Code, Context, and Epigenetic Catalysis in Gene Expression

Rodrick Wallace[1] and Deborah Wallace[2]

[1] The New York State Psychiatric Institute,
549 W. 123 St., Suite 16F, New York, NY, 10027. Tel.: (212) 865-4766
wallace@pi.cpmc.columbia.edu
[2] Consumers Union
rdwall@ix.netcom.com

Abstract. We examine a class of probability models describing how epigenetic context affects gene expression and organismal development, using the asymptotic limit theorems of information theory in a highly formal manner. Taking classic results on spontaneous symmetry breaking abducted from statistical physics in groupoid, rather than group, circumstances, the work suggests that epigenetic information sources act as analogs to a tunable catalyst, directing development into different characteristic pathways according to the structure of external signals. The results have significant implications for epigenetic epidemiology, in particular for understanding how environmental stressors, in a large sense, can induce a broad spectrum of developmental disorders in humans.

1 Introduction

1.1 Toward New Tools

Researchers have begun to explore a de-facto cognitive paradigm for gene expression in which contextual factors determine the behavior of what Cohen calls a 'reactive system', not at all a deterministic, or even stochastic, mechanical process (e.g., [18, 19, 74]). The different approaches, while highly formal, are nonetheless much in the spirit of the pioneering efforts of Maturana and Varela [53, 54] who foresaw the essential role that cognitive process must play in a vast realm of biological phenomena.

O'Nuallain [57] has recently placed gene expression firmly in the realm of complex linguistic behavior, for which context imposes meaning, claiming that the 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. First O'Nuallain argues that, at the orthographic or phonological level, depending on whether the language is written or spoken, we can map from phonetic elements to nucleotide sequence. His second claim is that Nature has designed highly ambiguous codes in both cases, and left disambiguation to the context.

C. Priami et al. (Eds.): Trans. on Comput. Syst. Biol. XI, LNBI 5750, pp. 283–334, 2009.

He notes that, given our concern with the Human Genome Project (HGP) and its implications for human health, only 2% of diseases can be traced back to a straightforward genetic cause. As a consequence the HGP will have to be redone for a variety of metabolic contexts in order to establish a sound technology of genetic engineering [58].

Here we investigate a broad class of probability models based on the asymptotic limit theorems of information theory that instantiate this perspective, finding a 'natural' means by which epigenetic context 'farms' gene expression in an inherently punctuated manner via a kind of tunable catalysis. The models will be used to study how normal developmental modes can be driven by external context into pathological trajectories often expressed, in humans, as comorbid psychiatric and physical disorders, expanding recent work [71]. It appears possible to convert such models to powerful tools for data analysis, much as those based on the Central Limit Theorem can be converted to parametric statistics. A more formal version of the underlying mathematics can be found in [34].

We will begin with a summary of the biological context, then examine the popular spinglass model of development taken from neural network studies that we will ultimately generalize using a cognitive paradigm. The expanded approach permits import of tools and methods from statistical physics via the homology between information source uncertainty and free energy density, and this leads directly to the idea of epigenetic catalysis.

It is worth keeping in mind throughout the formal mathematics that Feynman's basic measure of information is simply the free energy needed to erase it [31].

1.2 Epigenetic Epidemiology

What we attempt is itself embedded in a large and lively intellectual context. Jablonka and Lamb [41, 42] have long 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 structures 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, in the view of [41, 42], often, 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.

Parenthetically, some authors, e.g., [39], disagree with the assumption of adaptiveness, inferring that input responsible for methylation effects simply produces a phenotypic variability then subject to selection. The matter remains open.

Jablonka and Lamb [42] 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, they feel, 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, this perspective received much empirical confirmation. Backdahl et al. [6], for example, write that epigenetic regulation of gene expression primarily works through modifying the secondary and tertiary structures of DNA (chromatin), making it more or less accessible to transcription. The sum and interaction of epigenetic modifications has been proposed to constitute an 'epigenetic code' which organizes the chromatin structure on different hierarchical levels [67]. Modifications of histones include acetylation, methylation, phosphorylation, ubiquitination, and sumoylation, but also other modifications have been observed. Some such modifications are quite stable and play an important part in epigenetic memory although DNA methylation is the only epigenetic modification that has maintenance machinery which preserves the marks through mitosis. This argues for DNA methylation to function as a form of epigenetic memory for the epigenome.

Codes and memory, of course, are inherent to any cognitive paradigm.

Jaenish and Bird [45] argue that cells of a multicellular organism are genetically homogeneous but structurally and functionally heterogeneous owing to the differential expression of genes. Many of these differences in gene expression arise during development and are subsequently retained through mitosis. External influences on epigenetic processes are seen in the effects of diet on long-term diseases such as cancer. Thus, epigenetic mechanisms seem to allow an organism to respond to the environment through changes in gene expression. Epigenetic modifications of the genome provide a mechanism that allows the stable propagation of gene activity states from one generation of cells to the next. Because epigenetic states are reversible they can be modified by environmental factors, which may contribute to the development of abnormal responses. What needs to be explained, from their perspective, is the variety of stimuli that can bring about epigenetic changes, ranging from developmental progression and aging to viral infection and diet.

Jaenish and Bird conclude that the future will see intense study of the chains of signaling that are responsible for epigenetic programming. As a result, we will

be able to understand, and perhaps manipulate, the ways in which the genome learns from experience.

Indeed, our central interest precisely regards the manner in which the asymptotic limit theorems of information theory constrain such chains of signaling, in the same sense that the Central Limit Theorem constrains sums of stochastic variates.

Crews et al. [21, 22] provide a broad overview of induced epigenetic change in phenotype, as do Guerrero-Bosagna et al. [39], who focus particularly on early development. They propose that changes arising because of alterations in early development processes, in some cases environmentally induced, can appear whether or not such changes could become fixed and prosper in a population. They recognize two ways for this to occur, first by dramatically modifying DNA aspects in the germ line with transgenerational consequences – mutations or persistent epigenetic modifications of the genome – or by inducing ontogenetical variation in every generation, although not inheritance via the germ line. From their perspective inductive environmental forces can act to create, through these means, new conformations of organisms which also implies new possibilities within the surrounding environment.

Foley et al. [32] take a very general perspective on the prospects for epigenetic epidemiology. They argue that epimutation is estimated to be 100 times more frequent than genetic mutation and may occur randomly or in response to the environment. Periods of rapid cell division and epigenetic remodeling are likely to be most sensitive to stochastic or environmentally mediated epimutation. Disruption of epigenetic profile is a feature of most cancers and is speculated to play a role in the etiology of other complex diseases including asthma, allergy, obesity, type 2 diabetes, coronary heart disease, autism spectrum disorders, bipolar disorders, and schizophrenia.

They find evidence that a small change in the level of DNA methylation, especially in the lower range in an animal model, can dramatically alter expression for some genes. The timing of nutritional insufficiency or other environmental exposures may also be critical. In particular low-level maternal care was associated with developmental dysfunction and altered stress response in the young. Foley et al. emphasize the potential implications of such findings, given how widely stress is implicated in disease onset and relapse.

They especially note that when epigenetic status or change in status over time is the outcome, then models for either threshold-based dichotomies or proportional data will be required. Threshold models, defined by a given level or pattern of methylation or a degree of change in methylation over time, will, in their view, benefit from relevant functional data to identify meaningful thresholds.

A special contribution of the approach taken here is that just such threshold behavior leads ‘naturally’ to a language-like ‘dual information source’ constrained by the necessary conditions imposed by information theory’s asymptotic limit theorems, allowing development of statistical models of complicated cognitive phenomena, including but not limited to cognitive gene expression.

A recent review by Weaver [77] focuses specifically on the epigenetic effects of glucocorticoids – stress hormones. In mammals, Weaver argues, the closeness or degree of positive attachment in parent-infant bonding and parental investment during early life has long-term consequences on development of inter-individual differences in cognitive and emotional development in the offspring. The long-term effects of the early social experience, he continues, particularly of the mother-offspring interaction, have been widely investigated. The nature of that interaction influences gene expression and the development of behavioral responses in the offspring that remain stable from early development to the later stages of life. Although enhancing the offspring's ability to respond according to environmental clues early in life can have immediate adaptive value, the cost, Weaver says, is that these adaptations serve as predictors of ill health in later life. He concludes that maternal influences on the development of neuroendocrine systems that underlie hypothalamic-pituitary-adrenal (HPA) axis and behavioral responses to stress mediate the relation between early environment and health in the adult offspring. In particular, he argues, exposure of the mother to environmental adversity alters the nature of mother-offspring interaction, which, in turn, influences the development of defensive responses to threat and reproductive strategies in the progeny.

