Version 4

Generalized inattentional blindness from a Global Workspace perspective

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

February 6, 2006

Abstract

We apply Baars' Global Workspace model of consciousness to the phenomenon of inattentional blindness, using the groupoid network method of Stewart et al. to explore modular structures defined by information measures associated with cognitive processes. Internal cross-talk between such processes breaks the fundamental groupoid symmetry, and, if sufficiently strong, creates, in a highly punctuated manner, a linked, shifting, giant component which instantiates the global workspace of consciousness. Embedding, exterior, information sources act as a kind of external field which breaks the groupoid symmetry in a somewhat different manner, defining the slowly-acting contexts of Baars theory and providing topological constraints for the manifestations of consciousness. This analysis significantly extends recent mathematical treatments of the global workspace, and identifies a shifting, topologically-determined syntactical and grammatical 'bottleneck' as a tunable Rate Distortion manifold which constrains what sensory or other signals can be brought to conscious attention, typically in a punctuated manner. Sensations outside the limits of that filter's syntactically and grammatically tuned 'bandpass' have lower probability of detection, regardless of their structure, accounting for inattentional blindness. The fundamental empirical implication of this model is that virtually every manifestation of conscious attention should, in some measure, exhibit a form of inattentional blindness. Thus 'residuals' from this model across systems, i.e. variations in time, place, and manner, as it were, can provide the basis for significant new science, in the same sense that residuals from a regression model often indicate novel directions for research. The theory, based on necessary conditions imposed by the asymptotic limit theorems of communication theory, does not suffer the 'sufficiency indeterminacy' which Krebs finds inherent to neural network simulations of complicated mental processes.

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

*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.

Introduction

Inattentional blindness is a remarkable phenomenon which occurs when concentrated focus of attention on a single aspect of a complicated perceptual field actually precludes detection of others, which may themselves be very strong and normally expected to register on consciousness. The excellent review by Mack (1998) provides background and examples, as does Simons and Chabris (1998). 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) explore in some detail a particular experiment which is worth recounting. 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 more discussion.

Other examples, for instance involving an aircraft crew which became fixated on an unexpectedly flashing control panel light during a landing, or of a man walking a railroad track while having a cell phone conversation, had less benign outcomes.

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 we will attempt to similarly use the Baars global workspace model of consciousness to address the phenomenon, but from 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). The path is not without mathematical difficulty. We begin with a brief recapitulation of the Baars model.

The Global Workspace consciousness model

Bernard Baars' Global Workspace Theory (Baars, 1988) has become the de facto standard model of consciousness in humans (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 have only recently, in the context of a deluge of empirical results from brain imaging experiments, come to the point of digesting the perspective and moving on.

Currently popular agent-based and artificial neural network (ANN) treatments of cognition, consciousness and other higher order mental functions, in 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 in his application of information theory to mental function (Dretske, 1981, 1988, 1993, 1994), give considerably more insight. Perhaps the most cogent example is the difference between the Ptolemaic and Newtonian pictures of the solar system: one need not always expand in epicycles.

Wallace (2005a) has addressed Baars' theme from Dretske's perspective, 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. 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 approach 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. We significantly extend that work here, explicitly invoking the groupoid formalism of Stewart et al. (2003), Weinstein (1996) and Connes (1994), as well as a topologically-driven 'renormalization' which provides the syntactic bottleneck.

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 examines 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' will lead to the results of this paper.

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. We shall adapt that approach in our treatment of inattentional blindness, without studying, however, the full influence of Baars' contexts, a matter treated briefly in the Appendix.

We particularly extend Wallace's (2005b) information theory modular network treatment 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). Internal and external linkages between information sources break the underlying groupoid symmetry, and introduce richer 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).

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 interest are paths x which trigger pattern recognition-and-response exactly once. That is, given a fixed initial state a_0, such that $h(a_0) \in B_0$, we examine all possible subsequent paths x beginning with a_0 and leading exactly once to the event $h(x) \in B_1$. Thus $h(a_0, ..., a_j) \in B_0$ for all $j < m$, but $h(a_0, ..., a_m) \in B_1$. Wallace (2005a) examines the possibility of more complicated schemes as well, and concludes that they, like the use of varying forms of distortion measures in the Rate Distortion Theorem, all lead to similar results.

