Version 3

# GENE EXPRESSION AND ITS DISCONTENTS

Developmental disorders as dysfunctions of epigenetic cognition

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

January 30, 2009

## Abstract

Systems biology presently suffers the same mereological and sufficiency fallacies that haunt neural network models of high order cognition. Shifting perspective from the massively parallel space of gene matrix interactions to the grammar/syntax of the time series of expressed phenotypes using a cognitive paradigm permits import of techniques from statistical physics via the homology between information source uncertainty and free energy density. This produces a broad spectrum of possible statistical models of development and its pathologies in which epigenetic regulation and the effects of embedding environment are analogous to a tunable enzyme catalyst. A cognitive paradigm naturally incorporates memory, leading directly to models of epigenetic inheritance, as affected by environmental exposures, in the largest sense. Understanding gene expression, development, and their dysfunctions will require data analysis tools considerably more sophisticated than the present crop of simplistic models abducted from neural network studies or stochastic chemical reaction theory.

**Key Words** developmental disorder, epigenetic cognition, gene expression, information theory, merological fallacy, phase transition

## 1 Introduction

Researchers have recently begun to explore a de-facto cognitive paradigm for gene expression in which contextual factors determine behavior of what Cohen calls a 'reactive system', not at all a mechanistic process (e.g., Cohen, 2006; Cohen and Harel, 2007; Wallace and Wallace, 2008). O'Nuallain (2008) has, in fact, placed gene expression firmly in the realm of linguistic behavior, for which context imposes meaning:

> ...[T]he analogy between gene expression and language production is useful, both as a fruitful research paradigm and also, given the relative lack of success of natural language processing (nlp) by computer, as a cautionary tale for molecular biology. In particular, given our concern with the Human Genome Project (HGP) and human health, it is noticeable that only 2% of diseases can be traced back to a straightforward genetic cause. As a consequence, we argue that the HGP will have to be redone for a variety of metabolic contexts in order to found a sound technology of genetic engineering (O'Nuallain and Strohman, 2007).
>
> In essence, the analogy works as follows: first of all, at the orthographic or phonological level, depending on whether the language is written or spoken, we can map from phonetic elements to nucleotide sequence. The claim is made that Nature has designed highly ambiguous codes in both cases, and left disambiguation to the context.

This work investigates a class of statistical models based on the asymptotic limit theorems of information theory that instantiates this perspective, and explores a 'natural' means by which epigenetic context 'farms' gene expression in an inherently punctuated manner via a kind of catalysis. These models will then be used to illuminate ways in which 'normal' developmental modes can be driven into pathological trajectories expressed as comorbid psychiatric and physical disorders, expanding recent work by Wallace (2008b).

We begin with a brief reconsideration of the current defacto standard neural network-analog model of development.

## 2 The spinglass model

Following closely Ciliberti et al. (2007), the spinglass model of development 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

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

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 $\Delta t$ is a constant and $\sigma$ a sigmodial function whose value lies in the interval $(-1, 1)$. In the spinglass limit $\sigma$ is the sign function, taking only the values ±1.

The networks of interest are those whose expression state begins from a prespecified initial state $\mathbf{S}(0)$ at time $t = 0$ and converge to a second prespecified stable equilibrium state $\mathbf{S}_\infty$. Such a network is termed *viable*, for obvious reasons. Such viable networks, of course, comprise a tiny fraction of possible ones, i.e., those that do not begin with $\mathbf{S}_0$ and end at $\mathbf{S}_\infty$.

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., Jaeger et al, 2004; Mjolsness et al., 1991; Reinitz and Sharp, 1995; Sharp and Reinitz, 1998; Toulouse et al., 1986). 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 (1994) eloquently argues, is constrained by the necessary conditions imposed by the asymptotic limit theorems of information theory.

The next sections use information theory methods to make the transition from crossectional $w$-space into that of serially correlated sequences of phenotypes, expanding 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).

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 $x = (a_0, a_1, ..., a_n, ...)$ – analogous to the sequence of $\mathbf{S}(t+\Delta t)$ of equation (1), 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, $h$, 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

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

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

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

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

(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 $P(a_0, ..., a_n)$ and $P(a_n|a_0, ..., a_{n-1})$ such that appropriate conditional and joint Shannon uncertainties satisfy the classic relations

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

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

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

(3)

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

*Adiabatic* means that the source has been parametized 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 much like equation (1).

In the developmental vernacular of Ciliberti et al., we now examine paths in phenotype space that begins at some $\mathbf{S}_0$ and converges $n = t/\Delta t \to \infty$ 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.

# 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 $H[X] = \lim_{n\to\infty} \log[N(n)]/n$, 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 $P[X_n = a_n|X_0 = a_0, ...X_{n-1} = a_{n-1}]$ 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)

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

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

(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 $H = \lim_{n \to \infty} \log[N(n)]/n$.

Ash's comment above then has an important 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 – parametizes the set of equivalence relations and creates the (very large) groupoid.

There is a hierarchy 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 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.

## 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 metabolic energy density available at a rate $M$, and an allowed development time $\tau$, let $K = 1/\kappa M \tau$ for some appropriate scaling constant $\kappa$, so that $M\tau$ 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 $\tau$. The developmental energy might be raised to some power, e.g., $K = 1/(\kappa (M\tau)^b)$, 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.

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

$$P[H_{D_i}] = \frac{\exp[-H_{D_i}K]}{\sum_j \exp[-H_{D_j}K]}, \tag{6}$$

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

$$Z_D \equiv \sum_j \exp[-H_{D_j}K]. \tag{7}$$

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

$$F_D[K] \equiv -\frac{1}{K}\log[Z_D[K]], \tag{8}$$

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

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

## 5.3 Spontaneous symmetry breaking

We have expressed the probability of an 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 dynamics.

# 6 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 approximately $\exp[nH(X)]$ 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 $M\tau$. Thus, for a given $M\tau$, 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 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).

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. 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.