In an updated review of epigenetic epidemiology, Jablonka [43] finds it clear that the health and general physiology of animals and people can be affected not only by the interplay of their own genes and conditions of life, but also by the inherited effects of the interplay of genes and environment in their ancestors. These ancestral influences on health, Jablonka says, depend neither on inheriting particular genes, nor on the persistence of the ancestral environment.

Significantly, Bossdorf et al. [11] invoke 'contexts' much like Baars' model of consciousness [68], and infer a need to expand the 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.

More formally, Scherrer and Jost [62, 63] use information theory arguments to extend the definition of the gene to include the local epigenetic machinery, something they characterize as the 'genon'. Their central point is that coding information is not simply contained in the coded sequence, but is, in their terms, *provided by* the genon that accompanies it on the expression pathway and controls in which peptide it will end up. In their view the information that counts is not about the identity of a nucleotide or an amino acid derived from it, but about the relative frequency of the transcription and generation of a particular type of coding sequence that then contributes to the determination of the types and numbers of functional products derived from the DNA coding region under consideration.

From our perspective the formal tools for understanding such phenomena involve asymptotic limit theorems affecting information sources – active systems

that generate or 'provide' information – and these are respectively the Rate Distortion Theorem and its zero error limit, the Shannon-McMillan Theorem, described in the Mathematical Appendix.

We begin with a reconsideration of the current de-facto standard systems biology neural network-analog model of development, and proceed to its generalization.

2 Models of Development

2.1 The Spinglass Model

Ciliberti et al.[16, 17], culminating a long series of papers, apply the spinglass model from statistical physics to organisimal development in an evolutionary context. We summarize their formalism and look at some of the less obvious topological implications – in particular the mapping of disjoint directed homotopy classes of phenotype paths into interaction matrix space. We then extend the approach by applying a cognitive paradigm for gene expression first developed in [74]. Analogs to phase transition arguments in physical systems generate punctuated equilibrium evolutionary transitions in a 'highly natural' manner, even for the spinglass treatment, and a hierarchical extension permits incorporation of epigenetic effects as a kind of tunable catalysis.

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 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

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

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

The networks of interest in the spinglass model are those whose expression state begins from a prespecified initial state $\mathbf{S}(0)$ at time $t = 0$ and converge to a prespecified stable equilibrium state $\mathbf{S}_\infty$. Such a network is termed *viable*, for obvious reasons.

After an elaborate and very difficult simulation exercise, a particular series of results emerges. Reference [16] finds that viable networks comprise a tiny

fraction of possible ones. They could be widely scattered in the space of all possible networks and occupy disconnected islands in this space. However, direct computation indicates precisely the opposite. The metagraph of viable networks has one 'giant' connected component that comprises most or all viable networks. Any two networks in this component can be reached from one another through gradual changes of one regulatory interaction at a time, changes that never leave the space of viable networks, for this calculation.

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

Define $0 \leq d \leq 1$ as the the fraction of genes that differ in their expression state between $\mathbf{S}_0$ and $\mathbf{S}_\infty$. A typical result is that for $N = 5$ genes, $6 \leq M \leq 7$ total regulatory interactions, and $d = 0.4$, full enumeration finds a total of only 37,338 viable networks out of 6.3×10^7 possible ones [16]. Long random walks through the space of viable networks, however, visit all but a very small fraction of the nodes of the metagraph, and this missing fraction decreases as N increases. Large N require elaborate Monte Carlo sampling for simulation, a difficult and computationally intensive enterprise.

In w-space [16, 17] define a metric characterizing the distance between two network topologies as

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

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

Several observations emerge directly.

1. This approach is formally similar to spinglass neural network models of learning by selection, e.g., as proposed by Toulouse et al. [66] nearly a generation ago. Subsequent work [4, 5], summarized in [23], suggests that such models are simply not sufficient to the task of understanding high level cognitive function, and these have been largely supplanted by complicated 'global workspace' concepts whose mathematical characterization is highly nontrivial [3].

2. What [16, 17] observe, in another idiom, is that in phenotype space, in **S**-space, the set of all paths associated with viable networks forms an equivalence class, closely analogous to the directed homotopy equivalence classes in the sense of [36, 37]. Directed homotopy differs from simple homotopy (e.g., [50]) in that one uses paths from one point to another rather than loops, and seeks continuous deformations between them. See [74] for discussion in a biological context. Thus there is, in this spinglass model, a mapping from **S**-space into (w_{ij}) space, characterized by the metric D, that associates a unique simply connected component with each dihomotopy-like equivalence class of paths connecting two particular phenotype points. Indeed, the w-space component might well be treated according to standard homotopy arguments, i.e., using loops.

3. What one does with homotopically simply connected components is patch them together to build larger, and more interesting, topological structures, using the Seifert-Van Kampen Theorem (SVKT) (e.g., [50], Ch. 10). If paths within S-space are not continuously transformable into one another, (if there are 'holes'), then several distinct dihomotopy classes will exist, e.g., as in figures 1 and 2 of [74], explored further below in terms of developmental critical periods and their 'shadows'. The obvious conjecture is that, under such a circumstance, very complex topological objects may lurk in w-space, not just the simply connected component discovered by by [16, 17]. These may, according to the SVKT, intersect as well as exist as isolated and disconnected sets.

In particular, if there are dihomotopy 'holes' in S-space, consequently reflected in disconnected patches in w-space, then punctuated transition events of various sorts may well become an evolutionary norm, as in [38], even for the spinglass model.

4. A large and increasing body of work surrounding coupled cell networks invokes groupoids, a natural generalization of symmetry groups. As [25] remarks, until recently the abstract theory of coupled cell systems has mainly focused on the effects of symmetry in the network and the consequent formation of spatial and spatiotemporal patterns. The formal setting for this theory centers upon the symmetry group of the network.

Reference [25] concludes that analysis of robust patterns of synchrony in general coupled cell systems – that is, dynamics in which sets of cells behave identically as a consequence of the network topology – leads to the fruitful notion of the 'symmetry groupoid' of a coupled cell network. A groupoid is a generalization of a group, in which products of elements are not always defined. The symmetry groupoid of a coupled cell network is a natural algebraic formalization of the 'local symmetries' that relate subsets of the network to each other. In particular 'admissible' vector fields – those specified by the network topology – are precisely those that are equivariant under the action of the symmetry groupoid.

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

5. Both of these – analogous – approaches can apparently be coarse-grained into a symbolic dynamics associated with (simple) information sources having particular grammar and syntax. The method is straightforward (e.g., [7, 55]). One could, thus, probably translate the spinglass results of Ciliberti et al. into symbolic dynamics, using groupoid methods to study the underlying topological objects.

6. The spinglass model of development is abstracted from longstanding (if ultimately unsucessful) attempts at similar treatments of neural networks involved in high level cognition (e.g., [44, 56, 61, 64]). Thus and consequently [16, 17] are invoking an implicit cognitive paradigm for gene expression (e.g., [18, 19, 74]). Cognitive process, as the philosopher Fred Dretske eloquently argues (e.g., [26]), is constrained by the necessary conditions imposed by the asymptotic limit theorems of information theory. A little work produces a very general cognitive gene expression metanetwork structure recognizably similar to that found

in [16, 17]. The massively parallel computations are hidden, somewhat, in the required empirical fitting of regression model analogs based on the asymptotic limit theorems of information theory rather than on the central limit theorem.

7. A salient characteristic of high level cognitive process is precisely its inherent punctuation (e.g., [4, 5, 68]), and this emerges directly using an information theory approach via the famous homology between information and free energy (e.g., [31]). 'Simple' neural network analogs will inevitably have more difficulty replicating such behavior, but as discussed, the mapping of disjoint dihomotopy equivalence classes from phenotype sequence space to disjoint sets in interaction matrix space provides a straightforward example for spinglass models.

The next sections use information theory methods to make the transition from crossectional w-space into that of serially correlated sequences of phenotypes, expanding on the results of [74].

2.2 Shifting Perspective: Cognition as an Information Source

Atlan and Cohen [2], in the context of a study of the immune system, argue that the essence of cognition is the comparison of a perceived signal with an internal, learned picture of the world, and then choice of a single response from a large repertoire of possible responses.

Such choice inherently involves information and information transmission since it always generates a reduction in uncertainty, as explained in [1] (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, a 'sudden threshold machine', in a sense, that generates an output $h(x)$ that is an element of one of two disjoint sets B_0 and B_1 of possible system responses. Let us define the sets B_k as

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

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

Assume a graded response, supposing that if

$$h(x) \in B_0,$$

the pattern is not recognized, and if

$$h(x) \in B_1,$$

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

The principal objects of formal interest are paths x triggering pattern recognition-and-response. That is, given a fixed initial state a_0, examine all possible subsequent paths x beginning with a_0 and leading to the event $h(x) \in B_1$. Thus $h(a_0, ..., a_j) \in B_0$ for all $0 < j < m$, but $h(a_0, ..., a_m) \in B_1$.

