

Version 5.1

# Evolution as a language that speaks itself

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

November 29, 2010

## Abstract

The ‘self-referential’ character of evolutionary process noted by Goldenfeld and Woese (2010) can be restated through a ‘language’ model in which genes, gene expression, and environment are represented as interacting information sources. The larger, composite, source that characterizes the high probability evolutionary paths then becomes, in a real sense, a language that speaks itself. The approach represents a significant extension of nonequilibrium condensed matter formalism in which the asymptotic limit theorems of information theory beat back the mathematical thicket a full step, providing necessary but not sufficient conditions for punctuated evolutionary transitions that can themselves be expressed as highly structured large deviations having their own grammar and syntax.

**Key Words:** evolution, groupoid, information theory, large deviations, Morse Function, punctuation, renormalization

## 1 Introduction

Wallace (2010a) has proposed expanding the modern synthesis by introducing ‘The principle of environmental interaction,’ i.e., that individuals and groups engage in powerful, often punctuated, dynamic mutual relations with their embedding environments that may include the exchange of heritage material between markedly different organisms. Escaping the intellectual straightjacket of mathematical population genetics – or at least exchanging it for a slightly larger one – that work characterizes genes, gene expression (as a cognitive process) and embedding environments as interacting information sources constrained by the asymptotic limit theorems of information theory. This leads to an inherently coevolutionary system described in terms of a formalism quite similar to that of Onsager’s nonequilibrium thermodynamics and having quasi-stable ‘coevolutionary’ states coupled by highly structured large deviations, all much in the sense of Champagnat et al. (2006). Wallace (2010b) restricted the general theory in terms of gene expression and gene transmission, producing

close analogs to the niche construction approach of Odling-Smee et al. (2003).

This body of work introduces powerful methods from the statistical physics of phase transitions into evolutionary theory, much in the spirit of the recent paper by Goldenfeld and Woese (2010), who focus on evolution ‘as a problem in nonequilibrium statistical mechanics, where the key dynamical modes are collective’. They provide a central insight:

...[T]he genome encodes the information which governs the response of an organism to its physics and biological environment. At the same time, this environment actually shapes genomes through gene transfer processes and phenotype selection. Thus, we encounter a situation where the dynamics must be self-referential: the update rules change during the time evolution of the system, and the way in which they change is a function of the state and thus the history of the system... self-referential dynamics is an inherent and probably defining feature of evolutionary dynamics and thus biological systems.

Here we explore such self-referential dynamics from the perspectives of Wallace (2010a, b), recognizing that the representation of fundamental biological processes in terms of information sources restrains, somewhat, the inherent nonequilibrium nature of those processes. That is, although the operation of information sources is both nonequilibrium and irreversible in the most fundamental sense (e.g., few and short palindromes), the asymptotic limit theorems of information theory beat back the mathematical thicket surrounding such phenomena one full step. The theorems permit, in some measure, an ‘equilibrium-like’ approximation to inherently nonequilibrium processes under proper circumstances, and, as in Wallace (2010a), allow the stochastic differential equation models inherent to nonequilibrium statistical mechanics to penetrate a full step deeper.

## 2 Basic formalism

Evolutionary process, as envisioned by Wallace (2010a), involves dynamic interplay between (at least) three information sources representing gene transmission, gene expression,

\*Contact: wallace@pi.cpmc.columbia.edu. Box 47, 1051 Riverside Drive, New York, New York, 10032.

and embedding environment, given that both genes and environment ‘remember’, producing serial correlations in time. We suppose it possible to coarse-grain observational measures of those three processes, representing the results in terms of some ‘alphabet’ of possible states. Our interest is in (properly characterized, and possibly very long) temporal paths beginning at some initial state  $a_0$ , and having the form

$$x_n \equiv \{a_0, a_1, \dots, a_n\},$$

where the  $a_j$  are possible elements of the coarsegrained alphabet.