# 7 Rate Distortion dynamics

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 $y^n = y_1, y_2, ....$ This is 'digitized' in terms of the observed behavior of the system with which it communicates, say a sequence of observed behaviors $b^n = b_1, b_2, ....$ The $b_i$ happen in real time. Assume each $b^n$ is then deterministically retranslated back into a reproduction of the original biological signal,

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

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 $H(..., ...)$ is the joint and $H(...|...)$ 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

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

(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 communication 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):

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

(12)

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

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

(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 $F(K) = -\log[Z(K)]/K$ where $Z(K)$ the partition function defined by the system Hamiltonian. More precisely, if the possible energy states of the system are a set $E_i, i = 1, 2, ...$ 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, $K = 1/\kappa M\tau$.

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

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

(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:

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

(15)

again taking $K = 1/\kappa M \tau$.

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

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

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 metabolic free 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 $1/K = \kappa M \tau$ suggests the possibility of evolutionary tradeoffs between development time and the rate of available metabolic free energy.

# 8 More topology

It seems possible to extend this treatment using standard topological arguments.

Taking $T = 1/K$ in equations (6) and (14) *as a product of eigenvalues*, we can define it as the determinant of a Hessian matrix representing a Morse Function, $f$, on some underlying, background, manifold, $\mathcal{M}$, characterized in terms of (as yet unspecified) variables $\mathcal{X} = (x^1, ..., x^n)$, so that

$$1/K = \det(\mathcal{H}_{i,j}),$$

$$\mathcal{H}_{i,j} \equiv \partial^2 f/\partial x^i \partial x^j.$$

See the Appendix for a brief outline of Morse Theory.

Thus $\kappa, M$, and the development time $\tau$ are seen as eigenvalues of $\mathcal{H}$ on the manifold $\mathcal{M}$ in an abstract space defined by some set of variables $\mathcal{X}$.

By construction $\mathcal{H}$ has everywhere only nonzero, and indeed, positive, eigenvalues, whose product thereby defines $T$ as a generalized volume. Thus, and accordingly, all critical points of $f$ have index zero, that is, no eigenvalues of $\mathcal{H}$ are ever negative at any point, and hence at any critical point $\mathcal{X}_c$ where $df(\mathcal{X}_c) = 0$.

This defines a particularly simple topological structure for $\mathcal{M}$: If the interval $[a, b]$ contains a critical value of $f$ with a single critical point $\mathcal{X}_c$, then the topology of the set $\mathcal{M}_b$ defined above differs from that of $\mathcal{M}_a$ in a manner determined by the index $i$ of the critical point. $\mathcal{M}_b$ is then homeomorphic to the manifold obtained from attaching to $\mathcal{M}_a$ an i-handle, the direct product of an i-disk and an $(m - i)$-disk.

One obtains, in this case, since $i = 0$, the two halves of a sphere with critical points at the top and bottom (Matsumoto, 2002; Pettini, 2007). This is, as in Ciliberti et al. (2007), a simply connected object. What one does then is to invoke the Seifert-Van Kampen Theorem (SVKT, Lee, 2000) and patch together the various simply connected subcomponents to construct the larger, complicated, topological object representing the full range of possibilities.

The physical natures of $\kappa, M$, and $\tau$ thus impose constraints on the possible complexity of this system, in the sense of the SVKT.

# 9 Inherited epigenetic memory

The cognitive paradigm for gene expression invoked here requires an internal picture of the world against which incoming signals are compared – algorithmically combined according to the rules of Section 3 – and then fed into a nonlinear decision oscillator that chooses one (or a few) action(s) from a much large repertoire of possibilities. Memory is inherent, and recent work suggests that epigenetic memory is indeed heritable. Jablonka and Lamb (1998), in a now-classic review, argued that information can be transmitted from one generation to the next in ways other than through the base sequence of DNA. It can be transmitted through cultural and behavioral means in higher animals, and by epigenetic means in cell lineages. All of these transmission systems allow the inheritance of environmentally induced variation. Such Epigenetic Inheritance Systems are the memory systems that enable somatic cells of different phenotypes but identical genotypes to transmit their phenotypes to their descendants, even when the stimuli that originally induced these phenotypes are no longer present.

In chromatin-marking systems information is carried from one cell generation to the next because it rides with DNA as binding proteins or additional chemical groups that are attached to DNA and influence its activity. When DNA is replicated, so are the chromatin marks. One type of mark is the methylation pattern a gene carries. The same DNA sequence can have several different methylation patterns, each reflecting a different functional state. These alternative patterns can be stably inherited through many cell divisions.

Epigenetic inheritance systems are very different from the genetic system. Many variations are directed and predictable outcomes of environmental changes. Epigenetic variants are frequently, although not necessarily, adaptive. The frequency with which variants arise and their rate of reversion varies widely and epigenetic variations induced by environmental changes may be produced coordinatedly at several loci.

Jablonka and Lamb (1998) conclude that epigenetic systems may therefore produce rapid, reversible, co-ordinated, heritable changes. However such systems can also underlie non-induced changes changes that are induced but non-adaptive, and changes that are very stable.

What is needed, in their view, is a concept of epigenetic heritability comparable to the classical concept of heritability, and a model similar to those used for measuring the effects of cultural inheritance on human behavior in populations.

Following a furious decade of research and debate, Bossdorf et al. (2008), for example, are able to conclude that heritable variation in ecologically relevant traits can be generated through a suite of epigenetic mechanisms, even in the absence of genetic variation. Moreover, recent studies indicate that epigenetic variation in natural populations can be independent from genetic variation, and that in some cases environmentally induced epigenetic changes may be inherited by future generations. They infer that we might need to expand our concept of variation and evolution in natural populations, taking into account several (likely interacting) ecologically relevant inheritance systems. Potentially, this may result in a significant expansion (though by all means not a negation) of the Modern Evolutionary Synthesis as well as in more conceptual and empirical integration between ecology and evolution.

The abduction of spinglass and other models from neural network studies to the analysis of development and its evolution carries with it the possibility of more than one system of memory. What Baars called 'contexts' channeling high level animal cognition may often be the influence of cultural inheritance, in a large sense. Our formalism suggests a class of statistical models that indeed greatly generalize those used for measuring the 'effects of cultural inheritance on human behavior in populations'.

Epigenetic machinery, as a dual information source to a cognitive process, serves as a heritable system, intermediate between (relatively) hard-wired classical genetics, and a (usually) highly Larmarckian embedding cultural context. In particular, the three heritable systems interact, in our model, through a crosstalk in which the epigenetic machinery acts as a kind of intelligent catalyst for gene expression.

