

Version 3
**Does ‘information control the living
state’?**

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

March 3, 2014

Abstract

We generalize the recently-uncovered Data Rate Theorem in the context of cognitive systems having a ‘dual’ information source, including those of the living state that is particularly characterized by cognition at every scale and level of organization. The unification of information theory and control theory via the Data Rate Theorem is not additive, but synergistic, generating new statistical tools that greatly constrain the possible dynamics of that state. Thus, in addition to providing novel conceptual approaches, this emerging body of theory permits construction of models that, like those of regression analysis, can provide benchmarks against which to compare experimental or observational data.

Key Words: cognition; critical phenomena; phase transition; renormalization

1 Introduction

The keynote session of the 2014 Gordon Research Conferences section on Complex Adaptive Matter, ‘How Can Information Control Matter?’ by Goldenfeld and Ramaswamy, raises a question that has been the focus of intense debate since the end of World War II. Information theory and control theory both emerged from the technological cauldron of that conflict, engaging seemingly separate problems in communication and system regulation. The Shannon Coding and Source Coding Theorems, and the Rate Distortion Theorem, provide statistical constraints on communication, while the Bode Integral Theorem constrains system control – noise energy suppressed in one frequency range emerges

*Wallace@nyspi.columbia.edu, rodrick.wallace@gmail.com

in another. The two disciplines remained largely separate over the succeeding half-century.

Beginning in the late 1990's, however, a series of studies delineated a central relation between information theory and control theory, a direct extension of Bode's result known now as the Data Rate Theorem that, at least partially, addresses Goldenfeld and Ramaswamy's question. Here we will explore in more detail the implications of the unification of information and control theories for understanding the dynamics of the living state.

The underlying conceit is that living systems are cognitive at every scale and level of organization, and that many cognitive phenomena can be represented by 'dual' information sources constrained by the four basic theorems (e.g., Wallace, 2012a, 2014a; Wallace and Wallace, 2013). These ideas, of course, are quite old, first articulated by Maturana and Varela (1980). Wallace (2014a) and Wallace and Wallace (2013) describe gene expression, the immune system, tumor control, wound healing, animal consciousness, sociocultural cognition, and other biological processes, as examples. Wallace (2012b), in fact, provides a detailed exploration of the glycan/lectin cell surface 'kelp bed' as a cognitive system that necessarily rivals high order neural phenomena in its sophistication – the argument is surprisingly direct. In a sense, then, the question asked in the title to this paper is something of a red herring, reflecting a set of popular misconceptions about the relation between information and life, since information dynamics are so deeply convoluted with life itself.

We will, ultimately, expand the perspective via an 'obvious' generalization of the Data Rate Theorem.

2 The Data-Rate Theorem

The data-rate theorem, based on the Bode integral theorem for linear control systems (e.g., Yu and Mehta, 2010; Kitano, 2007; Csete and Doyle, 2002), describes the stability of linear feedback control under data rate constraints (e.g., Mitter, 2001; Tatikonda and Mitter, 2004; Sahai, 2004; Sahai and Mitter, 2006; Minero et al., 2009; Nair et al., 2007; You and Xie, 2013). 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 $x_{t+1} = \mathbf{A}x_t + \dots$, where x_t is the n-dimensional state vector at time t , then the necessary condition for stabilizability is

$$\mathcal{H} > \log[|\det \mathbf{A}^u|] \quad (1)$$

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

There is, then, 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, are as fundamental as the Shannon Coding and Source Coding Theorems, and the Rate Distortion Theorem (Cover and Thomas, 2006; Ash, 1990; Khinchin, 1957).

It is possible to significantly extend the argument. The essential analytic tool will be an analog to Pettini's (2007) 'topological hypothesis' – a version of Landau's spontaneous symmetry breaking insight for physical systems (Landau and Lifshitz, 2007) – which 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. The tool for the study of such topological changes is Morse Theory (Pettini, 2007; Matsumoto, 2002).

The first step is a recapitulation of an approach to cognition using the asymptotic limit theorems of information theory (Wallace 2000, 2005, 2007, 2012a, 2014a).

