Culture and the Trajectories of Developmental Pathology: Insights from Control and Information Theories

Rodrick Wallace
Division of Epidemiology
The New York State Psychiatric Institute
Wallace@nyspi.columbia.edu, rodrick.wallace@gmail.com

March 27, 2018

Abstract

Cognition in living entities – and their social groupings or institutional artifacts – is necessarily as complicated as their embedding environments, which, for humans, includes a particularly rich cultural milieu. The asymptotic limit theorems of information and control theories permit construction of a new class of empirical 'regression-like' statistical models for cognitive developmental processes, their dynamics, and modes of dysfunction. Such models may, as have their simpler analogs, prove useful in the study and remediation of cognitive failure at and across the scales and levels of organization that constitute and drive the phenomena of life. These new models particularly focus on the roles of sociocultural environment and stress, in a large sense, as both trigger for the failure of the regulation of biocognition and as 'riverbanks' determining the channels of pathology, with implications across life-course developmental trajectories. We examine the effects of an embedding cultural milieu and its socioeconomic implementations using the 'lenses' of metabolic optimization, control system theory, and an extension of symmetry-breaking appropriate to information systems. A central implication is that most, if not all, human developmental disorders are fundamentally culture-bound syndromes. This has deep implications for both individual treatment and public health policy.

Key Words: cognition; culture; developmental trajectory; global broadcast; living state; statistical models

1   Introduction

Contrary to much of current Western medical ideology, both the underlying structure and external manifestations of mental disorder are heavily influenced,

if not substantially determined, by culture. In essence, mental dysfunctions reflect a famous aphorism by the evolutionary anthropologist Robert Boyd: 'Culture is as much a part of human biology as the enamel on our teeth'. We argue here that other 'developmental disorders', via the central role of cognition at every scale and level of organization, will also be heavily influenced by the embedding cultural milieu.

Following Kohrt et al. (2014), something of this has been partly recognized within the current revision of the US-based Diagnostic and Statistical Manual of Mental Disorders (DSM-5), which refers to 'Cultural Concepts of Distress' as ways that cultural groups experience, understand, and communicate suffering, behavioral problems or troubling thoughts or emotions (APA 2013; Lewis-Fernandez et al., 2017). But beyond that, cultural differences in schizophrenia outcomes, identified in three successive studies (Craig et al. 1997; Harrison et al. 2001; Sartorius et al. 1972, 1977; WHO 1981; Jablonsky et al. 1992) have, in the words of Kohrt et al., "...done little to alter conceptualizations or treatment of the disorder...".

Here, we will offer such an altered conceptualization, using new statistical models of cognitive processes based on the asymptotic limit theorems of information and control theories.

Kleinman and Good (1985, p. 492) outline something of the cross-cultural subtleties affecting the study of depression that argue against simple evolutionary or genetic interpretations. They find that, when culture is treated as a constant, as is common when studies are conducted in Western society, it is relatively easy to view depression as a biological disorder, triggered by social stressors in the presence of ineffective support, and reflected in a set of symptoms or complaints that map back onto the biological substrate of the disorder. However, when culture is treated as a significant variable, for example, when the researcher seriously confronts the world of meaning and experience of members of non-Western societies, many Western assumptions about the nature of emotions and illness are cast in sharp relief.

Dramatic differences are found across cultures in the social organization, personal experience, and consequences of such emotions as sadness, grief, and anger, of behaviors such as withdrawal or aggression, and of psychological characteristics such as passivity and helplessness or the resort to altered states of consciousness. They are organized differently as psychological realities, communicated in a wide range of idioms, related to quite varied local contexts of power relations, and are interpreted, evaluated, and responded to as fundamentally different meaningful realities. Depressive illness and dysphoria are thus not only interpreted differently in non-Western societies and across cultures; they are constituted as fundamentally different forms of social reality.

A number of comprehensive overviews support that perspective, for example, Bebbington (1993), Jenkins et al. (1990), Manson (1995). It is now clear that cultural variations exist in the areas of meaning, perceived causes, onset patterns, epidemiology, symptom expression, course, and outcome, variations having important implications for understanding clinical activities including conceptualization, assessment, and therapy.

Kleinman and Cohen (1997) argue that several myths have become central to Western psychiatry. The first is that the forms of mental illness everywhere display similar degrees of prevalence. The second is an excessive adherence to a principle known as the pathogenic/pathoplastic dichotomy, which holds that biology is responsible for the underlying structure of a malaise, whereas cultural beliefs shape the specific ways in which a person experiences it. The third myth maintains that various unusual culture-specific disorders whose biological bases are uncertain occur only in exotic places outside the West. In an effort to base psychiatry in 'hard' science and thus raise its status to that of other medical disciplines, psychiatrists have narrowly focused on the biological underpinnings of mental disorders while discounting the importance of such 'soft' variables as culture and socioeconomic status.

Some further examination of mental disorders will prove useful, and we follow something of (Wallace 2015a).

The criteria for a diagnosis of antisocial personality disorder (ASPD) in the US medical nosology *Diagnostic and Statistical Manual of Mental Disorders* (DSM III-R 1987, pp. 344-346) lists a set of chronic disruptive, irresponsible and antisocial behaviors, including lack of remorse and empathy, and so on (e.g., Hare et al. 1991). Kessler et al. (1994), in a study of the USA, observed a prevalence of about 3.5% of ASPD. Later work found similar prevalence (Grant et al. 2004; Compton et al. 2005). In stark contrast, Hwu et al. (1989), using DSM III instruments translated into Chinese, found for Taiwan prevalences of 0.03, 0.07, and 0.14%, respectively in village, small town, and Taipei settings, rates that are markedly lower than in the USA.

Weissman et al. (1994) used DSM III criteria to quantify obsessive compulsive disorder (OCD) across a number of international settings. While the lifetime rates for Western societies were found to be approximately 2.3%, in Taiwan, the observed rate was 0.7%. As they comment,

> ...[T]he lifetime... and... annual prevalence rates (cases per 100) of obsessive compulsive disorder in seven international communities were remarkably consistent (with the exception of Taiwan)... The prevalence rates for Taiwan were substantially lower than all of the other sites, paralleling Taiwan's low published rates for all psychiatric disorders.

South Korea, in the aftermath of attempts at deculturation carried out during the Japanese occupation between 1910 and 1945, followed by the Korean War, was found to suffer an OCD rate of 1.9% in that study, still lower than the USA at 2.3%.

These results suggest a meditation on the differences between East Asian and Western modes of thinking. The cultural psychologist S. Heine (2001) argues that the extreme nature of American individualism implies that a psychology based on late 20th century American research not only stands the risk of developing models that are particular to that culture, but of developing an understanding of the self that is peculiar in the context of the world's cultures.

Indeed, Western atomistic thinking, which pervades a spectrum of disciplines ranging from economics and evolutionary theory to psychology and psychiatry, has deep cultural roots (Wallace 2015b, Ch. 1).

In contrast, Nisbett et al. (2001), following in a long line of research (Markus and Kitayama 1991; Heine 2001), review an extensive literature on empirical studies of basic cognitive differences between individuals raised in East Asian and Western cultural heritages, which they characterize, respectively, as 'holistic' and 'analytic', finding that

1. Social organization directs attention to some aspects of the perceptual field at the expense of others.

2. What is attended to influences metaphysics.

3. Metaphysics guides tacit epistemology, that is, beliefs about the nature of the world and causality.

4. Epistemology dictates the development and application of some cognitive processes at the expense of others.

5. Social organization can directly affect the plausibility of metaphysical assumptions, such as whether causality should be regarded as residing in the field vs. in the object.

6. Social organization and social practice can directly influence the development and use of cognitive processes such as dialectical vs. logical ones.

Nisbett et al. (2001) thus conclude that tools of thought embody a culture's intellectual history, that tools have theories built into them, and that users accept these theories, albeit unknowingly, when they use these tools.

Matsuda and Nisbett (2006) find that, whereas East Asians view the world holistically, attending to the entire field and relations among objects, Westerners view the world analytically, focusing on the attributes of salient objects. Compared to Americans, East Asians were more sensitive to contextual changes than to focal object changes, These results suggest that there can be cultural variation in what may seem to be basic perceptual processes.

Similarly, Nisbett and Miyamoto (2005) argue that fundamental perceptual processes are influenced by culture. These findings establish a dynamic relationship between the cultural context and perceptual processes. They suggest that perception can no longer be regarded as consisting of processes that are universal across all people at all times.

Wallace (2007) explores analogous dynamics involving inattentional blindness and culture.

Here, we will assert that the embedding influence of culture can generalize across cognitive physiological subsystems in humans. Indeed, it is well known that schizophrenia, for example, is parallel to significant medical comorbidities such as hypertension, diabetes, and inflammation, leading to reduced lifespan by a variety of mechanisms (e.g., Silverstein et al. 2017). The syndrome has been associated with systemic mitochondrial dysfunction (e.g., Ben-Shachar 2002, Prabakaran et al. 2004; Shao et al. 2008; Scaglia 2010; Clay et al. 2011). Similarly, mitochondrial dysfunction has also been found in autism spectrum disorder (e.g., Palmieri and Persico 2010; Giulivi et al. 2010; Rossignol and Frye 2014; Goh et al. 2014). We shall make much of such observations in Section 3

below, introducing explicit mechanisms for the impact of social interaction and cultural milieu.

Indeed, the living state, it has been argued, must be cognitive at almost every scale and level of organization. As Maturana and Varela (1980) put it,

> A cognitive system is a system whose organization defines a domain of interactions in which it can act with relevance to the maintenance of itself, and the process of cognition is the actual (inductive) acting or behaving in this domain. Living systems are cognitive systems, and living as a process is a process of cognition. This statement is valid for all organisms, with and without a nervous system.

The argument is consistent with the perspective of Atlan and Cohen (1998), who characterize the immune system as cognitive using an information theory approach. The underlying inference is that, from basic biochemical mechanism, through gene expression and development, multicellularity, clonal/social structure, to the elaborate institutions of human populations, living systems must receive 'environmental' signals (in a large sense), compare them with a learned or inherited 'picture of the world', and then choose a small set of actual responses from a larger range of those available to them. Such choice reduces uncertainty in a formal manner, and such reduction implies the existence of an information source 'dual' to the cognitive process. Wallace (2005, 2012, 2015a, b, 2017, 2018) provides particular mathematical instantiations.

The Baars model of consciousness (1988) introduces another essential mechanism that appears to generalize across physiological subsystems (Wallace 2012), i.e., that a set of underlying 'unconscious cognitive modules' must come together to create a rapid, shifting, highly tunable coalition that responds in a local 'real time' to the dynamic challenges facing an organism. Baars' time frame for consciousness is of the order of 100ms. Clearly, similar mechanisms must operate at and across other time domains, scales and levels of organization for other adaptive coalitions of cooperating 'lower level' cognitive modules.

Cognition, coalition, and tunability must all be highly regulated to function reliably in the context of the changing demands of the rapidly-shifting 'roadway' topologies confronting living systems. These will act at and across various scales and levels of organization that may be within or outside the organism. It is the nature of such regulation and its failure that is the focus of this work, using currently-available theoretical tools adapted from information and control theories, most centrally involving the linkages between the two. We will be particularly concerned with constructing statistical and other models of the effects of toxic exposure and 'psychosocial' stress (or its equivalent) on patterns of cognitive failure at and across the various scales and levels of organization (Wallace 2012).

We will adapt some results from information and control theories to explore necessary conditions for stability in cognitive processes at and across different scales and levels of physiological organization, with a particular emphasis on the potential role of embedding cultural information sources. Similar approaches are

already appearing in the literature (e.g., Silverstein et al. 2017) and promise considerable advance.

