Selection pressure and organizational cognition: implications for the social determinants of health

Rodrick Wallace, PhD
The New York State Psychiatric Institute*

June, 2001

Abstract

We model the effects of Schumpeterian 'selection pressures' – in particular Apartheid and the neoliberal 'market economy' – on organizational cognition in minority communities, given the special role of culture in human biology. Our focus is on the dual-function social networks by which culture is imposed and maintained on individuals and by which immediate patterns of opportunity and threat are recognized and given response. A mathematical model based on recent advances in complexity theory displays a joint cross-scale linkage of social, individual central nervous system, and immune cognition with external selection pressure through mixed and synergistic punctuated 'learning plateaus.' This provides a natural mechanism for addressing the social determinants of health at the individual level. The implications of the model, particularly the predictions of synergistic punctuation, appear to be empirically testable.

KEY WORDS Evolutionary punctuation, health, inequality, information theory, phase transition, social networks.

*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. Affiliations are for identification only.

Introduction

Lack of detailed mechanism haunts current discussions concerning the social determinants of health. While Wilkinson (1996) and his colleagues present a compelling epidemiological picture regarding some effects of stress and inequality on individual health – a picture sufficient in many respects to begin the design of corrective public policy – their arguments would be significantly strengthened and their policy recommendations greatly sharpened by a more detailed picture of the mechanisms by which social structure and process affects individual health and illness.

Our earlier empirical work and in this direction (e.g. R Wallace and D Wallace, 1977, 1997; D Wallace and R Wallace, 1998, 2000) suggests that deliberate policy, path dependence and the enduring burdens of history, material deprivation, and the relentless spread of effects by contagious process upwards along the social heirarchy (i.e. the US system of Apartheid cannot contain its effects), play far more central roles than is comfortably recognized by current US academic research, whose political economy is not particularly consonant with the needs of public policy. To paraphrase the opinions of more than one high level US public health official, "Every decade or so the academics come down to Washington en masse and tell us that poverty is bad for health. We already know this".

Under the US system, at least, changing policy depends on making the case that 'bad things' are not confined to marginalized populations – the basis for necessary large-scale public relations efforts to create consensus across traditional political divisions – and providing a roadmap which can be used to create a synergistically effective 'more-bang-for-the-buck' intervention strategy. Although the public relations effort is beyond us here, we can attempt a more thorough treatment of the mechanisms underlying the social determinants of health. The tools for this involve recent advances in complexity theory, given the special role which culture plays in human biology. Here we significantly expand a recent theoretical study in that direction (R Wallace, RG Wallace, D Wallace and M Fullilove, 2001).

"Culture," the evolutionary anthropologist Robert Boyd has asserted, "is as much a part of human biology as the enamel on our teeth" (Boyd, 1995). Indeed, the current dual vision of human biology among evolutionary anthropologists is summarized by Durham (1991) as follows:

> "...[G]enes and culture constitute two distinct but interacting systems of inheritance within human populations... [and] infor-

mation of both kinds has influence, actual or potential, over ... behaviors [which] creates a real and unambiguous symmetry between genes and phenotypes on the one hand, and culture and phenotypes on the other...

[G]enes and culture are best represented as two parallel lines or 'tracks' of hereditary influence on phenotypes..."

Elsewhere (Wallace et al., 2001) we have adapted this perspective to address culture as implemented by an immediate embedding 'sociocultural network' whose cognitive functions interact with individual central nervous system (CNS) and immune cognition, in the sense of IR Cohen (2000), to produce a system in which social factors are intimately related to individual health and illness.

Recently D Wallace and R Wallace (2000) published work which suggested evolutionary selection pressures can determine social network structure: public policies of 'planned shrinkage' directed against minority voting blocks in New York City triggered widespread, catastrophic, contagious housing destruction which shredded organizational structures capable of pulling the minority vote in primary elections, particularly in the city's Bronx section (e.g. R Wallace and D Wallace, 1977; D Wallace and R Wallace, 1998). With regard to the Bronx, Wallace and Wallace (2000) write

"...[T]he 1970's 'planned shrinkage' catastrophe in the Bronx was so extensive and extreme that only highly resilient social subsystems have survived this draconian selection... Evolutionary process relies on the interaction between variation and selection... [and a] key concept in [its] study is 'path dependence', the idea that the future development of a system is determined not only by its present macroscopic state... but by the way in which that state was reached, in other words the revolutionary idea for Americans that history, indeed, counts. Community structure, from this viewpoint, is seen as largely determined by prior adaptations to past selection pressures, so that possible adaptations to present pressures are strongly constrained by, and will be built upon, the complex structures, often nested hierarchies, resulting form the past. Future changes may be seen as taking place on the 'surface' of the nested legacy of the past.