3 Cognition as an information source

Atlan and Cohen (1998) argue that the essence of 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. That is, cognitive pattern recognition-and-response proceeds by an algorithmic combination of an incoming external sensory signal with an internal ongoing activity – incorporating the internalized picture of the world – and triggering an appropriate action based on a decision that the pattern of sensory activity requires a response.

Incoming sensory input is thus mixed in an unspecified but systematic manner with internal ongoing activity to create a path of combined signals $x = (a_0, a_1, \dots, a_n, \dots)$. 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 some 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

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

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

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.

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 $h(a_0, \dots, a_j) \in B_0$ for all $0 \leq j < m$, but $h(a_0, \dots, a_m) \in B_1$.

For each positive integer n , take $N(n)$ as 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 $h(x) \in B_1$.

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 (e.g., Beck and Schlogl, 1993).

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 $H \equiv \lim_{n \rightarrow \infty} \log[N(n)]/n$ both exists and is independent of the path x . Again, $N(n)$ 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 (e.g., Wallace, 2005, pp. 31-32).

Invoking the Shannon-McMillan Theorem (Cover and Thomas, 2006; Khinchin, 1957), it becomes possible to define an adiabatically, piecewise stationary, ergodic information source \mathbf{X} associated with stochastic variates X_j having joint and conditional probabilities $P(a_0, \dots, a_n)$ and $P(a_n|a_0, \dots, a_{n-1})$ such that appropriate joint and conditional Shannon uncertainties satisfy the classic relations

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

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

‘Adiabatic’ means that, when the information source is properly parameterized, 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’ that cross-sectional means converge to long-time averages. Between pieces, as will be described below, it is necessary to invoke phase change formalism, a ‘biological’ renormalization that generalizes Wilson’s (1971) approach to physical phase transition (Wallace, 2005).

Shannon uncertainties $H(\dots)$ are cross-sectional law-of-large-numbers sums of the form $-\sum_k P_k \log[P_k]$, where the P_k constitute a probability distribution.

See Cover and Thomas (2006), Ash (1990), or Khinchin (1957) for the standard details.

We are not, however, constrained in this approach to the Atlan-Cohen model of cognition that, through the comparison with an internal picture of the world, invokes representation. The essential inference is that a broad class of cognitive phenomena – with and without representation – can be associated with a dual information source. The argument is direct, since cognition inevitably involves choice, choice reduces uncertainty, and this implies the existence of an information source.

Extension to non-ergodic information sources can be done using the methods of Wallace (2005, Sec. 3.1).

4 Groupoid symmetries

For cognitive systems, 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 high probability meaningful paths connecting them to the same origin point. Disjoint partition by equivalence class, analogous to orbit equivalence classes for a dynamical system, defines the vertices of a network of cognitive dual languages that interact to actually constitute the system of interest. Each vertex then represents a different information source dual to a cognitive process. This is not a representation of a network of interacting physical systems as such, in the sense of network systems biology (e.g., Arrell and Terzic, 2010). It is an abstract set of languages dual to the set of cognitive processes of interest, that may become linked into higher order structures.

Topology is now an object of algebraic study, so-called algebraic topology, via the fundamental underlying symmetries of geometric spaces. Rotations, mirror transformations, simple (‘affine’) displacements, and the like, uniquely characterize topological spaces, and the networks inherent to cognitive phenomena having dual information sources also have complex underlying symmetries: characterization via equivalence classes defines a groupoid, an extension of the idea of a symmetry group, as summarized by Brown (1987) and Weinstein (1996). Linkages across this set of languages occur via the groupoid generalization of Landau’s spontaneous symmetry breaking arguments that will be used below (Landau and Lifshitz, 2007; Pettini, 2007).

5 ‘Environment’ as an information source

Multifactorial cognitive and behavioral systems interact with, affect, and are affected by, embedding ‘environments’, in a large sense, that remember interaction by various mechanisms. It is possible to reexpress environmental dynamics in terms of a grammar and syntax that represent the output of an information source – another generalized language.

