J Math Chem (2014) 52:2663–2679
DOI 10.1007/s10910-014-0418-x

ORIGINAL PAPER

Extending Swerdlow's hypothesis: statistical models of mitochondrial deterioration and aging

Rodrick Wallace

Received: 12 May 2014 / Accepted: 20 September 2014 / Published online: 25 September 2014

Abstract Consistent with Swerdlow's mitochondrial cascade hypothesis on the etiology of Alzheimer's disease, we find the rate of mitochondrial free energy availability serves as a temperature analog for the 'spontaneous symmetry breaking' of the group structure associated with the error minimization coding scheme of protein folding, characterizing a phase transition that collapses normal folding to pathological amyloid production. Generalization of the mitochondrial argument is possible to the groupoids that are central to the study of analogous, often highly punctuated, 'ground state' failures in more complex biological processes, taking Maturana's perspective on the essential role of cognition across the living state. The development is remarkably straightforward.

Keywords Alzheimer's disease · Cognitive system · Groupoid · Protein folding · Spontaneous symmetry breaking

1 Introduction

Swerdlow's mitochondrial cascade hypothesis for Alzheimer's disease (AD) [1] proposes that a person's genes determine their baseline mitochondrial function and durability. While both parents influence one's lifetime AD risk, since mtDNA is maternally inherited, mothers have a greater impact than fathers. Following [1], it is generally accepted that mitochondrial function declines with age, and data suggest this drives a variety of age-associated physiological changes. It is likely cell physiology initially compensates for and adapts to this change, but eventually a point is reached at which adequate compensation is no longer possible. The mitochondrial cascade hypothesis

R. Wallace (✉)
Division of Epidemiology, The New York State Psychiatric Institute, New York, NY, USA
e-mail: Wallace@nyspi.columbia.edu; rodrick.wallace@gmail.com

proposes a person's genetically determined mitochondrial starting line, in conjunction with their genetically and environmentally determined rate of mitochondrial decline, determines the age at which clinical disease ensues.

Here, we will propose a strikingly direct model of this dynamic that generalizes across much of the phenotype and pathology of aging.

We begin far afield indeed.

A remarkable, but seemingly underappreciated, theoretical development has been the finding of a close relation between information theory inequalities and a spectrum of results in the theory of finite groups [2]:

Given two random variables X_1 and X_2 having Shannon uncertainties $H(X_1)$ and $H(X_2)$ defined in the usual manner [3], the information theory chain rule states that, for the joint uncertainty $H(X_1, X_2)$,

$$H(X_1) + H(X_2) \geq H(X_1, X_2) \tag{1}$$

Similarly, let G be any finite group, and G_1, G_2 be subgroups of G. Let $|G|$ represent the order of a group, i.e., the number of elements. Then it is easy to show the intersection $G_1 \cap G_2$ is also a subgroup, and that

$$\log\left[\frac{|G|}{|G_1|}\right] + \log\left[\frac{|G|}{|G_2}\right] \geq \log\left[\frac{|G|}{|G_1 \cap G_2|}\right] \tag{2}$$

Defining a probability for a 'random variate' associated with a group G as $Pr\{X = a\} = 1/|G|$ permits construction of a group-characterized information source, noting that, in general, the joint uncertainty of a set of random variables in not necessarily the logarithm of a rational number. The surprising ultimate result, however, is that there is a one-to-one correspondence between unconstrained information inequalities and group inequalities. Indeed, unconstrained inequalities can be proved by techniques in group theory, and certain group-theoretic inequalities can be proven by techniques of information theory.

More generally, the theory of error-correcting codes, usually called algebraic coding theory [4–6], seeks particular redundancies in message coding over noisy channels that enable efficient reconstruction of lost or distorted information. The full-bore panoply of groups, ideals, rings, algebras, and finite fields is brought to bear on the problem to produce a spectrum of codes having different capabilities and complexities: BCH, Goppa, Hamming, Linear, Reed-Muller, Reed-Solomon, and so on.

Here, we will provide two examples suggesting that the relations between groups, groupoids, and a broad spectrum of information-related phenomena of interest in biology are, similarly, surprisingly intimate.

Group symmetries associated with an error-minimization coding scheme—as opposed to error correction coding—will dominate a necessary conditions statistical model of a 'spontaneous symmetry breaking' phase transition that drives the collapse of protein folding to pathological amyloid production, and groupoids emerge as central in the study of a similar wide-ranging 'ground state' failure of cognitive process, adopting the Maturana/Varela [7] perspective on the necessity of cognition at every scale and level of organization of the living state.

