Is consciousness inherently unstable? An iterated Data Rate Theorem model of high metabolic demand in neural tissues

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

July 19, 2015

Abstract

Evolutionary process has selected for inherently unstable physiological systems in higher animals that can react swiftly to patterns of threat or affordance, for example blood pressure and the immune response. However, these require ongoing strict regulation: unregulated blood pressure is fatal, and immune cells can attack 'self' tissues. Consciousness, perhaps the most sophisticated rapid large-scale neural process, demands high rates of metabolic free energy to both operate and regulate the basic physiological machinery. Both the 'stream of consciousness' and the 'riverbanks' that confine it to useful realms are constructed and reconstructed moment-by-moment in response to highly dynamic internal and environmental circumstances. Using an information bottleneck method that links control and information theories, it is relatively easy to show that rapid response based on instability and its stabilization will always require high rates of metabolic free energy. In sum, neural structures in higher animals are highly 'coevolutionary', responding both to environmental signals and to signals from other physiological systems, and stabilizing coevolutionary cognitive structures is as difficult as programming them. Consciousness appears fundamentally unstable, and the necessary synergism between conscious action and its regulation underlies the ten-fold higher rate of metabolic energy consumption in human neural tissues. Implications for the etiology of certain psychiatric disorders are obvious.

Key Words: control system, Data Rate Theorem, information bottleneck, metabolic free energy, rate distortion theorem

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

1 Introduction

Why do neural tissues in humans consume metabolic free energy at ten times the rate of other tissues (Clarke and Sokoloff 1999)? The simplistic answer is that, in humans, consciousness must operate with a time constant of a hundred milliseconds, and straightforward adaptation of the Arrhenius reaction rate law suggests an exponential increase of neural reaction rate with the rate of adenosine triphosphate (ATP) consumption. While this surely accounts for substantial increase in energy demand, as we will show below, an order of magnitude difference in energy consumption seems somewhat excessive. Is more going on? Here, we will argue that, in addition to the matter of a short time constant, metabolic free energy must be supplied to regulatory systems that stabilize consciousness as an inherently unstable phenomenon constrained by the Data Rate Theorem, necessitating independent metabolic free energy supplies for control purposes.

A control theory digression. At the beginning of World War II, according to report, British fighters were close adaptations of training aircraft that were inherently stable, in the sense that the aerodynamic center of pressure (CP) was well behind the machine's center of gravity (CG). Thus, hands-off, a perturbed aircraft would, after a few oscillations, return to stable flight. Early German fighters had less separation between CG and CP and were far harder to fly, but, in consequence, could turn on a dime, and were significantly better in high-speed combat maneuvers than early British fighters, with serious consequences for the Allies. Current fighter jets are, by contrast, inherently unstable, thus even more highly maneuverable, and must be flown-by-wire using three independent computers which operate controls using high-speed majority-voting rules. For such machines, regulation is everything.

Indeed, for many physiological systems, regulation is likewise – almost – everything. Three examples.

Even our basic multicellularity seems inherently unstable: cancerous 'cheating' is expected to be an ongoing threat to multi-celled organisms (Aktipis et al. 2015). Nunney (1999) looked at cancer occurrence as a function of animal size, arguing that, in larger animals, whose lifespan grows as about the 4/10 power of their cell count, prevention of cancer in rapidly proliferating tissues becomes more difficult in proportion to size. Cancer control requires the development of additional mechanisms and systems to address tumorigenesis as body size increases – a synergistic effect of cell number and organism longevity. This pattern may represent a real barrier to the evolution of large, long-lived animals, and Nunney predicts that those that do evolve have recruited additional controls over those of smaller animals to prevent cancer. Different tissues may have evolved markedly different tumor control strategies, all energetically expensive, using different complex signaling strategies, and subject to a multiplicity of reactions to signals, including, in social animals like humans, those related to psychosocial stress.

The immune system, seen as an independent subcomponent of the more general tumor control system (Atlan and Cohen 1998), is also inherently unstable:

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1236v2 | CC-BY 4.0 Open Access | rec: 21 Jul 2015, publ: 21 Jul 2015

failure of differentiation between 'self' and 'nonself' leads to carcinogenic chronic inflammation (Rakoff-Nahoum 2006) and autoimmune disorders (Mackey and Rose 2014). The immune system must, then, both respond quickly to injury or pathogenic challenge and yet be closely regulated to avoid self-attack.

Unregulated blood pressure would be quickly fatal in any animal with a circulatory system. The associated baroreceptor control reflex is not simple (Rau and Elbert 2001), but can be inhibited through peripheral processes, for example under conditions of high metabolic demand. Higher brain structures modulate the reflex, for instance, when threat is detected and fight or flight responses are being prepared. This suggests, then, that blood pressure control is a broad and actively regulated modular physiological system.

The stream of consciousness, or whatever metaphor one prefers, seems similarly regulated, and high speed mechanisms, in concert with high speed regulators, will require a high rate of metabolic free energy.

2 Reaction rate

Physiological processes such as wound healing, the immune response, tumor control, and animal consciousness all represent the evolutionary exaptation of inevitable information crosstalk into dynamic processes that recruit sets of simpler cognitive modules to build temporary working coalitions that address particular patterns of threat and opportunity confronting an organism (Wallace 2012). Such tunable coalitions operate, however, at markedly different rates. Wound healing, depending on the extent of injury, may take 18 months to complete its work (Mindwood et al 2004). Animal consciousness typically operates with a time constant of a few hundred milliseconds. How can phenomena acting on such different rates be subsumed under the same underlying mechanism? Adaptation of Arrhenius' law (Laidler 1987), which predicts exponential differences in reaction rate with 'temperature', in a large sense, produces a first approximation to the result, recognizing that cognitive phenomena are inherently nonequilibrium. That is, a large class of cognitive processes can be associated with dual information sources (Wallace 2012) for which palindromes are highly improbable. The rate of biocognition, however, appears exponentially driven by the rate of available metabolic free energy as a temperature analog.

