Version 2

Culture and generalized inattentional blindness

Rodrick Wallace, Ph.D.
The New York State Psychiatric Institute*

April 12, 2006

Abstract

A recent mathematical treatment of Baars' Global Workspace consciousness model, much in the spirit of Dretske's communication theory analysis of high level mental function, is used to study the effects of embedding cultural heritage on a generalized form of inattentional blindenss. Culture should express itself quite distinctly in this basic psychophysical phenomenon, acting across a variety of sensory and other modalities, because the limited syntactic and grammatical 'bandpass' of the topological rate distortion manifold characterizing conscious attention is itself strongly sculpted by the constraints of cultural context.

Key words cognition, consciousness, culture, global workspace, groupoid, inattentional blindness, information theory, orbit equivalence class, random network, rate distortion manifold.

INTRODUCTION

Inattentional blindness occurs when focus of attention on a single aspect of a complicated perceptual field precludes detection of others, which may be quite strong and normally expected to register on consciousness. Mack (1998) and Simons and Chabris (1998) provide background. The phenomenon was apparently well known in the early part of the 20th century, but its study languished thereafter, seemingly for many of the reasons that consciousness studies fell into disfavor for nearly a century.

Simons and Chabris (1999) detail a particularly spectacular example. A videotape was made of a basketball game between teams in white and black jerseys. Experimental subjects who viewed the tape were asked to keep silent mental counts of either the total number of passes made by one or the other of the teams, or separate counts of the number of bounce and areal passes. During the game, a figure in a full gorilla suit appears, faces the camera, beats its breast, and walks off the court. About one half of the experimental subjects completely failed to notice the Gorilla during the experiment. See Simons (2000) for an extended discussion.

Other case histories, involving an aircraft crew which became fixated on an unexpectedly flashing control panel light during a landing, or a man walking a railroad track while having a cell phone conversation, are less benign.

Dehaene and Changeux (2005) recently reported a neural network simulation of Baars' global workspace model of consciousness in which ignition of a coherent, spontaneous, excited state blocked external sensory processing, an observation they relate to inattentional blindness. Here, by contrast, we use a Dretske-style 'necessary conditions' analytic treatment of Baars' model to address the phenomenon, taking a modular network/information theory perspective which does not suffer the 'sufficiency indeterminacy' inherent to neural network simulations of high level mental phenomena (Krebs, 2005). This approach explicitly includes the potential influence of 'cultural factors' on inattentional blindness.

The necessity for such inclusion lies in the observations of Nisbett et al. (2001), and others, following the tradition of Markus and Kitayama (1991), regarding fundamental differences in perception between test subjects of 'Southeast Asian' and 'Western' cultural heritage across an broad realm of experiments. East Asian perspectives are characterized as 'holistic' and Western as 'analytic'. Nisbett et al. (2001) find:

(1) Social organization directs attention to some aspects of the perceptual field at the expense of others.

(2) What is attended to influences metaphysics.

(3) Metaphysics guides tacit epistemology, that is, beliefs about the nature of the world and causality.

(4) Epistemology dictates the development and application of some cognitive processes at the expense of others.

(5) Social organization can directly affect the plausibility of metaphysical assumptions, such as whether causality should be regarded as residing in the field vs. in the object.

(6) Social organization and social practice can directly influence the development and use of cognitive processes such as dialectical vs. logical ones.

Nisbett et al. (2001) conclude that tools of thought embody a culture's intellectual history, that tools have theories build into them, and that users accept these theories, albeit unknowingly, when they use these tools.

We begin by invoking a detailed mathematical model of consciousness in humans, which is elaborated up to and including the influences of embedding culture.

THE FORMAL THEORY

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

The Global Workspace consciousness model Bernard Baars' Global Workspace Theory (Baars, 1988, 2005) is rapidly becoming the de facto standard model of consciousness (e.g. Dehaene and Naccache, 2001; Dehaene and Changeaux, 2005). The central ideas are as follows (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.

Although this basic approach has been the focus of work by many researchers for two decades, consciousness studies has only recently, in the context of a deluge of empirical results from brain imaging experiments, begun digesting the perspective and preparing to move on.

We reiterate that currently popular agent-based and artificial neural network (ANN) treatments of cognition, consciousness and other higher order mental functions, to take Krebs' (2005) view, are little more than sufficiency arguments, in the same sense that a Fourier series expansion can be empirically fitted to nearly any function over a fixed interval without providing real understanding of the underlying structure. Necessary conditions, as Dretske argues (Dretske, 1981, 1988, 1993, 1994), give considerably more insight. Perhaps the most cogent example is the difference between the Ptolemaic and Copernican models of the solar system: one need not always expand in epicycles, but can seek the central motion. Dretske's perspective provides such centrality. Keplerian and Newtonian treatments, unfortunately, still lie ahead of us.

Wallace (2005a, b) has, in fact, addressed Baars' theme from Dretske's viewpoint, examining the necessary conditions which the asymptotic limit theorems of information theory impose on the Global Workspace. A central outcome of this work has been the incorporation, in a natural manner, of constraints on individual consciousness, i.e. what Baars calls contexts. Using information theory methods, extended by an obvious homology between information source uncertainty and free energy density, it is possible to formally account for the effects on individual consciousness of parallel physiological modules like the immune system, embedding structures like the local social network, and, most importantly, the all-encompassing cultural heritage which so uniquely marks human biology (e.g. Richerson and Boyd, 2004). This embedding evades the mereological fallacy which fatally bedevils brain-only theories of human consciousness (Bennett and Hacker, 2003).