We begin with some fairly general remarks on the relation between information and control in the context of 'global broadcast' models, and then, moving beyond the idea of 'culturally-specific generalized inflammation' used in Wallace (2015b, Section 8.5), introduce three new parallel 'lenses' as provided by optimization theory applied to metabolic free energy, by examination of the failure of a control system under sociocultural environment and stress, and by an analog to symmetry-breaking that is appropriate to information processes in a cultural milieu.

For completeness, we attach a summary of the 'generalized inflammation' argument in the Appendix.

An implication of this work is that deeper understanding of cognition and its pathologies requires new formal tools, perhaps different from, but certainly no less sophisticated than, those now used at the far ends of theoretical physics. At the not inconsiderable price of learning such tools, we will more deeply cultivate the stunningly rich landscapes of evolutionary, biological, and sociocultural process. Many will find the challenges and opportunities energizing.

2 Basic formalism

2.1 The Data Rate Theorem

To reiterate, biological systems and subsystems must function on complex, rapidly-shifting real-time 'roadway' topologies and are inherently unstable precisely because of those embedding topologies. The canonical metaphor is of driving a fast car on a twisting, pot-holed road at night. This requires not only quick reflexes and a reliable vehicle, but really good headlights. All important physiological processes are inherently unstable in the control theory sense: development must be closely regulated to activate and deactivate a relatively small number of genes in exactly the right sequence and at the right stage, in concert with and response to powerful, often rapidly-changing, epigenetic and environmental signals.

Failure of control – onset of instability – produces serious developmental disorders. The cognitive immune function (Atlan and Cohen 1998) engages in pathogen attack and routine maintenance according to highly irregular patterns of need, but must be closely regulated to avoid autoimmune dysfunction. Similarly, the stream of animal consciousness must be held within contextual 'riverbanks' to restrict its meanderings to useful realms. Inattentional blindness can lead to individual predation, while group survival among social animals requires conformance to group norms.

Thus for important biological and social processes, instability and its draconian regulation are always implicit.

The Data Rate Theorem (DRT) (Nair et al. 2007) links control and information theories in the study of regulation and its failure for such systems. To

use the metaphor, the DRT tells how good the headlights must be for driving on a given twisting, pot-holed road at night. In detail, the Data Rate Theorem establishes the minimum rate at which externally-supplied control information must be provided for an inherently unstable system to maintain stability.

For simplicity, we assume a linear expansion near a nonequilibrium steady state (nss), i.e., a condition that requires constant inflow of free energy for maintenance. An nss may be inherently stable or unstable, at a given energy flow, but depends on that constant flow for continued existence.

Thus an n-dimensional vector of parameters at time t, say x_t, determines the state at time $t + 1$ according to the relation

$$x_{t+1} = \mathbf{A}x_t + \mathbf{B}u_t + W_t \tag{1}$$

$\mathbf{A}, \mathbf{B}$ are fixed $n \times n$ matrices, u_t is the vector of control information, and W_t is an n-dimensional vector of white noise. See Figure 1.

Following Nair et al. (2007, Theorem 1), the DRT states that the minimum rate of control information $\mathcal{H}$ necessary to stabilize the system is

$$\mathcal{H} > \log[|\det[\mathbf{A}]|] \equiv a_0 \tag{2}$$

The right hand side of Eq.(2) is interpreted as the rate at which the underlying system generates 'topological information'.

According to the DRT, stability will be lost if the inequality of Eq.(2) is violated. For the night driving metaphor, if the headlights go out, a twisting road cannot be navigated. Here we will examine in more detail the dynamics of control failure under constraints of both available information and the ability to implement regulatory action. A more comprehensive derivation of the DRT is given in Wallace (2017, Section 7.10), based on an application of the Rate Distortion Theorem that varies with the control channel, but is based on the necessary convexity of the Rate Distortion Function for all channels.

2.2   No free lunch

Many, if not all, algorithmic systems – particularly those cognitive in the Atlan/Cohen sense – are constrained by the famous 'no free lunch' theorem (NFLT) (e.g., Wallace 2017, Section 4.11; Wolpert and MacReady 1995, 1997). An algorithmic approach to real-time problems usually centers on optimization of some chosen objective value function: total income, loss, cost by some measure, and so on. The NFLT argues there is no generally superior function optimizer.

That is, according to the theorem, an optimizer 'pays' for superior performance on some functions with inferior performance everywhere else. For those systems where the theorem applies, gains and losses balance and all optimizers have identical average performance: superiority on one subset of functions, by and large, implies inferiority on the complementary subset. This implies the necessity of tuning in the face of dynamic challenge, i.e., of the existence of a higher executive, or hierarchy of them, able to sense when to 'change gears' in a shifting 'roadway' environment. Thus Baars' model – a shifting 'spotlight' of
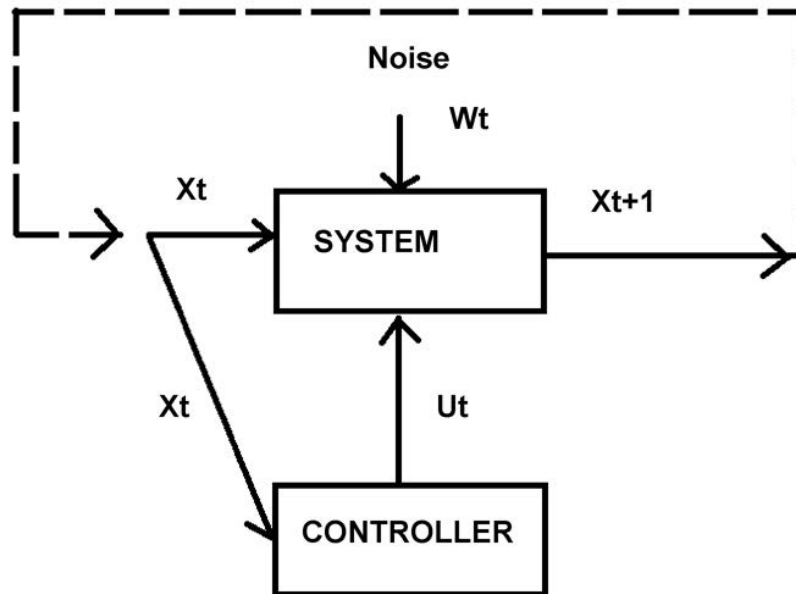
7

Figure 1: A linear expansion near a nonequilibrium steady state of an inherently unstable control system, for which $x_{t+1} = \mathbf{A}x_t + \mathbf{B}u_t + W_t$. $\mathbf{A}, \mathbf{B}$ are square matrices, x_t the vector of system parameters at time t, u_t the control vector at time t, and W_t a white noise vector. The Data Rate Theorem states that the minimum rate at which control information must be provided for system stability is $\mathcal{H} > \log[|\det[\mathbf{A}]|]$. This is characterized as saying that the rate of control information must exceed the rate at which the unstable system generates topological information.

cooperating lower level cognitive modules – emerges as a necessary evolutionary response to a critical 'resource limitation'.

A parallel approach follows the arguments of Wallace (2017, Section 4.10): it is straightforward to invert the famous Shannon Coding Theorem, fixing the distribution of the message sent along some channel, but tuning the probability distribution of that channel so as to maximize the information transmitted at the fixed message distribution. In a sense, the message is interpreted as 'transmitting the channel', and, according to the Shannon Coding Theorem, there will then be a channel distribution that will maximize such a dual channel capacity. Channel maximization for one set of control signals is thus highly specific to the message transmitted, and must be retuned for a different message.

Shannon (1959), in fact, described something of this in terms of a duality between the properties of an information source with a distortion measure and those of a channel, particularly for channels in which there is a cost associated with the different 'letters' transmitted. Solving the problem, Shannon claims, is equivalent to finding a source that is right for the channel and the desired cost. In a dual way, evaluating the rate distortion function for a source corresponds to finding a channel that is just right for the source and the allowed distortion level.

The implication is that, under dynamic circumstances, there must be a higher executive, an overriding cognitive system, able to retune the underlying control system of Figure 1 according to shifting demands, i.e., a dynamic 'optimizer of optimizers'.

2.3   Global broadcast

Such an optimizer of optimizers emerges via the evolutionary exaptation of the inevitable crosstalk between the information sources dual to the 'lower level' cognitive processes of interest (Wallace 2012). The simplest model is of a network of information sources whose probability of interaction is measured by the strength of the crosstalk. For a random network, it is well known that a 'giant component' involving most of the network nodes emerges in a punctuated manner once the strength of interaction exceeds a well-defined threshold. By contrast, a linked stars-of-stars configuration has a zero threshold: everything is connected. Network topology counts.

A random network analysis assumes a variable average number of fixed-strength linkages between the information sources composing the network nodes. Mutual information, by contrast, can continuously vary in magnitude, suggesting the need for a parameterized renormalization. The modular network structure linked by crosstalk has a topology depending on the degree of interaction. The central argument, following Wallace (2012) closely, is as follows. Define an interaction parameter, say ω, a real positive number, and look at geometric structures defined by linkages set to zero if crosstalk mutual information is less than that value, and set to unity if greater than ω. A given ω will define a regime of giant components of network elements – cognitive submodules – linked by mutual information greater than or equal to it.

The argument can be inverted. Some given topology for the giant component will define a critical value ω_C such that network elements interacting by crosstalk mutual information less than this value will be unable to participate, that is, will be locked out and not 'consciously' perceived on the time scale of the shifting 'spotlight' global broadcast of interest. As a result, ω_C is a tunable, syntactically dependent, detection limit that depends on the instantaneous topology of the corresponding shifting, tunable, giant component of linked cognitive modules defining the dynamic global broadcasts required by the no-free-lunch constraints. Wallace (2012) shows how this argument can be extended to a multiple set of simultaneous global broadcasts, characterized by a vector $\Omega = (\omega_1, ..., \omega_m)$.

All global broadcast systems, because of their necessary dynamic flexibility, are inherently unstable in the control theory sense and require fairly draconian regulation, as constrained by the Data Rate Theorem. What is the metabolic cost of such regulation?

3 Through the lens of metabolic optimization

Heinrich et al. (1991) describe enzymatic reaction systems using optimization principles. Their introductory remarks are relevant as well to cognitive systems:

> ...[A]ny approach towards the explanation of... the structural design of metabolic systems must take into account the fact that these systems are, in contrast to chemical systems of inanimate nature, the outcomes of evolution... In many... studies evolution is considered as an optimization process. This view is supported by a number of observations, for example by the fact that mutations or other changes... lead in most cases to a worse functioning of cellular metabolism... Much theoretical work has been devoted to inquire whether the structural design and functioning of present-day biochemical systems can be rationalized on the basis of optimization principles... In quantitative investigations the crucial point is the formulation of appropriate performance functions whose maximum (or minimum) value might correspond to the outcome of... evolution...

Wallace (2016) adapts an Arrhenius chemical reaction rate relation to model the rate at which biological cognition can proceed, treating the rate at which metabolic free energy is delivered to the system of interest as a temperature analog. For comparison, at 300 K, molecular energies are about 2.5 kJ/mol. By contrast, the hydrolysis of ATP at 300K represents about 50 kJ/mol, an effective 'temperature' of 6000 K.

The Rate Distortion Function R of the cognitive channel of interest is a 'free energy' index leading to a Boltzmann-like probability argument that a

normalized, dimensionless, rate of cognition, as measured by R will be given as

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

where M is the rate at which metabolic free energy is supplied, κ is a local constant and R_0 is the threshold channel capacity necessary for initiation of a cognitive function. In real systems this value will never be zero, and in many global broadcast settings will involve sharp punctuation much like the onset of consciousness: on or off.