# 10 Multiple processes

The argument to this point has, in large measure, been directly abducted from recent formal studies of high level cognition – consciousness – based on a Dretske-style information theoretic treatment of Bernard Baars' global workspace model (Wallace, 2005; Atmanspacher, 2006). A defining and grossly simplifying characteristic of that phenomenon is its rapidity: typically the global broadcasts of consciousness occur in a matter of a few hundred milliseconds, limiting the number of processes that can operate simultaneously. Slower cognitive dynamics can, therefore, be far more complex than individual consciousness. One well known example is institutional distributed cognition that encompasses both individual and group cognition in a hierarchical structure typically operating on timescales ranging from a few seconds or minutes in combat or hunting groups, to years at the level of major governmental structures, commercial enterprises, religious organizations, or other analogous large scale cultural artifacts. Wallace and Fullilove (2008) provide the first formal mathematical analysis of institutional distributed cognition.

Clearly cognitive gene expression is not generally limited to a few hundred milliseconds, and something much like the distributed cognition analysis may be applied here as well. Extending the analysis requires recognizing an individual cognitive actor can participate in more than one 'task', synchronously, asynchronously, or strictly sequentially. Again, the analogy is with institutional function whereby many individuals often work together on several distinct projects: Envision a multiplicity of possible cognitive gene expression dual 'languages' that themselves form a higher order network linked by crosstalk.

Next, describe crosstalk measures linking different dual languages on that meta-meta (MM) network by some characteristic magnitude $\omega$, and *define a topology on the MM network by renormalizing the network structure to zero if the crosstalk is less than $\omega$ and set it equal to one if greater or equal to it.* A particular $\omega$, of sufficient magnitude, defines a giant component of network elements linked by mutual information greater or equal to it, in the sense of Erdos and Renyi (1960), as more fully described in Wallace and Fullilove (2008, Section 3.4).

The fundamental trick is, in the Morse Theory sense (Matsumoto, 2002), to invert the argument so that a given topology for the giant component will, in turn, define some critical value, $\omega_C$, so that network elements interacting by mutual information less than that value will be unable to participate, will be locked out and not active. $\omega$ becomes an epigenetically syntactically-dependent detection limit, and depends critically on the instantaneous topology of the giant component defining the interaction between possible gene interaction MM networks.

Suppose, now, that a set of such giant components exists at some generalized system 'time' $k$ and is characterized by a set of parameters $\Omega_k \equiv \omega_1^k, ..., \omega_m^k$. Fixed parameter values define a particular giant component set having a particular set of topological structures. Suppose that, over a sequence of times the set of giant components can be characterized by a possibly coarse-grained path $\gamma_n = \Omega_0, \Omega_1, ..., \Omega_{n-1}$ having significant serial correlations that, in fact, permit definition of an adiabatically, piecewise stationary, ergodic (APSE) information source $\Gamma$.

Suppose that a set of (external or internal) epigenetic signals impinging on the set of such giant components can also be characterized by another APSE information source $Z$ that interacts not only with the system of interest globally, but with the tuning parameters of the set of giant components characterized by $\Gamma$. Pair the paths $(\gamma_n, z_n)$ and apply the joint information argument above, generating a splitting criterion between high and low probability sets of pairs of paths. We now have a multiple workspace cognitive genetic expression structure driven by epigenetic catalysis.

# 11 Multiple models

Recently R.G. Wallace and R. Wallace (2009) have argued that consciousness may have undergone the characteristic branching and pruning of evolutionary development, particularly in view of the rapidity of currently surviving conscious mechanisms. They write

> Evolution is littered with polyphyletic parallelisms: many roads lead to functional Romes. We propose that consciousness [as a particular form of high order cognitive process operating in real time] embodies one such example, [represented by] ... an equivalence class structure that factors the broad realm of necessary conditions information theoretic realizations of Baars' global workspace model... [M]any different physiological systems can support rapidly shifting, highly tunable, and even simultaneous assemblages of interacting unconscious cognitive modules... The variety of possibilities suggests minds today may be only a small surviving fraction of ancient evolutionary radiations – bush phylogenies of consciousness pruned by selection and chance extinction.

Even in the realms of rapid global broadcast inherent to real time cognition, R.G. Wallace and R. Wallace (2009), following a long tradition, speculate that ancient backbrain structures instantiate rapid emotional responses, while the newer forebrain harbors rapid 'reasoned' responses in animal consciousness. The cooperation and competition of these two rapid phenomena produces, of course, a plethora of systematic behaviors.

Since consciousness is necessarily restricted to realms of a few hundred milliseconds, evolutionary pruning may well have resulted in only a small surviving fraction of previous evolutionary radiations. Processes operating on longer timescales may well be spared such draconian evolutionary selection. That is, the vast spectrum of mathematical models of cognitive gene expression inherent to our analysis here, in the context of development times much longer than a few hundred milliseconds, implies current organisms may simultaneously harbor several, possibly many, quite different cognitive gene expression mechanisms.

It seems likely that slower cognitive phenomena, like institutional distributed cognition and cognitive gene expression, may well permit the operation of very many quite different cognitive processes simultaneously.

One inference is, then, that cognitive gene expression is far more complex than individual consciousness, currently regarded as one of the 'really big' unsolved scientific problems.

Neural network models adapted from the cognition studies of a generation ago are unlikely to cleave the Gordian Knot of scientific inference surrounding gene expression.

# 12 Epigenetic focus

The Tuning Theorem analysis of the Appendix permits an inattentional blindness/concentrated focus perspective on the famous computational 'no free lunch' theorem of Wolpert and Macready (1995, 1997). Following closely the arguments of English (1996), Wolpert and Macready have established that there exists no generally superior function optimizer. There is no 'free lunch' in the sense that an optimizer 'pays' for superior performance on some functions with inferior performance on others. if the distribution of functions is uniform, then gains and losses balance precisely, and all optimizers have identical average performance. The formal demonstration depends primarily upon a theorem that describes how information is conserved in optimization. This Conservation Lemma states that when an optimizer evaluates points, the posterior joint distribution of values for those points is exactly the prior joint distribution. Put simply, observing the values of a randomly selected function does not change the distribution: An optimizer has to 'pay' for its superiority on one subset of functions with inferiority on the complementary subset.