Obviously, within an organism, social assemblage of organisms, or ecosystem,

different structures may, at different times, play the role of ‘system’ and of ‘environment’, of controllee and of controller.

Some examples:

1. The turn-of-the seasons in a temperate climate, for many ecosystems, looks remarkably the same year after year: the ice melts, the migrating birds return, the trees bud, the grass grows, plants and animals reproduce, high summer arrives, the foliage turns, the birds leave, frost, snow, the rivers freeze, and so on.

2. Human interactions take place within fairly well defined social, cultural, and historical constraints, depending on context: birthday party behaviors are not the same as cocktail party behaviors in a particular social set, but both will be characteristic.

3. Gene expression during development is highly patterned by embedding environmental context via ‘norms of reaction’ (e.g., Wallace and Wallace, 2010, 2013; Wallace, 2014).

Suppose it possible to coarse-grain the generalized ‘ecosystem’ at time t , in the sense of symbolic dynamics (e.g., Beck and Schlogl, 1993) according to some appropriate partition of the phase space in which each division A_j represent a particular range of numbers of each possible fundamental actor in the generalized ecosystem, along with associated larger system parameters. What is of particular interest is the set of longitudinal paths, system statements, in a sense, of the form $x(n) = A_0, A_1, \dots, A_n$ defined in terms of some natural time unit of the system. Thus n corresponds to an again appropriate characteristic time unit T , so that $t = T, 2T, \dots, nT$.

Again, the central interest is in serial correlations along paths.

Let $N(n)$ be the number of possible paths of length n that are consistent with the underlying grammar and syntax of the appropriately coarsegrained embedding ecosystem, in a large sense. As above, the fundamental assumptions are that – for this chosen coarse-graining – $N(n)$, the number of possible grammatical paths, is much smaller than the total number of paths possible, and that, in the limit of (relatively) large n , $H = \lim_{n \rightarrow \infty} \log[N(n)]/n$ both exists and is independent of path.

These conditions represent a parallel with parametric statistics in that systems for which the assumptions are not true will require specialized approaches.

Nonetheless, not all possible ecosystem coarse-grainings are likely to work, and different such divisions, even when appropriate, might well lead to different descriptive quasi-languages for the ecosystem of interest. Thus, empirical identification of relevant coarse-grainings for which this theory will work may represent a difficult scientific problem.

Given an appropriately chosen coarse-graining, define joint and conditional probabilities for different ecosystem paths, having the form $P(A_0, A_1, \dots, A_n)$, $P(A_n|A_0, \dots, A_{n-1})$, such that appropriate joint and conditional Shannon uncertainties can be defined on them that satisfy equation (2).

Taking the definitions of Shannon uncertainties as above, and arguing backwards from the latter two parts of equation (2), it is indeed possible to recover the first, and divide the set of all possible ecosystem temporal paths into two

subsets, one very small, containing the grammatically correct, and hence highly probable paths, that we will call ‘meaningful’, and a much larger set of vanishingly low probability.

6 Interacting information sources

Given a set of cognitive modules (having dual information sources) that are linked to solve a problem, the ‘no free lunch’ theorem (English, 1996; Wolpert and Macready, 1995, 1997) extends a network theory-based theory (e.g., Arrell and Terzic, 2010). Wolpert and Macready show there exists no generally superior computational function optimizer. That is, there is no ‘free lunch’ in the sense that an optimizer pays for superior performance on some functions with inferior performance on others gains and losses balance precisely, and all optimizers have identical average performance. In sum, an optimizer has to pay for its superiority on one subset of functions with inferiority on the complementary subset.

This result is well-known using another description. Shannon (1959) recognized a powerful duality between the properties of an information source with a distortion measure and those of a channel. This duality is enhanced if we consider channels in which there is a cost associated with the different letters. Solving this problem corresponds to finding a source that is right for the channel and the desired cost. Evaluating the rate distortion function for a source corresponds to finding a channel that is just right for the source and allowed distortion level.

