

Version 8.1

Punctuated equilibrium in statistical models of generalized coevolutionary resilience: How sudden ecosystem transitions can entrain both phenotype expression and Darwinian selection

Rodrick Wallace, Ph.D.

The New York State Psychiatric Institute

Deborah N. Wallace, Ph.D.

Consumers Union*

February 2, 2008

Abstract

We argue that mesoscale ecosystem resilience shifts akin to sudden phase transitions in physical systems can entrain similarly punctuated events of gene expression on more rapid time scales, and, in part through such means, slower changes induced by selection pressure, triggering punctuated equilibrium Darwinian evolutionary transitions on geologic time scales. The approach reduces ecosystem, gene expression, and Darwinian genetic dynamics to a least common denominator of information sources interacting by crosstalk at markedly differing rates. Pettini's 'topological hypothesis', via a homology between information source uncertainty and free energy density, generates a statistical model of sudden coevolutionary phase transition based on the Rate Distortion and Shannon-McMillan Theorems of information theory which links all three levels. Holling's (1992) extended keystone hypothesis regarding the particular role of mesoscale phenomena in entraining both slower and faster dynamical structures produces the result. A main theme is the necessity of a cognitive paradigm for gene expression, mirroring I. Cohen's cognitive approach to immune function. Invocation of the necessary conditions imposed by the asymptotic limit theorems of communication theory enables us to penetrate one layer more deeply before needing to impose a phenomenological system of 'Onsager relation' recursive coevolutionary stochastic differential equations. Extending the development to second order via a large deviations argument may permit modeling the influence of human cultural structures on ecosystems.

Key words: coevolution, cognitive paradigm, gene expression, information theory, phase transition, punctuated equilibrium, rate distortion, source coding,

1 Introduction

In the early twentieth century, evolutionists debated whether change occurred gradually or as a result of massive catastrophes. At that time, the gradualists mainly won the debate, and the catastrophists were marginalized. Speciation was viewed as a gradual process of incremental changes in response to incremental environmental challenges. However, Eldredge and Gould (Gould, 2002), after study of the fossil record, concluded that speciation occurred suddenly. Species appeared in the fossil record, remained in the fossil record largely unchanged, and then disappeared. There was little or no evidence of gradual incremental changes that led to speciation. Eldredge and Gould called the process that they saw "punctuated equilibrium", a term that referred to the sudden changes (punctuations) and the quiet interims (equilibria), a combination of gradualism and catastrophism. Eldredge and Gould published their initial findings in the late 1960's-early 1970's.

At about the same time C. S. Holling published the ecosystem equivalent of punctuated equilibrium, namely ecosystem resilience theory (Holling, 1973). Ecosystem resilience theory views each ecosystem as in a quasi-equilibrium normally. As the ecosystem receives various impacts, it shows no obvious changes in structure or function but the relationships between the species become tighter as the impacts chip away at the more delicate peripheral relationships. Finally, either a more intense impact occurs or the aggregated impacts over time shatter so many loose relationships that the remaining ones become brittle and shatter. The ecosystem then flips into a different dynamic domain, a different quasi-equilibrium with different structure and function. Examples of domain change include natural ones such as change of forest into prairie after drought and major forest fires in areas marginal for forests and unnatural ones such as eutrophication of waterbodies from agricultural runoff and discharge of urban wastewater.

Ecosystems provide the niches for species. If ecosystems flip suddenly into different configurations, then species are

*Address correspondence to: R. Wallace, PISCS Inc., 549 W. 123 St., Suite 16F, New York, NY, 10027. Telephone (212) 865-4766, rd-wall@ix.netcom.com. Affiliations are for identification only.

confronted with sudden changes. Perhaps Holling's theory provides an explanation for Eldredge and Gould's reading of the fossil record. Besides the fossil record, the climatological and geological records also show major changes in temperature, atmospheric composition, and geological processes such as volcanoes, earthquakes, and movements of tectonic plates. These, of course, form the macroscale of ecosystems. Local topography, geology, hydrology, and microclimate lead to ecological niches. Organisms by their activities modify their own niches and the niches of other organisms (Lewontin, 1993). These localized processes form the microscale of ecosystems. Landscape processes such as wildfires which spread and affect large numbers of niches form the mesoscale (Holling, 1992).

Niches within ecosystems select for the fittest phenotypes for them. Not all genes of an organism are expressed. Thus, the genetic variability within the population of a particular niche may be far greater than the relatively uniform phenotype presented to the examining ecologist. If a characteristic may potentially be influenced by multiple genes, the niche may select for a phenotype consonant with the expression of only a single gene or only a couple of them. The species in the fossil record reflect only phenotypes, not the full range of genetic variability. Ecosystem domain shift would lead to selection for different phenotypes. Those individuals with the genes that can express these newly 'preferred' phenotypes will supplant the old phenotypes in the new ecosystem configuration. Voila! Apparent speciation! As time hardens the new ecoconfiguration, the genetic composition of the 'new species' will indeed shift toward the old alleles and new mutations expressing the new phenotypes most efficiently, and true speciation occurs. The wonderful book *Animal Traditions* (Avital and Jablonka, 2000) describes in detail how behavioral phenotypes end up encoded in the genome.

Ecosystem theorists now recognize several different kinds of resilience (e.g. Gunderson, 2000). The first, termed 'engineering resilience', since it is particularly characteristic of machines and man-machine interactions, involves the rate at which a disturbed system returns to a presumed single, stable, equilibrium condition, following perturbation. From that limited perspective, a resilient system is one which quickly reestablishes its one stable state.

Not many biological (or social) phenomena are resilient in this simplistic sense.

Holling's (1973) particular contribution was to recognize that sudden transitions between different, at best quasi-stable, domains of relation among ecosystem variates were possible, i.e. that more than one 'stable' state was possible for real ecosystems. Gunderson (2000) puts the matter as follows:

“One key distinction between these two types of resilience lies in assumptions regarding the existence of multiple [quasi-]stable states. If it is assumed that only one stable state exists or can be designed to exist, then the only possible definition and measures for resilience are near equilibrium ones – such as characteristic return time... The concept of eco-

logical resilience presumes the existence of multiple stability domains and the tolerance of the system to perturbations that facilitate transitions among stable states. Hence, ecological resilience refers to the width or limit of a stability domain and is defined by the magnitude of disturbance that a system can absorb before it changes stable states... The presence of multiple [quasi-]stable states and transitions among them [has] been [empirically] described in a [large] range of ecological systems...”

An often presumed difference between 'natural' and human-dominated ecosystems is, however, the particular role of both individual and collective cognitive action: human ecosystems are not simply reflex-driven, but can oppose reasoned, organized, responses to perturbation, as indeed can many individually intelligent animals and collectively cognitive animal groupings.

This paper presents a formal treatment of ecosystem resilience which can be extended upward in a highly natural manner to include a large class of explicitly cognitive individual and collective phenomena, (and indeed, to certain cognitive biological submodules including gene expression). The analysis indeed recovers most of classical ecosystem resilience theory, but, somewhat remarkably, can also be expanded downward in the sense of Adami et al. (2000) and Ofria et al. (2003) to encompass the Darwinian system of genetic heritage. The three levels – genetic, ecosystem, cognitive – can then be crosslinked into a class of statistical coevolutionary models inherently subject to multiple forms of phase transition – punctuated equilibria – in the sense of Pettini's (2007) topological hypothesis.

The means for this involve describing genetic heritage, ecosystem dynamics, and cognitive phenomena in terms of a least common denominator – as information sources – and then invoking the asymptotic limit theorems of communications theory to create a spectrum of necessary conditions statistical models in similar spirit to the construction of regression models which are constrained by the Central Limit and related theorems. A hierarchy of coevolutionary structure emerges when information sources are linked by crosstalk to become each other's principal environments, operating at different rates. By Holling's (1992) arguments, the mesoscale of the ecosystem entrains both slower and faster phenomena via punctuation.

We shall find quasi-stable resilience states represent a hierarchy of modes, including topological manifolds associated with equivalence classes of information sources, and equivalence classes of paths within individual manifolds. Equivalence class arguments can be expanded in a natural manner using a topological groupoid formalism which, however we will not pursue here. Wallace and Fullilove (2008) and Glazebrook and Wallace (2008) provide details.

Gunderson (2007) has suggested, as an example, the spruce budworm ecosystem. Following Fleming and Shoemaker (1992),

“The spruce budworm, *Choristoneura fumifer-*

ana... is the most damaging defoliator in North America's boreal forests. This native insect attacks balsam fir... and spruce... and can kill almost all the trees in dense, mature stands of fir during uncontrolled outbreaks. These periods of high population densities typically last 5-15 yr; between outbreaks the budworm can remain rare for 20-60 yr.

The economic impact of budworm outbreaks has prompted a variety of innovative modeling efforts... The most comprehensive models were developed... as 'laboratory worlds' for exploring the consequences of ecological hypotheses... and in particular alternative management strategies... [Many of these models] have roots in Holling's... seminal notions of resilience..."

In short, resilience theory applied here generates a system having two basic quasi-stable modes: [1] the 'normal' state of the forest in the absence of budworm eruptions, and [2], relatively brief but devastating budworm outbreaks which are self-limiting and relatively infrequent. This two-fold classification of states and possible dynamic paths represents the fundamental topology (homotopy) of the dynamic manifold. Central questions involve the transition between these two topological modes.

Imagine, however, a large scale perturbation like a major series of crown forest fires or other catastrophes driven, for example, by climate change (e.g. Volney and Fleming, 2007). Under such conditions there are no large-scale forests of balsam fir and spruce left to host budworm outbreaks: A new ecosystem will emerge and constitute a far different dynamic manifold, having quite different quasi-stable topological equivalence classes of states or dynamic system paths. Spruce budworms must suddenly evolve to graze on new plant species, or become extinct.

It is this hierarchy which we will attempt to model, in the context of inevitable topological phase transitions giving a punctuated equilibrium to patterns of broadly coevolutionary interacting between information sources.

To reiterate, the strategy is to express genetic heritage, ecosystem structure, and cognition according to an information theoretic least common denominator, gaining generality at the expense of specificity. This structure will then be extended through topological punctuation triggered by increasing crosstalk between these basic, now similar, if indeed simplistic, elements.

This is not, overall, a particularly new approach, although the use of modern mathematical tools to generate phase change is new.

For example Jimenez-Montano (1989) describes the well-known language metaphor of theoretical biology as follows:

"In his epilogue to the fourth volume of papers issuing from the IUBS Symposia at Villa Serbellion, inspired by papers of Pattee (1972) and Thom (1972), among others, Waddington concluded that...[in] situations which arise when there

is mutual interaction between the complexity-out-of-simplicity (self-assembly), and simplicity-out-of-complexity (self-organization) processes are...to be discussed most profoundly at the present time with the help of the analogy of language...

...Waddington [thus] suggested that it is a *language* that may become a paradigm for [a Theory of General Biology], but a language in which basic sentences are programs, not statements..."

Jimenez-Montano goes on to summarize the elegant, if rather limited, applications of formal language theory to molecular genetics, quoting Pattee (1972) to the effect that a molecule becomes a message only in the context of larger constraints which he has called a 'language.'

Perhaps the essential defining constraint of any language is that not all possible sequences of symbols are meaningful, that is, systems of symbols must be arranged according to grammar and syntax in the context of even higher levels of structure, a matter to which we will repeatedly return.

The full implications of the Rate Distortion Theorem (RDT), and its zero-error limit, the Shannon-McMillan Theorem (SMT), otherwise known as the Asymptotic Equipartition Theorem, for Waddington's program seem to have been largely overlooked. A development based on these theorems is useful, in particular, for examining the ways in which vastly different 'languages' can interact to create complicated structures. The central problem then becomes the characterization of ecosystems, the genetic heritage, and the cognitive processes of biology and social structure as such (generalized) languages.

Again, our approach, based on robust application of the RDT and SMT, has a 'necessary conditions' flavor recognizably similar to that surrounding the role of the Central Limit Theorem (CLT) in parametric statistics and regression theory. Regardless of the probability distribution of a particular stochastic variate, the CLT ensures that long sums of independent realizations of that variate will follow a Normal distribution. Application of the CLT gives regression theory, whose empirical models have been central to much of practical scientific research. Analogous constraints exist on the behavior of the information sources – both independent and interacting – that we will infer characterize ecosystems, and these are described by the RDT/SMT. Imposition of phase transition formalism from statistical physics onto the SMT, in the spirit of the Large Deviations Program of applied probability, permits concise and unified description of evolutionary, ecosystem, and cognitive 'learning plateaus' which, in the evolutionary case, are interpreted as evolutionary punctuation (e.g. Wallace, 2002a, b), and for ecosystems as resilience domains. This perspective provides a 'natural' means of exploring punctuated processes in the effects of habitat disruption on individuals, their social groups, their interactions, and indeed on the cognitive subprocesses of individual animal biology like immune function, tumor control, and gene expression.

Indeed, generating a 'cognitive paradigm' for gene expres-

sion, much in the spirit of Cohen’s (2000) cognitive paradigm for immune function, appears to be one of the more interesting subplots of this monograph.

Another recurring subplot is that of punctuation – sudden biological (or social) phase transition. Punctuated biological processes are found across temporal scales. Evolutionary punctuation is a modern extension of Darwinian evolutionary theory that accounts for the relative stability of a species’ fossil record between the time it first appears and its extinction (e.g. Eldredge, 1985; Gould, 2002). Species appear (relatively) suddenly on a geologic time scale, persist (again, relatively) unchanged for a fairly long time, and then disappear suddenly, again on a geologic time scale. Evolutionary process is vastly speeded up in tumorigenesis, which nonetheless seems amenable to similar analysis (e.g. Wallace et al., 2003).

In any event, the statistical models we generate here should be viewed as having primary scientific utility in the empirical comparison of different structures under similar conditions, or of similar structures under different conditions.

Our spectrum of models, then, as in the relation of the CLT to parametric statistical inference, is almost independent of the detailed structure of the interacting information sources inevitably associated with genetic, cognitive, and ecosystem processes, important as such structure may be in other contexts. This finesses some of the profound ambiguities associated with dynamic systems theory and deterministic chaos treatments in which the existence of dynamic attractors depends on very specific kinds of differential equation models akin to those used to describe test-tube population dynamics, chemical processes, or physical systems of weights-on-springs. Cognitive and natural ecosystem phenomena, like ecosystems, are neither well-stirred Erlenmeyer flasks of reacting agents, nor distorted, noisy, clocks, and the application of nonlinear dynamic systems theory to cognition, evolution, or ecology will, in all likelihood, ultimately be found to involve little more than hopeful metaphor. Indeed, as shown below, much of nonlinear dynamics can be subsumed within a larger information theory formalism through symbolic dynamics coarse-graining discretization techniques (e.g. Beck and Schlogl, 1995).

While idiosyncratic approaches analogous to nonparametric models in statistical theory may be required in some cases, the relatively straightforward formalism we develop here, like its cousins of parametric statistics and regression theory, may well robustly capture the essence of much relevant phenomena.

2 Ecosystems as information sources

2.1 Coarse-graining a simple model

We begin with a simplistic picture of an elementary predator/prey ecosystem which, nonetheless, provides a useful pedagogical starting point. Let X represent the appropriately scaled number of predators, Y the scaled number of prey, t the time, and ω a parameter defining their interaction. The

model assumes that the ecologically dominant relation is an interaction between predator and prey, so that

$$dX/dt = \omega Y$$

$$dY/dt = -\omega X$$

Thus the predator populations grows proportionately to the prey population, and the prey declines proportionately to the predator population.

After differentiating the first and using the second equation, we obtain the relation

$$d^2 X/dt^2 + \omega^2 X = 0$$

having the solution

$$X(t) = \sin(\omega t); Y(t) = \cos(\omega t).$$

with

$$X(t)^2 + Y(t)^2 = \sin^2(\omega t) + \cos^2(\omega t) \equiv 1.$$

Thus in the two dimensional phase space defined by $X(t)$ and $Y(t)$, the system traces out an endless, circular trajectory in time, representing the out-of-phase sinusoidal oscillations of the predator and prey populations.

Divide the $X - Y$ phase space into two components – the simplest coarse graining – calling the halfplane to the left of the vertical Y -axis A and that to the right B . This system, over units of the period $1/(2\pi\omega)$, traces out a stream of A ’s and B ’s having a very precise grammar and syntax, i.e.

ABABABAB...

Many other such statements might be conceivable, e.g.

AAAAA..., BBBBB..., AAABAAAB..., ABAABAAAB...,

and so on, but, of the obviously infinite number of possibilities, only one is actually observed, is ‘grammatical’, i.e. *ABABABAB...*

More complex dynamical system models, incorporating diffusional drift around deterministic solutions, or even very elaborate systems of complicated stochastic differential equations, having various domains of attraction, i.e. different sets of grammars, can be described by analogous symbolic dynamics (e.g. Beck and Schlogl, 1993, Ch. 3).

2.2 Ecosystems and information

Rather than taking symbolic dynamics as a simplification of more exact analytic or stochastic approaches, it proves useful, as it were, to throw out the Cheshire cat, but keep the cat’s smile, generalizing symbolic dynamics to a more comprehensive information dynamics: Ecosystems may not have identifiable sets of dynamic equations like noisy, nonlinear clocks, but, under appropriate coarse-graining, they may still have recognizable sets of grammar and syntax over the long-term.

Examples abound. The turn-of-the seasons in a temperate climate, for many natural communities, looks remarkably the same year after year: the ice melts, the migrating birds return, the trees bud, the grass grows, plants and animals reproduce, high summer arrives, the foliage turns, the birds leave, frost, snow, the rivers freeze, and so on.

Suppose it is indeed possible to empirically characterize an ecosystem at a given time t by observations of both habitat parameters such as temperature and rainfall, and numbers of various plant and animal species.

Traditionally, one can then calculate a cross-sectional species diversity index at time t using an information or entropy metric of the form

$$H = - \sum_{j=1}^M (n_j/N) \log[(n_j/N)],$$

$$N \equiv \sum_{j=1}^M n_j$$

(1)

where n_j is the number of observed individuals of species j and N is the total number of individuals of all species observed (e.g. Pielou, 1977; Ricotta, 2003; Fath et al., 2003).

This is not the approach taken here. Quite the contrary, in fact. Suppose it is possible to coarse grain the ecosystem at time t according to some appropriate partition of the phase space in which each division A_j represent a particular range of numbers of each possible species in the ecosystem, along with associated parameters such as temperature, rainfall, and the like. What is of particular interest to our development is not cross sectional structure, but rather longitudinal paths, i.e. ecosystem statements of the form

$$x(n) = A_0, A_1, \dots, A_n$$

defined in terms of some natural time unit of the system, i.e. n corresponds to an again appropriate characteristic time unit T , so that $t = T, 2T, \dots, nT$.