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

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

$$H = \lim_{n \to \infty} \frac{\log[N(n)]}{n} \tag{2}$$

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 [73].

Invoking the spirit of the Shannon-McMillan Theorem, whose content is described in more detail in the Appendix, 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

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

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

$$\lim_{n \to \infty} \frac{H(X_0, ..., X_n)}{n+1}. \tag{3}$$

See the Mathematical Appendix for a summary of basic information theory results.

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 [68, 74].

Using the developmental vernacular of [16, 17], 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.

This transformation carries heavy computational burdens, as well as providing deeper 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 [47] characterizes as the 'E-property' of a stationary, ergodic information source. This property is that, 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 [20, 24, 47].

Second, information source uncertainty has an important heuristic interpretation that [1] describes as follows:

> ...[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., [49])

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

$$\lim_{V \to \infty} \frac{\log[\hat{Z}(K,V)]}{V}, \tag{4}$$

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

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

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

where E_i is the energy of state i, k a constant, and T the system temperature, and $P[E_i]$ is the probability of state i.

Feynman [31], following the classic arguments of [9] that present idealized machines using information to do work, concludes *the information contained in a message is most simply measured by the free energy needed to erase it.* The arguments of [9] are clever indeed, and the Feynman treatment of them in [31] is well worth reading.

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.

3 Symmetry Arguments

A formal equivalence class algebra, in the sense of the groupoid section 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 [16]. 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 [78]. 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 an implicit hierarchy. First, there is structure *within the system having the same base and end points*, as in [16]. 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.

We begin with the simple case, the set of dual information sources associated with a fixed pair of beginning and end states.

3.1 The First Level

The spinglass model of [16, 17] 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 τ, let $K = 1/\kappa M\tau$ for some appropriate scaling constant κ, so that $M\tau$ is total developmental free energy. Then the probability of a particular H_α will be determined by the standard expression (e.g., [49]),

$$P[H_\beta] = \frac{\exp[-H_\beta K]}{\sum_\alpha \exp[-H_\alpha K]}, \tag{5}$$

where the sum may, in fact, be a complicated abstract integral.

This is just a version of the fundamental probability relation from statistical mechanics, as above. The sum in the denominator, the partition function in statistical physics, is a crucial normalizing factor that allows the definition of of $P[H_\beta]$ as a probability.

A basic requirement, then, is that the sum/integral always converges. K is the inverse product of a scaling factor, a metabolic energy density rate term, and

a characteristic development time τ. The developmental energy might be raised to some power, e.g., $K = 1/(\kappa(M\tau)^b)$, suggesting the possibility of allometric scaling.

Thus, in this formulation, there must be structure *within* a (cross sectional) connected component in the w-space of [16, 17], 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.

3.2 The Second Level

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 [52, 59]), 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.

Equivalence classes define groupoids, by standard mechanisms [13, 35, 78]. 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

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

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

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

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

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

again following the standard arguments from statistical physics [31, 49].

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

3.3 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, 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 [49, 59, 65].

The essence of 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, with one phase being far more symmetric than the other. A symmetry is lost in the transition, a phenomenon called spontaneous symmetry breaking. 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 – ensures that higher symmetry/energy states – mixed transitive groupoid structures – will then be accessible to the system. Absent high metabolic free energy, 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., [10, 14]).

Somewhat more rigorously, the biological renormalization schemes of the Appendix to [74] 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 Morse-Theory-based topological hypothesis [59] 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 [16, 17] is unlikely to be the full story, and that more complete models will likely be plagued – or graced – by highly punctuated dynamics.

Several matters are worth noting. First, Landau's spontaneous symmetry breaking arguments are perhaps the simplest approach possible here. The formal mathematical development requires invoking holonomy groups and groupoids, as in [34].

Second, one need not be restricted to terms of the form $\exp[-H_j K]$, as any $f(H_j, K)$ such that the sum over j converges will serve, although the resulting 'thermodynamic' relations between variates of central interest may then be less elegant.

Third, there may be some allometric scaling tradeoff between metabolic energy rate and development time determined by a relation of the form $K \propto (\tau M)^{\alpha}$.

4 Tunable Epigenetic Catalysis

Incorporating the influence of embedding contexts – epigenetic effects – is most elegantly done by invoking the Joint Asymptotic Equipartition Theorem (JAEPT) [20]. 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, from the information theory 'chain rule' [20],

$$H(X,Y) = H(X) + H(Y|X) \leq H(X) + H(Y),$$

while there are approximately $\exp[nH(X)]$ typical X sequences, and $\exp[nH(Z)]$ typical Z sequences, and hence $\exp[n(H(x) + H(Y))]$ independent joint sequences, there are only about $\exp[nH(X,Z)] \leq \exp[n(H(X) + H(Y))]$ jointly typical sequences, so that the effect of the embedding context, in this model, is to lower the *relative* free energy of a particular developmental channel.

Thus the effect of epigenetic regulation is to channel development into pathways that might otherwise be inhibited by an energy barrier. 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 [73, 74].

It is worth emphasizing that this is indeed a relative energy argument, since, metabolically, two systems must now be supported, i.e., that of the 'reaction' itself and that of its catalytic regulator. 'Programming' and stabilizing inevitably intertwined, as it were.

This elaboration allows a spectrum of possible 'final' phenotypes, what [33] calls developmental or phenotype plasticity. Thus gene expression is seen as, in part, responding to environmental or other, internal, developmental signals.

West-Eberhard [79] argues that 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, according to [79], 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. There cannot, [79] concludes, be a change in the phenotype, a novel phenotypic state, without an altered developmental pathway.

These mechanisms are accomplished in our formulation by allowing the set B_1 in section 2.2 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}.

Again, [62, 63] use information theory arguments to suggest something similar to epigenetic catalysis, finding the information in a sequence is not contained in the sequence but has been provided by the machinery that accompanies it on the expression pathway. That work does not, however, invoke a cognitive paradigm, its attendant groupoid symmetries, or the homology between information source uncertainty and free energy density that drives dynamics.

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

5 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 [20].

Define a distortion measure $d(y, \hat{y})$ which compares the original to the retranslated path. Many distortion measures are possible, as described in the Mathematical Appendix.

The distortion between *paths* y^n and $\hat{y}^n$ is defined as

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

A remarkable fact of the Rate Distortion Theorem is that *the basic result is independent of the exact distortion measure chosen* [20, 24].

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

$$D = \sum_{y^n} p(y^n) d(y^n, \hat{y}^n). \tag{9}$$

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:

$$I(Y, \hat{Y}) = H(Y) - H(Y|\hat{Y}) = H(Y) + H(\hat{Y}) - H(Y, \hat{Y}), \tag{10}$$

where $H(..., ...)$ is the joint and $H(...|...)$ the conditional uncertainty [1, 20].

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

$$R(D) = \min_{p(y,\hat{y}); \sum_{(y,\hat{y})} p(y)p(y|\hat{y})d(y,\hat{y}) \leq D} I(Y, \hat{Y}). \tag{11}$$

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. References [20, 24] 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 [1, 20]:

$$H[\mathbf{X}] \leq C, \tag{12}$$

where H is the uncertainty of the source X and C the channel capacity, defined according to the relation [1, 20]

$$C = \max_{P(X)} I(X|Y). \tag{13}$$

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

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

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:

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

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

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

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

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 [74] 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.

6 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

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

$$\mathcal{H}_{i,j} = \partial^2 f/\partial x^i \partial x^j. \tag{17}$$

Again, see the Appendix for a brief outline of Morse Theory.

Thus κ, M, and the development time τ 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 [52, 59]. This is, as in [16], a simply connected object. What one does then is to invoke the Seifert-Van Kampen Theorem (SVKT, [50]) 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 κ, M, and τ thus impose constraints on the possible complexity of this system, in the sense of the SVKT.

7 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 2.2 – and then fed into a sharply stepwise decision oscillator that chooses one (or a few) action(s) from a much large repertoire of possibilities. Memory is inherent, and much recent work, as described in the introduction, suggests that epigenetic memory is indeed heritable.

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.

8 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 [3, 68]. 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. Reference [73] provides 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 ω, and *define a topology on the MM network by renormalizing the network structure to zero if*

the crosstalk is less than ω and set it equal to one if greater or equal to it. A particular ω, of sufficient magnitude, defines a giant component of network elements linked by mutual information greater or equal to it, in the sense of [29], as more fully described in [73] (Section 3.4).

The fundamental trick is, in the Morse Theory sense [52], to invert the argument so that a given topology for the giant component will, in turn, define some critical value, ω_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. ω 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 $\Omega_k = \omega_1^k, ..., \omega_m^k$. 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 Γ.

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 Γ. Pair the paths (γ_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.

