CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Nature Precedings

Version 7

# Extending Astrobiology: Consciousness and Culture

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

January 5, 2011

Nature Precedings : hdl:10101/npre.2011.5286.2 : Posted 5 Jan 2011

## Abstract

The Stanley Miller experiment suggests that amino acid-based life is ubiquitous in our universe, although its varieties will not have followed the particular, highly contingent and path-dependent, evolutionary trajectory found on Earth. Are many alien organisms likely to be conscious in ways we would recognize? Almost certainly. Will some develop high order technology? Less likely, but still fairly probable. If so, will we be able to communicate with them? Only on a basic level, and only with profound difficulty. The argument is fairly direct.

**Key words:** astroethology, astropsychology, information theory, technology

## 1 Introduction

In spite of a popular social construction as such (e.g., Penrose, 1994), individual consciousness is no great mystery, constituting a basic evolutionary adaptation likely a half-billion years old (R.G. Wallace and R. Wallace, 2009, and references therein). Bernard Baars' global workspace/global broadcast model – the current front-runner in the Darwinian competition between consciousness theories (e.g., Dehaene and Naccache, 2001) – is itself nearly a generation old and accounts neatly, in a qualitative manner, for individual consciousness-as-we-know-it on Earth (Baars, 1988, 2005):

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

2. Individual 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, scientific consciousness study 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 (Baars, 2005).

The first essential point in developing a quantitative theory of individual consciousness based on Baars' model is to recognize Dretske's central argument (Dretske, 1994) that high level mental process of any nature inevitably involves the generation and transmission of information, both of which are constrained by the asymptotic limit theorems of information theory: Shannon Coding, Shannon-McMillan Source Coding, and the Rate Distortion theorems (Khinchin, 1957; Cover and Thomas, 2006).

The second fundamental point is that, in the sense of Feynman (2000) and Bennett (1988), information is simply another form of free energy, and arguments-by-abduction from statistical physics are not unreasonable. Indeed, it is not at all difficult to construct a simple (ideal) machine that turns the information contained in a message into work.

Third, cognitive process, in the sense of Atlan and Cohen (1998), involves comparison of a perceived signal with an internal picture of the world, and then, on that comparison, choice of a response from a large repertoire of those possible, causing a reduction in a formal measure of uncertainty. It is then easy to show (Wallace, 2000; Wallace and Wallace, 2008, 2009, summarized in the Mathematical Appendix) that a substantial class of such cognitive phenomena is necessarily associated with a well-behaved information source: If there are $N(n)$ possible behavioral and hence temporal output paths of that information source having length $n$, then there will be a path-independent limit $H$ such that

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

*Box 47, 1051 Riverside Dr., New York, NY, 10032, wallace@pi.cpmc.columbia.edu

Nature Precedings : hdl:10101/npre.2011.5286.2 : Posted 5 Jan 2011

This is the Shannon uncertainty of the (stationary, ergodic) information source dual to the cognitive process.

We envision the evolution of a broad set of unconscious cognitive modules within a reproducing organism that serve a number of independent purposes, ranging from the search for food and habitat, and the avoidance of predation, to modalities of reproduction. Thus even a simple organism will have a large network of unconscious cognitive modules that, necessarily, interact through some kind of crosstalk, indexed by an average measure $P$.

## 2 Cognitive Phase Transitions

Recall that the free energy density of a physical system is defined as

$$F = \lim_{V \to \infty} -T \frac{\log[Z[T,V]]}{V} \equiv \frac{\log[\hat{Z}[T,V]]}{V}, \tag{2}$$

where $Z$ is the partition function of the Hamiltonian of a physical system

$$Z = \sum_i \exp[\frac{-E_i(V)}{\kappa T}]. \tag{3}$$

$E_i(V)$ is the energy of state $i$ at system volume $V$, $T$ the temperature, and $\kappa$ a constant.

Recall also Landau's perspective on phase transition (Pettini, 2007). The essence of his insight was that certain phase transitions took place in the context of a significant symmetry change, with one phase being more symmetric than the other. A symmetry is lost in the transition, i. e., spontaneous symmetry breaking. The greatest possible set of symmetries being that of the Hamiltonian describing the energy states. Usually, states accessible at lower temperatures will lack the symmetries available at higher temperatures, so that the lower temperature state is less symmetric, with the transition occurring in a punctuated manner.

Larger scale information sources can thus emerge in a punctuated manner from the crosstalk-enabled interaction of underlying unconscious cognitive biological modules – increasing $P$, parameterizing the average strength of crosstalk. Then equation (1), representing the source uncertainty of this larger information source, written as $H$, becomes

$$H[P] = \lim_{n \to \infty} \frac{\log[N(n,P)]}{n}, \tag{4}$$

similar to the second part of equation (2), and we can apply Landau's argument, assigning $P$ as the temperature-analog.