To reiterate, unlike the traditional use of information theory in ecology, our interest is in the *serial correlations along paths*, and not at all in the cross-sectional entropy calculated for of a single element of a path.

Let $N(n)$ be the number of possible paths of length n which are consistent with the underlying grammar and syntax of the appropriately coarsegrained ecosystem, e.g. spring leads to summer, autumn, winter, back to spring, etc. but never something of the form spring to autumn to summer to winter in a temperate ecosystem.

The fundamental assumptions are that – for this chosen coarse-graining – $N(n)$, the number of possible grammatical

paths, is much smaller than the total number of paths possible, and that, in the limit of (relatively) large n ,

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

(2)

both exists and is independent of path.

This is a critical foundation to, and limitation on, the modeling strategy and its range of strict applicability, but is, in a sense, fairly general since it is *independent of the details of the serial correlations along a path*.

Again, these conditions are the essence of the parallel with parametric statistics. Systems for which the assumptions are not true will require special nonparametric approaches. We are inclined to believe, however, that, as for parametric statistical inference, the methodology will prove robust in that many systems will sufficiently fulfill the essential criteria.

This being said, some further comment does seem necessary. Not all possible ecosystem coarse-grainings are likely to work, and different such divisions, even when appropriate, might well lead to different descriptive quasi-languages for the ecosystem of interest. The example of Markov models is relevant. The essential Markov assumption is that the probability of a transition from one state at time T to another at time $T + \Delta T$ depends only on the state at T , and not at all on the history by which that state was reached. If changes within the interval of length ΔT are plastic, or path dependent, then attempts to model the system as a Markov process *within* the natural interval ΔT will fail, even though the model works quite well for phenomena separated by natural intervals.

Thus empirical identification of relevant coarse-grainings for which this body of theory will work is clearly not trivial, and may, in fact, constitute the hard scientific core of the matter.

This is not, however, a new difficulty in ecosystem theory. Holling (1992), for example, explores the linkage of ecosystems across scales, finding that mesoscale structures – what might correspond to the neighborhood in a human community – are ecological keystones in space, time, and population, which drive process and pattern at both smaller and larger scales and levels of organization. This will, in fact, be a core argument of our development.

Levin (1989) writes

“...[T]here is no single correct scale of observation: the insights one achieves from any investigation are contingent on the choice of scales. Pattern is neither a property of the system alone nor of the observer, but of an interaction between them... pattern exists at all levels and at all scales, and recognition of this multiplicity of scales is fundamental to describing and understanding ecosystems... there can be no

‘correct’ level of aggregation... We must recognize explicitly the multiplicity of scales within ecosystems, and develop a perspective that looks across scales and that builds on a multiplicity of models rather than seeking the single ‘correct’ one.”

Given an appropriately chosen coarse-graining, whose selection in many cases will be the difficult and central trick of scientific art, suppose it possible to define joint and conditional probabilities for different ecosystem paths, having the form

$$(3) \quad P(A_0, A_1, \dots, A_n), P(A_n | A_0, \dots, A_{n-1})$$

such that appropriate joint and conditional Shannon uncertainties can be defined on them. For paths of length two these would be of the form

$$H(X_1, X_2) \equiv - \sum_j \sum_k P(A_j, A_k) \log[P(A_j, A_k)]$$

$$H(X_1 | X_2) \equiv - \sum_j \sum_k P(A_j, A_k) \log[P(A_j | A_k)],$$

where the X_j represent the stochastic processes generating the respective paths of interest.

The essential content of the Shannon-McMillan Theorem is that, for a large class of systems characterized as information sources, a kind of law-of-large numbers exists in the limit of very long paths, so that

$$(4) \quad \begin{aligned} H[X] &= \lim_{n \rightarrow \infty} \frac{\log[N(n)]}{n} = \\ \lim_{n \rightarrow \infty} H(X_n | X_0, \dots, X_{n-1}) &= \\ \lim_{n \rightarrow \infty} \frac{H(X_0, X_1, \dots, X_n)}{n+1}. \end{aligned}$$

Taking the definitions of Shannon uncertainties as above, and arguing backwards from the latter two equations (e.g. Khinchin, 1957), it is indeed possible to recover the first, and divide the set of all possible temporal paths of our ecosystem into two subsets, one very small, containing the grammatically correct, and hence highly probable paths, which we will

call ‘meaningful’, and a much larger set of vanishingly low probability.

Basic material on information theory can be found in any number of texts, e.g. Ash (1990), Khinchin (1957), Cover and Thomas (1991).

The next task is to show how the cognitive processes which so distinguish much individual and collective animal activity, as well as many basic physiological processes, can be fitted into a similar context, i.e. characterized as information sources.

3 Cognition as an information source

Atlan and Cohen (1998) argue that the essence of cognition is comparison of a perceived external signal with an internal, learned picture of the world, and then, upon that comparison, the choice of one response from a much larger repertoire of possible responses.

Following the approach of Wallace (2000, 2005a), or Wallace and Fullilove (2008) it is possible to make a very general model of this process as an information source. One commentator has termed this model ‘trivial but not unimportant’. We concur.

Cognitive pattern recognition-and-selected response, as conceived here, proceeds by convoluting an incoming external ‘sensory’ signal with an internal ‘ongoing activity’ – the learned picture of the world – and, at some point, triggering an appropriate action based on a decision that the pattern of sensory activity requires a response. It is not necessary to specify how the pattern recognition system is trained, and hence possible to adopt a weak model, regardless of learning paradigm, which can itself be more formally described by the Rate Distortion Theorem. Fulfilling Atlan and Cohen’s (1998) criterion of meaning-from-response, we define a language’s contextual meaning entirely in terms of system output.

The model, an extension of that first presented in Wallace (2000), is as follows.

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

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

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

(5)

Thus we permit a graded response, supposing that if

$$h(x) \in B_0$$

(6)

the pattern is not recognized, and if

$$h(x) \in B_1$$

(7)

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

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

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

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

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

Note that disjoint partition of state space may be possible according to sets of states which can be connected by meaningful paths from a particular base point, leading to a natural coset algebra of the system, a groupoid. This is a matter of some mathematical importance pursued in Wallace and Fullilove (2008) in in Glazebrook and Wallace (2007).

It is thus possible to define an ergodic information source \mathbf{X} associated with stochastic variates X_j having joint and conditional probabilities $P(a_0, \dots, a_n)$ and $P(a_n|a_0, \dots, a_{n-1})$ such that appropriate joint and conditional Shannon uncertainties may be defined which satisfy the relations of equation 4 above.

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

As stated, the Shannon-McMillan Theorem and its variants provide ‘laws of large numbers’ which permit definition of the Shannon uncertainties in terms of cross-sectional sums of the form

$$H = - \sum P_k \log[P_k],$$

where the P_k constitute a probability distribution.

It is important to recognize that different quasi-languages will be defined by different divisions of the total universe of possible responses into various pairs of sets B_0 and B_1 . Like the use of different distortion measures in the Rate Distortion Theorem (e.g. Cover and Thomas, 1991), however, it seems obvious that the underlying dynamics will all be qualitatively similar.

Nonetheless, dividing the full set of possible responses into the sets B_0 and B_1 may itself require higher order cognitive decisions by another module or modules, suggesting the necessity of choice within a more or less broad set of possible quasi-languages. This would directly reflect the need to shift gears according to the different challenges faced by the organism or social group. A critical problem then becomes the choice of a normal zero-mode language among a very large set of possible languages representing the excited states accessible to the system. This is a fundamental matter which mirrors, for isolated cognitive systems, the resilience arguments applicable to more conventional ecosystems, i.e. the possibility of more than one zero state to a cognitive system. Identification of an excited state as the zero mode becomes, then, a kind of generalized autoimmune disorder which can be triggered by linkage with external ecological information sources of structured psychosocial stress, a matter we explore at length elsewhere (Wallace and Fullilove, 2008).

In sum, meaningful paths – creating an inherent grammar and syntax – have been defined entirely in terms of system response, as Atlan and Cohen (1998) propose.

This formalism can be applied to the stochastic neuron in a neural network: A series of inputs $y_i^j, i = 1, \dots, m$ from m nearby neurons at time j to the neuron of interest is convoluted with ‘weights’ $w_i^j, i = 1, \dots, m$, using an inner product

$$a_j = \mathbf{y}^j \cdot \mathbf{w}^j \equiv \sum_{i=1}^m y_i^j w_i^j$$

(8)

in the context of a transfer function $f(\mathbf{y}^j \cdot \mathbf{w}^j)$ such that the probability of the neuron firing and having a discrete output $z^j = 1$ is $P(z^j = 1) = f(\mathbf{y}^j \cdot \mathbf{w}^j)$.

Thus the probability that the neuron does not fire at time j is just $1 - P$. In the usual terminology the m values y_i^j constitute the ‘sensory activity’ and the m weights w_i^j the ‘ongoing activity’ at time j , with $a_j = \mathbf{y}^j \cdot \mathbf{w}^j$ and the path

$x \equiv a_0, a_1, \dots, a_n, \dots$. A more elaborate example is given in Wallace (2002a).

A little work leads to a standard neural network model in which the network is trained by appropriately varying \mathbf{w} through least squares or other error minimization feedback. This can be shown to replicate rate distortion arguments, as we can use the error definition to define a distortion function which measures the difference between the training pattern y and the network output \hat{y} as a function, for example, of the inverse number of training cycles, K . As we will discuss in another context, learning plateau behavior emerges naturally as a phase transition in the mutual information $I(Y, \hat{Y})$ driven by the parameter K .

This leads eventually to parametrization of the information source uncertainty of the dual information source to a cognitive pattern recognition-and-response with respect to one or more variates, writing, e.g. $H[\mathbf{K}]$, where $\mathbf{K} \equiv (K_1, \dots, K_s)$ represents a vector in a parameter space. Let the vector \mathbf{K} follow some path in time, i.e. trace out a generalized line or surface $\mathbf{K}(t)$. We will, following the argument of Wallace (2002b), assume that the probabilities defining H , for the most part, closely track changes in $\mathbf{K}(t)$, so that along a particular piece of a path in parameter space the information source remains as close to stationary and ergodic as is needed for the mathematics to work. Between pieces we will, below, impose phase transition characterized by a renormalization symmetry, in the sense of Wilson (1971). See the Mathematical Appendix for further details.

Such an information source can be termed adiabatically piecewise stationary ergodic (APSE). To reiterate, the ergodic nature of the information sources is a generalization of the law of large numbers and implies that the long-time averages we will need to calculate can, in fact, be closely approximated by averages across the probability spaces of those sources. This is no small matter.

The reader may have noticed parallels with Dretske’s speculations on the the role of the asymptotic limit theorems of information theory in constraining high level mental function (Dretske, 1994, Wallace, 2005a, Wallace and Fullilove, 2008).

Wallace (2004, 2005a) and Wallace and Fullilove (2008) describe in some detail how, for larger animals, immune function, tumor control, the hypothalamic-pituitary-adrenal (HPA) axis (the flight-or-fight system), emotion, conscious thought, and embedding group (and sometimes cultural) structures are all cognitive in this simple sense. In general these cognitive phenomena will occur at far faster rates than embedding ecosystem changes.

It is worth a more detailed recounting of the arguments for characterizing a number of physiological subsystems as cognitive in the sense of this section.

3.1 Immune cognition

Atlan and Cohen (1998) have proposed an information-theoretic cognitive model of immune function and process, a paradigm incorporating cognitive pattern recognition-and-response behaviors analogous to those of the central nervous

system. This work follows in a very long tradition of speculation on the cognitive properties of the immune system (e.g. Tauber, 1998; Podolsky and Tauber, 1998; Grossman, 2000).

From the Atlan/Cohen perspective, the meaning of an antigen can be reduced to the type of response the antigen generates. That is, the meaning of an antigen is functionally defined by the response of the immune system. The meaning of an antigen to the system is discernible in the type of immune response produced, not merely whether or not the antigen is perceived by the receptor repertoire. Because the meaning is defined by the type of response there is indeed a response repertoire and not only a receptor repertoire.

To account for immune interpretation Cohen (1992, 2000) has reformulated the cognitive paradigm for the immune system. The immune system can respond to a given antigen in various ways, it has ‘options.’ Thus the particular response we observe is the outcome of internal processes of weighing and integrating information about the antigen. In contrast to Burnet’s view of the immune response as a simple reflex, it is seen to exercise cognition by the interpolation of a level of information processing between the antigen stimulus and the immune response. A cognitive immune system organizes the information borne by the antigen stimulus within a given context and creates a format suitable for internal processing; the antigen and its context are transcribed internally into the ‘chemical language’ of the immune system.

The cognitive paradigm suggests a language metaphor to describe immune communication by a string of chemical signals. This metaphor is apt because the human and immune languages can be seen to manifest several similarities such as syntax and abstraction. Syntax, for example, enhances both linguistic and immune meaning.

Although individual words and even letters can have their own meanings, an unconnected subject or an unconnected predicate will tend to mean less than does the sentence generated by their connection.

The immune system creates a ‘language’ by linking two ontogenetically different classes of molecules in a syntactical fashion. One class of molecules are the T and B cell receptors for antigens. These molecules are not inherited, but are somatically generated in each individual. The other class of molecules responsible for internal information processing is encoded in the individual’s germline.

Meaning, the chosen type of immune response, is the outcome of the concrete connection between the antigen subject and the germline predicate signals.

The transcription of the antigens into processed peptides embedded in a context of germline ancillary signals constitutes the functional ‘language’ of the immune system. Despite the logic of clonal selection, the immune system does not respond to antigens as they are, but to abstractions of antigens-in-context.

More recently Cohen (2006) summarizes these matters as follows:

“Maintenance, including defense, requires the dynamic deployment of varied inflammatory pro-

cesses based on reliable information about cells in flux. The inflammatory response suited to repair a broken bone, for example, is clearly different from the inflammatory response required to hold one's gut bacteria in check or to cure a bout of influenza – which cells and molecules are to take part in the process, when, where, how, in what order, in which intensity, and with what dynamics?... The immune system mines information about the state of the various cells of the body (Is there a problem here? What kind?), integrates the body information into immune system information (antibody repertoires, cell repertoires, cell differentiation and numbers, cell movements and migrations, secreted molecules, and so forth). The modified state of the immune system, expressed locally at the site of injury and to some extent globally, is key to the inflammatory process. Immune inflammation, in turn, triggers a response of body cells in the area of injury leading, usually, to healing and restoration of function. As the process evolves, the immune system updates the inflammatory response to match the particular circumstances that emerge on the way to healing, maintaining and/or defending the body. The general success of physiological inflammation in keeping us fit is highlighted by the occasional disease caused by pathogenic inflammation – inflammation that is not properly managed by the immune system ... can cause autoimmune diseases such as multiple sclerosis, degenerative diseases such as Alzheimer's disease or allergic diseases such as asthma.

At the operational level, it is now clear that clones of lymphocytes do not function in isolation, as taught by the classical clonal selection theory. The immune system works as an integrated, whole system, and can respond potentially in many different, and even contradictory ways when it detects an injury or an antigen. The outcome of any immune response involves a choice between many alternative types of possible response, and many different types of cells take part in the response choice. This immune decision-making process uses strategies similar to those observed in nervous system cognition..."

3.2 Tumor control

We argue that the next larger cognitive submodule after the immune system must be a tumor control mechanism which may include immune surveillance, but clearly transcends it. Nunney (1999) has explored cancer occurrence as a function of animal size, suggesting that in larger animals, whose lifespan grows as about the 4/10 power of their cell count, prevention of cancer in rapidly proliferating tissues becomes more difficult in proportion to size. Cancer control requires the development of additional mechanisms and systems to address tumorigenesis as body size increases – a synergistic effect of cell number and organism longevity. Nunney (1999, p. 497)

concludes

“This pattern may represent a real barrier to the evolution of large, long-lived animals and predicts that those that do evolve ... have recruited additional controls [over those of smaller animals] to prevent cancer.”

Different tissues may have evolved markedly different tumor control strategies. All of these, however, are likely to be energetically expensive, permeated with different complex signaling strategies, and subject to a multiplicity of reactions to signals, including those related to psychosocial stress. Forlenza and Baum (2000) explore the effects of stress on the full spectrum of tumor control in higher animals, ranging from DNA damage and control, to apoptosis, immune surveillance, and mutation rate. Elsewhere (R. Wallace et al., 2003) we argue that this elaborate tumor control strategy, particularly in large animals, must be at least as cognitive as the immune system itself, which is one of its components. That is, some comparison must be made with an internal picture of a ‘healthy’ cell, and a choice made as to response: none, attempt DNA repair, trigger programmed cell death, engage in full-blown immune attack. This is, from the Atlan/Cohen perspective, the essence of cognition.

3.3 A cognitive paradigm for gene expression

While modes of genetic inheritance are assumed well understood since the Grand Evolutionary Synthesis of the early 20th Century, the mechanisms of gene activation, regulation, and expression remain largely hidden. A random reading of the literature illuminates a stark and increasingly mysterious landscape.

Liu and Ringner (2007):

“Gene expression signatures consisting of tens to hundreds of genes have been found to be informative for different biological states. Recently, many computational methods have been proposed for biological interpretation of such signatures. However, there is a lack of methods for identifying cell signaling pathways whose deregulation results in an observed expression signature... Genetic aberrations and variations in cellular processes are usually reflected in the expression of many genes... Signal transduction is at the core of many regulatory systems. Cellular functions such as growth, proliferation, differentiation, and apoptosis are regulated by signaling pathways... To uncover mechanisms underlying cellular phenotypes, therefore, it is crucial to systematically analyze gene expression signatures in the context of signaling pathways...”

Soyer et al. (2006):

“Biological signal transduction allows a cell or organism to sense its environment and react accordingly. This is achieved through cascades of proteins that interact via activation and inhibition to convert an external signal into a physiological response. In order to understand such a cascade (or network) one first needs to define its protein components and their interactions. Given the appropriate experimental data, a signal transduction network can be described by a mathematical model in order to obtain a quantitative understand of its behavior... Typically this understanding is limited to the specific network under investigation and cannot be easily used to extrapolate the behavior of other types of networks. Combined with the fact that the experimental work needed to gather enough quantitative information to develop accurate mathematical models is highly labor intensive, the modeling of specific networks may be limited in developing a broad understand of the general properties of biological signaling networks...”