Yet another approach to the same result is the through the ‘tuning theorem’ (Wallace, 2005, Sec. 2.2), which inverts the Shannon Coding Theorem by noting that, formally, one can view the channel as ‘transmitted’ by the signal. Then a dual channel capacity can be defined in terms of the channel probability distribution that maximizes information transmission assuming a fixed message probability distribution.

From the no free lunch argument, Shannon’s insight, or the ‘tuning theorem’, it becomes clear that different challenges facing any cognitive system, distributed collection of them, or interacting set of other information sources, that constitute an organism must be met by different arrangements of cooperating modules represented as information sources.

It is possible to make a very abstract picture of this phenomenon based on the network of linkages between the information sources dual to the individual ‘unconscious’ cognitive modules (UCM), and those of related information sources with which they interact. That is, a shifting, task-mapped, network of information sources is continually reexpressed: given two distinct problems classes confronting the organism, there must be two different wirings of the information sources, including those dual to the available UCM, with the network graph edges measured by the amount of information crosstalk between sets of nodes representing the different sources.

Thus living systems involve interaction between very general sets of infor-

mation sources assembled into a ‘task-specific device’ in the sense of Bingham (1988) that is necessarily highly tunable. This mechanism represents a broad generalization of the ‘shifting spotlight’ characterizing the global neuronal workspace model of consciousness (Baars, 1988; Wallace, 2005, 2012).

The mutual information measure of cross-talk is not inherently fixed, but can continuously vary in magnitude. This suggests a parameterized renormalization: the modular network structure linked by crosstalk has a topology depending on the degree of interaction of interest.

Define an interaction parameter ω , a real positive number, and look at geometric structures defined in terms of linkages set to zero if mutual information is less than, and ‘renormalized’ to unity if greater than, ω . Any given ω will define a regime of giant components of network elements linked by mutual information greater than or equal to it.

Now invert the argument: a given topology for the giant component will, in turn, define some critical value, ω_C , so that network elements interacting by mutual information less than that value will be unable to participate, i.e., will be locked out and not be consciously or otherwise perceived. See Wallace (2005, 2012a) for details. Thus ω is a tunable, syntactically-dependent, detection limit that depends critically on the instantaneous topology of the giant component of linked information sources defining the analog to a global broadcast of consciousness. That topology is the basic tunable syntactic filter across the underlying modular structure, and variation in ω is only one aspect of more general topological properties that can be described in terms of index theorems, where far more general analytic constraints can become closely linked to the topological structure and dynamics of underlying networks, and, in fact, can stand in place of them (Atyah and Singer, 1963; Hazewinkel, 2002).

Given a cognitive system by an information source X , in the context of a set of ‘environmental’ information sources Y_1, \dots, Y_k , we are particularly interested in the joint source defined by $H(X, Y_1, \dots, Y_k)$, and next examine some details of how the mutually embedded system might operate in real time, focusing on the role of rapidly-changing feedback information via an extension of the Data Rate Theorem. That is, for an ‘organism’ (or identifiable subsystem of one) interacting with an ‘environment’ (in a large sense that may include other subsystems of that organism), we are interested in dynamics for which $\omega \propto \mathcal{H}$.

7 Punctuated critical phenomena

The homology between the information source uncertainty dual to a cognitive process and the free energy density of a physical system arises from the formal similarity between their definitions in the asymptotic limit. Information source uncertainty can be defined as in the first part of equation (2). This is quite analogous to the free energy density of a physical system in terms of the thermodynamic limit of infinite volume (e.g., Wilson, 1971; Wallace, 2005). Feynman (2000) provides a series of physical examples, based on Bennett’s (1988) work, where this homology is an identity, at least for very simple systems. Bennett

argues, in terms of idealized 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.

It is possible to model a cognitive system interacting with an embedding environment using an extension of the language-of-cognition approach above. Recall that cognitive processes can be formally associated with information sources, and how a formal equivalence class algebra can be constructed for a complicated cognitive system by choosing different origin points in a particular abstract ‘space’ and defining the equivalence of two states by the existence of a high probability meaningful path connecting each of them to some defined origin point within that space.