2 Results

2.1 The group structure of biological codes

Tlusty's [8] analysis of deterministic-but-for-error codes (DBFE) that minimize the impact of coding errors provides a basis for examining the problem of amyloid protein misfolding. Tlusty [8] models the emergence of the genetic code as a transition in a noisy information channel, using a Rate Distortion Theorem methodology. After some development [9,10] he finds the number of possible amino acids in a coding scheme is analogous to the well-known topological coloring problem. But while in the coding problem one desires maximal similarity in the colors of neighboring 'countries', in the coloring problem one must color neighboring countries by different colors. Explicitly, one uses Heawood's formula [11] to determine the number of possible 'amino acids' given a codon graph designed to minimize errors in coding:

$$chr(\gamma) = Int\left[\frac{1}{2}(7+\sqrt{1+48\gamma})\right] \tag{3}$$

where $chr(\gamma)$ is the number of 'colored' regions, Int is the integer value of the enclosed expression, and γ is the genus of the surface of the underlying code network—basically the number of 'holes' in the code network. In general, $\gamma = 1 - (1/2)(V - E + F)$, where V is the number of code network vertices, E the number of network edges, and F the number of enclosed faces.

The central trick is that one can obtain, for any DBFE code, a basic group theoretic characterization by noting that the fundamental group (FG) of a closed, orientable surface of genus γ—in which the code network is taken as embedded—is the quotient of the free group on the 2γ generators $a_1, \ldots, a_\gamma, b_1, \ldots, b_\gamma$ by the normal subgroup generated by the product of the commutators

$$a_1 b_1 a_1^{-1} b_1^{-1} \ldots a_\gamma b_\gamma a_\gamma^{-1} b_\gamma^{-1} \tag{4}$$

This is a standard construction [12]. For example, the FG of a sphere, an orientable surface with zero holes, is trivial, having only one element, while that of the torus—a donut-like orientable surface with one hole—is isomorphic to the direct product of the integers, written as $\mathcal{Z} \times \mathcal{Z}$, and so on.

That is, every DBFE error-minimization biological code is associated with a fundamental group. The more complex the code, the richer the symmetries of the associated error network, seen as embedded in a smooth surface of genus γ. Indeed, a weakened 'groupoid' version of the argument will prove central to understanding the structure of cognitive process, as developed in a following section.

Wallace [13] suggests that the overall scheme applies to a 'protein folding code' as well. Hecht et al. [14] note that protein α-helices have the underlying 'code' 101100100110... where 1 indicates a polar and 0 a non-polar amino acid. Protein β-sheets, by contrast, have the simpler basic 'code' 10101010...

Equation (3), most directly, produces the table

γ (# surface holes)	chr(γ) (# error classes)
0	4
1	7
2	8
3	9
4	10
5	11
6, 7	12
8, 9	13

In Tlusty's scheme, the second column represents the maximal possible number of product classes that can be reliably produced by error-prone codes having γ holes in the underlying coding error network.

Normal irregular protein symmetries were first classified by Levitt and Chothia [15], following a visual study of polypeptide chain topologies in a limited dataset of globular proteins. Four major classes emerged; all α-helices; all β-sheets; α/β; and $\alpha + \beta$, with the latter two having the obvious meaning.

While this scheme strongly dominates observed irregular protein forms, Chou and Maggiora [16], using a much larger data set, recognize three more 'minor' symmetry equivalence classes; μ (multi-domain); σ (small protein); and ρ (peptide), and a possible three more subminor groupings.

We infer that, from Tlusty's perspective, the normal globular 'protein folding code error network' is, essentially, a large connected 'sphere'—producing the four dominant structural modes—but having as many as three more attachment handles, in the Morse Theory sense [17]. These basic entities then act to produce an almost unlimited set of functional proteins under normal conditions.

2.2 The amyloid condensation

What happens to the fundamental group of a deterministic-but-for-error code under conditions that are not normal?

Recent work has correlated aging in general with failure of the mitochondrial machinery providing metabolic free energy at the cellular level of organization [18,19]. As Lee and Wei [20] argue, aging is a degenerative process that is associated with progressive accumulation of deleterious changes with time, reduction of physiological function and increase in the chance of disease and death. Studies reveal a wide spectrum of alterations in mitochondria and mitochondrial DNA with aging. Mitochondria are the main cellular energy sources that generate the cellular energy source ATP through respiration and oxidative phosphorylation in the inner membrane of mitochondria. The respiratory chain of that system is also the primary intracellular source of reactive oxygen species and free radicals under normal physiological and pathological conditions. In addition, mitochondria play a central role in a great variety of cellular processes.

Numerous biochemical studies on isolated mitochondria revealed that the electron transport activities of respiratory enzyme complexes gradually decline with age in the

brain, skeletal muscle, liver and skin fibroblasts of normal human subjects. Numerous molecular studies demonstrated that somatic mutations in mitochondrial DNA accumulate with age in a variety of tissues in humans. These age-associated changes in mitochondria are well correlated with the deteriorative processes of tissues in aging.