The efficiency of a cognitive function indexed by j can thus be measured as

$$f_j(M_j) = \frac{\exp[-K_j/M_j]}{M_j} \tag{4}$$

for an associated local parameter K_j.

More generally, some $j = 1, ..., n$ cognitive modules must work together within an organism. They must communicate with each other, reflected at each module by some 'bandwidth' C_j representing the channel capacity available to the module. In addition, the embedding environment, represented by an information source H, must impose signals on each module at some rate H_j. For humans, there will be an additional embedding cultural signal module. Then Eq.(4) becomes

$$f_j(M_j, C_j, H_j) = \frac{\exp[-\frac{K_j}{M_j C_j H_j}]}{M_j C_j H_j} \tag{5}$$

Across the organism there will be significant constraints,

$$M = \sum_{j=1}^{n} M_j, \ C = \sum_{j=1}^{n} C_j, \ H = \sum_{j=1}^{n} H_j \tag{6}$$

Note that we will assume n cognitive modules and m constraints.

Given a particular cognitive module and its embedding regulatory apparatus – which must also be cognitive – we can now make a linear scalarization of a multi-objective optimization problem using unit weighting (Hwang and Massud 1979), and examine a standard Lagrangian optimization problem using

$$L = \sum_{j=1}^{n} f_j(\Pi_{k=1}^{m} x_k^j) - \sum_{k=1}^{m} \lambda_k (X_k - \sum_j x_k^j)$$
$$\partial L/\partial x_k^j = \lambda_k \tag{7}$$

where the x are variates, the X their limiting sums, and the λ are dual 'cost variates' often treated as 'undetermined multipliers'. In Eqs.(5) and (6), $m = 3$.

We can use algebraic regularities to project down, in a sense, on a single variate, here the rate of available metabolic free energy provided by mitochondrial and related machineries. This is somewhat subtle. Examining the set of

relations determining a stationary point, the second half of Eq.(7), and then *eliminating the full derivative of f*, by virtue of the product of independent variates, leads to conditions of the form $x_p^j \lambda_q = x_q^j \lambda_p$. Upon summation, a projection symmetry emerges as

$$x_p^j = \frac{X_p}{X_q} x_q^j \qquad (8)$$

Then

$$f_i(\Pi_{j=1}^m x_j^i) = f_i\left(\frac{[x_q^i]^m}{X_q^{m-1}}\Pi_{p\neq q}X_p\right) \qquad (9)$$

For $p \neq q$ we now normalize to $X_p \equiv 1$, and examine a module and its regulatory apparatus, taking each as having the same threshold K in Eq.(5).

We obtain the relation

$$f(M_i) = \frac{\exp[-\frac{K_i}{M_i^3/M^2}]}{M_i^3/M^2} \qquad (10)$$

and begin by examining a particular module and its regulatory system under metabolic free energy constraints. That is, we look at

$$\frac{\exp[-\frac{K_1}{M_1^3/M^2}]}{M_1^3/M^2} + \frac{\exp[-\frac{K_2}{(M-M_1)^3/M^2}]}{(M-M_1)^3/M^2} \qquad (11)$$

which would be appropriate to animal models, i.e., in the absence of culture, but taking social interaction as part of the environment.

Thus $M = M_1 + M_2$ is the constrained rate of metabolic free energy available to the cognitive system and its (similarly cognitive) associated regulatory apparatus.

Figure 2 shows Eq.(11) for $K_1 = K_2 = 0.5$, over the range $0 \le M \le 5$.

Below a threshold rate for the supply of metabolic free energy, the efficiency measure has a pathological bifurcation, a separation into two disjoint 'legs'. Above that threshold rate, the efficiency measure greatest when metabolic free energy is supplied equally to the system and its regulator, since we have taken the threshold levels $K_1 = K_2$ in this example.

The second condition of Eq.(7) is of independent interest, i.e., $\partial L/\partial M_i = 0$ so that $\partial f/\partial M_i = \lambda$. This is because, in a physical model, λ can be interpreted as an inverse temperature (e.g., Schrodinger 1989, Ch.II). Taking the derivative of Eq.(10) gives

$$\partial f/\partial M_i = 3\frac{M^2\left(KM^2 - M_i^3\right)}{M_i^7}\mathrm{e}^{-\frac{KM^2}{M_i^3}} = \lambda \qquad (12)$$

Figure 3 shows this relation for M_1, $M_2 = (M - M_1)$ and $K = 0.5$ from Figure 2, at $M = 5$. Of particular note is the fact that the derivatives are equal at $M_1 = M_2 = M/2$ and that the value of λ is negative. In a physical system, a negative temperature represents an unstable 'pumped' state in which
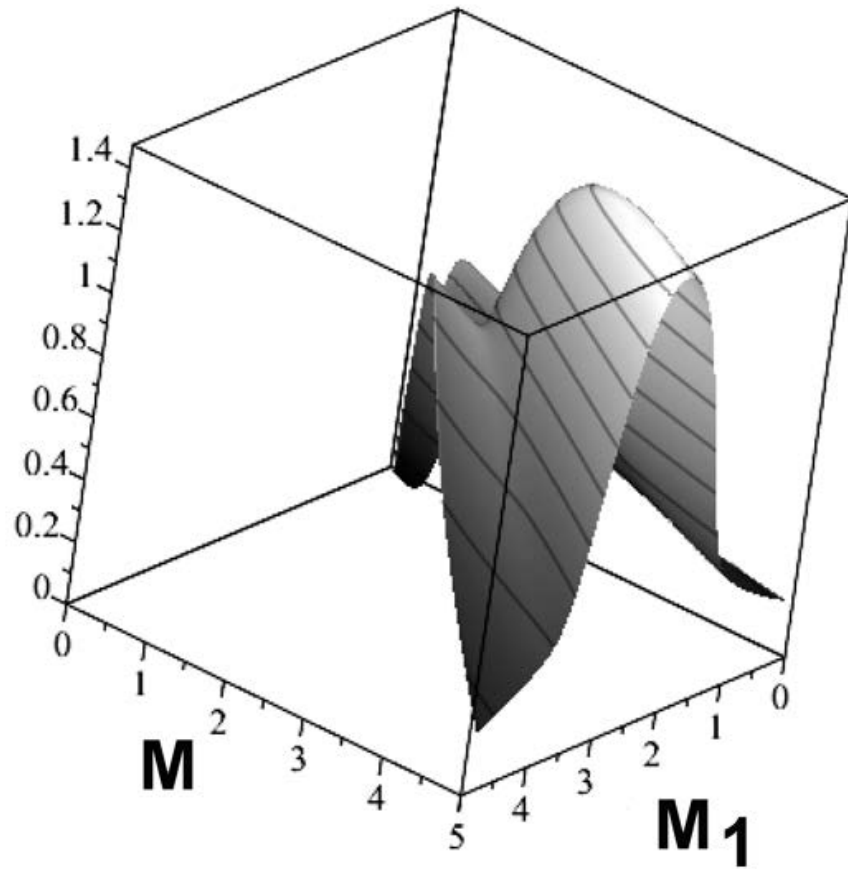
12

Figure 2: Below a threshold rate for the supply of metabolic free energy, the efficiency measure has a pathological separation into two disjoint sections. Above the threshold, efficiency is greatest when metabolic energy is supplied equally to the system and its regulator, since, here, the threshold K values are the same.

raising the energy causes the entropy to decline. For these models λ may be an independent system parameter. As Jin et al. (2008) note, if the circumstances surrounding λ are sufficiently pathological, *no optimal state may exist.*

Stochastic versions of this model are possible. The most straightforward is to average Eq.(7) across some of the variates and carry out the resulting optimization on the averaged expression. Here we focus on distributions in the threshold parameter K from the Arrhenius model.

Averaging Eq.(10) over K_j produces blunted versions of Figures 2 and 3 for inverted-U distributions that have modes greater than zero, or are indeed censored at zero.

Of more interest are inverse-J distributions for which the mode is zero, regardless of the mean value. Taking K in Eq.(10) as having an exponential distribution, so that $dP(K) = \omega \exp[-\omega K]dK, \quad \omega > 0$, gives the surprising result

$$< f_i >= \frac{M^2}{< K_i > M^2 + M_i^3} \tag{13}$$

The graph using this expression that corresponds to Figure 2 – recalling that we are optimizing the sum $f(< K >, M_1) + f(< K >, M - M_1)$ – is shown in Figure 4, and is similar for other reverse-J distributions in K, i.e., with zero mode.

An HPA axis might well be conditioned to have a higher probability of activation at low thresholds under pathologies of posttraumatic stress disorder (PTSD). The co-occurrence of depression with PTSD might then be seen as a protective measure designed to raise threshold activation to more tolerable levels, with the added costs of depressive symptoms (e.g., Dekel et al. 2014, 2017; Oquendo et al. 2003).

Similarly, Figure 4 might represent a schizophrenic inability to sort signals that is driven toward the more normal form of Figure 2 by the onset of the 'negative symptoms' that blunt emotional expression and the ability to engage in social exchange. Silverstein et al. (2017, pp.88-89) postulate something similar in terms of an increased level of noise in schizophrenic neural processing, represented here by the reverse-J distribution in activation threshold.

This argument involves a third-order metabolic free energy model that incorporates only one level of impinging 'environmental' information, basically an animal model. The human version, according to our arguments, would be a fourth order model, i.e., Eq.(10)-(12) would be in terms of M_i^3, M^4, so that Figures 2-4 would be considerably sharpened. For humans, small changes in the rate of metabolic free energy may then have far greater consequences than in animal models, a consequence of having 'culture as much a part of human biology as the enamel on our teeth'.

