

EXTENDING INFORMATION THEORY TO MODEL DEVELOPMENTAL DYSFUNCTION

Rodrick Wallace

Division of Epidemiology

The New York State Psychiatric Institute

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

(Dated: May 4, 2017)

A combination of directed homotopy topological and Morse theoretic methods can significantly extend control and information theories, permitting deeper understanding of ‘developmental’ pathologies afflicting a broad spectrum of biological, psychological, socioeconomic, machine, and hybrid processes across different time scales and levels of organization. Such pathologies emerge as phase transitions driven by synergistic forms of environmental insult under stochastic circumstances, causing ‘comorbid condensations’ through groupoid symmetry breaking. The resulting statistical models should be useful for the analysis of experimental and observational data in many fields.

Key Words: algebraic topology; control theory; directed homotopy; fog-of-war; groupoid; ontology; pathology

I. INTRODUCTION

Developmental process – ontology – is ubiquitous across vast biological, social, economic, and machine realms. Rosen (2012) characterizes this as ‘...anticipatory behavior at all levels of... organization’. Maturana and Varela (1980) see cognition permeating biology. Atlan and Cohen (1998) invoke a ‘cognitive paradigm’ for the immune system that generalizes to wound healing, blood pressure regulation, neural dynamics, and so on (Wallace 2012). West-Eberhard (2003; 2005) sees ontology as a matter of ‘choice’ at developmental branch points. Traffic flow involves repeated ‘ontological’ choices by atomistic vehicles at road junctions, as well as during ordinary passage in heavy traffic (Wallace 2016a Ch.9). Indeed, machine cognition quite generally requires repeated choice of response to environmental cues (Wallace 2016a). A firm responding to market pressures must, at least annually, reconfigure product lines and marketing strategies, also a cognitive process (e.g., Wallace 2015 and references therein). Democratic state actors confronted by changing patterns of threat and affordance must, at least during elections, repeatedly choose among the different patterns of response made available by the contending parties and candidates. Active warfare involves constantly repeated choice at all levels of organization leading up to, and during, combat operations.

All developmental phenomena are, however, subject to patterns of failure and dysfunction. These range from neurodevelopmental disorders such as autism and schizophrenia (Wallace 2016b) to collapse of vehicle flow in traffic jams (Kerner and Klenov 2009), and catastrophes of governance like Brexit, or the US occupation of Iraq. Here, we attempt to extend results from information and control theories to statistical tools useful in un-

derstanding developmental failure.

II. THE BASIC IDEA

The underlying model of development is that of figure 1, in which a system begins at some initial ‘phenotype’ S_0 confronting a branch point C_0 leading to two (or more) possible subsequent ‘phenotypes’ S_1 and S_2 , where new branch points C_1 and C_2 will be confronted, and at which choices must again be made, and so on.

Two of the three essential components of this model are intrinsically linked.

The first component is that of *directed homotopy*, in the sense of Grandis (2009) and Fajstrup et al. (2016). That is, there are equivalence classes of paths leading from ‘phenotype’ S_n to S_{n+1} , as defined by the branch conditions C_n . A group structure – the so-called ‘fundamental group’ – is imposed on a geometric object by convolution of loops within it that can be reduced without crossing a hole (e.g., Hatcher 2001). An algebraic topology of directed homotopy in figure 1 can be constructed from the composition of paths that constitutes a groupoid (Weinstein 1996), an object in which a product need not be defined between every possible object, here the equivalence classes of possible linear paths. See the Mathematical Appendix for an introduction to groupoid theory. As Weinstein (1996) emphasizes, almost every interesting equivalence relation on a space B arises in a natural way as the orbit equivalence relation of some groupoid G over that space. Instead of dealing directly with the orbit quotient space B/G as an object in the category of sets and mappings, one should consider instead the groupoid G itself as an object in the category of groupoids and homotopy classes of morphisms. An exactly similar perspective involves use of the homotopy and homology groups of algebraic topology to characterize complicated geometric objects (Hatcher 2001).

The second component is recognition that choice at developmental branch points involves active selection of one possible subsequent path from a larger number that

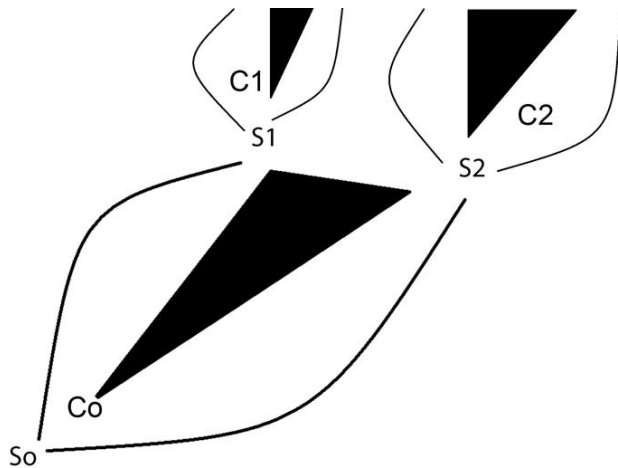


FIG. 1: Starting at an initial developmental ‘phenotype’ So , a critical point Co casts a developmental shadow. There are two directed homotopy equivalence classes of deformable paths leading to ‘phenotypes’ $S1$ and $S2$, followed by similar critical points $C1$ and $C2$. Convolution of equivalence classes creates a groupoid, an algebraic structure in which a product between classes is not necessarily defined, unlike the group properties of homotopy loops on geometric objects. Repeated critical points C_n over the developmental trajectory systematically enlarge the fundamental groupoid characterizing the developmental process.