One possible mathematical characterization of these difficulties is found in Sayyed-Ahmad et al. (2007):

“The state of a cell is specified by a set of variables Ψ for which we know the governing equations and a set T which is at the frontier of our understanding (i.e. for which we do not know the governing equations). The challenge is that the dynamics of Ψ is given by a cell model, e.g.

$$d\Psi/dt = G(\Psi, T(t), \Lambda)$$

(9)

in which the rate G depends not only on many rate and equilibrium constants Λ , but also on the time-dependent frontier variables $T(t)$. The descriptive variables, Ψ , can only be determined as a functional of the unknown time courses $T(t)$. Thus the model cannot be simulated...”

Liao et al (2003):

“High-dimensional data sets generated by high-throughput technologies, such as DNA microarray, are often the outputs of complex networked systems driven by hidden regulatory signals. Traditional statistical methods for computing low-dimensional or hidden representations of these data sets, such as principal component analysis and independent component analysis, ignore the underlying network structures and provide decompositions based on *a priori* statistical constraints on the computed component signals. The resulting decomposition thus

provides a phenomenological model for the observed data and does not necessarily contain physically or biologically meaningful signals.”

Baker and Stock (2007), however, pose the questions in a more general manner, using an information metaphor:

“Signal transduction systems that mediate adaptive changes in gene expression to specific sensory inputs have been well characterized. Recent studies have focused on mechanisms that allow crosstalk between different information processing modalities...”

What are the decision making mechanisms by which a bacterium controls the activities of its genes and proteins to adapt to changing environmental conditions? How do cells regulate the expression and activities of thousands of different genes and proteins to efficiently control motility, metabolism, growth and differentiation?... [H]ow do disparate regulatory modules interact to function as a cell? How are bacterial information processing networks organized?... Research on bacterial signal transduction is shifting from a focus on individual genes and proteins *in vitro* to the study of whole systems *in vivo*. Each component is now regarded as a node, the essential character of which can only be fully appreciated in terms of its connections to other nodes. In this context, an individual *E. coli* cell is a network with about 10^8 nodes composed of the products of about 10^3 different genes... The... task is to understand how these elements are connected to form a dynamic, adaptive cell. How is information converted into knowledge, and how is knowledge sorted, evaluated and combined to guide action, morphogenesis and growth?”

O’Nuallain (2006) provides an important perspective on this approach:

“The categorical failure to solve the general problem of natural language processing... by computer... [is] prognostic of the future of gene expression work. After what seemed like a promising start, the field was stalled by an inability to handle, or even define coherently, ‘contextual’ factors. Currently, the field is gradually being taken over by Bayesian ‘methods’ that simply look for the statistical incidence of co-occurrence of lexical items in the source (analogous to gene) and target (analogous to protein) languages. Contextual factors in the case of gene expression include the bioenergetic status of the cell, a status that can be assessed properly only with painstaking work; yet it determines what genes are being turned on and off at any particular moment...”

It seems clear that 18th Century dynamical models using 19th Century differential equation generalizations of equation (9) have little to offer in addressing fundamental questions of gene activation and regulation. More sophisticated work

must clearly move in the direction of an Atlan/Cohen cognitive paradigm for gene expression, characterizing the processes, and their embedding contexts, in terms of nested sets of interacting dual information sources, whose behavior is constrained by the necessary conditions imposed by the asymptotic limit theorems of communications theory.

That is, properly coarse-grained and nested biochemical networks will have an observed grammar and syntax, and, limited by powerful probability limit theorems, such description can enable construction of robust and meaningful statistical models of gene expression which may then be used for real scientific inference.

In sum, generalizing symbolic dynamics to a more inclusive, and less restrictive, cognitive paradigm for gene expression in terms of the model of equations (5) - (7), while invoking the inherent complexities of topological groupoids described in Wallace and Fullilove (2008) and Glazebrook and Wallace (2007), seems likely to provide badly needed illumination for this dark and confusing realm.

Not uncharacteristically, I. Cohen and colleagues (e.g. Cohen and Harel, 2007) have, in fact, already proposed something much in this direction:

“The discovery that DNA embodies the genetic code has drawn attention to DNA both as the cell’s primary reservoir of information and as the informatic vehicle for evolutionary change... Biologic explanations often start with DNA: biological information is seen to originate in the genome, and this genome DNA information is translated, by the way of RNA transcripts, into the diversity of expressed proteins – the proteome. The proteome then fashions the phenotype that defines the functioning organism. The genome, from such a viewpoint, appears as the master plan – the explanation – that encodes the organism... [But] [t]he explanation of the living system, from this viewpoint, is obtained not by reducing its complexity to a simple underlying ‘one dimensional’ genetic code... rather by reducing its complexity to an orderly, sequential transformation of information from genes to phenotypes... This transformational plan is not static, but homeostatic [taking this perspective] – the transformation of information from genome to phenotype, with the help of controlling feedback loops, generates the evolution of a stable, balanced adaptation of the living system to its changing environment...

But there is another way of explaining the living system; not as a hierarchical, sequential transformational system, but as a highly concurrent *reactive system*... A reactive system, in contrast to a transformational system, does not behave according to a pre-programmed chain of linked instructions. Rather, such a system reacts in parallel to many concurrent inputs, and its behaviours, outputs and effects, are not just a function of the values of its inputs but also of their variety, of the order in which

they arrive, of their timing, of their arrival speeds, and so forth. A reactive system expresses a dynamic narrative in which the DNA code is one of the many formative inputs... Structural proteins, enzymes, carbohydrates, lipids, intracellular signals, hormones and other molecules play key roles in forming and informing the system. The environment of the living system is a most critical source of information... True, DNA serves as a special repository of information because it is replicated and transmitted across generations, but DNA is meaningless without the proteins and other molecules that selectively activate segments of the DNA sequence in variable and alternative ways to create genes. The activation of specific genes emerges from the dynamic state of the cell. One could argue that DNA is just as much a servant of the cell’s state as it is the cell’s master; there is no hierarchical master plan...

Note that, unlike a transformational system, a reactive system does not seek equilibrium, has no set point and no state of rest... A reactive system holds itself together as a system just by reacting. A reactive system succeeds not by reaching homeostasis; a brain in homeostasis is clinically dead. A reactive system succeeds by being both robust and resilient. The reactive system responds to simultaneous perturbations and continues to survive; thanks to its reactive dynamics...

Reactive systems call our attention to their *emergent properties*. An emergent property of a system is a behavior of the system, taken as a whole, that is not expressed by any one of the lower-scale components that comprise it.”

Although Cohen and Harel (2007) then attempt to develop a complicated computer modeling strategy to address such reactive systems, Cohen (2006) describes the essential differences between them and conventional computer architecture as follows:

“No external operator or programmer;

No programs, algorithms, or software distinct from the system’s hardware – its cells or molecules; (Parenthetically... DNA is definitely not a program or set of algorithms... DNA is information whose meaning is defined by the way the DNA is used by the whole cell and its component molecules.)

No central processing unit (CPU);

No standard operating system; no two [biological] systems are identical, even in identical twins (since the maintenance histories of their bodies differ, their immune [and other biological] systems must differ);

No formal, mathematical logic;

No termination criteria; the system does not halt its operation;

No verification procedures.”

Zhu et al. (2007), by contrast, take an explicit kinetic chemical reaction approach to gene expression involving delayed stochastic differential equations. They begin by coarse-graining multi-step biochemical processes with single-step delayed reactions. Thus their coarse-graining involves not only collapsing biochemical steps, but collapsing as well the inevitable associated serial correlations into a small number of ‘time delays’:

“The key feature [or our model] is that the complex multiple-step biochemical processes, such as transcription, translation, and even the whole gene expression, are simplified to single-step time delayed reactions.”

While there are sufficiently many gene expression mechanisms so that some of them, at least, will yield to this method, we are interested in those which are more complex, and indeed broadly cognitive, subject to emergent patterns which cannot be modeled simply as bifurcations of stochastically-perturbed mechanistic models.

Indeed, rather than pursuing the computer models that Cohen and Harel (2007) and Zhu et al. (2007) invoke, here we will attempt to extend our statistical and dynamic analytic treatment of the cognitive paradigm to a structure incorporating gene expression in a broadly coevolutionary manner. As Richard Hamming so famously put it, “The purpose of computing is insight, not numbers”, and analytic models offer transparency as well as insight. We will, however, recover a phenomenological stochastic differential equation formalism as a kind of generalized system of Onsager relations, but at a later, and far more global, stage of structure. That is, invocation of the necessary conditions imposed by the limit theorems of communication theory enables us to penetrate one layer deeper before it becomes necessary to impose a phenomenological model.

4 Darwinian genetic inheritance as an information source

Adami et al. (2000) make the following case for reinterpreting the Darwinian transmission of genetic heritage in terms of a formal information process:

“A recent information-theoretic... definition identifies genomic complexity with the amount of information a sequence stores about its environment... genetic complexity can be defined in a consistent information-theoretic manner... [I]nformation cannot exist in a vacuum... [and] must have an instantiation... In biological systems the instantiation of information is DNA... To some extent it is the blueprint of an organism and thus information about its own structure. More specifically, it is a blueprint of how to build an organism that can best survive in its native environment, and pass on that information to its progeny... An organism’s DNA thus is not only

a ‘book’ about the organism, but also a book about the environment it lives in, including the species it co-evolves with... We... identify the complexity of geomes by the amount of information they encode about the world in which they have evolved...”

Ofria et al. (2003) continue as follows:

“...[G]enomic complexity can be defined rigorously within standard information theory as the information the genome of an organism contains about its environment... From the point of view of information theory, it is convenient to view Darwinian evolution on the molecular level as a collection of information transmission channels, subject to a number of constraints. In these channels, the organism’s genomes code for the information (a message) to be transmitted from progenitor to offspring, and are subject to noise due to an imperfect replication process. Information theory is concerned with analysing the properties of such channels, how much information can be transmitted and how the rate of perfect information transmission of such a channel can be maximized.”

Adami and Cerf (2000) argue, using simple models of genetic structure,

“Thus... the information content, or complexity, of a genomic string *by itself* (without referring to an environment) is a meaningless concept... [and] a change in environment (catastrophic or otherwise) generally leads to a [pathological] reduction in complexity.”

The transmission of genetic information is thus a contextual matter which involves operation of an information source which, according to this development, must interact with embedding (ecosystem) structures.

Such interaction is, as we show next, often highly punctuated.

5 Interacting information sources: punctuated crosstalk

Suppose that a cognitive or Darwinian information process of interest can be represented by a sequence of states in time, the path $x \equiv x_0, x_1, \dots$. Similarly, we assume an embedding ecosystem with which that process interacts can also be represented by a path $y \equiv y_0, y_1, \dots$. These paths are both very highly structured and, within themselves, are serially correlated and can, in fact, be represented by information sources \mathbf{X} and \mathbf{Y} . We assume the process of interest and the embedding ecosystem interact, so that these sequences of states are not independent, but are jointly serially correlated. We can, then, define a path of sequential pairs as $z \equiv (x_0, y_0), (x_1, y_1), \dots$

The essential content of the Joint Asymptotic Equipartition Theorem (JAEPT) version of the Shannon-McMillan Theorem is that the set of joint paths z can be partitioned into a relatively small set of high probability which is termed *jointly typical*, and a much larger set of vanishingly small probability. Further, according to the JAEPT, the *splitting criterion* between high and low probability sets of pairs is the mutual information

$$I(X, Y) = H(X) - H(X|Y) = H(X) + H(Y) - H(X, Y) \quad (10)$$

where $H(X), H(Y), H(X|Y)$ and $H(X, Y)$ are, respectively, the Shannon uncertainties of X and Y , their conditional uncertainty, and their joint uncertainty. Again, see Cover and Thomas (1991) or Ash (1990) for mathematical details. As stated above, the Shannon-McMillan Theorem and its variants permit expression of the various uncertainties in terms of cross sectional sums of terms of the form $-P_k \log[P_k]$ where the P_k are appropriate direct or conditional probabilities. Similar approaches to neural process have been recently adopted by Dimitrov and Miller (2001).

The high probability pairs of paths are, in this formulation, all equiprobable, and if $N(n)$ is the number of jointly typical pairs of length n , then, according to the Shannon-McMillan Theorem and its ‘joint’ variants,

$$I(X, Y) = \lim_{n \rightarrow \infty} \frac{\log[N(n)]}{n}. \quad (11)$$

Generalizing the earlier language-on-a-network models of Wallace and Wallace (1998, 1999), suppose there is a coupling parameter P representing the degree of linkage between the cognitive human subsystem of interest and the structured quasi-language of the embedding ecosystem, and set $K = 1/P$, following the development of those earlier studies. Then we have

$$I[K] = \lim_{n \rightarrow \infty} \frac{\log[N(K, n)]}{n}.$$

The essential homology between information theory and statistical mechanics lies in the similarity of this expression with the infinite volume limit of the free energy density. If $Z(K)$ is the statistical mechanics partition function derived from the system’s Hamiltonian, then the free energy density is determined by the relation

$$F[K] = \lim_{V \rightarrow \infty} \frac{\log[Z(K)]}{V}.$$

(12)

F is the free energy density, V the system volume and $K = 1/T$, where T is the system temperature.

Various authors argue at some length (e.g. Wallace, 2005a; Wallace and Fullilove, 2008; Rojdestvensky and Cottam, 2000; Feynman, 1996) that this is indeed a systematic mathematical homology which, as described in the Appendix, permits importation of renormalization symmetry into information theory. Imposition of invariance under renormalization on the mutual information splitting criterion $I(X, Y)$ implies the existence of phase transitions analogous to learning plateaus or punctuated evolutionary equilibria in the relations between cognitive mechanism and the embedding ecosystem. An extensive mathematical treatment of these ideas is presented elsewhere (e.g. Wallace, 2000, 2002a, b; Wallace 2005a; Wallace and Fullilove, 2008; Pettini, 2007) and in the Mathematical Appendix. A detailed example will be given in a subsequent section. Much of the uniqueness of the system under study will be expressed in the renormalization relations associated with that punctuation. See Wallace (2005a) or Wallace and Fullilove (2008) for fuller discussion.

Elaborate developments are possible. From a the more limited perspective of the Rate Distortion Theorem, a selective corollary of the Shannon-McMillan Theorem, we can view the onset of a punctuated interaction between the cognitive process and embedding ecosystem as the literal writing of distorted image of those systems upon each other, Lewontin’s (2000) interpenetration:

Suppose that two (adiabatically, piecewise stationary, ergodic) information sources \mathbf{Y} and \mathbf{B} begin to interact, to talk to each other, i.e. to influence each other in some way so that it is possible, for example, to look at the output of \mathbf{B} – strings b – and infer something about the behavior of \mathbf{Y} from it – strings y . We suppose it possible to define a retranslation from the B-language into the Y-language through a deterministic code book, and call $\hat{\mathbf{Y}}$ the translated information source, as mirrored by \mathbf{B} .

Define some distortion measure comparing paths y to paths \hat{y} , $d(y, \hat{y})$ (Cover and Thomas, 1991). We invoke the Rate Distortion Theorem’s mutual information $I(Y, \hat{Y})$, which is the splitting criterion between high and low probability pairs of paths. Impose, now, a parametrization by an inverse coupling strength K , and a renormalization symmetry representing the global structure of the system coupling.

Extending the analyses, triplets of sequences, Y_1, Y_2, Z , for which one in particular, here Z , is the ‘embedding context’ affecting the other two, can also be divided by a splitting criterion into two sets, having high and low probabilities respectively. The probability of a particular triplet of sequences is then determined by the conditional probabilities

$$I(X_1, \dots, X_i|Y_1, \dots, Y_j|Z_1, \dots, Z_k).$$

$$P(Y_1 = y^1, Y_2 = y^2, Z = z) = \prod_{j=1}^n p(y_j^1|z_j)p(y_j^2|z_j)p(z_j). \quad (16)$$

(13)

That is, Y_1 and Y_2 are, in some measure, driven by their interaction with Z .

For large n the number of triplet sequences in the high probability set will be determined by the relation (Cover and Thomas, 1992, p. 387)

$$N(n) \propto \exp[nI(Y_1; Y_2|Z)], \quad (14)$$

where splitting criterion is given by

$$I(Y_1; Y_2|Z) \equiv$$

$$H(Z) + H(Y_1|Z) + H(Y_2|Z) - H(Y_1, Y_2, Z).$$

It is then possible to examine mixed cognitive/adaptive phase transitions analogous to learning plateaus (Wallace, 2002b) in the splitting criterion $I(Y_1, Y_2|Z)$. We reiterate that these results are almost exactly parallel to the Eldredge/Gould model of evolutionary punctuated equilibrium (Eldredge, 1985; Gould, 2002).

The model is easily extended to any number of interacting information sources, Y_1, Y_2, \dots, Y_s conditional on an external context Z in terms of a splitting criterion defined by

$$I(Y_1, \dots, Y_s|Z) = H(Z) + \sum_{j=1}^s H(Y_j|Z) - H(Y_1, \dots, Y_s, Z), \quad (15)$$

where the conditional Shannon uncertainties $H(Y_j|Z)$ are determined by the appropriate direct and conditional probabilities.

If we assume interacting information sources can be partitioned into three different sets, perhaps fast, X_i , medium, Y_j and slow Z_k relative transmission rates, then mathematical induction on this equation produces a complicated expression of the form

In general, then, it seems fruitful to characterize the mutual interpenetration of cognitive biopsychosocial and non-cognitive ecosystem and genetic structures within the context a single, unifying, formal perspective summarized by a ‘larger’ information source, more precisely, invoking a mutual information between cognitive, genetic, and ecosystem information sources.

6 Dynamic manifolds

A fundamental homology between the information source uncertainty dual to a cognitive process and the free energy density of a physical system arises, in part, from the formal similarity between their definitions in the asymptotic limit. Information source uncertainty can be defined as in equation (4). This is, as noted, quite analogous to the free energy density of a physical system, equation (12).

Feynman (1996) provides a series of physical examples, based on Bennett’s work, where this homology is, in fact, an identity, at least for very simple systems. Bennett argues, in terms of idealized irreducibly elementary computing machines, that the information contained in a message can be viewed as the work saved by not needing to recompute what has been transmitted.

Feynman explores in some detail Bennett’s ideal microscopic machine designed to extract useful work from a transmitted message. The essential argument is that computing, in any form, takes work. Thus the more complicated a cognitive process, measured by its information source uncertainty, the greater its energy consumption, and the ability to provide energy is limited. Inattentive blindness, Wallace (2007) argues, emerges as a thermodynamic limit on processing capacity in a topologically-fixed global workspace, i.e. one which has been strongly configured about a particular task.

