Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

Version 20.1

Lost in translation: Toward a formal model of multilevel, multiscale medicine

Rodrick Wallace
Division of Epidemiology
The New York State Psychiatric Institute *

March 7, 2012

Abstract

For a broad spectrum of low level cognitive regulatory and other biological phenomena, isolation from signal crosstalk between them requires more metabolic free energy than permitting correlation. This allows an evolutionary exaptation leading to dynamic global broadcasts of interacting physiological processes at multiple scales. The argument is similar to the well-studied exaptation of noise to trigger stochastic resonance amplification in physiological subsystems. Not only is the living state characterized by cognition at every scale and level of organization, but by multiple, shifting, tunable, cooperative larger scale broadcasts that link selected subsets of functional modules to address problems. This multilevel dynamical viewpoint has implications for initiatives in translational medicine that have followed the implosive collapse of pharmaceutical industry 'magic bullet' research. In short, failure to respond to the inherently multilevel, multiscale nature of human pathophysiology will doom translational medicine to a similar implosion.

Key Words: index theorem, information theory, phase transition, sufficient conditions, topology

1 Introduction

1.1 Magic bullets

The collapse of pharmaceutical research productivity (e.g., Paul et al., 2010) has spawned attempts to speed the 'bench to bedside' translation of basic re-

*Box 47, 1051 Riverside Dr., New York, NY, 10032, Wallace@nyspi.columbia.edu

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

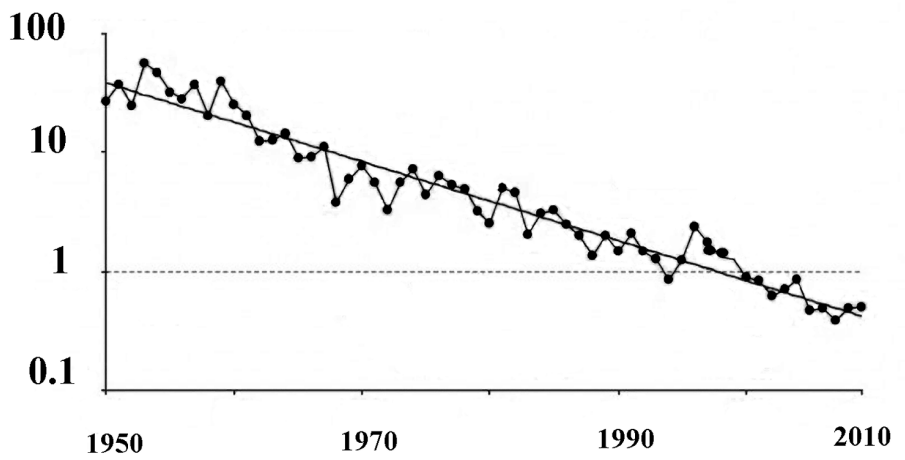

Figure 1: Adapted from Bernstein (2010). The inverse Moore's Law for pharmaceuticals. The number of small molecule and biological USFDA approvals per inflation-adjusted $ billion in research investment, 1950-2010. The apparent log-linear 'decline in research productivity' represents the failure of complex physiological phenomena to respond to simple interventions. Western medicine, as defined in the latter half of the 20th Century, has hit a brick wall, a catastrophic regime of exponential cost increase.

search into therapeutic instruments, usually seen as new magic bullets, drugs or otherwise.

The context for this effort can be seen in figure 1, adapted from Bernsetin (2010). It shows the number of small molecule and biologic USFDA approvals per inflation-adjusted billion dollars in research investment between 1950 and 2010. The cost per 'magic bullet' has increased exponentially from about $ 200 million to over $ 1.2 billion, and many pharmaceutical firms have markedly cut their research efforts as a consequence of this 'inverse Moore's Law' that represents the failure of complex physiological phenomena to respond to simple interventions.

In the words of the publisher of the recent volume by Littman and Krishna (2011), translational medicine

> ...seeks to translate biological and molecular knowledge of disease and how drugs work into innovative strategies that reduce the cost and increase the speed of delivering new medicines for patients.

Marincola (2011) has responded to current efforts with a scathing critique;

> Translational research is caught in a feedback cycle whereby complex, multi-factorial disease is confronted without sufficient understanding of human pathophysiology... It has been suggested that

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

> preclinical models do not represent human disease because of differences among species... However, this is not the principal reason they often fail to provide suitable models of human disease; the fundamental difference between preclinical and clinical testing is that in the former, the researcher can carefully select the model, whereas in the latter the clinician has to confront the unpredictable nature of human genetics and diseases, as well as environmental factors.

Wehling (2011) is more direct:

> Approximately a decade ago, translational medicine was invented both as a catchword and as a novel approach to improve success in drug development and ameliorate the low-output syndrome from collapsing pipelines. However, no major breakthroughs regarding rates of expensive late attrition or market approvals have been detected, and drug industry condensation continues to accelerate... [T]ranslational efforts so far seem to be driven mainly by claims, rather than by structure and systematic approaches. In addition, institutional structures also often seem to be only virtual or proclaiming in nature. This is simply not enough.

Horrobin (2003) lays out the lack of congruence between laboratory-level *in vitro* and patient-level *in vivo* models:

> An important distinction must be made between what might be called the anatomical biochemistry of the cell and its functional biochemistry... [I]f a particular biochemical step is present *in vitro*, then that particular biochemical step is also likely to be present... *in vivo*. We can therefore construct a network of all possible biochemical events *in vivo* by examining all possible biochemical events *in vitro*. But what the *in vitro* system cannot do is construct a functional and valid *in vivo* biochemistry. And that is potentially a fatal flaw. For in most human diseases it is the functional biochemistry and not the anatomical biochemistry which goes wrong.

Horrobin provides details, and we reproduce his Table 1:

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

***In Vitro* vs. *In Vivo* models**

1. The anatomical constraints and the cellular populations present in culture and *in vivo* are different. There is no circulation *in vitro*.

2. The types and rates of nutrient and oxygen supply, and carbon dioxide and metabolite removal, are different.

3. The restraints on cell multiplication are different.

4. The endocrine environment is different, both in terms of the amounts and patterns of hormones present and their kinetic changes.

5. The antibiotic environment is different: *in vivo* cells are not normally bathed in penicillin, streptomycin and other antibiotics, but there has been no systematic evaluation of the effects of any of these exogenous agents on metabolism.

6. The lipid environment is different. The phospholipid composition of cells in culture is quite different from the phospholipid composition of the parent *in vivo* cells. As phospholipid composition determines the quaternary structure and therefore function of a high proportion of a cell's proteins, and also determines signal transduction responses to most protein changes, it is likely that the functions of proteins *in vitro* will be, for the most part, somewhat different from the functions of those same proteins *in vivo*.

7. Even when appropriate constituents are present in culture fluid, their concentrations may be dramatically different from anything seen *in vivo*.

Horrobin (2003) goes on to liken current reductionist biomedicine to Herman Hesse's *Glasperlenspeil*, a Glass Bead Game, in which troublesome intellectuals have been seduced from real world problems into an elaborate, heavily subsidized glass cage, lucrative for those who become skilled at grantsmanship.

Here, in some contrast to current biomedical 'magic bullet' ideology, we will attempt to lay a foundation for the development of multilevel, multiscale 'magic strategies' that may, at least in their initial stages, better fit the inherently complex underlying patterns of multifactorial human pathophysiology. This is not an effort for the faint of heart, and we must begin far afield.

1.2 Beyond magic bullets

Researchers have long speculated and experimented on the role of noise in biological process via models of stochastic resonance (e.g., Park and Neelakanta, 1996; Gluckman et al., 1996; Ward, 2009; Kawaguchi et al., 2011). The necessary ubiquity of noise affecting information transmission underwent an evolutionary exaptation (e.g., Gould, 2002) to become a tool for amplification of weak signals. Here we examine the parallel necessary circumstance of information leakage between 'adjacent' communication channels or information sources, a generally unwelcome signal correlation that the electrical engineers call 'crosstalk'. The evolutionary exaptation of crosstalk appears to be nested systems of shifting

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

global biological broadcasts analogous to, but both slower and more general than, consciousness.