may be available. This is often done, in the sense of Atlan and Cohen (1998), by comparison of ‘sensory’ data with an internalized – learned or inherited – picture of the world, and upon that comparison, an active choice of response is made from a larger number of those possible. Rosen (2012) invokes ‘anticipatory models’ for such processes. Following the Atlan/Cohen model, choice involves reduction in uncertainty, and reduction in uncertainty implies the existence of an information source that we will call ‘dual’ to the underlying cognitive process. See Wallace(2012) for a somewhat more formal treatment.

What is clear is that the dual information source or sources associated with developmental process must be deeply coupled with the underlying groupoid symmetries characterizing development. As development proceeds, the groupoid symmetry becomes systematically richer.

Two points:

First, information, as Feynman (1996) argues, is not ‘entropy’, rather it can be viewed as a form of free energy. Indeed, Feynman (1996), following Bennett, constructs a quite nice little idealized machine that turns the information within a message into useful work.

Second, groupoids are almost groups, and it becomes possible to apply Landau’s symmetry breaking/making arguments to the dual information sources characterizing developmental process (Pettini 2007). In that theory, phase transitions are recognized in terms of sudden shifts in the underlying group symmetries available to the sys-

tem at different temperatures. High temperatures, with the greatest available energies, have the greatest possible symmetries. Symmetry breaking occurs in terms of the sudden nonzero value of some ‘order parameter’ like magnetization at a sufficiently low critical temperature.

For a road network based on figure 1, the ‘order parameter’ would be the number of road turnoffs blocked by a traffic jam. The temperature analog is an inverse function of the linear vehicle density (Kerner and Klenov 2009; Wallace 2016a).

The third component of the model explored here looks in detail at the embedding regulatory apparatus that must operate at each branch point to actively choose a path to the desired ‘phenotype’. This requires exploration of the intimate connection between control and information theories represented by the Data Rate Theorem (Nair et al. 2007).

III. CONTROL THEORY

There is no cognition without regulation: the rapidly fluctuating ‘stream of choice’ must be constrained within regulatory ‘riverbanks’ that represent an image of the real world in which the cognitive system is embedded (Atlan and Cohen 1998). Cognitive choice is then, for the most part, inherently unstable in the formal control theory sense of the Data Rate Theorem (Nair et al. 2007), and must be stabilized by provision of control information at a critical rate – constant making and remaking of the riverbanks. Free elections have guarded polling places. Traffic streams face traffic lights and laws, and vehicles must be actively driven. Outrageous advertising claims, except in politics, can meet draconian legal interventions, and law enforcement generally attempts to ensure conformance with the ‘riverbanks’ of government directives.

Failure to provide control information at or above a critical rate initiates characteristic modes of system failure that are – in a large sense – expressed as ‘developmental disorders’.

Assuming an approximate nonequilibrium steady state, the simplest model of deviations from that state – described in terms of an n -dimensional vector of observables x_t at time t – has a linear form:

$$x_{t+1} = \mathbf{A}x_t + \mathbf{B}u_t + W_t \quad (1)$$

where x_{t+1} is the state at time $t + 1$, u_t is an imposed n -dimensional control signal vector at time t , W_t is an added noise signal, and \mathbf{A} and \mathbf{B} are, in this approximation, fixed $n \times n$ matrices. See figure 2 for a schematic.

According to the Data Rate Theorem (Nair et al. 2007), for an inherently unstable system, the control information represented by the vector u_t must be provided at a rate \mathcal{H} that is greater than the rate at which the system produces what is called topological information. For the system of Eq.(1) and figure 2, that rate is

$$\mathcal{H} > \log[|\det(\mathbf{A}^u)|] \equiv \alpha_0 \quad (2)$$

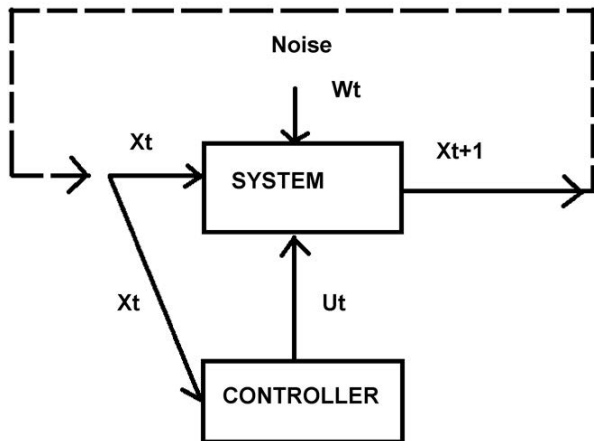


FIG. 2: A control system near some nonequilibrium steady state. x_t is system output at time t , u_t the control signal, and W_t an added noise term.

where \det is the determinant and \mathbf{A}^u is the component submatrix of \mathbf{A} that has eigenvalues ≥ 1 .