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 having $h(a_0) \in B_0$ 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

$$H \equiv \lim_{n \to \infty} \frac{\log[N(n)]}{n} \quad (1)$$

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).

Remember 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.

This is not a review article, and is not the place for an extensive tutorial discussion of general topics, nor for analysis of roughly parallel results (e.g. Freeman, 2003), but one particular point seems worth emphasizing. Quite often the statement is made that information theory is fundamentally defective because it has nothing to do with semantic meaning, or as Shannon puts it (Shannon, 1949)

> "[The] semantic aspects of communication are irrelevant to the engineering problem [addressed by information theory]."

A number of authors have misinterpreted this statement and attempted 'semantic' reformulations of Shannon's theory. This is unnecessary. The central core of Shannon's results lies in the asymptotic limit theorems which bear his name, extended by the Rate Distortion Theorem. These are, in effect, infinite, i.e. thermodynamic, limits. 'Semantics' applies primarily to short sequences of symbols, which are simply outside the purview of information theory. Within its realm of application, however, – very long path sequences – information source uncertainty serves as the splitting criterion between small high, and very large low, probability sets of those paths. In the limit of infinite sequences, except for a set of measure zero, virtually all syntactically and grammatically correct 'meaningful' paths will be contained within the high probability set. This result, with elaboration and the notable import of technologies from topological manifold and groupoid theory, provides the necessary conditions needed for a fairly complete treatment of cognition and consciousness. Traditional semantics simply becomes irrelevant under the asymptotic conditions of the information theory limit theorems, in much the same sense that parametric statistics relies on the asymptotic Central Limit Theorem to analyze very long sums of random variables.

We work, then, entirely at what amounts to a thermodynamic limit.

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 left and right identity elements λ_g, ρ_g such that $\lambda_g g = g = g\rho_g$ (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.

Breaking the symmetry groupoid: the giant component of consciousness

Symmetry groups are made to be broken: think of a (spherically symmetric) hydrogen atom. Neglecting electron spin, the emission spectrum of such a system is easily calculated using elementary quantum mechanics. Introduction of electron spin, however, gives internal 'spin-orbit interactions' which split many of those elementary spectral levels, generating a more complex structure.

Next, consider the atom in a powerful magnetic field. That field generates a 'natural direction' which further splits the spectral energy levels according to their inherent angular momenta, which are otherwise degenerate in the absence of that field, and further affects the atomic emission spectrum, creating a much richer pattern.

In a very similar sense, symmetry groupoids defining modular networks are also made to be broken, by internal cross-linkages with each other, and by linkage with external information sources. The first process, which can be very rapid, can generate consciousness in a punctuated manner, the second, much slower, is treated in the Appendix, and determines the influence of Baars' contexts.

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 $M \rightarrow \infty$, and discovered that such graphs almost surely have the following properties:

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 $\mu = -\alpha^{-1}\ln(1-\alpha) - 1$. Thus, for example, a random graph with approximately $M \ln(2)$ edges will have a giant component containing $\approx M/2$ vertices.

More explicitly, as Corless et al. (1996) discuss, when a graph with M vertices has $m = (1/2)aM$ edges chosen at random, for $a > 1$ it almost surely has a giant connected component having approximately gM vertices, with

$$g(a) = 1 + W(-a\exp(-a))/a, \quad (2)$$

where W is the Lambert-W function defined implicitly by the relation

$$W(x)\exp(W(x)) = x. \quad (3)$$

Figure 1 shows $g(a)$, displaying the sharp phase transition at $a = 1$.

Such a phase transition initiates a new, collective, cognitive phenomenon: the Global Workspace 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 defines the richness of the cognitive language of the workspace, will grow as some function of g, as more and more unconscious processes are incorporated into it. Wallace (2005a) examines what, in effect, are the functional forms $H \propto \exp(\alpha g), \alpha \ln[1/(1-g)]$, and $(1/(1-g))^{\delta}$, letting $R = 1/1 - g$ define a 'characteristic length' in the renormalization scheme. While these all have explicit solutions for the renormalization calculation (mostly in terms of the Lambert-W function), other, less tractable, expressions are certainly plausible, for example $H \propto g^{\gamma}, \gamma > 0$, γ real.