Transfer of phase change approaches from statistical physics to information theory via the same homology generates the punctuated nature of accession to consciousness in a similarly natural manner. The necessary renormalization calculation focuses on a phase transition driven by variation in the average strength of nondisjunctive 'weak ties' (Granovetter, 1973) linking unconscious cognitive submodules. A second-order 'universality class tuning' allows for adaptation of conscious attention via 'rate distortion manifolds' which generalize the idea of a retina. Aversion of the Baars model emerges as an almost exact parallel to hierarchical regression, based, however, on the Shannon-McMillan rather than the Central Limit Theorem.

Wallace (2005b) recently proposed a somewhat different approach, using classic results from random and semirandom network theory (Erdos and Renyi, 1960; Albert and Barabasi, 2002; Newman, 2003) applied to a modular network of cognitive processors. The unconscious modular network structure of the brain is, of course, not random. However, in the spirit of the wag who said "all mathematical models are wrong, but some are useful", the method serves as the foundation of a different, but roughly parallel, treatment of the Global Workspace to that given in Wallace (2005a), and hence as another basis for a benchmark model against which empirical data can be compared.

The first step is to argue for the existence of a network of loosely linked cognitive unconscious modules, and to characterize each of them by the 'richness' of the canonical language – information source – associated with it. This is in some contrast to attempts to explicitly model neural structures themselves using network theory, e.g. the 'neuropercolation' approach of Kozma et al. (2004, 2005), which nonetheless uses many similar mathematical techniques. Here, rather, we look at the necessary conditions imposed by the asymptotic limits of information theory on any realization of a cognitive process, be it biological 'wetware', silicon dryware, or some direct or systems-level hybrid. All cognitive processes, in this formulation, are to be associated with a canonical 'dual information source' which will be constrained by the Rate Distortion Theorem, or, in the zero-error limit, the Shannon-McMillan Theorem. It is interactions between nodes in this abstractly defined network which will be of interest here, rather than whatever mechanism or biological system, or mixture of them, actually constitute the underlying cognitive modules.

The second step is to examine the conditions under which a giant component (GC) suddenly emerges as a kind of phase transition in a network of such linked cognitive modules, to determine how large that component is, and to define the relation between the size of the component and the richness of the cognitive language associated with it. This is the candidate for Baars' shifting Global Workspace of consciousness.

While Wallace (2005a) examines the effect of changing the average strength of nondisjunctive weak ties acting across linked unconscious modules, Wallace (2005b) focuses on changing the average *number* of such ties having a fixed strength, a complementary perspective whose extension via a kind of 'renormalization' leads to a far more general approach.

The third step, following Wallace (2005b), is to tune the threshold at which the giant component comes into being, and to tune vigilance, the threshold for accession to consciousness.

Wallace's (2005b) information theory modular network treatment can be enriched by introducing a groupoid formalism which is roughly similar to recent analyses of linked dynamic networks described by differential equation models (e.g. Stewart et al., 2003, Stewart, 2004; Weinstein, 1996; Connes, 1994). Internal and external linkages between information sources break the underlying groupoid symmetry, and introduce more structure, the global workspace and the effect of contexts, respectively. The analysis provides a foundation for further mathematical exploration of linked cognitive processes.

Cognition as 'language' Cognition is not consciousness. Most mental, and many physiological, functions, while cognitive in a formal sense, hardly ever become entrained into the Global Workspace of consciousness: one seldom is able to consciously regulate immune function, blood pressure, or the details of binocular tracking and bipedal motion, except to decide 'what shall I look at', 'where shall I walk'. Nonetheless, many cognitive processes, conscious or unconscious, appear intimately related to 'language', broadly speaking. The construction is fairly straightforward (Wallace, 2000, 2005a, b).

Atlan and Cohen (1998) and Cohen (2000) argue, in the context of immune cognition, that the essence of cognitive function involves comparison of a perceived signal with an internal, learned picture of the world, and then, upon that comparison, choice of one response from a much larger repertoire of possible responses.

Cognitive pattern recognition-and-response proceeds by an algorithmic combination of an incoming external sensory signal with an internal ongoing activity – incorporating the learned picture of the world – and triggering an appropriate action based on a decision that the pattern of sensory activity requires a response.

More formally, a pattern of sensory input is mixed in an unspecified but systematic algorithmic 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 internal and external signals. Wallace (2005a) provides two neural network examples.

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

While combining algorithm, the form of the nonlinear oscillator, and the details of grammar and syntax, are all unspecified in this model, the critical assumption which permits inference 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.

We 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 seems possible (Wallace, 2005a).

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

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

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

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

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

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 Khinchin (1957), Ash (1990), or Cover and Thomas (1991) for the standard details.

The cognitive modular network symmetry groupoid A formal equivalence class algebra can be constructed by choosing different origin points a_0 and defining equivalence by the existence of a high probability meaningful path connecting two points. 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 processes instantiated by either biological wetware, mechanical dryware, or their direct or systems-level hybrids.

This structure is a groupoid, in the sense of Weinstein (1996). States a_j, a_k in a set A are related by the groupoid morphism if and only if there exists a high probability grammatical path connecting them, and tuning across the various possible ways in which that can happen – the different cognitive languages – parametizes the set of equivalence relations and creates the groupoid. This assertion requires some development.