The energetics of biological reactions are remarkable: at 300 K, molecular energies represent approximately 2.5 KJ/mol in available free energy. By contrast, the basic biological energy reaction – the hydrolysis of adenosine triphosphate (ATP) to adenosine diphosphate – under proper conditions at 300 K, produces some 50 KJ/mol. in reaction energy. This is equivalent to a 'reaction temperature' of 6000 K. Increasing the rate of ATP delivery to one kind of tissue an order of magnitude over any others provides sufficient energy for very rapid biocogniton.

The question is how such rapid biocognition is parceled out between consciousness itself and the mechanisms that must regulate and stabilize it.

In more detail, given a chemical reaction of the form $aA+bB \rightarrow pP+qQ$, the

rate of change in (for example) the concentration of chemical species P (written $[P]$) is often determined by an equation like

$$d[P]/dt = k(T)[A]^n[B]^m \tag{1}$$

where n and m depend on the reaction details. The rate constant k is expressed by the Arrhenius relation as

$$k = \alpha \exp[-E_a/RT] \tag{2}$$

where α is another characteristic constant, E_a is the reaction activation energy, T is the Kelvin temperature and R a universal constant. $\exp[-E_a/RT]$ is, using the Boltzmann distribution, the fraction of molecular interactions having energy greater than E_a.

The inherently nonequilibrium nature of cognition, however, requires a slightly more sophisticated treatment. Following Wallace (2005, 2012), cognition can often be associated with a dual information source. Consciousness appears to be largely an all-or-nothing phenomenon (Sergeant and Dehaene 2004), so that conscious signal perception must exceed a threshold before becoming entrained into the characteristic general broadcast.

A direct information theory argument focuses on the Rate Distortion Function (RDF) $R(D)$ associated with the channel connecting the cognitive individual with an embedding and embodying environment. $R(D) \geq 0$, a convex function (Cover and Thomas 2006), defines the minimum rate of information transmission needed to ensure that the average distortion between what is sent and what is received is less than or equal to $D \geq 0$, according to an appropriate distortion measure. Assuming a threshold R_0 for conscious perception of an incoming signal, we can, following Feynman's (2000) identification of information as a form of free energy, write a Boltzmann-like probability for the rate of cognition as

$$P[R \geq R_0] = \frac{\int_{R_0}^{\infty} \exp[-R/\omega M]dR}{\int_0^{\infty} \exp[-R/\omega M]dR} = \exp[-R_0/\omega M] \tag{3}$$

where M is the supplied rate of metabolic free energy, ω a constant (representing entropic loss-in-translation), and figure 1 follows.

If we define an efficiency measure as (cognition rate)$/M$, we see that energy efficiency peaks at a relatively low cognition rate. Indeed, a simple calculation shows that $\exp[-k/M]/M$ has its maximum at $M = k$. The form of $\exp[-k/M]/M$ is distinct, and closely similar to what has been found in other recent work. Using numerical models of optimal coding and information transmission in Hodgkin-Huxley neurons under metabolic constraints, Kostal and Kobayashi (2015) find an almost exactly similar efficiency curve. Their treatment, however, goes beyond ours and involves regimes determined by the critical value of the effective reversal potential of their neural model.

For mammals, since body temperature remains constant, the rate of available metabolic free energy – dependent on mitochondrial function – serves as a

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1236v2 | CC-BY 4.0 Open Access | rec: 21 Jul 2015, publ: 21 Jul 2015

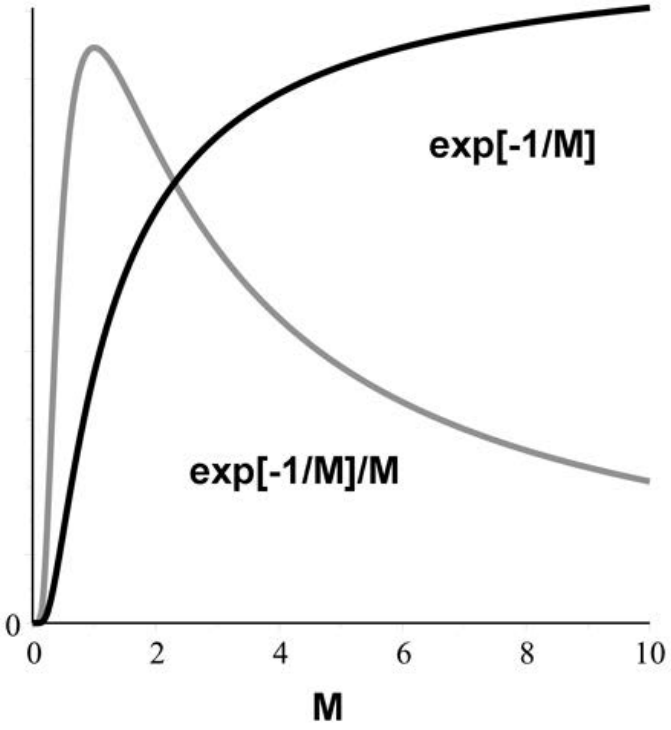

Figure 1: 'Arrhenius' relation for rate of cognition as a function of the rate of available metabolic free energy M, along with the efficiency measure of rate per unit metabolic energy. An order of magnitude increase in such free energy can enable several orders of magnitude increase in the rate of cognition, although the point of greatest efficiency is at relatively low values of M: it is easy to show that the efficiency measure $\exp[-k/M]/M$ has its maximum at $M = k$. Decline in M below the shoulder of the curve triggers catastrophic collapse of cognition.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1236v2 | CC-BY 4.0 Open Access | rec: 21 Jul 2015, publ: 21 Jul 2015

temperature index for rates of biocognition. This determines the characteristic rate of chemically-generated consciousness, or of the individual lower-level cognitive modules that come together in a temporary assemblage to form such an analog. Neural tissues, in humans consuming metabolic energy at an order of magnitude greater rate than other tissues, thus can provide cognitive function many orders of magnitude faster than similar physiological phenomena.

