

IMMUNE COGNITION AND PATHOGENIC CHALLENGE: SUDDEN AND CHRONIC INFECTION

Rodrick Wallace, PhD
The New York State Psychiatric Institute*

January 10, 2002

Abstract

We continue to study the implications of IR Cohen's theory of immune cognition, in the presence of both sudden and chronic pathogenic challenge, through a mathematical model derived from the Large Deviations Program of applied probability. The analysis makes explicit the linkage between an individual's 'immunocultural condensation' and embedding social or historical structures and processes, in particular power relations between groups. We use methods adapted from the theory of ecosystem resilience to explore the consequences of the sudden 'perturbation' caused by infection in the context of such embedding, and examine a 'stage' model for chronic infection involving multiple phase transitions analogous to 'learning plateaus' in neural networks or 'punctuated equilibria' in adaptive systems.

Introduction

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

A recent paper by R Wallace and RG Wallace (2001) broadly explored the use of Atlan and Cohen's (1998) information theory treatment of immune cognition for reinterpreting two important population case histories of chronic infection in the context of 'cultural variation,' namely malaria in Burkina Faso and 'heterosexual AIDS' in the US. Reductionist biomedical paradigm had assigned differences of immune response and disease expression between populations in a manner so as to reify 'race'. That is, different responses to pathogenic challenge were attributed entirely to genetic differences between ethnic groups, ignoring the considerable impact of 'path dependent' patterns of past and current power relations between coresident communities.

Reductionist perspective, at least on this matter, thus appears to incorporate a systematic logical fallacy which could significantly, and possibly profoundly, distort attempts to develop effective vaccine strategies against a number of disorders.

Here we present details of the mathematical model described in R Wallace and RG Wallace (2001) which allows the interactive condensation of cognitive systems with the 'selection pressures' of embedding social constraints. The work relies on foundations supplied by the Large Deviations Program of applied probability, which provides a unified vocabulary for addressing language structures, in a large sense, using mathematical techniques adapted from statistical mechanics and the theory of fluctuations.

The essence of the approach is to express the cognitive pattern recognition-and-response described as characterizing immune cognition by Atlan and Cohen (1998) in terms of a 'language,' in a broad sense, and then to show how that language can interact and coalesce with similar cognitive languages at larger scales – central nervous system (CNS) and the embedding local socio-cultural network. The next step is to demonstrate that these may, in turn, interact – and indeed coalesce – with a non-cognitive structured language of externally imposed constraint. The process of developing the formalism will make clear that change in such languages takes place 'on the surface,' in a sense, of what has gone before, so that path dependence – the persisting burden of history – becomes a natural outcome.

Next we describe the behavior of condensed cognitive systems under the perturbation of a sudden pathogenic challenge, using a generalization of Ives' (1995) and Holling's (1973, 1992) ecosystem resilience analysis. Individuals within more resilient social structures – typically those of the powerful rather than of the subordinate – will usually suffer less 'amplification' of symptom patterns caused by the infection.

Finally we examine chronic pathogenic challenge from this perspective, and propose a ‘stage’ model formally analogous to learning plateaus in cognitive, and ‘punctuated equilibrium’ in adaptive, systems.

We begin with a summary of relevant theory.

Information theory preliminaries

Suppose we have an ordered set of random variables, X_k , at ‘times’ $k = 1, 2, \dots$ – which we call \mathbf{X} – that emits sequences taken from some fixed alphabet of possible outcomes. Thus an output sequence of length n , x_n , termed a path, will have the form

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

where α_k is the value at step k of the stochastic variate X_k ,

$$X_k = \alpha_k.$$

A particular sequence x_n will have the probability

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

with associated conditional probabilities

$$(2) \quad P(X_n = \alpha_n | X_{n-1} = \alpha_{n-1}, \dots, X_0 = \alpha_0).$$

Thus substrings of x_n are not, in general, stochastically independent. That is, there may be powerful serial correlations along the x_n . We call \mathbf{X}

an information source, and are particularly interested in sources for which the long run frequencies of strings converge stochastically to their time-independent probabilities, generalizing the law of large numbers. These we call *ergodic* (Ash, 1990; Cover and Thomas, 1991; Khinchine, 1957, hereafter known as ACTK). If the probabilities of strings do not change in time, the source is called *memoryless*.

We shall be interested in sources which can be parametrized and that are, with respect to that parameter, *piecewise adiabatically memoryless*, i.e. probabilities closely track parameter changes within a ‘piece,’ but may change suddenly between pieces. This allows us to apply the simplest results from information theory, and to use renormalization methods to examine transitions between ‘pieces.’ Learning plateaus represent regions where, with respect to the parameter, the system is, to first approximation, adiabatically memoryless in this sense, analogous to adiabatic physical systems in which rapidly changing phenomena closely track slowly varying driving parameters. In what follows we use the term ‘ergodic,’ to mean ‘piecewise adiabatically memoryless ergodic.’

For any ergodic information source it is possible to divide all possible sequences of output, in the limit of large n , into two sets, S_1 and S_2 , having, respectively, very high and very low probabilities of occurrence. Sequences in S_1 we call *meaningful*.

The content of information theory’s Shannon-McMillan Theorem is twofold:

First, if there are $N(n)$ meaningful sequences of length n , where $N(n) \ll$ than the number of all possible sequences of length n , then, for each ergodic information source \mathbf{X} , there is a unique, path-independent number $H[\mathbf{X}]$ such that

$$\lim_{n \rightarrow \infty} \frac{\log[N(n)]}{n} = H[\mathbf{X}]. \quad (3)$$

See ACTK for details.

Thus, for large n , the probability of *any* meaningful path of length $n \gg 1$ – independent of path – is approximately

$$P(x_n \in S_1) \propto \exp(-nH[\mathbf{X}]) \propto 1/N(n).$$

(3)

This is the *asymptotic equipartition property* and the Shannon-McMillan Theorem is often called the Asymptotic Equipartition Theorem (AEPT).

$H[\mathbf{X}]$ is the *splitting criterion* between the two sets S_1 and S_2 , and the second part of the Shannon-McMillan Theorem involves its calculation. This requires introduction of some nomenclature.

Suppose we have stochastic variables X and Y which take the values x_j and y_k with probability distributions

$$P(X = x_j) = P_j$$

$$P(Y = y_k) = P_k$$

Let the joint and conditional probability distributions of X and Y be given, respectively, as

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

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

The *Shannon uncertainties* of X and of Y are, respectively

$$H(X) = - \sum_j P_j \log(P_j)$$

$$H(Y) = - \sum_k P_k \log(P_k)$$

(4)

The *joint uncertainty* of X and Y is defined as

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

(5)

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

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

(6)

Note that by expanding $P(y_k|x_j)$ we obtain

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

The second part of the Shannon-McMillan Theorem states that the – path independent – splitting criterion, $H[\mathbf{X}]$, of the ergodic information source \mathbf{X} , which divides high from low probability paths, is given in terms of the sequence probabilities of equations (1) and (2) as

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

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

(7)

The AEPT is one of the most unexpected and profound results of 20th Century applied mathematics.

Ash (1990) describes the uncertainty of an ergodic information source as follows;

“...[W]e may regard a portion of text in a particular language as being produced by an information source. the probabilities $P[X_n = \alpha_n | X_0 = \alpha_0, \dots, X_{n-1} = \alpha_{n-1}]$ may be estimated from the available data about the language. A large uncertainty means, by the AEPT, a large number of ‘meaningful’ sequences. Thus given two languages with uncertainties H_1 and H_2 respectively, if $H_1 > H_2$, then in the absence of noise it is easier to communicate in the first language; more can be said in the same amount of time. On the other hand, it will be easier to reconstruct a scrambled portion of text in the second language, since fewer of the possible sequences of length n are meaningful.”

Languages can affect each other, or, equivalently, systems can translate from one language to another, usually with error. The Rate Distortion Theorem, which is one generalization of the SMT, describes how this can take place. As IR Cohen (2001) has put it, in the context of the cognitive immune system (IR Cohen, 1992, 2000),

“An immune response is like a key to a particular lock; each immune response amounts to a functional image of the stimulus that elicited the response. Just as a key encodes a functional image of its lock, an effective [immune] response encodes a functional image of its stimulus; the stimulus and the response fit each other. The immune system, for example, has to deploy different types of inflammation to heal a broken bone, repair an infarction, effect neuroprotection, cure hepatitis, or contain tuberculosis. Each aspect of the response is a functional representation of the challenge.

Self-organization allows a system to adapt, to update itself in the image of the world it must respond to... The immune system, like the brain... aim[s] at representing a part of the world.”

These considerations suggest that the degree of possible back-translation between the world and its image within a cognitive system represents the profound and systematic coupling between a biological system and its environment, a coupling which may particularly express the way in which the system has ‘learned’ the environment. We attempt a formal treatment, from which it will appear that both cognition and response to systematic patterns of selection pressure are – almost inevitably – highly punctuated by ‘learning plateaus’ in which the two processes can become inextricably intertwined.