Recall that disjoint partition by equivalence class is analogous to orbit equivalence relations for dynamical systems, and defines the vertices of a network of cognitive dual languages available to the system: each vertex represents a different information source dual to a cognitive process. The structure creates a large groupoid, with each orbit corresponding to a transitive groupoid whose disjoint union is the full groupoid, and each subgroupoid associated with its own dual information source. Larger groupoids will, in general, have ‘richer’ dual information sources than smaller.

We can now begin to examine the relation between system cognition and the feedback of information from the rapidly-changing real-time environment, \mathcal{H} , in the sense of equation (1).

With each subgroupoid G_i of the (large) cognitive groupoid we can associate a joint information source uncertainty $H(X_{G_i}, Y) \equiv H_{G_i}$, where X is the dual information source of the cognitive phenomenon of interest, and Y that of the embedding environmental context.

Real time dynamic responses of a cognitive system can now be represented by high probability paths connecting ‘initial’ multivariate states to ‘final’ configurations, across a great variety of beginning and end points. This creates a similar variety of groupoid classifications and associated dual cognitive processes in which the equivalence of two states is defined by linkages to the same beginning and end states. Thus, we will show, it becomes possible to construct a ‘groupoid free energy’ driven by the quality of rapidly-changing, real-time information coming from the embedding ecosystem, represented by the information rate \mathcal{H} , to be taken as a temperature analog.

The argument-by-abduction from physical theory is, then, that \mathcal{H} constitutes a kind of thermal bath for the processes of channeled cognition. Thus we can, in analogy with the standard approach from physics (Pettini, 2007; Landau and Lifshitz, 2007) construct a Morse Function by writing a pseudo-probability for the jointly-defined information sources X_{G_i}, Y having source uncertainty H_{G_i} as

$$P[H_{G_i}] = \frac{\exp[-H_{G_i}/\kappa\mathcal{H}]}{\sum_j \exp[-H_{G_j}/\kappa\mathcal{H}]} \quad (3)$$

where κ is an appropriate dimensionless constant characteristic of the particular system. The sum is over all possible subgroupoids of the largest available sym-

metry groupoid. Again, compound sources, formed by the (tunable, shifting) union of underlying transitive groupoids, being more complex, will have higher free-energy-density equivalents than those of the base transitive groupoids.

A possible Morse Function for invocation of Pettini’s topological hypothesis or Landau’s spontaneous symmetry breaking is then a ‘groupoid free energy’ F defined by

$$\exp[-F/\kappa\mathcal{H}] \equiv \sum_j \exp[-H_{G_j}/\kappa\mathcal{H}] \quad (4)$$

It is possible, using the free energy-analog F , to apply Landau’s spontaneous symmetry breaking arguments, and Pettini’s topological hypothesis, to the groupoid associated with the set of dual information sources.

Many other Morse Functions might be constructed here, for example based on representations of the cognitive groupoid(s). The resulting qualitative picture would not be significantly different.

Again, Landau’s and Pettini’s insights regarding phase transitions in physical systems were that certain critical phenomena take place in the context of a significant alteration in symmetry, with one phase being far more symmetric than the other (Landau and Lifshitz, 2007; Pettini, 2007). A symmetry is lost in the transition – spontaneous symmetry breaking. The greatest possible set of symmetries in a physical system is that of the Hamiltonian describing its energy states. Usually states accessible at lower temperatures will lack the symmetries available at higher temperatures, so that the lower temperature phase is less symmetric. The randomization of higher temperatures ensures that higher symmetry/energy states will then be accessible to the system. The shift between symmetries is highly punctuated in the temperature index.

The essential point is that decline in the richness of real-time environmental feedback \mathcal{H} , or in the ability of that feedback to influence response, as indexed by κ , can lead to punctuated decline in the complexity of cognitive process within the entity of interest, according to this model.

