

# Chapter 2

## Models of Development

### 2.1 The Spinglass Model

Ciliberti et al. (2007a,b), culminating a long series of papers, apply the spinglass model from statistical physics to organismal development in an evolutionary context. We summarize their results and look at some of the less obvious topological implications—in particular the mapping of disjoint directed homotopy classes of phenotype paths into interaction matrix space. We then extend the approach by applying a cognitive paradigm for gene expression first developed in Wallace and Wallace (2009a). Analogs to phase change arguments for physical systems generate punctuated equilibrium evolutionary transitions in a “highly natural” manner, even for the spinglass treatment, and a hierarchical extension permits incorporation of epigenetic effects as a kind of tunable catalysis.

The spinglass model of development assumes that  $N$  transcriptional regulators are represented by their expression patterns

$$\mathbf{S}(t) = [S_1(t), \dots, S_N(t)]$$

at some time  $t$  during a developmental or cell-biological process and in one cell or domain of an embryo. The transcriptional regulators influence each other's expression through cross-regulatory and autoregulatory interactions described by a matrix  $w = (w_{ij})$ . For nonzero elements, if  $w_{ij} > 0$ , the interaction is activating, if  $w_{ij} < 0$ , it is repressing.  $w$  represents, in this model, the regulatory genotype of the system, while the expression state  $\mathbf{S}(t)$  is the phenotype. These regulatory interactions change the expression of the network  $\mathbf{S}(t)$  as time progresses according to a difference equation

$$S_i(t + \Delta t) = \sigma \left[ \sum_{j=1}^N w_{ij} S_j(t) \right] \quad (2.1)$$

where  $\Delta t$  is a constant and  $\sigma$  a sigmoidal function whose value lies in the interval  $(-1, 1)$ . In the spinglass limit  $\sigma$  is the sign function, taking only the values  $\pm 1$ .

The networks of interest in the spinglass model are those whose expression state begins from a prespecified initial state  $\mathbf{S}(0)$  at time  $t = 0$  and converge to a prespecified stable equilibrium state  $\mathbf{S}_\infty$ . Such a network is termed *viable*, for obvious reasons.

After an elaborate and very difficult simulation, Ciliberti et al. (2007b) find that viable networks comprise a tiny fraction of possible ones. These could be widely scattered in the space of all possible networks and occupy disconnected islands in this space. However, direct computation indicates precisely the opposite. The metagraph of viable networks has one “giant” connected component that comprises most or all viable networks. Any two networks in this component can be reached from one another through gradual changes of one regulatory interaction at a time, changes that never leave the space of viable networks, for this calculation.

In general, within the giant component, randomly chosen pairs of networks with the same phenotype will have vastly different organization, in terms of the matrix  $(w_{ij})$ .

Define  $0 \leq d \leq 1$  as the fraction of genes that differ in their expression state between  $\mathbf{S}_0$  and  $\mathbf{S}_\infty$ .

A typical result is that for  $N = 5$  genes,  $6 \leq M \leq 7$  total regulatory interactions, and  $d = 0.4$ , full enumeration finds a total of only 37,338 viable networks out of  $6.3 \times 10^7$  possible ones (Ciliberti et al. 2007b). Long random walks through the space of viable networks, however, visit all but a very small fraction of the nodes of the metagraph, and this missing fraction decreases as  $N$  increases. Large  $N$  require elaborate Monte Carlo sampling for simulation, a difficult and computationally intensive enterprise.

In  $w$ -space Ciliberti et al. define a metric characterizing the distance between two network topologies as

$$D(w, w') = \frac{1}{2M_+} \sum_{i,j} |\text{sign}(w_{ij}) - \text{sign}(w'_{ij})|,$$

where  $M_+$  is the maximum number of regulatory interactions, and  $\text{sign}(x) = \pm 1$  depends on the sign of  $x$ , and is 0 for  $x = 0$ .

Several observations emerge directly.