Suppose we have a ergodic information source \mathbf{Y} , a generalized language having grammar and syntax, with a source uncertainty $H[\mathbf{Y}]$ that ‘perturbs’ a system of interest. A chain of length n , a path of perturbations, has the form

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

Suppose that chain elicits a corresponding chain of responses from the system of interest, producing another path $b^n = (b_1, \dots, b_n)$, which has some ‘natural’ translation into the language of the perturbations, although not, generally, in a one-to-one manner. The image is of a continuous analog audio signal which has been ‘digitized’ into a discrete set of voltage values. Thus, there may well be several different y^n corresponding to a given ‘digitized’ b^n . Consequently, in translating back from the b-language into the y-language, there will generally be information loss.

Suppose, however, that with each path b^n we specify an inverse code which identifies exactly one path \hat{y}^n . We assume further there is a measure of distortion which compares the real path y^n with the inferred inverse \hat{y}^n . Below we follow the nomenclature of Cover and Thomas (1991).

The *Hamming distortion* is defined as

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

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

(8)

For continuous variates the *Squared error distortion* is defined as

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

(9)

Possibilities abound.

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

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

(10)

We suppose that with each path y^n and b^n -path translation into the y -language, denoted \hat{y}^n , there are associated individual, joint and conditional probability distributions $p(y^n)$, $p(\hat{y}^n)$, $p(y^n, \hat{y}^n)$ and $p(y^n|\hat{y}^n)$.

The *average distortion* is defined as

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

(11)

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 appropriate Shannon uncertainties:

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

(12)

If there is no uncertainty in Y given \hat{Y} , then no information is lost. In general, this will not be true.

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

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

(13)

where 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.

The Rate Distortion Theorem states that $R(D)$, as we have defined it, is the maximum achievable rate of information transmission which does not exceed distortion D . Note that the result is *independent of the exact form of the distortion measure* $d(y, \hat{y})$.

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 , 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 we have for those pairs the condition

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

(14)

Thus, roughly speaking, $I(Y, \hat{Y})$ embodies the splitting criterion between high and low probability pairs of paths. These pairs are, again, the input ‘training’ paths and corresponding output path.

Note that, in the absence of a distortion measure, this result remains true for two interacting information sources, the principal content of the *joint asymptotic equipartition theorem*, (Cover and Thomas, 1991, Theorem 8.6.1).

Thus the imposition of a distortion measure results in a limitation in the number of possible jointly typical sequences to those satisfying the distortion criterion.

For the theory we will explore later – of pairwise interacting information sources – $I(Y, \hat{Y})$ (or $I(Y_1, Y_2)$ without the distortion restriction), can play the role of H in the critical development of the next section.

The RDT is a generalization of the Shannon-McMillan Theorem which examines the interaction of two information sources under the constraint of a fixed average distortion. For our development we will require one more iteration, studying the interaction of three ‘languages’ under particular conditions, and require a similar generalization of the SMT in terms of the splitting criterion for triplets as opposed to single or double stranded patterns. The tool for this is at the core of what is termed *network information theory* (Cover and Thomas, 1991, Theorem 14.2.3). Suppose we have (piecewise memoryless) ergodic information sources Y_1, Y_2 and Y_3 . We assume Y_3 constitutes a critical embedding context for Y_1 and Y_2 so that, given three sequences of length n , the probability of a particular triplet of sequences is determined by *conditional probabilities with respect to Y_3* :

$$P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) =$$

$$\prod_{i=1}^n p(y_{1i}|y_{3i})p(y_{2i}|y_{3i})p(y_{3i}).$$

(15)

That is, Y_1 and Y_2 are, in some measure, driven by their interaction with Y_3

Then, in analogy with the previous two cases, triplets of sequences can be divided by a splitting criterion into two sets, having high and low probabilities respectively. For large n the number of triplet sequences in the high probability set will be determined by the relation (Cover and Thomas, 1991, p. 387)

$$N(n) \propto \exp[nI(Y_1; Y_2|Y_3)],$$

(16)

where splitting criterion is given by

$$I(Y_1; Y_2|Y_3) \equiv$$

$$H(Y_3) + H(Y_1|Y_3) + H(Y_2|Y_3) - H(Y_1, Y_2, Y_3)$$

Below we examine phase transitions in the splitting criteria H , which we will generalize to both $I(Y_1, Y_2)$ and $I(Y_1, Y_2|Y_3)$. The former will produce punctuated cognitive and non-cognitive learning plateaus, while the latter characterizes the interaction between selection pressure and sociocultural cognition.

Phase transition and coevolutionary condensation

The essential homology relating information theory to statistical mechanics and nonlinear dynamics is twofold (R Wallace and RG Wallace, 1998, 1999; Rojdestvenski and Cottam, 2000):

(1) A ‘linguistic’ equipartition of probable paths consistent with the Shannon-McMillan and Rate Distortion Theorems serves as the formal connection with nonlinear mechanics and fluctuation theory – a matter we will not fully explore here, and

(2) A correspondence between information source uncertainty and statistical mechanical free energy density, rather than entropy. See R Wallace and RG Wallace (1998, 1999) for a fuller discussion of the formal justification for this assumption, described by Bennett (1988) as follows:

“...[T]he value of a message is the amount of mathematical or other work plausibly done by the originator, which the receiver is saved from having to repeat.”

This is a central insight: In sum, we will generally impose invariance under renormalization symmetry on the ‘splitting criterion’ between high and low probability states from the Large Deviations Program of applied probability (e.g. Dembo and Zeitouni, 1998). Free energy density (which can be reexpressed as an ‘entropy’ in microscopic systems) is the splitting criterion for statistical mechanics, and information source uncertainty is the criterion for ‘language’ systems. Imposition of renormalization on free energy density gives phase transition in a physical system. For information systems it gives interactive condensation.

This analogy is indeed a mathematical homology:

The definition of the free energy density for a parametrized physical system is

$$F(K_1, \dots, K_m) = \lim_{V \rightarrow \infty} \frac{\log[Z(K_1, \dots, K_m)]}{V} \quad (17)$$

where the K_j are parameters, V is the system volume and Z is the ‘partition function’ defined from the energy function, the Hamiltonian, of the system.

For an ergodic information source the equivalent relation associates source uncertainty with the number of ‘meaningful’ sequences $N(n)$ of length n , in the limit

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

We will *parametrize* the information source to obtain the crucial expression on which our version of information dynamics will be constructed;

$$H[K_1, \dots, K_m, \mathbf{X}] = \lim_{n \rightarrow \infty} \frac{\log[N(K_1, \dots, K_m)]}{n}. \quad (18)$$

The essential point is that while information systems do not have ‘Hamiltonians’ allowing definition of a ‘partition function’ and a free energy density, they may have a source uncertainty obeying a limiting relation like that of free energy density. Importing ‘renormalization’ symmetry gives phase transitions at critical points (or surfaces), and importing a Legendre transform in a ‘natural’ manner gives dynamic behavior far from criticality. The first of these will be addressed here. The second is the subject of the ecological resilience analysis of the latter sections.

As neural networks demonstrate so well, it is possible to build larger pattern recognition systems from assemblages of smaller ones. We abstract this process in terms of a generalized linked array of subcomponents which ‘talk’ to each other in two different ways. These we take to be ‘strong’ and ‘weak’ ties between subassemblies. ‘Strong’ ties are, following arguments from sociology (Granovetter, 1973), those which permit disjoint partition of the system into equivalence classes. Thus the strong ties are associated with some reflexive, symmetric, and transitive relation between components. ‘Weak’ ties do not permit such disjoint partition. In a physical system these might be viewed, respectively, as ‘local’ and ‘mean field’ coupling.

We fix the magnitude of strong ties, but vary the index of weak ties between components, which we call P , taking $K = 1/P$.

We assume the information source of interest depends on three parameters, two explicit and one implicit. The explicit are K as above and an ‘external field strength’ analog J , which gives a ‘direction’ to the system. We may, in the limit, set $J = 0$.

The implicit parameter, which we call r , is an inherent generalized ‘length’ on which the phenomenon, including J and K , are defined. That is, we can write J and K 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 degree of niche partitioning in ecosystems.

Rather than specify complicated patterns of individual dependence or interaction between system subcomponents, we instead work entirely within the domain of the uncertainty of the parametrized information source associated with the system as a whole, which we write as

$$H[K, J, \mathbf{X}]$$

Imposition of invariance of H under a renormalization transform in the implicit parameter r leads to expectation of both a critical point in K , which we call K_C , reflecting a phase transition to or from collective behavior across the entire structure, and of power laws for behavior near K_C . Addition of other parameters to the system, e.g. some parameter V , results in a ‘critical line’ or surface $K_C(V)$.