This permits a Landau-analog phase transition analysis in which the quality of incoming information from the embedding ecosystem – feedback – serves to raise or lower the possible richness of cognitive response to patterns of challenge. If $\kappa\mathcal{H}$ is relatively large – a rich and varied real-time environment – then there are many possible cognitive responses. If, however, noise or simple constraint limit the magnitude of $\kappa\mathcal{H}$, then behavior collapses in a highly punctuated manner to a kind of ground state in which only limited responses are possible, represented by a simplified cognitive groupoid structure.

These results represent a significant generalization of the Data Rate Theorem, as expressed in equation (1).

8 Renormalization

Certain details of information phase transitions can be calculated using ‘biological’ renormalization methods (Wallace, 2005, Section 4.2) analogous to, but

much different from, those used in the determination of physical phase transition universality classes (Wilson, 1971).

Given F as a free energy analog, what are the transitions between adiabatic realms? Suppose, in classic manner, it is possible to define a characteristic ‘length’, say r , on the system, as described in the Mathematical Appendix. It is then possible to define renormalization symmetries in terms of the ‘clumping’ transformation, so that, for clumps of size R , in an external ‘field’ of strength J (that can be set to 0 in the limit), one can write, in the usual manner (e.g., Wilson, 1971)

$$\begin{aligned} F[Q(R), J(R)] &= f(R)F[Q(1), J(1)] \\ \chi(Q(R), J(R)) &= \frac{\chi(Q(1), J(1))}{R} \end{aligned} \quad (5)$$

where χ is a characteristic correlation length and Q is an ‘inverse temperature measure’, i.e., $\propto 1/\kappa\mathcal{H}$.

As described in Wallace (2005), very many ‘biological’ renormalizations, $f(R)$, are possible that lead to a number of quite different universality classes for phase transition. Indeed, a ‘universality class tuning’ can be used as a tool for large-scale regulation of the system. While Wilson (1971) necessarily uses $f(R) \propto R^3$, following Wallace (2005), it is possible to argue that, since F is a kind of information measure, it is likely to ‘top out’ at different rates with increasing system size, so other forms of $f(R)$ must be explored. Indeed, standard renormalization calculations for $f(R) \propto R^\delta$, $m \log(R) + 1$, and $\exp[m(R-1)/R]$ all carry through.

9 Discussion and conclusions

The argument leading to equations (4) and (5) significantly extends the Data Rate Theorem of equation (1) in the context of cognitive systems, including those of a living state characterized – even defined – by cognition at every scale and level of organization. The unification of information theory and control theory via the Data Rate Theorem is not additive, but synergistic, encompassing a new body of statistical models that, in the sense of Dretske (1994), greatly constrain the possible dynamics of the living state. That is, in addition to providing important new conceptual approaches, this emerging body of theory may permit construction of new analytic tools that, like regression models, can provide benchmarks against which to compare experimental or observational data (e.g., Wallace, 2014b, Wallace and Wallace, 2014).

To put it another way, as has been long understood in robotics, walking across a crowded room, an exercise in embodied cognition, is far more difficult than playing a good game of chess (e.g., Brooks, 1991; Wilson and Golonka, 2013). Wallace (2012b) argues that a similar paradox exists at the cellular level, involving the many ‘glycosynapses’ (Cohen and Varki, 2010) operating at a literally astronomical number of cell surfaces in higher animals. The very concept of ‘information that controls the living state’ is, in fact, a misnomer,

since the living state is cognitive at every scale and level of organization, and cognition, inherently involving active choices that reduce uncertainty, is characterized by nested, interacting information sources. This fact places necessary condition restrictions on the dynamics of life that are expressed in the Shannon Coding and Source Coding Theorems, the Rate Distortion Theorem, and the new Data Rate Theorem.

These are not trivial matters, either conceptually or formally. It took nearly half a century to unify information and control theories, and further development of useful mathematical tools is not likely to be less arduous. Even filling in the groupoid/Morse Theory outline presented here will not be easy. That outline, however, holds great promise across a spectrum of cognitive disciplines (e.g., Wallace, 2014c, d).