Baars' global workspace model of animal consciousness attributes the phenomenon to a dynamic array of unconscious cognitive modules that unite to become a global broadcast having a tunable perception threshold not unlike a theater spotlight, but whose range of attention is constrained by embedding contexts (e.g., Baars, 1988, 2005; Baars and Franklin, 2003):

1. The brain can be viewed as a collection of distributed specialized networks (processors).

2. Consciousness is associated with a global workspace in the brain – a fleeting memory capacity whose focal contents are widely distributed – 'broadcast' – to many unconscious specialized networks.

3. Conversely, a global workspace can also serve to integrate many competing and cooperating input networks.

4. Some unconscious networks, called contexts, shape conscious contents, for example unconscious parietal maps modulate visual feature cells that underlie the perception of color in the ventral stream.

5. Such contexts work together jointly to constrain conscious events.

6. Motives and emotions can be viewed as goal contexts.

7. Executive functions work as hierarchies of goal contexts.

The basic mechanism emerges 'naturally' from a relatively simple application of the asymptotic limit theorems of information theory, once a broad range of unconscious cognitive processes is recognized as inherently characterized by information sources – generalized languages (Wallace, 2000, 2005, 2007). The approach allows mapping physiological unconscious cognitive modules onto an abstract network of interacting information sources. This, in turn, permits a simplified mathematical attack based on phase transitions in network topology that, in the presence of sufficient linkage – crosstalk – permits rapid, shifting, global broadcasts.

While the mathematical description of consciousness is itself relatively simple, the evolutionary trajectories leading to its emergence seem otherwise. Here we argue that this is not the case, and that physical restrictions on the availability of metabolic free energy provide sufficient conditions for the emergence, not only of consciousness, but of a spectrum of analogous 'global' broadcast phenomena acting across a variety of biological scales of space, time, and levels of organization.

The argument is, in a sense, an extension of Gould and Lewontin's (1979) famous essay "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme". Spandrels are the triangular sectors of the intersecting arches that support a cathedral roof. They are simple byproducts of the need for arches, and their occurrence is in no way fundamental to the construction of a cathedral. Our assertion is that crosstalk between 'low level' cognitive biological modules is a similar inessential byproduct that evolutionary process has exapted to construct the dynamic global broadcasts of consciousness and a spectrum of roughly analogous physiological phenomena: Evolution built many new arches from a single spandrel.

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

We provide a minimal formal overview that will be reexpressed in more complex form, much like Onsager's nonequilibrium thermodynamics, and then, using these ideas, examine recent initiatives on 'translational medicine' (e.g., Littman and Krishna, 2011) that seek to overcome the recent collapse of pharmaceutical industry productivity (Paul et al., 2010).

2 Cognition as 'language'

Atlan and Cohen (1998) argue, in the context of a cognitive paradigm for the immune system, that the essence of cognitive function involves comparison of a perceived signal with an internal, learned or inherited picture of the world, and then, upon that comparison, choice of one response from a much larger repertoire of possible responses. That is, cognitive pattern recognition-and-response proceeds by an algorithmic combination of an incoming external sensory signal with an internal ongoing activity – incorporating the internalized picture of the world – and triggering an appropriate action based on a decision that the pattern of sensory activity requires a response.

More formally, incoming sensory input is mixed in an unspecified but systematic manner with a pattern of internal ongoing activity to create a path of combined signals $x = (a_0, a_1, ..., a_n, ...)$. Each a_k thus represents some functional composition of the internal and the external. An application of this perspective to a standard neural network is given in Wallace (2005, p.34).

This path is fed into a highly nonlinear, but otherwise similarly unspecified, decision oscillator, h, which generates an output $h(x)$ that is an element of one of two disjoint sets B_0 and B_1 of possible system responses. Let

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

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

Assume a graded response, supposing that if

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

the pattern is not recognized, and if

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

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

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

For each positive integer n, let $N(n)$ be the number of high probability grammatical and syntactical paths of length n that begin with some particular

a_0 and lead to the condition $h(x) \in B_1$. Call such paths 'meaningful', assuming, not unreasonably, that $N(n)$ will be considerably less than the number of all possible paths of length n leading from a_0 to the condition $h(x) \in B_1$.

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

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

(1)

both exists and is independent of the path x.

Call such a pattern recognition-and-response cognitive process *ergodic*. Not all cognitive processes are likely to be ergodic, implying that H, if it indeed exists at all, is path dependent, although extension to nearly ergodic processes, in a certain sense, seems possible (e.g., Wallace, 2005, pp. 31-32).

Invoking the spirit of the Shannon-McMillan Theorem, it is possible to define an adiabatically, piecewise stationary, 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 classic relations

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

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

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

(2)

This information source is defined as *dual* to the underlying ergodic cognitive process, in the sense of Wallace (2000, 2005).

The essence of 'adiabatic' is that, when the information source is parameterized according to some appropriate scheme, within continuous 'pieces', changes

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

in parameter values take place slowly enough so that the information source remains as close to stationary and ergodic as needed to make the fundamental limit theorems work. By 'stationary' we mean that probabilities do not change in time, and by 'ergodic' (roughly) that cross-sectional means converge to long-time averages. Between 'pieces' one invokes various kinds of phase change formalism, for example renormalization theory in cases where a mean field approximation holds (Wallace, 2005), or variants of random network theory where a mean number approximation is applied. More will be said of this latter approach below.

Recall that the Shannon uncertainties $H(...)$ are cross-sectional law-of-large-numbers sums of the form $-\sum_k P_k \log[P_k]$, where the P_k constitute a probability distribution. See Cover and Thomas (2006), Ash (1990), or Khinchin (1957) for the standard details.

A formal equivalence class algebra can be constructed by choosing different origin points a_0 and defining equivalence of two states by the existence of a high probability meaningful path connecting each of them with the same origin. Disjoint partition by equivalence class, analogous to orbit equivalence classes for dynamical systems, defines the vertices of the proposed network of cognitive dual languages. Each vertex then represents a different information source dual to a cognitive process. This is not a representation of a neural network as such, or of some circuit in silicon. It is, rather, an abstract set of 'languages' dual to the cognitive biological processes.

This structure generates a groupoid, leading to complicated algebraic properties we will not examine further (Wallace and Fullilove, 2008, Section 3.2).

A recent series of articles has applied this perspective to cognitive paradigms for gene expression (Wallace and Wallace, 2009, 2010), the regulation of protein folding (Wallace, 2010, 2011a, b), and the production and regulation of the glycan determinants that coat cellular surfaces and, in fact, constitute the principal means of biological information transmission (Wallace, 2012). The essential point is that such regulatory machineries can become nodes on a network of interacting information sources whose connections, by crosstalk, become the means for shifting, tunable, global broadcasts analogous to neural consciousness that dedicate chosen sets of physiological subsystems to selected problems.

3 No Free Lunch

Given a set of biological cognitive modules that become linked to solve a problem – e.g., riding a bicycle in heavy traffic, followed by localized wound healing – the famous 'no free lunch' theorem of Wolpert and Macready (1995, 1997) illuminates the next step in the argument. As English (1996) states the matter,

> ...Wolpert and Macready... have established that there exists no generally superior [computational] function optimizer. There is no 'free lunch' in the sense that an optimizer 'pays' for superior performance on some functions with inferior performance on others...

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

> gains and losses balance precisely, and all optimizers have identical average performance... [That is] an optimizer has to 'pay' for its superiority on one subset of functions with inferiority on the complementary subset...

Another way of stating this conundrum is to say that a computed solution is simply the product of the information processing of a problem, and, by a very famous argument, information can never be gained simply by processing. Thus a problem X is transmitted as a message by an information processing channel, Y, a computing device, and recoded as an answer. By the extended argument of the Mathematical Appendix, there will be a channel coding of Y which, when properly tuned, is *itself* most efficiently 'transmitted', in a purely formal sense, by the problem – the 'message' X. In general, then, the most efficient coding of the transmission channel, that is, the best algorithm turning a problem into a solution, will necessarily be highly problem-specific. Thus there can be no best algorithm for all sets of problems, although there will likely be an optimal algorithm for any given set.