Let $\kappa = (K_C - K)/K_C$ and take χ as the ‘correlation length’ defining the average domain in r -space for which the dual information source is primarily dominated by ‘strong’ ties. We begin by averaging across r -space in terms of ‘clumps’ of length R , defining J_R, K_R as J, K for $R = 1$. Then, following Wilson’s (1971) physical analog, we choose the renormalization relations as

$$H[K_R, J_R, \mathbf{X}] = R^{\mathcal{D}} H[K, J, \mathbf{X}]$$

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

(19)

where \mathcal{D} is a non-negative real constant, possibly reflecting fractal network structure. The first of these equations states that ‘processing capacity,’ as indexed by the source uncertainty of the system which represents the ‘richness’ of the inherent language, grows as $R^{\mathcal{D}}$, while the second just states that the correlation length simply scales as R .

Other, very subtle, symmetry relations – not necessarily based on elementary physical analogs – may well be possible. For example McCauley,

(1993, p.168) describes the counterintuitive renormalization relations needed to understand phase transition in simple ‘chaotic’ systems.

For K near K_C , if $J \rightarrow 0$, a simple series expansion and some clever algebra (e.g. Wilson, 1971; R Wallace and RG Wallace, 1998) gives

$$H = H_0 \kappa^{s\mathcal{D}}$$

$$\chi = \chi_0 \kappa^{-s}$$

(20)

where s is a positive constant. The series expansion argument which generates our equations (20) from equations (19) is precisely the same as that in K Wilson’s famous 1971 paper which produces his equations (22) and (23) from his equations (4) and (5), setting ‘external field strength’ $h \rightarrow 0$.

To reiterate, imposing renormalization symmetries different from equations (19) would give significantly different results.

Some rearrangement of equations (20) produces, near K_C ,

$$H \propto \frac{1}{\chi^{\mathcal{D}}}$$

(21)

This suggests that the ‘richness’ of the language is inversely related to the domain dominated by disjointly partitioning strong ties near criticality. As the nondisjunctive weak ties coupling declines, the efficiency of the coupled system as an information channel declines precipitously near the transition point: see ACTK for discussion of the relation between channel capacity and information source uncertainty.

Far from the critical point matters are considerably more complicated, apparently driven by appropriate generalizations of a physical system's 'Onsager relations', a matter treated in some detail below.

The essential insight is that *regardless of the particular renormalization symmetries involved, sudden critical point transition is possible in the opposite direction for this model*, that is, from a number of independent, isolated and fragmented systems operating individually and more or less at random, into a single large, interlocked, coherent system, 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 weak ties between them can bind several different 'language' processes into a single, embedding hierarchical metalanguage which contains those languages as linked subdialects.

This heuristic insight can be made exact using a rate distortion argument:

Suppose that two 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} .

Take some distortion measure d comparing paths y to paths \hat{y} , defining $d(y, \hat{y})$. We invoke the Rate Distortion Theorem's mutual information $I(Y, \hat{Y})$, which is a 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. 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.

We will make much of this below; cultural and genetic heritages are generalized languages, as are neural, immune, and sociocultural pattern recognition.

Pattern recognition as language

The task of this section is to express Atlan and Cohen's (1998) cognitive pattern recognition-and-response in terms of an ergodic information source

constrained by the AEPT. This general approach would then apply to the immune system, the CNS and sociocultural networks. Pattern recognition, as we will characterize it here, proceeds by convoluting an incoming ‘sensory’ signal with an internal ‘ongoing activity’ and, at some point, triggering an appropriate action based on a decision that the pattern of the sensory input requires a response. For the purposes of this work we do not need to model in any particular detail the manner in which the pattern recognition system is ‘trained,’ and thus adopt a ‘weak’ model which may have considerable generality, regardless of the system’s particular learning paradigm, which can be more formally described using the RDT.

We will, fulfilling Atlan and Cohen’s (1998) criterion of meaning-from-response, define a language’s contextual meaning entirely in terms of system output.

The model is as follows: A pattern of sensory input is convoluted with a pattern of internal ‘ongoing activity’ to create a path

$$x = (a_0, a_1, \dots, a_n, \dots).$$

This is fed into a (highly nonlinear) ‘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

$$B_0 = b_0, \dots, b_k$$

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

Thus we permit a graded response, supposing that if

$$h(x) \in B_0$$

the pattern is not recognized, and

$$h(x) \in B_1$$

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

We are interested in paths which trigger pattern recognition exactly once. 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 exactly once to the event $h(x) \in B_1$. Thus $h(a_0, a_1, \dots, a_j) \in B_0$ for all $j < m$ but $h(a_0, \dots, a_m) \in B_1$.

For each positive integer n , let $N(n)$ be the number of 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 is comparatively rare – and in particular assume that the finite limit

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

exists and is independent of the path x . We will – not surprisingly – call such a pattern recognition process ergodic.

We may thus define a 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 satisfy the relations

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

We say this ergodic information source is *dual* to the pattern recognition process.

Different ‘languages’ will, of course, be defined by different divisions of the total universe of possible responses into different pairs of sets B_0 and B_1 , or perhaps even by requiring more than one response in B_1 along a path. Like the use of different distortion measures in the RDT, however, it seems obvious that the underlying dynamics will all be qualitatively similar.

Meaningful paths – creating an inherent grammar and syntax – are defined entirely in terms of system response, as Atlan (1983, 1987, 1998) and Atlan and Cohen (1998) propose, quoting Atlan (1987)

“...[T]he perception of a pattern does not result from a two-step process with first perception of a pattern of signals and then processing by application of a rule of representation. Rather, a given pattern in the environment is perceived at the time when signals are received by a kind of resonance between a given structure of the environment – not necessarily obvious to the eyes of an observer – and an internal structure of the cognitive system. It is the latter which defines a possible functional meaning – for the system itself – of the environmental structure.”

Elsewhere (R Wallace, 2000a, b) we have termed this process an ‘information resonance.’

Although we do not pursue the matter here, the ‘space’ of the a_j can be partitioned into disjoint equivalence classes according to whether states can be connected by meaningful paths. This is analogous to a partition into domains of attraction for a nonlinear or chaotic system, and imposes a ‘natural’ algebraic structure which can, among other things, enable multitasking (R Wallace, 2000b).

We can apply this formalism to the stochastic neuron: A series of inputs $y_i^j, i = 1...m$ from m nearby neurons at time j is convoluted with ‘weights’ $w_i^j, i = 1...m$, using an inner product

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

(22)

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 $1 - f(\mathbf{y}^j \cdot \mathbf{w}^j)$.

In the terminology of this section the m values y_i^j constitute ‘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 $x = a_0, a_1, \dots, a_n, \dots$

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

Park et al. (2000) treat the stochastic neural network in terms of a space of related probability density functions $[p(\mathbf{x}, \mathbf{y}; \mathbf{w}) | \mathbf{w} \in \mathcal{R}^m]$, where \mathbf{x} is the input, \mathbf{y} the output and \mathbf{w} the parameter vector. The goal of learning is to find an optimum \mathbf{w}^* which maximizes the log likelihood function. They define a loss function of learning as

$$L(\mathbf{x}, \mathbf{y}; \mathbf{w}) \equiv -\log p(\mathbf{x}, \mathbf{y}; \mathbf{w}),$$

and one can take as a learning paradigm the gradient relation

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta_t \partial L(\mathbf{x}, \mathbf{y}; \mathbf{w}) / \partial \mathbf{w},$$

where η_t is a learning rate.

Park et al. (2000) attack this optimization problem by recognizing that the space of $p(\mathbf{x}, \mathbf{y}; \mathbf{w})$ is Riemannian with a metric given by the Fisher information matrix

$$G(\mathbf{w}) = \int \int \partial \log p / \partial \mathbf{w} [\partial \log p / \partial \mathbf{w}]^T p(\mathbf{x}, \mathbf{y}; \mathbf{w}) d\mathbf{y} d\mathbf{x}$$

where T is the transpose operation. A Fisher-efficient on-line estimator is then obtained by using the ‘natural’ gradient algorithm

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta_t G^{-1} \partial L(\mathbf{x}, \mathbf{y}; \mathbf{w}) / \partial \mathbf{w}.$$

Again, through the synergistic family of probability distributions $p(\mathbf{x}, \mathbf{y}; \mathbf{w})$, this can be viewed as a special case – a ‘representation’, to use physics jargon – of the general ‘convolution argument’ given above.

Again, it seems that a rate distortion argument between training language and network response language will nonetheless produce learning plateaus, even in this rather elegant special case.

The foundation of our mathematical modeling exercise is to assume that both the immune system and a sociocultural network's pattern recognition behavior, like that of other pattern recognition systems, can also be represented by the language arguments given above, and is thus dual to an ergodic information source, a context-defining language in Atlan and Cohen's (1998) sense, having a grammar and syntax such that meaning is explicitly defined in terms of system response.