As English puts it, anyone slightly familiar with the evolutionary computing literature recognizes the paper template 'Algorithm $X$ was treated with modification $Y$ to obtain the best known results for problems $P_1$ and $P_2$.' Anyone who has tried to find subsequent reports on 'promising' algorithms knows that they are extremely rare. Why should this be?

A claim that an algorithm is the very best for two functions is a claim that it is the very worst, on average, for all but two functions. It is due to the diversity of the benchmark set of test problems that the 'promise' is rarely realized. Boosting performance for one subset of the problems usually detracts from performance for the complement.

English concludes that hammers contain information about the distribution of nail-driving problems. Screwdrivers contain information about the distribution of screw-driving problems. Swiss army knives contain information about a broad distribution of survival problems. Swiss army knives do many jobs, but none particularly well. When the many jobs must be done under primitive conditions, Swiss army knives are ideal.

Thus, according to English, the tool literally carries information about the task optimizers are literally tools-an algorithm implemented by a computing device is a physical entity.

Another way of looking at this is to recognize that a computed solution is simply the product of the information processing of a problem, and, by a very famous argument, information can never be gained simply by processing. Thus a problem $X$ is transmitted as a message by an information processing channel, $Y$, a computing device, and recoded as an answer. By the Tuning Theorem argument of the Appendix there will be a channel coding of $Y$ which, when properly tuned, is most efficiently transmitted by the problem. In general, then, the most efficient coding of the transmission channel, that is, the best algorithm turning a problem into a solution, will necessarily be highly problem-specific. Thus there can be no best algorithm for all equivalence classes of problems, although there may well be an optimal algorithm for any given class. The tuning theorem form of the No Free Lunch theorem will apply quite generally to cognitive biological and social structures, as well as to massively parallel machines.

Rate distortion, however, occurs when the problem is collapsed into a smaller, simplified, version and then solved. Then there must be a tradeoff between allowed average distortion and the rate of solution: the retina effect. In a very fundamental sense – particularly for real time systems – rate distortion manifolds present a generalization of the converse of the Wolpert/Macready no free lunch arguments. The neural corollary is known as inattentional blindness (Wallace, 2007).

We are led to suggest that there may well be a considerable set of no free lunch-like conundrums confronting highly parallel real-time structures, including epigenetic control of gene expression, and that they may interact in distinctly nonlinear ways.

# 13 Developmental disorders

Let $U$ be an information source representing a systematic embedding environmental 'program' interacting with the process of cognitive gene expression, here defined as a complicated information set of sources having source joint uncertainty $H(Z_1, ..., Z_n)$ that guides the system into a particular equivalence class of desired developmental behaviors and trajectories.

Most simply, one can directly invoke the network information theory result of equations (3) and (4) of Wallace (2008a), so that

$$H(Z_1, ..., Z_n|U) = H(U) + \sum_{j=1}^{n} H(Z_j|U) - H(Z_1, ..., Z_n, U)$$

(17)

Again, the $Z_i$ represent internal information sources and $U$ that of the embedding environmental context.

The central point is that a one step extension of that system via the results of network information theory (Cover and Thomas, 1991) allows incorporating the effect of an external environmental 'farmer' in guiding cognitive developmental gene expression.

The environmental farming of development may not always be benign.

Suppose we can operationalize and quantify degrees of both overfocus/inattentional blindness (IAB) and of overall structure/environment distortion (D) in the actions of a highly parallel cognitive epigenetic regulatory system. The essential assumption is that the (internal) dual information source of a cognitive structure that has low levels of both IAB overfocus and structure/environment distortion will tend to be richer than that of one having greater levels. This is shown in figure 1a, where $H$ is the source uncertainty dual to internal cognitive process, $X = IAB$, and $Y = D$. Regions of low $X, Y$, near the origin, have greater source uncertainty than those nearby, so $H(X, Y)$ shows a (relatively gentle) peak at the origin, taken here as simply the product of two error functions.

We are, then, particularly interested in the internal cognitive capacity of the structure itself, as paramatized by degree of overfocus and by the (large scale) distortion between implementation and impact. That capacity, a purely internal quantity, need not be convex in the parameter $D$, which is taken to characterize interaction with an external environment, and thus becomes a context for internal measures.

The generalized Onsager argument, based on the homology between information source uncertainty and free energy, as explained more fully in the Appendix, is shown in figure 1b. $S = H(X, Y) - XdH/dX - YdH/dY$, the 'disorder' analog to entropy in a physical system, is graphed on the $Z$ axis against the $X - Y$ plane, assuming a gentle peak in $H$ at the origin. Peaks in $S$, according to theory, constitute repulsive system barriers, which must be overcome by external forces. In figure 1b there are three quasi-stable topological resilience modes, in the sense of Wallace (2008b), marked as $A, B$, and $C$. The $A$ region is locked in to low levels of both overfocus and distortion, as it sits in a pocket. Forcing the system in either direction, that is, increasing either IAB or D, will, initially, be met by homeostatic attempts to return to the resilience state $A$, according to this model.

If overall distortion becomes severe in spite of homeostatic developmental mechanisms, the system will then jump to the quasi-stable state $B$, a second pocket. According to the model, however, once that transition takes place, there will be a tendency for the system to remain in a condition of high distortion. That is, the system will become locked-in to a structure with high distortion in the match between structure implementation and structure impact, but one having lower overall cognitive capacity, i.e., a lower value of $H$ in figure 1a.