Some evolved neural substructures of organisms may be expected to operate at relatively high speed, in the realm of milliseconds. This implies, in turn, the inevitable and highly punctuated emergence of a rapid global cognitive process from the interaction of unconscious cognitive submodules if $P$ becomes large enough.

Evolution can then take this emergent phenomenon and run with it.

Given a network of interacting unconscious cognitive modules, we can construct an interacting network of dual information sources, and define a metric, say $r$, on it, using the methods of Wallace (2005). On that network it is then possible to define renormalization symmetries in terms of the usual 'clumping' transformation, so that, for clumps of size $R$, in an external 'field' of strength $J$ (that we can set to 0 in the limit), one can write, in the usual manner (e.g., Wilson, 1971)

$$H[P(R), J(R)] = f(R) H[P(1), J(1)],$$

$$\chi(P(R), J(R)) = \chi(P(1), J(1))/R, \tag{5}$$

where $\chi$ is a characteristic correlation length on the underlying network of interacting dual information sources.

As Wallace (2005) shows, following Wilson (1971), very many 'biological' renormalizations, $f(R)$, are possible that lead to a number of quite different classes of phase transformation. Baars' shifting global workspace emerges, in this model, *through the tuning of the renormalization symmetries* (Wallace, 2005). In particular, Pettini's (2007) topological argument can be used to create topologically-defined thresholds for detection of sensory signals by the shifting global information source representing consciousness (Wallace and Fullilove, 2008, Section 3.7; Wallace, 2007).

Nothing in this set of realizations of the Baars model seems restricted to terrestrial biological entities. Indeed, even on Earth, we know of widely different brain structures that instantiate conscious behaviors, ranging from familiar mammalian and avian forms, to reptiles, cephalopods, and perhaps insect colonies (Griffin and Speck, 2004; Edelman et al., 2005; Cartmill, 2000).

The essential point, from an astrobiological perspective, is that any organism evolving a set of unconscious cognitive submodules is likely to undergo an evolutionary transformation into something having a shifting, tunable, global workspace/broadcast mechanism. Those operating in the realm of a few hundred milliseconds would be analogous to consciousness-as-we-know-it.

# 3 Quantum Systems

The theory above is quite classical, and produces Baars' results directly. Wallace (2005, Chapter 5, Section 6), however, does examine how quantum versions of the asymptotic limit theorems of information theory (e.g., Bjelakovic et al., 2003, 2004) might be used to generalize the model. Unfortunately, the quantum results are not well characterized, and an exact treatment is lacking, Nonetheless, it becomes quite clear that consciousness in quantum systems – at least those supporting relatively large coherence lengths – would be to consciousness-as-we-know-it much as a flask of superfluid helium is to a glass of water.

Tegmark (2000), of course, has convincingly shown the impossibility of quantum treatments of consciousness at normal biological temperatures.

Since information is a form of free energy, even quantum systems having large coherence lengths will suffer second law heating through information transmission and transformation that will inherently limit the possible size of quantum-conscious structures. Typically $10^9 - 10^{10}$ interacting components are needed for high level mental function, involving large-scale information transfer. This scale of activity is likely to generate much heat, and unlikely to be attained in quantum realms by evolutionary process in the natural world.

The inference is, then, that conscious quantum systems are likely to remain in the realm of perpetual motion machines of the second kind.

# 4 Culture, Technology, and Collective Consciousness

The evolutionary anthropologist Robert Boyd has asserted that 'Culture is as much a part of human biology as the enamel in our teeth,' (e.g., Richerson and Boyd, 2004) and, while many other animals on Earth display some measure of culture as learned and transmitted behavior (e.g., Avital and Jablonka, 2000), nothing defines humans quite like the interpenetration of mind and self with cultural milieu. Technology and its artifacts are, of course, one part of that milieu.

It is not difficult to extend the Baars model to include interaction with an embedding culture and with a hierarchical set of institutions within that culture seen as a generalized transmissible language associated with a nested set of information sources. This includes both a form of niche construction (Wallace, 2010), and distributed cognitive institutions acting on various scales (Wallace and Fullilove, 2008). This is most easily done by invoking the set of interacting information sources dual to cognitive process via network information theory (e.g., El Gamal and Kim, 2010, p.2-26): Given a basic set of such dual information sources, say $(X_1, ..., X_k)$, that can be partitioned into two ordered sets, say $X(\mathcal{J})$ and $X(\mathcal{J}')$, then the splitting criterion of the larger system becomes $H(X(\mathcal{J})|X(\mathcal{J}'))$. Generalization to three or more such ordered sets seems direct, and leads to a Baars-like theory of collective consciousness in which different global workspaces act at different scales of size and time.