Note that not all possible pairs of states (a_j, a_k) can be connected by such a morphism, i.e. by a high probability, grammatical and syntactical cognitive path, but those that can define the groupoid element, a morphism $g = (a_j, a_k)$ having the 'natural' inverse $g^{-1} = (a_k, a_j)$. Given such a pairing, connection by a meaningful path, it is possible to define 'natural' end-point maps $\alpha(g) = a_j, \beta(g) = a_k$ from the set of morphisms G into A, and a formally associative product in the groupoid $g_1 g_2$ provided $\alpha(g_1 g_2) = \alpha(g_1), \beta(g_1 g_2) = \beta(g_2)$, and $\beta(g_1) = \alpha(g_2)$. Then the product is defined, and associative, i.e. $(g_1 g_2) g_3 = g_1(g_2 g_3)$.

In addition there are 'natural' left and right identity elements λ_g, ρ_g such that $\lambda_g g = g = g \rho_g$ whose characterization is left as an exercise (Weinstein, 1996).

An orbit of the groupoid G over A is an equivalence class for the relation $a_j \sim G a_k$ if and only if there is a groupoid element g with $\alpha(g) = a_j$ and $\beta(g) = a_k$.

The isotopy group of $a \in X$ consists of those g in G with $\alpha(g) = a = \beta(g)$.

In essence a groupoid is a category in which all morphisms have an inverse, here defined in terms of connection by a meaningful path of an information source dual to a cognitive process.

If G is any groupoid over A, the map $(\alpha, \beta) : G \to A \times A$ is a morphism from G to the pair groupoid of A. The image of (α, β) is the orbit equivalence relation $\sim G$, and the functional kernel is the union of the isotropy groups. If $f : X \to Y$ is a function, then the kernel of f, $ker(f) = [(x_1, x_2) \in X \times X : f(x_1) = f(x_2)]$ defines an equivalence relation.

As Weinstein (1996) points out, the morphism (α, β) suggests another way of looking at groupoids. A groupoid over A identifies not only which elements of A are equivalent to one another (isomorphic), but *it also parametizes the different ways (isomorphisms) in which two elements can be equivalent*, i.e. all possible information sources dual to some cognitive process. Given the information theoretic characterization of cognition presented above, this produces a full modular cognitive network in a highly natural manner.

The groupoid approach has become quite popular in the study of networks of coupled dynamical systems which can be defined by differential equation models, e.g. Stewart et al. (2003), Stewart (2004). Here we have outlined how to extend the technique to networks of interacting information sources which, in a dual sense, characterize cognitive processes, and cannot at all be described by the usual differential equation models. These latter, it seems, are much the spiritual offspring of 18th Century mechanical clock models. Cognitive and conscious processes in humans involve neither computers nor clocks, but remain constrained by the limit theorems of information theory, and these permit scientific inference on necessary conditions.

Internal forces breaking the symmetry groupoid The symmetry groupoid, as we have constructed it for unconscious cognitive submodules in 'information space', is parametized across that space by the possible ways in which states a_j, a_k can be 'equivalent', i.e. connected by a meaningful path of an information source dual to a cognitive process. These are different, and in this approximation, non-interacting unconscious cognitive processes. But symmetry groupoids, like symmetry groups, are made to be broken: by internal crosstalk akin to spin-orbit interactions within a symmetric atom, and by cross-talk with slower, external, information sources, akin to putting a symmetric atom in a powerful magnetic or electric field.

As to the first process, suppose that linkages can fleetingly occur between the ordinarily disjoint cognitive modules defined by the network groupoid. In the spirit of Wallace (2005a), this is represented by establishment of a nonzero mutual information measure between them: a cross-talk which breaks the strict groupoid symmetry developed above.

Wallace (2005a) describes this structure in terms of fixed magnitude disjunctive strong ties which give the equivalence class partitioning of modules, and nondisjunctive weak ties which link modules across the partition, and parametizes the overall structure by the average strength of the weak ties, to use Granovetter's (1973) term. By contrast the approach of Wallace (2005b), which we outline here, is to simply look at the average number of fixed-strength nondisjunctive links in a random topology. These are obviously the two analytically tractable limits of a much more complicated regime.

Since we know nothing about how the cross-talk connections can occur, we will – at first – assume they are random and construct a random graph in the classic Erdos/Renyi manner. Suppose there are M disjoint cognitive modules – M elements of the equivalence class algebra of languages dual to some cognitive process – which we now take to be the vertices of a possible graph.

For M very large, following Savante et al. (1993), when edges (defined by establishment of a fixed-strength mutual information measure between the graph vertices) are added at random to M initially disconnected vertices, a remarkable transition occurs when the number of edges becomes approximately $M/2$. Erdos and Renyi (1960) studied random graphs with M vertices and $(M/2)(1+\mu)$ edges as $M \to \infty$, and discovered that such graphs almost surely have the following properties (Molloy and Reed, 1995, 1998; Grimmett and Stacey, 1998; Luczak, 1990; Aiello et al., 200; Albert and Barabasi, 2002):

If $\mu < 0$, only small trees and 'unicyclic' components are present, where a unicyclic component is a tree with one additional edge; moreover, the size of the largest tree component is $(\mu - \ln(1+\mu))^{-1} + \mathcal{O}(\log\log n)$.