Given a particular tripartite starting point,  $a_0$ , evolution, being inherently path dependent, must build on what has gone before. Thus, crudely, subsequent paths can be divided into two classes, a vast set having vanishingly small probability, and a much smaller set that, we suppose, follows something like the regularities of information theory that govern the three component information sources. That is, if  $N(n)$  is the number of high probability paths of length  $n$ , then there exists a *path independent* limit  $H$  such that

$$H = \lim_{n \rightarrow \infty} \frac{\log[N(n)]}{n}.$$

(1)

Below we will indicate how the restriction of strict path independence might be lifted, somewhat.

We assume that, associated with each path  $x_n$  of length  $n$ , there is an information source  $X_n$  producing it that is defined in terms of the joint and conditional probabilities

$$P(a_0, a_1, \dots, a_n)$$

and

$$P(a_n | a_{n-1}, \dots, a_1, a_0),$$

such that appropriate Shannon uncertainties may be defined (e.g., Ash, 1990; Khinchin, 1957; Cover and Thomas, 2006), and that the Shannon-McMillan Theorem holds:

$$H = \lim_{n \rightarrow \infty} \frac{\log[N(n)]}{n} = \lim_{n \rightarrow \infty} H(X_n | X_{n-1}, \dots, X_0) = \lim_{n \rightarrow \infty} \frac{H(X_0, X_1, \dots, X_n)}{n+1}.$$

(2)

We now shift perspective, defining equivalence classes of paths, and an associated symmetry groupoid (simplest example, a disjoint union of groups) that will be needed for the characterization of collective phenomena, much in the sense that a symmetry group is needed for Landau’s theory of phase transition.

We call two states  $a_j$  and  $a_k$  equivalent if there is a high probability path beginning with  $a_0$  that connects them. The set of high probability paths beginning at  $a_0$  defines the possible evolutionary processes that start at that state, and the set of equivalence classes defines a groupoid in a standard manner that characterizes the information source  $H$  associated with them. See the Mathematical Appendix for a summary explanation of groupoids.

We can now index the set of possible evolutionary information sources by the groupoids defining the equivalence classes of high probability paths associated with them.

Next, allow the initial state to vary, that is, allow different starting points,  $a_0$ , across the system. This produces an even larger groupoid that will enable our analysis of certain collective phenomena.

### 3 Punctuated equilibrium: metabolic phase transitions in evolution

As Feynman (2000) argues, based on work by Bennett (1988), information is simply another form of free energy, and the information in a message is quite precisely measured by the free energy needed to erase it. Indeed, Feynman (2000) shows how to construct an (idealized) machine that directly converts the information in a message to work.

But there are subtleties. First, information sources are already inherently irreversible dynamic systems. For example, in spoken or written English, the short sequence ‘ the ’ has much higher probability than its time reversed ‘ eht ’. There is no local reversibility, and adaptation of methods from nonequilibrium statistical mechanics or thermodynamics will not be graced with ‘Onsager reciprocal relations’.

Another subtlety is that, in spite of the inherently nonequilibrium dynamic nature of an information source, the asymptotic limit theorems defining information source uncertainty appear to permit ‘nonequilibrium equilibria’ in a certain sense.

We suppose there to be some monotonic increasing measure of available metabolic free energy,  $Q(M)$ ,  $Q(0) = 0$ . We assume that possible evolutionary trajectories are constrained by the availability of metabolic resources, so that the probability of an (inherently irreversible and highly dynamic) information source associated with groupoid element  $G_j$ , at a fixed  $Q(M)$ , is given, in a first approximation, by the standard expression for the Gibbs distribution

$$P[H_{G_j}] = \frac{\exp[-H_{G_j}/Q]}{\sum_i \exp[-H_{G_i}/Q]}.$$

(3)

As Goldenfeld (2010) has pointed out, this appears to be not really appropriate for a system evolving in an open manner, and we will generalize the treatment somewhat, using an adiabatic approximation in which the dynamics remain ‘close enough’ to a form in which Morse Theory can work, adapting standard phase transition formalism for transitions between adiabatic realms.

In essence, however, by adopting an information source perspective on evolutionary process we implicitly incorporate the possibility of ‘nonequilibrium equilibria’ in the sense of Eldredge and Gould (1972).