As the anthropologists will attest, an astounding variety of culturally-driven institutions, associated forms of mind and self, and dynamics of interaction, graces the world. Typically, humans, whose overall genetic structure is more uniform than that of chimpanzee populations, do not communicate well across the many different cultural modes. Wallace and Fullilove (2008) suggest that stabilizing complex systems of interacting cognitive institutions is exceedingly difficult, given that canonical inability to communicate, and the planet seems to be facing a serious crisis of sustainability.

# 5 Consciousness, Culture, and Astrobiology

Certain matters seem clear from this line of argument:

Life in the cosmos is likely to be ubiquitous. Organisms that must react on timescales of a few hundred milliseconds should host many 'neural-like' structures that, in the presence of sufficient crosstalk, provide evolutionary process with the basic material to produce adaptive tunable/shifting global workspace/broadcast phenomena that become fixed in reproduction and that we would likely recognize as conscious.

A significant number of alien organisms will, over sufficient time, become synergistic with learned, transmissible, language-like patterns of adaptation analogous to culture that include collective structures of various forms acting under distributed cognition that are capable of large-scale cooperative activity. Assuming some few of these creatures able to stabilize the resulting systems of collective consciousness, emergence of high technology seems likely, although communication with them would probably be limited to exchanges of Balmer series symbols and schematics for amino acids.

In conclusion, astrobiology and, with reservations, astropsychology, appear relatively straightforward. Astroethology, by contrast, would be a profound intellectual challenge.

# 6 Mathematical Appendix

Cognitive pattern recognition-and-selected response, following the model of Atlan and Cohen (1998), proceeds by convoluting an incoming external 'sensory' signal with an internal 'ongoing activity' – which includes, but is not limited to, the learned picture of the world – and, at some point, triggering an appropriate action based on a decision that the pattern of sensory activity requires a response. It is not necessary

Nature Precedings : hdl:10101/npre.2011.5286.2 : Posted 5 Jan 2011

Nature Precedings : hdl:10101/npre.2011.5286.2 : Posted 5 Jan 2011

to specify how the pattern recognition system is trained, and hence possible to adopt a weak model, regardless of learning paradigm, that can be more formally described by the Rate Distortion Theorem. Fulfilling Atlan and Cohen's criterion of meaning-from-response, it is possible to define a language's contextual meaning entirely in terms of system output.

The model, a simplification of the standard neural network, is as follows.

A pattern of 'sensory' input – incorporating feedback from the external world – is expressed as an ordered sequence $y_0, y_1, ....$ This is mixed in a systematic (but unspecified) algorithmic manner with internal 'ongoing' activity, a sequence $w_0, w_1, ...$, to create a path of composite signals $x = a_0, a_1, ..., a_n, ...$, where $a_j = f(y_j, w_j)$ for some function $f$. This path is then fed into a highly nonlinear, but otherwise similarly unspecified, decision oscillator generating an output $h(x)$ that is an element of one of two (presumably) disjoint sets $B_0$ and $B_1$. We take $B_0 \equiv \{b_0, ..., b_k\}, B_1 \equiv \{b_{k+1}, ..., b_m\}$.

Thus the model permits 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.

This approach is broadly analogous to, but simpler than, the Hopfield/Hebb stochastic neuron in which series of inputs $y_i^j, i = 1...m$ from $m$ nearby neurons at time $j$ is convoluted with 'weights' $w_i^j, i = 1...m$, using an inner product $a_j = \mathbf{y}^j \cdot \mathbf{w}^j = \sum_{i=1}^m y_i^j w_i^j$ in the context of a 'transfer function' $f(\mathbf{y}^j \cdot \mathbf{w}^j)$ such that the probability of the neuron firing and having a discrete output $z^j = 1$ is $P(z^j = 1) = f(\mathbf{y}^j \cdot \mathbf{w}^j)$. Thus the probability that the neuron does not fire at time $j$ is $1 - f(\mathbf{y}^j \cdot \mathbf{w}^j)$.

The $m$ values $y_i^j$ constitute 'sensory activity' and the $m$ weights $w_i^j$ the 'ongoing activity' at time $j$, with $a_j = \mathbf{y}^j \cdot \mathbf{w}^j$ and $x = a_0, a_1, ...a_n, ....$ A little more work leads to a fairly standard neural network model in which the network is trained by appropriately varying the $\mathbf{w}$ through least squares or other error minimization feedback.

The principal focus of the simpler model given here is the composite paths $x$ that trigger pattern recognition-and-response. 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 to the event $h(x) \in B_1$. Thus $h(a_0, ..., a_j) \in B_0$ for all $0 \leq j < m$, but $h(a_0, ..., a_m) \in B_1$. Remember, the $y_j$, the 'sensory' input convoluted with the internal $w_j$, contains feedback from the external world, i.e., how well $h$ matches intent with need.