9 'Coevolutionary' Development

The model can be applied to multiple interacting information sources representing simultaneous gene expression processes, for example across a spatially differentiating organism as it develops. This is, in a broad sense, a 'coevolutionary' phenomenon in that the development of one segment may affect that of others.

Most generally we assume that different cognitive developmental subprocesses of gene expression characterized by information sources H_m interact through chemical or other signals and assume that *different processes become each other's principal environments*, a broadly coevolutionary phenomenon.

We write

$$H_m = H_m(K_1...K_s, ...H_j...), \tag{18}$$

where the K_s represent other relevant parameters and $j \neq m$.

The dynamics of such a system is driven by a recursive network of stochastic differential equations, similar to those used to study many other highly parallel dynamic structures (e.g., [83]).

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

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

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

$$dQ_t^j = \sum_i [L_{j,i}(t, ...\partial S^m / \partial Q^i ...)dt + \sigma_{j,i}(t, ...\partial S^m / \partial Q^i ...)dB_t^i], \tag{19}$$

where, again, for notational simplicity only, we have expressed both the H_j and the external K's in terms of the same symbols Q_j.

m ranges over the H_m and we could allow different kinds of 'noise' dB_t^i, having particular forms of quadratic variation that may, in fact, represent a projection of environmental factors under something like a rate distortion manifold [73, 74].

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

Second, however, there may well be analogs to fragmentation when the system exceeds the critical values of K_c according to the approach of [74]. That is, the K-parameter structure will represent full-scale fragmentation of the entire structure, and not just punctuation within it.

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

There are other possible patterns:

1. Setting equation (19) equal to zero and solving for stationary points again gives attractor states since the noise terms preclude unstable equilibria.

2. This system may converge to limit cycle or 'strange attractor' behaviors in which the system seems to chase its tail endlessly, e.g., the cycle of climate-driven phenotype changes in persistent temperate region plants.

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

Here we become enmeshed in a system of highly recursive phenomenological stochastic differential equations, but at a deeper level than the standard stochastic chemical reaction model (e.g., [84]), and in a dynamic rather than static manner: the objects of this system are equivalence classes of information sources and their crosstalk, rather than simple final states of a chemical system.

10 Multiple Models

Recent work [75] argues that consciousness may have undergone the characteristic branching and pruning of evolutionary development, particularly in view of the rapidity of currently surviving conscious mechanisms. According to that study, evolution is littered with polyphyletic parallelisms: many roads lead to functional Romes, and 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. Many different physiological systems, then, can support rapidly shifting, highly tunable, and even simultaneous assemblages of interacting unconscious cognitive modules. Thus [75] concludes 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, [75] speculates, following a long tradition, 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, then, that, with some generality, slow phenomena, ranging from institutional distributed cognition to cognitive gene expression, permit the operation of very many quite different cognitive processes simultaneously or in rapid succession.

One inference is, then, that gene expression and its epigenetic regulation are, for even very simple organisms, far more complex than individual human consciousness, currently regarded as one of the 'really big' unsolved scientific problems.

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

11 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 [81, 82]. Following closely the arguments of [28], [81, 82] 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 [28] describes, 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.

Reference [28] argues 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 [28], 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 no free lunch arguments. The neural corollary is known as inattentional blindness [69].

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 complicated ways.

12 Developmental Disorders

12.1 Network Information Theory

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.

To model the effect of U on development one can, most simply, invoke results from network information theory, ([20], p. 388). Given three interacting information sources, say Y_1, Y_2, Z, the splitting criterion between high and low probability sets of states, taking Z as the external context, is given by

$$I(Y_1, Y_2|Z) = H(Z) + H(Y_1|Z) + H(Y_2|Z) - H(Y_1, Y_2, Z),$$

where, again, $H(...|...)$ and $H(..., ..., ...)$ represent conditional and joint uncertainties. This generalizes to the relation

$$I(Y_1, ..., Y_n|Z) = H(Z) + \sum_{j=1}^{n} H(Y_j|Z) - H(Y_1, ..., Y_n, Z).$$

Thus the fundamental splitting criterion between low and high probability sets of joint developmental paths becomes

$$I(Z_1, ..., Z_n|U) = H(U) + \sum_{j=1}^{n} H(Z_j|U) - H(Z_1, ..., Z_n, U). \tag{20}$$

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 [20] allows incorporating the effect of an external environmental 'farmer' in guiding cognitive developmental gene expression.

12.2 Embedding Ecosystems as Information Sources

The principal farmer for a developing organism is the ecosystem in which it is embedded, in a large sense. Summarizing briefly the arguments of [74], ecosystems, under appropriate coarse graining, often have reconizable grammar and

syntax. For example, the turn-of-the-seasons in a temperate climate, for most natural communities, is remarkably similar from year to year in the sense that the ice melts, migrating birds return, trees bud, flowers and grass grow, plants and animals reproduce, the foliage turns, birds migrate, frost, snow, the rivers freeze, and so on in a predictable manner from year to year.

Suppose, then, that we can coarse grain an ecosystem at time t according to some appropriate partition of the phase space in which each division A_j represents a particular range of numbers for each possible species in the ecosystem, along with associated parameters such as temperature, rainfall, humidity, insolation, and so on. We examine longitudinal paths, statements of the form

$$x(n) = A_0, A_1, ..., A_n$$

defined in terms of some 'natural' time unit characteristic of the system. Then n corresponds to a time unit T, so that $t = T, 2T, ..., nT$. Our interest is in the serial correlation along paths. If $N(n)$ is the number of possible paths of length n that are consistent with the underlying grammar and syntax of the appropriately coarse grained ecosystem, for example, spring leads to summer, autumn, winter, back to spring, etc., but never spring to autumn to summer to winter in a temperate climate.

The essential assumption is that, for appropriate coarse graining, $N(n)$, the number of possible grammatical paths, is much smaller than the total conceivable number of paths, and that, in the limit of large n,

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

both exists and is independent of path.

Not all possible ecosystem coarse grainings are likely to lead to this result, as is sometimes the case with Markov models. Reference [40] in particular emphasizes that mesoscale ecosystem processes are most likely to entrain dynamics at larger and smaller scales, a process [74] characterizes as *mesoscale resonance*, a generalization of the Baldwin effect. See that reference for details, broadly based on the Tuning Theorem.

12.3 Ecosystems Farm Organismal Development

The environmental and ecosystem farming of development may not always be benign.

Suppose we can operationalize and quantify degrees of both overfocus or inattentional blindness (IAB) and of overall structure or 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. Such measures need not themselves be convex in D.

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 [71], 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 resilience, 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 [70, 71] for discussions of ecological 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.

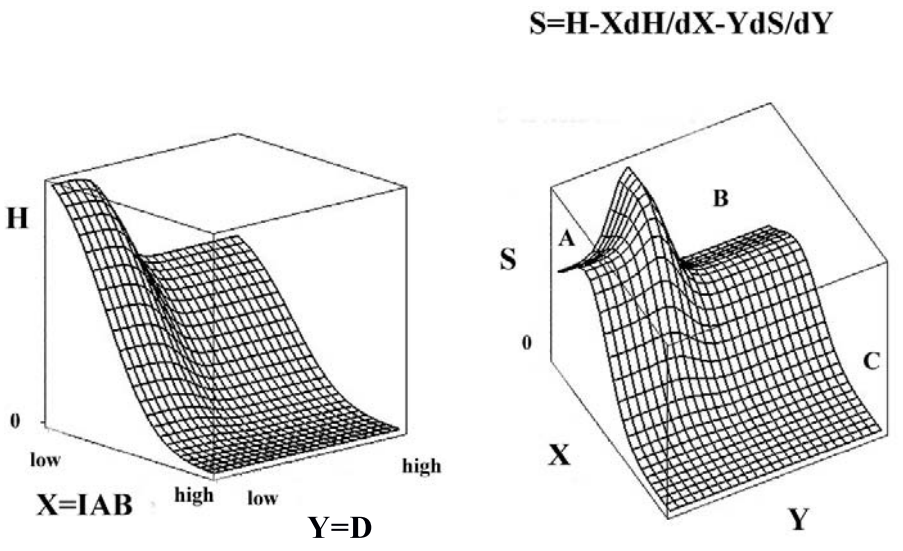

Fig. 1. a. Source uncertainty, H, of the dual information source of epigenetic cognition, as parametized by degrees of focus, $X = IAB$ and distortion, $Y = D$, between implementation and actual impact. Note the relatively gentle peak at low values of X, Y. Here H is generated as the product of two error 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.

In all of this, as in equation (19), 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' [72].

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 [71, 73]. 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. This might particularly occur under circumstances of rapid growth or long-term energy demand, since the peaks of figure 1 are inherently energy maxima by the duality between information source uncertainty and free energy density. During such a time the peaks of figure 1 might be relatively suppressed, and the system would become highly sensitive to perturbation, and to the onset of a subsequent pathological developmental trajectory.