An alternate derivation of Eq.(2) is given in the Mathematical Appendix in terms of a formal Rate Distortion argument.

The most direct generalization of Eq.(2) in the context of a scalar integrated environmental insult ρ – a fog-of-war index, for example, taken as the magnitude of the largest vector of an empirical principal component analysis – is

$$\mathcal{H}(\rho) > f(\rho)\alpha_0 \quad (3)$$

$f(0)\alpha_0$ is then the rate at which the system generates topological information in the absence of an integrated environmental exposure.

What are $\mathcal{H}(\rho)$ and $f(\rho)$? The Mathematical Appendix calculates $\mathcal{H}(\rho)$ in terms of the ‘cost’ of control information, given the ‘investment’ ρ , using the exactly-solvable Black-Scholes approximation (Black and Scholes 1973):

$$\mathcal{H}(\rho) \approx \kappa_1\rho + \kappa_2 \quad (4)$$

where the κ_i are positive or zero.

Taking the same level of approximation for $f(\rho)$ as in Eq.(3), so that $f(\rho) \approx \kappa_3\rho + \kappa_4$, the stability condition becomes

$$\mathcal{T} \equiv \frac{\kappa_1\rho + \kappa_2}{\kappa_3\rho + \kappa_4} > \alpha_0 \quad (5)$$

If ρ is small, the necessary condition for stability is $\kappa_2/\kappa_4 > \alpha_0$. At high ρ the necessary condition for stability is $\kappa_1/\kappa_3 > \alpha_0$. If $\kappa_2/\kappa_4 \gg \kappa_1/\kappa_3$, then at some intermediate value of ρ the stability inequality may be violated, leading to failure of developmental regulation, as in figure 3.

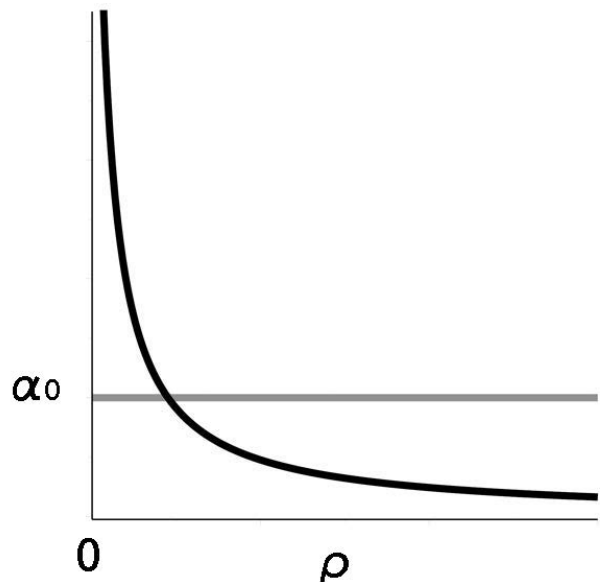


FIG. 3: The horizontal line is the limit α_0 . If $\kappa_2/\kappa_4 \gg \kappa_1/\kappa_3$, at an intermediate value of the environmental insult index ρ , $\mathcal{T} = (\kappa_1\rho + \kappa_2)/(\kappa_3\rho + \kappa_4)$ falls below criticality, and control of the developmental process fails catastrophically.

Developmental trajectories are embedded not only in complex, interacting environments of direct exposure, but of multimodal inheritance, in a large sense, through various synergistic mechanisms: it is implicit that environmental and historical factors must interact along the developmental trajectory of the system of interest. As a consequence, rather than a simple scalar ρ we are confronted by an $m \times m$ matrix having elements $\rho_{i,j}$, $i, j = 1 \dots m$ that is not likely to be symmetric.

However, square matrices of order m have m scalar invariants. That is, they have m real numbers that characterize the matrix regardless of how it is expressed in different coordinate systems. The first invariant is the trace, and the last, typically, \pm the determinant. These invariants are the coefficients of the characteristic polynomial $\mathcal{P}(\lambda)$:

$$\mathcal{P}(\lambda) = \det(\rho - \lambda I) = \lambda^m + r_1\lambda^{m-1} + \dots + r_{m-1}\lambda + r_m \quad (6)$$

where the parameter λ is an element of a ring, \det is the determinant, and I the $m \times m$ identity matrix. λ can be taken as the matrix ρ itself, since square matrices form a ring, producing the classic matrix polynomial expression $\mathcal{P}(\rho) = 0 \times I$.