The third pocket, marked $C$, is a broad plain in which both IAB and D remain high, a highly overfocused, poorly linked pattern of behavior which will require significant intervention to alter once it reaches such a quasi-stable resilience mode. The structure's cognitive capacity, measured by $H$ in figure 1a, is the lowest of all for this condition of pathological re-

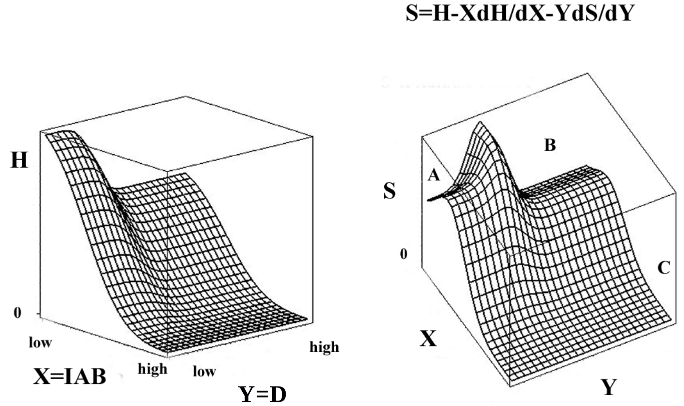

Figure 1: a. Source uncertainty, $H$, of the dual information source of epigenetic cognition, as parametized by degrees of focus, $X = IAB$ and distortion, $Y = D$, between implementation and actual impact. Note the relatively gentle peak at low values of $X, Y$. Here $H$ is generated as the product of two error functions. b. Generalized Onsager treatment of figure 1a. $S = H(X,Y) - XdH/dX - YdH/dY$. The regions marked $A, B$, and $C$ represent realms of resilient quasi-stability, divided by barriers defined by the relative peaks in $S$. Transition among them requires a forcing mechanism. From another perspective, limiting energy or other resources, or imposing stress from the outside – driving down $H$ in figure 1a, would force the system into the lower plain of $C$, in which the system would then become trapped in states having high levels of distortion and inattentional blindness/overfocus.

silience, and attempts to correct the problem – to return to condition $A$, will be met with very high barriers in $S$, according to figure 1b. That is, mode $C$ is very highly resilient, although pathologically so, much like the eutrophication of a pure lake by sewage outflow. See Wallace (2008a, b) for a discussion of resilience and literature references.

We can argue that the three quasi-equilibrium configurations of figure 1b represent different dynamical states of the system, and that the possibility of transition between them represents the breaking of the associated symmetry groupoid by external forcing mechanisms. That is, three manifolds representing three different kinds of system dynamics have been patched together to create a more complicated topological structure. For cognitive phenomena, such behavior is likely to be the rule rather than the exception. 'Pure' groupoids are abstractions, and the fundamental questions will involve linkages which break the underlying symmetry.

In all of this, system convergence is not to some fixed state, limit cycle, or pseudorandom strange attractor, but rather to some appropriate set of highly dynamic information sources, i.e., behavior patterns constituting, here, developmental trajectories, rather than to some fixed 'answer to a computing problem' (Wallace, 2009).

What this model suggests is that sufficiently strong external perturbation can force a highly parallel real-time cognitive epigenetic structure from a normal, almost homeostatic, developmental path into one involving a widespread, comorbid, developmental disorder. This is a well studied pattern for humans and their institutions, reviewed at some length elsewhere (Wallace and Fullilove, 2008; Wallace, 2008b). Indeed, this argument provides the foundation of a fairly comprehensive model of chronic developmental dysfunction across a broad class of cognitive systems, including, but not limited to, cognitive epigenetic control of gene expression. One approach might be as follows:

A developmental process can be viewed as involving a sequence of surfaces like figure 1, having, for example, 'critical periods' when the barriers between the normal state A and the pathological states B and C are relatively low. During such a time the system would become highly sensitive to perturbation, and to the onset of a subsequent pathological developmental trajectory. Critical periods might occur during times of rapid growth and/or high system demand for which an energy limitation imposes the need to focus via something like a rate distortion manifold. Cognitive process requires energy through the homologies with free energy density, and more focus at one end necessarily implies less at some other. In a distributed zero sum developmental game, as it were, some cognitive processes must receive more attentional metabolic free energy than others.

A structure trapped in region C might be said to suffer something much like what Wiegand (2003) describes as the loss of gradient problem, in which one part of a multiple population coevolutionary system comes to dominate the others, creating an impossible situation in which the other participants do not have enough information from which to learn. That is, the cliff just becomes too steep to climb. Wiegand

(2003) also characterizes focusing problems in which a two-population coevolutionary process becomes overspecialized on the opponent's weaknesses, effectively a kind of inattentional blindness.

Thus there seems some consonance between our asymptotic analysis of cognitive structural function and current studies of pathologies affecting coevolutionary algorithms (e.g. Ficici et al., 2005; Wallace, 2009). In particular the possibility of historic trajectory, of path dependence, in producing individualized failure modes, suggests there can be no one-size-fits-all amelioration strategy.

Equation (17) basically enables a kind of environmental catalysis to cognitive gene expression, in a sense closely similar to the arguments of Section 6. This is analogous to, but more general than, the 'mesoscale resonance' invoked by Wallace and Wallace (2008): during critical periods, according to these models, environmental signals can have vast impact on developmental trajectory.

# 14 Discussion and conclusions

We have hidden the massively parallel neural network-like 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).

Analogies with neural network studies suggest, however, 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. Krebs' particular characterization is that 'neural possibility is not neural plausibility'.

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, cultural, or environmental milieu. This suggests a particular necessity for the formal inclusion of the effects of embedding contexts – the epigenetic $Z$ and the environmental $U$ – 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, cultural, environments. As described above, our formulation can include such influences in a highly natural manner, as they influence epigenetic catalysis. In addition, multiple, and quite different, cognitive gene expression mechanisms may operate simultaneously, or in appropriate sequence, given sufficient development time.

Developmental disorders, in a broad sense that must include comorbid mental and physical characteristics, emerge as pathological 'resilience' modes, in the sense of Wallace (2008b). Environmental farming through an embedding information source affecting internal epigenetic regulation of gene expression, can, as a kind of programming of a highly parallel cognitive system, place the organism into a quasi-stable pathological developmental behavior pattern.

The model of developmental disorder presented here is, most fundamentally, a statistical one based on the asymptotic limit theorems of information theory, in the same sense that regression models are really a broad class based on the Central Limit Theorem. We have not, then, given 'a' model of developmental disorder in cognitive gene expression, but, rather, outlined a general strategy for fitting empirically-determined statistical models of developmental disorder to real data, in precisely the sense that one would fit regression models to data.

