BioSystems 117 (2014) 54–59

Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

Statistical models of critical phenomena in fuzzy biocognition

Rodrick Wallace

Division of Epidemiology, The New York State Psychiatric Institute, Box 47, 1051 Riverside Drive, New York, NY 10032, United States

ARTICLE INFO

Article history:
Received 26 November 2013
Received in revised form 3 January 2014
Accepted 12 January 2014

Keywords:
Crosstalk
Groupoid
Index theorem
Information theory
Phase transition

ABSTRACT

A recent line of study exploring statistical models of punctuated global broadcasts associated with attention states has focused on the evolutionary exaptation of the inevitable signal crosstalk between related sets of unconscious cognitive modules (UCM). This work invokes a groupoid treatment of the equivalence classes arising from information sources 'dual', in a formal sense, to the UCM, via a standard spontaneous symmetry breaking/lifting methodology abducted from statistical physics. A related approach involves an index theorem based on a stochastic empirical Onsager-like entropy-analog gradient model. Surprisingly, similar arguments may apply to 'fuzzy groupoid' generalizations likely to better fit biological complexities.

1. Introduction

Many broadly cognitive physiological systems engage in a highly punctuated, dynamic recruitment of lower-level 'unconscious' cognitive modules to form attention-directed, temporary working coalitions that address patterns of threat and opportunity confronting organisms, and do so across a variety of modalities, levels of organization, scales, and time domains (Wallace, 2012a; Wallace and Wallace, 2013).

This set of mechanisms represents the evolutionary exaptation (Gould, 2002) of the crosstalk inevitable to information transmission. Crosstalk – correlation between interacting channels – is known from electrical engineering practice to require extraordinary efforts for mitigation, and is observed across a broad spectrum of biological processes (e.g., McCarthy, 2010; McNeill and Woodgett, 2010; Attisano and Wrana, 2013).

Previous studies of such cognitive global broadcasts have been based on necessary conditions models arising from the asymptotic limit theorems of information theory, following the footsteps of Dretske (Adams, 2003; Dretske, 1994). These have assumed well-defined, tiling-like, quasi-symmetries inherent to cognitive dynamics, and invoked spontaneous symmetry-breaking arguments akin to those now standard in physical theory to derive the highly punctuated accession to 'consciousness' and overt attention characterizing such phenomena (Wallace, 2005, 2007, 2012a). Here, we extend these arguments to inherently 'fuzzy' systems, where underlying form is smeared out, in a certain sense, and may hence be in better consonance with biological realities.

Fuzzy sets, algorithms, and control systems, were first studied by Zadeh (1965, 1968), and have received increasing attention (e.g., Jantzen, 2007). Biological global broadcasts that range from the immune system, wound healing, the HPA axis, emotional response, consciousness, and sociocultural distributed cognition (Wallace, 2012a), are basically control processes, and the application of Zadeh's perspective appears straightforward. In essence, the 'fuzzification' of algebraic structures and relations is based on an extension of the idea of the characteristic function, mapping an arbitrary set G onto the set of integers $\{0, 1\}$, so that $f: G \to 0, 1$. Then, if $x \in G$, $f(x) = 1$. Otherwise, $f(x) = 0$. The fundamental idea involves letting f map onto the real interval [0, 1] rather than onto a set of integers. Rosenfeld (1971) first applied the method to defining fuzzy groups and groupoids, and the construction of group/groupoid representations is relatively direct, although modified by some complexities (Bos, 2007; Houghton, 1975).

2. Cognition as an information source

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

E-mail addresses: Wallace@nyspi.columbia.edu, wallace@pi.cpmc.columbia.edu

0303-2647/$ – see front matter © 2014 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.biosystems.2014.01.004

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

This path is fed into a highly nonlinear, but otherwise similarly unspecified, decision function, h, generating an output $h(x)$ that is an element of one of two disjoint sets B_0 and B_1 of possible system responses. Let $B_0 \equiv \{b_0, \ldots, b_k\}$, and $B_1 \equiv \{b_{k+1}, \ldots, b_m\}$.

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, $k+1 \le j \le m$ takes place.

Interest focuses on paths x triggering pattern recognition-and-response: given a fixed initial state a_0, examine all possible subsequent paths x beginning with a_0 and leading to the event $h(x) \in B_1$. Thus $h(a_0, \ldots, a_j) \in B_0$ for all $0 \le j < m$, but $h(a_0, \ldots, a_m) \in B_1$.

For each positive integer n, let $N(n)$ be the number of high probability paths of length n that begin with some particular a_0 and lead to the condition $h(x) \in B_1$. Call such paths 'meaningful', assuming that $N(n)$ will be considerably less than the number of all possible paths of length n leading from a_0 to the condition $h(x) \in B_1$.