For a $m \times m$ matrix there are m invariants r_1, \dots, r_m from Eq.(6), and an appropriate scalar exposure index can be constructed as a monotonic increasing function of the r_i :

$$\hat{\rho} = \hat{\rho}(r_1, \dots, r_m) \quad (7)$$

This can be used to determine \mathcal{T} as

$$\mathcal{T}(\hat{\rho}) = \frac{\kappa_1 \hat{\rho} + \kappa_2}{\kappa_3 \hat{\rho} + \kappa_4} \quad (8)$$

This construction invokes the rate distortion manifold of Glazebrook and Wallace (2009), formulated as a ‘generalized retina’ in Wallace and Wallace (2010). It serves to project a complicated information manifold onto a lower dimensional tangent space-analog tuned across that manifold so as to preserve much of the underlying information. Here, the model assumes a scalar tangent space. Higher dimensional tangent spaces can, of course, be constructed from the matrix invariants in the standard manner of differential geometry, but at the cost of a considerable increase in mathematical detail.

In a sense, \mathcal{T} has become the ‘temperature’ of the complicated fog-of-war interaction matrix $\rho_{i,j}$.

What happens to $\mathcal{T}(\hat{\rho})$ under stochastic circumstances? We can explore something of this by examining how a control signal u_t in figure 2 is expressed in the system response x_{t+1} . To be explicit, suppose it possible to deterministically retranslate a sequence of system outputs $X^i = x_1^i, x_2^i, \dots$ into a sequence of possible control signals $\hat{U}^i = \hat{u}_0^i, \hat{u}_1^i, \dots$ and then to compare that sequence with the original control sequence $U^i = u_0^i, u_1^i, \dots$. The difference between them is a real number measured by some appropriate distortion measure, allowing calculation of an average distortion as

$$\langle d \rangle = \sum_i p(U^i) d(U^i, \hat{U}^i) \quad (9)$$

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

The next step is to apply a Rate Distortion Theorem (RDT) argument. According to the RDT, there is a Rate Distortion Function (RDF) that determines the minimum channel capacity, $R(D)$, necessary to keep the average distortion $\langle d \rangle$ below some fixed limit D (Cover and Thomas 2006). Again, applying Feynman’s (1996) interpretation of information as a form of free energy, it is possible to construct a Boltzmann-like pseudoprobability in the ‘temperature’ \mathcal{T} :

$$dP(R, \mathcal{T}) = \frac{\exp[-R/\mathcal{T}]dR}{\int_0^\infty \exp[-R/\mathcal{T}]dR} \quad (10)$$

since it is clear that higher \mathcal{T} necessarily implies greater channel capacity.

The denominator integral is similar to a statistical mechanical partition function, allowing definition of a ‘free energy’ Morse Function F analog (Pettini 2007) as

$$\exp[-F/\mathcal{T}] = \int_0^\infty \exp[-R/\mathcal{T}]dR = \mathcal{T} \quad (11)$$

Thus $F(\mathcal{T}) = -\mathcal{T} \log[\mathcal{T}]$.

See the Mathematical Appendix for a brief outline of Morse Theory.

An entropy-analog can now be defined in terms of the Legendre transform of F as

$$\mathcal{S} \equiv F(\mathcal{T}) - \mathcal{T}dF/d\mathcal{T} = \mathcal{T} \quad (12)$$

At first approximation, it is possible to use Onsager’s treatment of nonequilibrium thermodynamics (de Groot and Mazur 1984). The system dynamics are then driven by the gradient of \mathcal{S} in the essential parameter \mathcal{T} under conditions of noise, giving a stochastic differential equation (SDE) (Protter 1990)

$$d\mathcal{T}_t \approx (\mu d\mathcal{S}/d\mathcal{T})dt + \beta \mathcal{T}_t dW_t = \mu dt + \beta \mathcal{T}_t dW_t \quad (13)$$

μ is a kind of ‘diffusion coefficient’ representing the efforts of the underlying control mechanism, and β is the magnitude of an inherent impinging white noise dW_t in the context of volatility, so that the effect of noise is proportional to the signal.

Applying the Ito chain rule (Protter 1990) to $\log(\mathcal{T})$ in Eq.(13) produces the stochastic differential equation

$$d \log[\mathcal{T}]_t = (\mu - \frac{1}{2}\sigma^2)dt + \sigma dW_t \quad (14)$$

Applying Jensen’s inequality for a concave function, the nonequilibrium steady state (nss) expectation for \mathcal{T} satisfies the limit condition

$$E(\mathcal{T}_t) \geq E(\log[|\mathcal{T}_t|]) = \frac{\mu}{\beta^2/2} \quad (15)$$

Recall that, in this model, μ indexes attempts by the embedding control apparatus to impose stability, to raise \mathcal{T} . Thus impinging noise can significantly increase the probability that \mathcal{T} falls below the critical limit of figure 3, triggering a failure in developmental control.

However, $E(\mathcal{T})$ is an expectation, so that there is always a nonzero probability that \mathcal{T} will fall below the critical value α_0 in the multimodal expression for $\mathcal{T}(\hat{\rho})$. This implies that sporadic control dysfunctions cannot be eliminated. While raising μ and lowering β decreases their probability, such interventions will not drive them to zero in this model, a matter of some importance for population-rates of developmental failure.