Based on the no free lunch argument, it is clear that different challenges facing an entity must be met by different arrangements of cooperating basic 'low level' cognitive modules. It is possible to make a very abstract picture of this phenomenon, not based on anatomy, but rather on the linkages between the information sources dual to the basic physiological and learned unconscious cognitive modules (UCM). That is, *the remapped network of lower level cognitive modules is reexpressed in terms of the information sources dual to the UCM.* Given two distinct problems classes (e.g., riding a bicycle vs. wound healing), there must be two different 'wirings' of the information sources dual to the available physiological UCM, as in figure 2, with the network graph edges measured by the amount of information crosstalk between sets of nodes representing the dual information sources. A more formal treatment of such coupling can be given in terms of network information theory (Cover and Thomas, 2006), particularly incorporating the effects of embedding contexts, implied by the 'external' information source Z – signals from the environment.

The possible expansion of a closely linked set of information sources dual to the UCM into a global workspace/broadcast – the occurrence of a kind of 'spandrel' – depends, in this model, on the underlying network topology of the dual information sources and on the strength of the couplings between the individual components of that network. For random networks the results are well known, based on the work of Erdos and Renyi (1960). Following the review by Spenser (2010) closely (see, e.g., Boccaletti et al., 2006, for more detail), assume there are n network nodes and e edges connecting the nodes, distributed with uniform probability – no nonrandom clustering. Let $G[n, e]$ be the state when there are e edges. The central question is the typical behavior of $G[n, e]$ as e changes from 0 to $(n-2)!/2$. The latter expression is the number of possible pair contacts in a population having n individuals. Another way to say this is to let $G(n, p)$ be the probability space over graphs on n vertices where each pair is adjacent with independent probability p. The behaviors of $G[n, e]$ and

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

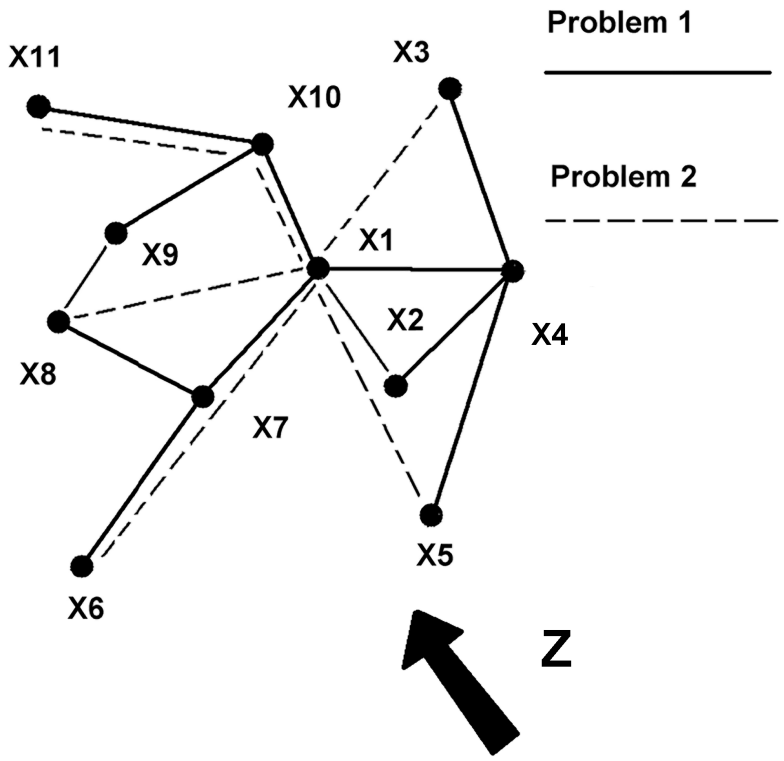

Figure 2: By the no free lunch theorem, two markedly different problems will be optimally solved by two different linkages of available lower level cognitive modules – characterized now by their dual information sources X_j – into different temporary networks of working structures, here represented by crosstalk among those sources rather than by the physiological UCM themselves. The embedding information source Z represents the influence of external signals whose effects can be at least formally accounted for by network information theory.

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

$G(n,p)$ where $e = p(n-2)!/2$ are asymptotically the same.

For 'real world' biological and social structures, one can have $p = f(e,n)$, where f may not be simple or even monotonic. For example, while low e would almost always be associated with low p, beyond some threshold, high e might drive individuals or nodal groups into isolation, decreasing p and producing an 'inverted-U' signal transduction relation akin to stochastic resonance. Something like this would account for Fechner's law which states that perception of sensory signals often scales as the log of the signal intensity.

For the simple random case, however, we can parameterize as $p = c/n$. The graph with $n/2$ edges then corresponds to $c = 1$. The essential finding is that the behavior of the random network has three sections:

[1] If $c < 1$ all the linked subnetworks are very small, *and no global broadcast can take place.*

[2] If $c = 1$ there is a single large interlinked component of a size $\approx n^{2/3}$.

[3] If $c > 1$ then there is a single large component of size yn – a global broadcast – where y is the positive solution to the equation

$$\exp(-cy) = 1 - y. \tag{3}$$

Then

$$y = \frac{W(-c/\exp(c)) + c}{c}, \tag{4}$$

where W is the Lambert W function.

The solid line in figure 3 shows y as a function of c, representing the fraction of network nodes that are incorporated into the interlinked giant component – a de-facto global broadcast for interacting UCM. To the left of $c = 1$ there is no giant component, and large scale cognitive process is not possible.

The dotted line, however, represents the fraction of nodes in the giant component for a highly nonrandom network, a star-of-stars-of-stars (SoS) in which every node is directly or indirectly connected with every other one. For such a topology there is no threshold, only a single giant component, showing that the emergence of a giant component in a network of information sources dual to the UCM is dependent on a network topology that may itself be tunable (Wallace and Fullilove, 2008). We will obtain a generalization of this result by means of an index theorem argument below.

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

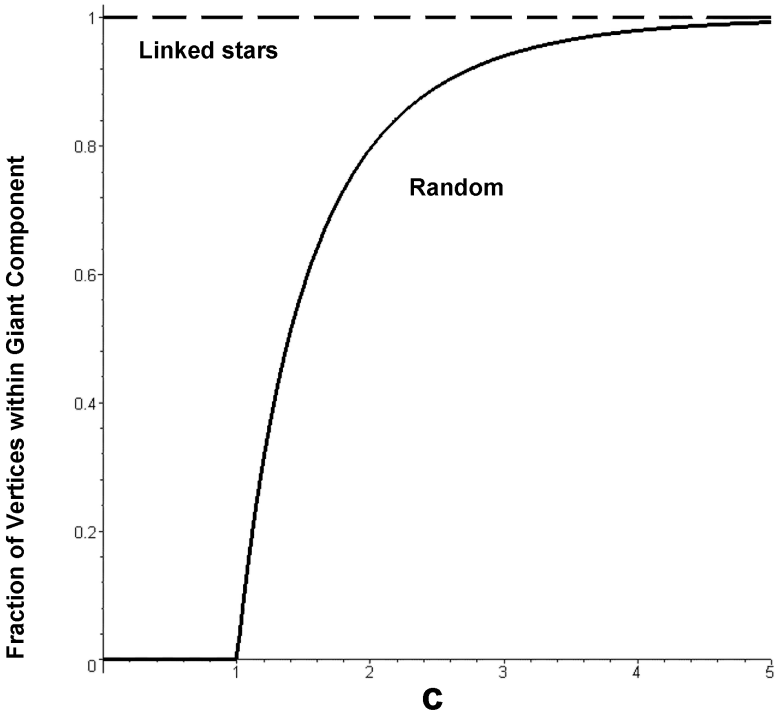

Figure 3: Fraction of network nodes in the giant component as a function of the crosstalk coupling parameter c. The solid line represents a random graph, the dotted line a star-of-stars-of-stars network in which all nodes are interconnected, showing that the dynamics of giant component emergence are highly dependent on an underlying network topology that, for UCM, may itself be tunable. For the random graph, a strength of $c < 1$ precludes emergence of a larger-scale 'global' broadcast.

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

4 Multiple broadcasts, punctuated detection

The random network development above is predicated on there being a variable average number of fixed-strength linkages between components. Clearly, the mutual information measure of cross-talk is not inherently fixed, but can continuously vary in magnitude. We address this by a parameterized renormalization. In essence the modular network structure linked by mutual information interactions has a topology depending on the degree of interaction of interest. Suppose we define an interaction parameter ω, a real positive number, and look at geometric structures defined in terms of linkages which are zero if mutual information is less than, and 'renormalized' to unity if greater than, ω. Any given ω will define a regime of giant components of network elements linked by mutual information greater than or equal to it.