We next examine the failure of regulation from a specifically control theory perspective, particularly in the context of environmental contamination and/or, for higher organisms, psychosocial stress.
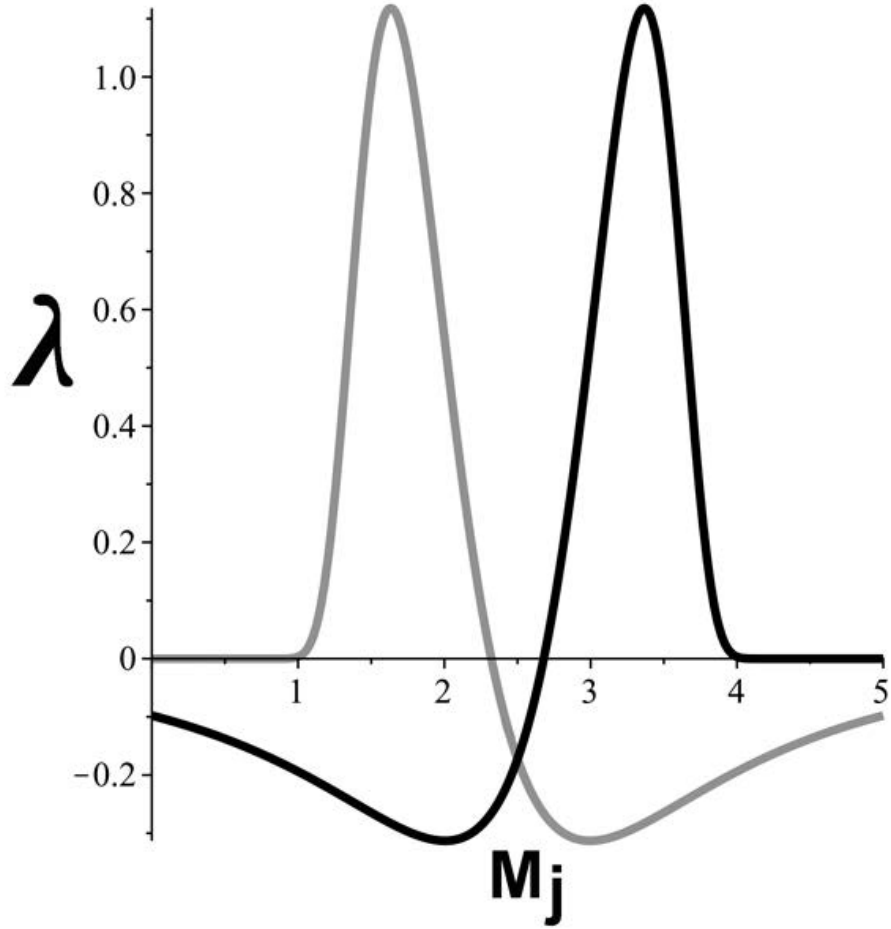
Figure 3: The condition $\partial f/\partial M_i = \lambda$ at $M = 5$, $K = 0.5$, in Figure 2. The derivatives are equal at $M_i = M/2$, but a negative λ implies this is a potentially unstable 'excited state' of the system. If, however, the system parameter λ is sufficiently positive or sufficiently negative, no optimal value will exist. The gray graph represents $\partial f/\partial M_1$ and the black $\partial f/\partial M_2$, remembering that $M_2 = M - M_1$.
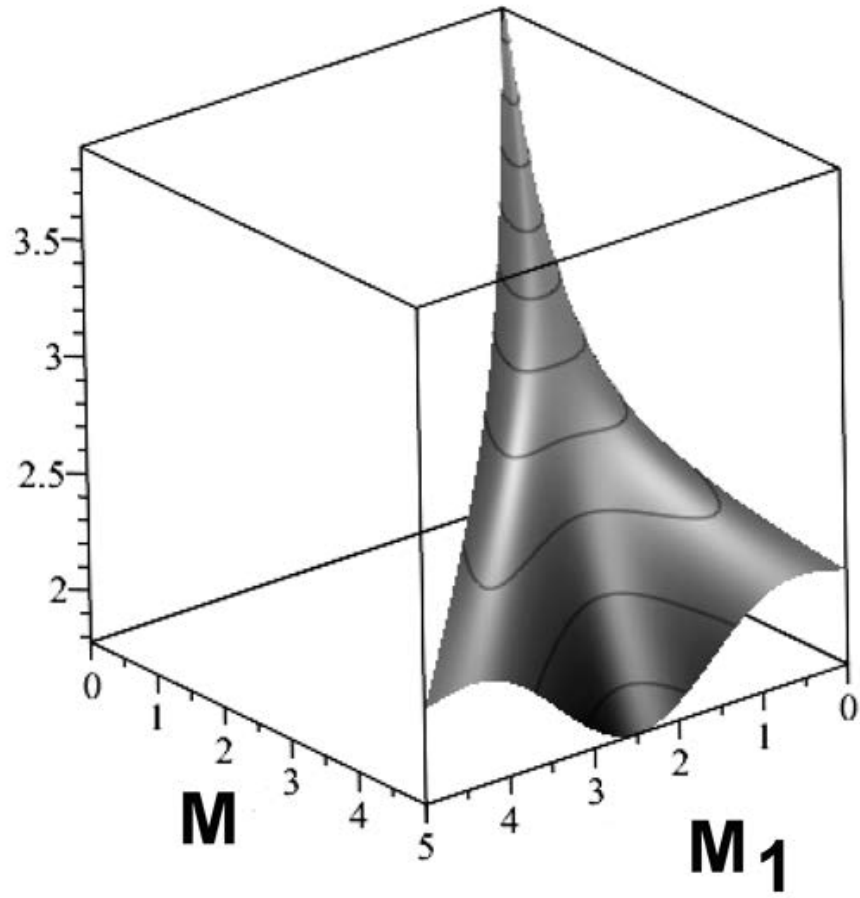
Figure 4: Optimization surface for an exponential distribution in activation threshold, K. Using Eq.(13), we examine $f(<K>, M_1) + f(<K>, M - M_1)$. While inverted-U distributions for K reproduce blunted versions of Figure 2, other reverse-J distributions – having a zero mode – reproduce this result.

4 Through the lens of control theory

4.1 Extending the Data Rate Theorem

How do elaborate control systems generally fail in other contexts? The military strategist and historian Carl von Clausewitz emphasized two factors leading to control failure in 'simple' systems: 'fog-of-war' and 'friction'. These are, first, the inevitability of limited intelligence regarding battlefield conditions, and, second, the difficulty of actually imposing control. For the night-driving metaphor, this might perhaps be a synergism between poor headlights and unresponsive steering. It is possible to generalize this insight to more complicated circumstances, at the expense of some further mathematical overhead.

Biological real-time cognitive systems can be expected to have several, perhaps many, constraints acting synergistically. We must thus envision, for each such, a relatively large nonsymmetric $n \times n$-matrix analog to a 'correlation matrix', a matrix ρ having elements $\rho_{i,j}$ representing those constraints and their pattern of interaction. Such matrices each have n invariants, $r_i, i = 1, ..., n$, that remain fixed when 'principal component' transformations are applied to data, and we can construct an invariant scalar measure from them, based on the standard polynomial relation

$$p(\beta) = \det(\rho - \beta I) = \beta^n + r_1 \beta^{n-1} + ... r_{n-1}\beta + r_n \tag{14}$$

det is the determinant, β a parameter, and I is the $n \times n$ identity matrix. The first invariant will be the trace of the matrix, and the last $\pm$ the determinant. Using these n invariants it is possible to then define an appropriate composite scalar index $\Gamma = \Gamma(r_1, ..., r_n)$ as a monotonic increasing real function. This is essentially the Rate Distortion Manifold of Glazebrook and Wallace (2009) and the Generalized Retina of Wallace and Wallace (2016).

Using the one-dimensional projection Γ, we heuristically extend the condition of Eq.(2) as

$$\mathcal{H}(\Gamma) > f(\Gamma)a_0 \tag{15}$$

The Mathematical Appendix, following Wallace (2017, Section 7.10), applies a Black-Scholes approximation to find that $\mathcal{H}(\Gamma)$ will have, in first order, the surprisingly simple form $\mathcal{H} \approx \kappa_1\Gamma + \kappa_2$. Taking $f(\Gamma)$ to similar order, so that $f(\Gamma) = \kappa_3\Gamma + \kappa_4$, the limit condition becomes

$$\mathcal{T} \equiv \frac{\kappa_1\Gamma + \kappa_2}{\kappa_3\Gamma + \kappa_4} > a_0 \tag{16}$$

where we characterize $\mathcal{T}$ as the 'control temperature' of the system. For $\Gamma = 0$ the stability condition is $\kappa_2/\kappa_4 > a_0$. At large Γ this becomes $\kappa_1/\kappa_3 > a_0$. If $\kappa_2/\kappa_4 \gg \kappa_1/\kappa_3$, the stability condition may be violated at sufficiently high Γ. Figure 5 shows the pattern.

How does $\mathcal{T}$ change in time?

The central point is how the control signal u_t in Figure 1 is expressed in the system response x_{t+1}, with a focus on $\mathcal{T}$.
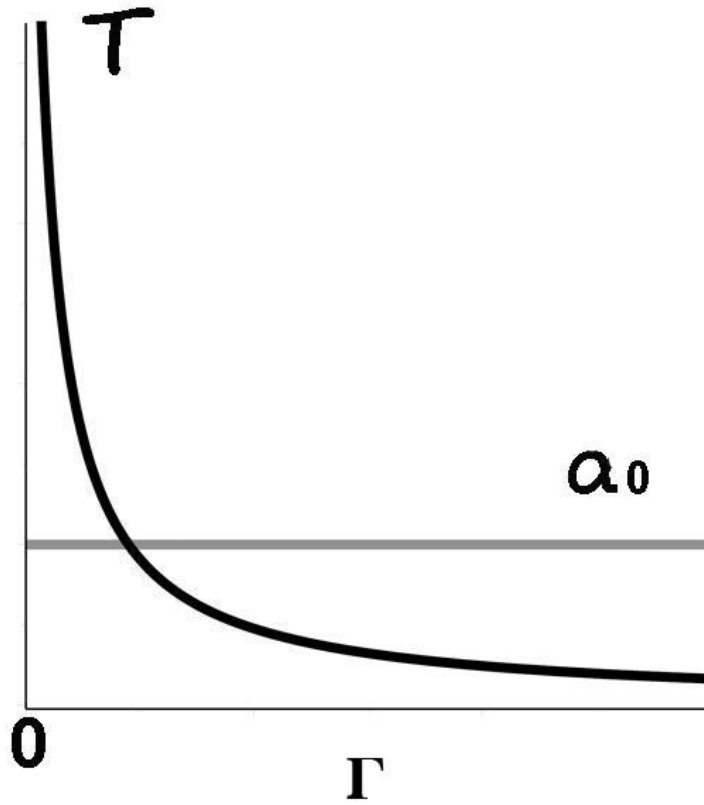
Figure 5: The horizontal line represents the critical limit a_0. For $\kappa_2/\kappa_4 \gg \kappa_1/\kappa_3$, at an intermediate value of the index Γ the 'control temperature' $\mathcal{T}$ falls below that limit, and control fails.

The basic idea is to deterministically retranslate an observed sequence of system outputs $X^i = x_1^i, x_2^i, ...$ into a sequence of possible control signals $\hat{U}^i = \hat{u}_0^i, \hat{u}_1^i, ...$ and compare that sequence with the original control sequence $U^i = u_0^i, u_1^i, ...$, with the difference between them having a particular value under some chosen distortion measure and hence having an average distortion (Cover and Thomas 2006)

$$D \equiv \sum_i p(U^i) d(U^i, \hat{U}^i) \qquad (17)$$

where $p(U^i)$ is the probability of the sequence U^i and $d(U^i, \hat{U}^i)$ measures the distortion between U^i and the sequence of control signals that has been deterministically reconstructed from the system output.

According to the Rate Distortion Theorem, there is a Rate Distortion Function (RDF), $R(D)$, that determines the minimum channel capacity necessary to keep the average distortion below a fixed real limit $D > 0$ in the presence of noise (Cover and Thomas 2006). Based on Feynman's (2000) interpretation of information as a form of free energy, using the system's channel-specific Rate Distortion Function R, we next construct a Boltzmann-like probability in the control temperature $\mathcal{T}$ as

$$dP(R, \mathcal{T}) = \frac{\exp[-R/\mathcal{T}]dR}{\int_0^\infty \exp[-R/\mathcal{T}]dR} \qquad (18)$$

since higher $\mathcal{T}$ must necessarily be associated with greater channel capacity.

The denominator can now be interpreted as a statistical mechanical partition function, and it becomes possible to define a 'free energy' Morse Function (Pettini 2007; see the Mathematical Appendix) $\mathcal{F}$ as

$$\exp[-\mathcal{F}/\mathcal{T}] \equiv \int_0^\infty \exp[-R/\mathcal{T}]dR = \mathcal{T} \qquad (19)$$

We can now define an 'entropy' in the free energy measure $\mathcal{F}$ as the Legendre Transform $\mathcal{S} \equiv \mathcal{F}(\mathcal{T}) - \mathcal{T}d\mathcal{F}/d\mathcal{T} = \mathcal{T}$. This allows use of a stochastic Onsager approximation for the dynamics of $\mathcal{T}$ in the gradient $d\mathcal{S}/d\mathcal{T}$ (de Groot and Mazur 1984). The resulting stochastic differential equation is

$$d\mathcal{T}_t = \mu dt + \sigma \mathcal{T}_t dW_t \qquad (20)$$

where μ is a 'diffusion coefficient' representing the attempts of the system to meet the demands of control, dW_t is Brownian white noise, and σ determines the magnitude of the volatility.