From this viewpoint the 'fittest' social structures remaining in the Bronx were those which persisted in the face of a degree of social disruption whose only analogs are massive natural disaster and the aftermath of modern war."

Here we will attempt to extend our earlier work on the effects of systematic social perturbation on individual health and illness, as modulated by the embedding sociocultural network, to include the fact that such networks may be simultaneously involved in both cognitive and evolutionary process. The picture which emerges will show unexpectedly punctuated mechanisms by which the impact 'selection pressures' can be expressed at the level of individual mental disorder and behavioral or immune dysfunction, through the intermediate mechanism of sociocultural cognition.

Some preliminary theoretical development is first necessary, introducing formalism from information theory and related fields, and expressing cognitive pattern recognition-and-response as a 'language' constrained by the basic limit theorems of information theory. A synergistic intertwining of punctuated learning plateaus in organizational cognition and of punctuation in response to selection pressure will emerge in a 'natural' manner.

Ergodic information sources, the Shannon-McMillan Theorem, and its generalizations

Suppose we have an ordered set of random variables, X_k, at 'times' $k = 1, 2, ...$ – 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

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

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

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

(1)

with associated conditional probabilities

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

(2)

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). If the probabilities of strings do not change in time, the source is called *memoryless*. We shall be interested in sources which can be parametized and that are, with respect to that parameter, *piecewise memoryless*, i.e. probabilities do not change markedly within a 'piece,' but may do so 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, memoryless in this sense. In what follows we use the term 'ergodic,' to mean 'piecewise 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

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

(3)

See Ash (1990), Cover and Thomas (1991) or Khinchine (1957) 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

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

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

(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

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

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

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

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

Suppose that chain elicits a corresponding chain of responses from the system of interest, producing another path $b^n = (b_1, ..., b_n)$, 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 $n \to \infty$ 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 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, Ch. 14, 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:

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

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

(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, 2001; Rojdestevnski 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, 2000) 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.

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

$$F(K_1, ...K_m) = \lim_{V \to \infty} \frac{\log[Z(K_1, .., K_m)]}{V}$$

(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

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

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

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

(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. Only the first will be needed to solve the problems we wish to address here.

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 array, sensory activity and ongoing activity depend 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 for sensory activity, ongoing activity and array components, we follow the direction suggested above and instead work entirely within the domain of the uncertainty of the ergodic information source dual to the large-scale pattern recognition process, 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 array, and of power laws for system behavior near K_C. Addition of other parameters to the system, e.g. some Q, results in a 'critical line' or surface $K_C(Q)$.

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 $J \to 0$, a simple series expansion and some clever algebra (e.g. Wilson, 1971; Binney et al., 1995; 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. Some rearrangement produces, near K_C,

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

(21)

This suggests that the 'richness' of the pattern recognition 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 (e.g.) Ash (1990) for discussion of the relation between channel capacity and information source uncertainty.

Further from the critical point matters are more complicated.

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 pattern recognition systems operating individually and more or less at random, into a single large, interlocked, coherent pattern recognition 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 pattern recognition or other 'language' processes into a single, embedding hierarchical metalanguage which contains the different 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 parametization 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 cognitive pattern recognition-and-response in terms of a ergodic information source constrained by the AEPT.

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

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

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

$$B_0 = b_0, ..., b_k$$

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

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 $h(a_0, a_1, ...a_j) \in B_0$ for all $j < m$ but $h(a_0, ..., a_m) \in B_1$.

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

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

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 $P(a_0, ...a_n)$ and

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

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

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

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

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, 1997) 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, 2000b) 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 $x = a_0, a_1, ...a_n, ...$

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.

Our scientific leap of faith – the foundation of the mathematical modeling exercise – is to claim that 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 a 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 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' R Wallace and M Fullilove (1999) postulate.

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 $0, 1, ...$, producing a path of the usual form $x = a_0, a_1, ..., a_n, ...$. 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. Arendt, 1963; Richerson and Boyd, 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.

Learning plateaus in 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 excitory 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 parametized 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 (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 sociocultural 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.

The canonical example would be a schoolchild with asthma living in a disintegrating or dysfunctional inner-city community.