However, although abundant experimental data have been gathered in the past decade to support the concept that decline in mitochondrial energy metabolism, reactive oxygen species overproduction and accumulation of mitochondrial DNA mutations in tissue cells are important contributors to human aging, the detailed mechanisms by which these biochemical events cause aging have remained to be established. Here we will examine the implications of mitochondrial deterioration for control of protein folding, finding, as described above, mechanisms consistent with Swerdlow's mitochondrial cascade hypothesis [1].

The prebiotic 'amyloid world' of Maury [21], in contrast to the current rich variety of normal protein structures and functions, is built on a single β-sheet lamination, and shows, by contrast to the normal protein world, a starkly simple eight-fold steric zipper [22].

As Goldschmidt et al. [23] put the matter,

> We found that [protein segments with high fibrillation propensity] tend to be buried or twisted into unfavorable conformations for forming beta sheets... For some proteins a delicate balance between protein folding and misfolding exists that can be tipped by changes in environment, destabilizing mutations, or even protein concentration...
> In addition to the self-chaperoning effects described above, proteins are also protected from fibrillation during the process of folding by molecular chaperones... Our genome-wide analysis revealed that self-complementary segments are found in almost all proteins, yet not all proteins are amyloids. The implication is that chaperoning effects have evolved to constrain self-complementary segments from interaction with each other.

Clearly, effective chaperoning requires considerable metabolic energy, and failure to provide levels adequate for both maintaining and operating such biochemical translation machinery would be expected to trigger a canonical 'code collapse', most likely in a highly punctuated manner. The formalism is classic.

The existence of a Tlusty-like error minimization coding structure implies the existence of some information source using that code-and-translator or code-and-chaperone channel. As Feynman [24], following Bennett [25] argues, it is possible to make a small (idealized) machine that transforms information received into work—free energy. Indeed, Feynman defines information precisely in terms of the free energy needed to erase it. Representing the intensity of available mitochondrial free energy as $\mathcal{H}$, we write a pseudoprobability for an information source X_j associated with coding mode j and having source uncertainty H_j as

$$\Pr[H_j] = \frac{\exp[-H_j/\omega\mathcal{H}]}{\sum_{i=1}^{n} \exp[-H_i/\omega\mathcal{H})]} \tag{5}$$

This leads to constructuion of a 'free energy' Morse Function, $\mathcal{F}$, defined in terms of the rate of available metabolic free energy as

$$\exp[-\mathcal{F}/\omega\mathcal{H}] = \sum_{i=1}^{n} \exp[-H_i/\omega\mathcal{H}] \tag{6}$$

See the Mathematical Appendix for a summary of standard material on Morse Functions.

The central insight regarding phase transitions in physical systems is 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 [26,27]. 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.

This line of argument suggests the existence of complex forms of highly punctuated phase transition in code/translator function with changes in demand for, or supply of, the rate of metabolic free energy needed to run the protein chaperone machine. That is, applying a spontaneous symmetry breaking argument to $\mathcal{F}$ generates topological transitions involving changes in the fundamental group defined by error code graph structure as the mitochondrial 'temperature' $\mathcal{H}$ decreases. As the rate of delivery of the free energy running the chaperone machines decreases, complex coding schemes can no longer be sustained, driving a punctuated shift of the fundamental group of the protein folding code to a degenerate, collapsed amyloid state.

Details of such an information phase transition may also be described using 'biological' renormalization methods [28] analogous to, but much different from, those used in the determination of physical phase transition universality classes [29]. Suppose, in classic manner, it is possible to define a characteristic 'length', say l, on the system. It is then possible to define renormalization symmetries in terms of the 'clumping' transformation, so that, for clumps of size L, in an external 'field' of strength J (that can be set to 0 in the limit), one can write, in the usual manner

$$\begin{aligned} \mathcal{F}[Q(L), J(L)] &= f(L)\mathcal{F}[Q(1), J(1)] \\ \chi(Q(L), J(L)) &= \frac{\chi(Q(1), J(1))}{L} \end{aligned} \tag{7}$$

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

As described in [28], very many 'biological' renormalizations, $f(L)$, are possible that lead to a number of quite different universality classes for biological phase transition. Indeed, a 'universality class tuning' can be used as a tool for large-scale regulation of the system. While [29] necessarily uses $f(L) \propto L^3$ for simple physical

systems, following [28], it is possible to argue that, since $\mathcal{F}$ is so closely related to information measures, it is likely to 'top out' at different rates with increasing system size, so other forms of $f(L)$ must be explored. Indeed, standard renormalization calculations for $f(L) \propto L^{\delta}$, $m\log(L)+1$, and $\exp[m(L-1)/L]$ all carry through.