The ‘E-property’ that Khinchin (1957) identifies – the division of paths into high and low probability sets – and the limiting relation

$$\lim_{n \rightarrow \infty} \frac{\log[N(n)]}{n} = H$$

for all high probability paths generated by an ergodic information source, permits imposition of a powerful regularity onto inherently nonequilibrium evolutionary processes.

The partition function-analog of this strange system can, as usual, be defined as

$$Z_G(Q) = \sum_i \exp[-H_{G_i}/Q].$$

(4)

We can now define an evolutionary groupoid source uncertainty,  $F_G$ , constructed over the full set of possible evolutionary trajectories as constrained by metabolic energy, as

$$\exp[-F_G/Q] \equiv \sum_i \exp[-H_{G_i}/Q],$$

(5)

so that

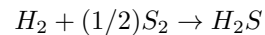
$$F_G(Q) = -Q \log[Z_G(Q)].$$

This is to be taken as a Morse Function, in the sense of the Mathematical Appendix.

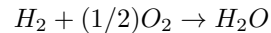
Argument by abduction from statistical physics (Landau and Lifshitz, 2007; Pettini, 2007) identifies this as the groupoid free energy of the evolutionary process, here constrained by metabolic energy availability, to which we now apply Landau’s theory of punctuated phase transition in terms of groupoid, rather than group, symmetries.

Recall, now, Landau’s perspective on phase transition (Pettini, 2007). The essence of his insight was that certain physical phase transitions took place in the context of a significant symmetry change, with one phase being more symmetric than the other. A symmetry is lost in the transition, i.e., spontaneous symmetry breaking. The greatest possible set of symmetries being that of the Hamiltonian describing the energy states. Usually, states accessible at lower temperatures will lack the symmetries available at higher temperatures, so that the lower temperature state is less symmetric, and transitions can be highly punctuated.

Here, we have characterized the dependence of evolutionary process on the availability of metabolic free energy in terms of groupoid, rather than group, symmetries, and the argument by abduction is essentially similar: Increasing availability of metabolic free energy – rising  $Q(M)$  – will allow richer interactions between gene expression, genes, and environment, and will do so in a highly punctuated manner, as in Eldredge and Gould (1972). Wallace (2009) examined the eukaryotic transition from this perspective, using a cruder version of the theory, and found that the aerobic transition was a likely prerequisite to the coming-together of individual organisms into nucleated cells. A sulfur-based metabolism, e.g.,



generates about 21 KJ/mol, while



provides about 241 KJ/mol, more than an order of magnitude greater, a typical result. Canfield et al. (2006) show a range of possible electron donors and acceptors likely available to early anaerobic metabolisms, and discuss likely early ecosystems based on them. These would have been orders of magnitude less active than an  $O_2$ -based biosphere. A ten-fold increase in available metabolic free energy might well be expected to permit very high orders of (groupoid) symmetries, including the eukaryotic transition.

## 4 Extending the model

### 4.1 Kadanoff theory

Given  $F_G$  as a free energy analog, we are interested in a mathematical treatment of transitions between adiabatic realms and suppose it possible to define a characteristic ‘length’, say  $r$ , on the system, more fully described below. We can then define renormalization symmetries in terms of the ‘clumping’ transformation, so that, for clumps of size  $R$ , in an external ‘field’ of strength  $J$  (that we can set to 0 in the limit), one can write, in the usual manner (e.g., Wilson, 1971)

$$F_G[Q(R), J(R)] = f(R)F_G[Q(1), J(1)],$$

$$\chi(Q(R), J(R)) = \frac{\chi(Q(1), J(1))}{R},$$

(7)

where  $\chi$  is a characteristic correlation length.

As Wallace (2005) shows, following Wilson (1971), very many ‘biological’ renormalizations,  $f(R)$ , are possible that lead to a number of quite different universality classes for phase transition.

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

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

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

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

(8)

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

Clearly, other approaches to metric construction on  $U$  seem possible, and other approaches to renormalization than outlined by equation (7).