We are suggesting that detailed empirical examination of the 'learning curves' of that child's immune system, school performance, and the ability of the child's immediate embedding sociocultural network to address problems, will form a unitary and synergistic whole.

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.

Non-cognitive 'learning plateaus' in evolutionary process

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

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

(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 parametized 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 equipartion 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 learning plateaus' 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.

Punctuated synergistic interaction 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 parametize 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.

Discussion and conclusions

This result has profound implications for the social determinants of individual health, since individual CNS and immune cognition are embedded in, and interact with, sociocultural processes of cognition. Sufficiently draconian external 'social selection pressures' – a euphemism for patterns of

Apartheid and the 'market' economy's systematic deprivation – should become manifest at the behavioral and cellular levels of an individual, through the intermediate mechanism of the embedding sociocultural network.

This is not a new observation. Franz Fanon (1966) describes what is an essentially similar phenomenon as follows:

> "[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, and that colonized man should have a vulnerability to immune stressors – microbiological or chemical – which should extend beyond, but may be synergistic with, the effects of deprivation alone. Differences in immune function heretofore attributed to genetic differences between populations may perhaps reflect this mechanism, which can have implications for attempts to develop vaccines against HIV and the like.

A particular implication of our work is that the ability of local sociocultural networks to respond to patterns of opportunity and threat can become closely convoluted with external selection pressures – the impacts of Apartheid and the depredations of neoliberal 'market economics'. The mathematical model suggests that increases in the selection pressure itself, or in 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.

A more general version of this work would examine how the decision processes of a 'firm', in the very largest sense of institutional economics (e.g. Hodgson, 1992), becomes convoluted with the Schumpterian selection pressures of a 'market,' suggesting a similar punctuated degradation or other changes in group cognition as the firm's position deteriorates.

These are all matters directly subject to evident empirical test.

Inherent to our approach is recognition of the burdens of history, the way in which the grammar and syntax of a culturally-determined individual 'behavioral language,' and its embodyment at cellular levels, encapsulate earlier adaptations to external selection pressures, even in the sudden absence of those pressures. Selection pressures themselves may involve deliberate policy, 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: Ethnic cleansing and pogrom by euphemism.

The mathematical model we have used to examine these matters may seem excessive to many readers. We can only reply using the words of the master mathematical ecologist EC Pielou (1977) whose answer to such criticism was that the principal value of mathematical models was to raise research questions for subsequent empirical test. Currently historical, public policy, and economic justice questions are increasingly excised from academic discussions of 'health disparities' and the 'social determinants of health' 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 extreme degree to which that excision limits the value of such work in the design of corrective policy.

References

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," *Psysica 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.

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

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.

Cohen IR, 2001, "Immunity, set points, reactive systems, and allograft rejection." To appear.

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.

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

Fanon F, 1966, *The Wretched of the Earth*, Grove Press, New York.

Gammaitoini A, P Hanggi, P Jung, and F Marchesoni, 1998, "Stochastic Resonance," *Reviews of Modern Physics*, **70**, 223-287.

Granovetter M, 1973, "The strength of weak ties," *American Journal of Sociology*, **78**, 1360-1380.

Hodgson G, 1993, *Economics and Evolution: Bringing Life Back Into Economics*, University of Michigan Press, Ann Arbor.

Holling C, 1973, "Resilience and stability of ecological systems," *Annual Reviews of Ecological Systematics*, **4**, 1-25.

Khinchine A, 1957, *The Mathematical Foundations of Information Theory*, Dover Publications, New York.

Levin S, 1989, "Ecology in theory and application," in *Applied Mathematical Ecology*, S Levin, T Hallam and L Gross (eds.), Springer-Verlag, New York.

Luce R, 1997, "Several unresolved conceptual problems of mathematical psychology," *Journal of Mathematical Psychology*, **41**, 79-87.

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

Wallace R and M Fullilove, 1999, unpublished.

Wallace R and R Fullilove, 1999, "Why simple regression models work so well describing 'risk behaviors' in the USA," *Environment and Planning A*, **31**, 719-734.

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, 2001, *The New Information Theory: applications in biology, medicine and social science*, Submitted book manuscript.

Wallace R, RG Wallace, D Wallace and M Fullilove, 2001, "Marginalization, environmental justice and generalized cognition: new perspectives on the social induction of medical disorder." Submitted.

Wilkinson R, 1996, *Unhealthy Societies*, Routledge, London.

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