2.3 The groupoid structure of cognition

The approach of the previous sections can be extended to larger-scale machinery that uses deterministic-but-for-error biological or other codes as relatively simple components in more complex systems. As many have argued, the living state involves cognitive processes at every scale of organization [7,30,31]. It not difficult to show that many forms of cognition are associated with groupoid-characterized dual information sources.

Atlan and Cohen [32] 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 $x = (a_0, a_1, \ldots, a_n, \ldots)$. 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 [28].

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

$$B_0 \equiv \{b_0, \ldots, b_k\},$$
$$B_1 \equiv \{b_{k+1}, \ldots, b_m\}.$$

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 $b_j, k+1 \leq j \leq m$ 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 $h(a_0, \ldots, a_j) \in B_0$ for all $0 \leq j < m$, but $h(a_0, \ldots, a_m) \in B_1$.

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.

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 $H \equiv \lim_{n\to\infty} \log[N(n)]/n$ 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 exists, may be path dependent, although extension to nearly ergodic processes, in a certain sense, seems possible [28].

Invoking the Shannon-McMillan Theorem [3], 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 $P(a_0, \ldots, a_n)$ and $P(a_n|a_0, \ldots, a_{n-1})$ such that appropriate joint and conditional Shannon uncertainties satisfy the classic relations

$$\begin{aligned} H[\mathbf{X}] &= \lim_{n\to\infty} \frac{\log[N(n)]}{n} \\ &= \lim_{n\to\infty} H(X_n|X_0, \ldots, X_{n-1}) \\ &= \lim_{n\to\infty} \frac{H(X_0, \ldots, X_n)}{n} \end{aligned} \tag{8}$$

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 described above, it is necessary to invoke phase change formalism, a 'biological' renormalization that generalizes Wilson's [29] approach to physical phase transition [28].

Again, Shannon uncertainties $H(...)$ are cross-sectional law-of-large-numbers sums of the form $-\sum_k P_k \log[P_k]$, where the P_k constitute a probability distribution [3].

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.

For cognitive systems, an equivalence class algebra can be now 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 a groupoid. See the Mathematical Appendix for a summary of material on groupoids. This is a weak version of a very standard argument in algebraic toplogy leading to the definition of fundamental and free groups [12,33]. One might call this construction the fundamental groupoid of the cognitive process.

The vertices of the resulting network of cognitive dual languages 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. It is an abstract set of language-analogs dual to the set of cognitive processes of interest, that may become linked into higher order structures through crosstalk.

As briefly touched upon above, topology has become an object of algebraic study—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. Again, characterization via equivalence classes defines a groupoid, an extension of the idea of a symmetry group, as summarized by [34,35]. Linkages across this set of languages occur via the groupoid generalization of Landau's spontaneous symmetry breaking arguments used above. As above, we use a standard approach to constructing a Morse Function parameterized in the rate of available metabolic free energy.

With each subgroupoid G_i of the fundamental groupoid associated with the cognitive process of interest we can associate source uncertainty $H(X_{G_i}) \equiv H_{G_i}$, where X is the dual information source of the cognitive phenomenon of interest.

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 it becomes possible to construct a 'groupoid free energy' driven by the quality of available metabolic free energy, represented by the mitochondrial rate $\mathcal{H}$, to be taken as a temperature analog.

The argument-by-abduction from physical theory is that $\mathcal{H}$ constitutes a kind of thermal bath for the processes of biological cognition. Thus we can construct another Morse Function by writing a pseudo-probability for the information sources X_{G_i} having source uncertainties H_{G_i} as

$$\Pr[H_{G_i}] = \frac{\exp[-H_{G_i}/\kappa\mathcal{H})]}{\sum_j \exp[-H_{G_j}/\kappa\mathcal{H}]} \tag{9}$$

where κ is an appropriate dimensionless constant characteristic of the particular system. The sum is over all possible subgroupiods of the largest available cognitive groupoid. Note that 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.

The Morse Function defined for invocation of Pettini's topological hypothesis or Landau's spontaneous symmetry breaking is then a 'groupoid free energy' F given by

$$\exp[-F/\kappa\mathcal{H}] \equiv \sum_j \exp[-H_{G_j}/\kappa\mathcal{H}] \tag{10}$$

Spontaneous symmetry breaking arguments are invoked here in terms of the groupoid associated with the set of dual information sources.

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

The essential point is that decline in the rate of available mitochondrial free energy $\mathcal{H}$, or in the ability to actually use that free energy as indexed by κ, can lead to punctuated decline in the complexity of cognitive process within the entity of interest, according to this model.