10 Mathematical appendix

In order to define the metric r used in equation (5), impose a topology on the system, so that, near a particular ‘language’ A defining some H_G there is (in an appropriate sense) an open set U of closely similar languages \hat{A} , such that $A, \hat{A} \subset U$.

Since the information sources are ‘similar’, for all pairs of languages A, \hat{A} in U , it is possible to:

1. Create an embedding alphabet which includes all symbols allowed to both of them.
2. Define an information-theoretic distortion measure in that extended, joint alphabet between any high probability (grammatical and syntactical) paths in A and \hat{A} , written as $d(Ax, \hat{A}x)$ (Cover and Thomas, 2006). Note that these languages do not interact in this approximation.
3. Define a metric on U , for example,

$$r(A, \hat{A}) = \left| \lim \frac{\int_{A, \hat{A}} d(Ax, \hat{A}x)}{\int_{A, A} d(Ax, Ax)} - 1 \right| \quad (6)$$

using an appropriate integration limit argument over the high probability paths. Note that the integration in the denominator is over different paths within A itself, while in the numerator it is between different paths in A and \hat{A} . Consideration suggests r is indeed a formal metric.

Other approaches to metric construction on U are possible, as are other approaches to renormalization and phase transition.

References

- Arrell, D., & Terzic, A., (2010). Network systems biology for drug discovery. *Clinical Pharmacology and Therapeutics*, 88, 120-125.
- Ash, R., (1990). *Information Theory*. Dover Publications.

- Atiyah, M., & Singer, I., (1963). The index of elliptical operators on compact manifolds. *Bulletin of the American Mathematical Society*, 69, 322-433.
- Atlan, H., & Cohen, I., (1998). Immune information, self-organization, and meaning. *International Immunology*. 10, 711-717.
- Baars, B., (1988). *A Cognitive Theory of Consciousness*. Cambridge University Press.
- Baillieu, J., (2001). Feedback designs in information based control. In Pasik-Duncan, B. (ed.), *Stochastic Theory and Control: Proceedings of a Workshop Held in Lawrence, Kansas*. Springer, pp.35-57.
- Beck, C., & Schlogl, F., (1995). *Thermodynamics of Chaotic Systems*. Cambridge University Press.
- Bennett, C., (1988). Logical depth and physical complexity. In Herkin R., (ed.), *The Universal Turing Machine: A Half-Century Survey*, pp. 227-257, Oxford University Press.
- Bingham, G., (1988). Task-specific devices and the perceptual bottleneck. *Human Movement Science*. 7, 225-264.
- Brooks, R., (1991). Intelligence without representation. *Artificial Intelligence*. 47, 139-159.
- Brown, R., (1987). From groups to groupoids: a brief survey. *Bulletin of the London Mathematical Society*. 19, 113-134.
- Cohen, M., & Varki, A., (2010). The sialome – far more than the sum of its parts. *OMICS*. 14, 455-464.
- Cover, T., & Thomas, J., (2006). *Elements of Information Theory, 2nd Edition*. John Wiley & Sons.
- Csete, M., & Doyle, J., (2002). Reverse engineering of biological complexity. *Science*. 295, 1664-1669.
- Desjarlais, R. et al. (1995), *World Mental Health*. Oxford University Press.
- Dembo, A., & Zeitouni, O., (1998). *Large Deviations: Techniques and Applications, 2nd ed.*. Springer.
- Dretske, F., (1994). The explanatory role of information. *Philosophical Transactions of the Royal Society A*. 349, 59-70.
- English, T., (1996). Evaluation of evolutionary and genetic optimizers: no free lunch. In Fogel, L., Angeline, P., Back T., (eds.), *Evolutionary Programming V: Proceedings of the Fifth Annual Conference on Evolutionary Programming*. MIT Press, pp. 163-169.
- Feynman, R., (2000). *Lectures on Computation*. Westview Press.
- Hazewinkel, M., (2002). Index Formulas, *Encyclopedia of Mathematics*. Springer.
- Khinchin, A., (1957). *The Mathematical Foundations of Information Theory*. Dover Publications.
- Landau, L., & Lifshitz, E., (2007). *Statistical Physics, 3rd Edition, Part I*. Elsevier.
- Matsumoto, Y., (2002). *An Introduction to Morse Theory*. American Mathematical Society.
- Maturana, H., Varela, F., (1980). *Autopoiesis and Cognition*. Reidel.