Sociogeographic or sociocultural networks – social networks embedded in place and embodying culture – serve a number of functions, including acting as the local tools for teaching cultural norms and processes to individuals. Thus, for the purposes of this work, a person's social network – family and friends, workgroup, church, etc. – becomes the immediate agency of cultural dynamics, and provides the foundation for the brain/culture 'condensation' of Nisbett et al. (2001).

Sociocultural networks serve also, however, as instruments for collective decision-making, a cognitive phenomenon. Such networks serve as hosts to a political, in the large sense, process by which a community recognizes and responds to patterns of threat and opportunity. To treat pattern recognition on sociocultural networks we impose a version of the structure and general formalism relating pattern recognition to a dual information source:

We envision problem recognition by a local sociocultural network as follows: A 'real problem,' in some sense, becomes convoluted with a community's internal sociocultural 'ongoing activity' to create the path of a 'perceived problem' at times $0, 1, \dots$, producing a path of the usual form $x = a_0, a_1, \dots, a_n, \dots$. That serially correlated path is then subject to a decision process across the sociocultural network, designated $h(x)$ which produces output in two sets B_0 and B_1 , as before. The problem is officially recognized and resources committed to if and only if $h(x) \in B_1$, a rare event made even more rare if resources must then be diverted from previously recognized problems.

For the purposes of this work, then, we will view 'culture' as, in fact, a sociocultural cognitive process which can entrain individual cognition, a matter on which there is considerable research (e.g. Richerson and Boyd, 1995, 1998)

Our next task is to apply phase transition dynamics to ergodic information sources dual to a pattern recognition language, using techniques of the sections above. Similar considerations will apply to 'non-cognitive' interaction between structured selection pressures and the affected system.

To reiterate, we have, following the discussion in part 1 of Atlan and Cohen's (1998) work, implicitly assumed that the immune cognition can likewise be expressed as a pattern recognition-and-response language characterized by an information source uncertainty. The essential point – see figure 1 of Atlan and Cohen (1998) – is that antigens do not themselves directly trigger inflammatory response, but rather through the convolutional medium of antigen-presenting cells which communicate an immune sequence to a T-cell. The sequence/sentence describes the nature of the antigen. The T-Cell does not recognize antigens in their native form directly, rather, a processed peptide in the MHC cleft is the subject of the immune sequence recognized by the somatically generated TCR. How the T-cell responds to the peptide-MHC subject defines the immune ‘meaning’ of the antigen.

This is quite precisely, then, an example of the pattern-recognition-as-language mechanism we have just analyzed.

Generalized cognitive condensations

We suppose a cognitive system – more generally a linked, hierarchically structured, and broadly coevolutionary condensation of several such systems – is exposed to a structured pattern of sensory activity – the training pattern – to which it must learn an appropriate matching response. From that response we can infer, in a direct manner, something of the form of the excitatory sensory activity. We suppose the training pattern to have sufficient grammar and syntax so as to itself constitute a ergodic information source Y . The output of the cognitive system, B , is deterministically backtranslated into the ‘language’ of Y , and we call that translation \hat{Y} . The rate distortion behavior relating Y and \hat{Y} , is, according to the RDT, determined by the mutual information $I(Y, \hat{Y})$. We take the index of coupling between the sensory input and the cognitive system to be the number of training cycles – an exposure measure – having an inverse K , and write

$$I(Y, \hat{Y}) = I[K]$$

(23)

$I[K]$ defines the splitting criterion between high and low probability pairs of training and response paths for a specified average distortion D , and is analogous to the parametrized information source uncertainty upon which we imposed renormalization symmetry to obtain phase transition.

We thus interpret the sudden changes in the measured average distortion $D \equiv \sum p(y)d(y, \hat{y})$ which determines ‘mean square error’ between training pattern and output pattern, e.g. the *ending* of a learning plateau, as representing onset of a phase transition in $I[K]$ at some critical K_C , consonant with our earlier developments.

Note that $I[K]$ constitutes an interaction between the cognitive system and the impinging sensory activity, so that its properties may be quite different from those of the cognitive condensation itself.

From this viewpoint learning plateaus are an inherently ‘natural’ phase transition behavior of pattern recognition systems. While one may perhaps, in the sense of Park et al, (2000), find more efficient gradient learning algorithms, our development suggests learning plateaus will be both ubiquitous and highly characteristic of a cognitive system. Indeed, it seems likely that proper analysis of learning plateaus will give deep insight into the structures underlying that system.

This is not a new thought: Mathematical learning models of varying complexity have been under constant development since the late 1940’s (e.g. Luce, 1997), and learning plateau behavior has always been a focus of such studies.

The particular contribution of our perspective to this debate is that the distinct coevolutionary condensation of immune, CNS, and local socio-cultural network cognition which distinguishes human biology must respond as a composite in a coherent, unitary and coupled manner to sensory input. Thus the ‘learning curves’ of the immune system, the CNS and the embedding sociocultural network are inevitably coupled and must reflect each other. Such reflection or interaction will, of necessity, be complicated.

Our analysis, however, has a particular implication. Learned cultural behavior – sociocultural cognition – is, from our viewpoint, a nested hierarchy of phase transition learning plateaus which carries within it the history of an individual’s embedding socioculture. Through the cognitive condensation which distinguishes human biology, that punctuated history becomes part of individual cognitive and immune function. Simply removing ‘constraints’ which have deformed individual and collective past is unlikely to have the desired impact: one never, really, forgets how to ride a bicycle, and a social

group, in the absence of affirmative redress, will not ‘forget’ the punctuated adaptations ‘learned’ from experiences of slavery or holocaust. Indeed, at the individual level, sufficiently traumatic events may become encoded within the CNS and immune systems to express themselves as Post Traumatic Stress Disorder.

Noncognitive condensation in response to selection pressure

As discussed above, sociocultural networks serve multiple functions and are not only decision making cognitive structures, but are cultural repositories which embody the history of a community. Sociocultural networks, like human biology in the large, and the immune system in the small, have a duality in that they make decisions based on recognizing patterns of opportunity and threat by comparison with an internalized picture of the world, and they respond to selection pressure in the sense that cultural patterns which cannot adapt to external selection pressures simply do not survive. This is not learning in the traditional sense of neural networks. Thus the immune system has both ‘innate’ genetically programmed and ‘learned’ components, and human biology in the large is a convolution of genetic and cultural systems of information transmission.

We suggest that sociocultural networks – the instrumentalities of culture – likewise contain both cognitive and selective systems of information transmission which are closely intertwined to create a composite whole.

We now examine processes of ‘punctuated evolution’ inherent to evolutionary systems of information transmission.

We suppose a self-reproducing cultural system – more specifically a linked, and in the large sense coevolutionary, condensation of several such systems – is exposed to a structured pattern of selective environmental pressures to which it must adapt if it is to survive. From that adaptive selection – changes in genotype and phenotype analogs – we can infer, in a direct manner, something, but not everything, of the form of the structured system of selection pressures. That is, the culture contains markers of past ‘selection events’.

We suppose the system of selection pressures to have sufficient internal structure – grammar and syntax – so as to itself constitute an ergodic information source Y whose probabilities are fixed on the timescale of analysis. The output of that system, B , is backtranslated into the ‘language’ of Y , and we call that translation \hat{Y} . The rate distortion behavior relating Y and \hat{Y} , is, according to the RDT, determined by the mutual information $I(Y, \hat{Y})$.

We take there to be a measure of the ‘strength’ of the selection pressure, P , which we use as an index of coupling with the culture of interest, having an inverse $K = 1/P$, and write

$$I(Y, \hat{Y}) = I[K]. \quad (24)$$

P might be measured by the rate of attack by predatory colonizers, or the response to extreme environmental perturbation, and so on.

$I[K]$ thus defines the splitting criterion between high and low probability pairs of input and output paths for a specified average distortion D , and is analogous to the parametrized information source uncertainty upon which we imposed renormalization symmetry to obtain phase transition. The result is robust in the absence of a distortion measure through the joint asymptotic equipartition theorem, as discussed above.

We thus interpret the sudden changes in the measured average distortion $D \equiv \sum p(y)d(y, \hat{y})$ which determines ‘mean error’ between pressure and response, i.e. the *ending* of a ‘learning plateau’, as representing onset of a phase transition in $I[K]$ at some critical K_C , consonant with our earlier developments. In the absence of a distortion measure, we may still expect phase transition in $I[K]$, according to the joint AEPT.

Note that $I[K]$ constitutes an interaction between the self-reproducing system of interest and the impinging ecosystem’s selection pressure, so that its properties may be quite different from those of the individual or conjoined subcomponents (R Wallace and RG Wallace, 1998, 1999).