If $\mu = 0$, however, the largest component has size of order $M^{2/3}$.

And if $\mu > 0$, there is a unique 'giant component' (GC) whose size is of order M; in fact, the size of this component is asymptotically αM, where $\mu = -\alpha^{-1}[\ln(1-\alpha) - 1]$, which has an explicit solution for α in terms of the Lambert W-function. Thus, for example, a random graph with approximately $M\ln(2)$ edges will have a giant component containing $\approx M/2$ vertices.

Such a phase transition initiates a new, collective, cognitive phenomenon: the Global Workspace of consciousness, emergently defined by a set of cross-talk mutual information measures between interacting unconscious cognitive submodules. The source uncertainty, H, of the language dual to the collective cognitive process, which characterizes the richness of the cognitive language of the workspace, will grow as some monotonic function of the size of the GC, as more and more unconscious processes are incorporated into it. Wallace (2005b) provides details.

Others have taken similar network phase transition approaches to assemblies of neurons, e.g. 'neuropercolation' (Kozma et al., 2004, 2005), but their work has not focused explicitly on modular networks of cognitive processes, which may or may not be instantiated by neurons. Restricting analysis to such modular networks finesses much of the underlying conceptual difficulty, and permits use of the asymptotic limit theorems of information theory and the import of techniques from statistical physics, a matter we will discuss later.

External forces breaking the symmetry groupoid Just as a higher order information source, associated with the GC of a random or semirandom graph, can be constructed out of the interlinking of unconscious cognitive modules by mutual information, so too external information sources, for example in humans the cognitive immune and other physiological systems, and embedding sociocultural structures, can be represented as slower-acting information sources whose influence on the GC can be felt in a collective mutual information measure. For machines these would be the onion-like 'structured environment', to be viewed as among Baars' contexts (Baars, 1988, 2005; Baars and Franklin, 2003). The collective mutual information measure will, through the Joint Asymptotic Equipartition Theorem which generalizes the Shannon-McMillan Theorem, be the splitting criterion for high and low probability joint paths across the entire system.

The tool for this is network information theory (Cover and Thomas, 1991, p. 388). Given three interacting information sources, Y_1, Y_2, Z, the splitting criterion, taking Z as the 'external context', is given by

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

(2)

where $H(..|..)$ and $H(..,..,..)$ represent conditional and joint uncertainties (Khinchin, 1957; Ash, 1990; Cover and Thomas, 1991).

This generalizes to

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

(3)

If we assume the Global Workspace/Giant Component to involve a very rapidly shifting, and indeed highly tunable, dual information source X, embedding contextual cognitive modules like the immune system will have a set of significantly slower-responding sources $Y_j, j = 1..m$, and external social, cultural and other 'environmental' processes will be characterized by even more slowly-acting sources $Z_k, k = 1..n$. Mathematical induction on equation (3) gives a complicated expression for a mutual information splitting criterion which we write as

$$I(X|Y_1, .., Y_m|Z_1, .., Z_n).$$

(4)

This encompasses a fully interpenetrating 'biopsychosociocultural' structure for individual consciousness, one in which Baars' contexts act as important, but flexible, boundary conditions, defining the underlying topology available to the far more rapidly shifting global workspace (Wallace, 2005a, b).

This result does not commit the mereological fallacy which Bennett and Hacker (2003) impute to excessively neurocentric perspectives on consciousness in humans, that is, the mistake of imputing to a part of a system the characteristics which require functional entirety. The underlying concept of this fallacy should extend to machines interacting with their environments, and its baleful influence probably accounts for a significant part of AI's failure to deliver. See Wallace (2005a) for further discussion.

Punctuation phenomena As a number of researchers have noted, in one way or another, – see Wallace, (2005a) for discussion – equation (1),

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

is homologous to the thermodynamic limit in the definition of the free energy density of a physical system. This has the form

$$F(K) = \lim_{V \to \infty} \frac{\log[Z(K)]}{V}, \qquad (5)$$

where F is the free energy density, K the inverse temperature, V the system volume, and $Z(K)$ is the partition function defined by the system Hamiltonian.

Wallace (2005a) shows at some length how this homology permits the natural transfer of renormalization methods from statistical mechanics to information theory. In the spirit of the Large Deviations Program of applied probability theory, this produces phase transitions and analogs to evolutionary punctuation in systems characterized by piecewise, adiabatically stationary, ergodic information sources. These 'biological' phase changes appear to be ubiquitous in natural systems and can be expected to dominate machine behaviors as well, particularly those which seek to emulate biological paradigms. Wallace (2002) uses these arguments to explore the differences and similarities between evolutionary punctuation in genetic and learning plateaus in neural systems.

Renormalizing the giant component: the second order iteration 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. This we address by a parametized 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.

The fundamental conceptual trick at this point is to 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 '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 defining the global workspace of consciousness. That topology is, fundamentally, the basic tunable syntactic filter across the underlying modular symmetry groupoid, and variation in ω is only one aspect of a much more general topological shift. More detailed analysis is given below in terms of a topological rate distortion manifold.

Suppose the giant component 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 having a particular topological structure (Wallace, 2005b). Suppose that, over a sequence of 'times' the giant component can be characterized by a (possibly coarse-grained) path $x_n = \Omega_0, \Omega_1, ..., \Omega_{n-1}$ having significant serial correlations which, in fact, permit definition of an adiabatically, piecewise stationary, ergodic (APSE) information source in the sense of Wallace (2005a). Call that information source $\mathbf{X}$.