But is this the whole story? Kostal and Kobayashi (2015) argue that efficiency matters in neural process, so that regimes of lower energy consumption may be favored over the highest cognitive rates. But, as Ristroph et al. (2013) argue, inherent instability, in itself, allows extremely rapid responses that have been strongly selected for. Here, we will argue that regulation of such phenomena must consume significant metabolic free energy, in addition to that needed for (relatively) rapid cognition.

How do we understand the regulation of inherently unstable control systems? Two fundamental relations, the Data Rate and Rate Distortion Theorems, are a necessary foundation. Their convolution, we shall show, provides further insight on metabolic energy demands of high-speed cognition.

3 The Data Rate Theorem

The Data Rate Theorem (DRT), based on an extension of the Bode Integral Theorem for linear control systems, describes the stability of feedback control under data rate constraints (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 $x_{t+1} = \mathbf{A}x_t + ...$, where x_t is the n-dimensional state vector at time t, then the necessary condition for stabilizability is that

$$\mathcal{H} > \log[|det\mathbf{A}^u|] \tag{4}$$

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 determinant represents a generalized volume. Thus there is a critical positive data rate below which there does not exist any quantization and control scheme able to stabilize an unstable system (Nair et al. 2007).

The new theorem, in its various forms, relates control theory to information theory and is as fundamental as the Shannon Coding and Source Coding Theorems, and the Rate Distortion Theorem for understanding complex biological phenomena.

Some thought suggests that, accepting Feynman's (2000) insight that information is simply a form of free energy, in biological circumstances, we can write that $M = m(\mathcal{H})$, where M is the rate of metabolic free energy used to generate the control information rate $\mathcal{H}$, and m is a sharply increasing monotonic function, a consequence of massive entropic losses necessarily associated with translation of metabolic energy to information. Equation (4) thus implies that

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1236v2 | CC-BY 4.0 Open Access | rec: 21 Jul 2015, publ: 21 Jul 2015

there is a minimum necessary rate of free energy consumption below which it is not possible to stabilize an inherently unstable biological control system. A little more calculation, however, provides a much deeper result.

4 The Rate Distortion Theorem

Suppose a sequence of signals is generated by a biological information source Y having output $y^n = y_1, y_2, ...$ This is 'digitized' in terms of the observed behavior of the system with which it communicates, for example a sequence of 'observed behaviors' $b^n = b_1, b_2, ...$ Assume each b^n is then deterministically retranslated back into a reproduction of the original biological signal, $b^n \rightarrow \hat{y}^n = \hat{y}_1, \hat{y}_2, ...$

Define a distortion measure $d(y, \hat{y})$ comparing the original to the retranslated path. Many distortion measures are possible. For example, the Hamming distortion is defined simply as $d(y, \hat{y}) = 1, y \neq \hat{y}, d(y, \hat{y}) = 0, y = \hat{y}$.

For continuous variates, the squared error distortion measure is just $d(y, \hat{y}) = (y - \hat{y})^2$.

There are many possible distortion measures. The distortion between paths y^n and $\hat{y}^n$ is defined as

$$d(y^n, \hat{y}^n) \equiv \frac{1}{n} \sum_{j=1}^{n} d(y_j, \hat{y}_j) \tag{5}$$

A remarkable characteristic of the Rate Distortion Theorem is that the basic result is independent of the exact distortion measure chosen (Cover and Thomas 2006). We shall iterate the Data Rate Theorem via the information bottleneck method of Tishby et al. (1999), and use $\mathcal{H}$ of the Data Rate Theorem as the distortion measure.

Suppose that with each path y^n and b^n-path retranslation into the y-language, denoted $\hat{y}^n$, there are associated individual, joint, and conditional probability distributions $p(y^n), p(\hat{y}^n), p(y^n, \hat{y}^n), p(y^n|\hat{y}^n)$.

The average distortion is defined as

$$D \equiv \sum_{y^n} p(y^n) d(y^n, \hat{y}^n) \tag{6}$$

It is possible to define the information transmitted from the Y to the $\hat{Y}$ process using the Shannon source uncertainty of the strings:

$$I(Y, \hat{Y}) \equiv H(Y) - H(Y|\hat{Y}) = H(Y) + H(\hat{Y}) - H(Y, \hat{Y}) \tag{7}$$

where $H(..., ...)$ is the joint, and $H(...|...)$ the conditional, Shannon uncertainties (Cover and Thomas 2006).

If there is no uncertainty in Y given the retranslation $\hat{Y}$, then no information is lost, and the systems are perfectly synchronous.

This will almost never be true.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1236v2 | CC-BY 4.0 Open Access | rec: 21 Jul 2015, publ: 21 Jul 2015

The rate distortion function $R(D)$ for a source Y with a distortion measure $d(y, \hat{y})$ is defined as

$$R(D) = \min_{p(y,\hat{y}); \sum_{(y,\hat{y})} p(y)p(y|\hat{y})d(y,\hat{y}) \leq D} I(Y, \hat{Y}) \tag{8}$$

The minimization is over all conditional distributions $p(y|\hat{y})$ for which the joint distribution $p(y, \hat{y}) = p(y)p(y|\hat{y})$ satisfies the average distortion constraint of having average distortion $\leq D$.