From this viewpoint highly punctuated ‘non-cognitive condensations’ are an inherently ‘natural’ phase transition behavior of evolutionary systems, even in the absence of a distortion measure. Again, while there may exist, in the sense of Park et al. (2000), more efficient convergence algorithms, our development suggests plateaus will be both ubiquitous and highly characteristic of evolutionary process and path. Indeed, it seems likely that proper analysis of non-cognitive evolutionary ‘learning’ plateaus – to the extent they can be observed or reconstructed – will give deep insight into the mechanisms underlying that system.

Convolution between selection pressure and sociocultural cognition

Selection pressure acting on sociocultural networks can be expected to affect their cognitive function, their ability to recognize and respond to relatively immediate patterns of threat and opportunity. In fact, those patterns themselves may in no small part represent factors of that selection pressure, conditionally dependent on it. We assume, then, the linkage of *three* information sources, two of which are conditionally dependent on and may indeed be dominated by, a highly structured embedding system of externally imposed selection pressure which we call Y_3 . Y_2 we will characterize as the pattern recognition-and-response language of the sociocultural network itself.

In IR Cohen’s (2000) sense, this involves comparison of sensory information with an internalized picture of the world, and choice of a response from a repertory of possibilities. Y_1 we take to be a more rapidly changing, but nonetheless structured, pattern of immediate threat-and-opportunity which demands appropriate response and resource allocation – the ‘training pattern’. We reiterate that Y_1 is likely to be conditionally dependent on the embedding selection pressure, Y_3 , as is the hierarchically layered history expressed by Y_2 .

According to the triplet version of the SMT which we discussed at the end of the theoretical section above, then, for large n , triplets of paths in Y_1, Y_2 and Y_3 may be divided into two sets, a smaller ‘meaningful’ one of high probability – representing those paths consistent with the ‘grammar’ and ‘syntax’ of the interaction between the selection pressure, the cognitive sociocultural process, and the pattern of immediate ‘sensory challenge’ it faces – and a very large set of vanishingly small probability. The splitting criterion is the conditional mutual information:

$$I(Y_1; Y_2 | Y_3).$$

We parametrize this splitting criterion by a variate K representing the inverse of the strength of the coupling between the system of selection pressure and the linked complex of the sociocultural cognitive process and the structured system of day-to-day problems it must address. $I[K]$ will, according to the ‘phase transition’ developments above, be highly punctuated by ‘mixed’ plateau behavior representing the synergistic and inextricably intertwined action of both externally imposed selection pressure and internal sociocultural cognition.

Modeling a condensed cognitive system far from phase transition

We suppose a linked and broadly coevolutionary condensation of several cognitive systems is exposed to a sudden perturbation, specifically, an episode of pathogenic challenge, and wish to estimate the response of that system, particularly within the context of an embedding ‘selection pressure’. Exploring this question requires some development. We will be interested, somewhat loosely, in the ‘thermodynamics’ and ‘generalized Onsager relations’ of the splitting criteria associated with the AEPT, JAEPT and RDT, and the network information theory variant. These are, respectively, the source uncertainty, mutual information, and conditional mutual information. Although we will couch the development in terms of the source uncertainty, the other splitting criteria are presumed to be treated similarly.

Suppose the source uncertainty of a coevolutionarily condensed behavioral language, $H[\mathbf{X}]$ is a function of system-wide average variables, K, J, M which represent the *ensemble* indices – associated with the entire individual-and-group.

Thus we may write

$$H = H[K, J, M, \mathbf{X}].$$

(25)

We assume that as K, J, M increase, that H decreases monotonically. We reiterate the essential trick that, in its definition,

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

(26)

is the analog of the *free energy density* of statistical mechanics: Take a physical system of volume V which can be characterized by an inverse temperature parameter $K = 1/T$. The ‘partition function’ for the system is (K. Wilson, 1971)

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

(27)

where E_j represents the energy of the individual state j . The probability of the state j is then

$$P_j = \frac{\exp[-K E_j]}{Z(K)}.$$

Then the ‘free energy density’ of the *entire* system is defined as

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

(28)

As described, the relation between information and thermodynamic free energy is long studied.

Equation (26) expresses one of the central theorems of information theory in a form similar to equation (28). It is this similarity which suggests that, for some systems under proper circumstances, there may be a ‘duality’ which maps Shannon’s ergodic source uncertainty onto thermodynamic free energy density, and perhaps vice versa.

The method we propose here, based entirely on equation (26), the Shannon-McMillan Theorem for ergodic information sources, and similar extensions

such as the Joint Asymptotic Equipartition, Rate Distortion Theorem, and network information theory, may prove to be more generally applicable to information systems, not requiring the explicit identification a ‘duality’ in each and every case.

The formal analogy – the homological duality – between free energy density for a physical system and ergodic source uncertainty, (or the mutual information splitting criteria associated with the JAEPT, RDT, and network information theory), is based on equations (26) and (28). This suggests we impose a *thermodynamics* on source uncertainty or the appropriate mutual information measures which provide splitting criteria between low and high probability sets of paths.

By a thermodynamics we mean, in the sense of Feigenbaum (1988, p. 530), the deduction of variables and their relations to one another that in some well-defined sense are averages of exponential quantities defined microscopically on a set.

In our context the relation between the number of meaningful sequences on length n , $N(n)$, for (fixed) large n and the source uncertainty, i.e.

$$N(n) \approx \exp(nH[\mathbf{X}])$$

for large n provides the exponential dependence exactly analogous to performing statistical mechanics. We are, to reiterate, going to express $H[\mathbf{X}]$ in terms of a number of parameters which characterize the underlying community which carries the behavioral language.

We suppose that H , or the various I , representing the splitting criteria defined by our coevolutionary condensation of CNS, immune and sociocultural cognition, is allowed to depend on a number of observable parameters, which we will not fully specify here.

If source uncertainty H is the analog to free energy density in a physical system, K is the analog to inverse temperature, the next ‘natural’ step is to apply a *Legendre transformation* to H so as to define a generalized ‘entropy,’ S , and other (very) rough analogs to classical thermodynamic entities, depending on the parameters.

Courant and Hilbert (1989, p.32) characterize the Legendre transformation as defining a surface as the envelope of its tangent planes, rather than as the set of points satisfying a particular equation.

Their development shows the Legendre transformation of a well-behaved function $f(Z_1, Z_2, \dots Z_w)$, denoted g , is

$$g = f - \sum_{i=1}^w Z_i \partial f / \partial Z_i \equiv f - \sum_{i=1}^w Z_i V_i.$$

(29)

with, clearly, $V_i \equiv \partial f / \partial Z_i$.

This expression is assumed to be invertible, hence the ‘duality:’

$$f = g - \sum_{i=1}^w V_i \partial g / \partial V_i.$$

Transformation from the ‘Lagrangian’ to the ‘Hamiltonian’ formulation of classical mechanics (Landau and Lifshitz, 1959) is via a Legendre transformation.

The generalization when f is not well-behaved is via a variational principle (Beck and Schlogl, 1993) rather than a tangent plane argument. Then

$$g(V) = \min_Z [f(Z) - V Z]$$

$$f(Z) = \min_V [g(V) - V Z].$$

(30)

In the first expression the variation is taken with respect to Z , in the second with respect to V .

For a badly behaved function it is usually possible to fix up a reasonable invertible structure since the singularities of f or g will usually belong to ‘a set of measure zero,’ for example a finite number of points on a line or lines on a surface where we may designate inverse values by fiat.

We first consider a very simple system in which the ergodic source uncertainty H depends *only* on the inverse strength of weak ties K , giving an analog to the ‘canonical ensemble’ of statistical mechanics which depends only on temperature. We define S , an entropy-analog which we term the ‘disorder,’ as the Legendre transform of the Shannon uncertainty $H[K, \mathbf{X}]$:

$$S = H - KdH/dK \equiv H - KU \quad (31)$$

where we take $U = dH/dK$ as an analog to the ‘internal energy’ of a system. Note that S and H have the same physical dimensionality.

Since

$$dS/dK = dH/dK - U - KdU/dK = -KdU/dK$$

we have

$$dS/dU = -K$$

and

$$dU \propto PdS.$$

This is the analog to the classic thermodynamic relation $dQ = TdS$ for physical systems. Thus what we have defined here as the disorder S is indeed a generalized entropy.

Note that since $dS/dU = -K$ we have

$$S = H - KU = H + UdS/dU$$

or

$$H = S - UdS/dU$$

which explicitly shows the dual relation between H and S .

Again let $N(n)$ represent the number of meaningful sequences of length n emitted by the source \mathbf{X} . Since

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

for large n , we have

$$U = dH/dK = \lim_{n \rightarrow \infty} \frac{1}{nN} dN/dK.$$

(32)

For fixed (large) n , U is thus the *proportionate* rate of change in number of meaningful sequences of length n with change in K . This is something like the rate of change of mass per unit mass for a person losing weight: A small value will not be much noticed, while a large one may represent a rigorous starvation causing considerable distress.