If $\kappa\mathcal{H}$ is relatively large—a rich and varied real-time free energy environment—then there are many possible cognitive responses. If, however, constraints of mitochondrial aging limit the magnitude of $\kappa\mathcal{H}$, then an essential cognitive system may or will begin to collapse 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, recognizably akin to amyloid collapse in the much simpler deterministic-but-for-error protein coding machineries.

2.4 Distortion as order parameter

As described in the Mathematical Appendix, the Rate Distortion Function (RDF) is the minimum rate of information transmission necessary to ensure that the average distortion between message sent and message received, using a particular distortion measure over a given channel, is less than $D \geq 0$. Usually written $R(D)$, it is always a decreasing convex function of D, a reverse J-shaped curve [3]. For example, a Gaussian channel under the squared distortion measure and in the presence of noise with zero mean and variance σ^2, has $R(D) = 1/2 \log[\sigma^2/D]$.

For protein folding in the cell, elaborate regulatory machinery is provided by the endoplasmic reticulum [36], implying the necessity of some comparison between what is desired and what is produced. In general, Maturana-like cognitive processes at every scale and level of organization of the living state must have regulatory systems that make similar comparisons. What we have argued in the previous two sections can be restated in terms of the collapse of the RDF with decreasing available metabolic free energy, or rather, via convexity, as the sudden appearance of a large average distortion D, as an analog to the usual order parameter in a physical system. That is, in the way magnetization disappears above a certain critical temperature in a ferromagnet, the average distortion declines in a punctuated manner in the presence of high enough rates of available metabolic free energy, driven by the underlying groupoid structure,

remembering that the simplest groupoid is the disjoint union of groups, including a set consisting of a single group.

3 Discussion

The unexpected correspondence between unconstrained information theory inequalities and the structure of finite groups appears to foreshadow a spectrum of deeper relations between the dynamics of information sources and sometimes hidden underlying biological symmetries. These can be simple groups, as with DBFE error-minimization biological codes, or subtle 'tilings' akin to Arabic decorations—cognitive groupoids. Wallace [37,38] extends the argument to intrinsically disordered proteins and their logic gates, via nonrigid molecular symmetries built on semidirect and wreath products of simpler groups. The satisfactory operation of such gates will then be a symmetry-constrained punctuated function of available rates of metabolic free energy, although mathematical description of such intermediate scales is likely to be typically more difficult than the two relatively simple limits described in this paper.

Indeed, using the methods of Houghton [39] it is possible to define wreath products of groupoids, leading to a 'nonrigid theory of cognition'—not mathematically trivial—that can be extended further via 'fuzzy' generalizations likely to better fit biological complexities [40].

What seems clear is that information and symmetries, of various sorts and subtleties, may have unexpected convolutions and intertwinings, and these, in the context of the living state, will in turn be driven by the availability of metabolic free energy. Inability to provide adequate rates of that resource expresses itself in punctuated failure of central physiological function, recognizably analogous to spontaneous symmetry breaking in simple physical systems. This, via deterioration of basic cellular mitochondrial energy mechanisms, appears to be a critical component in the phenomenon of aging.

4 Mathematical appendix

4.1 Morse theory

Morse Theory explores relations between analytic behavior of a function—the location and character of its critical points—and the underlying topology of the manifold on which the function is defined. We are interested in a number of such functions, for example information source uncertainty on a parameter space and possible iterations involving parameter manifolds determining critical behavior. An example might be the sudden onset of a giant component. These can be reformulated from a Morse Theory perspective [27].

The basic idea of Morse Theory is to examine an n-dimensional manifold M as decomposed into level sets of some function $f : M \to \mathbf{R}$ where $\mathbf{R}$ is the set of real numbers. The a-level set of f is defined as

$$f^{-1}(a) = \{x \in M : f(x) = a\},$$

the set of all points in M with $f(x) = a$. If M is compact, then the whole manifold can be decomposed into such slices in a canonical fashion between two limits, defined by the minimum and maximum of f on M. Let the part of M below a be defined as

$$M_a = f^{-1}(-\infty, a] = \{x \in M : f(x) \leq a\}.$$

These sets describe the whole manifold as a varies between the minimum and maximum of f.

Morse functions are defined as a particular set of smooth functions $f : M \to \mathbf{R}$ as follows. Suppose a function f has a critical point x_c, so that the derivative $df(x_c) = 0$, with critical value $f(x_c)$. Then, f is a Morse function if its critical points are nondegenerate in the sense that the Hessian matrix of second derivatives at x_c, whose elements, in terms of local coordinates are

$$\mathcal{H}_{i,j} = \partial^2 f / \partial x^i \partial x^j,$$

has rank n, which means that it has only nonzero eigenvalues, so that there are no lines or surfaces of critical points and, ultimately, critical points are isolated.