To reiterate, then, during times of rapid growth, embryonic de- and re- methylation, and/or other high system demand, metabolic 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 or metabolic processes must receive more free energy than others, and these may then be more easily affected by external chemical, biological, or social stressors, or by simple stochastic variation. Something much like this has indeed become a standard perspective (e.g., [76]).

A structure trapped in region C might be said to suffer something much like what [80] 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. Reference [80] 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. [30, 72]). 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 (20) basically enables a kind of environmental catalysis to cognitive gene expression, in a sense closely similar to the arguments of Section 4. This is analogous to, but more general than, the 'mesoscale resonance' invoked by [74]: during critical periods, according to these models, environmental signals can have vast impact on developmental trajectory.

12.4 A Simple Probability Argument

Again, critical periods of rapid growth require energy, and by the homology between free energy density and cognitive information source uncertainty, that energy requirement may be in the context of a zero-sum game so that the barriers of figure 1 may be lowered by metabolic energy constraints or high energy demand. In particular the groupoid structure of equation (5) changes progressively as the organism develops, with new equivalence classes being added to $A = \cup\alpha$. If metabolic energy remains capped, then

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

must decrease with increase in α, i.e., with increase in the cardinality of A. Thus, for restricted K, barriers between different developmental paths must fall as the system becomes more complicated.

A precis of these results can be more formally captured using methods closely similar to recent algebraic geometry approaches to concurrent, i.e., highly parallel, computing [26, 37, 60].

13 Reconsidering Directed Homotopy: Shadows

Here we reconsider directed homotopy in a developmental context, as shadowed by critical developmental periods. First, we restrict the analysis to a two dimensional phenotype space, and begin development at some $\mathbf{S}_0$ as in figure 2.

If one requires temporal path dependence – no reverse development – then figure 2 shows two possible final states, $\mathbf{S}_1$ and $\mathbf{S}_2$, separated by a critical point $\mathbf{C}$ *that casts a path-dependent developmental shadow* in time. There are, consequently, two separate 'ways' of reaching a final state in this model. The $\mathbf{S}_i$ thus represent (relatively) static phenotypic expressions of the solutions to equation (19) that are, of themselves, highly dynamic information sources.

Elements of each 'way' can be transformed into each other by continuous deformation without crossing the impenetrable shadow cast by the critical period $\mathbf{C}$.

These ways are the equivalence classes defining the system's topological structure, a groupoid analogous to the fundamental homotopy group in spaces that admit of loops [50] rather than time-driven, one-way paths. That is, the closed loops needed for classical homotopy theory are impossible for this kind of system because of the 'flow of time' defining the output of an information source – one goes from $\mathbf{S}_0$ to some final state. The theory is thus one of *directed homotopy*, dihomotopy, and the central question revolves around the continuous deformation of paths in development space into one another, without crossing the shadow cast by the critical period $\mathbf{C}$. Reference [36] provides another introduction to the formalism.

Thus the external signals U of equation (20), as a catalytic mechanism, can define quite different developmental dihomotopies.

Such considerations suggest that a multitasking developmental process that becomes trapped in a particular pattern cannot, in general, expect to emerge from it in the absence of external forcing mechanisms or the stochastic resonance/mutational action of 'noise'. Emerging from such a trap involves large-scale topological changes, and this is the functional equivalent of a first order phase transition in a physical systems and requires energy.

The fundamental topological insight is that environmental context – the U in equation (20) – can be imposed on the 'natural' groupoids underlying massively parallel gene expression. This sort of behavior is, as noted in [71], central to ecosystem resilience theory.

Apparently the set of developmental manifolds, and its subsets of directed homotopy equivalence classes, formally classifies quasi-equilibrium states, and

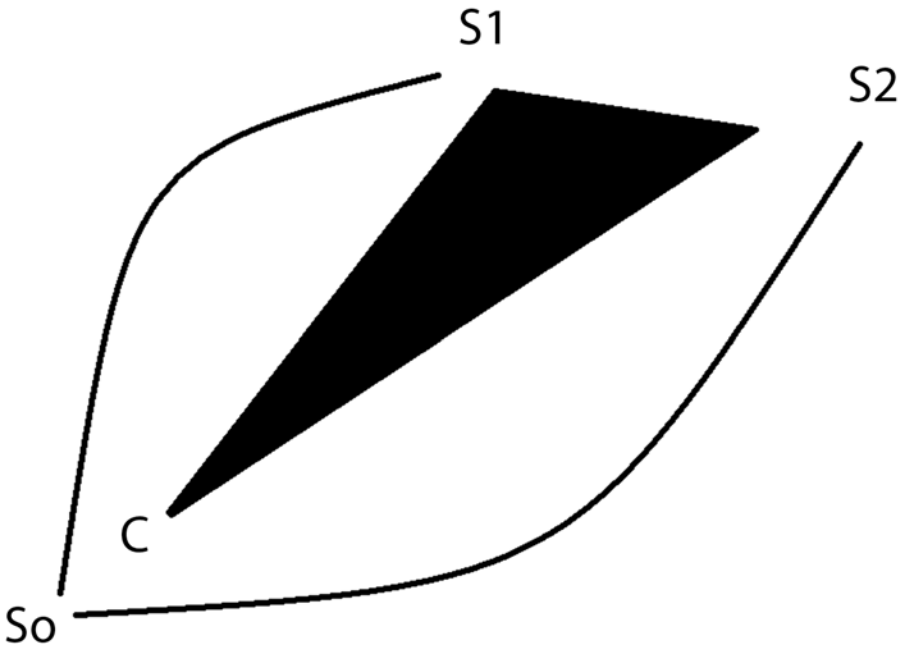

Fig. 2. Given an initial developmental state $\mathbf{S}_0$ and a critical period $\mathbf{C}$ casting a path-dependent developmental shadow, there are two directed homotopy equivalence classes of deformable paths leading, respectively, to final phenotype states $\mathbf{S}_1$ and $\mathbf{S}_2$ that are expressions of the highly dynamic information source solutions to equation (19). These equivalence classes define a topological groupoid on the developmental system.

thus characterizes the different possible developmental resilience modes. Some of these may be highly pathological.

Shifts between markedly different topological modes appear to be necessary effects of phase transitions, involving analogs to phase changes in physical systems.

It seems clear that both 'normal development' and possible pathological states can be represented as topological resilience/phase modes in this model, suggesting a real equivalence between difficulties in carrying out gene expression and its stabilization. This mirrors recent results on the relation between programming difficulty and system stability in highly parallel computing devices [70].

14 Epigenetic Programming of Artificial Systems for Biotechnology

Reference [72] examines how highly parallel 'Self-X' computing machines – self-programming, protecting, repairing, etc. – are inevitably coevolutionary in the sense of Section 9 above, since elements of a dynamic structural hierarchy always interact, an effect that will asymptotically dominate system behavior at great

scale. The 'farming' paradigm provides a model for programming such devices, that, while broadly similar to the liquid state machines of [51], differs in that convergence is to an information source, a systematic dynamic behavior pattern, rather than to a computed fixed 'answer'. As the farming metaphor suggests, stabilizing complex coevolutionary mechanisms appears as difficult as programming them. Sufficiently large networks of even the most dimly cognitive modules will become emergently coevolutionary, suggesting the necessity of 'second order' evolutionary programming that generalizes the conventional Nix/Vose models.

Although we cannot pursue the argument in detail here, very clearly such an approach to programming highly parallel coevolutionary machines – equivalent to deliberate epigenetic farming – should be applicable to a broad class of artificial biological systems/machines for which some particular ongoing behavior is to be required, rather than some final state 'answer'. Examples might include the manufacture, in a large sense, of a dynamic product, e.g., a chemical substance, anti-cancer or artificial immune search-and-destroy strategy, biological signal detection/transduction process, and so on.

Tunable epigenetic catalysis lowers an 'effective energy' associated with the convergence of a highly coevolutionary cognitive system to a final dynamic behavioral strategy. Given a particular 'farming' information source acting as the program, the behavior of the final state of interest will become associated with the lowest value of the free energy-analog, possibly calculable by optimization methods. If the retina-like rate distortion manifold has been properly implemented, a kind of converse to the no free lunch theorem, then this optimization procedure should converge to an appropriate solution, fixed or dynamic. Thus we invoke a synergism between the focusing theorem and a 'tunable epigenetic catalysis theorem' to raise the probability of an acceptable solution, particularly for a real-time system whose dynamics will be dominated by rate distortion theorem constraints.

The degree of catalysis needed for convergence in a real time system would seem critically dependent on the rate distortion function $R(D)$ or on its product with an acceptable reaction time, τ, that is, on there being sufficient bandwidth in the communication between a cognitive biological 'machine' and its embedding environment. If that bandwidth is too limited, or the available reaction time too short, then the system will inevitably freeze out into what amounts to a highly dysfunctional 'ground state'.