Understanding the time dynamics of cognitive systems away from the kind of phase transition critical points described above requires a phenomenology similar to the Onsager relations of nonequilibrium thermodynamics. This will lead to a more general phase transition theory involving large-scale topological changes in the sense of Morse theory, summarized in the Mathematical Appendix.

If the dual source uncertainty of a cognitive process is parameterized by some vector of quantities $\mathbf{K} \equiv (K_1, \dots, K_m)$, then, in analogy with nonequilibrium thermodynamics, gradients in the K_j of the *disorder*, defined as

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

(17)

become of central interest.

Equation (17) is similar to the definition of entropy in terms of the free energy density of a physical system, as suggested by the homology between free energy density and information source uncertainty described above.

Pursuing the homology further, the generalized Onsager relations defining temporal dynamics become

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

(18)

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 $\partial S / \partial K$ are analogous to thermodynamic forces in a chemical system, and may be subject to override by external physiological driving mechanisms (Wallace, 2005c).

An essential contrast with simple physical systems driven by (say) entropy maximization is that cognitive systems make decisions about resource allocation, to the extent resources are available. That is, resource availability is a context for cognitive function, in the sense of Baars, not a determinant.

Equations (17) and (18) 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. Cognitive, genetic, and ecosystem phenomena are, according to our development, to be associated with particular information sources, and we are interested in the local properties of the system near a particular reference state. We impose a topology on the system, so that, near a particular ‘language’ A , dual to an underlying cognitive process, there is (in some sense) an open set U of closely similar languages \hat{A} , such that $A, \hat{A} \subset U$. Note that it may be necessary to coarse-grain the system’s responses to define these information sources. The problem is to proceed in such a way as to preserve the underlying essential topology, while eliminating ‘high frequency noise’. The formal tools for this can be found, e.g., in Chapter 8 of Burago et al. (2001).

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

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

[2] Define an information-theoretic distortion measure in that extended, joint alphabet between any high probability (i.e. grammatical and syntactical) paths in A and \hat{A} , which

we write as $d(Ax, \hat{A}x)$ (Cover and Thomas, 1991). Note that these languages do not interact, in this approximation.

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

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

(19)

using an appropriate integration limit argument over the high probability paths. Note that the integration in the denominator is over different paths within A itself, while in the numerator it is between different paths in A and \hat{A} .

Consideration suggests \mathcal{M} is a formal metric, having

$$\mathcal{M}(A, B) \geq 0, \mathcal{M}(A, A) = 0, \mathcal{M}(A, B) = \mathcal{M}(B, A),$$

$$\mathcal{M}(A, C) \leq \mathcal{M}(A, B) + \mathcal{M}(B, C).$$

Other approaches to metric construction on U seem possible.

Structures weaker than a conventional metric would be of more general utility, but the mathematical complications are formidable (Glazebrook and Wallace, 2007).

Note that these conditions can be used to define equivalence classes of *languages*, where previously, in cognitive process, we could define equivalence classes of *states* which could be linked by high probability, grammatical and syntactical, paths to some base point. This led to the groupoid characterization of different information sources described in Wallace and Fullilove (2008). Here we construct an entity, formally a topological manifold, *which is an equivalence class of information sources*. This is, provided \mathcal{M} is a conventional metric, a classic differentiable manifold. We shall be interested in topological states within such manifolds, and in the possibilities of transition between manifolds.

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

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

(20)

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

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

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

$$S \equiv H - \mathcal{M}dH/d\mathcal{M}. \quad (21)$$

The simplest possible Onsager relation – here seen as an empirical, fitted, equation like a regression model – in this case becomes

$$d\mathcal{M}/dt = LdS/d\mathcal{M}, \quad (22)$$

where t is the time and $dS/d\mathcal{M}$ represents an analog to the thermodynamic force in a chemical system. This is seen as acting on the reference state A_0 . For

$$\begin{aligned} dS/d\mathcal{M}|_{A_0} &= 0, \\ d^2S/d\mathcal{M}^2|_{A_0} &> 0 \end{aligned} \quad (23)$$

the system is quasistable, a Black hole, if you will, and externally imposed forcing mechanisms will be needed to effect a transition to a different state. We shall explore this circumstance below in terms of topological considerations analogous to the concept of ecosystem resilience.

Conversely, changing the direction of the second condition, so that

$$dS^2/d\mathcal{M}^2|_{A_0} < 0,$$

leads to a repulsive peak, a White hole, representing a possibly unattainable realm of states.

Explicit parametrization of \mathcal{M} introduces standard – and quite considerable – notational complications (e.g. Burago et al., 2001; Auslander, 1967): Imposing a metric for different cognitive dual languages parametrized by \mathbf{K} leads to Riemannian, or even Finsler, geometries, including the usual geodesics. See the Mathematical Appendix for details.

The dynamics, as we have presented them so far, have been noiseless, while neural systems, from which we are abducting theory, are well known to be very noisy, and indeed may be subject to mechanisms of stochastic resonance. Equation (22) might be rewritten as

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

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

$$d\mathcal{M}_t = L(t, dS/d\mathcal{M})dt + \sigma(t, dS/d\mathcal{M})dB_t, \quad (24)$$

where L and σ are appropriately regular functions of t and \mathcal{M} , and dB_t represents the noise structure.

In the sense of Emery (1989), this leads into deep realms of stochastic differential geometry and related topics. The obvious inference is that noise, which need not be ‘white’, can serve as a tool to shift the system between various equivalence classes, i.e. as a kind of crosstalk and the source of a generalized stochastic resonance.

Deeply hidden in equation (24) is a very complicated pattern of equivalence class dynamics, since flows are defined on a manifold of languages having particular relations between H , S , and \mathcal{M} . Many possible information sources may, in fact, correspond to any particular ‘point’ on this manifold. Although we cannot pursue this matter in much detail, as it involves subtle matters of ‘topological groupoids’ and the like, some implications are clear. In particular, setting equation (24) to zero and solving for ‘stationary points’ find a set of stable attractors, since the noise terms will perturb the structure from unstable equilibria. Second, what is converged to is not some ‘stable state’ in any sense, but rather is an equivalence class of highly dynamic information sources. We will have more to say on this below.

We have defined a particular set of equivalence classes of information sources dual to cognitive processes, ecosystems, and genetic heritage. That set parsimoniously characterizes the available dynamical manifolds, and, breaking of the associated groupoid symmetry creates more complex objects of considerable interest. This leads to the possibility, indeed, the necessity, of *Deus ex Machina* mechanisms to force transitions between the different possible modes within and across dynamic manifolds.

Equivalence classes of *states* gave dual information sources to cognitive systems. Equivalence classes of *information sources* give different characteristic system dynamics. Below we will examine equivalence classes of *paths*, which will produce different directed homotopy topologies characterizing those dynamical manifolds. This introduces the possibility of having different quasi-stable resilience modes *within* individual dynamic manifolds. Pink or white noise might provide a tunable means of creating crosstalk between different topological states within a dynamical manifold, or between different dynamical manifolds altogether.

Effectively, topological shifts between and within dynamic manifolds constitute a theory of phase transitions. Indeed, similar considerations have become central in the study of phase changes for physical systems. Franzosi and Pettini (2004), for instance, write

“The standard way of [studying phase transition in physical systems] is to consider how the values of thermodynamic observables, obtained in laboratory experiments, vary with temperature, volume, or an external field, and then to associate the experimentally observed discontinuities at a PT [phase transition] to the appearance of some kind of singularity entailing a loss of analyticity... However, we can wonder whether this is the ultimate level of mathematical understanding of PT phenomena, or if some reduction to a more basic level is possible... [Our] new theorem says that nonanalyticity is the ‘shadow’ of a more fundamental phenomenon occurring in configuration space: *a topology change*... [Our] theorem means that a topology change [in a particular energy manifold] is a *necessary* condition for a phase transition to take place... The topology changes implied here are those described within the framework of Morse theory through *attachment handles*...

The converse of our Theorem is not true. There is not a one-to-one correspondence between phase transitions and topology changes... an open problem is that of *sufficiency* conditions, that is to determine which kinds of topology changes can entail the appearance of a [phase transition].”

The phenomenological Onsager treatment would also be enriched by adoption of a Morse theory perspective on topological transitions, following Michel and Mozrzymas (1977).

The next section introduces a further topological complication.

7 Directed homotopy

To reiterate, we can define equivalence classes of states according to whether they can be linked by grammatical/syntactical high probability ‘meaningful’ paths, generating ‘languages’, as done in Wallace and Fullilove (2008). Above we developed equivalence classes of languages constituting dynamic manifolds. Next we ask the precisely complementary question regarding paths on dynamical manifolds: For any two particular given states, is there some sense in which we can define equivalence classes across the set of meaningful paths linking them?

This is of particular interest in a second order hierarchical model which, effectively, describes a universality class tuning of the renormalization parameters characterizing the dancing, flowing, tunably punctuated process of cognitive function.

A closely similar question is central to recent algebraic geometry approaches to concurrent, i.e. highly parallel, comput-

ing (e.g. Pratt, 1991; Goubault and Raussen, 2002; Goubault, 2003), which we adapt.

For the moment we restrict the analysis to a system characterized by two Morse-theoretic parameters, say w_1 and w_2 , and consider the set of meaningful paths connecting two particular points, say a and b , in the two dimensional w -space plane of figure 1. The arguments surrounding equations (17), (18) and (23) suggests that there may be regions of fatal attraction and strong repulsion, Black holes and White holes, which can either trap or deflect the path of the system.

Figures 1 and 2 show two possible configurations for a Black and a White hole, diagonal and cross-diagonal. If one requires path monotonicity – always increasing or remaining the same – then, following, e.g. Goubault (2003, figs. 6,7), there are, intuitively, two direct ways, without switchbacks, that one can get from a to b in the diagonal geometry of figure 1, without crossing a Black or White hole, but there are three in the cross-diagonal structure of figure 2.

Elements of each ‘way’ can be transformed into each other by continuous deformation without crossing either the Black or White hole. Figure 1 has two additional possible monotonic ways, involving over/under switchbacks, which are not drawn. Relaxing the monotonicity requirement generates a plethora of other possibilities, e.g. loopings and backwards switchbacks, whose consideration is left as an exercise. It is not clear under what circumstances such complex paths can be meaningful, a matter for further study.

These ways are the equivalence classes defining the topological structure of the two different w -spaces, analogs to the fundamental homotopy groups in spaces which admit of loops (e.g. Lee, 2000). 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 a to b , although, for nonmonotonic paths, intermediate looping would seem possible. The theory is thus one of directed homotopy, dihomotopy, and the central question revolves around the continuous deformation of paths in w -space into one another, without crossing Black or White holes. Goubault and Raussen (2002) provide another introduction to the formalism.

It seems likely that cultural heritage or developmental history can define quite different dihomotopies in natural ecosystems, cognitive process, and genetic heritage. That is, the topology will be developmentally modulated.

Such considerations, and indeed the Black Hole development of equation (23), suggest that a system which becomes trapped in a particular pattern of behavior 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 phase transition in a physical system.

This sort of behavior is central to ecosystem resilience theory (Gunderson, 2000; Holling, 1973). The essential idea is that equivalence classes of dynamic manifolds, and the directed homotopy classes within those manifolds, each and together create domains of quasi-stability requiring action of

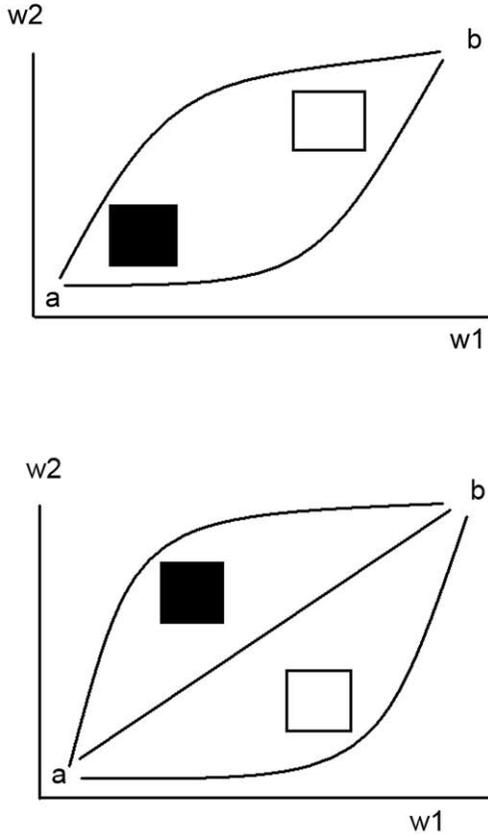


Figure 1: Diagonal Black and White holes in the two dimensional w -plane. Only two direct paths can link points a and b which are continuously deformable into one another without crossing either hole. There are two additional monotonic switchback paths which are not drawn.

Figure 2: Cross-diagonal Black and White holes. Three direct equivalence classes of continuously deformable paths can link a and b . Thus the two spaces are topologically distinct. Here monotonic switchbacks are not possible, although relaxation of that condition can lead to ‘backwards’ switchbacks and intermediate loopings.

some external factor for change.

Apparently the set of dynamic manifolds, and its subsets of directed homotopy equivalence classes, formally classifies quasi-equilibrium states, and thus characterizes the different possible resilience modes.

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

Equivalence classes of quasi-languages generated dynamical manifolds, which Wallace and Fullilove (2008) and Glazebrook and Wallace (2007) use to construct a groupoid structure, and equivalence classes of paths on those manifolds constitute dihomotopy topological states. We have recovered the essential material of the Spruce budworm example described in the introduction. Shifts between dihomotopy modes represent transitions within manifolds, but larger scale shifts, between manifolds, are also possible, in this model.

Next we consider a particular canonical form of interaction between rapid, mesoscale, and slow information sources, which will produce the principal results.

8 Red Queen coevolution

8.1 Fragmentation

Natural systems subject to coevolutionary interaction may become enmeshed in the Red Queen dilemma of Alice in Wonderland, in that they must undergo constant evolutionary change in order to avoid extinction – they must constantly run just to stay in the same place. An example would be a competitive arms race between predator and prey: Each evolutionary advance in predation must be met with a coevolutionary adaptation which allows the prey to avoid the more efficient predator. Otherwise the system will become extinct, since a highly specialized predator can literally eat itself out of house and home. Similarly, each prey defense must be matched by a predator adaptation for the system to persist.

Here we present a fairly elaborate model of coevolution, in terms of interacting information sources. Interaction events, we will argue, can be highly punctuated. These may be between Darwinian genetic, broadly cognitive, or embedding ecosystem structures.

We begin by examining ergodic information sources and their dynamics under the self-similarity of a renormalization transformation near a punctuated phase transition. We then study the linked interaction of two information sources in which the richness of the quasi-language of each affects the other, that is, when two information sources have become one another’s primary environments. This leads directly and naturally to a coevolutionary Red Queen. We will conclude by generalizing the development to a ‘block diagonal’ structure of several interacting sources.

The structures of interest to us here can be most weakly, and hence universally, described in terms of an adiabatically, piecewise stationary, ergodic information source involving a stochastic variate X which, in some general sense, sends symbols a in correlated sequences $\alpha_0, \alpha_1 \dots \alpha_{n-1}$ of length n (which

may vary), according to a joint probability distribution, and its associated conditional probability distribution,

$$P[X_0 = \alpha_0, X_1 = \alpha_1, \dots, X_{n-1} = \alpha_{n-1}],$$

$$P[X_{n-1} = \alpha_{n-1} | X_0 = \alpha_0, \dots, X_{n-2} = \alpha_{n-2}].$$

If the conditional probability distribution depends only on m previous values of X , then the information source is said to be of order m (Ash, 1990).

By ‘ergodic’ we mean that, in the long term, correlated sequences of symbols are generated at an average rate equal to their (joint) probabilities. ‘Adiabatic’ means that changes are slow enough to allow the necessary limit theorems to function, ‘stationary’ means that, between pieces, probabilities don’t change (much), and ‘piecewise’ means that these properties hold between phase transitions, which are described using renormalization methods.

As the length of the (correlated) sequences increases without limit, the Shannon-McMillan Theorem permits division of all possible streams of symbols into two groups, a relatively small number characterized as meaningful, whose long-time behavior matches the underlying probability distribution, and an increasingly large set of gibberish with vanishingly small probability. Let $N(n)$ be the number of possible meaningful sequences of length n emitted by the source \mathbf{X} . Again, uncertainty of the source, $H[\mathbf{X}]$, can be defined by the subadditive relation

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

The Shannon-McMillan Theorem shows how to characterize $H[\mathbf{X}]$ directly in terms of the joint probability distribution of the source \mathbf{X} : $H[\mathbf{X}]$ is observable and can be calculated from the inferred pattern of joint probabilities.

Let $P[x_i|y_j]$ be the conditional probability that stochastic variate $X = x_i$ given that stochastic variate $Y = y_j$ and let $P[x_i, y_j]$ be the joint probability that $X = x_i$ and $Y = y_j$. Then the joint uncertainty of X and Y , $H(X, Y)$, is given by

$$H(X, Y) = - \sum_i \sum_j P(x_i, y_j) \log[P(x_i, y_j)].$$

The conditional uncertainty of X given Y is defined as

$$H(X|Y) = - \sum_i \sum_j P(x_i, y_j) \log[P(x_i|y_j)].$$

Again, the Shannon-McMillan Theorem states that the subadditive function for $H[\mathbf{X}]$ is given by the limits

$$\begin{aligned} H[\mathbf{X}] &= \lim_{n \rightarrow \infty} H(X_n | X_0, \dots, X_{n-1}) \\ &= \lim_{n \rightarrow \infty} \frac{H(X_0, X_1, \dots, X_n)}{n+1}. \end{aligned}$$

Estimating the probabilities of the sequences $\alpha_0, \dots, \alpha_{n-1}$ from observation, the ergodic property allows us to use them

to estimate the uncertainty of the source, i.e. of the behavioral language \mathbf{X} . That is, $H[\mathbf{X}]$ is directly measurable.

Some elementary consideration (e.g. Ash, 1990; Cover and Thomas, 1991) shows that source uncertainty has a least upper bound, a supremum, defined by the capacity of the channel along which information is transmitted. That is, there exists a number C defined by externalities such that $H[\mathbf{X}] \leq C$.

C is the maximum rate at which the external world can transmit information originating with the information source, or that internal workspaces can communicate. Much of the subsequent development could, in fact, be expressed using this relation.

Again recall the relation

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

and its analog with a physical system having volume V which undergoes a phase transition depending on an inverse temperature parameter $K = 1/T$ at a critical temperature T_C . Remember that the free energy density $F(K)$ of a physical system is defined by

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

where V is as above and $Z(K)$ is the *partition function* defined from the system’s energy distribution.