For each positive integer $n$ let $N(n)$ be the number of grammatical and syntactic high probability 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' and assume $N(n)$ to be considerably less than the number of all possible paths of length $n$ – pattern recognition-and-response is comparatively rare. Again assume that the longitudinal finite limit $H \equiv \lim_{n\to\infty} \log[N(n)]/n$ both exists and is independent of the path $x$. Call such a cognitive process *ergodic*.

Note that disjoint partition of state space may be possible according to sets of states which can be connected by meaningful paths from a particular base point, leading to a natural coset algebra of the system defining a groupoid.

It is thus possible to define an 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 may be defined which satisfy the standard relations (Cover and Thomas, 2006).

This information source is taken as *dual* to the ergodic cognitive process.

Recall that the Shannon-McMillan Theorem and its variants provide 'laws of large numbers' that permit definition of the Shannon uncertainties in terms of cross-sectional sums of the form $H = -\sum P_k \log[P_k]$, where the $P_k$ constitute a probability distribution (Ash, 1990; Cover and Thomas, 2006).

Different quasi-languages will be defined by different divisions of the total universe of possible responses into various pairs of sets $B_0$ and $B_1$. Like the use of different distortion measures in the Rate Distortion Theorem, however, it seems obvious that the underlying dynamics will all be qualitatively similar.

Nonetheless, dividing the full set of possible responses into the sets $B_0$ and $B_1$ may itself require higher order cognitive decisions by another module or modules, suggesting the necessity of choice within a more or less broad set of possible quasi-languages. This would directly reflect the need to shift gears according to the different challenges faced by the organism or organic subsystem. A critical problem then becomes the choice of a normal zero-mode language among a very large set of possible languages representing accessible excited states. This is a fundamental matter that mirrors, for isolated cognitive systems, the resilience arguments applicable to more conventional ecosystems, that is, the possibility of more than one zero state to a cognitive system. Identification of an excited state as the zero mode becomes, then, a kind of generalized autoimmune disorder that can be triggered by linkage with external ecological information sources representing various kinds of structured stress.

In sum, meaningful paths – creating an inherent grammar and syntax – have been defined entirely in terms of system response, as Atlan and Cohen (1998) propose.

# 7 References

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,

Avital, E., and E. Jablonka, 2000, Animal Traditions: Behavioral Inheritance in Evolution, Cambridge University Press, New York.

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

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

Bjelakovic, I., T. Kruger, R. Siegmund-Schultz, and A. Szkola, 2003, Chained typical subspaces a quantum version of Brieman's theorem, ArXiv, quant-ph/0301177.

Bjelakovic, I., T. Kruger, R. Siegmund-Schultze, and A. Szkola, 2004, The Shannon-McMillan theorem for ergodic quantum lattice systems, Inventiones Mathematicae, 155:203-222.

Bennett, C., 1988, Logical depth and physical complexity. In The Universal Turing Machine: A Half-Century Survey. R. Herkin (ed.), pp. 227-257. Oxford University Press, New York..

Cartmill, M., 2000, Animal consciousness: some philosophical, methodological, and evolutionary problems, American Zoologist, 40:835-846.

Cover, T., and J. Thomas, 2006, Elements of Information Theory, Second Edition, 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.

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

Edelman, D., B. Baars, and A. Seth, 2005, Identifying hallmarks of consciousness in non-mammalian species, Consciousness and Cognition, 14:169-187.

El Gamal, A., and Y. Kim, 2010, Lecture notes on network information theory. arXiv:1001.3404v4 [cs.IT]

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

Griffin, D., and G., Speck, 2004, New evidence of animal consciousness, Animal Cognition, 7:5-18.

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

Penrose, R., 1994, Shadows of the Mind: A Search for the Missing Science of Consciousness, Oxford University Press, New York.

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

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

Tegmark, M., 2000, Importance of quantum decoherence in brain processes, Physical Review E, 61:4194-4206.

Wallace, R.G., and R.Wallace, 2009, Evolutionary radiation and the spectrum of consciousness, Consciousness and Cognition, 18:160-167.

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

Wallace, R., and D. Wallace, 2008, Punctuated equilibrium in statistical models of generalized coevolutionary resilience: how sudden ecosystem transitions can entrain both phenotype expression and Darwinian selection, Transactions on Computational Systems Biology IX, LNBI 5121, 23-85.

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

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

Wallace, R., 2007, Culture and inattentional blindness, Journal of Theoretical Biology, 245:378-390.

Wallace, R., 2010, Expanding the modern synthesis II: Formal perspectives on the inherent role of niche construction. http://precedings.nature.com/documents/5059/version/2

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

Nature Precedings : hdl:10101/npre.2011.5286.2 : Posted 5 Jan 2011