The base equation

$$d\mathcal{T}/dt = \mu > 0 \qquad (21)$$

'explodes' with increasing time. By the Stochastic Stabilization Theorem (Mao 2007; Appleby et al. 2008), an 'exploding' function $|f(x, t)| \to \infty$ for which

$$|f(x, t)| \leq |x|\omega, \ \omega > 0 \qquad (22)$$

can be stabilized by a volatility term $\sigma x_t dW_t$ in the sense that

$$\limsup_{t \to \infty} \frac{\log[|x(t)|]}{t} \to -\frac{\sigma^2}{2} + \omega \tag{23}$$

almost surely. If $\sigma^2/2 > \omega$, $x(t) \to 0$.

Thus, for fixed μ, rising 'volatility' – increasing σ – can trigger a downward ratchet leading to violation of the DRT condition for $\mathcal{T}$ in a highly punctuated manner.

A variant of this model assumes a volatility expression of the form

$$\sigma\sqrt{\mathcal{T}^2 + \beta^2}dW_t, \ \beta > 0$$

in Eq.(20), so that there is an internal source of variation, β, independent of $\mathcal{T}$. Then expansion of $\log[\mathcal{T}]$ via the Ito relation, using Jensen's inequality for a concave function, leads to the expectation condition

$$E(\mathcal{T}) \geq E(\log[\mathcal{T}]) = \frac{\mu \pm \sqrt{\mu^2 - \beta^2\sigma^4}}{\sigma^2} \tag{24}$$

Some exploration shows the upper limit is stable, while the lower either rises to the upper or collapses to zero. An evident necessary condition for any stability is $\mu > \beta\sigma^2$, independent of, and in addition to, the DRT stability requirement that $\mathcal{T} > a_0$.

4.2   A first stress model

We now extend the analysis to include the effects of sociocultural influence and environmental stress, via introduction of an inverse index – ϕ – into Eq.(20), making the replacement $\mathcal{T} \to \phi\mathcal{T}$. More generally, any positive, monotonically increasing function of ϕ and $\mathcal{T}$ could be used, at the expense of algebraic complications. Their simple product serves as a first example.

Adapting the model from the Appendix of Wallace (2018), the entropy expression then becomes

$$\mathcal{S} = \mathcal{F}(\mathcal{T}, \phi) - \mathcal{T}\partial\mathcal{F}/\partial\mathcal{T} - \phi\partial\mathcal{F}/\partial\phi =$$
$$\phi\mathcal{T}(\log[\phi\mathcal{T}] + 2) \tag{25}$$

Taking ϕ as a fixed external index, the dynamic relation becomes

$$d\mathcal{T}_t = (\mu\partial\mathcal{S}/\partial\mathcal{T})dt + \sigma\mathcal{T}_t dW_t =$$
$$(\mu\phi(\log[\phi\mathcal{T}_t] + 3)dt + \sigma\mathcal{T}_t dW_t \tag{26}$$

Expanding $\log[\mathcal{T}_t]$ using the Ito Chain Rule gives

$$d\mathcal{T}/dt = \mu\phi\log[\mathcal{T}(t)\phi] + 2\mu\phi - \frac{1}{2}\mathcal{T}(t)\sigma^2 \tag{27}$$

There are, again, two nonequilibrium steady state solutions. Again, these are very roughly constrained by Jensen's inequality, with the larger one stable and the smaller either collapsing to zero or increasing toward the larger. The relations are

$$E(\mathcal{T}_L) \geq -\frac{1}{\sigma^2} 2\phi W[-1, \frac{-\sigma^2 \exp(-3)}{\mu\phi^2}]$$

$$E(T_S) \geq -\frac{1}{\sigma^2} 2\phi W[0, \frac{-\sigma^2 \exp(-3)}{\mu\phi^2}] \tag{28}$$

$W[-1, x]$, $W[0, x]$ are the -1 and 0 branches of the Lambert W-function. As above, a sufficiently large value of σ forces the upper and lower limits to coalesce, causing $\mathcal{T}$ to collapse to zero. Figure 6 displays such coalescence with increasing σ for the relation

$$\frac{-1}{\sigma^2} W[j, -\sigma^2], \ j = -1, 0$$

Setting the two different expressions for W in Eq.(28) equal to each other and solving for ϕ gives a stability condition in terms of σ and μ. The trick is to recognize that $W[-1, -x] = W[0, -x]$ at a branch point $x = \exp[-1]$. The stability condition for the inverse exposure index ϕ is then linear in σ:

$$\phi > \frac{\sigma \exp[-1]}{\sqrt{\mu}} \tag{29}$$

It is clearly necessary to make the same kind of expansion for ϕ as was done for Γ in Section 4.1.

4.3 A second stress model

Next, we explore the interaction between $\mathcal{T}$ and ϕ using a version of the classic 'vector borne infection' formalism.

First, normalize variates as $\hat{\mathcal{T}} \equiv \mathcal{T}/\mathcal{T}_{max}$, $\hat{\phi} \equiv \phi/\phi_{max}$. The interaction between them is then taken as (e.g., Bailey 1975)

$$d\hat{\mathcal{T}}/dt = \mu_1 \hat{\phi}(t)[1 - \hat{\mathcal{T}}(t)] - \gamma_1 \hat{\mathcal{T}}(t)$$

$$d\hat{\phi}/dt = \mu_2 \hat{\mathcal{T}}(t)[1 - \hat{\phi}(t)] - \gamma_2 \hat{\phi}(t) \tag{30}$$

The μ_i indicate positive feedback and the γ_i represent the rate of 'entropy' effects that decrease $\hat{\mathcal{T}}$ and $\hat{\phi}$.

Elementary calculation finds nonequilibrium steady state values for this system as

$$\hat{\mathcal{T}} \to \frac{\mu_1\mu_2 - \gamma_1\gamma_2}{\mu_2(\mu_1 + \gamma_1)}$$

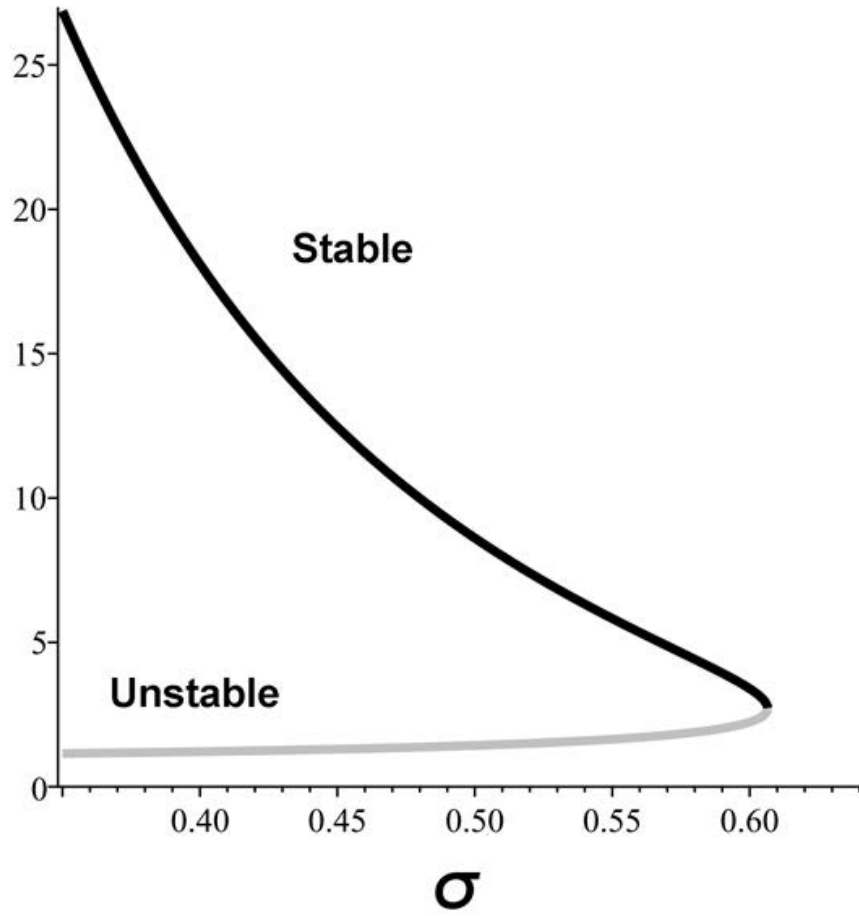$$\hat{\phi} \to \frac{\mu_1\mu_2 - \gamma_1\gamma_2}{\mu_1(\mu_2 + \gamma_2)} \tag{31}$$

Figure 6: Coalescence of stable and unstable nonequilibrium steady state modes of $-(1/\sigma^2)W[i, -\sigma^2], i = -1, 0$ with increasing σ. If σ becomes sufficiently large, regulation fails. The black line refers to the first expression of Eq.(28).

Clearly, no nonequilibrium steady state is possible unless $\mu_1\mu_2 > \gamma_1\gamma_2$. That is, the system collapses unless positive reinforcement effects are greater than entropy effects, in this model.

A simplified stochastic variant of Eq.(30) can be made by fixing the value of ϕ. Then

$$
\begin{aligned}
d\hat{\mathcal{T}}_t &= [\mu\phi(1 - \hat{\mathcal{T}}_t) - \gamma\hat{\mathcal{T}}_t]dt + \sigma\hat{\mathcal{T}}_t dW_t \\
&= [\mu\phi - (\mu\phi + \gamma)\hat{\mathcal{T}}_t]dt + \sigma\hat{\mathcal{T}}_t dW_t
\end{aligned}
\tag{32}
$$

This has the expectation

$$
E(\hat{\mathcal{T}}) = \frac{\mu\phi}{\mu\phi + \gamma}
\tag{33}
$$

Two points.

First, since this is an expectation, there will always be some probability that the system falls below the critical value for $\hat{\mathcal{T}}$ determined by the DRT.

Second, as γ rises, this probability significantly increases.

However, applying the Ito Chain Rule to $\hat{\mathcal{T}}^2$, after some calculation, finds

$$
E(\hat{\mathcal{T}}^2) = [\frac{\mu\phi}{\mu\phi + \gamma - \sigma^2/2}]^2
\tag{34}
$$

an expression that explodes if σ becomes large enough.

That is, the condition for stable variance in this model is

$$
\begin{aligned}
\sigma &< \sqrt{2(\mu\phi + \gamma)} \\
\phi &> \frac{1}{\mu}[\frac{\sigma^2}{2} - \gamma]
\end{aligned}
\tag{35}
$$

Rising σ can thus again trigger instability leading to rapid violation of the DRT condition at a fixed value of ϕ, although the expression, involving stability in variance, is qualitatively different from Eq.(29) and depends on the square of σ.

These results appear to suggest that the overall modeling approach might involve more general approximations of the form

$$
\phi > \sum_{k=0}^{m} a_k (\frac{\sigma}{\sqrt{\mu}})^k
$$

for constants $a_k, k = 0, ..., m$. Such series expansions may or may not have closed functional form.

Clearly, using these methods, there can be no single, one-size-fits-all, 'physics envy' format appropriate to all circumstances. These are regression-analog statistical models that must be fitted individually to data.