Now *invert the argument*: A given topology for the giant component will, in turn, define some critical value, ω_C, so that network elements interacting by mutual information less than that value will be unable to participate, i.e., will be locked out and not be 'consciously' perceived. We hence are assuming that the ω is a tunable, syntactically-dependent, detection limit, and depends critically on the instantaneous topology of the giant component of linked cognitive modules defining the global broadcast. That topology is, fundamentally, the basic tunable syntactic filter across the underlying modular structure (Wallace and Fullilove, 2008), and variation in ω is only one aspect of a much more general topological shift. Further analysis can be given in terms of a topological rate distortion manifold (Wallace and Fullilove, 2008; Glazebrook and Wallace, 2009).

There is considerable empirical evidence from fMRI brain imaging and many other experiments to show that individual animal consciousness – restricted by necessity of a time constant near 100 milliseconds – involves a single, shifting and tunable, global broadcast, a matter leading necessarily to the phenomenon of inattentional blindness. Multiple cognitive submodules within systems not constrained to the 100 ms time range, for example institutions – individuals, departments, formal and informal workgroups – by contrast, can do more than one thing, and indeed, are usually required to multitask. Clearly, then, multiple workspace global broadcasts lessen the probability of inattentional blindness, if there is time to support them, but do not eliminate it, and introduce critical failure modes related to the degradation of information transmitted between global broadcasts.

We must postulate a set of crosstalk information measures between cognitive submodules, each associated with its own tunable giant component having its own special topology.

Again, although animal consciousness, with its 100ms time constant, seems restricted to a single tunable global broadcast, it is clear that slower physiological consciousness global broadcast analogs would permit individual subsystems, or localized sets of such subsystems, to engage in more than one global broadcast at a time, to multitask, in the same sense that workgroups within an institution will usually be given more than one task at a time. Thus the immune system

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

can be expected to simultaneously engage in wound healing, attack on invading microorganisms, neuroimmuno dialog, and routine tissue maintenance tasks.

Following, again, the arguments of Wallace and Fullilove (2008), we further expand the argument.

Suppose, now, a set of giant components of interacting cognitive physiological submodules at some time k is characterized by a set of parameters $\Omega_k \equiv (\omega_1^k, ..., \omega_m^k)$. Fixed parameter values define a particular giant component set having a particular set of topological structures. Suppose that, over a sequence of times, the set of giant components can be characterized by a possibly coarse-grained path $x_n = (\Omega_0, \Omega_1, ..., \Omega_{n-1})$ having significant serial correlations permitting definition of an adiabatically, piecewise stationary, ergodic (APSE) information source that we call $\mathbf{X}$. Suppose a set of (external or internal) signals impinging on the set of giant components is also highly structured and forms another APSE information source $\mathbf{Y}$. Then we can define joint and conditional Shannon uncertainties leading to an iterated chain rule argument as above, complicated by the necessity of information transfer between the multiple, shifting spotlights characterizing the interacting giant components. To reiterate, a major possible source of pathology would be distortion in the transmission of information between interacting global broadcasts.

5 Information and metabolic free energy

The information sources dual to unconscious cognitive modules represented in figure 2 are not independent, but are correlated, so that a joint information source can be defined having the properties

$$H(X_1, ..., X_n) \leq \sum_{j=1}^{n} H(X_j). \tag{5}$$

This result is known as the information chain rule (e.g., Cover and Thomas, 2006), and has profound implications: Feynman (2000) describes in great detail how information and free energy have an inherent duality. Feynman, in fact, defines information precisely as the free energy needed to erase a message. The argument is surprisingly direct (e.g., Bennett, 1988), and for very simple systems it is easy to design a small (idealized) machine that turns the information within a message directly into usable work – free energy. Information is a form of free energy and the construction and transmission of information within living things consumes metabolic free energy, with inevitable losses via the second law of thermodynamics.

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

Information catalysis arises most simply via the information theory chain rule. Restricting the argument to two information sources, X and Y, one can define jointly typical paths $z_i = (x_i, y_i)$ having the joint information source uncertainty $H(X,Y)$ satisfying $H(X,Y) = H(X) + H(Y|X) \leq H(X) + H(Y)$.

Of necessity, then, $H(X,Y) \leq H(X) + H(Y)$ if $H(Y) \neq 0$.

Within a biological structure, however, there will be an ensemble of possible reactions, driven by available metabolic free energy, so that, taking $\hat{H}$ as representing an average,

$$\hat{H}(X,Y) < \hat{H}(X) + \hat{H}(Y).$$

(6)

This is a very general result that, by the equivalence of information and free energy, leads to a model in which interacting biological signals can 'canalize' the overall behavior of the system: Interaction consumes less metabolic free energy than signal isolation.

Typically, letting $Q(\kappa M) \geq 0, Q(0) = 0$ represent an intensity measure of available metabolic free energy, and C be the maximum channel capacity available to the cognitive biological processes of interest, one would expect

$$\hat{H} = \frac{\int_0^C H \exp[-H/Q] dH}{\int_0^C \exp[-H/Q] dH} = \frac{Q[\exp(C/Q) - 1] - C}{\exp(C/Q) - 1}.$$

(7)

κ is an inverse energy intensity scaling constant that may be quite small indeed, a consequence of entropic translation losses between metabolic free energy and the expression of information. Note that, near $M = 0$, we can expand Q as a Taylor series, with a first term $Q \approx \kappa M$.

This expression tops out quite rapidly with increases in either C or Q, producing energy- and channel capacity- limited results

$$\hat{H} = Q(\kappa M), C/2.$$

(8)

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

Then, expanding Q near zero, the two limiting relations imply

$$Q(\kappa M_{X,Y}) < Q(\kappa M_X) + Q(\kappa M_Y) \rightarrow M_{X,Y} < M_X + M_Y,$$

$$C_{X,Y} < C_X + C_Y.$$

(9)

The channel capacity constraint can be parsed further for a noisy Gaussian channel. Then (Cover and Thomas, 2006)

$$C = 1/2 \log[1 + \mathcal{P}/\sigma^2] \approx 1/2\mathcal{P}/\sigma^2$$

(10)

for small $\mathcal{P}/\sigma^2$, where $\mathcal{P}$ is the 'power constraint' such that $E(X^2) < \mathcal{P}$ and σ^2 is the noise variance. Assuming information sources X and Y act on the same scale, so that noise variances are the same and quite large, then we may take $\mathcal{P} = Q(\kappa M)$ – channel power is determined by available metabolic free energy – and we recover the expression

$$Q(\kappa M_{X,Y}) < Q(\kappa M_X) + Q(\kappa M_Y).$$

Both limiting inequalities are, then, free energy relations leading to a kind of 'reaction canalization' in which a set of lower level cognitive modules consumes less metabolic free energy if interaction among them is permitted than under conditions of individual signal isolation.

The global broadcast mechanisms of consciousness and its slower physiological generalizations make an arch of this spandrel, using the lowered free energy requirement of crosstalk interaction between low level cognitive modules as the springboard for launching (sometimes) rapid, tunable, more highly correlated, multiple global broadcasts that link those modules to solve problems.

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

6 Environmental signals

Lower level cognitive modules operate within larger, highly structured, environmental signals and other constraints whose regularities may also have a recognizable grammar and syntax, represented in figure 2 by an embedding information source Z. Under such a circumstance the splitting criterion for three jointly typical sequences is given by the classic relation of network information theory (Cover and Thomas, 2006, Theorem 15.2.3)

$$I(X_1, X_2|Z) = H(Z) + H(X_1|Z) + H(X_2|Z) - H(X_1, X_2, Z) \tag{11}$$

that generalizes as

$$I(X_1, ..., X_n|Z) = H(Z) + \sum_{j=1}^{n} H(X_j|Z) - H(X_1, ..., X_n, Z) \tag{12}$$

More complicated multivariate typical sequences are treated much the same (e.g., El Gamal and Kim, 2010, p.2-26). Given a basic set of interacting information sources $(X_1, ..., X_k)$ that one partitions into two ordered sets $X(\mathcal{J})$ and $X(\mathcal{J}')$, then the splitting criterion becomes $H[X(\mathcal{J}|\mathcal{J}')]$. Extension to a greater number of ordered sets is straightforward.

Then the joint splitting criterion – I, H above – however it may be expressed as a composite of the underlying information sources and their interactions, satisfies a relation like the first expression in equation (2), where $N(n)$ is the number of high probability jointly typical paths of length n, and the theory carries through, now incorporating the effects of external signals as the information source Z.