### 4.2 Nonergodic information sources

The ergodic nature of an information source is a generalization of the law of large numbers and implies that the long-time averages can be closely approximated by averages across the probability spaces of those sources. For non-ergodic information sources, a function,  $\mathcal{H}(x_n)$ , of each path  $x_n \rightarrow x$ , may be defined, such that  $\lim_{n \rightarrow \infty} \mathcal{H}(x_n) = \mathcal{H}(x)$ , but  $\mathcal{H}$  will not in general be given by the simple cross-sectional laws-of-large numbers analogs above (Khinchin, 1957).

Let  $s \equiv d(x, \hat{x})$  for high probability paths  $x$  and  $\hat{x}$ , where  $d$  is a distortion measure, as described in Cover and Thomas (2006). For ‘nearly’ ergodic systems one might use something of the form

$$\mathcal{H}(\hat{x}) \approx \mathcal{H}(x) + sd\mathcal{H}/ds|_{s=0}$$

for  $s$  sufficiently small. The idea is to take a distortion measure as a kind of Finsler metric, imposing a resulting ‘global’ structure over an appropriate class of non-ergodic information sources. One question obviously revolves around what properties are metric-independent, in much the same manner as the Rate Distortion Theorem.

These heuristics can be made more precise:

Take a set of ‘high probability’ paths  $x_n \rightarrow x$ .

Suppose, for all such  $x$ , there is an open set,  $U$ , containing  $x$ , on which the following conditions hold:

1. For all paths  $\hat{x}_n \rightarrow \hat{x} \in U$ , a distortion measure  $s_n \equiv d_U(x_n, \hat{x}_n)$  exists.
2. For each path  $x_n \rightarrow x$  in  $U$  there exists a pathwise invariant function  $\mathcal{H}(x_n) \rightarrow \mathcal{H}(x)$ , in the sense of Khinchin (1957, p.72). While such a function will almost always exist, only in the case of an ergodic information source can it be identified as an ‘entropy’ in the usual sense.
3. A function  $F_U(s_n, n) \equiv f_n \rightarrow f$  exists, for example,

$$f_n = s_n, \log[s_n]/n, s_n/n,$$

and so on.

4. The limit

$$\lim_{n \rightarrow \infty} \frac{\mathcal{H}(x_n) - \mathcal{H}(\hat{x}_n)}{f_n} \equiv \nabla_F \mathcal{H}|_x$$

exists and is finite.

Under such conditions, standard global atlas/manifold constructions are possible. Again,  $\mathcal{H}$  is not simply given by the usual expressions for source uncertainty if the source is not ergodic, and the phase transition development above may be correspondingly more complicated. Restriction to high probability paths simplifies matters considerably, although precisely characterizing them may be difficult, requiring extension of the Shannon-McMillan Theorem and its Rate Distortion generalization.

An essential question is under what circumstances this differential treatment for ‘almost’ ergodic information sources permits something very much like what Khinchin (1957, p. 54) calls the ‘E property’ enabling classification of paths into a small set of high probability and a vastly larger set of vanishingly small probability (Khinchin, 1957, p. 74).

### 4.3 Morse Theory

As Goldenfeld (2010) has pointed out, equation (3), the Gibbs distribution, seems not really appropriate for a system evolving in an open manner, although, as we have argued, the regularities imposed by the asymptotic limit theorems of information theory permit study of ‘nonequilibrium equilibria’ in a standard way. Here we attempt to significantly generalize that treatment.

Letting  $t$  represent time, for the groupoid information source ensemble we propose a probability density function of the form

$$P[H_{G_j}] = \frac{f_t(H_{G_j}/Q)}{\sum_i f_t(H_{G_i}/Q)},$$

(9)

with the provisos that the sum always converges and that the dependence on time is adiabatic, in the sense that the system changes slowly enough so that the information theory mathematics ‘almost’ works all the time. Between such realms the Kadanoff formalism is to be applied.