More specifically, invoking the Stochastic Stabilization Theorem (Mao 2007),

$$\lim_{t \rightarrow \infty} \frac{\log[|\mathcal{T}_t|]}{t} \rightarrow < 0 \quad (16)$$

almost surely unless

$$\mu > \frac{1}{2}\sigma^2 \quad (17)$$

IV. COGNITIVE THEORY

A second approach to the dynamics of developmental regulation and its failure is through the cognitive paradigm of Atlan and Cohen (1998). Cognition, in their view, requires active choice of a particular response to an incoming signal, here an insult, from a larger repertoire of those possible. Recall that choice reduces uncertainty and directly implies the existence of an underlying information source (Wallace 2012).

Given an information source associated with an inherently unstable, rapidly-acting developmental control system – ‘dual’ to it – an equivalence class algebra can be constructed by choosing different system origin states a_0 and defining the equivalence of two subsequent states at times $m, n > 0$, written as a_m, a_n , by the existence of high-probability meaningful paths connecting them to the same origin point. Disjoint partition by equivalence class, analogous to orbit equivalence classes in dynamical systems, defines a symmetry groupoid associated with the cognitive process. As described, groupoids are deep generalizations of the group concept in which there is not necessarily a product defined for each possible element pair (Weinstein 1996).

The equivalence classes define a set of cognitive dual information sources available to the inherently unstable development regulation system, creating a large groupoid, with each orbit corresponding to a transitive groupoid whose disjoint union is the full groupoid. Each subgroupoid is associated with its own dual information source, and larger groupoids will have richer dual information sources than smaller.

Let X_{G_i} be the control system’s dual information source associated with the groupoid element G_i , and let Y be the information source associated with embedding ‘normal’ environmental variation that impinges on development. Again, Wallace (2012) details how environmental regularities imply the existence of an environmental information source that, for humans, particularly includes cultural and socioeconomic factors (e.g., Wallace 2015).

We again build a ‘free energy’ Morse Function. Let $H(X_{G_i}, Y) \equiv H_{G_i}$ be the joint uncertainty of the two information sources. A Boltzmann-like pseudoprobability is written as

$$P[H_{G_i}] = \frac{\exp[-H_{G_i}/\mathcal{T}]}{\sum_j \exp[-H_{G_j}/\mathcal{T}]} \quad (18)$$

Again, \mathcal{T} is the ‘temperature’ from Eq.(8), via the $\hat{\rho}$ of Eq.(7), and the sum is over the different possible cognitive modes of the full system, characterized by the groupoid inherent to the developmental trajectory of figure 1.

A new Morse Function \mathcal{F} is defined by

$$\exp[-\mathcal{F}/\mathcal{T}] \equiv \sum_j \exp[-H_{G_j}/\mathcal{T}] \quad (19)$$

Given the inherent developmental groupoid as a generalization of the simpler symmetry group, again it is possible to invoke an extension of Landau’s picture of phase transition (Pettini 2007). To reiterate, in Landau’s ‘spontaneous symmetry breaking’, phase transitions driven by temperature changes occur as alteration of system symmetry, with higher energies at higher temperatures being more symmetric.

For this model, the shift between symmetries is highly punctuated in \mathcal{T} under the Data Rate Theorem for unstable control systems. Typically, there are only a very limited number of possible phases, which may or may not coexist under particular circumstances. For example, Wallace (2016b) argues that comorbid autism spectrum and schizophreniform disorders represent the simultaneous presence of two such ‘condensations’ affecting the recruitment of lower-level neural modules into the ongoing stream of high-order real-time cognitive function.

More generally, then, reduction in \mathcal{T} can lead to punctuated decline in the complexity of cognitive process possible within the developmental control system, driving it into a ‘ground state’ (or states) collapse in which essential systems fail to develop normally.

Again, the essential feature is the integrated environmental insult $\hat{\rho}$ constructed as a scalar projection of a complicated matrix of interacting ‘toxic’ modes. Most of the topology of the inherently unstable developmental system has been factored out so that $\hat{\rho}(r_1, \dots, r_m)$ remains the only possible index of the rate of topological information generation for the DRT. Thus, in Eqs.(18) and (19), $\mathcal{T}(\hat{\rho})$ is again the driving parameter. Again, more complicated, higher dimensional ‘tangent spaces’ can be constructed from the r_i at the cost of the standard mathematical overhead associated with differential geometry.

Increasing $\hat{\rho}$ is equivalent to lowering the ‘temperature’ \mathcal{T} , and the system passes from high symmetry ‘free flow’ to different forms of ‘crystalline’ structure – broken symmetries representing the punctuated onset of significant developmental failure characteristic of different systems.

Sufficient conditions for the intractability – stability – of pathological low \mathcal{T} ‘condensates’ can be explored using the methods of Wallace (2016c).