7 The simplest 'regression' model

Given the splitting criteria $I(X_1, ..., X_n|Z)$ or $H[X(\mathcal{J}|\mathcal{J}')]$ as above, the essential point is that these are the limit, for large n, of the expression $\log[N(n)]/n$, where $N(n)$ is the number of jointly typical paths of the interacting information

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

sources of length n. Again, as Feynman (2000) argues at great length, information is simply another form of free energy, and its dynamics can be expressed using a formalism similar to Onsager's nonequilbrium thermodynamics.

The argument is direct.

First, the physical model. Let $F(K)$ be the free energy density of a physical system, K the normalized temperature, V the volume and $Z(K,V)$ the *partition function* defined from the Hamiltonian characterizing energy states E_i. Then

$$Z(V,K) \equiv \sum_i \exp[-E_i(V)/K],$$

(13)

and

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

If a nonequilibrium physical system is parameterized by a set of variables $\{K_i\}$, then the *empirical Onsager equations* are defined in terms of the gradient of the entropy $S \equiv F - \sum_j K_j dF/dK_j$ as

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

(14)

where the $L_{i,j}$ are empirical constants. For a physical system having microreversibility, $L_{i,j} = L_{j,i}$. For an information source where, for example, ‘ the ’ has a much different probability than ‘ eht ’, no such microreversibility is possible, and no ‘reciprocity relations’ can apply.

For stochastic systems this generalizes to the set of stochastic differential equations

$$dK_t^j = \sum_i [L_{j,i}(t,...\partial S/\partial K^i...)dt + \sigma_{j,i}(t,...\partial S/\partial K^i)dB_t^i]$$

$$= L(t, K^1, ..., K^n)dt + \sum_i \sigma(t, K^1, ..., K^n)dB_t^i,$$

(15)

where terms have been collected and expressed in terms of the driving parameters. The dB_t^i represent different kinds of 'noise' whose characteristics are usually expressed in terms of their quadratic variation. See any standard text for definitions, examples, and details.

The essential trick is to recognize that, for the splitting criteria $I(X_1, ..., X_n|Z)$ or $H[X(\mathcal{J}|\mathcal{J}')]$, the role of information as a form of free energy, and the corresponding limit in $\log[N(n)]/n$, make it possible to define entropy-analogs as

$$S \equiv I(...K^i...) - \sum_j K^j \partial I/\partial K^j$$

$$S \equiv H[X(\mathcal{J}|\mathcal{J}')] - \sum_j K^j \partial H[X(\mathcal{J}|\mathcal{J}')]/\partial K^j.$$

$$S \propto M_{\mathcal{J}|\mathcal{J}'} - \sum_j K^j \partial M_{\mathcal{J}|\mathcal{J}'}/\partial K^j.$$

(16)

where, for the last relation, we have invoked the embedding metabolic free energies that instantiate the actual mechanisms by which information is transmitted.

The basic information theory 'regression equations' for the system of figures 2 and 3, driven by a set of external 'sensory' and other, internal, signal parameters $\mathbf{K} = (K^1, ..., K^n)$ *that may be measured by the information source uncertainty of other information sources* is then precisely the set of equations (15) above.

Several features emerge directly from invoking this 'coevolutionary' approach.

The first involves Pettini's (2007) topological hypothesis: A fundamental change in the underlying topology of a system characterized by any free energy-like 'Morse Function' is a necessary condition for the kind of phase transition shown in figure 3. What seems clear from the neurological context is that a converse topological tuning of the threshold for the global broadcast phase transition is possible.

Second, there are several obvious possible dynamic patterns:

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

1. Setting equation (15) equal to zero and solving for stationary points gives attractor states since the noise terms preclude unstable equilibria.

2. This system may converge to limit cycle or pseudorandom 'strange attractor' behaviors in which the system seems to chase its tail endlessly within a limited venue – a kind of 'Red Queen' pathology.

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

We are deeply enmeshed in a highly recursive phenomenological stochastic differential equations (as in, e.g., Zhu et al. 2007), but in a dynamic rather than static manner. The objects of this dynamical system are equivalence classes of information sources, rather than simple 'stationary states' of a dynamical or reactive chemical system. The necessary conditions of the asymptotic limit theorems of communication theory have beaten the mathematical thicket back one layer.

Third, as Champagnat et al. (2006) note, shifts between the quasi-equilibria of a coevolutionary system can be addressed by the large deviations formalism. They find that the issue of dynamics drifting away from trajectories predicted by the canonical equation can be investigated by considering the asymptotic of the probability of 'rare events' for the sample paths of the diffusion.

By 'rare events' they mean diffusion paths drifting far away from the direct solutions of the canonical equation. The probability of such rare events is governed by a large deviation principle: when a critical parameter (designated ϵ) goes to zero, the probability that the sample path of the diffusion is close to a given rare path ϕ decreases exponentially to 0 with rate $\mathcal{I}(\phi)$, where the 'rate function' $\mathcal{I}$ can be expressed in terms of the parameters of the diffusion. This result, in their view, can be used to study long-time behavior of the diffusion process when there are multiple attractive singularities. Under proper conditions the most likely path followed by the diffusion when exiting a basin of attraction is the one minimizing the rate function $\mathcal{I}$ over all the appropriate trajectories. The time needed to exit the basin is of the order $\exp(\mathcal{V}/\epsilon)$ where $\mathcal{V}$ is a quasi-potential representing the minimum of the rate function $\mathcal{I}$ over all possible trajectories.

An essential fact of large deviations theory is that the rate function $\mathcal{I}$ which Champagnat et al. invoke can almost always be expressed as a kind of entropy, that is, having the canonical form

$$\mathcal{I} = -\sum_j P_j \log(P_j) \tag{17}$$

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

for some probability distribution. This result goes under a number of names; Sanov's Theorem, Cramer's Theorem, the Gartner-Ellis Theorem, the Shannon-McMillan Theorem, and so forth (Dembo and Zeitouni, 1998).

These considerations lead very much in the direction of equation (15), but now seen as subject to internally-driven large deviations *that are themselves described as information sources*, providing $K = f(\mathcal{I})$-parameters that can trigger punctuated shifts between quasi-stable modes. Thus both external signals, characterized by the information source Z, and internal 'ruminations', characterized by the information source $\mathcal{I}$, can provide K-parameters that serve to drive the system to different quasi-equilibrium 'conscious attention states' in a highly punctuated manner, if they are of sufficient magnitude to overcome the topological renormalization ω-constraints described above.

More generally, following the topological arguments of Section 4, setting equation (15) to zero generates an *index theorem* (Hazewinkel, 2002), in the sense of Atiyah and Singer (1963), that relates analytic results – the solutions to the equations – to an underlying set of topological structures representing the eigenmodes of a complicated Ω-network geometric operator whose spectrum represents the possible multiple global broadcast states of the system.

Finally, an essential perspective of the Baars global workspace/global broadcast model of animal consciousness is the role of contexts in defining the 'riverbanks' between which the stream of individual consciousness flows. The most essential context for the dynamic global broadcasts associated with human pathophysiology is the embedding cultural milieu that most distinguishes humans from other animals. Richerson and Boyd (2006), for example, argue persuasively that culture is as much a part of human biology as the enamel on our teeth and bipedal locomotion. That is, culture and human biology are inextricably linked. Pathophysiology involves developmental trajectories driven by cognitive processes of gene expression (e.g., Wallace and Wallace, 2010) that, in the sense of figure 2, respond to environmental signals largely defined by cultural context as mitigated by social interaction and the power relations between groups. Expressing the regularities of sociocultural interaction in terms of the grammar and syntax of an information source would permit their incorporation into the dynamics of equation (15) in a natural manner.

Clearly, then, the 'riverbank' nature of cultural pattern and power relations that directs the stream of human pathophysiology arises from the difference in time scales between normal physiological process and the rate of change of culture, catastrophic events aside.

The general argument implied by equation (17) can be directly expressed in terms of an inherently multi-scale therapeutic intervention, or the 'farming' of pathophysiology within the interpenetrating, shifting, global broadcasts of the mind/body system.