5 Through the lens of symmetry-breaking

5.1 Basic ideas

An almost-classic phase transition model for less regular circumstances emerges if it is possible to identify equivalence classes of a system's developmental pathways, for example 'healthy' and 'pathological'. This allows definition of a developmental symmetry groupoid for the developmental process (Wallace 2017; Weinstein 1996; Golubitsky and Stewart 2006). A groupoid is a generalization of an algebraic group in which a product is not necessarily defined between each element. The simplest example is, perhaps, a disjoint union of separate symmetry groups, but sets of equivalence classes also define a groupoid. See Wallace (2012, 2015a, 2017) or Weinstein (1996) for details and the Mathematical Appendix for a summary.

Another 'free energy' can then be defined that is liable to an analog of Landau's classical spontaneous symmetry breaking, in the Morse Theory sense (Pettini 2007; Matsumoto 2002). Under symmetry breaking, higher 'temperatures' are associated with more symmetric higher energy states in physical systems. Cosmology makes much of such matters in the first moments after the 'big bang': different physical phenomena break out as the universe cools. Here, for the control of cognitive processes, decline in $\mathcal{T}$, ϕ, or M from Section 3 can result in a sharply punctuated collapse from higher to lower symmetry states, causing serious failures across a broad spectrum of control processes.

Again, following Atlan and Cohen (1998), a system is taken as cognitive if it compares incoming signals with a learned or inherited picture of the world, then actively chooses a response from a larger set of those possible to it. Choice implies the existence of an information source, since it reduces uncertainty in a formal manner. The basic model holds even in the absence of an internal picture of the world.

Given a 'dual' information source associated with the inherently unstable cognitive system of interest, an equivalence class algebra can be constructed by choosing different system origin states and defining the equivalence of subsequent states at a later time by the existence of a high probability path connecting them to the same origin state. Disjoint partition by equivalence classes, analogous to orbit equivalence classes in dynamical systems, defines a symmetry groupoid associated with the cognitive process (Wallace 2017; Weinstein 1996; Golubitsky and Stewart 2006).

The equivalence classes across possible origin states define a set of information sources dual to different cognitive states available to the inherently unstable cognitive system or coalition of systems of interest. These create a large groupoid, with each orbit corresponding to an elementary 'transitive' groupoid whose disjoint union is the full groupoid. Each subgroupoid is associated with its own dual information source, and larger groupoids must have richer dual information sources than smaller.

Take X_{G_i} as the system's dual information source associated with groupoid element G_i. A Morse Function can now be built using an appropriate temper-

ature analog.

Perhaps the most obvious such 'temperature' involves some positive, monotonic increasing function of $\phi, \mathcal{T}$, the variables M, λ from Section 3, and $|\Omega|$ from Section 2.3, i.e.,

$$\hat{T} \equiv \hat{T}(\phi, \mathcal{T}, M, \lambda, |\Omega|).$$

Recall, however, that negative λ necessarily implies a 'pumped' state requiring high rates of M and may thus signal collapse in $\mathcal{T}$ with falling ϕ: these parameters are not independent. As a consequence, determining the functional form of $\hat{T}$ for a particular system may require detailed and difficult empirical or observational study, particularly for rapidly-shifting global broadcast systems characterized by the vector Ω.

The devil, as they say, is in the details.

Take $H(X_{G_i}) \equiv H_{G_i}$ as the Shannon uncertainty of the information source associated with the groupoid element G_i. It is possible to define another Boltzmann-like probability as

$$P[H_{G_i}] \equiv \frac{\exp[-H_{G_i}/\hat{T}]}{\sum_j \exp[-H_{G_j}/\hat{T}]} \tag{36}$$

The sum is over the different possible cognitive modes of the full system. The 'free energy' Morse Function F is then defined as

$$\exp[-F/\hat{T}] \equiv \sum_j \exp[-H_{G_j}/\hat{T}]$$
$$F = -\hat{T} \log[\sum_j \exp[-H_{G_j}/\hat{T}]] \tag{37}$$

Given the underlying groupoid generalized symmetries associated with high-order cognition, as opposed to simple control theory, it is possible to apply a version of Landau's symmetry-breaking approach to phase transition (Pettini 2007). The shift between such symmetries should remain highly punctuated in the 'temperature' $\hat{T}$, but in the context of what are likely to be far more complicated groupoid rather than group symmetries.

Arguing by abduction from physical theory, there should be only a few possible phases, with sharp and sudden transitions between them as $\hat{T}$ decreases.

5.2 Culture, pathology, and treatment

'Culture', the evolutionary anthropologist Robert Boyd asserts, 'is as much a part of human biology as the enamel on our teeth'. The symmetry-breaking perspective can be extended to include the effects of an embedding cultural environment – via an information source Z – on global broadcast mechanisms at different scales and levels of organization. The dual information source X_{G_i} then becomes a joint information source whose individual components are linked

by crosstalk, having a joint source uncertainty $H(X_{G^1}^i, X_{G^2}^j, ..., X_{G^m}^q)$. Given the embedding cultural information source Z, then the splitting criterion between high and low probability developmental trajectories is given by network information theory as the complicated sum

$$I(X_{G^1}^i, X_{G^2}^j, ..., X_{G^m}^q | Z) =$$
$$H(Z) + \sum H(X_{G^n}^j | Z) - H(X_{G^1}^i, X_{G^2}^j, ..., X_{G^m}^q | Z) \tag{38}$$

Eqs.(36) and (37) are then rewritten in terms of the splitting criterion

$$I((X_{G^1}^i, X_{G^2}^j, ..., X_{G^m}^q | Z).$$

We will call the new 'free energy' index – now influenced by embedding culture – $\mathcal{F}$.

We have, in essence, produced a kind of 'Mathematical Kleinman Theory', representing in a formal way something of the observations of Kleinman (1991), Kleinman and Cohen (1997), and their colleagues who have explored the profound differences in the expression and experience of mental disorders across cultures.

Indeed, Wallace et al. (2009, Section 6.2.4) examine the influence of culture on malaria in the Fulani lands of West Africa, finding that former masters and former slaves suffer markedly different disease phenotypes, presumably related to the divergence in their cultural circumstances and historical trajectories.

It is possible to examine sufficient conditions for the intractable stability of a pathological 'ground state' condensation representing control system collapse via the Stochastic Stabilization Theorem (Appleby et al. 2008; Mao 2007) in a particular embedding cultural milieu.

We assume a multidimensional vector of parameters associated with that phase, J, *that measures deviations from the pathological ground state*. The free energy measure from the generalization in Eq.(38) allows definition of another 'entropy' in terms of the Legendre transform

$$\hat{S} \equiv \mathcal{F}(J) - J \cdot \nabla_J \mathcal{F} \tag{39}$$

Now we write another first-order 'Onsager' dynamic equation in the gradients of $\hat{S}$:

$$dJ_t = f(J_t, t)dt + \sigma(J_t, t)dW_t \tag{40}$$

dW_t is multi-dimensional white noise vector and $\sigma(J_t, t)$ is a multidimensional matrix function.

Again, $f(J_t, t)$ is a first-order 'diffusion' equation in the gradients of $\hat{S}$ by J.

The base equation $dX/dt = f(J, t)$, after some delay, under normal conditions of recovery from a pathological state, will have a solution $|J(t)| \to \infty$, implying that there must be a transition to a more healthy stable but nonequilibrium steady state. Recall that even a stable nonequilibrium state may remain stable only so long as it is provided a sufficient rate of metabolic free energy.

Successful organisms, species, and/or more complex colonial systems, all have long-evolved remedial mechanisms akin to the immune system, cancer suppression, wound healing, and suchlike.

However, the multidimensional version of the Stochastic Stabilization Theorem (Appleby et al. 2008) ensures that, under very broad conditions, sufficient noise – a big enough 'symmetric' form of the multidimensional noise matrix σ – will drive $|J(t)|$ logarithmically to zero, stabilizing the pathological mode in spite of internal remedial efforts. Damage accumulation, aging, starvation and so on come to mind.

Conversely, however, Appleby et al. (2008) also show that, for a system of dimension ≥ 2, a noise matrix can always be found that destabilizes an inherently stable system, i.e., one for which $|J_t| \to 0$, in this context, a persistent pathological condition for the organism or colony. That is, a 'treatment' can be found that causes a transition to a different nonequilibrium steady state. Iatrogenic intervention makes the individual sicker, proper treatment heals. In many cases, of course, successful treatment is simply not realistic: 'First, do no harm'.

What should be evident from the previous section is that, for humans, culture will be inherently convoluted not only with patterns of health and illness, but with successful modalities of treatment (Wallace 2015b; Wallace et al. 2009, Section 6.2; Kleinman 1991). Treatment, in the sense of Kleinman, will itself be 'culture-bound'.

For nonergodic systems, where time averages are not the same as ensemble averages, the groupoid symmetries become 'trivial', associated with the individual high probability paths for which an H-value may be defined, although it cannot be represented in the form of the usual Shannon 'entropy' (Khinchin 1957, p. 72). Then equivalence classes must be defined in terms of other similarity measures for different developmental pathways. The arguments of this section regarding pathological modes and their treatment then follow through. See Wallace (2018) for details.

6 Discussion and conclusions

A most characteristic marker of the living state is its dynamic cognition at almost every scale and level of organization, from the subcellular to the sociocultural. For humans in particular, cognition at and across scales and levels of organization is deeply embedded in a cultural milieu that affects even the most basic modalities (e.g., Nisbett and Miyamoto 2005; Wallace 2007, 2015a). Correct biocognitive function is particularly important during critical developmental periods of an organism, and the failure of regulation of gene expression causes a broad spectrum of serious developmental disorders.

Machine systems have yet to match either this cross-scale cognitive density, or the ubiquity of Barrs' tunable global broadcast mechanism that recruits underlying simpler cognitive modules into working coalitions, again at various scales and levels of organization (Wallace 2012). Indeed, while cutting-edge com-

puters can play the game of Go very well indeed, they cannot quickly switch to chess, poker, playing with a cat, or driving home in rush-hour traffic, all of which are within reach of a human Go master. Indeed, the Go master can do all of these while his or her immune system is successfully fighting off some minor infection and conducting routine cellular maintenance.

Cognition in living entities (and their social or sociocultural constructs or artifacts) is necessarily as complex and dynamic as their embedding environments. Only since the middle of the last century have formal tools been available that can address such complicated dynamics. The asymptotic limit theorems of probability, and of information and control theories, permit construction of empirical 'regression models' for biocognition and its dynamics. These models may, as have their simpler analogs, become useful in the study and remediation of cognitive failure at and across the scales, levels of organization, and dynamic coalitions, that constitute and drive the phenomena of life.

Here, we have outlined the basic argument, and provided a number of examples of such 'regression models' that might be fitted to observational or experimental data. Many more can, of course, be constructed from the essential formal building blocks provided, and via the introduction of other such elements. One particular interest has been the impact of environmental contamination and/or psychosocial stress on biocognitive failure, in a large sense, affecting developmental trajectories across scale and organizational level. According to the model, such factors can precipitate collapse to treatment-resistant pathological 'ground states' across those scales and levels of organization, suggesting a basis for both environmental disorders and for the 'side effects' of treatments addressing them. Such matters are, of course, well-known in other vocabularies, but the analysis here suggests a comprehensive and systematic approach to understanding the impact of environment and of treatment-as-environment on developmental disorders and their remediation.

We are led to suggest that, for humans, the central influences of culture and its socioeconomic instantiations in developmental processes at all scales and levels of organization suggests that virtually all developmental pathologies must be viewed as fundamentally culture-bound. This has deep and obvious implications for both individual treatment modalities and for large-scale public health policies.