V. DISCUSSION AND CONCLUSIONS

In a sense, the underlying argument is by abduction from recent advances in evolutionary theory: West-Eberhard (2003, 2005) sees development as a key, but often poorly appreciated, element of evolutionary process, in that a new input, whether it comes from a genome, like a mutation or from the external environment, like a temperature change, a pathogen, or a parental opinion, has a developmental effect only if the preexisting phenotype can respond. A novel input causes a reorganization of the phenotype, a ‘developmental recombination’ in which phenotypic traits are expressed in new or distinctive combinations during ontogeny, or undergo corre-

lated quantitative changes in dimensions. Developmental recombination can result in evolutionary divergence at all levels of organization.

Most importantly, perhaps, West-Eberhard characterizes individual development as a series of branching pathways. Each branch point is a developmental decision, a switch point, governed by some regulatory apparatus, and each switch point defines a modular trait. Developmental recombination implies the origin or deletion of a branch and a new or lost modular trait. The novel regulatory response and the novel trait originate simultaneously, and their origins are inseparable events: there cannot be a change in the phenotype without an altered developmental pathway.

Thus, there are strong arguments for the great evolutionary potential of environmentally induced novelties. An environmental factor can affect numerous individuals, whereas a mutation initially can affect only one, a perspective having implications, not only for evolutionary economics, but across a full spectrum of ubiquitous ‘developmental’ phenomena: even traffic streams ‘evolve’ under changing selection pressures, and, indeed, such pressures act at every level of biological, social, or economic organization, as well as across rapidly expanding realms of machine cognition.

That is, just as the Atlan/Cohen ‘cognitive paradigm’ for the immune system generalizes across many different systems (Wallace 2012), so too does the West-Eberhard model of development: repeated branching under the control of an embedding regulatory apparatus responding to environmental cues is widely observed. Here, we have applied a control theory formalism via the Data Rate Theorem, and using information theory, invoked the dual information source necessarily associated with regulatory cognition. The intent has been to examine developmental disorders, in a large sense, over a spectrum that ranges from cellular to socioeconomic and emerging machine levels of organization, and across time scales from those of biological evolution to extremely rapid machine response.

The main focus here has been on exploring the influence of environmental insult on developmental dysfunction, where insult itself has been measured by a projected scalar ‘tangent space’ defined in terms of the invariants of the unsymmetric fog-of-war matrix $\rho_{i,j}$. The synergism between control and information theories via the Data Rate Theorem, and the extensions using topological and ‘free energy’ Morse Theory methods, provide a new theoretical window into the dynamics of many developmental processes, via the construction of statistical models that, like more familiar regression procedures, can be applied to a broad range of experimental and observational data.

VI. MATHEMATICAL APPENDIX

A. Groupoids

A groupoid, G , is defined by a base set A upon which some mapping – a morphism – can be defined (Weinstein 1996). Importantly, not all possible pairs of states (a_j, a_k) in the base set A can be connected by that morphism. Those that can define the groupoid element, a morphism $g = (a_j, a_k)$ having the natural inverse $g^{-1} = (a_k, a_j)$. Given such a pairing, there are natural end-point maps $\alpha(g) = a_j, \beta(g) = a_k$ from the set of morphisms G into A , and a formally associative product in the groupoid $g_1 g_2$ provided $\alpha(g_1 g_2) = \alpha(g_1), \beta(g_1 g_2) = \beta(g_2)$, and $\beta(g_1) = \alpha(g_2)$. Then the product is defined, and associative, $(g_1 g_2) g_3 = g_1 (g_2 g_3)$.

Further, there are natural left and right identity elements λ_g, ρ_g such that $\lambda_g g = g = g \rho_g$.

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

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

If G is any groupoid over A , the map $(\alpha, \beta) : G \rightarrow A \times A$ 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 $f : X \rightarrow Y$ 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 (Weinstein 1996). 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 sometimes criticism of 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. It is possible to introduce a metric structure on manifolds of related information sources, producing such interaction.

Fundamentally, a groupoid is a category in which all morphisms have an inverse.

As Weinstein (1996) emphasizes, 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 also parameterizes the different ways (isomorphisms) in which two elements can be equivalent, e.g., all possible information sources

dual to some cognitive process.

Brown (1987) describes the fundamental 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 λ , $\lambda \in \Lambda$.

EXAMPLE 2. An equivalence relation R on [a set] X becomes a groupoid with $\alpha, \beta : R \rightarrow X$ 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 central 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.

This approach underlies the ‘topological’ perspective of directed homotopy developmental process.

Groupoids have become quite popular in the study of networks of coupled dynamical systems which can be defined by differential equation models, (e.g., Golubitsky and Stewart 2006).

B. An RDT proof of the DRT

The Rate Distortion Theorem of information theory asks how much a signal can be compressed and have average distortion, according to an appropriate measure, less than some predetermined limit $D > 0$. The result is an expression for the minimum necessary channel capacity, R , as a function of D . See Cover and Thomas (2006) for details. Different channels have different expressions. For the Gaussian channel under the squared distortion measure,

$$\begin{aligned} R(D) &= \frac{1}{2} \log\left[\frac{\sigma^2}{D}\right] \quad D < \sigma^2 \\ R(D) &= 0 \quad D \geq \sigma^2 \end{aligned} \quad (20)$$

where σ^2 is the variance of channel noise having zero mean.