Given a particular $H(g)$, the approach of Wallace (2005a) involves adjusting universality class parameters of the phase transition.

By contrast, in the new class of models introduced by Wallace (2005b), the degree of clustering of the graph of cognitive modules is, itself, tunable, producing a variable threshold for consciousness: a topological shift, which should be observable from brain-imaging studies. Second order iteration (Wallace, 2005b) leads to an analog of the hierarchical cognitive model of Wallace (2005a).

Renormalizing the giant component: the second order iteration

The random network calculation 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 detection limit, and depends critically on the instantaneous topology of the giant component defining the global workspace of consciousness.

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).

In the Appendix we outline the effects of embedding contexts on consciousness via network information theory. These contexts, like a magnetic field applied to a hydrogen atom, vary much more slowly that the fleeting, shifting, global workspace of consciousness, which is more analogous to atomic spin-orbit interaction.

The next step in the development is to examine more closely the 'energy' arguments associated with cognitive phenomena, as we have described them.

The quasi-thermodynamics of cognition

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), i.e.

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

where $N(n)$ is the number of 'meaningful' paths of length n. This is quite analogous to the free energy density of a physical system,

$$F = \lim_{V\to\infty} \frac{\log[Z]}{V}$$

(4)

where V is the system volume and Z is the 'partition function' defined by the system's Hamiltonian energy function. 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.

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

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

(5)

become of central interest.

Equation (5) 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

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

(6)

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, as Wallace (2005c) argues, may be subject to override by external physiological driving mechanisms.

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).

Figure 2 treats the example of the giant component, with the information source uncertainty/channel capacity taken as directly proportional to the component's size:

$$H(a) \propto g(a) = 1 + W(-ae^{-a})/a,$$

where, as before, W is the Lambert W-function. The top trace is H and the lower one is the 'thermodynamic force' defined by $-dS/da$. As the system rides up the top curve from left to right, H 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 cross-talk 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. This can be modeled in 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

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

(7)

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

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

(8)

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, $\mathbf{Y}$ in the Appendix, 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.

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

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

(9)

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

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

(10)

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

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

(11)

$\mathbf{J}$ 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

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

(12)

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 non-orthogonal 'eigenmodes' of response are sufficient to characterize a huge plethora of color sensation. Here, if such a 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.

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, of course, another matter.

Note that Rate Distortion Manifolds can be quite formally described using standard techniques from topological manifold theory (Glazebrook, 2005), a matter left as an exercise for the reader.

Discussion and conclusions

If the mind's eye is 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.

One empirical implication of this analysis is that the phenomenon should not be restricted to the visual system, but, in one form or another, is likely to be ubiquitous across conscious experience. Possible examples abound, ranging from acupuncture to the Lamaze birthing technique ('Concentrate on your breathing').

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.

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.

An obvious, and somewhat disturbing, corollary of this latter point is that the development of dependable computing devices based on biological paradigms seems severely limited by our present understanding of, and ability to replicate, that elaborate regulatory machinery.

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.

To reiterate, a 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 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 filter for grammatical and syntactical 'meaningful' paths, which affects what can be brought to consciousness. Phenomena outside the syntax/grammar bandpass of this manifold are subject to lessened probability of punctuated conscious detection: generalized inattentional blindness.

As Glazebrook (2005) has pointed out, 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. This 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. Expressing this result mathematically, at least in terms of the necessary conditions imposed by the asymptotic limit theorems of information theory on cognitive process, is apparently not as blood-simple as one would like it to be.

Appendix: contexts and consciousness

The preceding sections suggest, more generally, that, 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 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. These are, then, to be viewed as among Baars' contexts. 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 formal analytic tool for this extension is network information theory (Cover and Thomas, 1991, p. 387). Given three interacting information sources, Y_1, Y_2, Z, the splitting criterion, taking Z as the 'external context', is given by

$$I(Y_1, Y_2|Z) = H(Z) + H(Y_1|Z) - H(Y_1, Y_2, Z), \quad (13)$$

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

This generalizes to

$$I(Y_1, ...Y_n|Z) = H(Z) + \sum_{j=1}^{n} H(Y_j|Z) - H(Y_1, ..., Y_n, Z). \quad (14)$$

If we assume the brain's Global Workspace/GC to involve a very rapidly shifting 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 and cultural processes will be characterized by even more slowly-acting sources $Z_k, k = 1..n$. Mathematical induction on equation (8) gives a complicated expression for a mutual information splitting criterion which we write as

$$I(X|Y_1, .., Y_m|Z_1, .., Z_n). \quad (15)$$

This encompasses a full biopsychosociocultural structure for individual human consciousness, one in which contexts act as important boundary conditions.