Suppose, again in the manner of Wallace (2005a), that a set of (external or else internal, systemic) signals impinging on consciousness, i.e. the giant component, is also highly structured and forms another APSE information source $\mathbf{Y}$ which interacts not only with the system of interest globally, but specifically with the tuning parameters of the giant component characterized by $\mathbf{X}$. $\mathbf{Y}$ is necessarily associated with a set of paths y_n.

Pair the two sets of paths into a joint path $z_n \equiv (x_n, y_n)$, and invoke some inverse coupling parameter, K, between the information sources and their paths. By the arguments of Wallace (2005a) this leads to phase transition punctuation of $I[K]$, the mutual information between $\mathbf{X}$ and $\mathbf{Y}$, under either the Joint Asymptotic Equipartition Theorem, or, given a distortion measure, under the Rate Distortion Theorem.

$I[K]$ is a splitting criterion between high and low probability pairs of paths, and partakes of the homology with free energy density described in Wallace (2005a). Attentional focusing then itself becomes a punctuated event in response to increasing linkage between the organism or device and an external structured signal, or some particular system of internal events. This iterated argument parallels the extension of the General Linear Model into the Hierarchical Linear Model of regression theory.

Call this the Hierarchical Cognitive Model (HCM).

The HCM version of Baars' global workspace model, as we have constructed it, stands in some contrast to other current work.

Tononi (2004), for example, takes a 'complexity' perspective on consciousness, in which he averages mutual information across all possible bipartitions of the thalamocortical system, and, essentially, demands an 'infomax' clustering solution. Other clustering statistics, however, may serve as well or better, as in generating phylogenetic trees, and the method does not seem to produce conscious punctuation in any natural manner.

Dehaene and Changeux (2005) take an explicit Baars global workspace perspective on consciousness, but use an elaborate neural network simulation to generate a phenomenon analogous to inattentional blindness. While their model does indeed display the expected punctuated behaviors, as noted above, Krebs (2005) unsparingly labels such constructions with the phrase 'neurological possibility does not imply neurological plausibility', suggesting that the method does little more than fit a kind of Fourier series construction to high level mental processes.

Here we have attempted a central motion model of consciousness, focusing on modular networks defined by function rather than by structure.

Cognitive quasi-thermodynamics A fundamental homology between the information source uncertainty dual to a cognitive process and the free energy density of a physical system arises, in part, from the formal similarity between their definitions in the asymptotic limit. Information source uncertainty can be defined as in equation (1). This is quite analogous to the free energy density of a physical system, equation (5).

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

Feynman explores in some detail Bennett's microscopic machine designed to extract useful work from a transmitted message. The essential argument is that computing, in any form, takes work, the more complicated a cognitive process, measured by its information source uncertainty, the greater its energy consumption, and our ability to provide energy to the brain is limited. Inattentional blindness emerges as an inevitable thermodynamic limit on processing capacity in a topologically-fixed global workspace, i.e. one which has been strongly configured about a particular task (Wallace, 2006).

Understanding the time dynamics of cognitive systems away from phase transition critical points requires a phenomenology similar to the Onsager relations of nonequilibrium thermodynamics. If the dual source uncertainty of a cognitive process is parametized by some vector of quantities $\mathbf{K} \equiv (K_1, ..., K_m)$, then, in analogy with nonequilibrium thermodynamics, gradients in the K_j of the *disorder*, defined as

$$S \equiv H(\mathbf{K}) - \sum_{j=1}^{m} K_j \partial H / \partial K_j \quad (6)$$

become of central interest.

Equation (6) is similar to the definition of entropy in terms of the free energy density of a physical system, as suggested by the homology between free energy density and information source uncertainty described above.

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

$$dK_j/dt = \sum_i L_{j,i} \partial S / \partial K_i, \quad (7)$$

where the $L_{j,i}$ are, in first order, constants reflecting the nature of the underlying cognitive phenomena. The L-matrix is to be viewed empirically, in the same spirit as the slope and intercept of a regression model, and may have structure far different than familiar from more simple chemical or physical processes. The $\partial S/\partial K$ are analogous to thermodynamic forces in a chemical system, and may be subject to override by external physiological driving mechanisms (Wallace, 2005c).

Imposing a metric for different cognitive dual languages parametized by $\mathbf{K}$ leads quickly into the rich structures of Riemannian, or even Finsler, geometries (Wallace, 2005c).

One can apply this formalism to the example of the giant component, with the information source uncertainty/channel capacity taken as directly proportional to the component's size, which increases monotonically with the average number of (renormalized) linkages, a, after the critical point. $H(a)$ then rises to some asymptotic limit.

As the system rides up with increasing a, $H(a)$ increases against the 'force' defined by $-dS/da$. Raising the cognitive capacity of the giant component, making it larger, requires energy, and is done against a particular kind of opposition. Beyond a certain point, the system just runs out of steam. Altering the topology of the network, no longer focusing on a particular demanding task, would allow detection of crosstalk signals from other submodules, as would the intrusion of a signal above the renormalization limit ω.