Our concern is how a control signal u_t is expressed in the system response x_{t+1} . We suppose it possible to deterministically retranslate an observed sequence of system outputs x_1, x_2, x_3, \dots into a sequence of possible control signals $\hat{u}_0, \hat{u}_1, \dots$ and to compare that sequence with the original control sequence u_0, u_1, \dots , with the difference between them having a particular value under the chosen distortion measure, and hence an observed average distortion.

The correspondence expansion is as follows.

Feynman (1996), following ideas of Bennett, identifies information as a form of free energy. Thus $R(D)$, the minimum channel capacity necessary for average distortion D , is also a free energy measure, and we may define an entropy S as

$$S \equiv R(D) - DdR/dD \quad (21)$$

For a Gaussian channel under the squared distortion measure,

$$S = 1/2 \log[\sigma^2/D] + 1/2 \quad (22)$$

Other channels will have different expressions.

The simplest dynamics of such a system are given by a nonequilibrium Onsager equation in the gradient of S , (de Groot and Mazur 1984) so that

$$dD/dt = -\mu dS/dD = -\frac{\mu}{2D} \quad (23)$$

By inspection,

$$D(t) = \sqrt{\mu t} \quad (24)$$

which is the classic outcome of the diffusion equation. For the ‘natural’ channel having $R(D) \propto 1/D$, $D(t) \propto$ the cube root of t .

This correspondence reduction allows an expansion to more complicated systems, in particular, to the control system of figure 2.

Let \mathcal{H} be the rate at which control information is fed into an inherently unstable control system, in the presence of a further source of control system noise β , in addition to the channel noise defined by σ^2 . The simplest generalization of Eq.(23), for a Gaussian channel, is the stochastic differential equation

$$dD_t = \left[\frac{\mu}{2D_t} - M(\mathcal{H}) \right] dt + \beta D_t dW_t \quad (25)$$

where dW_t represents white noise and $M(\mathcal{H}) \geq 0$ is a monotonically increasing function.

This equation has the nonequilibrium steady state expectation

$$D_{nss} = \frac{\mu}{2M(\mathcal{H})} \quad (26)$$

measuring the average distortion between what the control system wants and what it gets. In a sense, this is

a kind of converse to the famous radar equation which states that a returned signal will be proportional to the inverse fourth power of the distance between the transmitter and the target. But there is a deeper result, leading to the DRT.

Applying the Ito chain rule to Eq.(25) (Protter 1990), it is possible to calculate the expected variance in the distortion as $E(D_t^2) - (E(D_t))^2$. But application of the Ito rule to D_t^2 shows that *no real number solution for its expectation is possible unless the discriminant of the resulting quadratic equation is ≥ 0* , so that a necessary condition for stability is

$$\begin{aligned} M(\mathcal{H}) &\geq \beta\sqrt{\mu} \\ \mathcal{H} &\geq M^{-1}(\beta\sqrt{\mu}) \end{aligned} \quad (27)$$

where the second expression follows from the monotonicity of M .

As a consequence of the correspondence reduction leading to Eq.(25), we have generalized the DRT of Eq.(2). Different ‘control channels’, with different forms of $R(D)$, will give different detailed expressions for the rate of generation of ‘topological information’ by an inherently unstable system.

C. A Black-Scholes model

We look at $\mathcal{H}(\rho)$ as the control information rate ‘cost’ of stability at the integrated environmental insult ρ . To determine the mathematical form of $\mathcal{H}(\rho)$ under conditions of volatility i.e., variability proportional to a signal, we must first model the variability of ρ , most simply taken as

$$d\rho_t = g(t, \rho_t)dt + b\rho_t dW_t \quad (28)$$

Here, dW_t is white noise and – counterintuitively – the function $g(t, \rho)$ will fall out of the calculation on the assumption of certain regularities.

$\mathcal{H}(\rho_t, t)$ is the minimum needed incoming rate of control information under the Data Rate Theorem. Expand \mathcal{H} in ρ using the Ito chain rule (Protter 1990):

$$\begin{aligned} d\mathcal{H}_t &= [\partial\mathcal{H}/\partial t + g(\rho_t, t)\partial\mathcal{H}/\partial\rho + \frac{1}{2}b^2\rho_t^2\partial^2\mathcal{H}/\partial\rho^2]dt \\ &\quad + [b\rho_t\partial\mathcal{H}/\partial\rho]dW_t \end{aligned} \quad (29)$$

It is now possible to define a Legendre transform, L , of the rate \mathcal{H} , by convention having the form

$$L = -\mathcal{H} + \rho\partial\mathcal{H}/\partial\rho \quad (30)$$

\mathcal{H} is an information index, a free energy measure in the sense of Feynman (1996), so that L is a classic entropy measure.