Some rearrangement gives

$$Q \equiv (S - H) = -K dH/dK = U dS/dU$$

(33)

We define $Q = S - H$ as the *instability* of the system.

If $-dH/dK$ is approximately constant – something like a heat capacity in a physical system – then we have the approximate linear relation

$$Q \approx b_K K$$

with

$$b_K \equiv -\partial H/\partial K.$$

We generalize this as follows:

Allow H to depend on a number of parameters, for example average probability of weak ties, the inverse level of community resources, and or other factors which we call $Z_i, i = 1....$ Then, taking $H = H[Z_i, \mathbf{X}]$, we obtain the equations of state

$$S = H - \sum_{j=1}^s Z_j \partial H / \partial Z_j = H - \sum_{j=1}^s Z_j V_j.$$

$$V_i \equiv \partial H / \partial Z_i$$

(34)

and the *instability relation*

$$Q = S - H = - \sum_{j=1}^s Z_j \partial H / \partial Z_j = \sum_{j=1}^s V_j \partial S / \partial V_j,$$

$$= - \sum_{j=1}^s V_j Z_j = -\mathbf{Z} \cdot \nabla_{|\mathbf{Z}} H,$$

(35)

taking $\mathbf{Z} = (Z_1, Z_2, \dots, Z_s)$.

Q represents the degree of disorder above and beyond that which is inherent to the ergodic information source itself.

Instability, as we have defined it, is driven by the declining capacity to convey messages across the system, since information theory tells us that information source uncertainty must be less than or equal to the capacity of the transmitting channel, C .

We suppose that the capacity, C , of the underlying communication channel declines with increasing K , so that $C = C(K)$ is monotonically decreasing in K . An ergodic information source can be transmitted without error by a channel only if $H[K, \mathbf{X}] < C(K)$ – again see ACTK – so that declining C will inevitably result in rising Q .

Q is, according to this development, driven by underlying parameters characterizing physiological and social systems – the Z_j and V_j .

For a social system, equation (35) is interpreted as stating that the rate of indices of distress is proportional a systemic experience of instability.

It may be possible to generalize the development to include temporal effects if we suppose that H depends on $d\mathbf{Z}/dt \equiv \dot{\mathbf{Z}}$ as well as on \mathbf{Z} . Note that terms of the form $\partial H/\partial t$ would violate ergodicity. Then we would take

$$Q = -(\mathbf{Z}, \dot{\mathbf{Z}}) \cdot \nabla|_{(\mathbf{Z}, \dot{\mathbf{Z}})} H \equiv$$

$$\sum_j -Z_j \partial H / \partial Z_j - \dot{Z}_j \partial H / \partial \dot{Z}_j.$$

(36)

This suggests that both parameter gradients and their rates of change can be globally destabilizing.

In linear approximation, assuming $-\partial H/\partial Z_i = -V_i = b_{Z_i} \approx \text{constant}$, equation (36) can be rewritten as

$$Q \approx b_K K + b_J J + b_M M.$$

The use of environmental index variates for critical system parameters will generally result in a nonzero intercept, giving the final equation

$$Q \approx b_K K + b_J J + b_M M + b_0.$$

(37)

Note that the intercept b_0 may, in fact, be quite complex, perhaps incorporating other parameters not explicitly included in the model. But it may include as well an ‘error term’ representing stochastic fluctuations not entirely damped by large population effects, or even some ‘nonlinear’ structure when the b_{Z_i} are not quite constant.

Most importantly for our analysis here, if the ‘potentials’ $V_i = \partial H / \partial Z_i$ cannot be approximated as constants, then simple linear regression will fail entirely, and equations (34) and (37) will represent an appropriate generic model – possibly with ‘error terms’ – however *the system will be both nonlinear and nonmonotonic*, hence representing signal transduction in physiological systems.

In sum, we claim the instability relation derived from a fairly simple quasi-thermodynamic argument applied to an ergodic information source parametrized by various significant indices, (as well as, perhaps, their rates of change), explains the high degree to which simple regression models based on those parameters account for observed patterns of physiological, psychological, psychosocial, or immune response to the perturbation of an infection.

Sudden pathogenic challenge: ecological resilience of cognitive condensations

In reality, matters are significantly more complex than we have described so far: physiological, psychological and (locally) sociocultural responses to infection may, in turn, affect the infection itself through feedback loops.

Thus the inherently nonlinear model for response as produced by increasing stimulation, $Q = -\sum_j V_j Z_j$ is replaced by an even more nonlinear structure:

$$Q(t) = -\sum_j V_j(Q(t)) Z_j(Q(t)).$$

(38)

In a first iteration using linear approximation, we can replace this equation with a series for which *each* of s variates – ‘independent’ as well as ‘dependent’ – is expressed in terms of s linear regressions on all the others:

$$x_i(t) = \sum_{j \neq i}^s b_{i,j} x_j(t) + b_{i,0} + \epsilon(t, x_1(t) \dots x_s(t)).$$

(39)

Here the $x_j, j = 1 \dots s$ are both ‘independent’ and ‘dependent’ variates involved in the feedback, $b_{i,0}$ is the intercept constant, and the ϵ terms are ‘error’ terms which may not be small, in this approximation.

In matrix notation this set of equations becomes

$$X(t) = \mathbf{B}X(t) + U(t)$$

(40)

where $X(t)$ is a s -dimensional vector, \mathbf{B} is an $s \times s$ matrix of regression coefficients having a zero diagonal and U is an s -dimensional vector containing the constant and ‘error’ terms which are not necessarily small.

We suggest that, on the timescale of the applied perturbation of infection, and of initial responses, the \mathbf{B} -matrix remains relatively constant.

Following the analysis of Ives (1995) this structure has a number of interesting properties which permit estimation of the effects of an infective perturbation on an individual embedded in larger social structures, particularly those defined by unequal power relationships.

We begin by rewriting the matrix equation as

$$[\mathbf{I} - \mathbf{B}]X(t) = U(t)$$

(41)

where \mathbf{I} is the $s \times s$ identity matrix and, to reiterate, \mathbf{B} has a zero diagonal.

We reexpress matter *in terms of the eigenstructure of \mathbf{B}* .

Let \mathbf{Q} be the matrix of eigenvectors which diagonalizes \mathbf{B} . Take $\mathbf{Q}Y(t) = X(t)$ and $\mathbf{Q}W(t) = U(t)$. Let \mathbf{J} be the diagonal matrix of eigenvalues of \mathbf{B} so that $\mathbf{B} = \mathbf{Q}\mathbf{J}\mathbf{Q}^{-1}$. In R Wallace, Y Huang, P Gould and D Wallace (1997) we show the eigenvalues of \mathbf{B} are all real. Then, for the eigenvectors Y_k of \mathbf{B} , corresponding to the eigenvalues λ_k ,

$$Y_k(t) = \mathbf{J}Y_k(t) + W_k(t).$$

(42)

Using a term-by-term shorthand for the components of Y_k , this becomes

$$y_k(t) = \lambda_k y_k(t) + w_k(t).$$

(43)

Define the mean $E[f]$ of a time-dependent function $f(t)$ over the time interval $0 \rightarrow \Delta T$ as

$$E[f] = \frac{1}{\Delta T} \int_0^{\Delta T} f(t) dt.$$

(44)

We assume an appropriately ‘rational’ structure as $\Delta T \rightarrow \infty$, probably some kind of ‘ergodic’ hypothesis.

Note that this form of expectation does not include the effects of differing timescales or lag times. Under such circumstances, increasing ΔT will begin to ‘pick up’ new effects, in a path-dependent manner: The mathematics of equation (44) suddenly becomes extremely complicated.

The variance $V[f]$ over the same interval is defined as $E[f - E[f]^2]$.

Again taking matters term-by-term, we take the variance of the $y_k(t)$ from the development above to obtain

$$V[(1 - \lambda_k)y_k(t)] = V[w_k(t)]$$

so that

$$V[y_k] = \frac{V[w_k]}{(1 - \lambda_k)^2}$$

or

$$\sigma(y_k) = \frac{\sigma(w_k)}{|1 - \lambda_k|}.$$

(45)

The y_k are the components of the eigentransformed pathology, income, crowding, community size etc. variates x_i and the w_k are the similarly transformed variates of the driving externalities u_i .

The eigenvector components y_i are characteristic but non-orthogonal combinations of the original variates x_i whose standard deviation is that of

the (transformed) externalities w_i , but *synergistically amplified* by the term $1/|1 - \lambda_i|$, a function of the eigenvalues of the matrix of regression coefficients \mathbf{B} . If $\lambda_i \rightarrow 1$ then any change in external driving factors – here an infection – will cause great instability within the affected individual.