The Rate Distortion Theorem states that $R(D)$ is the minimum necessary rate of information transmission which ensures the communication between the biological vesicles does not exceed average distortion D. Thus $R(D)$ defines a minimum necessary channel capacity. Cover and Thomas (2006) or Dembo and Zeitouni (1998) provide details. The rate distortion function has been calculated for a number of systems, often using Lagrange multiplier or Khun-Tucker optimization methods.

Cover and Thomas (2006, Lemma 13.4.1) show that $R(D)$ is necessarily a decreasing convex function of D for any reasonable definition of distortion. That is, $R(D)$ is always a reverse J-shaped curve. This will prove crucial for the overall argument: convexity is an exceedingly powerful mathematical condition, and permits deep mathematical inference (Rockafellar 1970, Ellis (1985, Ch. VI). This is, indeed, the point from which all else follows. We will use the Gaussian channel as an easily calculated example, but the central results are quite general, and will drive the final argument.

For the standard Gaussian channel, having white noise with zero mean and variance σ^2, and using the squared distortion measure,

$$\begin{aligned} R(D) = 1/2 \log[\sigma^2/D], 0 \leq D \leq \sigma^2 \\ R(D) = 0, D > \sigma^2 \end{aligned} \tag{9}$$

5 Elementary Rate Distortion dynamics

Following Wallace (2015a), for the Gaussian channel, we can define a 'Rate Distortion entropy' as the Legendre transform

$$S_R = R(D) - DdR(D)/dD = 1/2 \log[\sigma^2/D] + 1/2 \tag{10}$$

The simplest nonequilibrium Onsager equation (de Groot and Mazur 1984) is

$$dD/dt = -\mu dS_R/dD = \mu/2D \tag{11}$$

where t is the time and μ is a diffusion coefficient. This has the solution

$$D(t) = \sqrt{\mu t} \tag{12}$$

which is the classic outcome of the diffusion equation. Such correspondence reduction serves as the foundation for arguing upward in both scale and complexity.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1236v2 | CC-BY 4.0 Open Access | rec: 21 Jul 2015, publ: 21 Jul 2015

Regulation, however, does not involve the diffusive drift of average distortion. Let M be the rate of metabolic free energy available for such regulation. Then a plausible model, in the presence of an internal system noise β^2 in addition to the environmental channel noise defined by σ^2, is the stochastic differential equation

$$dD_t = (\frac{\mu}{2D_t} - F(M))dt + \frac{\beta^2}{2} D_t dW_t \tag{13}$$

where dW_t represents unstructured white noise and $F(M) \geq 0$ is a monotonically increasing function in the rate of metabolic free energy M.

This relation has the nonequilibrium steady state expectation

$$D_{nss} = \frac{\mu}{2F(M)} \tag{14}$$

Using the Ito chain rule on equation (13) (Protter 1990; Khasminskii 2012), it is possible to calculate the variance in the distortion as $E(D_t^2) - (E(D_t))^2$.

Letting $Y_t = D_t^2$ and applying the Ito relation,

$$dY_t = [2\sqrt{Y_t}(\frac{\mu}{2\sqrt{Y_t}} - F(M) + \frac{\beta^4}{4} Y_t]dt + \beta^2 Y_t dW_t \tag{15}$$

where $(\beta^4/4)Y_t$ is the Ito correction to the time term of the SDE.

No real number solution for the expectation of $Y_t = D_t^2$ is possible unless the discriminant of the resulting quadratic equation is ≥ 0, producing a minimum necessary rate of available metabolic free energy for regulatory stability defined by

$$F(M) \geq \frac{\beta^2}{2}\sqrt{\mu} \tag{16}$$

Values of $F(M)$ below this limit will trigger a phase transition into a disintegrated, pathological, system dynamic in a highly punctuated manner. Wallace (2015a) uses a Black-Scholes model to calculate the form of $F(M)$, and to solve for M in terms of system parameters. Applying the inverse of the function F to equation (16) gives a slightly different form of the Data Rate Theorem, in terms of minimum necessary metabolic free energy. Similar models can be constructed using the 'natural' channel having the Rate Distortion Function $R(D) = \beta/D$.

We can extend the Data Rate Theorem by iterating the argument for any convex Rate Distortion Function, via the information bottleneck method.

6 A Data Rate index theorem model

The Data Rate Theorem states that there is a minimum necessary rate of control information needed to stabilize an inherently unstable system. Is this, in itself, a stable condition? That is, once a stabilizing control information rate has been identified, is that the whole story, or are there other dynamic processes to consider? A control system has at least three components: the structure to be controlled, the mechanism for control, and the underlying 'program' of that

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1236v2 | CC-BY 4.0 Open Access | rec: 21 Jul 2015, publ: 21 Jul 2015

mechanism. The 'mechanism', in our case, is the interacting set of high-speed neural systems that becomes the global broadcast of consciousness.

An approach to the dynamics of control stability in such a system – and the rate of metabolic free energy required – is possible using a variant of the information bottleneck of Tishby et al. (1999).