1. This approach is formally similar to spinglass neural network models of learning by selection, e.g., as proposed by Toulouse et al. (1986) nearly a generation ago. Subsequent work by Baars (1988, 2005), summarized in Dehaene and Naccache (2001), suggests that such models are simply not sufficient to the task of understanding high level cognitive function, and these have been largely supplanted by complicated “global workspace” concepts whose mathematical characterization is highly nontrivial (Atmanspacher 2006).
2. What Ciliberti et al. observe, in another idiom, is that in phenotype space, in  $\mathbf{S}$ -space, the set of all paths associated with viable networks forms an equivalence class, closely analogous to the directed homotopy equivalence classes in the

sense of Goubault and Raussen (2002) and Goubault (2003). Directed homotopy differs from simple homotopy (e.g., Lee 2000) in that one uses paths from one point to another rather than loops, and seeks continuous deformations between them. See Wallace and Wallace (2009a) for another discussion, in a biological context. Thus there is, in this spinglass model, a mapping from  $S$ -space into  $(w_{ij})$  space, characterized by the metric  $D$ , that associates a unique simply connected component with each dihomotopy-like equivalence class of paths connecting two particular phenotype points. Indeed, the  $w$ -space component might well be treated according to standard homotopy arguments, i.e., using loops.

3. What one does with homotopically simply connected components is patch them together to build larger, and more interesting, topological structures, using the Seifert–Van Kampen Theorem (SVKT) (e.g., Lee 2000, Chap. 10). If paths within  $S$ -space are not continuously transformable into one another (if there are “holes”), then several distinct dihomotopy classes will exist, e.g., as in Fig. 1 and 2 of Wallace and Wallace (2009a), explored further below in terms of developmental critical periods and their “shadows.” The obvious conjecture is that, under such a circumstance, very complex topological objects may lurk in  $w$ -space, not just the simply connected component discovered by Ciliberti et al. These may, according to the SVKT, intersect as well as exist as isolated and disconnected sets.

If there are dihomotopy “holes” in  $S$ -space, consequently reflected in disconnected patches in  $w$ -space, then punctuated transitions of various sorts may well become an evolutionary norm (as in Gould 2002), even for the spinglass model.

4. A large and increasing body of work surrounding coupled cell networks invokes groupoids, a natural generalization of symmetry groups. As Dias and Stewart (2004) remark, until recently the abstract theory of coupled cell systems has mainly focused on the effects of symmetry in the network and the consequent formation of spatial and spatiotemporal patterns. The formal setting for this theory centers upon the symmetry group of the network.

Dias and Stewart conclude that analysis of robust patterns of synchrony in general coupled cell systems—that is, dynamics in which sets of cells behave identically as a consequence of the network topology—leads to the fruitful notion of the “symmetry groupoid” of a coupled cell network. A groupoid is a generalization of a group, in which products of elements are not always defined. The symmetry groupoid of a coupled cell network is a natural algebraic formalization of the “local symmetries” that relate subsets of the network to each other. In particular “admissible” vector fields—those specified by the network topology—are precisely those that are equivariant under the action of the symmetry groupoid.

The Appendix provides a summary of standard material on groupoids that will be of later use.

5. Both approaches can apparently be coarse-grained into a symbolic dynamics associated with (simple) information sources having particular grammar and syntax. The method is straightforward (e.g., Beck and Schlogl 1995; McCauly 1994). One could, thus, probably translate the spinglass results of Ciliberti et al. into symbolic dynamics, using groupoid methods to study the underlying topological objects.

6. The spinglass model of development is abstracted from longstanding (if ultimately unsuccessful) attempts at similar treatments of neural networks involved in high level cognition (e.g., Jaeger et al. 2004; Mjolsness et al. 1991; Reinitz and Sharp 1995; Sharp and Reinitz 1998). Thus and consequently Ciliberti et al. are invoking an implicit cognitive paradigm for gene expression. Cognitive process, as the philosopher Fred Dretske (1994) eloquently argues, is constrained by the necessary conditions imposed by the asymptotic limit theorems of information theory. A little work produces a very general cognitive gene expression metanetwork structure recognizably similar to that found in Ciliberti et al. The massively parallel computations are hidden, somewhat, in the required empirical fitting of regression model analogs based on the asymptotic limit theorems of information theory rather than on the central limit theorem.
7. A salient characteristic of high level cognitive process is precisely its inherent punctuation (e.g., Baars 1988, 2005; Wallace 2005b), and this emerges directly using an information theory approach via the famous homology between information and free energy. “Simple” neural network analogs will inevitably have more difficulty replicating such behavior, but as discussed, the mapping of disjoint dihomotopy equivalence classes from phenotype sequence space to disjoint sets in interaction matrix space provides a straightforward example for spinglass models.

The next sections use information theory methods to make the transition from crosssectional  $w$ -space into that of serially correlated sequences of phenotypes, expanding on the results of Wallace and Wallace (2009a).

## 2.2 Shifting Perspective: Cognition as an Information Source

As described above, Atlan and Cohen (1998), in the context of a study of the immune system, argue that the essence of cognition is the comparison of a perceived signal with an internal, learned or inherited picture of the world, and then choice of a single response from a large repertoire of possible responses.