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

8 Multiscale therapeutic intervention

8.1 The generalized retina

Cohen (2000) argues for an 'immunological homunculus' as the immune system's perception of the body as a whole. The particular utility of such a thing, in his view, is that sensing perturbations in a bodily self-image can serve as an early warning sign of pending necessary inflammatory response – expressions of tumorigenesis, acute or chronic infection, parasitization, and the like. Thayer and Lane (2000) argue something analogous for emotional response as a quick internal index of larger patterns of threat or opportunity.

It seems obvious that the tunable, shifting global broadcasts of interacting cognitive submodules explored above must also have coherent internal self-images of the states of the mind/body and its social relationships. Such an inferred picture might be termed a 'generalized retina' (GR). It is possible to use the responses of the GR to characterize physiological/mental responses to both illness and to medical interventions used to treat that illness. Illness and treatment may then come to reflect one another in a hall of mirrors reminiscent of Jerne's idiotypic network proposed for the dynamics of the immune system.

Suppose, rather than measuring either stress or cognitive submodule and broadcast function directly, it is possible to determine the concentrations of hormones, neurotransmitters, certain cytokines, and other biomarkers, or else macroscopic behaviors, beliefs, feelings, or other responses associated with the function of cognitive submodules according to some natural time frame inherent to the system. This would typically be the circadian cycle in both men and women, and the hormonal cycle in premenopausal women. Suppose, in the absence of extraordinary meaningful psychosocial stress, it is possible to measure a series of n biomarker concentrations, behavioral characteristics, other indices at time t which we represent as an n-dimensional vector X_t. Suppose it possible to conduct a number of experiments, and create a regression model so that, in the absence of perturbation, it becomes possible to write, to first order, the set of markers at time $t+1$ in terms of that at time t using a matrix equation of the form

$$X_{t+1} \approx \mathbf{R} X_t, \tag{18}$$

where $\mathbf{R}$ is the matrix of regression coefficients, with normalization to a zero vector of constant terms.

Write a GR response to short-term perturbation as

$$X_{t+1} = (\mathbf{R}_0 + \delta\mathbf{R}_{t+1}) X_t,$$

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

where $\delta\mathbf{R}$ represents variation of the generalized cognitive self-image about the basic state $\mathbf{R}_0$.

Now impose a Jordan block diagonalization in terms of the matrix of (generally nonorthogonal) eigenvectors $\mathbf{Q}_0$ of some 'zero reference state' $\mathbf{R}_0$, obtaining, for an initial condition that is an eigenvector $Y_t \equiv Y_k$ of $\mathbf{R}_0$,

$$Y_{t+1} = (\mathbf{J}_0 + \delta\mathbf{J}_{t+1})Y_k = \lambda_k Y_k + \delta Y_{t+1} =$$

$$\lambda_k Y_k + \sum_{j=1}^{n} a_j Y_j,$$

(19)

where $\mathbf{J}_0$ is a (block) diagonal matrix as above, $\delta\mathbf{J}_{t+1} \equiv \mathbf{Q}_0\delta\mathbf{R}_{t+1}\mathbf{Q}_0^{-1}$, and δY_{t+1} *has been expanded in terms of a spectrum of the eigenvectors of* $\mathbf{R}_0$, with

$$|a_j| \ll |\lambda_k|, |a_{j+1}| \ll |a_j|.$$

(20)

The essential point is that, provided $\mathbf{R}_0$ has been properly tuned, so that this condition is true, the first few terms in the spectrum of the plieotropic iteration of the eigenstate will contain almost all of the essential information about the perturbation, i.e., most of the variance. This is precisely similar to the detection of color in the optical retina, where three overlapping non-orthogonal 'eigenmodes' of response suffice to characterize a vast array of color sensations. Here, if a concise spectral expansion is possible, a very small number of (typically nonorthogonal) 'generalized cognitive eigenmodes' permit characterization of a vast range of external perturbations, and rate distortion constraints become very manageable indeed. Thus GR responses – the spectrum of excited eigenmodes of $\mathbf{R}_0$, provided it is properly tuned – can be a very accurate and precise gage of environmental perturbation.

The choice of zero reference state $\mathbf{R}_0$, the 'base state' from which perturbations are measured, is, apparently, a highly nontrivial task, necessitating a specialized apparatus.

This is no small matter. According to current theory, the adapted human mind functions through the action and interaction of distinct mental modules which evolved fairly rapidly to help address special problems of environmental

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

and social selection pressure faced by our Pleistocene ancestors (Barkow et al., 1992). It appears necessary to postulate other physiological and social cognitive modules. As is well known in computer engineering, calculation by specialized submodules – numeric processor chips – can be a far more efficient means of solving particular well-defined classes of problems than direct computation by a generalized system. It appears, then, that generalized physiological cognition has evolved specialized submodules to speed the address of certain commonly recurring challenges. Nunney (1999) has argued that, as a power law of cell count, specialized subsystems are increasingly required to recognize and redress tumorigenesis, mechanisms ranging from molecular error-correcting codes, to programmed cell death, and finally full-blown immune attack.

It seems that identification of the designated normal state of the GR – generalized cognition's self-image of the body and its social relationships – is difficult, requiring a dedicated cognitive submodule within overall generalized cognition. This is essentially because, for the vast majority of information systems, unlike mechanical systems, there are no restoring springs whose low energy state automatically identifies equilibrium: relatively speaking, all states of the GR are 'high energy' states. That is, active comparison must be made of the state of the GR with some stored internal reference picture, and a decision made about whether to reset to zero, which is a cognitive process. The complexity of such a submodule may also follow something like Nunney's power law with animal size, as the overall generalized cognition and its image of the self, become increasingly complicated with rising number of cells and levels of linked cognition.

Failure of that cognitive submodule can result in identification of an excited state of the GR as normal, triggering the collective patterns of systemic activation which, following the argument of Wallace (2004), constitute certain comorbid mental and chronic physical disorders. This would result in a relatively small number of characteristic eigenforms of comorbidity, which would typically become more mixed with increasing disorder.

In sum, since such 'zero mode identification' (ZMI) is a (presumed) cognitive submodule of overall generalized cognition, it involves convoluting incoming 'sensory' with 'ongoing' internal memory data in choosing the zero state, i.e., defining $\mathbf{R}_0$. The dual information source defined by this cognitive process can then interact in a punctuated manner with 'external information sources' according to the Rate Distortion and related arguments above. From a Rate Distortion Theorem perspective (Cover and Thomas, 2006), then, those external information sources literally write a distorted image of themselves onto the ZMI, often in a punctuated manner: (relatively) sudden onset of a developmental trajectory to comorbid mental disorders and pathophysiology.

Different systems of external signals – including but not limited to structured psychosocial stress – will, presumably, write different characteristic images of themselves onto the ZMI cognitive submodule, i.e., trigger different patterns of comorbid mental/physical disorder.

A brief reformulation in abstract terms may be of interest. Recall that the essential characteristic of cognition in this formalism involves a function h which maps a (convolutional) path $x = a_0, a_1, ..., a_n, ...$ onto a member of one of two

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

disjoint sets, B_0 or B_1. Thus respectively, either (1) $h(x) \in B_0$, implying no action taken, or (2), $h(x) \in B_1$, and some particular response is chosen from a large repertoire of possible responses. Some 'higher order cognitive module' might be needed to identify what constituted B_0, the set of 'normal' states. Again, this is because there is no low energy mode for information systems: virtually all states are more or less high energy states, and there is no way to identify a ground state using the physicist's favorite variational or other minimization arguments on energy.

Suppose that higher order cognitive module, now recognizable as a kind of Zero Mode Identification, interacts with an embedding language of structured psychosocial stress (or other systemic perturbation) and, instantiating a Rate Distortion image of that embedding stress, begins to include one or more members of the set B_1 into the set B_0. Recurrent 'hits' on that aberrant state would be experienced as episodes of highly structured comorbid mind/body pathology.

Empirical tests of this hypothesis, however, quickly lead again into real-world regression models involving the interrelations of measurable biomarkers, beliefs, behaviors, feelings, and so on, requiring formalism much like that used above. The GR can, then, be viewed as a generic heuristic device typifying such regression approaches.

The generalized retina is more appropriately characterized as a 'Rate Distortion Manifold', a local projection that, through overlap, has global structure, much like the tangent planes to a complicated geometric object. Glazebrook and Wallace (2009a) provide more detailed, indeed cutting-edge, mathematical treatment. Some thought will show that the GR and the more abstract Rate Distortion Manifold are explicit examples of the general 'tuning theorem' argument of the Mathematical Appendix.