- Minero, P., Franceschetti, M., Dey, S., Nair, G., (2009). Data Rate Theorem for stabilization over time-varying feedback channels. *IEEE Transactions on Automatic Control*. 54, 243-255.
- Mitter, S., (2001). Control with limited information. *European Journal of Control*. 7, 122-131.
- Nair, G., Fagnani, F., Zampieri, S., Evans, R., (2007). Feedback control under data rate constraints: an overview. *Proceedings of the IEEE*. 95, 108-137.
- Pettini, M., (2007). *Geometry and Topology in Hamiltonian Dynamics and Statistical Mechanics*. Springer.
- Sahai, A., (2004). The necessity and sufficiency of anytime capacity for control over a noisy communication link. *Decision and Control, 43rd IEEE Conference on CDC, Vol. 2*, 1896-1901.
- Sahai, A., & Mitter, S., (2006). The necessity and sufficiency of anytime capacity for control over a noisy communication link Part II: vector systems, <http://arxiv.org/abs/cs/0610146>.
- Shannon, C., (1959). Coding theorems for a discrete source with a fidelity criterion. *Institute of Radio Engineers International Convention Record 7*, 142-163.
- Tatikonda, S., & Mitter, S., (2004). Control over noisy channels. *IEEE Transactions on Automatic Control*. 49, 1196-1201.
- Touchette, H., & Lloyd, S., (2004). Information-theoretic approach to the study of control systems. *Physica A*. 331, 140-172.
- 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.
- Wallace, R., (2007). Culture and inattentive blindness: a global workspace perspective. *Journal of Theoretical Biology*. 245, 378-390.
- Wallace, R., (2012a). Consciousness, crosstalk, and the mereological fallacy: an evolutionary perspective. *Physics of Life Reviews*. 9, 426-453.
- Wallace, R., (2012b). Extending Tlusty's rate distortion index theorem method to the glycome: Do even 'low level' biochemical phenomena require sophisticated cognitive paradigms? *BioSystems*. 107, 145-152.
- Wallace, R., (2014a). Cognition and biology: perspectives from information theory. *Cognitive Processing*. 15, 1-12.
- Wallace, R., (2014b). A new formal perspective on 'Cambrian explosions'. *Comptes Rendus Biologies*. 337, 1-5.
- Wallace, R., (2014c). Canonical failure modes of real-time control systems. *International Journal of Systems Science*. In press.
- Wallace, R., (2014d). Embodied cognition, embodied regulation, and the Data Rate Theorem. *Biologically Inspired Cognitive Architectures*. In press.
- Wallace, R., & Wallace, D., (2010). *Gene Expression and its Discontents: The Social Production of Chronic Disease*. Springer.

Wallace, R., & Wallace, D., (2013). *A Mathematical Approach to Multilevel, Multiscale Health Interventions: Pharmaceutical industry decline and policy response*. Imperial College Press.

Wallace, R., & Wallace, R.G., (2014). Blowback: new formal perspectives on agriculturally driven pathogen evolution and spread. *Epidemiology and Infection*. doi: 10.1017/S0950268814000077.

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

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

Wilson, A., & Golonka, A., (2013). Embodied cognition is not what you think it is. *Frontiers in Psychology*. doi: 10.3389/fpsyg.2013.00058.

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

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

Wong, W., & Brockett, R., (1999). Systems with finite communication bandwidth constraints II: stabilization with limited information feedback. *IEEE Transactions on Automation and Control*. 44, 1049-1053.

You, K., & Xie, L., (2013). Survey of recent progress in networked control systems. *Acta Automatica Sinica*. 39, 101-117.

Yu, S., & Mehta, P., (2010). Bode-like fundamental performance limitations in control of nonlinear systems. *IEEE Transactions on Automatic Control*. 55, 1390-1405.