Such statistical models do not, in themselves, do science. That is done by comparing fitted models for similar systems under different, or different systems under similar, conditions, and by examining the structure of residuals.

A particular inference of this work, then, is that understanding complicated processes of gene expression and development – and their pathologies – will require construction of data analysis tools considerably more sophisticated than now available, including the present crop of simplistic systems biology models abducted from neural network studies or stochastic chemical reaction theory.

# 15 Mathematical appendix

## 15.1 Groupoids

### 15.1.1 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 $g_1g_2$ provided $\alpha(g_1g_2) = \alpha(g_1), \beta(g_1g_2) = \beta(g_2)$, and $\beta(g_1) = \alpha(g_2)$. Then the product is defined, and associative, $(g_1g_2)g_3 = g_1(g_2g_3)$.

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 $a_j \sim Ga_k$ 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 $(\alpha, \beta) : G \to A \times A$ 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 $f : X \to Y$ 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 parametizes 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 $\alpha, \beta : R \to X$ 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).

### 15.1.2 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 $R^2 \backslash X$, 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), $P_2 = B \backslash P_1$ (the complement of $P_1$ in $B$, which is the tiles), and $P_3 = R^2 \backslash B$ (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 $O_2$.

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.

## 15.2 Morse Theory

Morse theory examines relations between analytic behavior of a function – the location and character of its critical points – and the underlying topology of the manifold on which the function is defined. We are interested in a number of such functions, for example information source uncertainty on a parameter space and 'second order' iterations involving parameter manifolds determining critical behavior, for example sudden onset of a giant component in the mean number model (Wallace and Wallace, 2008), and universality class tuning in the mean field model of the next section. These can be reformulated from a Morse theory perspective. Here we follow closely the elegant treatments of Pettini (2007) and Kastner (2006).

The essential idea of Morse theory is to examine an $n$-dimensional manifold $M$ as decomposed into level sets of some function $f : M \to \mathbf{R}$ where $\mathbf{R}$ is the set of real numbers. The $a$-level set of $f$ is defined as

$$f^{-1}(a) = \{x \in M : f(x) = a\},$$

the set of all points in $M$ with $f(x) = a$. If $M$ is compact, then the whole manifold can be decomposed into such slices in a canonical fashion between two limits, defined by the minimum and maximum of $f$ on $M$. Let the part of $M$ below $a$ be defined as

$$M_a = f^{-1}(-\infty, a] = \{x \in M : f(x) \leq a\}.$$

These sets describe the whole manifold as $a$ varies between the minimum and maximum of $f$.

Morse functions are defined as a particular set of smooth functions $f : M \to \mathbf{R}$ as follows. Suppose a function $f$ has a critical point $x_c$, so that the derivative $df(x_c) = 0$, with critical value $f(x_c)$. Then $f$ is a Morse function if its critical points are nondegenerate in the sense that the Hessian matrix of second derivatives at $x_c$, whose elements, in terms of local coordinates are

$$\mathcal{H}_{i,j} = \partial^2 f / \partial x^i \partial x^j,$$

has rank $n$, which means that it has only nonzero eigenvalues, so that there are no lines or surfaces of critical points and, ultimately, critical points are isolated.

The index of the critical point is the number of negative eigenvalues of $\mathcal{H}$ at $x_c$.

A level set $f^{-1}(a)$ of $f$ is called a critical level if $a$ is a critical value of $f$, that is, if there is at least one critical point $x_c \in f^{-1}(a)$.

Again following Pettini (2007), the essential results of Morse theory are:

[1] If an interval $[a, b]$ contains no critical values of $f$, then the topology of $f^{-1}[a, v]$ does not change for any $v \in (a, b]$. Importantly, the result is valid even if $f$ is not a Morse function, but only a smooth function.

[2] If the interval $[a, b]$ contains critical values, the topology of $f^{-1}[a, v]$ changes in a manner determined by the properties of the matrix $H$ at the critical points.

[3] If $f : M \to \mathbf{R}$ is a Morse function, the set of all the critical points of $f$ is a discrete subset of $M$, i.e., critical points are isolated. This is Sard's Theorem.

[4] If $f : M \to \mathbf{R}$ is a Morse function, with $M$ compact, then on a finite interval $[a, b] \subset \mathbf{R}$, there is only a finite number of critical points $p$ of $f$ such that $f(p) \in [a, b]$. The set of critical values of $f$ is a discrete set of $\mathbf{R}$.

[5] For any differentiable manifold $M$, the set of Morse functions on $M$ is an open dense set in the set of real functions of $M$ of differentiability class $r$ for $0 \leq r \leq \infty$.

[6] Some topological invariants of $M$, that is, quantities that are the same for all the manifolds that have the same topology as $M$, can be estimated and sometimes computed exactly once all the critical points of $f$ are known: Let the Morse numbers $\mu_i (i = 0, ..., m)$ of a function $f$ on $M$ be the number of critical points of $f$ of index $i$, (the number of negative eigenvalues of $H$). The Euler characteristic of the complicated manifold $M$ can be expressed as the alternating sum of the Morse numbers of any Morse function on $M$,

$$\chi = \sum_{i=1}^{m} (-1)^i \mu_i.$$

The Euler characteristic reduces, in the case of a simple polyhedron, to

$$\chi = V - E + F$$

where $V, E$, and $F$ are the numbers of vertices, edges, and faces in the polyhedron.

[7] Another important theorem states that, if the interval $[a, b]$ contains a critical value of $f$ with a single critical point $x_c$, then the topology of the set $M_b$ defined above differs from that of $M_a$ in a way which is determined by the index, $i$, of the critical point. Then $M_b$ is homeomorphic to the manifold obtained from attaching to $M_a$ an $i$-handle, i.e., the direct product of an $i$-disk and an $(m - i)$-disk.

Again, see Pettini (2007) or Matusmoto (2002) for details.

## 15.3 Generalized Onsager Theory

Understanding the time dynamics of groupoid-driven information systems away from the kind of phase transition critical points described above requires a phenomenology similar to the Onsager relations of nonequilibrium thermodynamics. This also leads to a general theory involving large-scale topological changes in the sense of Morse theory.