We envision an iterated application of the Rate Distortion Theorem to a control system in which a series of 'orders' $y^n = y_1, ..., y_n$, having probability $p(y^n)$, is sent through and the outcomes monitored as $\hat{y}^n = \hat{y}_1, ..., \hat{y}_n$. The distortion measure, however, is now taken as the minimum necessary control information $\mathcal{H}(y^n, \hat{y}^n)$, defining an average 'distortion' $\hat{\mathcal{H}}$ as

$$\hat{\mathcal{H}} = \sum_{y^n} p(y^n)\mathcal{H}(y^n, \hat{y}^n) \geq 0 \tag{17}$$

We can then define a new, iterated, Rate Distortion Function $\mathcal{R}(\hat{\mathcal{H}})$ and a new 'entropy' as

$$\mathcal{S} = \mathcal{R}(\hat{\mathcal{H}}) - \hat{\mathcal{H}} d\mathcal{R}/d\hat{\mathcal{H}} \tag{18}$$

We next invoke the analog to the 'diffusion' equation (11),

$$d\hat{\mathcal{H}}/dt = -\mu d\mathcal{S}/d\hat{\mathcal{H}} \tag{19}$$

where t is the time and μ a diffusion coefficient.

Since $\mathcal{R}$ is always a convex function of $\hat{\mathcal{H}}$ (Tishby et al. 1999; Cover and Thomas 2006), this relation has the solution

$$\hat{\mathcal{H}}(t) = f(t) \tag{20}$$

where $f(t)$ is a monotonic increasing function of t. Thus, in the absence of continuous regulation, in this model the needed control signal will relentlessly rise in time, surpassing all possible bounds, and hence triggering a failure to control an inherently unstable system.

The next stage involves generalization of equation (18) in the direction of equation (13) to allow calculation of a nonequilibrium steady state in a model for ongoing, continuous investment metabolic energy and other resources at a rate M in order to regulate the system.

As a further consequence of the convexity of the Rate Distortion Function, this too will have an expectation analogous to equations (14) and (16), so that – for the moment ignoring stochastic effects –

$$\hat{\mathcal{H}}_{nss} \propto 1/g(M) \tag{21}$$

where $g(M)$ is monotonic increasing in the metabolic energy rate M.

Thus the magnitude of the 'control distortion' $\hat{\mathcal{H}}_{nss}$ needed to stabilize high level animal cognition – consciousness – can be constrained by the delivery of regulatory metabolic free energy at a sufficiently high rate, in addition to the rate of metabolic free energy necessary to operate the underlying machinery of consciousness itself.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1236v2 | CC-BY 4.0 Open Access | rec: 21 Jul 2015, publ: 21 Jul 2015

Since we do not know the mathematical form of $\mathcal{R}(\hat{\mathcal{H}})$, we cannot carry out the calculation leading to equation (16) or the explicit results of Wallace (2015a), based on the assumption of a Gaussian channel.

We can, however, take stochastic effects into account and generalize equation (19) in the direction of equation (13) to derive an index theorem, in the sense of Atiyah and Singer (1963) and Hazewinkel (2002). An index theorem is an analytic relation whose solutions represent different topological modes of an underlying manifold, in a large sense. The argument is straightforward. Imposing white noise on the system, equation (19) becomes

$$d\hat{\mathcal{H}}_t = (-\mu d\mathcal{S}/d\hat{\mathcal{H}}_t - F(M))dt + \frac{\beta^2}{2}\hat{\mathcal{H}}_t dW_t \tag{22}$$

Setting the time-average expectation of $d\hat{\mathcal{H}}_t$ to zero

$$< d\hat{\mathcal{H}}_t >= 0 \tag{23}$$

defines the index theorem, whose multiple possible solutions represent quasi-stable modes of the regulatory system. These may range from simple fixed points to closed 'Red Queen' cycles or pseudorandom 'strange attractors' within a bounded region. Below, we will examine 'directed transitions' between such modes representing large deviations in the sense of Champagnat et al. (2006).

7 A phase change model

Indeed, more subtle regulatory behaviors, following Parker et al. (2002), seem likely to involve bifurcations or phase transitions, since biological systems are cognitive at every scale and level of organization, and cognition can often be represented in terms of groupoid symmetries that generalize the group structures familiar to the dynamics of simple physical systems (e.g., Wallace 2012). An outline of the argument is as follows.

Let $\mathcal{T} \equiv 1/g(M)$, so that the nonequilibrium steady state 'information cost' of stabilizing an inherently unstable (and inherently cognitive) biological system grows with the 'temperature' measure $\mathcal{T}$. This is, we argue, analogous to phase transitions arguments in statistical physics that use a Morse Function argument, as follows.

A pseudoprobability over the unstable, regulated cognitive system characterized by a groupoid $\{G_i\}$ can be constructed as

$$P[\hat{\mathcal{H}}_{G_i}] = \frac{\exp[-\hat{\mathcal{H}}_{G_i}/\kappa\mathcal{T}]}{\sum_j \exp[-\hat{\mathcal{H}}_{G_j}/\kappa\mathcal{T}]} \tag{24}$$

where κ is a characteristic constant. The sum is to be taken over all possible subgroupoids of the largest embedding cognitive symmetry groupoid.

A relatively simple Morse Function, leading to Pettini's (2007) topological hypothesis in this system, that can be built from this is the 'groupoid free

energy' $\mathcal{F}$ defined as

$$\exp[-\mathcal{F}/\kappa\mathcal{T}] \equiv \sum_j \exp[-\hat{\mathcal{H}}_{G_j}/\kappa\mathcal{T}] \tag{25}$$

Using $\mathcal{F}$, we impose a modified version of Landau's spontaneous symmetry breaking argument on the groupoid associated with the cognitive process regulated by the average control information cost $\hat{\mathcal{H}}$.