Imposition of a renormalization symmetry on $F(K)$ describes, in the infinite volume limit, the behavior of the system at the phase transition in terms of scaling laws (K. Wilson, 1971). After some development, taking the limit $n \rightarrow \infty$ as an analog to the infinite volume limit of a physical system, we will apply this approach to a source uncertainty associated with crosstalking internal information sources. We will examine changes in their structure as a fundamental ‘inverse temperature’ changes across the underlying system.

We use three parameters to describe the relations between an information source and its environment or between different interacting sources.

The first, $J \geq 0$, measures the degree to which acquired characteristics are transmitted. For systems without memory $J = 0$. $J \approx 0$ thus represents a high degree of genetic as opposed to cultural inheritance.

J will always remain distinguished, a kind of inherent direction or external field strength in the sense of Wilson (1971).

The second parameter, $Q = 1/C \geq 0$, represents the inverse availability of resources. $Q \approx 0$ thus represents a high ability to renew and maintain an particular system, in a large sense.

The third parameter, $K = 1/T$, is an inverse index of a generalized temperature T , which we will more directly specify below in terms of the richness of interacting information sources.

We suppose further that the structure of interest is implicitly embedded in, and operates within the context of, a larger manifold stratified by metric distances.

Take these as multidimensional vector quantities \mathbf{A} , \mathbf{B} , \mathbf{C} \mathbf{A} may represent location in space, time delay, or the like, and \mathbf{B} may be determined through multivariate analysis

of a spectrum of observed behavioral or other factors, in the largest sense, etc.

It may be possible to reduce the effects of these vectors to a function of their magnitudes $a = |\mathbf{A}|$, $b = |\mathbf{B}|$ and $c = |\mathbf{C}|$, etc. Define the *generalized distance* r as

$$r^2 = a^2 + b^2 + c^2 + \dots \quad (25)$$

To be more explicit, we assume an ergodic information source \mathbf{X} is associated with the reproduction and/or persistence of a population, ecosystem, cognitive language or other structure. The source \mathbf{X} , its uncertainty $H[J, K, Q, \mathbf{X}]$ and its parameters J, K, Q all are assumed to depend *implicitly* on the embedding manifold, in particular on the metric r of equation (25).

A particularly elegant and natural formalism for generating such punctuation in our context involves application of Wilson's (1971) program of renormalization symmetry – invariance under the renormalization transform – to source uncertainty defined on the r -manifold. The results predict that language in the most general sense, which includes the transfer of information within a system, or between it and an embedding context, will undergo sudden changes in structure analogous to phase transitions in physical systems.

We must, however, emphasize that this approach is argument by abduction, in Hodgson's (1993) sense, from physical theory: Much current development surrounding self-organizing physical phenomena is based on the assumption that at phase transition a system looks the same under renormalization. That is, phase transition represents a stationary point for a renormalization transform in the sense that the transformed quantities are related by simple scaling laws to the original values.

Renormalization is a clustering semigroup transformation in which individual components of a system are combined according to a particular set of rules into a 'clumped' system whose behavior is a simplified average of those components. Since such clumping is a many-to-one condensation, there can be no unique inverse renormalization, and, as the Appendix shows, many possible forms of condensation.

Assume it possible to redefine characteristics of the information source \mathbf{X} and J, K, Q as functions of averages across the manifold having metric r , which we write as R . That is, 'renormalize' by clustering the entire system in terms of blocks of different sized R .

Let $N(K, J, Q, n)$ be the number of high probability meaningful correlated sequences of length n across the entire community in the r -manifold, given parameter values K, J, Q . We study changes in

$$H[K, J, Q, \mathbf{X}] \equiv \lim_{n \rightarrow \infty} \frac{\log[N(K, J, Q, n)]}{n}$$

as $K \rightarrow K_C$ and/or $Q \rightarrow Q_C$ for critical values K_C, Q_C at which the community begins to undergo a marked transformation from one kind of structure to another.

Given the metric of equation (25), a *correlation length*, $\chi(K, J, Q)$, can be defined as the average length in r -space over which structures involving a particular phase dominate.

Now clump the 'community' into blocks of average size R in the multivariate r -manifold, the 'space' in which the cognitive enterprise is implicitly embedded.

Following the classic argument of Wilson (1971), reproduced and expanded in the Appendix, it is possible to impose renormalization symmetry on the source uncertainty on H and χ by assuming at transition the relations

$$H[K_R, J_R, Q_R, \mathbf{X}] = R^D H[K, J, Q, \mathbf{X}] \quad (26)$$

and

$$\chi(K_R, J_R, Q_R) = \frac{\chi(K, J, Q)}{R} \quad (27)$$

hold, where K_R, J_R and Q_R are the transformed values of K, J and Q after the clumping of renormalization. We take $K_1, J_1, Q_1 \equiv K, J, Q$ and permit the characteristic exponent D to be nonintegral. The Mathematical Appendix provides examples of other possible relations.

Equations (26) and (27) are assumed to hold in a neighborhood of the transition values K_C and Q_C .

Differentiating these with respect to R gives complicated expressions for $dK_R/dR, dJ_R/dR$ and dQ_R/dR depending simply on R which we write as

$$\begin{aligned} dK_R/dR &= \frac{u(K_R, J_R, Q_R)}{R} \\ dQ_R/dR &= \frac{w(K_R, J_R, Q_R)}{R} \\ dJ_R/dR &= \frac{v(K_R, J_R, Q_R)}{R} J_R. \end{aligned} \quad (28)$$

Solving these differential equations gives K_R , J_R and Q_R as functions of J, K, Q and R .

Substituting back into equations (26) and (27) and expanding in a first order Taylor series near the critical values K_C and Q_C gives power laws much like the Widom-Kadanoff relations for physical systems (Wilson, 1971). For example, letting $J = Q = 0$ and taking $\kappa \equiv (K_C - K)/K_C$ gives, in first order near K_C ,

$$\begin{aligned} H &= \kappa^{D/y} H_0 \\ \chi &= \kappa^{-1/y} \chi_0 \end{aligned} \quad (29)$$

where y is a constant arising from the series expansion.

Note that there are only two fundamental equations – (26) and (27) – in $n > 2$ unknowns: The critical ‘point’ is, in this formulation, most likely to be a complicated implicitly defined critical surface in J, K, Q, \dots -space. The ‘external field strength’ J remains distinguished in this treatment, i.e. the inverse of the degree to which acquired characteristics are inherited, but *neither K , Q nor other parameters are, by themselves, fundamental*, rather their joint interaction defines critical behavior along this surface.

That surface is a fundamental object, not the particular set of parameters (except for J) used to define it, which may be subject to any set of transformations which leave the surface invariant. Thus whatever parameters drive the system are inextricably intertwined and mutually interacting, according to the form of this critical evolutionary transition surface. That surface, in turn, is unlikely to remain fixed, and should vary with time or other extrinsic parameters, including, but not likely limited to, J .

At the critical surface a Taylor expansion of the renormalization equations (26) and (27) gives a first order matrix of derivatives whose eigenstructure defines fundamental system behavior. For physical systems the surface is a saddle point (Wilson, 1971), but more complicated behavior seems likely in what we study. See Binney et al., (1986) for some details of this differential geometry.

Taking, for the moment, the simplest formulation, ($J = Q = 0$), as K increases toward a threshold value K_C , the source uncertainty of the reproductive, behavioral or other language common across the community declines and, at K_C , the average regime dominated by the ‘other phase’ grows. That is, the system begins to freeze into one having a large correlation length for the second phase. The two phenomena are linked at criticality in physical systems by the scaling exponent y .

Assume the rate of change of $\kappa = (K_C - K)/K_C$ remains constant, $|d\kappa/dt| = 1/\tau_K$. Analogs with physical theory suggest there is a characteristic time constant for the phase transition, $\tau \equiv \tau_0/\kappa$, such that if changes in κ take place on a

timescale longer than τ for any given κ , we may expect the correlation length $\chi = \chi_0 \kappa^{-s}$, $s = 1/y$, will be in equilibrium with internal changes and result in a very large fragment in r -space. Following Zurek (1985, 1996), the ‘critical’ freezeout time, \hat{t} , will occur at a ‘system time’ $\hat{t} = \chi/|d\chi/dt|$ such that $\hat{t} = \tau$. Taking the derivative $d\chi/dt$, remembering that by definition $d\kappa/dt = 1/\tau_K$, gives

$$\frac{\chi}{|d\chi/dt|} = \frac{\kappa \tau_K}{s} = \frac{\tau_0}{\kappa}$$

so that

$$\kappa = \sqrt{s\tau_0/\tau_K}.$$

Substituting this value of κ into the equation for correlation length, the expected size of fragments in r -space, $d(\hat{t})$, becomes

$$d \approx \chi_0 \left(\frac{\tau_K}{s\tau_0} \right)^{s/2}$$

with $s = 1/y > 0$. The more rapidly K approaches K_C the smaller is τ_K and the smaller and more numerous are the resulting r -space fragments. Thus rapid change produces small fragments more likely to risk extinction in a system dominated by economies of scale.

8.2 Recursion

Extending the theory above involves envisioning reciprocally interacting genetic, cognitive or ecosystem information sources as subject to a coevolutionary Red Queen by treating their respective source uncertainties as recursively parameterized by each other. That is, *assume the information sources are each other’s primary environments*. These are, respectively, characterized by information sources \mathbf{X} and \mathbf{Y} , whose uncertainties are parameterized

[1] by measures of both inheritance and inverse resources – J, Q as above – and, most critically,

[2] by each other’s inverse uncertainties, $\mathcal{H}_X \equiv 1/H[\mathbf{X}]$ and $\mathcal{H}_Y \equiv 1/H[\mathbf{Y}]$, i.e.

$$H[\mathbf{X}] = H[Q, J, \mathcal{H}_Y, \mathbf{X}]$$

$$H[\mathbf{Y}] = H[Q, J, \mathcal{H}_X, \mathbf{Y}].$$

(30)

This is a recursive system having complex behaviors.

Assume a strongly heritable genetic system, i.e. $J = 0$, with fixed inverse resource base, Q , for which $H[\mathbf{X}]$ follows something like the lower graph in figure 3, a reverse S-shaped curve with $K \equiv \mathcal{H}_Y = 1/H[\mathbf{Y}]$, and similarly $H[\mathbf{Y}]$ depends on \mathcal{H}_X . That is, increase or decline in the source uncertainty

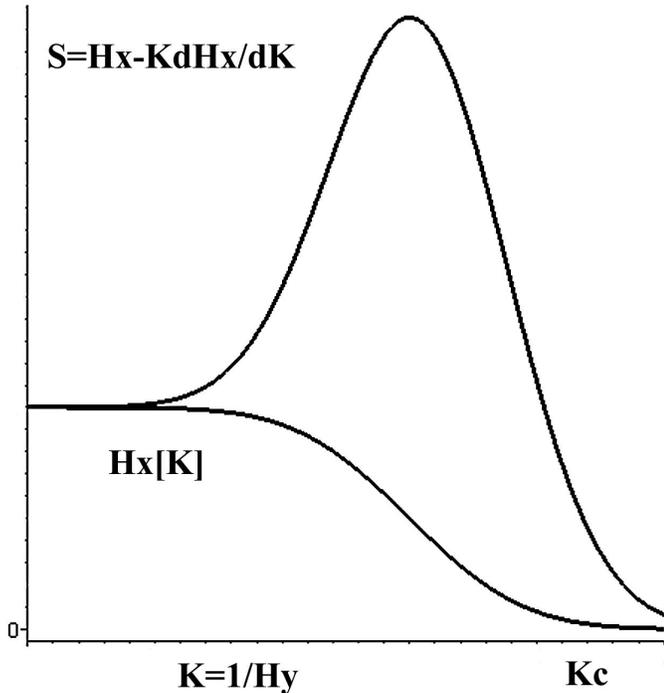


Figure 3: A reverse-S-shaped curve for source uncertainty $H[\mathbf{X}]$ – measuring language richness – as a function of an inverse temperature parameter $K = 1/H[\mathbf{Y}]$. To the right of the critical point K_C the system breaks into fragments in r -space whose size is determined by the rate at which K approaches K_C .

A collection of fragments already to the right of K_C ,

however, would be seen as condensing into a single unit as K declined below the critical point. If K is an inverse source uncertainty itself, i.e. $K = 1/H[\mathbf{Y}]$ for some information source \mathbf{Y} , then under such conditions a Red Queen dynamic can become enabled, driving the system strongly to the left.

No intermediate points are asymptotically stable, given a genetic heritage in this development, although generalized

Onsager/dynamical arguments suggest that the repulsive peak in $S = H - K/dH/dK$ can serve to create quasi-stable resilience realms. To the right of the critical point K_C the system is locked into disjoint fragments.

of one system leads to increase or decline in the source uncertainty of the other, and vice versa.

Start at the right of the lower graph for $H[\mathbf{X}]$ in figure 3, the source uncertainty of one system, but to the left of the critical point K_C . Assume $H[\mathbf{Y}]$ increases so \mathcal{H}_Y decreases, and thus $H[\mathbf{X}]$ increases, walking up the lower curve of figure 3 from the right: the richness of the system’s internal language increases – or the interaction between structures increases the richness of their associated information sources – they get smarter or faster or more poisonous, or their herd behavior becomes more sophisticated in the presence of a predator.

The increase of $H[\mathbf{X}]$ leads, in turn, to a decline in \mathcal{H}_X and triggers an increase of $H[\mathbf{Y}]$, whose increase leads to a further increase of $H[\mathbf{X}]$ and vice versa: The Red Queen, taking the system from the right of figure 3 to the left, up the lower curve as the two systems mutually interact.

The upper graph of figure 3 represents the disorder

$$S = H[K, \mathbf{X}] - KdH[K, \mathbf{X}]/dK, K \equiv 1/H[\mathbf{Y}].$$

According to the dynamical manifold analysis, the peak in S represents a repulsive barrier for transition between high and low values of $H[\mathbf{X}]$. This leads to the expectation of *hysteresis*. That is, the two realms, to the left and right of the peak in S for figure 3, thus represent quasi-stable resilience modes, in this model.

8.3 Extension

The model directly generalizes to multiple interacting information sources.

First consider a matrix of crosstalk measures between a set of information sources. Assume the matrix can be block diagonalized into two major components, characterized by network information source measures like equation (16),

$$I_m(X_1 \dots X_i | Y_1 \dots Y_j | Z_1 \dots Z_k), m = 1, 2.$$

Then apply the two-component theory above.

Extending the development to multiple, recursively interacting information sources resulting from a more general block diagonalization seems direct. First use inverse measures $\mathcal{I}_j \equiv 1/I_j, j \neq m$ as parameters for each of the other blocks, writing

$$I_m = I_m(K_1 \dots K_s, \dots \mathcal{I}_j \dots), j \neq m$$

where the K_s represent other relevant parameters.

Next segregate the \mathcal{I}_j according to their relative rates of change, as in equation (16). Cognitive gene expression would be among the most rapid, followed by ecosystem dynamics and selection.

The dynamics of such a system, following the pattern of equations (18) and (24), becomes a recursive network of stochastic differential equations, similar to those used to study many other highly parallel dynamic structures (e.g. Wymer, 1997).

Letting the K_j and \mathcal{I}_m all be represented as parameters Q_j , (with the caveat that I_m not depend on \mathcal{I}_m), one can define

$$S_I^m \equiv I_m - \sum_i Q_i \partial I_m / \partial Q_i$$

to obtain a complicated recursive system of phenomenological ‘Onsager relations’ stochastic differential equations like (24),

$$dQ_t^j = \sum_i [L_{j,i}(t, \dots \partial S_I^m / \partial Q^i \dots) dt + \sigma_{j,i}(t, \dots \partial S_I^m / \partial Q^i \dots) dB_t^i], \quad (31)$$

where, again, for notational simplicity only, we have expressed both the reciprocal \mathcal{I} ’s and the external K ’s in terms of the same Q_j .

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

Indeed, the I_m and/or the derived S^m might, in some circumstances, be combined into a Morse function, permitting application of Pettini’s Topological Hypothesis.

The model rapidly becomes unwieldy, probably requiring some clever combinatorial or groupoid convolution algebra and related diagrams for concise expression, much as in the usual field theoretic perturbation expansions (Hopf algebras, for example). The virtual reaction method of Zhu et al. (2007) is another possible approach.

As in the simple model above, there will be, first, multiple quasi-stable points within a given system’s I_m , representing a class of generalized resilience modes accessible via punctuation and enmeshing gene selection, gene expression, and ecological resilience – analogous to the simple model of figure 3.

Second, however, will be analogs to the fragmentation of figure 3 when the system exceeds the critical value K_c . 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, both of which could entrain ecosystem resilience shifts, gene expression, and gene selection, although the latter kind would seem to be the far more dramatic.

There are other possible patterns: [1] Setting equation (31) equal to zero and solving for stationary points again gives attractor states since the noise terms preclude unstable equilibria. [2] Unlike equation (24), however, this system may converge to limit cycle behavior in which the system seems to chase its tail endlessly. [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 structure represents particular patterns of ongoing dynamics rather than some identifiable ‘state’.

Here we are, at last and indeed, deeply enmeshed in a highly recursive phenomenological stochastic differential equations, but at a deeper level than Zhu et al. (2007) envisioned for gene expression alone, and in a dynamic rather than static manner: the objects of this dynamical system are equivalence classes of information sources and their crosstalk, rather than simple ‘states’ of a dynamical or reactive chemical system.

Imposition of necessary conditions from the asymptotic limit theorems of communication theory has, at least in theory, beaten the thicket back one full layer.

Other formulations may well be possible, but our work here serves to illustrate the method.

It is, however, interesting to compare our results to those of Dieckmann and Law (1996), who invoke evolutionary game dynamics to obtain a first order canonical equation having the form

$$ds_i/dt = K_i(s) \partial W_i(s'_i, s) |_{s'_i=s_i}. \quad (32)$$

Dieckmann and Law describe this as follows:

“The s_i , with $i = 1, \dots, N$ denote adaptive trait values in a community comprising N species. The $W_i(s'_i, s)$ are measures of fitness of individuals with trait values s'_i in the environment determined by the resident trait values s , and the $K_i(s)$ are non-negative coefficients, possibly distinct for each species, that scale the rate of evolutionary change. Adaptive dynamics [of this kind] have frequently been postulated, based either on the notion of a hill-climbing process on an adaptive landscape or some other sort of plausibility argument...”