A simple example suffices. Suppose we have the two empirical regression equations

$$x_1 = b_1 x_2 + b_{01}$$

and

$$x_2 = b_2 x_1 + b_{02}$$

where x_1 is, for example, an index of violent crime and x_2 is an index of the ‘strength of weak ties.’ These equations say that weak ties affect violence and violence affects weak ties. Then, after normalizing x_1 and x_2 to zero mean and unit variance, the \mathbf{B} -matrix becomes

$$\mathbf{B} = \begin{pmatrix} 0 & R \\ R & 0 \end{pmatrix}$$

where $R = b_1 = b_2$ is simply the correlation between x_1 and x_2 .

This matrix has eigenvalues $\pm|R|$ and eigenvectors $[\pm 1/\sqrt{2}, 1/\sqrt{2}]$. As the variates become more closely correlated, $R \rightarrow 1$ and the ratio of the standard deviation of the eigenvector with positive components and that of the external perturbations, $1/[1 - R]$, diverges.

There is a kind of physical picture for this model. Imagine a violin strung with limp, wet cotton twine. Then $R \approx 0$ and no amount of bowing – an external perturbation – will excite any sound from the instrument. Now restring that violin with finely tuned catgut (to be somewhat old fashioned). Then $R \approx 1$ and external perturbation – bowing – will now excite loud and brisant eigenresonances.

Ives (1995) defines an ecosystem for which $\lambda \approx 0$, so that $1/|1 - \lambda| \approx 1$, as *resilient* in the sense that applied perturbations will have no more effect than their own standard deviation.

Such treatments are now routine in population and community ecology, but are still rare in epidemiological or physiological studies. The essential point is that a nonlinear deterministic ‘backbone’ serves to amplify external perturbations – in our case an infection. As Higgins et al. (1997) put it,

“...[R]elatively small environmental perturbations can markedly alter the dynamics of deterministic biological mechanisms, producing very large fluctuations...”

This is a version of Holling’s (1973, 1992) mechanism for the loss of ecological resilience by which the small can influence the large. In our case the ‘fluctuation’ is the sudden sickening of an individual by a pathogen, within the context of social structure and process.

To be more explicit, we propose that pathogenic challenge or infection is a perturbative ‘event’ affecting an individual embedded in a cognitive socioculture which may itself be subject to highly structured external ‘selection pressures’. The resilience of the individual-in-context will determine the response to that perturbation – the eigenpattern of infection process and symptom. The response of the (former) masters will generally be significantly different from that of the (former) slaves. Characteristic forms of physiological, psychological, and psychosocial burden – structured ‘selection pressures’ – experienced by the powerful will generally be far less intrusive than those experienced by the powerless. We propose, then, that socially dominant individuals tend to be ecologically resilient to infection in Ives’ sense, so that eigenpatterns of symptoms excited by infection, and related processes, will be relatively muted.

The powerless will generally find both infection process – virulence – and symptom response to be greatly amplified.

Chronic pathogenic challenge: stages of disease as ‘punctuated equilibria’

Chronic infectious disease presents a conundrum for immunology. To paraphrase remarks by Grossman (2002), the pathogenesis of HIV disease is not well understood, partly because of both the limited accessibility of the sites in the patients body where essential processes take place, and partly due to limitations of the non-human primate model of the disease. More fundamentally, however, according to Grossman, the root problem is an incomplete understanding of both the rules of immunity and of the reaction of the immune system to chronic infection. Here we attempt to broadly address these matters.

The first paper in this series (R Wallace and RG Wallace, 2002) examined cultural variation in malaria pathology and in rates of heterosexual

HIV transmission. HIV is too simple to be cognitive, and responds to immunogenic challenge as an evolution machine which hides deep in tissues. *P. falciparum* engages in analogous rapid clonal antigenic variation, and cytoadherence and sequestration in the deep vasculature, primary mechanisms to escape from antibody-mediated mechanisms of the host's immune system (e.g. Allred, 1998).

Recently Adami et al. (2000) applied an information theory approach to conclude that genomic complexity resulting from evolutionary adaptation can be identified with the amount of information a gene sequence stores about its environment. Elsewhere (R Wallace, 2002) we have used a Rate Distortion argument in the context of imposed renormalization symmetry to obtain 'punctuated equilibrium' in evolution as a consequence of their mechanism. Here we apply the more general Joint Asymptotic Equipartition Theorem to conclude that analogous pathogenic adaptive response to immune challenge will generally be characterized by relatively rapid changes which can be interpreted as phase transitions, in a large sense, suggesting a 'punctuated stage' model for many chronic infections. The result is formally analogous to learning plateau behavior in neural networks (R Wallace, 2002).

The last section examined host response to sudden pathogenic challenge. Suppose the pathogen is not extirpated by that response, but, changing its coat or hiding within spatial refugia, becomes an established invading population, having a particular genotype and phenotype which may, in fact, change in time. The immune system is cognitive, the pathogen is not, for this analysis.

We suppose that the sequence of 'states' of the host, a path $x \equiv x_0, x_1, \dots$ and the sequence of 'states' of the pathogen population, a path $y \equiv y_0, y_1, \dots$, are very highly structured and serially correlated and can, in fact, be represented by 'information sources' \mathbf{X} and \mathbf{Y} . Since the host and pathogen population interact, these sequences of states are not independent, but are jointly serially correlated. We 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 is that the set of joint paths z can be partitioned into a relatively small set of relatively high probability which is termed *jointly typical*, and a much larger set of vanishingly small probability, in much the same manner as equation (3). Further, according to the JAEPT, the *splitting criterion* between high and low probability sets of pairs is the mutual information

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

which represents the degree of interaction between \mathbf{X} and \mathbf{Y} .

What we propose is that I can be further parametrized by some external measurable index, P , of coupling between the host and the pathogen population, or perhaps by some set of them. P can, for example, represent the selection pressure exerted by the host immune system on the pathogen population, or, vice versa, of the physiological impact of the pathogen. We let $K = 1/P$. The basic argument is that, following the results of the previous sections, the effects of changing K on the mutual information I can be very highly punctuated, most likely multiply so. Such ‘learning plateaus’, in this model, represent different stages of disease.

In essence, then, the pathogen population *cognitively* writes itself onto the host’s immune system, which in turn, *adaptively* writes itself onto the pathogen population’s genotype and phenotype, in an endless but highly punctuated cycle, an example of what Levins and Lewontin (Lewontin, 2000) have called ‘interpenetration’. The ‘punctuated equilibria’ of such interpenetration constitute, in this model, the different stages of the chronic infection.

This section, and the one above, represent two limiting cases of interaction between host and pathogen population, i.e. ‘instant’ and ‘adiabatic’. Clearly more work is needed on intermediate time scales, which will be typically complicated.

A criticism of mathematical models

Mathematical models of complex biological phenomena, like the one we have presented, do not have a good history. From ‘mathematical ecology’ to ‘mathematical epidemiology’, and more recently ‘theoretical immunology’, Erlenmeyer flask models based on systems of 19th Century differential or difference equations have been touted as ‘the Next Big Thing’ by many engaged in a kind of slash-and-burn academic plantation farming. While our adaptation of the Large Deviations Program is perhaps of more compelling mathematical interest than this work, the whole enterprise is fraught with a certain intellectual peril. Pielou (1977) has attempted to correct some of the excesses of ‘mathematical ecology’, and her comments are significant for all such studies:

“...[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 answering questions but in raising them*. Models can be used to inspire new field observations and these are the only source of new knowledge as opposed to new speculation.”

Here, then, we are hoping to encourage a new class of speculation regarding the effect of embedding cultural structures on the manifestations of acute and chronic infectious disease, particularly as they relate to the development of vaccine strategy. In spite of the considerable formal overhead, this must remain a relatively modest goal: Mathematical modeling is, at best, a distinctly subordinate voice in a dialog with data.

Discussion: history and the immunocultural condensation

Individual CNS and immune cognition are embedded in, and interact with, sociocultural processes of cognition. Sufficiently draconian external ‘social selection pressures’ – a euphemism for the burdens of history, patterns of Apartheid, or the systematic deprivation of a neoliberal ‘market economy’ – should become manifest at the behavioral and cellular levels of an individual, through the intermediate mechanism of the embedding socio-cultural network. This manifestation will often amplify the perturbative or chronic impact of exposure to pathogens, parasites, or chemical stressors.

This is not entirely a new observation. Franz Fanon (1966) has described an essentially similar phenomenon:

“[Under an Apartheid system the] world [is] divided into compartments, a motionless, Manicheistic world... The native is being hemmed in; apartheid is simply one form of the division into compartments of the colonial world... his dreams are of action

and aggression... The colonized man will first manifest this aggressiveness which has been deposited in his bones against his own people. This is the period when the niggers beat each other up, and the police and magistrates do not know which way to turn when faced with the astonishing waves of crime...