Following Pettini (2007), changes in $\mathcal{T}$, an inverse function of available levels of metabolic free energy, can lead to punctuated phase transition/bifurcation changes in the average control signal needed to stabilize an inherently unstable, but consequently highly responsive, dynamic biological system.

Failure of metabolic free energy supply – from developmental disorder to senescence – will lead to punctuated onset of dysfunction.

8 An optimization model

Something much like the multiple quasi-stable states implied by the developments leading to equations (23) and (25) can also be derived via an optimization argument applied to the rate calculation of equation (3). The essential point is that both consciousness and its necessary cognitive regulatory system(s) will follow similar metabolic scaling functions, so that we can seek to maximize a joint efficiency measure subject to constraint, applying the usual Lagrange multiplier argument. That is, letting the subscript C represent consciousness and R its regulatory machinery, we seek to maximize an efficiency functional

$$\frac{\exp[-k_C/M_C]}{M_C} + \frac{\exp[-k_R/M_R]}{M_R} \tag{26}$$

subject to the constraint

$$M_C + M_R = M \tag{27}$$

The k_X are appropriate constants and M_X is the metabolic free energy rate for process X.

Taking

$$\mathbf{\Lambda}(M_C, M_R, \lambda) \equiv \frac{\exp[-k_C/M_C]}{M_C} + \frac{\exp[-k_R/M_R]}{M_R} + \lambda(M_C + M_R - M) \tag{28}$$

gives the Lagrange optimization conditions as

$$\nabla_{M_C, M_R, \lambda} \mathbf{\Lambda} = 0 \tag{29}$$

The resulting complicated third order equation for solution pairs M_C, M_R implies the existence of several different possible (M-dependent) optimization points for the system. In a sense, this represents yet another form of index theorem relating solutions of an analytic equation to underlying topological modes.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1236v2 | CC-BY 4.0 Open Access | rec: 21 Jul 2015, publ: 21 Jul 2015

Extending the perspective somewhat, as is well known, there are a number of different stages to sleep, most notably NREM (non-rapid eye movement) which involves low rates of blood flow to the brain, and REM sleep which can rival or exceed conscious state blood flows. Again, sleep states must, like other neural processes, also be highly regulated, and similar arguments may well carry through.

Indeed, crudely, if we assume $k_C \approx k_R$ in equation (26), direct calculation shows a symmetric efficiency curve with equal peaks at two ends of the relation $M = M_C + M_R$, as in figure 2. There, taking $k_X = 1, M_C + M_R = M = 10$ we obtain maximum efficiency at the symmetric points $(M_C, M_R) = (1.03, 8.97), (8.97, 1.03)$, suggesting an on/off conscious/sleep mode for such a system in which 'sleep' may represent a parsimonious assumption of essential maintenance duties by systems otherwise dedicated to the regulation of awake consciousness. This is analogous to the immune system which, when not extinguishing the fires of infection, wound healing and malignancy, is deeply involved with processes of routine cellular maintenance (Cohen 2000). This suggests that sleep dysfunction may be very serious indeed.

Note that values of $k_C = k_R > 1$ generate much broader symmetric curves with far less well-defined peaks, while values less than 1 are much more sharply peaked. Unequal values raise or lower one or the other peak. Thus 'tuning' these parameters would provide significant added system control.

This argument can be generalized. The essence of the information bottleneck method is use of an information measure as a distortion parameter. From our viewpoint, however, metabolic free energy can also be taken as an information index, and we can redo part of the calculation above from that perspective. Let us define, then, a new 'distortion' parameter $\hat{M}_X, X = C, R$ using something like equation (17)

$$\hat{M}_X = \sum_{y^n} p(y^n) M_X(y^n, \hat{y}^n) \geq 0 \tag{30}$$

Now we can, as in the information bottleneck method, define new convex 'rate distortion functions' based on $\hat{M}_R$ and $\hat{M}_C$. The sum of the new RDF's also represents a free energy equivalent, constrained by the condition $M = \hat{M}_C + \hat{M}_R$. The energy equivalent of that RDF sum can then be written as a function of $\hat{M}_C$, and, some thought shows, will necessarily be bi-convex, i.e., U-shaped, in $\hat{M}_C$, since $\hat{M}_R = M - \hat{M}_C$. A horizontal line at M then defines the possible operating points of $\hat{M}_C$ at high and low values of the index.

The question arises, how is the transition made between such operating modes?

9 Transition dynamics

How does the consciousness/regulator structure – however we choose to characterize it – make changes between the quasi-stable states that the different modeling strategies above imply are central to the regulatory process? Indeed, recent primate experiments imply that even routine conscious decision-making

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1236v2 | CC-BY 4.0 Open Access | rec: 21 Jul 2015, publ: 21 Jul 2015