When this equation is set equal to zero, so there is no time dependence, one obtains what are characterized as ‘evolutionary singularities’, i.e. stationary points.

Dieckmann and Law contend that their formal derivation of this equation satisfies four critical requirements: [1] The evolutionary process needs to be considered in a coevolutionary context. [2] A proper mathematical theory of evolution should be dynamical. [3] The coevolutionary dynamics ought to be underpinned by a microscopic theory. [4] The evolutionary process has important stochastic elements.

Our equation (31) seems clearly within this same ballpark, although we have taken a much different route, one which indeed produces elaborate patterns of phase transition punctuation in a highly natural manner. Champagnat et al. (2006),

in fact, derive a higher order canonical approximation extending equation (32) which is very much closer equation to (31), that is, a stochastic differential equation describing evolutionary dynamics. Champagnat et al. (2006) go even further, using a large deviations argument to analyze dynamical co-evolutionary paths, not merely evolutionary singularities:

“In general, the issue of evolutionary dynamics drifting away from trajectories predicted by the canonical equation can be investigated by considering the asymptotic of the probability of ‘rare events’ for the sample paths of the diffusion. By ‘rare events’ we mean diffusion paths drifting far away from the canonical equation. The probability of such rare events is governed by a *large deviation principle*...: when [a critical parameter ϵ] goes to zero, the probability that the sample path of the diffusion is close to a given rare path ϕ decreases exponentially to 0 with rate $I(\phi)$, where the ‘rate function’ I can be expressed in terms of the parameters of the diffusion...

This result can be used to study long-time behavior of the diffusion process when there are multiple attractive evolutionary singularities... [under proper conditions] the most likely path followed by the diffusion when exiting [a basin of attraction] is the one minimizing the rate function I over all the [appropriate] trajectories... The time needed to exit [the basin is] of the order $[\exp(H/\epsilon)]$ where H is a quasi-potential representing the minimum of the rate function I over all possible trajectories]...”

An essential fact of large deviations theory is that the rate function I which Champagnat et al. (2006) invoke can almost always be expressed as a kind of entropy, that is, in the form $I \propto \sum_j P_j \log(P_j)$ for some probability distribution. This result goes under a number of names; Sanov’s Theorem, Cramer’s Theorem, the Garnter-Ellis Theorem, the Shannon-McMillan Theorem, and so forth (e.g. Dembo and Zeitouni, 1998). Here we will use it, in combination with the cognitive paradigm for gene expression, to suggest the possibility of second order effects in coevolutionary process. That is, gene expression, because of its underlying cognitive nature, may be an even more central aspect of coevolutionary process than is currently understood: The fluctuational paths defined by the system of equations in (31) may, under some conditions, become serially correlated outputs of an information source driven by cognitive gene expression. In particular, the co-evolutionary pressures inherent to equation (31) may in fact strongly select for significant cognition in gene expression.

8.4 Higher order coevolution

We begin with a recapitulation of large deviations and fluctuation formalism.

Information source uncertainty, according to the Shannon-McMillan Theorem, serves as a splitting criterion between high and low probability sequences (or pairs of them) and displays the fundamental characteristic of a growing body of

work in applied probability often termed the Large Deviations Program, (LDP) which seeks to unite information theory, statistical mechanics and the theory of fluctuations under a single umbrella.

Following Dembo and Zeitouni, (1998, p.2), let X_1, X_2, \dots, X_n be a sequence of independent, standard Normal, real-valued random variables and let

$$S_n = \frac{1}{n} \sum_{j=1}^n X_j.$$

(33)

Since S_n is again a Normal random variable with zero mean and variance $1/n$, for all $\delta > 0$

$$\lim_{n \rightarrow \infty} P(|S_n| \geq \delta) = 0,$$

(34)

where P is the probability that the absolute value of S_n is greater or equal to δ . Some manipulation, however, gives

$$P(|S_n| \geq \delta) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} \exp(-x^2/2) dx,$$

(35)

so that

$$\lim_{n \rightarrow \infty} \frac{\log P(|S_n| \geq \delta)}{n} = -\delta^2/2$$

(36)

This can be rewritten for large n as

$$P(|S_n| \geq \delta) \approx \exp(-n\delta^2/2).$$

(37)

That is, for large n , the probability of a large deviation in S_n follows something much like the asymptotic equipartition relation of the Shannon-McMillan Theorem, i.e. that meaningful paths of length n all have approximately the same probability $P(n) \propto \exp(-nH[\mathbf{X}])$.

Questions about meaningful paths appear suddenly as formally isomorphic to the central argument of the LDP which encompasses statistical mechanics, fluctuation theory, and information theory into a single structure (Dembo and Zeitouni, 1998).

Perhaps the cardinal tenet of large deviation theory is that the rate function $-\delta^2/2$ can, under proper circumstances, be expressed as a mathematical entropy having the standard form

$$-\sum_k p_k \log p_k, \quad (38)$$

for some set of probabilities p_k . Again, this striking result goes under various names at various levels of approximation – Sanov’s Theorem, Cramer’s Theorem, the Gartner-Ellis Theorem, the Shannon-McMillan Theorem, and so on (Dembo and Zeitouni, 1998).

Next we briefly recapitulate part of the standard treatment of large fluctuations (Onsager and Machlup, 1953; Fredlin and Wentzell, 1998).

The macroscopic behavior of a complicated physical system in time is assumed to be described by the phenomenological Onsager relations giving large-scale fluxes as

$$\sum_i R_{i,j} dK_j/dt = \partial S/\partial K_i, \quad (39)$$

where the $R_{i,j}$ are appropriate constants, S is the system entropy and the K_i are the generalized coordinates which parametrize the system’s free energy.

Entropy is defined from free energy F by a Legendre transform – more of which follows below:

$$S \equiv F - \sum_j K_j \partial F/\partial K_j,$$

where the K_j are appropriate system parameters.

Neglecting volume problems for the moment, free energy can be defined from the system’s partition function Z as

$$F(K) = \log[Z(K)].$$

The partition function Z , in turn, is defined from the system Hamiltonian – defining the energy states – as

$$Z(K) = \sum_j \exp[-K E_j],$$

where K is an inverse temperature or other parameter and the E_j are the energy states.

Inverting the Onsager relations gives

$$dK_i/dt = \sum_j L_{i,j} \partial S/\partial K_j = L_i(K_1, \dots, K_m, t) \equiv L_i(K, t). \quad (40)$$

The terms $\partial S/\partial K_i$ are macroscopic driving forces dependent on the entropy gradient.

Let a white Brownian noise $\epsilon(t)$ perturb the system, so that

$$\begin{aligned} dK_i/dt &= \sum_j L_{i,j} \partial S/\partial K_j + \epsilon(t) \\ &= L_i(K, t) + \epsilon(t), \end{aligned} \quad (41)$$

where the time averages of ϵ are $\langle \epsilon(t) \rangle = 0$ and $\langle \epsilon(t)\epsilon(0) \rangle = D\delta(t)$. $\delta(t)$ is the Dirac delta function, and we take K as a vector in the K_i .

Following Luchinsky (1997), if the probability that the system starts at some initial macroscopic parameter state K_0 at time $t = 0$ and gets to the state $K(t)$ at time t is $P(K, t)$, then a somewhat subtle development (e.g. Feller, 1971) gives the forward Fokker-Planck equation for P :

$$\partial P(K, t)/\partial t = -\nabla \cdot (L(K, t)P(K, t)) + (D/2)\nabla^2 P(K, t). \quad (42)$$

In the limit of weak noise intensity this can be solved using the WKB, i.e. the eikonal, approximation, as follows: take

$$P(K, t) = z(K, t) \exp(-s(K, t)/D). \quad (43)$$

$z(K, t)$ is a prefactor and $s(K, t)$ is a classical action satisfying the Hamilton-Jacobi equation, which can be solved by integrating the Hamiltonian equations of motion. The equation reexpresses $P(K, t)$ in the usual parametrized negative exponential format.

Let $p \equiv \nabla s$. Substituting and collecting terms of similar order in D gives

$$dK/dt = p + L, dp/dt = -\partial L/\partial K p \quad (44)$$

and

$$-\partial s/\partial t \equiv h(K, p, t) = pL(K, t) + \frac{p^2}{2}, \quad (45)$$

with $h(K, t)$ the Hamiltonian for appropriate boundary conditions.

Again following Luchinsky (1997), these Hamiltonian equations have two different types of solution, depending on p . For $p = 0$, $dK/dt = L(K, t)$ which describes the system in the absence of noise. We expect that with finite noise intensity the system will give rise to a distribution about this deterministic path. Solutions for which $p \neq 0$ correspond to *optimal paths* along which the system will move with overwhelming probability.

These results can, however, again be directly derived as a special case of a Large Deviation Principle based on generalized entropies mathematically similar to Shannon's uncertainty from information theory, bypassing the Hamiltonian formulation entirely (Dembo and Zeitouni, 1998).

For a cognitive system characterized by a dual information source, of course, there is no Hamiltonian, but the generalized entropy or splitting criterion treatment still works. The trick is to do with information source uncertainty what is done here with a Hamiltonians.

Here we are concerned, not with a random Brownian distortion of simple physical systems, but, invoking cognitive gene expression, with a possibly complex behavioral structure, in the largest sense, composed of quasi-independent actors for which *meaningful/optimal paths have extremely structured serial correlation, amounting to a grammar and syntax,*

precisely the fact which allows definition of an information source and enables the use of the very sparse equipartition of the Shannon-McMillan and Rate Distortion Theorems.

In sum, to again paraphrase Luchinsky (1997), large fluctuations, although infrequent, are fundamental in a broad range of processes, and it was recognized by Onsager and Machlup (1953) that insight into the problem could be gained from studying the distribution of fluctuational paths along which the system moves to a given state. This distribution is a fundamental characteristic of the fluctuational dynamics, and its understanding leads toward control of fluctuations. Fluctuational motion from the vicinity of a stable state may occur along different paths. For large fluctuations, the distribution of these paths peaks sharply along an optimal, most probable, path. In the theory of large fluctuations, the pattern of optimal paths plays a role similar to that of the phase portrait in nonlinear dynamics.

In this development meaningful paths driven by cognitive gene expression can play something of the role of optimal paths in the theory of large fluctuations which Champagnat et al. (2006) have invoked, but without benefit of a Hamiltonian.

The spread of the possible spectrum of cognitive gene expression within a species, affecting the ability to adapt to changing ecological niches, then becomes central to the mitigation of selection pressures generated by coevolutionary dynamics: too limited a response repertoire will cause a species to become fully entrained into high probability dynamical fluctuational paths leading to punctuated extinction events. A broad spectrum allows a species to ride out much more of the coevolutionary selection pressure.

A sufficiently broad repertoire of cognitive gene expression responses leads, however, to the necessity of a second order coevolution model in which the high probability fluctuational paths defined by the system of equations (31) *are, in fact, themselves the output of some information source*. This is a model closely analogous to the second order cognitive structures needed to explain animal consciousness (e.g. Wallace, 2005a). Intuitively, this transition to 'cognitive coevolution' would be particularly likely under the influence of a strong system of epigenetic inheritance, that is, an animal culture extending the niche spectrum offered by cognitive gene expression alone. Thus we could expand this development to one encompassing biocultural coevolution, in particular the development of agriculture, matters to be pursued in subsequent work.

9 Discussion and conclusions

The basic point is the inevitability of punctuation in generalized coevolutionary interactions, representing fundamental structural changes in underlying manifolds, roughly analogous to the topological hypothesis of Pettini (2007). Thus evolution, resilience, and cognitive phenomena, which can all be (at least crudely) represented by information sources, are inherently subject to punctuated equilibrium phenomena essentially similar to ecosystem resilience. This pattern will

involve each individually, as well as their interactions, a consequence of the fundamental homology between information source uncertainty and free energy density.

Holling (1992) describes the particular role of the mesoscale in ecological structures as follows:

“[Analysis suggests] *The Extended Keystone Hypothesis*: All ecosystems are controlled and organized by a small number of key plant, animal, and abiotic processes that structure the landscape at different scales.

The Entrainment Hypothesis: Within any one ecosystem, the periodicities and architectural attributes of the critical structuring processes will establish a nested set of periodicities and spatial features that become attractors for other variables...

...The degree to which small, fast events influence larger, slower ones is critically dependent upon... mesoscale disturbance processes.”

Our lowest common denominator information theoretic approach to coevolutionary interaction between genes, embedding ecosystem, and cognitive process identifies ecosystem phenomena as the driving mesoscale: cognitive phenomena are much faster, and (for large animals) genetic change much slower.

That is, punctuated changes in ecosystem structure, the traditional purview of ecological resilience, appear able to entrain both Darwinian genetic and cognitive phenomena – including gene expression, triggering similarly punctuated outcomes, on top of the punctuation naturally inherent to these information systems.

Thus, while discontinuous phase transitions are ‘natural’ at all scales of biological information process, we argue here that punctuated changes in embedding ecosystem resilience regime will be particularly effective at entraining faster cognitive and slower Darwinian genetic structural phenomena. In particular, punctuated changes in ecosystem structure can write images of themselves onto genetic sequence structure in a punctuated manner, resulting in punctuated population extinction and/or speciation events on geologic timescales, and in sudden changes in gene expression and other cognitive phenomena on more rapid timescales.

This is not an entirely new idea. Laland et al. (1999) have used a different methodology to reach similar conclusions:

“There is increasing recognition that all organisms modify their environments... a process that we call ‘niche construction’. Such modifications can have profound effects on the distribution and abundance of organisms, the influence of keystone species, the control of energy and material flows, residence and return times, ecosystem resilience, and specific trophic relationships... The consequences of environment modification by organisms, however, are not restricted to ecology, and organisms can affect both their own and each other’s evolution by modifying sources of natural selection in their environments... Lewontin... points out that many of the

activities of organisms, such as migration, hoarding of food resources, habitat selection, or thermoregulatory behavior, are adaptive precisely because they dampen statistical variation in the availability of environmental resources...

Hitherto, it has not been possible to apply evolutionary theory to ecosystems, because of the presence of nonevolving abiota in ecosystems. We suspect this obstacle has been largely responsible for preventing the full integration of ecosystem ecology with population-community ecology... However... adding the new process of niche construction to the established process of natural selection... enables the incorporation of both abiotic environmental components and interactions among populations and abiota in ecosystems into evolutionary models... [an approach] equally applicable to both population-community ecology and ecosystem-level ecology...”

More recently Dercole et al. (2006) have addressed the problem using their version of equation (32) to produce very complex dynamical patterns:

“Understanding the determinants of population dynamics is an important theme throughout biology, from human health to conservation. In studying population dynamics, much research has addressed how ecological interactions affect population stability yet ignoring the genetic diversity and ensuing evolvability of populations... The dynamical interplay of ecology and evolution prompts three general questions...: (i) how does evolution of adaptive traits affect the ecological stability of a community? (ii) Under which conditions are ecological interactions expected to beget fluctuations in a population’s genetic state? (iii) How do eco-evolutionary dynamics respond to environmental change?... [A] unified analysis of eco-evolutionary dynamics in communities containing ‘slow’ and ‘fast’ populations... allows us to relax the ecological equilibrium assumption.

Slow-fast systems are composed of populations whose ecological fluctuations develop on contrasting time-scales...”

Whitham et al. (2006), in parallel with our approach, take a genetic framework associated with ecologically-dominant keystone species to examine what they call community and ecosystem phenotypes:

“Can heritable traits in a single species affect an entire ecosystem? Recent studies show that such traits... have predictable effects on community structure and ecosystem processes. Because these community and ecosystem phenotypes have a genetic basis and are heritable, we can begin to apply the principles of population and quantitative genetics to place the study of complex communities and ecosystems within an evolutionary framework.

This... could allow us to understand, for the first time, the genetic basis of ecosystem processes, and the effect of such phenomena as climate change and introduced transgenetic organisms on entire communities.”

Whitham et al. (2006) go on to define community evolution as a genetically based change in the ecological interactions that occur between species over time.

Here, by contrast, although we too focus on keystone scales, our particular innovation has been to reduce the dynamics of genetic inheritance, ecosystem persistence, and gene expression to a least common denominator as information sources operating at markedly different rates, but coupled by crosstalk into a broadly coevolutionary phenomenon marked at all scales by emergent ‘phase transition’ phenomena generating patterns of punctuated equilibrium.

We have, at times, grossly simplified the mathematical analysis. Invocation of equivalence class arguments leads naturally into deep groupoid structures and related topological generalizations, including Morse theory (Wallace and Fullilove, 2008). Taking a ‘mean number’ rather than the mean field approach of the Mathematical Appendix generates a qualitatively different class of exactly solvable models, based on giant component phase transitions in networks. Hybrids of the two are possible, and evolutionary process is unlikely to be at all constrained by formal mathematical tractability. In addition higher cognitive phenomena like individual or group consciousness require second order models analogous to hierarchical regression. Much of this is described in Wallace and Fullilove (2008).

From a larger perspective, however, what we have developed is really a class of ‘necessary conditions’ statistical models based on the Shannon McMillan and Rate Distortion Theorems. These are, in spirit, much like regression models based on the Central Limit Theorem. The scientific utility of such things is not as some universal theoretical solvent, or as the biological equivalent of a unified field theory, but rather as empirical models fitted to data. The science then emerges, with the usual difficulties, from comparisons of one system under different conditions, or between different systems under similar conditions. At best, treating what we have developed as conceptual models, E.C. Pielou’s (1977) important warning regarding the role of mathematical speculation in biological theory remains relevant:

“...[Mathematical models] are easy to devise; even though the assumptions of which they are constructed may be hard to justify, the magic phrase ‘let us assume that...’ overrides objections temporarily. One is then confronted with a much harder task: How is such a model to be tested? The correspondence between a model’s predictions and observed events is sometimes gratifyingly close but this cannot be taken to imply the model’s simplifying assumptions are reasonable in the sense that neglected complications are indeed negligible in their effects...”

In my opinion the usefulness of models is great... [however] it consists *not in answer questions but in raising them*. Models can be used to inspire new field investigations and these are the only source of new knowledge as opposed to new speculation.”

The principal model-based speculation of this work is that mesoscale ecosystem resilience shifts can entrain punctuated events of gene expression and other cognitive phenomena on more rapid time scales, and, in large part through such mechanisms of phenotype expression, slower genetic selection-induced changes, triggering punctuated equilibrium Darwinian evolutionary transitions on geologic time scales.

For human populations, several other layers of information sources, those of (Lamarckian) culture, and of individual and group consciousness and learning, become manifest, producing a rich stew of complicated and interesting phenomena (Wallace, 2004, 2005b; Wallace and Fullilove, 2008).