Then, as above, we take

$$Z(Q) \equiv \sum_i f_t(H_{G_i}/Q)$$

(10)

and define a free energy-analog by

$$f_t(F_G/Q) = Z(Q),$$

$$F_G = Qf_t^{-1}(Z(Q)).$$

(11)

The essential point is that  $F_G$  is to be taken as a Morse Function, in the sense of the Mathematical Appendix, so that Pettini’s (2007) topological hypothesis applies, and Landau’s symmetry breaking arguments carry through, albeit in a groupoid context. If  $H_{G_i}$  and  $Q$  are not easily separable, then the algebra becomes more complicated, but the principle remains the same.

The essential point is that, if a system evolving in an open manner can be described in terms of information sources, then the asymptotic limit theorems constraining the behavior of such sources permit construction of a Morse Function inherently generating punctuated equilibria for that system.

### 4.4 Large deviations

Wallace (2010a) has taken a more recognizable nonequilibrium statistical mechanics approach to evolutionary dynamics. In that work the interaction of genes, (cognitive) gene expression, and environmental information sources is expressed using the coevolutionary formalism of Chapagnat et al. (2006). The basic idea is to write each information source as a function of those with which it interacts:

$$H_m = H_m(Q_1, \dots, Q_s, \dots, H_j \dots), j \neq m.$$

where the  $Q_k$  represent other relevant parameters. The dynamics of such a system is defined by the usual recursive network of stochastic differential equations, using gradients in a ‘disorder’ construct as analogs to the more usual gradients in entropy, the thermodynamic forces:

$$S_m \equiv H_m - \sum_j \partial H_m / \partial K_j,$$

(12)

where we have expressed both the  $H_j$  and  $Q_j$  as driving parameters  $K_j$ , again with the proviso that one not express  $H_m$  directly as a function of itself.

Then, via the homology between information and free energy, the dynamics become driven by the usual Onsager set of stochastic differential equations,

$$dK_t^j = \sum_i [L_{i,j}(t, \dots, \partial S_m / \partial K_i \dots) dt + \sigma_{i,j}(t, \dots, \partial S_m / \partial K_i \dots) dB_t^i] =$$

$$L_j(t, K_1, \dots, K_n) dt + \sum_i \sigma_{i,j}(t, K_1, \dots, K_n) dB_t^i$$

(13)

where we have collected and simplified terms.  $L_j$  and the  $\sigma_{i,j}$  are functions, and the terms  $dB_t^j$  represent different kinds of ‘noise’ constrained by particular forms of quadratic variation, in the usual manner. Standard texts abound.

Again, since information sources are not locally reversible, there are no ‘Onsager reciprocal relations’.

Several patterns are obvious.

1. Setting this system of equations to zero and solving for stationary points gives quasi-equilibrium attractor states since the noise terms preclude unstable equilibria. The system then undergoes diffusive drift about the equilibrium configuration.

2. The system may converge to a limit cycle or a pseudo-random strange attractor.

3. What is converged to, however, is not a simple state or set of such states. Rather, this system, via the constraints imposed by the asymptotic limit theorems of information theory, converges to an equivalence class of highly dynamic information sources coupled by mutual crosstalk, and equivalence classes define groupoids, as above. In effect, via the Shannon-McMillan Theorem that defines the information source uncertainty, we have driven the mathematical thicket one layer back, expressing a dynamical system in terms of a relatively simple formalism abducted from nonequilibrium statistical mechanics.

As Champagnat et al. (2006) note, however, shifts between the quasi-equilibria of this system of equations can be addressed by the large deviations formalism. They find that the issue of evolutionary dynamics drifting away from trajectories predicted by the canonical equation can be investigated by considering the asymptotic of the probability of ‘rare events’ for the sample paths of the diffusion.

By ‘rare events’ they mean diffusion paths drifting far away from the canonical equation. The probability of such rare events is governed by a large deviation principle: when a critical parameter (designated  $\epsilon$ ) goes to zero, the probability that the sample path of the diffusion is close to a given rare path  $\phi$  decreases exponentially to 0 with rate  $I(\phi)$ , where the ‘rate function’  $I$  can be expressed in terms of the parameters of the diffusion. This result, in their view, can be used to study long-time behavior of the diffusion process when there are multiple attractive evolutionary singularities. Under proper conditions the most likely path followed by the diffusion when exiting a basin of attraction is the one minimizing the rate function  $I$  over all the appropriate trajectories. The time needed to exit the basin is of the order  $\exp(H/\epsilon)$  where  $H$  is a quasi-potential representing the minimum of the rate function  $I$  over all possible trajectories.