7   Mathematical appendix

7.1   The Black-Scholes model

Take $\mathcal{H}(\Gamma)$ as the control information rate 'cost' of stability at the index level Γ. What is the mathematical form of $\mathcal{H}(\Gamma)$ under conditions of volatility i.e., variability in Γ proportional to it? Let

$$d\Gamma_t = g(t, \Gamma_t)dt + b\Gamma_t dW_t \tag{41}$$

where dW_t is taken as white noise and the function $g(t, \Gamma)$ will 'fall out' of

the calculation on the assumption of certain regularities outlined in Black and Scholes (1973).

Let $\mathcal{H}(\Gamma_t, t)$ be the minimum needed incoming rate of control information under the Data Rate Theorem, and expand in Γ using the Ito chain rule (Protter 1990)

$$
\begin{aligned}
d\mathcal{H}_t = [\partial\mathcal{H}/\partial t + g(\Gamma_t, t)\partial\mathcal{H}/\partial\Gamma + \frac{1}{2}b^2\Gamma_t^2\partial^2\mathcal{H}/\partial\Gamma^2]dt \\
+ [b\Gamma_t\partial\mathcal{H}/\partial\Gamma]dW_t
\end{aligned} \tag{42}
$$

Define a quantity L as a Legendre transform of the rate $\mathcal{H}$, by convention having the form

$$
L = -\mathcal{H} + \Gamma\partial\mathcal{H}/\partial\Gamma \tag{43}
$$

Since $\mathcal{H}$ is an information index, it is a kind of free energy in the sense of Feynman (2000) and L is a classic entropy measure.

Heuristically, replacing dX with ΔX in these expressions and applying Eq.(42),

$$
\Delta L = (-\partial\mathcal{H}/\partial t - \frac{1}{2}b^2\Gamma^2\partial^2\mathcal{H}/\partial\Gamma^2)\Delta t \tag{44}
$$

As in the classical Black-Scholes model (Black and Scholes 1973), the terms in g and dW_t 'cancel out', and the effects of noise are subsumed into the Ito correction factor, a regularity assumption making this an exactly solvable but highly approximate model.

The conventional Black-Scholes calculation takes $\Delta L/\Delta T \propto L$. Here, at nonequilibrium steady state, we assume $\Delta L/\Delta t = \partial\mathcal{H}/\partial t = 0$, so that

$$
-\frac{1}{2}b^2\Gamma^2\partial^2\mathcal{H}/\partial\Gamma^2 = 0 \tag{45}
$$

By inspection,

$$
\mathcal{H} = \kappa_1\Gamma + \kappa_2 \tag{46}
$$

where the κ_i are nonnegative constants.

7.2 Groupoids

Given a pairing, connection by a meaningful path to the same basepoint, it is possible to define 'natural' end-point maps $\alpha(g) = a_j, \beta(g) = a_k$ from the set of morphisms G into the set A containing the a_k, and a formally associative product in the groupoid g_1g_2 provided $\alpha(g_1g_2) = \alpha(g_1), \beta(g_1g_2) = \beta(g_2)$, and $\beta(g_1) = \alpha(g_2)$. Then the product is defined, and associative, i.e., $(g_1g_2)g_3 = g_1(g_2g_3)$, with inverse defined by $g = (a_j, a_k), g^{-1} \equiv (a_k, a_j)$.

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 $a_j \sim Ga_k$ 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. As Weinstein (1996) explains, 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. A criticism sometimes applied to groupoid theory is that their classification up to isomorphism is nothing other than the classification of equivalence relations via the orbit equivalence relation and groups via the isotropy groups. The imposition of a compatible topological structure produces a nontrivial interaction between the two structures. Below we will introduce a metric structure on manifolds of related information sources, producing such interaction.

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

As Weinstein (1996) points out, 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 (isomorphic), but *it also parameterizes the different ways (isomorphisms) in which two elements can be equivalent*, i.e., 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.

Brown (1987) describes the basic structure as follows:

> A groupoid should be thought of as a group with many objects, or with many identities... A groupoid with one object is essentially just a group. So the notion of groupoid is an extension of that of groups. It gives an additional convenience, flexibility and range of applications...
>
> EXAMPLE 1. A disjoint union [of groups] $G = \cup_\lambda G_\lambda, \lambda \in \Lambda$, is a groupoid: the product ab is defined if and only if a, b belong to the same G_λ, and ab is then just the product in the group G_λ. There is an identity 1_λ for each $\lambda \in \Lambda$. The maps α, β coincide and map G_λ to $\lambda, \lambda \in \Lambda$.
>
> EXAMPLE 2. An equivalence relation R on [a set] X becomes a groupoid with $\alpha, \beta : R \to X$ the two projections, and product $(x, y)(y, z) = (x, z)$ whenever $(x, y), (y, z) \in R$. There is an identity,

namely (x, x), for each $x \in X$...

Weinstein (1996) makes the following fundamental point:

> Almost every interesting equivalence relation on a space B arises in a natural way as the orbit equivalence relation of some groupoid G over B. Instead of dealing directly with the orbit space B/G as an object in the category S_{map} of sets and mappings, one should consider instead the groupoid G itself as an object in the category G_{htp} of groupoids and homotopy classes of morphisms.

It is, in fact, possible to explore homotopy in paths generated by information sources.

7.3   Morse Theory

Morse theory examines 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 'second order' iterations involving parameter manifolds determining critical behavior, for example sudden onset of a giant component in a network model. We follow Pettini (2007).

The central argument 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

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

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 $\mathcal{J}$ of second derivatives at x_c, whose elements, in terms of local coordinates are

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

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 $\mathcal{J}$ 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 Pettini (2007), the essential results of Morse theory are as follows:

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 $\mathcal{J}$ 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 $0 \le r \le \infty$.

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 $\mu_i (i = 0, ..., m)$ 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.

Matsumoto (2002) and Pettini (2007) provide details and further references.

8 Culturally-specific generalized inflammation

We summarize Wallace (2015b, Sec. 8.5), which is not widely available. The stresses that activate cognitive modules and their shifting coalitions are not random sequences of perturbations and not independent of culturally-modulated perception. That is, stress responses involve highly correlated, grammatical, syntactical processes by which an embedding psychosocial environment communicates with an individual's cognitive physiological and mental hierarchy, strongly structured by the power relations between groups. The stress experienced by an individual can thus be taken as another adiabatically piecewise stationary ergodic (APSE) information source, interacting with a set of dual information sources defined by cognition at different scales and levels of organization.

We now define stress by empirical observation, determining the concentrations of HPA axis hormones and other biochemical signals according to an appropriate natural time frame, taken as the inherent period of the system. Under normal circumstances, i.e., in the absence of extraordinary meaningful psychosocial stress, we measure a series of n biomarker concentrations at time t represented as an n-dimensional vector X_t. We conduct numerous experiments and/or observations, creating a regression model so that, in the absence of extreme perturbation, and to first order, the concentration of biomarkers at time $t+1$ can be written in terms of that at time t using a matrix equation of the form

$$X_{t+1} \approx < \mathbf{R} > X_t + b_0 \tag{47}$$

where $< \mathbf{R} >$ is a matrix of regression coefficients, b_0 a vector of constant terms.

Under modest perturbation by a non-random, structured, stress,

$$\begin{aligned} X_{t+1} &= (< \mathbf{R} > + \delta \mathbf{R}_{t+1}) X_t + b_0 \\ &\equiv < \mathbf{R} > X_t + \epsilon_{t+1} \end{aligned} \tag{48}$$

where b_0 and $\delta \mathbf{R}_{t+1} X_t$ have been absorbed into a vector ϵ_{t+1} of 'error terms' that, importantly, *are not necessarily small*.

Recognize that this is not a population process whose continuous analog is exponential growth. Rather this expression involves the passage of a signal, structured psychosocial stress, through some physiological and/or mental filter that can change the structure of that signal.

Taking the matrix of regression coefficients $< \mathbf{R} >$ as sufficiently regular, it becomes possible to Jordan block diagonalize it using the matrix of its column eigenvectors $\mathbf{Q}$ so that

$$\mathbf{Q} X_{t+1} = (\mathbf{Q} < \mathbf{R} > \mathbf{Q}^{-1}) \mathbf{Q} X_t + \mathbf{Q} \epsilon_{t+1} \tag{49}$$

or,

$$Y_{t+1} = < \mathbf{J} > Y_t + W_{t+1} \tag{50}$$

where $Y_t \equiv \mathbf{Q}X_t, W_{t+1} \equiv \mathbf{Q}\epsilon_{t+1}$, and $< \mathbf{J} > \equiv \mathbf{Q} < \mathbf{R} > \mathbf{Q}^{-1}$ is a block diagonal matrix in terms of the eigenvalues of $< \mathbf{R} >$.

Thus the rate-distorted writing of structured stress on the affected physiological submodules through $\delta\mathbf{R}_{t+1}$ is reexpressed in terms of the vector W_{t+1}.

The sequence of W_{t+1} is then the rate-distorted image of the information source defined by the system of external structured psychosocial stress. This formulation permits estimation of the nonequilibrium steady-state effects of that image on underlying cognitive physiological and mental modules and dynamics. Since everything is APSE, it becomes possible to either time or ensemble average both sides of Eq.(50), so that the one-period offset is absorbed in the averaging, giving a nonequilibrium steady state relation

$$< Y >=< \mathbf{J} >< Y > + < W > \tag{51}$$

equivalent to

$$< Y >= (\mathbf{I}- < \mathbf{J} >)^{-1} < W > \tag{52}$$

where $\mathbf{I}$ is the $n \times n$ identity matrix.

It is now possible to *reverse the argument*, choosing Y_k to be a fixed eigenvector of $< \mathbf{R} >$. Expressing the diagonalization of $< \mathbf{J} >$ in terms of its eigenvalues gives the average excitation of the generalized physiological stress response in terms of an appropriate eigentransformed pattern of exciting perturbations as

$$< Y_k >= \frac{< W_k >}{1- < \lambda_k >} \tag{53}$$

where $< \lambda_k >$ is the eigenvalue of $< Y_k >$, and $< W_k >$ is a similarly appropriately transformed set of ongoing perturbations by structured psychosocial stress.

In consequence, there will be a culturally characteristic form of perturbation by structured psychosocial stress, by the W_k, that will resonantly excite a particular eigenmode of the generalized physiological stress response. Conversely, by tuning the eigenmodes of $< \mathbf{R} >$ the generalized stress response can be trained to galvanized excitation in the presence of particular forms of perturbation.

This is because, if $< \mathbf{R} >$ has been appropriately determined from regression relations, then the λ_k will be a kind of multiple correlation coefficient (Wallace and Wallace 2000) so that particular eigenpatterns of perturbation will have greatly amplified impact on the generalized inflammatory response. If $\lambda = 0$ then perturbation has no more effect than its own magnitude. However, if $\lambda \to 1$, then the written image of a perturbing psychosocial stressor will have very great impact. A system with $\lambda \approx 0$ can be characterized as locally-resilient since its response is no greater than the perturbation itself.

Learning by the cognitive physiological modules is, in this model, the process of tuning response to perturbation. This is why we have written $< \mathbf{R} >$ instead of simply $\mathbf{R}$: the regression matrix is a tunable set of variables. The argument could be greatly simplified by invoking the tuning theorem variant of the coding theorem (Wallace 2015b, Sec. 1.1).

9   Acknowledgments

References

APA 2013, American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders: DSM-5. Washington, DC.