10 Mathematical Appendix

10.1 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 the random variable X above, so that an output sequence of length n , with the form

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

has joint and conditional probabilities

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

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

Using these probabilities we may calculate the conditional uncertainty

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

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

$$H[\mathbf{X}] \equiv \lim_{n \rightarrow \infty} H(X_n | X_0, X_1, \dots, X_{n-1}). \quad (46)$$

In general

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

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

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

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

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

In general the outputs of the $X_j, j = 0, 1, \dots, n$ 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 stationary in time, that is, the 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

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

(47)

More formally,

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

(48)

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

10.2 The Rate Distortion Theorem

The Shannon-McMillan Theorem can be expressed as the ‘zero error limit’ of the Rate Distortion Theorem (Dembo and Zeitouni, 1998; Cover and Thomas, 1991), which defines a splitting criterion that identifies high probability pairs of sequences. We follow closely the treatment of Cover and Thomas (1991).

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

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

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

$$b^n = b_1, \dots, b_n,$$

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 :

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

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

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

$$(49) \quad d(y, \hat{y}) = 0, y = \hat{y}.$$

$$(53) \quad R(D) = \min_{p(y, \hat{y}); \sum_{(y, \hat{y})} p(y)p(\hat{y})d(y, \hat{y}) \leq D} I(Y, \hat{Y}).$$

For continuous variates the *Squared error distortion* is

$$(50) \quad d(y, \hat{y}) = (y - \hat{y})^2.$$

There are many possibilities.

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

$$(51) \quad d(y^n, \hat{y}^n) = \frac{1}{n} \sum_{j=1}^n d(y_j, \hat{y}_j).$$

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

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

The *average distortion* is defined as

$$(52) \quad D = \sum_{y^n} p(y^n) d(y^n, \hat{y}^n).$$

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:

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

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

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

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 . Cover and Thomas (1991) or Dembo and Zeitouni (1998) provide 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 $n \rightarrow \infty$, the smaller set approaches unit probability, and, for those pairs,

$$(54) \quad p(y^n) \geq p(\hat{y}^n | y^n) \exp[-nI(Y, \hat{Y})].$$

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 Cover and Thomas (1991).

10.3 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, and universality class tuning in the mean field model. These can be reformulated from a Morse theory perspective. Here we follow closely the elegant treatments of Pettini (2007) and Kastner (2006).

The essential idea of Morse theory is to examine an n -dimensional manifold M as decomposed into level sets of some function $f : M \rightarrow \mathbf{R}$ 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 $f : M \rightarrow \mathbf{R}$ 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

$$H_{i,j} = \partial^2 f / \partial x^i \partial x^j,$$

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 H at x_c .

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

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

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

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

[3] If $f : M \rightarrow \mathbf{R}$ 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 $f : M \rightarrow \mathbf{R}$ 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

$\mu_i (i = 1, \dots, m)$ 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 ,

$$\chi = \sum_{i=0}^m (-1)^i \mu_i.$$

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, Pettini (2007) contains both mathematical details and further references. See, for example, Matusmoto (2002) or the classic by Milnor (1963).

10.4 The mean field model

Wallace and Wallace (1998; 1999) have addressed how a language, in a large sense, 'spoken' on a network structure, responds as properties of the network change. The language might be speech, pattern recognition, or cognition. The network might be social, chemical, or neural. The properties of interest were the magnitude of 'strong' or 'weak' ties which, respectively, either disjointly partitioned the network or linked it across such partitioning. These would be analogous to local and mean-field couplings in physical systems.

Fix the magnitude of strong ties – again, those which disjointly partition the underlying network into cognitive or other submodules – but vary the index of nondisjunctive weak ties, P , between components, taking $K = 1/P$.

Assume the piecewise, adiabatically stationary ergodic information source (or sources) dual to cognitive process depends on three parameters, two explicit and one implicit. The explicit are K as above and, as a calculational device, an 'external field strength' analog J , which gives a 'direction' to the system. We will, in the limit, set $J = 0$. Note that many other approaches may well be possible, since renormalization techniques are more philosophy than prescription.

The implicit parameter, r , is an inherent generalized 'length' characteristic of the phenomenon, on which J and K are defined. That is, J and K are written as functions of averages of the parameter r , which may be quite complex, having nothing at all to do with conventional ideas of space. For example r may be defined by the degree of niche partitioning in ecosystems or separation in social structures.

For a given generalized language of interest having a well defined (adiabatically, piecewise stationary) ergodic source uncertainty, $H = H[K, J, \mathbf{X}]$.

To summarize a long train of standard argument (Binney et al., 1986; Wilson, 1971), imposition of invariance of H under a renormalization transform in the implicit parameter r leads to expectation of both a critical point in K , written K_C , reflecting a phase transition to or from collective behavior across the entire array, and of power laws for system behavior near K_C . Addition of other parameters to the system results in a ‘critical line’ or surface.

Let $\kappa \equiv (K_C - K)/K_C$ and take χ as the ‘correlation length’ defining the average domain in r -space for which the information source is primarily dominated by ‘strong’ ties. The first step is to average across r -space in terms of ‘clumps’ of length $R = \langle r \rangle$. Then $H[J, K, \mathbf{X}] \rightarrow H[J_R, K_R, \mathbf{X}]$.

Taking Wilson’s (1971) analysis as a starting point – not the only way to proceed – the ‘renormalization relations’ used here are:

$$H[K_R, J_R, \mathbf{X}] = f(R)H[K, J, \mathbf{X}]$$

$$\chi(K_R, J_R) = \frac{\chi(K, J)}{R},$$

(55)

with $f(1) = 1$ and $J_1 = J, K_1 = K$. The first equation significantly extends Wilson’s treatment. It states that ‘processing capacity,’ as indexed by the source uncertainty of the system, representing the ‘richness’ of the generalized language, grows monotonically as $f(R)$, which must itself be a dimensionless function in R , since both $H[K_R, J_R]$ and $H[K, J]$ are themselves dimensionless. Most simply, this requires replacing R by R/R_0 , where R_0 is the ‘characteristic length’ for the system over which renormalization procedures are reasonable, then setting $R_0 \equiv 1$, hence measuring length in units of R_0 .

Wilson’s original analysis focused on free energy density. Under ‘clumping,’ densities must remain the same, so that if $F[K_R, J_R]$ is the free energy of the clumped system, and $F[K, J]$ is the free energy density before clumping, then Wilson’s equation (4) is $F[K, J] = R^{-3}F[K_R, J_R]$,

$$F[K_R, J_R] = R^3 F[K, J].$$

Remarkably, the renormalization equations are solvable for a broad class of functions $f(R)$, or more precisely, $f(R/R_0), R_0 \equiv 1$.

The second equation just states that the correlation length simply scales as R .

Again, the central feature of renormalization in this context is the assumption that, at criticality, the system looks the same at all scales, that is, it is *invariant under renormalization* at the critical point. All else flows from this.

There is no unique renormalization procedure for information sources: other, very subtle, symmetry relations – not necessarily based on the elementary physical analog we use here – may well be possible. For example, McCauley (1993, p.168) describes the highly counterintuitive renormalizations needed to understand phase transition in simple ‘chaotic’ systems. This is important, since biological or social systems may well alter their renormalization properties – equivalent to tuning their phase transition dynamics – in response to external signals. We will make much use of a simple version of this possibility, termed ‘universality class tuning,’ below.

To begin, following Wilson, take $f(R) = R^d$, d some real number $d > 0$, and restrict K to near the ‘critical value’ K_C . If $J \rightarrow 0$, a simple series expansion and some clever algebra (Wilson, 1971; Binney et al., 1986) gives

$$H = H_0 \kappa^\alpha$$

$$\chi = \frac{\chi_0}{\kappa^s},$$

(56)

where α, s are positive constants. More biologically relevant examples appear below.

Further from the critical point, matters are more complicated, appearing to involve Generalized Onsager Relations, ‘dynamical groupoids’, and a kind of thermodynamics associated with a Legendre transform of $H: S \equiv H - KdH/dK$ (Wallace, 2002a). Although this extension is quite important to describing behaviors away from criticality, the mathematical detail is cumbersome. A more detailed discussion appears at the end of this chapter.

An essential insight is that *regardless of the particular renormalization properties, sudden critical point transition is possible in the opposite direction for this model*. That is, going from a number of independent, isolated and fragmented systems operating individually and more or less at random, into a single large, interlocked, coherent structure, once the parameter K , the inverse strength of weak ties, falls below threshold, or, conversely, once the strength of weak ties parameter $P = 1/K$ becomes large enough.

Thus, increasing nondisjunctive weak ties between them can bind several different cognitive ‘language’ functions into a single, embedding hierarchical metalanguage containing each as a linked subdialect, and do so in an inherently punctuated manner. This could be a dynamic process, creating a shifting, ever-changing pattern of linked cognitive submodules, according to the challenges or opportunities faced by the organism.

This heuristic insight can be made more exact using a rate distortion argument (or, more generally, using the Joint Asymptotic Equipartition Theorem) as follows:

Suppose that two ergodic information sources \mathbf{Y} and \mathbf{B} begin to interact, to ‘talk’ to each other, to influence each other

in some way so that it is possible, for example, to look at the output of \mathbf{B} – strings b – and infer something about the behavior of \mathbf{Y} from it – strings y . We suppose it possible to define a retranslation from the B-language into the Y-language through a deterministic code book, and call $\hat{\mathbf{Y}}$ the translated information source, as mirrored by \mathbf{B} .

Define some distortion measure comparing paths y to paths \hat{y} , $d(y, \hat{y})$. Invoke the Rate Distortion Theorem’s mutual information $I(Y, \hat{Y})$, which is the splitting criterion between high and low probability pairs of paths. Impose, now, a parametrization by an inverse coupling strength K , and a renormalization representing the global structure of the system coupling. This may be much different from the renormalization behavior of the individual components. If $K < K_C$, where K_C is a critical point (or surface), the two information sources will be closely coupled enough to be characterized as condensed.

In the absence of a distortion measure, the Joint Asymptotic Equipartition Theorem gives a similar result.

Detailed coupling mechanisms will be sharply constrained through regularities of grammar and syntax imposed by limit theorems associated with phase transition.

10.4.1 Biological renormalization

Next the mathematical detail concealed by the invocation of the asymptotic limit theorems emerges with a vengeance. Equation (55) states that the information source and the correlation length, the degree of coherence on the underlying network, scale under renormalization clustering in chunks of size R as

$$H[K_R, J_R]/f(R) = H[J, K]$$

$$\chi[K_R, J_R]R = \chi(K, J),$$

with $f(1) = 1, K_1 = K, J_1 = J$, where we have slightly rearranged terms.

Differentiating these two equations with respect to R , so that the right hand sides are zero, and solving for dK_R/dR and dJ_R/dR gives, after some consolidation, expressions of the form

$$dK_R/dR = u_1 d\log(f)/dR + u_2/R$$

$$dJ_R/dR = v_1 J_R d\log(f)/dR + \frac{v_2}{R} J_R.$$

(57)

The $u_i, v_i, i = 1, 2$ are functions of K_R, J_R , but not explicitly of R itself.

We expand these equations about the critical value $K_R = K_C$ and about $J_R = 0$, obtaining

$$dK_R/dR = (K_R - K_C)y d\log(f)/dR + (K_R - K_C)z/R$$

$$dJ_R/dR = w J_R d\log(f)/dR + x J_R/R.$$

(58)

The terms $y = du_1/dK_R|_{K_R=K_C}, z = dv_2/dK_R|_{K_R=K_C}, w = v_1(K_C, 0), x = v_2(K_C, 0)$ are constants.

Solving the first of these equations gives

$$K_R = K_C + (K - K_C)R^z f(R)^y,$$

(59)

again remembering that $K_1 = K, J_1 = J, f(1) = 1$.

Wilson’s essential trick is to iterate on this relation, which is supposed to converge rapidly near the critical point (Binney et al., 1986), assuming that for K_R near K_C , we have

$$K_C/2 \approx K_C + (K - K_C)R^z f(R)^y.$$

(60)

We iterate in two steps, first solving this for $f(R)$ in terms of known values, and then solving for R , finding a value R_C that we then substitute into the first of equations (55) to obtain an expression for $H[K, 0]$ in terms of known functions and parameter values.

The first step gives the general result

$$f(R_C) \approx \frac{[K_C/(K_C - K)]^{1/y}}{2^{1/y} R_C^{z/y}}.$$

(61)

Solving this for R_C and substituting into the first expression of equation (55) gives, as a first iteration of a far more general procedure (Shirkov and Kovalev, 2001), the result

$$H[K, 0] \approx \frac{H[K_C/2, 0]}{f(R_C)} = \frac{H_0}{f(R_C)}$$

$$\chi(K, 0) \approx \chi(K_C/2, 0)R_C = \chi_0 R_C,$$

(62)

which are the essential relationships.

Note that a power law of the form $f(R) = R^m, m = 3$, which is the direct physical analog, may not be biologically reasonable, since it says that ‘language richness’ can grow very rapidly as a function of increased network size. Such rapid growth is simply not observed.

Taking the biologically realistic example of non-integral ‘fractal’ exponential growth,

$$f(R) = R^\delta,$$

(63)

where $\delta > 0$ is a real number which may be quite small, equation (6.17) can be solved for R_C , obtaining

$$R_C = \frac{[K_C/(K_C - K)]^{1/(\delta y + z)}}{2^{1/(\delta y + z)}}$$

(64)

for K near K_C . Note that, for a given value of y , one might characterize the relation $\alpha \equiv \delta y + z = \text{constant}$ as a ‘tunable universality class relation’ in the sense of Albert and Barabasi (2002).

Substituting this value for R_C back into equation (61) gives a complex expression for H , having three parameters: δ, y, z .

A more biologically interesting choice for $f(R)$ is a logarithmic curve that ‘tops out’, for example

$$f(R) = m \log(R) + 1.$$

(65)

Again $f(1) = 1$.

Using Mathematica 4.2 or above to solve equation (61) for R_C gives

$$R_C = \left[\frac{Q}{\text{LambertW}[Q \exp(z/my)]} \right]^{y/z},$$

(66)

where

$$Q \equiv (z/my)2^{-1/y}[K_C/(K_C - K)]^{1/y}.$$

The transcendental function $\text{LambertW}(x)$ is defined by the relation

$$\text{LambertW}(x) \exp(\text{LambertW}(x)) = x.$$

It arises in the theory of random networks and in renormalization strategies for quantum field theories.

An asymptotic relation for $f(R)$ would be of particular biological interest, implying that ‘language richness’ increases to a limiting value with population growth. Such a pattern is broadly consistent with calculations of the degree of allelic heterozygosity as a function of population size under a balance between genetic drift and neutral mutation (Hartl and Clark, 1997; Ridley, 1996). Taking

$$f(R) = \exp[m(R - 1)/R]$$

(67)

gives a system which begins at 1 when $R = 1$, and approaches the asymptotic limit $\exp(m)$ as $R \rightarrow \infty$. Mathematica 4.2 finds

$$R_C = \frac{my/z}{\text{LambertW}[A]},$$

(68)

where

$$A \equiv (my/z) \exp(my/z) [2^{1/y} [K_C/(K_C - K)]^{-1/y}]^{y/z}.$$

These developments indicate the possibility of taking the theory significantly beyond arguments by abduction from simple physical models, although the notorious difficulty of implementing information theory existence arguments will undoubtedly persist.

10.4.2 Universality class distribution

Physical systems undergoing phase transition usually have relatively pure renormalization properties, with quite different systems clumped into the same ‘universality class,’ having fixed exponents at transition (Binney et al., 1986). Biological and social phenomena may be far more complicated:

If the system of interest is a mix of subgroups with different values of some significant renormalization parameter m in the expression for $f(R, m)$, according to a distribution $\rho(m)$, then the first expression in equation (55) should generalize, at least to first order, as

$$\begin{aligned} H[K_R, J_R] &= \langle f(R, m) \rangle H[K, J] \\ &\equiv H[K, J] \int f(R, m) \rho(m) dm. \end{aligned} \tag{69}$$

If $f(R) = 1 + m \log(R)$ then, given any distribution for m ,

$$\langle f(R) \rangle = 1 + \langle m \rangle \log(R) \tag{70}$$

where $\langle m \rangle$ is simply the mean of m over that distribution.

Other forms of $f(R)$ having more complicated dependencies on the distributed parameter or parameters, like the power law R^δ , do not produce such a simple result. Taking $\rho(\delta)$ as a normal distribution, for example, gives

$$\langle R^\delta \rangle = R^{\langle \delta \rangle} \exp[(1/2)(\log(R^\sigma))^2], \tag{71}$$

where σ^2 is the distribution variance. The renormalization properties of this function can be determined from equation (61), and the calculation is left to the reader as an exercise, best done in Mathematica 4.2 or above.

Thus the information dynamic phase transition properties of mixed systems will not in general be simply related to those of a single subcomponent, a matter of possible empirical importance: If sets of relevant parameters defining renormalization universality classes are indeed distributed, experiments

observing pure phase changes may be very difficult. Tuning among different possible renormalization strategies in response to external signals would result in even greater ambiguity in recognizing and classifying information dynamic phase transitions.

Important aspects of mechanism may be reflected in the combination of renormalization properties and the details of their distribution across subsystems.

In sum, real biological, social, or interacting biopsychosocial systems are likely to have very rich patterns of phase transition which may not display the simplistic, indeed, literally elemental, purity familiar to physicists. Overall mechanisms will, however, still remain significantly constrained by the theory, in the general sense of probability limit theorems.

10.4.3 Punctuated universality class tuning

The next step is to iterate the general argument onto the process of phase transition itself, producing a model of consciousness as a tunable neural workspace subject to inherent punctuated detection of external events.

As described above, an essential character of physical systems subject to phase transition is that they belong to particular ‘universality classes’. Again, this means that the exponents of power laws describing behavior at phase transition will be the same for large groups of markedly different systems, with ‘natural’ aggregations representing fundamental class properties (Binney et al., 1986).

It appears that biological or social systems undergoing phase transition analogs need not be constrained to such classes, and that ‘universality class tuning’, meaning the strategic alteration of parameters characterizing the renormalization properties of punctuation, might well be possible. Here we focus on the tuning of parameters within a single, given, renormalization relation. Clearly, however, wholesale shifts of renormalization properties must ultimately be considered as well, a matter for future work.

Universality class tuning has been observed in models of ‘real world’ networks. As Albert and Barabasi (2002) put it,

“The inseparability of the topology and dynamics of evolving networks is shown by the fact that [the exponents defining universality class] are related by [a] scaling relation..., underlying the fact that a network’s assembly uniquely determines its topology. However, in no case are these exponents unique. They can be tuned continuously...”