An essential fact of large deviations theory is that the rate function  $I$  which Champagnat et al. invoke can almost always be expressed as a kind of entropy, that is, having the canonical form

$$I = - \sum_j P_j \log(P_j)$$

for some probability distribution. This result goes under a number of names; Sanov’s Theorem, Cramer’s Theorem, the Gartner-Ellis Theorem, the Shannon-McMillan Theorem, and so forth (Dembo and Zeitouni, 1998; R. Wallace and R.G. Wallace, 2008).

These considerations lead very much in the direction of equation (13), but now seen as subject to internally-driven large deviations *that are themselves described as information sources*, providing  $H$ -parameters that can trigger punctuated shifts between quasi-stable modes, in addition to resilience transitions driven by ‘catastrophic’ external events or the exchange of heritage information between different classes of organisms.

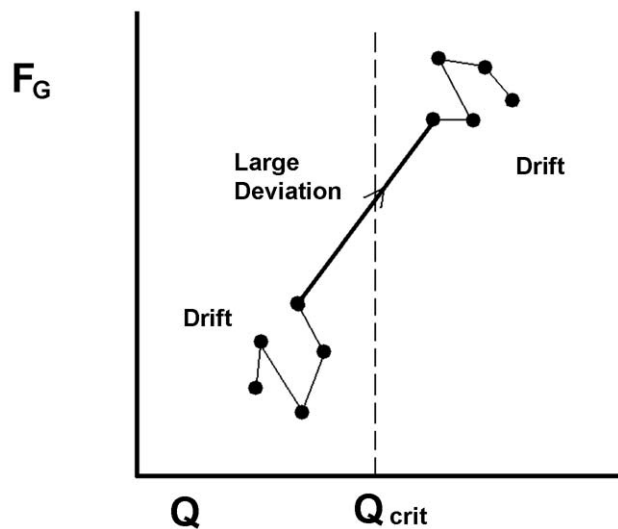


Figure 1: Spontaneous symmetry breaking in  $F_G$  as an approximation to a structured large deviations transition driven by increase in an available energy parameter. Unlike a simple physical system, such a transition can occur if  $Q$  increases beyond  $Q_{crit}$ , but will not do so in the absence of a highly structured large deviation. Increase in  $Q$  is therefore a necessary, but not sufficient, condition.

Figure 1 is a schematic that links this perspective to the Morse Theory treatment of section 4.3.  $F_G$ , as constructed, is a kind of systemic information source subject to punctuated transitions in a driving ‘metabolic’ parameter that we call  $Q$ . As  $Q$  increases, spontaneous symmetry breaking permits, say, a transition to eukaryotic structures via serial endosymbiosis: the transition from the lower cluster to the higher. But this is seen to take place via a highly structured large deviation that is itself constrained as being the output of an information source.

The spontaneous symmetry breaking argument is thereby seen as a simplified approximation to the coevolutionary formalism of Champagnat et al. (2006), as adapted by Wallace (2010a). Such transitions can occur, but, unlike simple physical systems, need not occur, in the absence of a large deviation that is itself highly structured. To reiterate, in figure 1, increase of available metabolic free energy is a necessary, but not sufficient, condition for punctuated evolutionary change that must be driven by a ‘self-dynamic’ and highly structured large deviation having its own grammar and syntax.

## 5 Discussion and conclusions

Clearly, something analogous to what Goldenfeld and Woese (2010) want to do can, in fact, be done, at least in terms of a theory of evolutionary collective phenomena that has roots

in physics. But life is not physics: the self-referential nature of evolutionary process is truly something different. While dependent on free energy and constrained by physical principles, from the perspectives of this analysis, evolution is a language that speaks itself. For example, available metabolic free energy, written as  $Q(M)$  above, can itself be an evolutionary product, as with the aerobic transition. The formal description of such bootstrapping will require more comprehensive methods than are available by abduction from relatively simple physical theory, as Goldenfeld and Woese (2010) have noted.