The index of the critical point is the number of negative eigenvalues of $\mathcal{H}$ at x_c.

A level set $f^{-1}(a)$ of f is called a critical level if a is a critical value of f, that is, if there is at least one critical point $x_c \in f^{-1}(a)$.

Again following [27], the essential results of Morse Theory are:

1. If an interval $[a, b]$ contains no critical values of f, then the topology of $f^{-1}[a, v]$ does not change for any $v \in (a, b]$. Importantly, the result is valid even if f is not a Morse function, but only a smooth function.
2. If the interval $[a, b]$ contains critical values, the topology of $f^{-1}[a, v]$ changes in a manner determined by the properties of the matrix H at the critical points.
3. If $f : M \to \mathbf{R}$ is a Morse function, the set of all the critical points of f is a discrete subset of M, i.e., critical points are isolated. This is Sard's Theorem.
4. If $f : M \to \mathbf{R}$ is a Morse function, with M compact, then on a finite interval $[a, b] \subset \mathbf{R}$, there is only a finite number of critical points p of f such that $f(p) \in [a, b]$. The set of critical values of f is a discrete set of $\mathbf{R}$.
5. For any differentiable manifold M, the set of Morse functions on M is an open dense set in the set of real functions of M of differentiability class r for $0 \leq r \leq \infty$.
6. Some topological invariants of M, that is, quantities that are the same for all the manifolds that have the same topology as M, can be estimated and sometimes computed exactly once all the critical points of f are known: let the Morse numbers $\mu_i (i = 0, \ldots, m)$ of a function f on M be the number of critical points of f of index i, (the number of negative eigenvalues of H). The Euler characteristic of the complicated manifold M can be expressed as the alternating sum of the Morse numbers of any Morse function on M,

$$\chi = \sum_{i=1}^{m} (-1)^i \mu_i.$$

The Euler characteristic reduces, in the case of a simple polyhedron, to

$$\chi = V - E + F$$

where V, E, and F are the numbers of vertices, edges, and faces in the polyhedron.

7. Another important theorem states that, if the interval $[a, b]$ contains a critical value of f with a single critical point x_c, then the topology of the set M_b defined above differs from that of M_a in a way which is determined by the index, i, of the critical point. Then M_b is homeomorphic to the manifold obtained from attaching to M_a an i-handle, i.e., the direct product of an i-disk and an $(m - i)$-disk.

Pettini [27] and Matsumoto [7] contain details and further references.

4.2 Groupoids

A groupoid, G, is defined by a base set A upon which some mapping—a morphism—can be defined. Note that not all possible pairs of states (a_j, a_k) in the base set A can be connected by such a morphism. Those that can define the groupoid element, a morphism $g = (a_j, a_k)$ having the natural inverse $g^{-1} = (a_k, a_j)$. Given such a pairing, it is possible to define 'natural' end-point maps $\alpha(g) = a_j$, $\beta(g) = a_k$ from the set of morphisms G into A, and a formally associative product in the groupoid $g_1 g_2$ provided $\alpha(g_1 g_2) = \alpha(g_1)$, $\beta(g_1 g_2) = \beta(g_2)$, and $\beta(g_1) = \alpha(g_2)$. Then, the product is defined, and associative, $(g_1 g_2)g_3 = g_1(g_2 g_3)$. In addition, there are natural left and right identity elements λ_g, ρ_g such that $\lambda_g g = g = g\rho_g$.

An orbit of the groupoid G over A is an equivalence class for the relation $a_j \sim G a_k$ if and only if there is a groupoid element g with $\alpha(g) = a_j$ and $\beta(g) = a_k$. A groupoid is called transitive if it has just one orbit. The transitive groupoids are the building blocks of groupoids in that there is a natural decomposition of the base space of a general groupoid into orbits. Over each orbit there is a transitive groupoid, and the disjoint union of these transitive groupoids is the original groupoid. Conversely, the disjoint union of groupoids is itself a groupoid.

The isotropy group of $a \in X$ consists of those g in G with $\alpha(g) = a = \beta(g)$. These groups prove fundamental to classifying groupoids.

If G is any groupoid over A, the map $(\alpha, \beta) : G \to A \times A$ is a morphism from G to the pair groupoid of A. The image of (α, β) is the orbit equivalence relation $\sim G$, and the functional kernel is the union of the isotropy groups. If $f : X \to Y$ is a function, then the kernel of f, $ker(f) = [(x_1, x_2) \in X \times X : f(x_1) = f(x_2)]$ defines an equivalence relation.