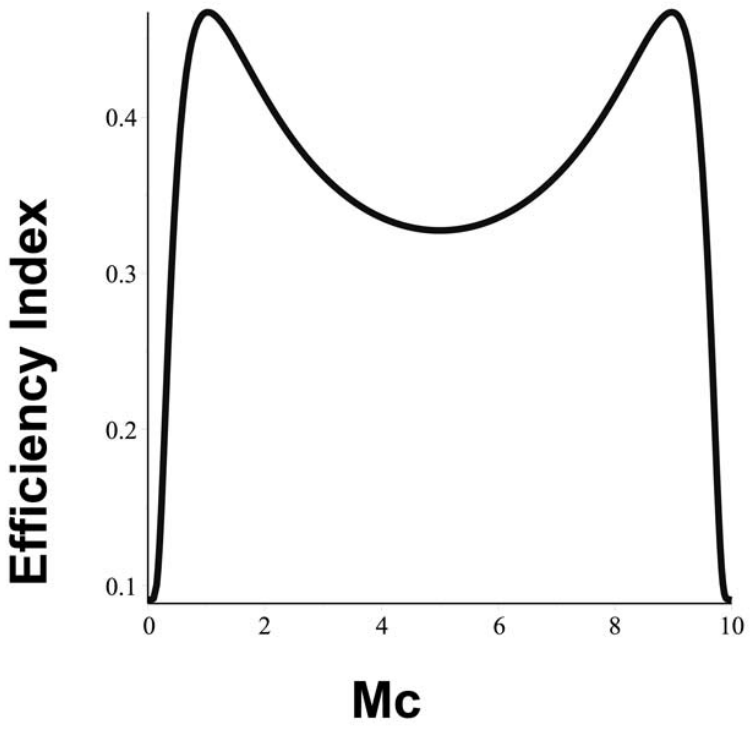

Figure 2: Total consciousness-and-regulator efficiency as a function of the metabolic free energy consumed by consciousness, M_C, for $k_X = 1, M_C + M_R = 10$. The maxima are at (1.03, 8.97) and (8.97, 1.03), suggesting sleep/awake modes for consciousness. During sleep, 'regulator' systems for consciousness may perform routine maintenance duties as does the immune system when not fighting fires of infection, wound healing, or malignancy. Changing the values of the k_X markedly shifts the relative heights and widths of the peaks, suggesting an added control mode.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1236v2 | CC-BY 4.0 Open Access | rec: 21 Jul 2015, publ: 21 Jul 2015

takes place in discrete steps (Latimer et al. 2015). Similar problems arise in evolutionary theory. Taking the approach of equations (22) and (23) in an evolutionary context, Champagnat et al. (2006) argue that the probability of a 'large deviation' driving the system from one quasi-stable mode to another is given by a negative exponential of an entropy-like function

$$\mathcal{I} = -\sum_j P_j \log[P_j] \tag{31}$$

where the P_j represent a particular probability distribution. This result – the large deviations argument – is well known in numerous contexts under various names – Sanov's Theorem, the Gartner/Ellis Theorem, etc. (Dembo and Zeitouni 1998). For the composite of human consciousness-and-regulation, we argue, the transition between 'states' involves the effect of impinging information sources. That is, $\mathcal{I}$ is not simply an 'entropy' in this case, but represents action of an external information source (or sources) that, iteratively, regulates the internal regulators controlling individual consciousness.

A variant on this kind of approach would, for the optimization model, make different values of the essential parameters M, k_C and k_R the outputs of another, embedding, information source.

10 Discussion and conclusions

Evolutionary process has selected for unstable control systems in higher animals that, among other things, can react swiftly to patterns of threat or affordance, but require ongoing strict regulation at different scales and levels of organization for their proper operation. Here, we have argued that consciousness, perhaps the most significant and sophisticated rapid large-scale neural process, must be supplied with high rates of metabolic free energy to both operate and regulate the basic physiological machinery. That is, both the 'stream of consciousness' and the 'riverbanks' that confine it to realms useful to the animal are constructed and reconstructed moment-by-moment in response to highly dynamic internal and environmental circumstances. High speed response, as the second part of this paper argues, requires considerable metabolic free energy.

Neural structures in higher animals are highly 'coevolutionary' in that they respond both to incoming signals and to signals from other neural systems. It has long been known that stabilizing coevolutionary computing systems is as inherently difficult as programming them (Wallace 2016). Here, we argue that consciousness is inherently unstable and that a necessary powerful synergism between conscious action and its regulation is a large underlying contributing factor to the ten-fold higher rate of metabolic energy consumption in human neural tissues.

Working out the full details of the energy consumption tradeoffs between rapid cognition and its necessary control machinery remains to be done, but punctuated transition seems inherent.

As has often been speculated, however, failure of regulation seems to underlie many psychiatric disorders.

Emotions, Thayer and Lane (2000) assert, are an integrative index of individual adjustment to changing environmental demands, an organismal response to an environmental event that allows rapid mobilization of multiple subsystems. Emotions allow the efficient coordination of the organism for goal-directed behavior. When the system works properly, it allows for flexible adaptation of the organism to changing environmental demands. An emotional response must be regulated to represent a proper selection of an appropriate response and the inhibition of other less appropriate responses from a more or less broad behavioral repertoire of possible responses. From their perspective, disorders of affect represent a condition in which the individual is unable to select the appropriate response, or to inhibit the inappropriate response, so that the response selection mechanism is somehow corrupted – regulation fails.

Gilbert (2001) similarly suggests that a canonical form of such corruption is the inappropriate excitation of modes that, in other circumstances, represent normal evolutionary adaptations, again representing a fundamental failure of regulation.

The formal development thus extends the perspective of Wallace (2015c) on the pathologies of mitochondrial dysfunction toward realms of psychiatric disorders.

However, atomistic, individual-scale regulation must be iterated to include social and cultural influences. 'Culture', to use the words of the evolutionary anthropologist Robert Boyd, 'is as much a part of human biology as the enamel on our teeth', and this leads to extensions of the transition arguments above: the principal environment of humans is other humans, and we are the naked mole rats of primates. Thus social interaction and cultural heritage tend to confine individual consciousness to realms leading to socially acceptable phenotypes. Failure of such constraint is then socially constructed as misbehavior or pathology.

References

Aktipis, C., A. Boddy, G. Jansen, U. Hibner, M. Hochberg. C. Maley, G. Wilkinson, 2015, Cancer across the tree of life: cooperation and cheating in multicellularity, Philosophical Transactions B 370:20140219.