If the Groupoid Free Energy of a biological process is parametized by some vector of quantities $\mathbf{K} \equiv (K_1, ..., K_m)$, then, in analogy with nonequilibrium thermodynamics, gradients in the $K_j$ of the *disorder*, defined as

$$S_G \equiv F_G(\mathbf{K}) - \sum_{j=1}^{m} K_j \partial F_G / \partial K_j$$

(18)

become of central interest.

Equation (18) is similar to the definition of entropy in terms of the free energy of a physical system.

Pursuing the homology further, the generalized Onsager relations defining temporal dynamics of systems having a GFE become

$$dK_j/dt = \sum_i L_{j,i} \partial S_G / \partial K_i,$$

(19)

where the $L_{j,i}$ are, in first order, constants reflecting the nature of the underlying cognitive phenomena. The L-matrix is to be viewed empirically, in the same spirit as the slope and intercept of a regression model, and may have structure far different than familiar from more simple chemical or physical processes. The $\partial S_G / \partial K$ are analogous to thermodynamic forces in a chemical system, and may be subject to override by external physiological or other driving mechanisms: biological and cognitive phenomena, unlike simple physical systems, can make choices as to resource allocation.

That is, an essential contrast with simple physical systems driven by (say) entropy maximization is that complex biological or cognitive structures can make decisions about resource allocation, to the extent resources are available. Thus resource availability is a context, not a determinant, of behavior.

Equations (18) and (19) can be derived in a simple parameter-free covariant manner which relies on the underlying topology of the information source space implicit to the development (e.g., Wallace and Wallace, 2008b). We will not pursue that development here.

The dynamics, as we have presented them so far, have been noiseless, while biological systems are always very noisy. Equation (19) might be rewritten as

$$dK_j/dt = \sum_i L_{j,i} \partial S_G / \partial K_i + \sigma W(t)$$

where $\sigma$ is a constant and $W(t)$ represents white noise. This leads directly to a family of classic stochastic differential equations having the form

$$dK_t^j = L^j(t, \mathbf{K})dt + \sigma^j(t, \mathbf{K})dB_t,$$

(20)

where the $L^j$ and $\sigma^j$ are appropriately regular functions of $t$ and $\mathbf{K}$, and $dB_t$ represents the noise structure, and we have readjusted the indices.

Further progress in this direction requires introduction of methods from stochastic differential geometry and related topics in the sense of Emery (1989). The obvious inference is that noise – not necessarily 'white' – can serve as a tool to shift the system between various topological modes, as a kind of crosstalk and the source of a generalized stochastic resonance.

Effectively, topological shifts between and within dynamic manifolds constitute another theory of phase transitions (Pettini, 2007), and this phenomenological Onsager treatment would likely be much enriched by explicit adoption of a Morse theory perspective.

## 15.4 The Tuning Theorem

Messages from an information source, seen as symbols $x_j$ from some alphabet, each having probabilities $P_j$ associated with a random variable $X$, are 'encoded' into the language of a 'transmission channel', a random variable $Y$ with symbols $y_k$, having probabilities $P_k$, possibly with error. Someone receiving the symbol $y_k$ then retranslates it (without error) into some $x_k$, which may or may not be the same as the $x_j$ that was sent.

More formally, the message sent along the channel is characterized by a random variable $X$ having the distribution

$$P(X = x_j) = P_j, j = 1, ..., M.$$

The channel through which the message is sent is characterized by a second random variable $Y$ having the distribution

$$P(Y = y_k) = P_k, k = 1, ..., L.$$

Let the joint probability distribution of $X$ and $Y$ be defined as

$$P(X = x_j, Y = y_k) = P(x_j, y_k) = P_{j,k}$$

and the conditional probability of $Y$ given $X$ as

$$P(Y = y_k|X = x_j) = P(y_k|x_j).$$

Then the Shannon uncertainty of $X$ and $Y$ independently and the joint uncertainty of $X$ and $Y$ together are defined respectively as

$$H(X) = -\sum_{j=1}^{M} P_j \log(P_j)$$

$$H(Y) = -\sum_{k=1}^{L} P_k \log(P_k)$$

$$H(X, Y) = -\sum_{j=1}^{M}\sum_{k=1}^{L} P_{j,k} \log(P_{j,k}).$$

(21)

The *conditional uncertainty* of $Y$ given $X$ is defined as

$$H(Y|X) = -\sum_{j=1}^{M}\sum_{k=1}^{L} P_{j,k} \log[P(y_k|x_j)].$$

(22)

For any two stochastic variates $X$ and $Y$, $H(Y) \geq H(Y|X)$, as knowledge of $X$ generally gives some knowledge of $Y$. Equality occurs only in the case of stochastic independence.

Since $P(x_j, y_k) = P(x_j)P(y_k|x_j)$, we have

$$H(X|Y) = H(X, Y) - H(Y).$$

The information transmitted by translating the variable $X$ into the channel transmission variable $Y$ – possibly with error – and then retranslating without error the transmitted $Y$ back into $X$ is defined as

$$I(X|Y) \equiv H(X) - H(X|Y) = H(X) + H(Y) - H(X, Y)$$

(23)

See, for example, Ash (1990), Khinchin (1957) or Cover and Thomas (1991) for details. The essential point is that if there is no uncertainty in $X$ given the channel $Y$, then there is no loss of information through transmission. In general this will not be true, and herein lies the essence of the theory.

Given a fixed vocabulary for the transmitted variable $X$, and a fixed vocabulary and probability distribution for the channel $Y$, we may vary the probability distribution of $X$ in such a way as to maximize the information sent. The capacity of the channel is defined as

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

(24)

subject to the subsidiary condition that $\sum P(X) = 1$.

The critical trick of the Shannon Coding Theorem for sending a message with arbitrarily small error along the channel $Y$ at any rate $R < C$ is to encode it in longer and longer 'typical' sequences of the variable $X$; that is, those sequences whose distribution of symbols approximates the probability distribution $P(X)$ above which maximizes $C$.