It would therefore seem that the colonial context is sufficiently original to give grounds for reinterpretation of the causes of criminality... The [colonized individual], exposed to temptations to commit murder every day – famine, eviction from his room because he has not paid the rent, the mother’s dried up breasts, children like skeletons, the building-yard which has closed down, the unemployed that hang about the foreman like crows – the native comes to see his neighbor as a relentless enemy. If he strikes his bare foot against a big stone in the middle of the path, it is a native who has placed it there... The [colonized individual’s] criminality, his impulsivity and the violence of his murders are therefore not of characterial originality, but the direct product of the colonial system.”

One of our contributions to this debate is to suggest that patterns of immune function should become entrained into this process as well through the immunocultural condensation, so that colonized man should have a vulnerability to sudden or chronic immune stressors – in particular microbiological or chemical challenges – which should extend beyond, but will likely be synergistic with, the effects of deprivation alone. Differences in immune function or response to infection heretofore attributed to genetic differences between populations will, in our view, closely reflect this mechanism. These matters have evident importance for development of vaccines against HIV, malaria, tuberculosis, other infectious diseases, and certain classes of malignancy, for example breast cancer.

A particular implication of our work is that the functioning of local socio-cultural networks can become closely convoluted with past or present external selection pressures – the burden of history, effects of continuing Apartheid, or the depredations of ‘the market’. The mathematical model suggests that increases of selection pressure itself, or of the coupling between selection pressure and sociocultural cognition, should manifest themselves through punctuated changes in the ability of sociocultural networks to meet the challenges of changing patterns of threat or opportunity, and, ultimately, in the ability

of the immunocultural condensation of individuals within those sociocultural networks to respond to patterns of microbiological, parasite, or mutagenic challenge.

In general, according to our model, the powerful will be more resilient than the powerless in the ecological sense of not resonantly amplifying the impacts of such challenge, although the underlying ‘eigenmodes’ of symptoms and physiological response may often be similar. Sufficiently different sociocultural systems, however, may, as Nisbett et al. (2001) found for CNS cognition, impose significantly different response eigenpatterns. At the other extreme, patterns of staging of chronic disease are likely to be similarly modulated.

These are all questions directly subject to empirical test.

Inherent to our approach is recognition of the burdens of history. By this we mean the way in which the grammar and syntax of a culturally-determined individual ‘behavioral language,’ and its embodiment at cellular levels through the interaction of the CNS and the immunocultural condensation, encapsulate earlier adaptations to external selection pressures, even in the sudden absence of those pressures. Selection pressures may involve deliberate policy, ranging, in the US, from an unrelieved history of slavery, to more contemporary depredations like the ‘urban renewal’ of the 1950’s, its evolutionary successor the ‘planned shrinkage’ of the 1970’s, or the subsequent explicit counterreformation against the successes of the Civil Rights Movement.

Our mathematical approach may seem excessive to some. We reply, as did the master mathematical ecologist EC Pielou above, that the principal value of mathematical models is in raising research questions for subsequent empirical test, rather than in answering them directly. Empirical study, to reiterate Pielou’s words, “Is the only true source of new knowledge, as opposed to new speculation.”

We have, then, created a mathematical model in the tradition of the Large Deviations Program of applied probability which makes explicit the entrainment of environment and development into the expression of pathogenic or chemical challenge. We have used that model to raise the speculation that naive and simplistic geneticism which, in effect, reifies ‘race’, and its parallel of naive cross-sectional environmentalism which neglects the path-dependent burdens of history, significantly hinder current research.

Matters of historical, political, economic, and social justice are largely and increasingly excised from academic discussions of health and illness in

the US, apparently for fear of angering funding agencies or those with power over academic advancement. Our modeling exercise suggests, among other things, the extraordinary degree to which that excision may limit the value of such work in the design of corrective policy, particularly the creation of effective vaccine strategies.

References

- Adame C, C Ofria and T Collier, 2000, "Evolution of biological complexity," *Proceedings of the National Academy of Sciences*, **97**, 4463-4468.
- Alred D, 1998, "Antigenic variation in *Babesia bovis*: how similar is it to that in *Plasmodium falciparum*?", *Annual Tropical Medicine and Parasitology*, **92**, 461-472.
- Ash R, 1990, *Information Theory*, Dover Publications, New York.
- Atlan H, 1983, "Information theory," pp. 9-41 in R Trappl (ed.), *Cybernetics: Theory and Application*, Springer-Verlag, New York.
- Atlan H, 1987, "Self creation of meaning," *Physica Scripta*, **36**, 563-576.
- Atlan H, 1998, "Intentional self-organization, emergence and reduction: towards a physical theory of intentionality," *Theses Eleven*, **52**, 6-34.
- Atlan H and IR Cohen, 1998, "Immune information, self-organization and meaning," *International Immunology*, **10**, 711-717.
- Bennett C, 1988, "Logical depth and physical complexity". In: Herkin R (Ed) *The Universal Turing Machine: A Half-Century Survey*, Oxford University Press, pp. 227-257.
- Cohen IR, 2001, "Immunity, set points, reactive systems, and allograft rejection." To appear.
- Cohen IR, 1992, "The cognitive principle challenges clonal selection," *Immunology Today*, **13**, 441-444.
- Cohen IR, 2000, *Tending Adam's Garden: evolving the cognitive immune self*, Academic Press, New York.
- Courant R and D Hilbert, 1989, *Methods of Mathematical Physics, Volume II*, John Wiley and Sons, New York.
- Cover T and J Thomas, 1991, *Elements of Information Theory*, Wiley, New York.
- Dembo A and O Zeitouni, 1998, *Large Deviations: Techniques and Applications, 2nd Ed.*, Springer-Verlag, New York.
- Fanon F, 1966, *The Wretched of the Earth*, Grove Press, New York.

- Feigenbaum M, 1988, "Presentation functions, fixed points and a theory of scaling function dynamics", *Journal of Statistical Physics*, **52**, 527-569.
- Granovetter M, 1973, "The strength of weak ties," *American Journal of Sociology*, **78**, 1360-1380.
- Grossman Z, 2002, personal communication.
- Holling C, 1973, "Resilience and stability of ecological systems", *Annual Reviews of Ecological Systematics*, **4**, 1-25.
- Holling C, 1992, "Cross-scale morphology, geometry, and dynamics of ecosystems", *Ecological Monographs*, **62**, 447-502.
- Ives A, 1995, "Measuring resilience in stochastic systems", *Ecological Monographs*, **65**, 217-233.
- Khinchine A, 1957, *The Mathematical Foundations of Information Theory*, Dover Publications, New York.
- Landau L and E Lifshitz, 1959, *Classical Mechanics*, Addison-Wesley, Reading MA.
- Lewontin R, 2000, *The Triple Helix: Gene, Organism, and Environment*, Harvard University Press, Cambridge, MA.
- Luce R, 1997, "Several unresolved conceptual problems of mathematical psychology," *Journal of Mathematical Psychology*, **41**, 79-87.
- McCauley L, 1993, *Chaos, Dynamics and Fractals: an algorithmic approach to deterministic chaos*, Cambridge University Press, Cambridge, UK.
- Nisbett R, K Peng, C Incheol and A Norenzayan, 2001, "Culture and systems of thought: holistic vs analytic cognition", *Psychological Review*, **108**, 291-310.
- Park H, S Amari and K Fukumizu, 2000, "Adaptive natural gradient learning algorithms for various stochastic models," *Neural Networks*, **13**, 755-764.
- Pielou E, 1977, *Mathematical Ecology*, John Wiley and Sons, New York.
- Richerson P and R Boyd, 1995, "The evolution of human hypersociality." Paper for Ringberg Castle Symposium on Ideology, Warfare and Indoctrinability (January, 1995), and for HBES meeting, 1995.
- Richerson P and R Boyd, 1998, "Complex societies: the evolutionary origins of a crude superorganism," to appear.
- Rojdestvenski I and M Cottam, 2000, "Mapping of statistical physics to information theory with applications to biological systems," *Journal of Theoretical Biology*, **202**, 43-54.
- Wallace R, Y Huang, P Gould and D Wallace, 1997, "The hierarchical diffusion of AIDS and violent crime among US metropolitan regions...", *Social*

Science and Medicine, **44**, 935-947.

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

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

Wallace R and RG Wallace, 2002, "Immune cognition and vaccine strategy: beyond geonomics." In press, *Microbes and Infection*

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

Wallace R, 2000b, "Information resonance and pattern recognition in classical and quantum systems: toward a 'language model' of hierarchical neural structure and process," available at

www.ma.utexas.edu/mp_arc-bin/mpa?yn=00-190.

Wallace R, 2002, "Adaptation, punctuation, and information: A rate-distortion approach to non-cognitive 'learning plateaus' in evolutionary process." Submitted. Available at

www.ma.utexas.edu/mp_arc-bin/mpa?yn=01-163.

Wilson K, 1971, "Renormalization group and critical phenomena. I Renormalization group and the Kadanoff scaling picture", *Physical Review B*, **3**, 3174-3183.