Atiyah, M., I. Singer, 1963, The index of elliptical operators on compact manifolds, Bulletin of the American Mathematical Society, 69:322-433.

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

Baars, B., 1989, A Cognitive Theory of Consciousness, Cambridge University Press, New York (1989).

Champagnat, N., R. Ferriere, S. Meleard, 2006, Unifying evolutionary dynamics: from individual stochastic process to macroscopic models, Theoretical Population Biology, 69:297-321.

Clarke, D., L. Sokoloff, 1999, Circulation and energy metabolism of the brain, Chapter 31 in Siegel, G., et al. (eds.), Basic Neurochemistry: Molecular, Cellular and Medical Aspects, 6th edition, Lippincott-Raven, New York.

Cohen, I., 2000, Tending Adam's Garden: Evolving the cognitive immune self, Academic Press, New York.

Cover, T., J. Thomas, 2006, Elements of Information Theory, Wiley, New York.

Damasio, A., 1998, Emotion in the perspective of an integrated nervous system, Brain Research Reviews 26:83-86.

de Groot, S., P. Mazur, 1984, Non-Equilibrium Thermodynamics, Dover Publications, New York.

Dembo, A., O. Zeitouni, 1998, Large Deviations and Applications, Springer, New York.

Ellis, R., 1985, Entropy, Large Deviations, and Statistical Mechanics, Springer, New York.

Feynman, R., 2000, Lectures on Computation, Westview Press, Boulder, CO.

Gilbert, P., 2001, Evolutionary approaches to psychopathology: the role of natural defenses, Austrailian and New Zealand Journal of Psychiatry 35:17-27.

Hazewinkel¡ M., 2002, Encyclopedia of Mathematics, 'Index Formulas', Springer, New York.

Khasminskii, R., 2012, Stochastic Stability of Differential Equations, 2nd Edition, Springer, New York.

Kostal, L., R. Kobayashi, 2015, Optimal decoding and information transmission in Hodgkin-Huxley neurons under metabolic cost constraints, BioSystems, doi: 10.1016/j.biosystems.2015.06.008.

Laidler, K., 1987, Chemical Kinetics, 3rd ed., Harper and Row, New York.

Latimer, K., J. Yates, M. Meister A. Huk, J. Pillow, 2015, Single-trial spike trains in parietal cortex reveal discrete steps during decision-making, Science, 349:184-187.

Mackay, I., N. Rose (eds.), 2014, The Autoimmune Diseases, 5th Edition, Academic Press, New York.

Mindwood, K., et al., 2004, Tissue repair and the dynamics of the extracellular matrix, International Journal of Biochemistry and Cell Biology, 36:1031-1037 (2004).

Nair, G., F. Fagnani, S. Zampieri, R. Evans, 2007, Feedback control under data rate constraints: an overview, Proceedings of the IEEE 95:108-137.

Nunny, L., 1999, Lineage selection and the evolution of multistage carcinogenesis, Proceedings of the Royal Society B 266:493-498.

Parker, A., T. Gedeon, A. Dimitrov, 2002, Annealing and the rate distortion problem. In Thrun, S., S. Becker, K. Obermeyer (eds.), Advances in Neural Information Processing Systems, 15:969-976, MI Press, Cambridge MA.

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

Protter, 1990, Stochastic Integration and Differential Equations, Springer, New York.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1236v2 | CC-BY 4.0 Open Access | rec: 21 Jul 2015, publ: 21 Jul 2015

Rakoff-Nahoum, S., 2006, Why cancer inflammation? Yale Journal of Biology and Medicine 76:123-130.

Rau, H., T. Elbert, 2001, Psychophysiology of arterial baroreceptors and the etiology of hypertension, Biological Psychology 57:179-201.

Ristroph, L., G. Ristroph, S. Morozova, A. Bergou, S. Chang, J. Guckenheimer, Z. Wang, I. Cohen, 2013, Active and passive stabilization of body pitch in insect flight, Journal of the Royal Society Interface 10:20130237.

Rockafellar, R., 1970, Convex Analysis, Princeton University Press, Princeton, NJ.

Sergeant C., S. Dehaene, 2004, Is consciousness a gradual phenomenon? Psychological Science, 15:11-19.

Thayer, J., R. Lane, 2000, A model of neurovisceral integration in emotional regulation and dysregulation, Journal of Affective Disorders 61:201-216.

Tishby, N., F. Pereira, W. Bialek, 1999, The information bottleneck method, 37th Allerton Conference on Communication, Control and Computing, pp. 368-377.

Wallace, R., D. Wallace, 2010, Gene Expression and Its Discontents: The social production of chronic disease, Springer, New York.

Wallace, R., 2005, Consciousness: A mathematical treatment of the global neuronal workspace model, Springer, New York.

Wallace, R., 2012, Consciousness, crosstalk and the mereological fallacy: an evolutionary perspective, Physics of Life Reviews, 9:426-453.

Wallace, R., 2015a, Metabolic free energy and biological codes: A 'Data Rate Theorem' aging model, Bulletin of Mathematical Biology 77:879-903.

Wallace, R., 2015b, An Ecosystem Approach to Economic Stabilization: Escaping the neoliberal wilderness, Vol. 21, Routledge Series in Heterodox Economics, New York.

Wallace, R., 2015c, An Information Approach to Mitochondrial Dysfunction: Extending Swerdlow's hypothesis, Vol. 4 of the World Scientific Series in Information Studies, Singapore.

Wallace, R., 2016, An Information Treatment of Instabilities in Critical Systems. In press, World Scientific, Singapore.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1236v2 | CC-BY 4.0 Open Access | rec: 21 Jul 2015, publ: 21 Jul 2015