The essential point would seem to be a convergence between emerging needs in biotechnology and general strategies for programming coevolutionary computing devices.

15 Discussion and Conclusions

We have hidden the kind of massive calculations made explicit in [16, 17], 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., [68, 73, 74]).

Analogies with neural network studies suggest, however, intractable conceptual difficulties for spinglass-type models of gene expression and development dynamics, much as claimed by [57]. In spite of nearly a century of sophisticated neural network model studies – including elegant treatments like [66] – Atmanspacher [3] claims 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 a very few contemporary approaches, including that of [68], are worth mentioning, in his view.

Furthermore, [48] 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. The particular characterization of [48] is that 'neural possibility is not neural plausibility'.

Likewise, [8] concludes 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 [4, 5]. That is, gene expression and development are conditioned by 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.

Although epigenetic catalysis, as we have explored it here, might seem worthy of special focus, this would be a kind of intellectual optical illusion akin to inattentional blindness. Epigenetic catalysis is only one aspect of a general cognitive paradigm for gene expression, and this larger, and very complicated 'perceptual field' should remain the center of intellectual attention, rather than any single element of that field. This is to take, perhaps, an 'East Asian' rather than 'Western' perspective on the matter [69].

Developmental disorders, in a broad sense that must include comorbid mental and physical characteristics, emerge as pathological 'resilience' modes, in the sense of [71], a viewpoint from ecosystem theory quite similar to that of epigenetic epidemiology [32, 76]. 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 pattern converging on a dysfunctional phenotype.

The probability models of cognitive process presented here will lead, most fundamentally, to statistical tools based on the asymptotic limit theorems of information theory, in the same sense that the usual parametric statistics are based on the Central Limit Theorem. We have not, then, given 'a' model of development and its disorders in cognitive gene expression, but, rather, outlined a possible general strategy for fitting empirically-determined statistical models to real data, in precisely the sense that one would fit the usual parametric statistical models to normally distributed data.

The fitting of statistical models does not, of itself, perform scientific inference. 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.

One implication 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 simple models abducted from neural network studies or stochastic chemical reaction theory. Most centrally, however, currently popular (and fundable) reductionist approaches to understanding gene expression must eventually exhaust themselves in the same desert of sand-grain hyperparticularity that appears to have driven James Crick from molecular biology into consciousness studies, a field now mature enough to provide tools for use in the other direction.

Acknowledgments

The author thanks Dr. C. Guerrero-Bosagna and two anonymous reviewers for comments useful in revision.

References

1. Ash, R.: Information Theory. Dover Publications, New York (1990)
2. Atlan, H., Cohen, I.: Immune information, self-organization, and meaning. International Immunology 10, 711–717 (1998)
3. Atmanspacher, H.: Toward an information theoretical implementation of contextual conditions for consciousness. Acta Biotheoretica 54, 157–160 (2006)
4. Baars, B.: A Cognitive Theory of Consciousness. Cambridge University Press, New York (1988)
5. Baars, B.: Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. Progress in Brain Research 150, 45–53 (2005)
6. Backdahl, L., Bushell, A., Beck, S.: Inflammatory signalling as mediator of epigenetic modulation in tissue-specific chronic inflammation. The International Journal of Biochemistry and Cell Biology (2009), doi:10.1016/j.biocel.2008.08.023

7. Beck, C., Schlogl, F.: Thermodynamics of Chaotic Systems. Cambridge University Press, Cambridge (1995)
8. Bennett, M., Hacker, P.: Philosophical Foundations of Neuroscience. Blackwell Publishing, Malden (2003)
9. Bennett, C.: Logical depth and physical complexity. In: Herkin, R. (ed.) The Universal Turing Machine: A Half-Century Survey, pp. 227–257. Oxford University Press, Oxford (1988)
10. Bos, R.: Continuous representations of groupoids. arXiv:math/0612639 (2007)
11. Bossdorf, O., Richards, C., Pigliucci, M.: Epigenetics for ecologists. Ecology Letters 11, 106–115 (2008)
12. Britten, R., Davidson, E.: Gene regulation for higher cells: a theory. Science 165, 349–357 (1969)
13. Brown, R.: From groups to groupoids: a brief survey. Bulletin of the London Mathematical Society 19, 113–134 (1987)
14. Buneci, M.: Representare de Groupoizi. Editura Mirton, Timisoara (2003)
15. Cannas Da Silva, A., Weinstein, A.: Geometric Models for Noncommutative Algebras. American Mathematical Society, RI (1999)
16. Ciliberti, S., Martin, O., Wagner, A.: Robustness can evolve gradually in complex regulatory networks with varying topology. PLoS Computational Biology 3(2), e15 (2007)
17. Ciliberti, S., Martin, O., Wagner, A.: Innovation and robustness in complex regulatory gene networks. Proceedings of the National Academy of Sciences 104, 13591–13596 (2007)
18. Cohen, I.: Immune system computation and the immunological homunculus. In: Nierstrasz, O., Whittle, J., Harel, D., Reggio, G. (eds.) MoDELS 2006. LNCS, vol. 4199, pp. 499–512. Springer, Heidelberg (2006)
19. Cohen, I., Harel, D.: Explaining a complex living system: dynamics, multi-scaling, and emergence. Journal of the Royal Society: Interface 4, 175–182 (2007)
20. Cover, T., Thomas, J.: Elements of Information Theory. John Wiley and Sons, New York (1991)
21. Crews, D., McLachlan, J.A.: Epigenetics, evolution, endocrine disruption, health, and disease. Endocrinology 147, S4–S10 (2006)
22. Crews, D., Gore, A., Hsu, T., Dangleben, N., Spinetta, M., Schallert, T., Anway, M., Skinner, M.: Transgenerational epigenetic imprints on mate preference. Proceedings of the National Academy of Sciences 104, 5942–5946 (2007)
23. Dehaene, S., Naccache, L.: Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. Cognition 79, 1–37 (2001)
24. Dembo, A., Zeitouni, O.: Large Deviations: Techniques and Applications, 2nd edn. Springer, New York (1998)
25. Dias, A., Stewart, I.: Symmetry groupoids and admissible vector fields for coupled cell networks. Journal of the London Mathematical Society 69, 707–736 (2004)
26. Dretske, F.: The explanatory role of information. Philosophical Transactions of the Royal Society A 349, 59–70 (1994)
27. Emery, M.: Stochastic Calculus on Manifolds. Springer, New York (1989)
28. English, T.: Evaluation of evolutionary and genetic optimizers: no free lunch. In: Fogel, L., Angeline, P., Back, T. (eds.) Evolutionary Programming V: Proceedings of the Fifth Annual Conference on Evolutionary Programming, pp. 163–169. MIT Press, Cambridge (1996)
29. Erdos, P., Renyi, A.: On the evolution of random graphs (1960); reprinted in The Art of Counting, pp. 574–618 (1973), and in Selected Papers of Alfred Renyi, pp. 482–525 (1976)

30. Ficici, S., Milnik, O., Pollak, J.: A game-theoretic and dynamical systems analysis of selection methods in coevolution. IEEE Transactions on Evolutionary Computation 9, 580–602 (2005)
31. Feynman, R.: Lectures on Computation. Westview Press, New York (2000)
32. Foley, D., Craid, J., Morley, R., Olsson, C., Dwyer, T., Smith, K., Saffery, R.: Prospects for epigenetic epidemiology. American Journal of Epidemiology 169, 389–400 (2009)
33. Gilbert, S.: Mechanisms for the environmental regulation of gene expression: ecological aspects of animal development. Journal of Bioscience 30, 65–74 (2001)
34. Glazebrook, J.F., Wallace, R.: Small worlds and red queens in the global workspace: an information-theoretic approach. Cognitive Systems Reserch (2009), doi:10.1016/j.cogsys.2009.01.002
35. Golubitsky, M., Stewart, I.: Nonlinear dynamics and networks: the groupoid formalism. Bulletin of the American Mathematical Society 43, 305–364 (2006)
36. Goubault, E., Raussen, M.: Dihomotopy as a tool in state space analysis. In: Rajsbaum, S. (ed.) LATIN 2002. LNCS, vol. 2286, pp. 16–37. Springer, Heidelberg (2002)
37. Goubault, E.: Some geometric perspectives on concurrency theory. Homology, Homotopy, and Applications 5, 95–136 (2003)
38. Gould, S.: The Structure of Evolutionary Theory. Harvard University Press, Cambridge (2002)
39. Guerrero-Bosagna, C., Sabat, P., Valladares, L.: Environmental signaling and evolutionary change: can exposure of pregnant mammals to environmental estrogens lead to epigenetically induced evolutionary changes in embryos? Evolution and Development 7, 341–350 (2005)
40. Holling, C.: Cross-scale morphology, geometry and dynamicsl of ecosystems. Ecological Monographs 41, 1–50 (1992)
41. Jablonka, E., Lamb, M.: Epigenetic Inheritance and Evolution: The Lamarckian Dimension. Oxford University Press, Oxford (1995)
42. Jablonka, E., Lamb, M.J.: Epigenetic inheritance in evolution. Journal of Evolutionary Biology 11, 159–183 (1998)
43. Jablonka, E.: Epigenetic epidemiology. International Journal of Epidemiology 33, 929–935 (2004)
44. Jaeger, J., Surkova, S., Blagov, M., Janssens, H., Kosman, D., Kozlov, K., Manu, M., Myasnikova, E., Vanario-Alonso, C., Samsonova, M., Sharp, D., Reintiz, J.: Dynamic control of positional information in the early Drosophila embryo. Nature 430, 368–371 (2004)
45. Jaenisch, R., Bird, A.: Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nature Genetics Supplement 33, 245–254 (2003)
46. Kastner, M.: Phase transitions and configuration space topology. ArXiv cond-mat/0703401 (2006)
47. Khinchin, A.: Mathematical Foundations of Information Theory. Dover, New York (1957)
48. Krebs, P.: Models of cognition: neurological possibility does not indicate neurological plausibility. In: Bara, B., Barsalou, L., Bucciarelli, M. (eds.) Proceedings of CogSci 2005, Stresa, Italy, pp. 1184–1189 (2005), http://cogprints.org/4498/
49. Landau, L., Lifshitz, E.: Statistical Physics, Part I, 3rd edn., Part I. Elsevier, New York (2007)
50. Lee, J.: Introduction to topological manifolds. Springer, New York (2000)