This result does not commit the mereological fallacy which Bennett and Hacker (2003) impute to excessively neurocentric perspectives on consciousness. See Wallace (2005a, b) for more discussion.

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.

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.

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/.

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

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.

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.

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/

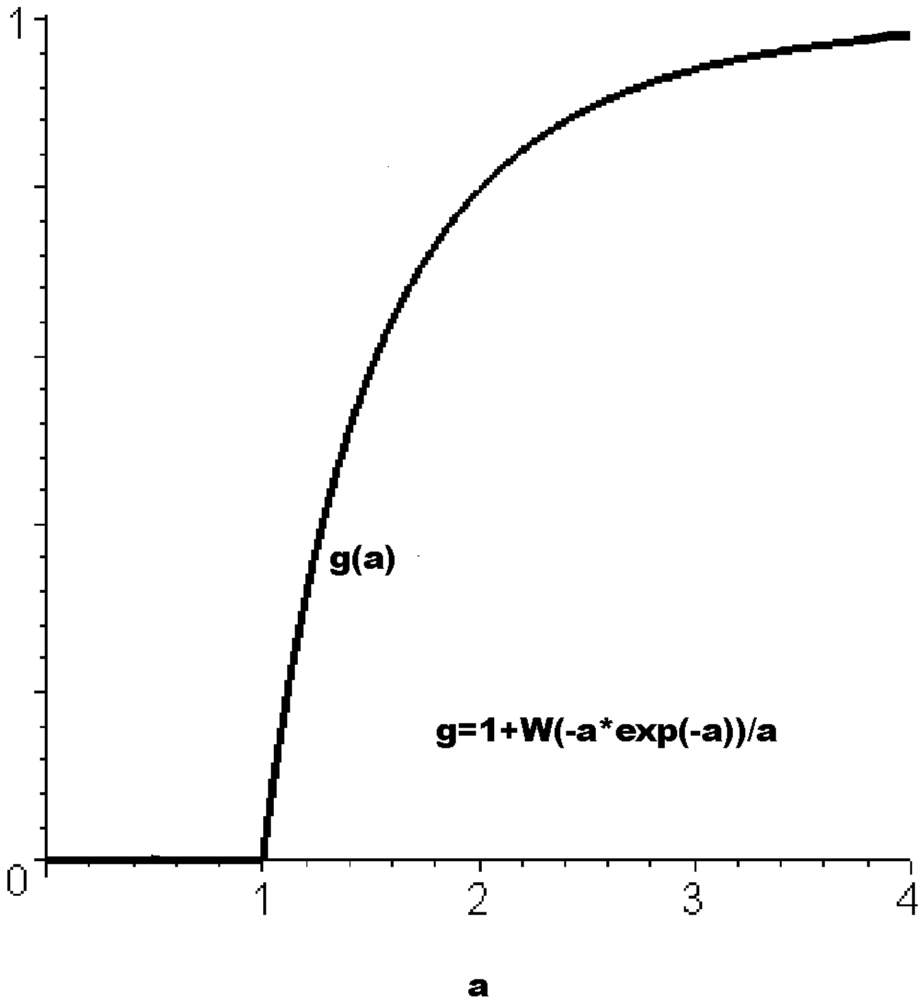
Weinstein A., 1996, Groupoids: unifying internal and external symmetry, *Notices of the American Mathematical Association*, 43:744-752.