8.2 Therapeutic efficacy

To reiterate, if $\mathbf{X}$ represents the information source dual to 'zero mode identification' in a generalized cognition, and if $\mathbf{Z}$ is the information source characterizing structured psychosocial stress or other noxious embedding context, the mutual information between them

$$I(\mathbf{X};\mathbf{Z}) = H(\mathbf{X}) - H(\mathbf{X}|\mathbf{Z}) \tag{21}$$

serves as a splitting criterion for pairs of linked paths of states.

Suppose it possible to parameterize the coupling between these interacting information sources by some appropriate index, ω, writing

$$I(\mathbf{X};\mathbf{Z}) = I[\omega],$$

(22)

with structured psychosocial stress or some other noxious condition as the embedding context.

Socioculturally constructed and structured psychosocial stress or other noxious exposure, in this model having both (generalized) grammar and syntax, can be viewed as entraining the function of zero mode identification when the coupling with stress exceeds a threshold, following the arguments of Section 4. More than one threshold appears likely, accounting in a sense for the typically staged nature of environmentally caused disorders. These should result in a synergistic – i.e., comorbidly excited – mixed affective, rationally cognitive, psychosocial, and inflammatory or other physical excited state of otherwise normal response, and represent the effect of stress on the linked decision processes of various cognitive functions, in particular through the identification of a false 'zero mode' of the GR. This is a collective, but highly systematic, 'tuning failure' that, in the Rate Distortion sense, represents a literal image of the structure of imposed pathogenic context written upon the ability of the GR to characterize a normal condition of excitation, causing a mixed, shifting, highly dynamic excited state of chronically comorbid mental and physical disorder.

In this model different eigenmodes Y_k of the GR regression model characterized by the matrix $\mathbf{R}_0$ can be taken to represent the 'shifting-of-gears' between different 'languages' defining the sets B_0 and B_1. That is, different eigenmodes of the GR would correspond to different required (and possibly mixed), highly dynamic characteristic systemic responses.

If there is a state (or set of states) Y_1 such that $\mathbf{R}_0 Y_1 = Y_1$, then the 'unitary kernel' Y_1 corresponds to the condition 'no response required', the set B_0.

Suppose pathology becomes manifest,

$$\mathbf{R}_0 \to \mathbf{R}_0 + \delta\mathbf{R} \equiv \hat{\mathbf{R}}_0,$$

so that, for example, some chronic excited state becomes the new 'unitary kernel', and

$$Y_1 \to \hat{Y}_1 \neq Y_1$$

$$\hat{\mathbf{R}}_0 \hat{Y}_1 = \hat{Y}_1.$$

This could represent chronic inflammation, autoimmune response, persistent depression/anxiety or HPA axis activation/burnout, and so on.

Medical intervention seeks to induce a sequence of therapeutic counterperturbations $\delta\mathbf{T}_k$ according to the pattern

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

$$[\hat{\mathbf{R}}_0 + \delta\mathbf{T}_1]\hat{Y}_1 = Y^1,$$

$$\hat{\mathbf{R}}_1 \equiv \hat{\mathbf{R}}_0 + \delta\mathbf{T}_1,$$

$$[\hat{\mathbf{R}}_1 + \delta\mathbf{T}_2]Y^1 = Y^2$$

...

(23)

so that, using an appropriate metric,

$$Y^j \to Y_1.$$

(24)

That is, the multilevel, highly dynamic, shifting, tunable system of global broadcasts –the mind/body system – as monitored by the GR, is driven to its original condition.

The condition $\hat{\mathbf{R}}_0 \to \mathbf{R}_0$ may or may not be met. That is, actual cure may not be possible, in which case palliation or control is the therapeutic aim.

The essential point is that the pathological state represented by $\hat{\mathbf{R}}_0$ and the sequence of inherently multiscale therapeutic interventions $\delta\mathbf{T}_k, k = 1, 2, ...$ are interactive and reflective, depending on the regression of the set of vectors Y^j to the desired state Y_1, much in the same spirit as Jerne's immunological idiotypic hall of mirrors.

The therapeutic problem revolves around minimizing the difference between Y^k and Y_1 over the course of treatment: that difference represents the inextricable convolution of 'treatment failure' with 'adverse reactions' to the course of treatment itself, and 'failure of compliance' attributed through social construction by provider to patient, i.e., failure of the therapeutic alliance.

It should be obvious that the treatment sequence $\delta\mathbf{T}_k$ represents a cognitive path of interventions having, in turn, a dual information source in the sense previously invoked.

Treatment may, then, interact in the usual Rate Distortion manner with patterns of structured pathogenic context that are, themselves, signals from

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

an embedding information source. Thus treatment failure, adverse reactions, and patient noncompliance will, of necessity, embody a distorted image of that context.

In sum, characteristic patterns of treatment failure, adverse reactions, and patient noncompliance reflecting collapse of the therapeutic alliance, will occur in virtually all therapeutic interventions – even those acting across scale – according to the manner in which structured psychosocial stress or other embedding noxious conditions are expressed as an image within the multiscale treatment process. This would most likely occur in a highly punctuated manner, depending in a quantitative way on the degree of coupling of the three-fold system of affected individual, patient/provider interaction, and multilevel treatment mode, with that stress or condition.

Given that the principal environment of humans is defined by interaction with other humans and their socioeconomic institutions, social effects in particular are likely to be dominant.

9 Discussion and conclusions

A tuning theorem variant of the Shannon Coding Theorem that expresses the no free lunch restriction allows construction of a broad spectrum of versions of Bernard Baars' global workspace/global broadcast model of animal consciousness that apply to many interacting 'low level' cognitive biological submodules, usually having much longer characteristic time constants than the 100ms of consciousness. Such generalized global broadcasts, via the giant component linking lower level 'unconscious' cognitive modules (and possibly inattentional blindness, via the no free lunch condition), emerges directly, and the effects of external signals and internal 'biological ruminations' can be incorporated through standard arguments leading to punctuated threshold detection.

The central conceit leading to this elaborate range of mechanisms is that the spandrel of crosstalk between 'unconscious', lower level, cognitive modules becomes a sufficient condition for evolutionary exaptation into the arch of global broadcasts through the information theory chain rule that implies it takes more metabolic free energy to prevent correlation than to allow it. Such generalization of neural consciousness, in terms of tunable, shifting global broadcasts, seems ubiquitous, as collective phenomena like wound healing and the many 'psychoneuroimmuno', gene expression, and other processes imply.

The parallel argument is, of course, that the similar necessary ubiquity of noise in the transmission of information has been exapted into mechanisms of stochastic resonance amplification at various scales.

It should be obvious that roughly similar evolutionary exaptations would be available under a broad variety of astrobiological circumstances, via the statistical regularities imposed by the asymptotic limit theorems of information theory.

In sum, since information is a form of free energy, a simple entropy gradient argument leads to an index theorem in which analytic solutions of an empirical

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

equation characterize different possible topological modes of cognitive processes that are linked at, and across, different scales and levels of hierarchy.

Thus, not only is the living state characterized by cognition at every scale and level of organization, but also by multiple, shifting, tunable, cooperative broadcasts analogous to, if more general than, consciousness, at and across those same structures.

This perspective, of an inherently multiscale and multilevel, interactive, and highly dynamic system, has profound implications for translational medicine. From the viewpoint of this study, the solution to the conundrum of figure 1 is to reconfigure interventions so as to encapsulate more than a single scale or level of organization. That is, it has now become necessary for the pharmaceutical industry – and its medical associates – to move beyond small molecule design to the principled construction of more comprehensive multifactorial or multiscale interventions designed to affect the interaction of complementary biochemical and information source networks, driving them from pathological to benign conformations, using externally-imposed and cleverly constructed 'large deviations' in the sense of equation (17), as expanded in the section on therapeutic intervention.

At the individual level this would appear to require seeking synergistic total strategies that act across levels of organization, rather than applying a sequence of scale-limited magic bullets, a difficult tectonic shift in scientific perspective, research, and practice not likely to prove popular with those embedded in current funding streams.