We propose, then, that the manner in which the system 'runs out of steam' involves a maxed-out, fixed topology for the giant component of consciousness. As argued above, the renormalization parameter ω then becomes an information/energy bottleneck. To keep the giant component at optimum function in its particular topology, i.e. focused on a particular task involving a necessary set of interacting cognitive submodules, a relatively high limit must be placed on the magnitude of a mutual information signal which can intrude into consciousness.

Consciousness is tunable, and signals outside the chosen 'syntactical/grammatical bandpass' are often simply not strong enough to be detected, accounting for the phenomena of inattentional blindness (Wallace, 2006). This basic focus mechanism can be modeled in far more detail.

Focusing the mind's eye: the simplest rate distortion manifold The second order iteration above – analogous to expanding the General Linear Model to the Hierarchical Linear Model – which involved paths in parameter space, can itself be significantly extended. This produces a generalized tunable retina model which can be interpreted as a 'Rate Distortion manifold', a concept which further opens the way for import of a vast array of tools from geometry and topology.

Suppose, now, that threshold behavior in conscious reaction requires some elaborate system of nonlinear relationships defining a set of renormalization parameters $\Omega_k \equiv \omega_1^k, ..., \omega_m^k$. The critical assumption is that there is a tunable 'zero order state,' and that changes about that state are, in first order, relatively small, although their effects on punctuated process may not be at all small. Thus, given an initial m-dimensional vector Ω_k, the parameter vector at time $k+1$, Ω_{k+1}, can, in first order, be written as

$$\Omega_{k+1} \approx \mathbf{R}_{k+1}\Omega_k, \tag{8}$$

where $\mathbf{R}_{t+1}$ is an $m \times m$ matrix, having m^2 components.

If the initial parameter vector at time $k = 0$ is Ω_0, then at time k

$$\Omega_k = \mathbf{R}_k\mathbf{R}_{k-1}...\mathbf{R}_1\Omega_0. \tag{9}$$

The interesting correlates of consciousness are, in this development, *now represented by an information-theoretic path defined by the sequence of operators* $\mathbf{R}_k$, each member having m^2 components. The grammar and syntax of the path defined by these operators is associated with a dual information source, in the usual manner.

The effect of an information source of external signals, $\mathbf{Y}$, is now seen in terms of more complex joint paths in Y and R-space whose behavior is, again, governed by a mutual information splitting criterion according to the JAEPT.

The complex sequence in m^2-dimensional R-space has, by this construction, been projected down onto a parallel path, the smaller set of m-dimensional ω-parameter vectors $\Omega_0, ..., \Omega_k$.

If the punctuated tuning of consciousness is now characterized by a 'higher' dual information source – an embedding generalized language – so that the paths of the operators $\mathbf{R}_k$ are autocorrelated, then the autocorrelated paths in Ω_k represent output of a parallel information source which is, given Rate Distortion limitations, apparently a grossly simplified, and hence highly distorted, picture of the 'higher' conscious process represented by the R-operators, having m as opposed to $m \times m$ components.

High levels of distortion may not necessarily be the case for such a structure, *provided it is properly tuned to the incoming signal.* If it is inappropriately tuned, however, then distortion may be extraordinary.

Let us examine a single iteration in more detail, assuming now there is a (tunable) zero reference state, $\mathbf{R}_0$, for the sequence of operators $\mathbf{R}_k$, and that

$$\Omega_{k+1} = (\mathbf{R}_0 + \delta\mathbf{R}_{k+1})\Omega_k, \tag{10}$$

where $\delta\mathbf{R}_k$ is 'small' in some sense compared to $\mathbf{R}_0$.

Note that in this analysis the operators $\mathbf{R}_k$ are, implicitly, determined by linear regression. We thus can invoke a quasi-diagonalization in terms of $\mathbf{R}_0$. Let $\mathbf{Q}$ be the matrix of eigenvectors which Jordan-block-diagonalizes $\mathbf{R}_0$. Then

$$\mathbf{Q}\Omega_{k+1} = (\mathbf{Q}\mathbf{R}_0\mathbf{Q}^{-1} + \mathbf{Q}\delta\mathbf{R}_{k+1}\mathbf{Q}^{-1})\mathbf{Q}\Omega_k. \tag{11}$$

If $\mathbf{Q}\Omega_k$ is an eigenvector of $\mathbf{R}_0$, say Y_j with eigenvalue λ_j, it is possible to rewrite this equation as a generalized spectral expansion

$$Y_{k+1} = (\mathbf{J} + \delta\mathbf{J}_{k+1})Y_j \equiv \lambda_j Y_j + \delta Y_{k+1}$$
$$= \lambda_j Y_j + \sum_{i=1}^{n} a_i Y_i. \tag{12}$$

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

$$|a_i| \ll |\lambda_j|, |a_{i+1}| \ll |a_i|. \tag{13}$$

The point is that, provided $\mathbf{R}_0$ has been tuned so that this condition is true, the first few terms in the spectrum of this iteration of the eigenstate will contain most of the essential information about $\delta \mathbf{R}_{k+1}$. This appears quite similar to the detection of color in the retina, where three overlapping nonorthogonal eigenmodes of response are sufficient to characterize a huge plethora of color sensation. Here, if such a tuned spectral expansion is possible, a very small number of observed eigenmodes would suffice to permit identification of a vast range of changes, so that the rate-distortion constraints become quite modest. That is, there will not be much distortion in the reduction from paths in R-space to paths in Ω-space. Inappropriate tuning, however, can produce very marked distortion, even inattentional blindness.