Note that identification of the 'alphabet' of the states a_j, B_k may depend on the proper system 'coarse graining' in the sense of symbolic dynamics (Beck and Schlogl, 1993), and this is often far from trivial.

Combining algorithm, the form of the function h, and the details of grammar and syntax, are all unspecified in this model. The assumption permitting inference on necessary conditions constrained by the asymptotic limit theorems of information theory is that the finite limit

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

both exists and is independent of the path x. Recall that $N(n)$ is the number of high probability paths of length n.

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

Invoking the spirit of the Shannon–McMillan Theorem, it is possible to define an adiabatically, piecewise stationary, ergodic information source **X** associated with stochastic variates X_j having joint and conditional probabilities $P(a_0, \ldots, a_n)$ and $P(a_n | a_0, \ldots, a_{n-1})$ such that appropriate joint and conditional Shannon uncertainties satisfy the classic relations (Cover and Thomas, 2006)

$$\begin{aligned} H[\mathbf{X}] &= \lim_{n \to \infty} \frac{\log[N(n)]}{n} = \lim_{n \to \infty} H(X_n | X_0, \ldots, X_{n-1}) \\ &= \lim_{n \to \infty} \frac{H(X_0, \ldots, X_n)}{n}. \end{aligned} \tag{2}$$

This information source is defined as dual to the underlying ergodic cognitive process.

'Adiabatic' means that, when the information source is parameterized according to some appropriate scheme, within continuous 'pieces', changes in parameter values take place slowly enough so that the information source remains as close to stationary and ergodic as needed to make the fundamental limit theorems work. 'Stationary' means that probabilities do not change in time, and 'ergodic' (roughly) that cross-sectional means converge to long-time averages. Between 'pieces' it is possible to invoke various kinds of phase change formalism (Wallace, 2005), usually much more complicated than has been found necessary for the study of physical systems.

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

An equivalence class algebra can be constructed by choosing different origin points, a_0, and defining the equivalence of two states, a_m, a_n, by the existence of a high probability meaningful path connecting them to the same origin point. Disjoint partition by equivalence class, analogous to orbit equivalence classes for dynamical systems, defines the vertices of a 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 set of cognitive biological processes.

Such a set of equivalence classes generates a groupoid, whose algebraic properties – an important extension of the idea of both a symmetry group and an equivalence class – are summarized in the Mathematical appendix. An essential point is that products need not be defined globally (Brown, 1987; Cannas da Silva and Weinstein, 1999; Weinstein, 1996).

We now allow generalization of these ideas to fuzzy groupoids, in Rosenfeld's sense (1971), in the context of the Data Rate Theorem and its extensions.

3. The Data-Rate Theorem

The recently-formalized Data-Rate Theorem, itself a generalization of the classic Bode integral theorem for linear control systems, describes the stability of linear feedback control under data rate constraints (e.g., Nair et al., 2007). Given a noise-free data link between a discrete linear plant and its controller, unstable modes can be stabilized only if the feedback data rate $\mathcal{H}$ is greater than the rate of 'topological information' generated by the unstable system. For the simplest incarnation, if the linear matrix equation of the plant is of the form $x_{t+1} = \mathbf{A}x_t + \ldots$, where x_t is the n-dimensional state vector at time t, then the necessary condition for stabilizability is

$$\mathcal{H} > \log[|\det \mathbf{A}^u|] \tag{3}$$

where *det* is the determinant and $\mathbf{A}^u$ is the decoupled unstable component of $\mathbf{A}$, i.e., the part having eigenvalues ≥ 1.

The essential matter is that there is a critical positive data rate below which there does not exist any quantization and control scheme able to stabilize an unstable (linear) feedback system.

This result, and its variations, for the first time linking control theory to information theory, are as fundamental as the Shannon Coding and Source Coding Theorems, and the Rate Distortion Theorem (Cover and Thomas, 2006; Ash, 1990; Khinchin, 1957). Since cognitive biological modules are dedicated to regulation and control, some version of this result will impose itself on their dynamics.

We extend the cognition-as-information source approach, exploring cognitive dynamics that inherently take place under data-rate constraints.

The means taken will be something much like Pettini's (2007) topological hypothesis which is a version of Landau's spontaneous symmetry breaking insight for physical systems (Landau and Lifshitz, 2007). The hypothesis infers that punctuated events often involve a change in the topology of an underlying configuration space, and the observed singularities in the measures of interest can be interpreted as a 'shadow' of major topological change happening at a more basic level.