At the population level, where public policy can be most effective, the increasing expense of individual level interventions – even if the rate of decline of figure 1 can be mitigated by following a multiscale or multilevel perspective at the individual level – would seem to imply the necessity of again recognizing what has been known for the last two hundred years, that patterns of health and illness are determined by living and working conditions and the power relations between groups (e.g., Kleinman, Das and Lock, 1994; Wallace and Fullilove, 2008; Wallace and Wallace, 2010).

10 Acknowledgments

The author thanks D.N. Wallace, R.G. Wallace and B. Baars for useful discussions.

11 References

Ash, R., 1990, Information Theory, Dover, New York.

Atiyah, M., I. Singer, 1963, The index of elliptical operators on compact manifolds, Bulletin of the American Mathematical Society, 69:322-433.

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

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

Baars, B., 1988, A Cognitive Theory of Consciousness, Cambridge University Press, New York.

Baars, B., 2005, Global workspace theory of consciousness: toward a cognitive neuroscience of human experience, Progress in Brain Research, 150:45-53.

Baars, B., S. Franklin, 2003, How conscious experience and working memory interact, Trends in Cognitive Science, 7:166-172.

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

Bernstein Research, 2010, The Long View - R & D Producitivty.

Boccaletti, S., V. Latora, Y. Moreno, M. Chavez, D. Hwang, 2006, Complex networks: structure and dynamics, Physics Reports, 424:175-208.

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

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

Cover, T., J. Thomas, 2006, Elements of Information Theory, Second Edition, Wiley, New York.

Dembo, A., O. Zeitouni, 1998, Large Deviations and Applications, Springer, New York.

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

El Gamal, A., Y. Kim, 2010, Lecture Notes on Network Information Theory, ArXiv:1001.3404v4.

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

Erdos, P., A. Renyi, 1960, On the evolution of random graphs, Magyar Tud. Akad. Mat. Kutato Int. Kozl, 5:17-61.

Feynman, R., 2000, Lectures on Computation, Westview Press, New York.

Glazebrook, J.F., R. Wallace, 2009, Rate distortion manifolds as model spaces for cognitive information, Informatica, 33:309-346.

Gluckman, B., T. Netoff, E. Neel, W. Ditto, M. Spano, S. Schiff, 1996, Stochastic resonance in a neuronal network from mammalian brain, Physical Review Letters, 77:4098-4101.

Gould, S., R. Lewontin, 1979, The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme, Proceedings of the Royal Society of London, B, 205:581-598.

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

Hazewinkel, M., 2002, Encyclopedia of Mathematics, 'Index Formulas', Springer, NY.

Horrobin, D., 2003, Modern biomedical research: an internally self-consistent

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

universe with little contact with medical reality?, Nature Reviews: Drug Discovery, 2:151-154.

Kawaguchi, M., H. Mino, D. Durand, 2011, Stochastic resonance can enhance information transmission in neural networks, IEEE Transactions on Biomedical Engineering, 58:(7) DOI 10.1109/TBME.2011.2126571.

Khinchin, A., 1957, The Mathematical Foundations of Information Theory, Dover, New York.

Kleinman, A., V. Das, M. Lock, 1994, Social Suffering, University of California Press, Berkeley, CA.

Littman, B., R. Krishna, 2011, Translational Medicine and Drug Discovery, Cambridge University Press, Cambridge, UK.

Marincola, F., 2011, The trouble with translational medicine, Journal of Internal Medicine, 270:123-127.

Nunney, L., 1999, Lineage selection and the evolution of multistage carcinogenesis, Proceedings of the Royal Society B, 266:493-498.

Park, J., P. Neelakanta, 1996, Information-theoretic aspects of neural stochastic resonance, Complex Systems, 10:55-71.

Paul, S., D. Mytelka, C. Dunwiddie, C. Persinger, B. Munos, S. Lindborg, A. Schact, 2010, How to improve R & D productivity: the pharmaceutical industry's grand challenge, Nature Reviews: Drug Discovery, 9:201-214.

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

Richerson, P., R. Boyd, 2006, Not by genes alone: how culture transformed human evolution, University of Chicago Press, Chicago, IL.

Spenser, J., 2010, The giant component: the golden anniversary, Notices of the AMS, 57:720-724.

Thayer, J., R. Lane, 2000, A model of neurovisceral integration in emotion regulation and dysregulation, Journal of Affective Disorders, 61:201-216.

Wallace, R., 2000, 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., 2004, Comorbidity and anticomorbidity; autocognitive developmental disorders of psychosocial stress, Acta Biotheoretica, 52:71-93.

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

Wallace, R., 2007, Culture and inattentional blindenss: a global workspace perspective, Journal of Theoretical Biology, 245:378-390.

Wallace, R., 2008, Toward formal models of biologically inspired, highly parallel machine cognition, International Journal of Parallel, Emergent, and Distributed Systems, 23:367-408.

Wallace, R., 2009, Programming coevolutionary machines: the emerging conundrum, International Journal of Parallel, Emergent, and Distributed Systems, 24:443-453.

Wallace, R., 2010, Tunable epigenetic catalysis: programming real-time cognitive machines, International Journal of Parallel, Emergent, and Distributed

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

Systems, 25:209-222.

Wallace, R., 2010, Protein folding disorders: toward a basic biological paradigm, Journal of Theoretical Biology, 267:582-594.

Wallace, R., 2011a, Structure and dynamics of the 'protein folding code' inferred using Tlusty's topological rate distortion approach, BioSystems, 103:18-26.

Wallace, R., 2011b, Multifunction moonlighting and intrinsically disordered proteins: information catalysis, non-rigid molecule symmetries and the 'logic gate' spectrum, Comptes Rendus Chimie, 14:1117-1121.

Wallace, R., 2012, Extending Tlusty's rate distortion index theorem method to the glycome: Do even 'low level' biochemical phenomena require sophisticated cognitive paradigms?, BioSystems, 107:145-152.

Wallace, R., M. Fullilove, 2008, Collective Consciousness and its Discontents, Springer, New York.

Wallace, R., D. Wallace, 2009, Code, context, and epigenetic catalysis in gene expression, Transactions on Computational Systems Biology XI, LNBI 5750, 283-334.

Wallace, R., D. Wallace, 2010, Gene Expression and its Discontents: The social production of chronic disease, Springer, New York.

Ward, L., 2009, Physics of neural synchronation mediated by stochastic resonance, Contemporary Physics, 50:563-574.

Wehling, M., 2011, Drug development in the light of translational science: shine or shade? Drug Discovery Today, 16:1076-1083.

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

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

Zhu, R., A. Rebirio, D. Salahub, S. Kaufmann, 2007, Studying genetic regulatory networks at the molecular level: delayed reaction stochastic models, Journal of Theoretical Biology, 246:725-745.

12 Mathematical Appendix

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

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

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

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

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

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

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

and the conditional probability of Y given X as

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

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

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

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

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

(25)

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

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

(26)

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

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

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

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

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

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

(27)

Again, see Ash (1990), Cover and Thomas (2006) or Khinchin (1957) for details. The essential point is that if there is no uncertainty in X given the channel Y, then there is no loss of information through transmission. In general this will not be true, and herein lies the essence of the theory.

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

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

(28)

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

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

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

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

where $H(X)$ is the uncertainty of the stochastic variable defined above. Some consideration shows that $S(n)$ is much less than the total number of possible messages of length n. Thus, as $n \to \infty$, only a vanishingly small fraction of all possible messages is meaningful in this sense. This observation, after some

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012

considerable development, is what allows the Coding Theorem to work so well. In sum, the prescription is to encode messages in typical sequences, which are sent at very nearly the capacity of the channel. As the encoded messages become longer and longer, their maximum possible rate of transmission without error approaches channel capacity as a limit. Again, the standard references provide details.

This approach can be, in a sense, inverted to give a 'tuning theorem' variant of the coding theorem.

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

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

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

$$C^* \equiv \max_{P(Y),P(Y|X)} I(X|Y) \tag{29}$$

But

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

since

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

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

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

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

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

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

Calculating C^* requires maximizing the complicated expression

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

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

If $M = L$, then choose

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

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

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

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

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

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

We have examined the two limits in which either the distributions of $P(Y)$ or of $P(X)$ are kept fixed. The first provides the usual Shannon Coding Theorem, and the second a tuning theorem variant, i.e. a tunable, retina-like, Rate Distortion Manifold, in the sense of Glazebrook and Wallace (2009).

Nature Precedings : hdl:10101/npre.2012.6973.1 : Posted 7 Mar 2012