotation group  $\mathcal{O}_2$ .

It is  $Z_2 \times Z_2$  for  $\mathcal{O}_2$ .

For  $\mathcal{O}_3$  it is the eight-element dihedral group  $D_4$ .

For  $\mathcal{O}_4, \mathcal{O}_5$  and  $\mathcal{O}_6$  it is simply  $Z_2$ .

These are the ‘local symmetries’ of the tile-in-context.

## 8 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.

Baars, B., 1988, *A Cognitive Theory of Consciousness*, Cambridge University Press, New York.

Baars, B., 2005, Global workspace theory of consciousness: toward a cognitive neuroscience of human experience, *Progress in Brain Research*, 150:45-53.

Beck, C. and F. Schlogl, 1995, *Thermodynamics of Chaotic Systems*, Cambridge University Press, Cambridge, UK.

Bennett, M., and P. Hacker, 2003, *Philosophical Foundations of Neuroscience*, Blackwell Publishing.

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.

Britten, R., and E. Davidson, 1969, Gene regulation for higher cells: a theory, *Science*, 165:349-357.

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.

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

Chouard, T., 2008, Beneath the surface, *Nature*, 456:300-303.

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, *Proceedings 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.

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

Dias, A., and I. Stewart, 2004, Symmetry groupoids and admissible vector fields for coupled cell networks, *Journal of the London Mathematical Society*, 69:707-736.

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

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

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

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

Gobault, E., 2003, Some geometric perspectives in concurrency theory, *Homology, Homotopy, and Applications*, 5:95-136.

Gobault, E., and M. Raussen, 2002, Dihomotopy as a tool in state space analysis. In, Rajsbaum, S., (ed.), *LATIN 2002*, LNCS, vol. 2286, pp. 16-37, Springer, Heidelberg.

Gould, S., 2002, *The Structure of Evolutionary Theory*, Harvard University Press, Cambridge, MA.

Jaeger, J., S. Surkova, M. Blagov, H. Janssens, D. Kosman, K. Kozlov, M. Manu, E. Myasnikova, C. Vanario-Alonso, M. Samsonova, D. Sharp, and J. Reintiz, 2004, Dynamic control of positional information in the early *Drosophila* embryo, *Nature*, 430:368-371.

Khinchin, A., 1957, *Mathematical Foundations of Information Theory*, Dover, New York.

Krebs, P., 2005, Models of cognition: neurological possibility does not indicate neurological plausibility. In Bara, B., L. Barsalou, and M. Bucciarelli (eds.), *Proceedings of CogSci 2005*, pp. 1184-1189, Stresa, Italy. Available at <http://cogprints.org/4498/>.

Landau, L., and E. Lifshitz, 2007, *Statistical Physics, 3rd Edition*, Part I, Elsevier, New York.

Matsumoto, Y., 2002, *An Introduction to Morse Theory*, American Mathematical Society, Providence, RI.

McCauly, J., 1994, *Chaos, Dynamics and Fractals*, Cambridge Nonlinear Science Series, Cambridge, UK.

Mjolsness, E., D. Sharp, and J. Reinitz, 1991, A connectionist model of development, *Journal of Theoretical Biology*, 152:429-458.

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

Raff, E., E. Popodi, J. Kaufmann, B. Sly, F. Turner, V. Morris, and R. Raff, 2003, Regulatory punctuated equilibrium and convergence in the evolution of developmental pathways in direct-developing sea urchins, *Evolution and Development*, 5:478-493.

Reinitz, J., and D. Sharp, 1995, Mechanisms of even stripe formation, *Mechanics of Development* 49:133-158.

Sharp, D., and J. Reinitz, 1998, Prediction of mutant expression patterns using gene circuits, *BioSystems*, 47:79-90.

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.

Waddington, C., 1942, The canalization of development and the inheritance of acquired characters, *Nature*, 150:563-565.

Waddington, C., 1957, *The Strategy of the Genes*, Macmillan, New York.

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

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

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.

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.