51. Maas, W., Natschlager, T., Markram, H.: Real-time computing without stable states: a new framework for neural computation based on perturbations. Neural Computation 14, 2531–2560 (2002)
52. Matsumoto, Y.: An Introduction to Morse Theory. American Mathematical Society, Providence (2002)
53. Maturana, H.R., Varela, F.J.: Autopoiesis and Cognition. Reidel Publishing Company, Dordrecht (1980)
54. Maturana, H.R., Varela, F.J.: The Tree of Knowledge. Shambhala Publications, Boston (1992)
55. McCauly, J.: Chaos, Dynamics, and Fractals. Cambridge Nonlinear Science Series, Cambridge, UK (1994)
56. Mjolsness, E., Sharp, D., Reinitz, J.: A connectionist model of development. Journal of Theoretical Biology 152, 429–458 (1991)
57. O'Nuallain, S.: Code and context in gene expression, cognition, and consciousness. In: Barbiere, M. (ed.) The Codes of Life: The Rules of Macroevolution, ch. 15, pp. 347–356. Springer, New York (2008)
58. O'Nuallain, S., Strohman, R.: Genome and natural language: how far can the analogy be extended? In: Witzany, G. (ed.) Proceedings of Biosemiotics. Tartu University Press, Umweb (2007)
59. Pettini, M.: Geometry and Topology in Hamiltonian Dynamics and Statistical Mechanics. Springer, New York (2007)
60. Pratt, V.: Modeling concurrency with geometry. In: Proceedings of the 18th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, pp. 311–322 (1991)
61. Reinitz, J., Sharp, D.: Mechanisms of even stripe formation. Mechanics of Development 49, 133–158 (1995)
62. Scherrer, K., Jost, J.: The gene and the genon concept: a functional and information-theoretic analysis. Molecular Systems Biology 3, 87–95 (2007)
63. Scherrer, K., Jost, J.: Gene and genon concept: coding versus regulation. Theory in Bioscience 126, 65–113 (2007)
64. Sharp, D., Reinitz, J.: Prediction of mutant expression patterns using gene circuits. BioSystems 47, 79–90 (1998)
65. Skierski, M., Grundland, A., Tuszynski, J.: Analysis of the three-dimensional time-dependent Landau-Ginzburg equation and its solutions. Journal of Physics A (Math. Gen.) 22, 3789–3808 (1989)
66. Toulouse, G., Dehaene, S., Changeux, J.: Spin glass model of learning by selection. Proceedings of the National Academy of Sciences 83, 1695–1698 (1986)
67. Turner, B.: Histone acetylation and an epigeneticv code. Bioessays 22, 836–845 (2000)
68. Wallace, R.: Consciousness: A Mathematical Treatment of the Global Neuronal Workspace Model. Springer, New York (2005)
69. Wallace, R.: Culture and inattentional blindness. Journal of Theoretical Biology 245, 378–390 (2007)
70. Wallace, R.: Toward formal models of biologically inspired, highly parallel machine cognition. International Journal of Parallel, Emergent, and Distributed Systems 23, 367–408 (2008)
71. Wallace, R.: Developmental disorders as pathological resilience domains. Ecology and Society 13(1), 29 (2008), http://www.ecologyandsociety.org/vol13/iss1/art29/
72. Wallace, R.: Programming coevolutionary machines: the emerging conundrum. International Journal of Parallel, Emergent, and Distributed Systems (in press, 2009)

73. Wallace, R., Fullilove, M.: Collective Consciousness and its Discontents: Institutional Distributed Cognition, Racial Policy, and Public Health in the United States. Springer, New York (2008)
74. Wallace, R., Wallace, D.: Punctuated equilibrium in statistical models of generalized coevolutionary resilience: how sudden ecosystem transitions can entrain both phenotype expression and Darwinian selection. In: Priami, C. (ed.) Transactions on Computational Systems Biology IX. LNCS (LNBI), vol. 5121, pp. 23–85. Springer, Heidelberg (2008)
75. Wallace, R.G., Wallace, R.: Evolutionary radiation and the spectrum of consciousness. Consciousness and Cognition (2009), doi:10.1016/j.concog.2008.12.002
76. Waterland, R., Michels, K.: Epigenetic epidemiology of the developmental origins hypothesis. Annual Reviews of Nutrition 27, 363–388 (2007)
77. Weaver, I.: Epigenetic effects of glucocorticoids. Seminars in Fetal and Neonatal Medicine (2009), doi:10.1016/j.siny.2008.12.002
78. Weinstein, A.: Groupoids: unifying internal and external symmetry. Notices of the American Mathematical Association 43, 744–752 (1996)
79. West-Eberhard, M.: Developmental plasticity and the origin of species differences. Proceedings of the National Academy of Sciences 102, 6543–6549 (2005)
80. Wiegand, R.: An analysis of cooperative coevolutionary algorithms. PhD Thesis, George Mason University (2003)
81. Wolpert, D., Macready, W.: No free lunch theorems for search. Santa Fe Institute, SFI-TR-02-010 (1995)
82. Wolpert, D., Macready, W.: No free lunch theorems for optimization. IEEE Transactions on Evolutionary Computation 1, 67–82 (1997)
83. Wymer, C.R.: Structural nonlinear continuous-time models in econometrics. Macroeconomic Dynamics 1, 518–548 (1997)
84. Zhu, R., Rebirio, A., Salahub, D., Kaufmann, S.: Studying genetic regulatory networks at the molecular level: delayed reaction stochastic models. Journal of Theoretical Biology 246, 725–745 (2007)

16 Mathematical Appendix

16.1 The Shannon-McMillan Theorem

According to the structure of the underlying language of which a message is a particular expression, some messages are more 'meaningful' than others, that is, are in accord with the grammar and syntax of the language. The Shannon-McMillan or Asymptotic Equipartition Theorem, describes how messages themselves are to be classified.

Suppose a long sequence of symbols is chosen, using the output of a random variable X, so that an output sequence of length n, with the form

$$x_n = (\alpha_0, \alpha_1, ..., \alpha_{n-1})$$

has joint and conditional probabilities

$$P(X_0 = \alpha_0, X_1 = \alpha_1, ..., X_{n-1} = \alpha_{n-1})$$

$$P(X_n = \alpha_n | X_0 = \alpha_0, ..., X_{n-1} = \alpha_{n-1}).$$

Using these probabilities we may calculate the conditional uncertainty

$$H(X_n|X_0, X_1, ..., X_{n-1}).$$

The uncertainty of the *information source*, $H[\mathbf{X}]$, is defined as

$$H[\mathbf{X}] = \lim_{n\to\infty} H(X_n|X_0, X_1, ..., X_{n-1}). \tag{21}$$

In general

$$H(X_n|X_0, X_1, ..., X_{n-1}) \leq H(X_n).$$

Only if the random variables X_j are all stochastically independent does equality hold. If there is a maximum n such that, for all $m > 0$

$$H(X_{n+m}|X_0, ..., X_{n+m-1}) = H(X_n|X_0, ..., X_{n-1}),$$

then the source is said to be of *order* n. It is easy to show that

$$H[\mathbf{X}] = \lim_{n\to\infty} \frac{H(X_0, ...X_n)}{n+1}.$$

In general the outputs of the $X_j, j = 0, 1, ..., n$ are *dependent*. That is, the output of the communication process at step n depends on previous steps. Such serial correlation, in fact, is the very structure which enables most of what is done in this paper.