Appleby, J., X. Mao, A. Rodkina, 2008, Stabilization and destabilization of nonlinear differential equations by noise, IEEE Transactions on Automatic Control, 53:126-132.

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

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

Bailey, N., 1975, The Mathematical Theory of Infectious Diseases and its Applications, Second Edition, Hafner, New York.

Bebbington, P., 1993, Transcultural aspects of affective disorders, International Review of Psychiatry 5:145-156.

Ben-Sharchar, D., 2002, Mitochondrial dysfunction in schizophrenia: a possible linkage to dopamine, Journal of Neurochemistry 83:1241-1251.

Black, F., M. Scholes, 1973, The pricing of options and corporate liabilities, Journal of Political Economy, 81:637-654.

Boyd, S., N. Parikh, E. Chu, B. Peleato, J. Eckstein, 2010, Distributed optimization and statistical learning via the alternating direction method of multipliers, Foundations and Trends in Machine Learning, 3:1-122.

Brown, R., 1987, From groups to groupoids: a brief survey, Bulletin of the London Mathematical Society, 19:113-134.

Clay, H., S. Sullivan, C. Konradi, 2011, Mitochondrial dysfunction and pathology in bipolar disorder and schizophrenia, International Journal of Developmental Neuroscience 29:311-324.

Compton, W., K. Conway, F. Stinson, J. Colliver, B. Grant, 2005, Prevalence, correlates, and comorbidity of DSM-IV antisocial personality syndromes and alcohol and specific drug disorders in the United States: Results from the national epidemiologic survey on alcohol and related conditions, Journal of Clinical Psychiatry 66:677-685.

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

Craig, T., C. Siegel, K. Hopper, S. Lin, N. Sartorius, 1997, Outcome in schizophrenia and related disorders compared between developing and developed countries A recursive partitioning re-analysis of the WHO DOSMD data, British Journal of Psychiatry, 170:22933.

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

Dekel, S., Z. Solomon, D. Horesh, T. Ein-Dor, 2014, Posttraumatic stress disorder and depressive sysmptoms: joined or independent sequelae of trauma? Journal of Psychiatric Research, 54:64-69.

Dekel, S., T. Ein-Dor, J. Rosen, G. Bonanno, 2017, Differences in cortisol response to trauma activation in individuals with and without comorbid PTSD and Depression, frontiers in Psychology, doi: 10.3389/fpsyg.2017.00797.

Eshel, N., M. Bukwich, V. Rao, V. Hemmeider, J. Tian, N. Uchida, 2015, Nature, 243-246.

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

Giulivi, C., Y. Zhang, A. Omanska-Klusek, C. Ross-Inta, S. Wong, I. Hertz-Picciotto, F. Tassone, I. Pessah, 2010, Mitochondrial dysfunction in autism, Journal of the American Medical Association 304:2389-2395.

Glazebrook, J.F., R. Wallace, 2009, Rate distortion manifolds as model spaces for cognitive information, Informatica, 33:309-346.

Goh, S., Z. Dong, Y. Zhang, S. DiMauro, B. Peterson, 2014, Mitochondrial dysfunction as a neurobiological subtype of Autism Spectrum Disorder, JAMA Psychiatry doiZ:10.1001/jamapsychiatry.2014.179.

Golubitsky, M., I. Stewart, 2006, Nonlinear dynamics and networks: the groupoid formalism, Bulletin of the American Mathematical Society, 43:305-364.

Grant, B., D. Hasin, F. Stinson, D. Dawson, S. Chou, W. Ruan, R. Pickering, 2004, Prevalence, correlates, and disability of personality disorders in the United States: Results from the national epidemiologic survey on alcohol and related conditions, Journal of Clinical Psychiatry 7:948-958.

Hare, R., S. Hart, T. Harpur, 1991, Psychopathology and the DSM-IV criteria for antisocial personality disorder, Journal of Abnormal Psychology 100:391-398.

Harrison, G., K. Hopper, T. Craig, et al., 2001, Recovery from psychotic illness: a 15- and 25-year international follow-up study, British Journal of Psychiatry, 178:50617.

Heine, S., 2001, Self as cultural product: An examination of East Asian and North American selves, Journal of Personality 69:881-906.

Heinrich, R., S. Schuster, H.G. Holzhutter, 1991, Mathematical analysis of enzymatic reaction systems using optimization principles, European Journal of Biochemistry, 201:1-21.

Hwang, C., A. Masud, 1979, Multiple Objective Decision Making, Methods and Applications, Springer, New York.

Hwu, H., E. Yeh, L. Chang, 1989, Prevalence of psychiatric disorders in Taiwan defined by the Chinese diagnostic interview schedule, Acta Psychiatrica Scandinavia 79:136-147.

Jablonsky A., N. Sartorius, G. Ernberg, M. Anker M, 1992, Schizophrenia: manifestations, incidence and course in different cultures: a World Health Organization ten-country study, Psycholical Medicine Monographs, Supplement, 20:197.

Jenkins, J., A. Kleinman, B. Good, 1990, Cross-cultural studies of depression. In Advances in Mood Disorders: Theory and Research, (Becker, J., A. Kleinman eds.), L. Erlbaum, Los Angeles CA.

Jin, H., Z. Xu, X. Zhou, 2008, A convex stochastic optimization problem arising from portfolio selection, Mathematical Finance, 18:171-183.

Kessler, R., K. McGonagle, S. Zhao, C. Nelson, M. Hughes, S. Eshleman, H. Wittchen, K. Kendler, 1994, Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey, Archives in General Psychiatry 51:8-19.

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

Kleinman, A., B. Good, 1985, Culture and Depression: Studies in the Anthropology of Cross-Cultural Psychiatry of Affect and Depression, University of California Press, Berkeley, CA.

Kleinman, A., 1991, Rethinking Psychiatry: From cultural category to personal experience, Free Press, New York.

Kleinman, A., A. Cohen, 1997, Psychiatry's global challenge, Scientific American 276(3):86-89.

Kohrt, B., A. Rasmussen, B. Kaiser, E .Haroz, S. Maharjan, B. Mutamba, J. de Jong, D. Hinton, 2014, Cultural concepts of distress and psychiatric disordes: literature review and research recommendations for global mental health epidemiology, International Journal of Epidemiology, 43:365-406.

Lewis-Fernandez, R., L. Kirmayer, P. Guarnaccia, P. Ruiz, 2017, Cultural Concepts of Distress and Assessment. Chapter 29 in Sadock et al. (eds.), Comprehensive Textbook of Psychiatry Vol II, 10th edition, Kulwer, New York.

Lobel, I., A. Ozdaglar, D. Feijer, 2011, Distributed multi-agent optimization with state-dependent communication, Mathematical Programming B, 129:255-284.

Manson, S., 1995, Culture and major depression: current challenges in the diagnosis of mood disorders, Psychiatric Clinics of North America 18:487-501.

Mao, X., 2007, Stochastic Differential Equations and Applications, Second Edition, Woodhead Publishing, Philadelphia.

Markus, H., S. Kitayama, 1991, Culture and self-implications for cognition, emotion, and motivation, Psychological Review 98:224-253.

Matsuda, T., R. Nisbett, 2006, Culture and change blindness, Cognitive Science: A Multidisciplinary Journal 30:381-399.

Matsumoto, Y., 2002, An Introduction to Morse Theory, American Mathematical Society, Providence, RI.

Maturana, H., F. Varela, 1980, Autopoiesis and Cognition, Reidel, the Netherlands.

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

Nisbett, R., K. Peng, C. Incheol, A. Norenzayan, 2001, Culture and the system of thought: Holistic vs. analytic cognition, Psychological Review 108:291-310.

Nisbett, R., Y. Miyamoto, 2005, The influence of culture: holistic versus analytic perception, TRENDS in Cognitive Sciences, 10:467-473.

Oquendo, M., G. Echavarra, H. Galfalvy, M. Grunebaum, A. Burke, A. Barrera, 2003, Lower cortisol levels in depressed patients with comorbid post-traumatic stress disorder, Neuropsychopharmacology, 28:591-598.

Palmieri, L., A. Persico, 2010, Mitochondrial dysfunction in autism spectrum disorders: cause or effect? Biochemica et Biophysica Acta 1797:1130-1137.

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

Prabakaran, S., J. Swatton, M. Ryan, S. Huffaker, J. Huang, J. Griffin, M. Wayland, et al., 2004, Mitochondrial dysfunction in schizophrenia: evidence for compromised brain metabolism and oxidative stress, Molecular Psychiatry 9:6840697.

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

Rossignol, R., R. Frye, 2010, Mitochondrial dysfunction in autism spectrum disorders: a ssytematic review and meta-analysis, Molecular Psychiatry 17:290-314.

Sartorius N., R. Shapiro, M. Kimura, K. Barrett, 1972, WHO international pilot study of schizophrenia, Psychological Medicine, 2: 42225.

Sartorius N., A. Jablensky, R. Shapiro, 1977, Two-year follow-up of the patients included in the WHO International Pilot Study of Schizophrenia, Psycholical Medicine, 7:52941.

Scaglia, F., 2010, The role of mitochondrial dysfunction in psychiatric disease, Developmental Disabilities Research Reviews 16:136-143.

Schrodinger, E., 1989, Statistical Thermodynamics, Dover Publications, New York.

Schultz, W., P. Dayan, P. Montague, 1997, A neural substrate of prediction and reward, Science, 1593-1598.

Shannon, C., 1959, Coding theorems for a discrete source with a fidelity criterion, Institute of Radio Engineers International Convention Record 7:142-163.

Shao, L., M. Martin, S. Watson, A. Schatzberg, H. Akil, R. Myers, E. Jones, W. Bunney, M. Vawter, 2008, Mitochondrial involvement in psychiatric disorders, Trends in Molecular Medicine 40:281-295.

Silverstein, S., M. Wibral, W. Phillips, 2017, Implications of information theory for computational modeling of schizophrenia, Computational Psychiatry, 1:82-101.

Wallace, D., R. Wallace, 2000, Life and death in Upper Manhattan and the Bronx: toward evolutionary perspectives on catastrophic social change, Environment and Planning A, 32:1245-1266.

Wallace, R., D. Wallace, R.G. Wallace, 2009, Farming Human Pathogens: Ecological resilience and evolutionary process, Springer, New York.

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

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

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

Wallace, R., 2015a, Closed-system 'economic' models for psychiatric disorders: Western atomism and its culture-bound syndromes Cognitive Processing, 16:279-290.

Wallace, R., 2015b, An Information Approach to Mitochondrial Dysfunction: Extending Swerdlow's Hypothesis, World Scientific, Singapore.

Wallace, R., 2016, High metabolic demand in neural tissues: information and control theory perspectives on the synergism between rate and stability, Journal of Theoretical Biology, 409:86-96.

Wallace, R., 2017, Information Theory Models of Instabilities in Critical Systems, World Scientific, Singapore.

Wallace, R., 2018, New statistical models of nonergodic cognitive systems and their pathologies, Journal of Theoretical Biology, 436:72-78.

Wallace, R., D. Wallace, 2016, Gene Expression and its Discontents: The Social Production of Chronic Disease, Second Edition, Springer, New York.

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

Weissman, M., R. Bland, G. Canino, S. Greenwald et al., 1994, The cross-national epidemiology of obsessive-compulsive disorder. The cross-national collaborative group. Journal of Clinical Psychiatry, March 55 Supplement pp.5-10.

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

Wolpert, D., W. MacReady, 1997, No free lunches, IEEE Transactions on Evolutionary Computation, 9:731-735.

World Health Organization, 1981, Schizophrenia: an International Follow-up Study. Chichester, Wiley, UK.