We make an approximation, replacing dX with ΔX and applying Eq.(29), so that

$$\Delta L = (-\partial\mathcal{H}/\partial t - \frac{1}{2}b^2\rho^2\partial^2\mathcal{H}/\partial\rho^2)\Delta t \quad (31)$$

According to the classical Black-Scholes model (Black and Scholes 1973), the terms in g and dW_t ‘cancel out’, and white noise has been subsumed into the Ito correction factor, a regularity assumption making this an exactly solvable but highly approximate model.

The conventional Black-Scholes calculation takes $\Delta L/\Delta t \propto L$. At nonequilibrium steady state, by some contrast, we can assume $\Delta L/\Delta t = \partial\mathcal{H}/\partial t = 0$, giving

$$-\frac{1}{2}b^2\rho^2\partial^2\mathcal{H}/\partial\rho^2 = 0 \quad (32)$$

so that

$$\mathcal{H} = \kappa_1\rho + \kappa_2 \quad (33)$$

The κ_i will be nonnegative constants.

D. Morse Theory

Morse theory studies how analytic behavior of a function – the location and character of its critical points – relates to the underlying topology of the manifold on which the function is defined. Here we roughly follow Pettini (2007). The basic approach is to examine an n -dimensional manifold M as decomposed into level sets of some function $f : M \rightarrow \mathbf{R}$ where \mathbf{R} is the set of real numbers. The a -level set of f is defined as $f^{-1}(a) = \{x \in M : f(x) = a\}$, the set of all points in M with $f(x) = a$. If M is compact, then the whole manifold can be decomposed into such slices in a canonical fashion between two limits, defined by the minimum and maximum of f on M . Let the part of M below a be defined as $M_a = f^{-1}(-\infty, a] = \{x \in M : f(x) \leq a\}$. These sets describe the whole manifold as a varies between the minimum and maximum of f .

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

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

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

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

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

The essential results of Morse theory are (Pettini 2007):

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

$v \in (a, b]$. The result is valid even if f is not a Morse function, but only a smooth function.

2. If the interval $[a, b]$ contains critical values, the topology of $f^{-1}[a, v]$ changes in a manner determined by the properties of the matrix H at the critical points.

3. If $f : M \rightarrow \mathbf{R}$ 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 $f : M \rightarrow \mathbf{R}$ is a Morse function, with M compact, then on a finite interval $[a, b] \subset \mathbf{R}$, there is only a finite number of critical points p of f such that $f(p) \in [a, b]$. The set of critical values of f is a discrete set of \mathbf{R} .

5. For any differentiable manifold M , the set of Morse functions on M is an open dense set in the set of real functions of M of differentiability class r for $0 \leq r \leq \infty$.

6. Some topological invariants of M , that is, quantities that are the same for all the manifolds that have the same topology as M , can be estimated and some-

times computed exactly once all the critical points of f are known: Let the Morse numbers $\mu_i (i = 0, \dots, m)$ 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 , $\chi = \sum_{i=0}^m (-1)^i \mu_i$.

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.

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

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

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

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

de Groot, S., R. Mazur, 1984, *Non-Equilibrium Thermodynamics*, Dover, New York.

Fajstrup, L., E. Goubault, A. Mourgues, S. Mimram, M. Raussen, 2016, *Directed Algebraic Topology and Concurrency*, Springer, New York.

Feynman, R., 1996, *Feynman Lectures on Computation*, Addison-Wesley, Reading, MA.

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

Grandis, M., 2009, *Directed Algebraic Topology: Models of Non-Reversible Worlds*, Cambridge University Press, New York.

Hatcher, A., 2001, *Algebraic Topology*, Cambridge University Press, New York.

Kerner, B., S. Klenov, 2009, Phase transitions in traffic flow on multilane roads, *Physics Reviews E*, 80:056101.

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

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

Nair, G. et al., 2007, Feedback control under data rate constraints: an overview, *Proceedings of the IEEE*, 95:108-137.

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

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

Rosen, R., 2012, *Anticipatory Systems: Philosophical, Mathematical, and Methodological Foundations*, Second Edition, Springer, New York.

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

Wallace, R., 2015, *An Ecosystem Approach to Economic Stabilization: Escaping the neoliberal wilderness*, Routledge Advances in Heterodox Economics, New York.

Wallace, R., 2016a, *Information Theory Models of Instabilities in Critical Systems*, Vol. 7 of the World Scientific Series in Information Studies, Singapore.

Wallace, R., 2016b, Environmental induction of neurodevelopmental disorders, *Bulletin of Mathematical Biology*, doi 10.1007/s11538-016-0226-5.

Wallace, R., 2016c, Subtle noise structures as control signals in high-order biocognition, *Physics Letters A*, 380:726-729.

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

West-Eberhard, M., 2003, *Developmental Plasticity and Evolution*, Oxford University Press, New York.

West-Eberhard, M., 2005, Developmental plasticity and the origin of species differences, *PNAS*, 102:6543-6549.