Here, however, the processes are all assumed statble in time, that is, the probabilities and serial correlations do not change in time, and the system is *stationary*.

A very broad class of such self-correlated, stationary, information sources, the so-called *ergodic* sources for which the long-run relative frequency of a sequence converges stochastically to the probability assigned to it, have a particularly interesting property:

It is possible, in the limit of large n, to divide all sequences of outputs of an ergodic information source into two distinct sets, S_1 and S_2, having, respectively, very high and very low probabilities of occurrence, with the source uncertainty providing the splitting criterion. In particular the Shannon-McMillan Theorem states that, for a (long) sequence having n (serially correlated) elements, the number of 'meaningful' sequences, $N(n)$ – those belonging to set S_1 – will satisfy the relation

$$\frac{\log[N(n)]}{n} \approx H[\mathbf{X}]. \tag{22}$$

More formally,

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

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

$$= \lim_{n\to\infty} \frac{H(X_0, ..., X_n)}{n+1}. \tag{23}$$

Using the internal structures of the information source permits *limiting attention only to high probability 'meaningful' sequences of symbols.*

16.2 The Rate Distortion Theorem

The Shannon-McMillan Theorem can be expressed as the 'zero error limit' of the Rate Distortion Theorem [20, 24] which defines a splitting criterion that identifies high probability pairs of sequences. We follow closely the treatment of [20].

The origin of the problem is the question of representing one information source by a simpler one in such a way that the least information is lost. For example we might have a continuous variate between 0 and 100, and wish to represent it in terms of a small set of integers in a way that minimizes the inevitable distortion that process creates. Typically, for example, an analog audio signal will be replaced by a 'digital' one. The problem is to do this in a way which least distorts the *reconstructed* audio waveform.

Suppose the original stationary, ergodic information source Y with output from a particular alphabet generates sequences of the form

$$y^n = y_1, ..., y_n.$$

These are 'digitized,' in some sense, producing a chain of 'digitized values'

$$b^n = b_1, ..., b_n,$$

where the b-alphabet is much more restricted than the y-alphabet.

b^n is, in turn, *deterministically retranslated* into a reproduction of the original signal y^n. That is, each b^m is mapped on to a unique n-length y-sequence in the alphabet of the information source Y:

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

Note, however, that many y^n sequences may be mapped onto the *same* retranslation sequence $\hat{y}^n$, so that information will, in general, be lost.

The central problem is to explicitly minimize that loss.

The retranslation process defines a new stationary, ergodic information source, $\hat{Y}$.

The next step is to define a *distortion measure*, $d(y, \hat{y})$, which compares the original to the retranslated path. For example the *Hamming distortion* is

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

$$d(y, \hat{y}) = 0, y = \hat{y}. \tag{24}$$

For continuous variates the *Squared error distortion* is

$$d(y, \hat{y}) = (y - \hat{y})^2. \tag{25}$$

There are many possibilities.

The distortion between paths y^n and $\hat{y}^n$ is defined as

$$d(y^n, \hat{y}^n) = \frac{1}{n} \sum_{j=1}^{n} d(y_j, \hat{y}_j). \tag{26}$$

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

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

The *average distortion* is defined as

$$D = \sum_{y^n} p(y^n) d(y^n, \hat{y}^n). \tag{27}$$

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

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

If there is no uncertainty in Y given the retranslation $\hat{Y}$, then no information is lost.

In general, this will not be true.

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

$$R(D) = \min_{p(y,\hat{y}); \sum_{(y,\hat{y})} p(y) p(y|\hat{y}) d(y,\hat{y}) \leq D} I(Y, \hat{Y}). \tag{28}$$

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 maximum achievable rate of information transmission which does not exceed the distortion* D. See [20, 24] details.

More to the point, however, is the following: Pairs of sequences $(y^n, \hat{y}^n)$ can be defined as *distortion typical*; that is, for a given average distortion D, defined in terms of a particular measure, pairs of sequences can be divided into two sets, a high probability one containing a relatively small number of (matched) pairs with $d(y^n, \hat{y}^n) \leq D$, and a low probability one containing most pairs. As $n \to \infty$, the smaller set approaches unit probability, and, for those pairs,

$$p(y^n) \geq p(\hat{y}^n|y^n) \exp[-nI(Y, \hat{Y})]. \tag{29}$$

Thus, roughly speaking, $I(Y, \hat{Y})$ embodies the splitting criterion between high and low probability pairs of paths.

For the theory of interacting information sources, then, $I(Y, \hat{Y})$ can play the role of H in the dynamic treatment above.

The rate distortion function can actually be calculated in many cases by using a Lagrange multiplier method – see Section 13.7 of [20].

16.3 Groupoids

Basic ideas. Following [78] closely, a groupoid, G, is defined by a base set A upon which some mapping – a morphism – can be defined. Note that not all possible pairs of states (a_j, a_k) in the base set A can be connected by such a morphism. Those that can define the groupoid element, a morphism $g = (a_j, a_k)$ having the natural inverse $g^{-1} = (a_k, a_j)$. Given such a pairing, it is possible to define 'natural' end-point maps $\alpha(g) = a_j, \beta(g) = a_k$ from the set of morphisms G into A, and a formally associative product in the groupoid $g_1 g_2$ provided $\alpha(g_1 g_2) = \alpha(g_1), \beta(g_1 g_2) = \beta(g_2)$, and $\beta(g_1) = \alpha(g_2)$. Then the product is defined, and associative, $(g_1 g_2) g_3 = g_1 (g_2 g_3)$.

In addition, there are natural left and right identity elements λ_g, ρ_g such that $\lambda_g g = g = g \rho_g$ [78].

An orbit of the groupoid G over A is an equivalence class for the relation $a_j \sim G a_k$ if and only if there is a groupoid element g with $\alpha(g) = a_j$ and $\beta(g) = a_k$. Following [15], 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 (α, β) 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 [78] explains, a groupoid G is a topological groupoid over a base space X if G and X are topological spaces and α, β 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. It is possible to 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 [78] points out, the morphism (α, β) 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 [13] 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_λ, and ab is then just the product in the group G_λ. There is an identity 1_λ for each $\lambda \in \Lambda$. The maps α, β coincide and map G_λ to λ, $\lambda \in \Lambda$.
>
> 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 $x \in X$...

[78] 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, [35].

Global and local symmetry groupoids. Here we follow [78] fairly closely, using the 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 Γ 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 in [78], much is lost in this coarse-graining, in particular the same symmetry group would arise if we replaced X entirely by the lattice Λ of corner points. Γ 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 Γ 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, γ, 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.

16.4 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 [74], 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 [46, 59].

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 [59], 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 [52, 59] for details.

16.5 Generalized Onsager Theory

Understanding the time dynamics of groupoid-driven information systems away from phase transition critical points 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 (GFE) of a biological process is parametized by some vector of quantities $\mathbf{K} = (K_1, ..., K_m)$, then, in analogy with nonequilibrium thermodynamics, gradients in the K_j of the *disorder*, defined as

$$S_G = F_G(\mathbf{K}) - \sum_{j=1}^{m} K_j \partial F_G / \partial K_j \tag{30}$$

become of central interest.

Equation (30) 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

$$dK_j/dt = \sum_i L_{j,i} \partial S_G / \partial K_i, \tag{31}$$

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 (30) and (31) 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 [74]. 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 (31) might be rewritten as

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

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

$$dK_t^j = L^j(t, \mathbf{K})dt + \sigma^j(t, \mathbf{K})dB_t, \tag{32}$$

where the L^j and σ^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 [27]. 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 [59], and this phenomenological Onsager treatment would likely be much enriched by explicit adoption of a Morse theory perspective.

16.6 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

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

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

$$H(X, Y) = -\sum_{j=1}^{M} \sum_{k=1}^{L} P_{j,k} \log(P_{j,k}). \tag{33}$$

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

$$H(Y|X) = -\sum_{j=1}^{M}\sum_{k=1}^{L} P_{j,k}\log[P(y_k|x_j)]. \tag{34}$$

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

$$I(X|Y) = H(X) - H(X|Y) = H(X) + H(Y) - H(X,Y) \tag{35}$$

See, for example, [1, 20, 47] 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

$$C = \max_{P(X)} I(X|Y) \tag{36}$$

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, [1, 20, 47] 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

$$C^* = \max_{P(Y),P(Y|X)} I(X|Y). \tag{37}$$

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

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

that 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

$$C^* = H(X),$$

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.