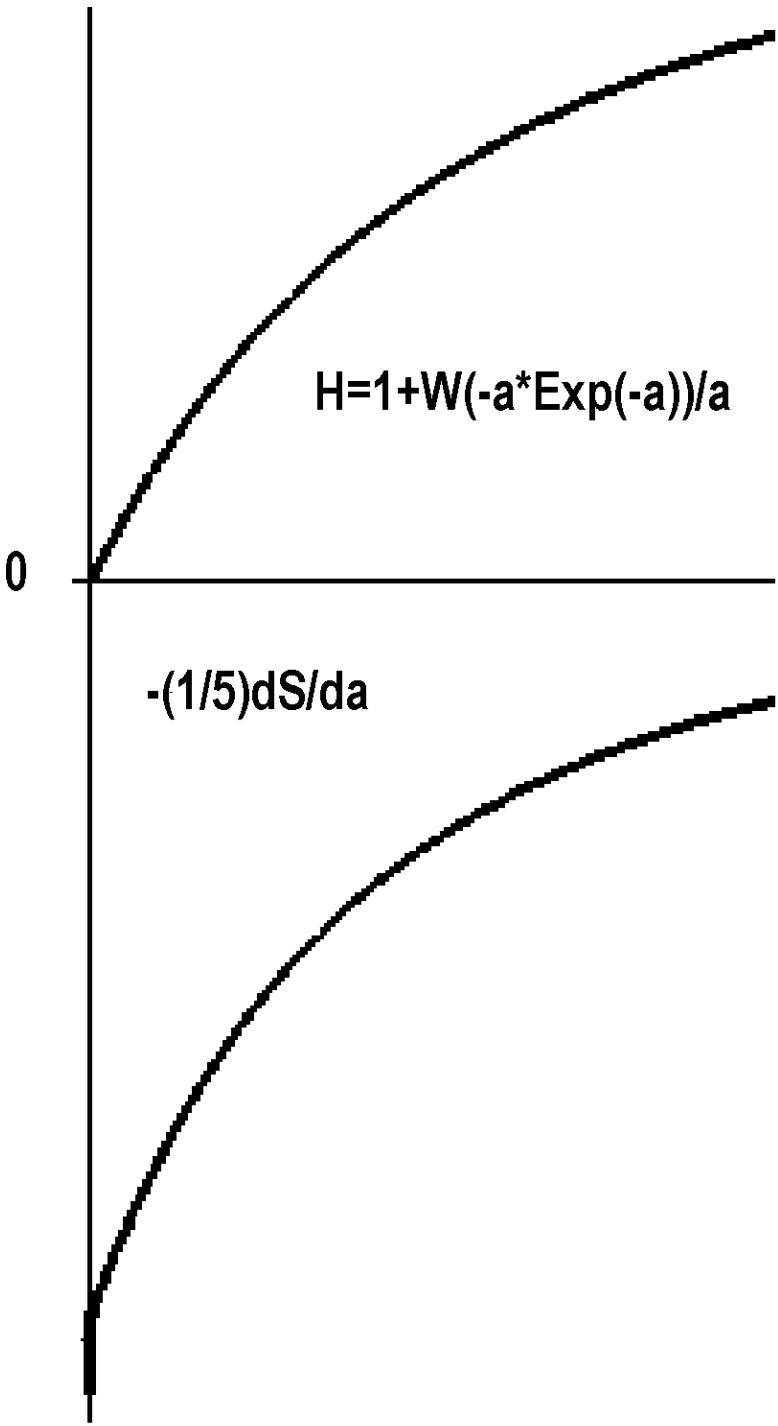
Figure Caption

Figure 1. Relative size of the largest connected component of a random graph, as a function of 2× the average number of fixed-strength connections between vertices. W is the Lambert-W function, or the ProductLog in Mathematica, which solves the relation $W(x)\exp[W(x)] = x$. Note the sharp threshold at $a = 1$, and the subsequent topping-out.'Tuning' the giant component by changing topology generally leads to a family of similar curves, those having progressively higher threshold with correspondingly lower asymptotic limits (e.g. Newman et al., 2001, fig. 4).

Figure 2. Assuming the source uncertainty, $H(a)$, of a conscious processes, the upper graph, is proportional to the size of the giant component, the thermodynamic force, $-dS/da$, with $S = H - dS/da$, will be against increase in a, the lower trace.

RELATIVE SIZE OF LARGEST CONNECTED COMPONENT





H=1+W(-a*Exp(-a))/a
0
-(1/5)dS/da