Groupoids may have additional structure. For example, a groupoid G is a topological groupoid over a base space X if G and X are topological spaces and α, β and multiplication are continuous maps.

In essence, a groupoid is a category in which all morphisms have an inverse, here defined in terms of connection to a base point by a meaningful path of an information source dual to a cognitive process.

The morphism (α, β) suggests another way of looking at groupoids. A groupoid over A identifies not only which elements of A are equivalent to one another (iso-

morphic), but *it also parameterizes the different ways (isomorphisms) in which two elements can be equivalent*, i.e., in our context, all possible information sources dual to some cognitive process. Given the information theoretic characterization of cognition presented above, this produces a full modular cognitive network in a highly natural manner.

It is interesting to conjecture that characterization of cognition in terms of groupoids generalizes in some sense with application of the groupoid version of the Seifert-van Kampen Theorem [41]. The question is how, when a number of cognitive processes both operate simultaneously and interact, does the groupoid associated with the joint information source relate to those of the underlying cognitive processes. The canonical example might be the global workspace of consciousness [28], but wound healing and the immune response provide other examples [30]. Each cognitive process X_j can be associated with an individual source uncertainty H_j. Then, by the information theory chain rule, $H[X_1, \ldots] \leq H[X_1] + \cdots$ as in Eq. (1), presumably leading to some groupoid version of Eq. (2). The choice of a fixed a_0 state as a starting point for all processes means that they all 'touch' at that base point, and this may permit definition of some appropriate 'free groupoid' in the spirit that a topological free group can be defined if several topological spaces touch at a basepoint. The resulting free group analog then would characterize the symmetry of the joint uncertainty in terms of the groupoids of the underlying sources. Details, however, do not appear to be at all simple (e.g., [42]).

4.3 The rate distortion theorem

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

Define a distortion measure $d(y, \hat{y})$ comparing the original to the retranslated path. Many distortion measures are possible. 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 $d(y, \hat{y}) = (y - \hat{y})^2$.

The distortion between paths y^n and $\hat{y}^n$ is defined as $d(y^n, \hat{y}^n) \equiv \frac{1}{n} \sum_{j=1}^{n} d(y_j, \hat{y}_j)$.

A remarkable characteristic of the Rate Distortion Theorem is that the basic result is independent of the exact distortion measure chosen [3].

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

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

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

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

where $H(..., ...)$ is the standard joint, and $H(...|...)$ the conditional, Shannon uncertainties [3].

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

In general, of course, this will not be true.

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

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

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 (i.e., 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 [3] or Dembo and Zeitouni [43] provide details. The rate distortion function has been calculated for a number of systems, often using Lagrange multiplier or Khun-Tucker optimization methods.

$R(D)$ is necessarily a decreasing convex function of D for any reasonable definition of distortion [3]. That is, $R(D)$ is always a reverse J-shaped curve.

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

$$R(D) = 1/2 \log[\sigma^2/D], 0 \leq D \leq \sigma^2$$
$$R(D) = 0, D > \sigma^2$$

Recall the relation between information source uncertainty and channel capacity [3]:

$$H[X] \leq C$$

where H is the uncertainty of the source X and C the channel capacity. Remember also that

$$C \equiv \max_{P(X)} I(X|Y)$$

where $P(X)$ is chosen so as to maximize the rate of information transmission along a channel Y.