Suppose that a structured external environment, itself an appropriately regular information source \mathbf{Y} , ‘engages’ a modifiable cognitive system. The environment begins to write an image of itself on the cognitive system in a distorted manner permitting definition of a mutual information $I[K]$ splitting criterion according to the Rate Distortion or Joint Asymptotic Equipartition Theorems. K is an inverse coupling parameter between system and environment. At punctuation – near some critical point K_C – the systems begin to interact

very strongly indeed, and, near K_C , using the simple physical model of equation (56),

$$I[K] \approx I_0 \left[\frac{K_C - K}{K_C} \right]^\alpha.$$

For a physical system α is fixed, determined by the underlying ‘universality class.’ Here we will allow α to vary, and, in the section below, to itself respond explicitly to signals.

Normalizing K_C and I_0 to 1,

$$I[K] \approx (1 - K)^\alpha. \quad (72)$$

The horizontal line $I[K] = 1$ corresponds to $\alpha = 0$, while $\alpha = 1$ gives a declining straight line with unit slope which passes through 0 at $K = 1$. Consideration shows there are progressively sharper transitions between the necessary zero value at $K = 1$ and the values defined by this relation for $0 < K, \alpha < 1$. The rapidly rising slope of transition with declining α is of considerable significance:

The instability associated with the splitting criterion $I[K]$ is defined by

$$Q[K] \equiv -KdI[K]/dK = \alpha K(1 - K)^{\alpha-1}, \quad (73)$$

and is singular at $K = K_C = 1$ for $0 < \alpha < 1$. Following earlier work (Wallace and Wallace, 1998, 1999; Wallace and Fullilove, 1999; Wallace, 2002a), we interpret this to mean that values of $0 < \alpha \ll 1$ are highly unlikely for real systems, since $Q[K]$, in this model, represents a kind of barrier for ‘social’ information systems, in particular interacting neural network modules, a matter explored further below.

On the other hand, smaller values of α mean that the system is far more efficient at responding to the adaptive demands imposed by the embedding structured environment, since the mutual information which tracks the matching of internal response to external demands, $I[K]$, rises more and more quickly toward the maximum for smaller and smaller α as the inverse coupling parameter K declines below $K_C = 1$. That is, systems able to attain smaller α are more responsive to external signals than those characterized by larger values, in this model, but smaller values will be harder to reach, probably only at some considerable physiological or opportunity cost. Focused conscious action takes resources, of one form or another.

Wallace (2005a) makes these considerations explicit, modeling the role of contextual and energy constraints on the relations between Q , I , and other system properties.

The more biologically realistic renormalization strategies given above produce sets of several parameters defining the universality class, whose tuning gives behavior much like that of α in this simple example.

Formal iteration of the phase transition argument on this calculation gives tunable consciousness, focusing on paths of universality class parameters.

Suppose the renormalization properties of a language-on-a-network system at some ‘time’ k are characterized by a set of parameters $A_k \equiv \alpha_1^k, \dots, \alpha_m^k$. Fixed parameter values define a particular universality class for the renormalization. We suppose that, over a sequence of ‘times,’ the universality class properties can be characterized by a path $x_n = A_0, A_1, \dots, A_{n-1}$ having significant serial correlations which, in fact, permit definition of an adiabatically piecewise stationary ergodic information source associated with the paths x_n . We call that source \mathbf{X} .

Suppose also, in the now-usual manner, that the set of external (or internal, systemic) signals impinging on consciousness is also highly structured and forms another information source \mathbf{Y} which interacts not only with the system of interest globally, but specifically with its universality class properties as characterized by \mathbf{X} . \mathbf{Y} is necessarily associated with a set of paths y_n .

Pair the two sets of paths into a joint path, $z_n \equiv (x_n, y_n)$ and invoke an inverse coupling parameter, K , between the information sources and their paths. This leads, by the arguments above, to phase transition punctuation of $I[K]$, the mutual information between \mathbf{X} and \mathbf{Y} , under either the Joint Asymptotic Equipartition Theorem or under limitation by a distortion measure, through the Rate Distortion Theorem. The essential point is that $I[K]$ is a splitting criterion under these theorems, and thus partakes of the homology with free energy density which we have invoked above.

Activation of universality class tuning, the mean field model’s version of attentional focusing, then becomes itself a punctuated event in response to increasing linkage between the organism and an external structured signal or some particular system of internal events.

This iterated argument exactly parallels the extension of the General Linear Model to the Hierarchical Linear Model in regression theory (Byrk and Raudenbusch, 2001).

Another path to the fluctuating dynamic threshold might be through a second order iteration similar to that just above, but focused on the parameters defining the universality class distributions given above.

10.4.4 A network of dynamic manifolds and its tuning

The set of universality class tuning parameters, A_k , defines a manifold whose topology could also be more fully analyzed using Morse theory. That is an equivalence class of dynamic manifolds is determined, not by universality class, which is tunable, but by the underlying form of the renormalization relation, in the sense of the many different possible renormalization symmetries described above. Thus the possible higher

level dynamic manifolds in this model are characterized by fixed renormalization relations, but tunable universality class parameters. One can then invoke a crosstalk coupling within a groupoid network of different dynamic manifolds defined by these renormalization relations, leading to the same kind of Morse theoretic analysis of the higher level topological structure.

11 References

Adami C., and N. Cerf, 2000, Physical complexity of symbolic sequences, *Physica D*, 137:62-69.

Adami C., C. Ofria, and T. Collier, 2000, Evolution of biological complexity, *Proceedings of the National Academy of Sciences*, 97:4463-4468.

Albert R., and A. Barabasi, 2002, Statistical mechanics of complex networks, *Reviews of Modern Physics*, 74:47-97.

Ash, 1990, *Information Theory*, Dover Publications, New York.

Atlan H. and I. Cohen, 1998, Immune information, self-organization and meaning, *International Immunology*, 10:711-717.

Auslander L., 1967, *Differential Geometry*, Harper and Row, New York.

Avital E., and E. Jablonka, 2000, *Animal Traditions: Behavioral inheritance in evolution*, Cambridge University Press, New York.

Baker M., and J. Stock, 2007, Signal transduction: networks and integrated circuits in bacterial cognition, *Current Biology*, 17(4):R1021-4.

Barkow J., L. Cosmides and J. Tooby, eds., 1992, *The Adapted Mind: Biological Approaches to Mind and Culture*, University of Toronto Press.

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

Binney J., N. Dowrick, A. Fisher, and M. Newman, 1986, *The theory of critical phenomena*, Clarendon Press, Oxford, UK.

Bonner J., 1980, *The evolution of culture in animals*, Princeton University Press, Princeton, NJ.

Burago D., Y. Burago, and S. Ivanov, 2001, *A Course in Metric Geometry*, American Mathematical Society, Providence, RI.

Byrk A., and S. Raudenbusch, 2001, *Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage Publications, New York.

Champagnat N., R. Ferriere, and S. Meleard, 2006, Unifying evolutionary dynamics: From individual stochastic processes to macroscopic models, *Theoretical Population Biology*, 69:297-321.

Cohen I., 1992, The cognitive principle challenges clonal selection, *Immunology Today*, 13:441-444.

Cohen I., 2000, *Tending Adam's Garden: Evolving the Cognitive Immune Self*, Academic Press, New York.

Cohen I., 2006, Immune system computation and the immunological homunculus, in O. Nierstrasz et al. (eds), MoD-

ELS 2006, *Lecture Notes in Computer Science*, 4199:499-512.

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

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

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

Dercole F., R. Ferriere, A. Gragnani and S. Rinaldi, 2006, Coevolution of slow-fast populations: evolutionary sliding, evolutionary pseudo-equilibria and complex Red Queen dynamics, *Proceedings of the Royal Society, B*, 273:983-990.

Diekmann U., and R. Law, 1996, The dynamical theory of coevolution: a derivation from stochastic ecological processes, *Journal of Mathematical Biology*, 34:579-612.

Dimitrov A., and J. Miller, 2001, Neural coding and decoding: communication channels and quantization, *Computation and Neural Systems*, 12:441-472.

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

Durham W., 1991, *Coevolution: Genes, Culture, and Human Diversity*, Stanford University Press, Palo Alto, CA.

Eldredge N., 1985, *Time Frames: The Rethinking of Darwinian Evolution and the Theory of Punctuated Equilibria*, Simon and Schuster, New York.

Emery M., 1989, *Stochastic Calculus in Manifolds*, Universitext series, Springer, New York.

Fath B., H. Cabezas, and C. Pawlowski, 2003, Regime changes in ecological systems: an information theory approach, *Journal of Theoretical Biology*, 222:517-530.

Feller W., 1971, *An Introduction to Probability Theory and Its Applications*, John Wiley and Sons, New York.

Feynman R., 1996, *Feynman Lectures on Computation*, Addison-Wesley, Reading, MA.

Flegal K., M. Carroll, C. Ogden, and C. Johnson, 2002, Prevalence and trends in obesity among US adults, 1999-2000, *Journal of the American Medical Association*, 288:1723-1727.

Fleming, R., and C. Shoemaker, 1992, Evaluating models for spruce budworm-forest management: comparing output with regional field data, *Ecological Applications*, 2:460-477.

Forlenza M., and A. Baum, 2000, Psychosocial influences on cancer progression: alternative cellular and molecular mechanisms, *Current Opinion in Psychiatry*, 13:639-645.

Franzosi R., and M. Pettini, 2004, Theorem on the origin of phase transitions, *Physical Review Letters*, 92:060601.

Fredlin M., and A. Wentzell, 1998, *Random Perturbations of Dynamical Systems*, Springer, New York.

Glazebrook J., and R. Wallace, 2007, Rate distortion manifolds as model spaces for cognitive information. Submitted.

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

Goubault E., and M. Raussen, 2002, Dihomotopy as a tool in state space analysis, *Lecture Notes in Computer Science*, Vol. 2286, April, 2002, pp. 16-37. Springer, New York.

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

- Grossman, Z., 2000, Round 3, *Seminars in Immunology*, 12:313-318.
- Gunderson L., 2000, Ecological resilience – in theory and application, *Annual Reviews of Ecological Systematics*, 31:425-439.
- Gunderson L., 2007, Personal communication.
- Hartl D., and A. Clark, 1997, *Principles of Population Genetics*, Sinaur Associates, Sunderland, MA.
- Holling C., 1973, 1973, Resilience and stability of ecological systems, *Annual Reviews of Ecological Systematics*, 4:1-23.
- Holling C., 1992, Cross-scale morphology, geometry and dynamics of ecosystems, *Ecological Monographs*, 41:1-50.
- Jimenez-Montano M., 1989, Formal languages and theoretical molecular biology, in *Theoretical Biology: Epigenetic and Evolutionary Order in Complex Systems*, B. Goodwin and P. Saunders (eds.), Edinburgh University Press.
- Kastner M., 2006, Phase transitions and configuration space topology, arXiv preprint cond-mat/0703401
- Khinchin A., 1957, *Mathematical Foundations of Information Theory*, Dover Publications, New York.
- Laland K., F. Odling-Smee, and M. Feldman, 1999, Evolutionary consequences of niche construction and their implications for ecology, *Proceedings of the National Academy of Sciences*, 96:10242-10247.
- Lee J., 2000, *Introduction to Topological Manifolds*, Springer, New York.
- Levin S., 1989, Ecology in theory and application, in *Applied Mathematical Ecology*, S. Levin, T. Hallam, and L. Gross (eds.), biomathematics Texts 18, Springer-Verlag, New York.
- Lewontin R., 1993, *Biology as Ideology: The Doctrine of DNA*, Harper Collins, New York.
- Lewontin R., 2000, *The Triple Helix: gene, organism, and environment*, Harvard University Press.
- Liao J., R. Biscolo, Y. Yang, L. My Tran, C. Sabatti, and V. Roychowdhury, 2003, Network component analysis: Reconstruction of regulatory signals in biological systems, *Proceedings of the National Academy of Sciences*, 100:15522-15527.
- Liu, Y., and M. Ringner, 2007, Revealing signaling pathway deregulation by using gene expression signatures and regulatory motif analysis, *Genome Biology*, 8:R77.
- Luchinsky D., 1997, On the nature of large fluctuations in equilibrium systems: observations of an optimal force, *Journal of Physics A*, 30:L577-L583.
- Matsumoto Y., 2002, *An Introduction to Morse Theory*, Translations of Mathematical Monographs, Vol. 208, American Mathematical Society.
- Michel L., and J. Mozrymas, 1977, Application of Morse Theory to the symmetry breaking in the Landau theory of the second order phase transition, in *Group Theoretical Methods in Physics: Sixth International Colloquium*, eds. P. Kramer, A. Rieckers, Lecture Notes in Physics, Vol. 79, pp. 447-461, Springer, New York.
- Milnor J., 1963, *Morse Theory*, Annals of Mathematical Studies, Vol. 51, Princeton University Press.
- Nunney L., 1999, Lineage selection and the evolution of multistage carcinogenesis, *Proceedings of the London Royal Society B*, 266:493-498.
- Ofria C., C. Adami, and T. Collier, 2003, Selective pressures on genomes in molecular evolution, *Journal of Theoretical Biology*, 222:477-483.
- Onsager L., and S. Machlup, 1953, Fluctuations and irreversible processes, *Physical Review*, 91:1505-1512.
- O’Nuallain S., 2006, Context in computational linguistics and gene expression, <http://www-cls.stanford.edu/events/Coglunch/nuallain-2006>.
- Pettini M., 2007, *Geometry and Topology in Hamiltonian Dynamics and Statistical Mechanics*, Springer, New York.
- Pielou E.C., 1977, *Mathematical Ecology*, John Wiley and Sons, New York.
- Pattee H., 1972, Laws and constraints, symbols, and languages, in C. Waddington (ed.), *Towards a Theoretical Biology: Essays*, Aldine-Atherton, Chicago.
- Pettini, M., 2007, *Geometry and Topology in Hamiltonian Dynamics and Statistical Mechanics*, Springer, New York.
- Podolsky S., and A. Tauber, 1998, *The generation of diversity: Clonal selection theory and the rise of molecular biology*, Harvard University Press.
- Pratt V., 1991, Modeling concurrency with geometry, *Proceedings of the 18th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, 311-322.
- Richerson P. and R. Boyd, 1995, The evolution of human hypersociality, Paper for Rindberg Castle Symposium on Ideology, Warfare, and Indoctrination, (January, 1995), and HBES meeting, 1995.
- Ricotta C., 2003, Additive partition of parametric information and its associated β -diversity measure, *Acta Biotheoretica*, 51:91-100.
- Ridley M., 1996, *Evolution*, Second Edition, Blackwell Science, Oxford University Press.
- Rojdestvensky I. and M. Cottam, 2000, Mapping of statistical physics to information theory with applications to biological systems, *Journal of Theoretical Biology*, 202:43-54.
- Sayed-Ahmad, A., K. Tuncay, and P. Ortoleva, 2007, Transcriptional regulatory network refinement and quantification through kinetic modeling, gene expression microarray data and information theory, *BMC Bioinformatics*, 8:20.
- Shirkov D., and V. Kovalev, 2001, The Bogoliubov renormalization group and solution symmetry in mathematical physics, *Physics Reports*, 352:219-249.
- Soyer O., M. Salathe, and S. Bonhoeffer, 2006, Signal transduction networks: Topology, response and biochemical processes, *Journal of Theoretical Biology*, 238:416-425.
- Tauber A., 1998, Conceptual shifts in immunology: Comments on the ‘two-way paradigm’, *Theoretical Medicine and Bioethics*, 19:457-473.
- Thom R., 1972, Structuralism and biology, in C. Waddington (ed.), *Towards a Theoretical Biology: Essays*, Aldine-Atherton, Chicago.
- Volney, W., and R. Fleming, 2007, Spruce budworm (*Choristoneura* spp.) biotype reactions to forest and climate characteristics, *Global Change Biology*, 13:1630-1643.
- Waddington C., 1972, Epilogue, in C. Waddington (ed.), *Towards a Theoretical Biology: Essays*, Aldine-Atherton, Chicago.

Wallace R., and R.G. Wallace, 2008, *Psychopathica Automatorum: A cognitive neuroscience perspective on highly parallel computation and its dysfunctions*. To appear.

Wallace R., and M. Fullilove, 2008, *Collective Consciousness and its Discontents: Institutional Distributed Cognition, Racial Policy and Public Health in the United States*, Springer, New York.

Wallace R. and R.G. Wallace, 1998, Information theory, scaling laws and the thermodynamics of evolution, *Journal of Theoretical Biology*, 192:545-559.

Wallace R., and R.G. Wallace, 1999, Organisms, organizations, and interactions: an information theory approach to biocultural evolution, *BioSystems*, 51:101-119.

Wallace R. and R.G. Wallace, 2002, Immune cognition and vaccine strategy: beyond genomics, *Microbes and Infection*, 4:521-527.

Wallace R., D. Wallace, and R.G. Wallace, 2003, Toward cultural oncology: the evolutionary information dynamics of cancer, *Open Systems and Information Dynamics*, 10:159-181.

Wallace R., 2000, Language and coherent neural amplification in hierarchical systems: renormalization and the dual information source of a generalized stochastic resonance, *International Journal of Bifurcation and Chaos*, 10:493-502.

Wallace R., 2002a, Immune cognition and vaccine strategy: pathogenic challenge and ecological resilience, *Open Systems and Information Dynamics*, 9:51-83.

Wallace R., 2002b, Adaptation, punctuation and rate distortion: non-cognitive 'learning plateaus' in evolutionary process, *Acta Biotheoretica*, 50:101-116.

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

Wallace R., 2005b, A global workspace perspective on mental disorders, *Theoretical Biology and Medical Modelling*, 2:49.

Whitham T., et al., 2006, A framework for community and ecosystem genetics: from genes to ecosystems, *nature reviews:genetics*, 7:510-523.

Wilson, K., 1971, Renormalization group and critical phenomena. I Renormalization group and the Kadanoff scaling picture, *Physical Review B*, 4:3174-3183.

Wymer C., 1997, Structural nonlinear continuous-time models in econometrics, *Macroeconomic Dynamics*, 1:518-548.

Zhu R., A. Riberio, D. Salahub, and S. Kauffman, 2007, Studying genetic regulatory networks at the molecular level: Delayed reaction stochastic models, *Journal of Theoretical Biology*, 246:725-745.

Zurek W., 1985, Cosomological experiments in superfluid helium? *Nature*, 317:505-508.

Zurek W., 1996, Shards of broken symmetry, *Nature*, 382:296-298.