Again, figure 1 suggests that changes in available metabolic free energy were a necessary, but not sufficient, condition for the eukaryotic transition. Evolution is indeed self referential.

Gene expression is a cognitive process (e.g., Wallace and Wallace, 2009) that takes cues from the embedding environment to produce a phenotype response. Modes of such expression having adaptive value can become fixed in genetic (or cultural) heritage by selection. Some very simple organisms can, in fact, increase their rates of mutation in response to environmental stress. Some might even use such signals to direct mutational strategies, as Goldenfeld and Woese (2010) suggest. But evolution will remain a self-dynamic, self-referential, continually bootstrapping phenomenon, in effect, a language that speaks itself, and life is not physics. Although statistical physics and more reductionist theories are certainly tools useful in the study of data on living systems, the underlying scientific discipline remains that of biology.

There is a cautionary note. One wag has put it as ‘All mathematical models are wrong, but some are useful’. Pielou (1977, p. 106) warns that mathematical models in biology and ecology are only useful as subordinate partners in a continuing dialog with data: models can only recommend perspectives for empirical test. Replacing the intellectual straightjacket of mathematical population genetics with one driven by the asymptotic limit theorems of information theory will not address the essential scientific problems now facing evolutionary theory. These will yield only to data-based empirical study in which mathematical models are only one among many possible tools: the word is not the thing.

## 6 Mathematical appendix

### 6.1 Groupoids

Following Weinstein (1996), states  $a_j, a_k$  in a set  $A$  are related by the groupoid morphism if and only if there exists a high-probability grammatical path connecting them to the same base point, and tuning across the various possible ways in which that can happen parameterizes the set of equivalence relations and creates the groupoid. This assertion requires some development.

Note that not all possible pairs of states  $(a_j, a_k)$  can be connected by such a morphism, that is, by a high-probability, grammatical and syntactical path linking them with some given base point. Those that can define the groupoid element, a morphism  $g = (a_j, a_k)$  having the natural inverse

$g^{-1} = (a_k, a_j)$ . Given such a pairing, it is possible to define ‘natural’ end-point maps  $\alpha(g) = a_j, \beta(g) = a_k$  from the set of morphisms  $G$  into  $A$ , and a formally associative product in the groupoid  $g_1 g_2$  provided  $\alpha(g_1 g_2) = \alpha(g_1), \beta(g_1 g_2) = \beta(g_2)$ , and  $\beta(g_1) = \alpha(g_2)$ . Then the product is defined, and associative,  $(g_1 g_2) g_3 = g_1 (g_2 g_3)$ .

In addition, there are natural left and right identity elements  $\lambda_g, \rho_g$  such that  $\lambda_g g = g = g \rho_g$  (Weinstein, 1996).

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  $\beta(g) = a_k$ . Following Cannas da Silva and Weinstein (1999), we note that 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  $(\alpha, \beta) : G \rightarrow A \times A$  is a morphism from  $G$  to the pair groupoid of  $A$ . The image of  $(\alpha, \beta)$  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. As Weinstein (1996) explains, a groupoid  $G$  is a topological groupoid over a base space  $X$  if  $G$  and  $X$  are topological spaces and  $\alpha, \beta$  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 to a base point by a meaningful path of an information source dual to a cognitive process.

As Weinstein (1996) points out, the morphism  $(\alpha, \beta)$  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 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  $G_{\lambda}$ , and  $ab$  is then just the product in the group  $G_{\lambda}$ . There is an identity  $1_{\lambda}$  for each  $\lambda \in \Lambda$ . The maps  $\alpha, \beta$  coincide and map  $G_{\lambda}$  to  $\lambda$ ,  $\lambda \in \Lambda$ .

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 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.

The groupoid approach has become quite popular in the study of networks of coupled dynamical systems which can be defined by differential equation models, (Golubitsky and Stewart, 2006).