References

1. R. Swerdlow et al., The Alzheimer's disease mitochondrial cascade hypothesis. J. Alzheimers Dis. **20**, S265–S279 (2010)
2. R. Yeung, *Information Theory and Network Coding* (Springer, New York, 2008)
3. T. Cover, J. Thomas, *Elements of Information Theory*, 2nd edn. (Wiley, New York, 2006)
4. J. van Lint, *Introduction to Coding Theory* (Springer, New York, 1999)
5. O. Pretzel, *Error-Correcting Codes and Finite Fields* (Clarendon Press, Oxford, 1996)
6. S. Roman, *Introduction to Coding and Information Theory* (Springer, New York, 1997)
7. H. Maturana, F. Varela, *Autopoiesis and Cognition* (Reidel, Netherlands, 1980)
8. T. Tlusty, A model for the emergence of the genetic code as a transition in a noisy information channel. J. Theor. Biol. **249**, 331–342 (2007)
9. T. Tlusty, A simple model for the evolution of molecular codes driven by the interplay of accuracy, diversity and cost. Phys. Biol. **5**, 016001 (2008)
10. T. Tlusty, Casting polymer nets to optimize noisy molecular codes. PNAS **105**, 8238–8243 (2008)
11. G. Ringel, J. Young, Solutions of the Heawood map-coloring problem. PNAS **60**, 438–445 (1968)
12. J. Lee, *Introduction to Topological Manifolds, Graduate Texts in Mathematics Series* (Springer, New York, 2000)
13. R. Wallace, A scientific open season. Phys. Life Rev. **7**, 377–378 (2010)
14. M. Hecht et al., De novo proteins from designed combinatorial libraries. Protein Sci. **13**, 1711–1723 (2004)
15. M. Levitt, C. Chothia, Structural patterns in globular proteins. Nature **261**, 552–557 (1976)
16. K.C. Chou, G. Maggiora, Domain structural class prediction. Protein Eng. **11**, 523–528 (1998)
17. Y. Matsumoto, *An Introduction to Morse Theory, Translations of Mathematical Monographs*, vol. 208 (American Mathematical Society, Providence, 2002)
18. D.C. Wallace, A mitocondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. Ann. Rev. Gen. **39**, 359–407 (2005)
19. D.C. Wallace, Mitochondrial DNA mutations and aging. Environ. Mol. Mutagen. **51**, 440–450 (2010)
20. H. Lee, Y. Wei, Mitochondria and Aging. Chapter 14 in Scantena et al. (eds.), Advances in Mitochondrial Medicine, Springer, New York (2012)
21. C. Maury, Self-propagating β-sheet polypeptide structures as prebiotic informational molecular entities: the amyloid world. Orig. Life Evol. Biosph. **39**, 141–150 (2009)
22. M. Sawaya et al., Atomic structures of amyloid cross-β splines reveal varied steric zippers. Nature **447**, 453–457 (2007)
23. L. Goldschmidt et al., Identifying the amylome, proteins capable of forming amyloid-like fibrils. PNASD **107**, 3487–3492 (2010)
24. R. Feynman, *Lectures on Computation* (Westview Press, New York, 2000)
25. C. Bennett, Logical depth and physical complexity, in *The Universal Turing Machine: A Half-Century Survey*, ed. by R. Herkin (Oxford University Press, New York, 1988), pp. 227–257
26. L. Landau, E. Lifshitz, *Statistical Physics, Part I* (Elsevier, New York, 2007)
27. M. Pettini, *Geometry and Topology in Hamiltonian Dynamics* (Springer, New York, 2007)
28. R. Wallace, *Consciousness: A Mathematical Treatment of the Global Neuronal Workspace Model* (Springer, NY, 2005)
29. K. Wilson, Renormalization group and critical phenomena I. Renormalization group and the Kadanoff scaling picture. Phys. Rev. B **4**, 3174–3183 (1971)
30. R. Wallace, Consciousness, crosstalk, and the mereological fallacy: an evolutionary perspective. Phys. Life Rev. **9**, 426–453 (2012)
31. R. Wallace, Cognition and biology: perspectives from information theory. Cogn. Process. **15**, 1–12 (2014)
32. H. Atlan, I. Cohen, Immune information, self-organization, and meaning. Int. Immunol. **10**, 711–717 (1998)
33. R. Crowell, R. Fox, *Introduction to Knot Theory* (Ginn and Company, New York, 1963)
34. R. Brown, From groups to groupoids: a brief survey. Bull. Lond. Math. Soc. **19**, 113–134 (1987)
35. A. Weinstein, Groupoids: unifying internal and external symmetry. Not. Am. Math. Assoc. **43**, 744–752 (1996)
36. Z. Budrikis et al. Protein accumulation in the endoplasmic reticulum as a non-equilibrium phase transition. Nat. Commun. **5**, 3620 (2014)

37. R. Wallace, Multifunction moonlighting and intrinsically disordered proteins: information catalysis, non-rigid molecule symmetries and the 'logic gate' spectrum. Comptes Rendus Chimie **14**, 1117–1121 (2011)
38. R. Wallace, Spontaneous symmetry breaking in a non-rigid molecule approach to intrinsically disordered proteins. Mol. BioSyst. **8**, 374–377 (2012)
39. C. Houghton, Wreath products of groupoids. J. Lond. Math. Soc. **10**, 179–188 (1975)
40. R. Wallace, Statistical models of critical phenomena in fuzzy biocognition. Biosystems **117**, 54–59 (2014)
41. R. Brown et al., Nonabelian Algebraic Topology: Filtered Spaces, Crossed Complexes, Cubical Homotopy Groupoids, EMS Tracts in Mathematics, vol. 15 (2011)
42. I. Baianu et al., Complex nonlinear biodynamics in categories: higher dimensional algebra, and Lukasiewicz–Moisil topos: transformations of neuronal, genetic and neoplastic networks. Axiomathes **16**, 65–122 (2005)
43. A. Dembo, O. Zeitouni, *Large Deviations and Applications* (Springer, New York, 1998)