If $S(n)$ is the number of such 'typical' sequences of length $n$, then

$$\log[S(n)] \approx nH(X),$$

where $H(X)$ is the uncertainty of the stochastic variable defined above. Some consideration shows that $S(n)$ is much less than the total number of possible messages of length $n$. Thus, as $n \to \infty$, only a vanishingly small fraction of all possible messages is meaningful in this sense. This observation, after some considerable development, is what allows the Coding Theorem to work so well. In sum, the prescription is to encode messages in typical sequences, which are sent at very nearly the capacity of the channel. As the encoded messages become longer and longer, their maximum possible rate of transmission without error approaches channel capacity as a limit. Again, Ash (1990), Khinchin (1957) and Cover and Thomas (1991) provide details.

This approach can be, in a sense, inverted to give a tuning theorem which parsimoniously describes the essence of the Rate Distortion Manifold.

Telephone lines, optical wave, guides and the tenuous plasma through which a planetary probe transmits data to earth may all be viewed in traditional information-theoretic terms as a *noisy channel* around which we must structure a message so as to attain an optimal error-free transmission rate.

Telephone lines, wave guides, and interplanetary plasmas are, relatively speaking, fixed on the timescale of most messages, as are most other signaling networks. Indeed, the capacity of a channel, is defined by varying the probability distribution of the 'message' process $X$ so as to maximize $I(X|Y)$.

Suppose there is some message $X$ so critical that its probability distribution must remain fixed. The trick is to fix the distribution $P(x)$ but *modify the channel* – i.e., tune it – so as to maximize $I(X|Y)$. The *dual* channel capacity $C^*$ can be defined as

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

(25)

But

$$C^* = \max_{P(Y), P(Y|X)} I(Y|X)$$

since

$$I(X|Y) = H(X) + H(Y) - H(X,Y) = I(Y|X).$$

Thus, in a purely formal mathematical sense, *the message transmits the channel*, and there will indeed be, according to the Coding Theorem, a channel distribution $P(Y)$ which maximizes $C^*$.

One may do better than this, however, by modifying the channel matrix $P(Y|X)$. Since

$$P(y_j) = \sum_{i=1}^{M} P(x_i)P(y_j|x_i),$$

$P(Y)$ is entirely defined by the channel matrix $P(Y|X)$ for fixed $P(X)$ and

$$C^* = \max_{P(Y), P(Y|X)} I(Y|X) = \max_{P(Y|X)} I(Y|X).$$

Calculating $C^*$ requires maximizing the complicated expression

$$I(X|Y) = H(X) + H(Y) - H(X,Y),$$

which contains products of terms and their logs, subject to constraints that the sums of probabilities are 1 and each probability is itself between 0 and 1. Maximization is done by varying the channel matrix terms $P(y_j|x_i)$ within the constraints. This is a difficult problem in nonlinear optimization. However, for the special case $M = L$, $C^*$ may be found by inspection:

If $M = L$, then choose

$$P(y_j|x_i) = \delta_{j,i},$$

where $\delta_{i,j}$ is 1 if $i = j$ and 0 otherwise. For this special case

$$C^* \equiv H(X),$$

with $P(y_k) = P(x_k)$ for all $k$. *Information is thus transmitted without error when the channel becomes 'typical' with respect to the fixed message distribution $P(X)$.*

If $M < L$, matters reduce to this case, but for $L < M$ information must be lost, leading to Rate Distortion limitations.

Thus modifying the channel may be a far more efficient means of ensuring transmission of an important message than encoding that message in a 'natural' language which maximizes the rate of transmission of information on a fixed channel.

We have examined the two limits in which either the distributions of $P(Y)$ or of $P(X)$ are kept fixed. The first provides the usual Shannon Coding Theorem, and the second a tuning theorem variant, a tunable retina-like Rate Distortion Manifold. It seems likely, however, than for many important systems $P(X)$ and $P(Y)$ will interpenetrate, to use Richard Levins' terminology. That is, $P(X)$ and $P(Y)$ will affect each other in characteristic ways, so that some form of mutual tuning may be the most effective strategy.

# 16 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.

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.

Bossdorf, O., C. Richards, and M. Pigliucci, 2008, Epigenetics for ecologists, *Ecology Letters*, 11:106-115.

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.

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.

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.

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.

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

English, T., 1996, Evaluation of evolutionary and genetic optimizers: no free lunch. In Fogel, L, P. Angeline, and T. Back (eds.), *Evolutionary Programming V: Proceedings of the Fifth Annual Conference on Evolutionary Programming*, 163-169, MIT Press, Cambridge, MA.

Erdos, P., and A. Renyi, 1960, On the evolution of random graphs. Reprinted in *The Art of Counting*, 1973, 574-618, and inn *Selected Papers of Alfred Renyi*, 1976, 482-525.

Ficici, S., O. Milnik, and J. Pollak, 2005, A game-theoretic and dynamical systems analysis of selection methods in coevolution, *IEEE Transactions on Evolutionary Computation*, 9:580-602.

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.

Jablonka, E., and M. Lamb, 1998, Epigenetic inheritance in evolution, *Journal of Evolutionary Biology*, 11:159-183.

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.

Kastner, M., 2006, Phase transitions and configuration space topology. ArXiv cond-mat/0703401.

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.

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

O'Nuallain, S., 2008, Code and context in gene expression, cognition, and consciousness. Chapter 15 in Barbiere, M., (ed.), *The Codes of Life: The Rules of Macroevolution*, Springer, New York, pp. 347-356.

O'Nuallain, S., and R. Strohman, 2007, Genome and natural language: how far can the analogy be extended? In Witzany, G., (ed.), *Proceedings of Biosemiotics*, Tartu University Press, Umweb, Finland.

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

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.

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., 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.

Wallace, R., 2009, Programming coevolutionary machines: the emerging conundrum. 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.G., and R. Wallace, 2009, Evolutionary radiation and the spectrum of consciousness. In press *Consciousness and Cognition*, doi:10.1026/j.concog.2008.12.002.

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.

Wiegand, R., 2003, *An analysis of cooperative coevolutionary algorithms*, PhD Thesis, George Mason University.

Wolpert, D., and W. Macready, 1995, No free lunch theorems for search, Santa Fe Institute, SFI-TR-02-010.

Wolpert, D., and W. Macready, 1997, No free lunch theorems for optimization, *IEEE Transactions on Evolutionary Computation*, 1:67-82.