Reflection suggests that, if consciousness indeed has something like a grammatically and syntactically-tunable retina, then appropriately chosen observable correlates of consciousness may, at a particular time and under particular circumstances, actually provide very good local characterization of conscious process. Large-scale global processes are, like hyperfocal tuning, another matter.

Note that Rate Distortion Manifolds can be quite formally described using standard techniques from topological manifold theory (Glazebrook, 2005). The essential point is that a rate distortion manifold is a topological structure which constrains the 'stream of consciousness' much the way a riverbank constrains the flow of the river it contains. This is a fundamental insight.

DISCUSSION AND CONCLUSIONS

The simple groupoid defined by underlying cognitive modular structure can be broken by intrusion of (rapid) crosstalk within it, and by the imposition of (slower) crosstalk from without it. The former, if strong enough, can initiate a topologically-determined giant component global workspace of consciousness, in a punctuated manner, while the latter deforms the underlying topology of the entire system, limiting what paths can actually be traversed by consciousness, the 'torus and sphere' argument of Wallace (2005a). Broken symmetry creates richer structure in systems characterized by groupoids, just as it does for those characterized by groups. Conscious attention acts through a Rate Distortion manifold, a kind of retina-like filter for grammatical and syntactical 'meaningful' paths, which affects what can be brought to consciousness. Signals outside the tunable syntax/grammar bandpass of this manifold are subject to lessened probability of punctuated conscious detection: generalized inattentional blindness. Culture will, according to this model, profoundly affect the phenomenon by imposing additional topological constraints defining the 'surface' along which consciousness can (and cannot) glide.

Glazebrook (2005) has suggested that, lurking in the background of this basic construction, is what Brown has called the groupoid atlas, i.e. an extension of topological manifold theory to groupoid mappings. Formalizing this insight should prove to be an arduous enterprise. Also lurking is identification and exploration of the 'natural' groupoid convolution algebra which so often marks these structures (e.g. Weinstein, 1996; Connes, 1994).

Consideration suggests, in fact, that a path may be 'meaningful' according to the groupoid parametization of all possible dual information sources, and that tuning is done across that parametization via a rate distortion manifold.

Baars' global workspace of consciousness is, in effect, a movable bucket of limited capacity. If it is already filled up by attention to a particular task, droplets from other tasks will likely overflow, and may not be consciously perceived.

If one is, then, intensely focused on watching a basketball game and counting passes, requiring a very particular fixed (but highly tunable) cognitive topology, a gorilla beating its chest may simply not be a strong enough syntactically/grammatically correct signal to intrude on consciousness. On the other hand, falling off one's chair, a hotfoot, or a particularly sharp comment from one's significant other, might prove intrusive enough – above the tunable syntax limit characterized by ω – to permit detection in the given topological configuration, or else powerful enough to shift conscious topology altogether, i.e. to retune the operator $\mathbf{R}_0$ in the rate distortion manifold argument above. Short of that, there remains a significant probability that signals outside the range of the grammar/syntax filter of conscious attention will not be meaningful and will simply not be detected: inattentional blindness.

Implicit, however, are the constraints imposed by embedding cultural heritage, which may further limit the properties of $\mathbf{R}_0$.

Clearly the phenomenon should not be restricted to the visual system, but, in one form or another, is likely to be ubiquitous across conscious experience, all of which should display particular cultural characteristics.

The mathematical ecologist E.C. Pieou (1977, p.106) describes the utility of mathematical models of complex ecosystem phenomena as follows:

> "...[Mathematical models] are easy to devise; even though the assumptions of which they are constructed may be hard to justify, the magic phrase 'let us assume that...' overrides objections temporarily. One is then confronted with a much harder task: How is such a model to be tested? The correspondence between a model's predictions and observed events is sometimes gratifyingly close but this cannot be taken to imply the model's simplifying assumptions are reasonable in the sense that neglected complications are indeed [always] negligible in their effects...
>
> In my opinion [in spite of these serious dangers] the usefulness of models is great... [however] it consists *not in answering questions but in raising them.* Models can be used to inspire new field investigations and these are the only source of new knowledge as opposed to new speculation."

Extending that perspective slightly, the model we have presented, like a regression analysis, would perhaps provide the most scientific value through its violation, i.e. new science is often found in the residuals (Kepler and Newton extending Copernicus).

Thus, for example, variations in the forms of inattentional blindness across vision, touch, taste, hearing, and their interactions, should give deeper understanding of consciousness. A second empirical implication is that the various forms of inattentional blindness are likely subject to elaborate regulation: too much distractibility while hunting, like too much fixation on one's prey while one is, in turn, being hunted, could be rapidly fatal.

Attentional focus is necessary for consciousness to be effective in learning new, or successfully carrying out old, skills. Too much focus, however, leads to inattentional blindness, which can be dangerous. Here we have attempted to reexpress this trade-off in terms of a syntactical/grammatical version of conventional signal theory, i.e. as a 'tuned meaningful path' form of the classic balance between sensitivity and selectivity.

The final, and perhaps central, empirical implication is that 'Western', 'East Asian', and other cultural heritages should impose observable differences in the manifestations of generalized inattentional blindness.

These speculations, in particular the latter, are all subject to explicit empirical test.

References

Aiello W., F. Chung, and L. Lu, 2000, A random graph model for massive graphs, in *Proceedings of the 32nd Annual ACM Symposium on the Theory of Computing*.