## 6.2 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 a ‘free energy’ constructed from information source uncertainties on a parameter space and ‘second order’ iterations involving parameter manifolds determining critical behavior. These can be reformulated from a Morse theory perspective. Here we follow closely the elegant treatments of Pettini (2007) and Kastner (2006).

The essential idea of Morse theory 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

$$H_{i,j} = \partial^2 f / \partial x^i \partial x^j,$$

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  $H$  at  $x_c$ .

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

Again following Pettini (2007), the essential results of Morse theory are:

1. If an interval  $[a, b]$  contains no critical values of  $f$ , then the topology of  $f^{-1}[a, v]$  does not change for any  $v \in (a, b)$ . Importantly, the result is valid even if  $f$  is not a Morse function, but only a smooth function.

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

3. If  $f : M \rightarrow \mathbf{R}$  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 sometimes computed exactly once all the critical points of  $f$  are known: Let the Morse numbers  $\mu_i (i = 1, \dots, m)$  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.

Again, Pettini (2007) contains both mathematical details and further references. See, for example, Matusmoto (2002) or the classic by Milnor (1963).

## 7 References

Ash, R., (1990), Information Theory, Dover, New York.

Bennett, C., (1988), Logical depth and physical complexity. In *The Universal Turing Machine: A Half-Century Survey*. R. Herkin (ed.), pp. 227-257, Oxford University Press.

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

Canfield, D., M. Rosing, and C. Bjerrum, (2006), Early anaerobic metabolisms. *Philosophical Transactions of the Royal Society B*, 361:1819-1836.

Cannas Da Silva, A., and A. Weinstein, (1999), *Geometric Models for Noncommutative Algebras*. American Mathematical Society, Providence, RI.

Cover, T., and J. Thomas, (2006), *Elements of Information Theory*, Second Edition, John Wiley and Sons, New York.

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

Dembo, A., and O. Zeitouni, (1998), *Large Deviations and Applications*, Second Edition, Springer, New York.

Eldredge, N., and S. Gould, (1972), Punctuated equilibrium: an alternative to pyletic gradualism. In T. Schopf (ed.), *Models in Paleobiology*, Freeman Cooper and Co., San Francisco.

Feynman, R., (2000), *Lectures on Computation*, Westview Press, New York.

Goldenfeld, N., and C. Woese, (2010), *Life is physics: evolution as a collective phenomenon far from equilibrium*. ArXiv:1011.4125v1 [q-bio.PE].

Goldenfeld, N., (2010), Personal communication.

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

Kastner, M., (2006), *Phase transitions and configuration space topology*. ArXiv cond-mat/0703401.

Khinchin, A., (1957), *Mathematical Foundations of Information Theory*. Dover, New York.

Landau, L., and L. Lifshitz, (2007), *Statistical Physics*, Third Edition, Part 1. Elsevier, New York.

Matsumoto, Y., (2002), *An Introduction to Morse Theory*. *Translations of Mathematical Monographs*, Vol. 208, American Mathematical Society.

Milnor, J., (1963), *Morse Theory*. *Annals of Mathematical Studies*, Vol. 51, Princeton University Press.

Odling-Smee, F., K. Laland, and M. Feldman, (2003), *Niche Construction: The Neglected Process in Evolution*. Princeton University Press, Princeton, NJ.

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

Pielou, E., (1977), *Mathematical Ecology*. John Wiley and Sons, New York.

Wallace, R., and D. Wallace, (2009), Code, context, and epigenetic catalysis in gene expression. *Transactions on Computational Systems Biology XI*, LNBI 5750: 283-334.

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

Wallace, R., (2009), Metabolic constraints on the eukaryotic transition. *Origins of Life and Evolution of Biospheres*, 38:165-176.

Wallace, R., (2010a), Expanding the modern synthesis. *Comptes Rendus Biologies*, 333:701-709.

Wallace, R., (2010b), Expanding the modern synthesis II: Formal perspectives on the inherent role of niche construction in fitness.

<http://precedings.nature.com/documents/5059/version/2>.

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

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