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

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

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

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

Baars B., and S. Franklin, 2003, How conscious experience and working memory interact, *Trends in Cognitive Science*, doi:10.1016/S1364-6613(03)00056-1.

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

Bennett M., and P. Hacker, 2003 *Philosophical Foundations of Neuroscience*, Blackwell Publishing, London.

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

Connes A., 1994, *Noncommutative Geometry*, Academic Press, San Diego.

Corless R., G. Gonnet, D. Hare, D. Jeffrey, and D. Knuth, 1996, On the Lambert W function, *Advances in Computational Mathematics*, 4:329-359.

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

Dehaene S., and L. Naccache, 2001, Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework, *Cognition*, 79:1-37.

Dehaene S., and J. Changeux, 2005, Ongoing spontaneous activity controls access to consciousness: a neuronal model for inattentional blindness, *PLOS Biology*, 3:e141.

Dretske F., 1981, *Knowledge and the Flow of Information*, MIT Press, Cambridge, MA.

Dretske F., 1988, *Explaining Behavior*, MIT Press, Cambridge, MA.

Dretske, F., 1993, Mental events as structuring causes of behavior, in *Mental Causation* (ed. by A. Mele and J. Heil), pp. 121-136, Oxford University Press.

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

Erdos P., and A. Renyi, 1960, On the evolution of random graphs, reprinted in *The Art of Counting*, 1973, 574-618 and in *Selected Papers of Alfred Renyi*, 1976, 482-525.

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

Freeman, W., 2003, The wave packet: an action potential of the 21st Century, *Journal of Integrative Neurosciences*, 2:3-30.

Glazebrook, J., 2005, Personal communication.

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

Grimmett G., and A. Stacey, 1998, Critical probabilities for site and bond percolation models, *The Annals of Probability*, 4:1788-1812.

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

Kozma R., M. Puljic, P. Balister, B. Bollobas, and W. Freeman, 2004, Neuropercolation: a random cellular automata approach to spatio-temporal neurodynamics, *Lecture Notes in Computer Science*, 3305:435-443.

Kozma R., M. Puljic, P. Balister, and B. Bollobas, 2005, Phase transitions in the neuropercolation model of neural populations with mixed local and non-local interactions, *Biological Cybernetics*, 92:367-379.

Krebs, P., 2005, Models of cognition: neurological possibility does not indicate neurological plausibility, in Bara, B., L. Barsalou, and M. Bucciarelli (eds.), *Proceedings of CogSci 2005*, pp. 1184-1189, Stresa, Italy. Available at http://cogprints.org/4498/.

Luczak T., 1990, *Random Structures and Algorithms*, 1:287.

Mack A., 1998, *Inattentional Blindness*, MIT Press, Cambridge, MA.

Markus H., and S. Kitayama, 1991, Culture and the self-implications for cognition, emotion, and motivation, *Psychological Review*, 98:224-253.

Molloy M., and B. Reed, 1995, A critical point for random graphs with a given degree sequence, *Random Structures and Algorithms*, 6:161-179.

Molloy M., and B. Reed, 1998, The size of the giant component of a random graph with a given degree sequence, *Combinatorics, Probability, and Computing*, 7:295-305.

Newman M., S. Strogatz, and D. Watts, 2001, Random graphs with arbitrary degree distributions and their applications, *Physical Review E*, 64:026118, 1-17.

Newman M., 2003, Properties of highly clustered networks, arXiv:cond-mat/0303183v1.

Nisbett R., K. Peng, C. Incheol, and A. Norenzayan, 2001, Culture and systems of thought: holistic vs. analytic cognition, *Psychological Review*, 108:291-310.

Pielou, E., 1977, *Mathematical Ecology*, John Wiley and Sons, New York.

Richerson P., and R. Boyd, 2004, *Not by Genes Alone: How Culture Transformed Human Evolution*, Chicago University Press.

Savante J., D. Knuth, T. Luczak, and B. Pittel, 1993, The birth of the giant component, arXiv:math.PR/9310236v1.

Shannon C., and W. Weaver, 1949, *The Mathematical Theory of Communmication*, University of Illinois Press, Chicago, IL.

Simons, D., and C. Chabris, 1999, Gorillas in our midst: sustained inattentional blindness for dynamic events, *Perception*, 28:1059-1074.

Simons D., 2000, Attentional capture and inattentional blindness, *Trends in Cognitive Sciences*, 4:147-155.

Stewart I., M. Golubitsky, and M. Pivato, 2003, Symmetry groupoids and patterns of synchrony in coupled cell networks, *SIAM Journal of Applied Dynamical Systems*, 2:609-646.

Stewart I., 2004, Networking opportunity, *Nature*, 427:601-604.

Tononi G., 2004, An information integration theory of consciousness, *BMC Neuroscience*, 5:42.

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

Wallace R., 2005b, A global workspace perspective on mental disorders, *Theoretical Biology and Medical Modelling*, 2:49, http://www.tbiomed.com/content/2/1/49.

Wallace R., 2005c, The sleep cycle: a mathematical analysis from a global workspace perspective, http://cogprints.org/4517/

Wallace R., 2006, Generalized inattentional blindness from a global workspace perspective, http://cogprints.org/4719/

Weinstein A., 1996, Groupoids: unifying internal and external symmetry, *Notices of the American Mathematical Association*, 43:744-752.