Such choice inherently involves information and information transmission since it always generates a reduction in uncertainty, as explained in Ash (1990, p. 21).

More formally, a pattern of incoming input—like the  $\mathbf{S}(t)$  of Eq. (2.1)—is mixed in a systematic algorithmic manner with a pattern of internal ongoing activity—like the  $(w_{ij})$  according to Eq. (2.1)—to create a path of combined signals  $x = (a_0, a_1, \dots, a_n, \dots)$ —analogous to the sequence of  $\mathbf{S}(t + \Delta t)$  of Eq. (2.1), with, say,  $n = t/\Delta t$ . Each  $a_k$  thus represents some functional composition of internal and external signals.

This path is fed into a decision function,  $h$  that generates an output,  $h(x)$ , that is an element of one of two disjoint sets  $B_0$  and  $B_1$  of possible system responses:

$$B_0 \equiv \{b_0, \dots, b_k\},$$

$$B_1 \equiv \{b_{k+1}, \dots, b_m\}.$$

Assume a graded response, supposing that if

$$h(x) \in B_0,$$

the pattern is not recognized, and if

$$h(x) \in B_1,$$

the pattern has been recognized, and some action  $b_j, k + 1 \leq j \leq m$  takes place.

The principal objects of formal interest are paths  $x$  triggering pattern recognition-and-response. That is, given a fixed initial state  $a_0$ , examine all possible subsequent paths  $x$  beginning with  $a_0$  and leading to the event  $h(x) \in B_1$ . Thus  $h(a_0, \dots, a_j) \in B_0$  for all  $0 < j < m$ , but  $h(a_0, \dots, a_m) \in B_1$ .

For each positive integer  $n$ , let  $N(n)$  be the number of high probability grammatical and syntactical paths of length  $n$  which begin with some particular  $a_0$  and lead to the condition  $h(x) \in B_1$ . Call such paths “meaningful,” assuming, not unreasonably, that  $N(n)$  will be considerably less than the number of all possible paths of length  $n$  leading from  $a_0$  to the condition  $h(x) \in B_1$ .

While the combining algorithm, the form of the nonlinear oscillator, and the details of grammar and syntax are all unspecified in this model, the critical assumption which permits inference of the necessary conditions constrained by the asymptotic limit theorems of information theory is that the finite limit

$$H \equiv \lim_{n \rightarrow \infty} \frac{\log[N(n)]}{n} \quad (2.2)$$

both exists and is independent of the path  $x$ .

Define such a pattern recognition-and-response cognitive process as *ergodic*. Not all cognitive processes are likely to be ergodic in this sense, implying that  $H$ , if it indeed exists at all, is path dependent, although extension to nearly ergodic processes seems possible (Wallace and Fullilove 2008).

Invoking the spirit of the Shannon–McMillan Theorem, described in more detail in the Appendix, as choice involves an inherent reduction in uncertainty, it is then possible to define an adiabatically, piecewise stationary, ergodic (APSE) information source  $\mathbf{X}$  associated with stochastic variates  $X_j$  having joint and conditional probabilities  $P(a_0, \dots, a_n)$  and  $P(a_n|a_0, \dots, a_{n-1})$  such that appropriate conditional and joint Shannon uncertainties satisfy the classic relations

$$\begin{aligned}
H[\mathbf{X}] &= \lim_{n \rightarrow \infty} \frac{\log[N(n)]}{n} \\
&= \lim_{n \rightarrow \infty} H(X_n | X_0, \dots, X_{n-1}) \\
&= \lim_{n \rightarrow \infty} \frac{H(X_0, \dots, X_n)}{n + 1}
\end{aligned} \tag{2.3}$$

See the Mathematical Appendix for a summary of basic information theory results.

This information source is defined as *dual* to the underlying ergodic cognitive process.

*Adiabatic* means that the source has been parameterized according to some scheme, and that, over a certain range, along a particular piece, as the parameters vary, the source remains as close to stationary and ergodic as needed for information theory's central theorems to apply. *Stationary* means that the system's probabilities do not change in time, and *ergodic*, roughly, that the cross sectional means approximate long-time averages. Between pieces it is necessary to invoke various kinds of phase transition formalisms, as described more fully in Wallace (2005b) and Wallace and Fullilove (2008).

Using the developmental vernacular of Ciliberti et al., we now examine paths in phenotype space that begins at some  $\mathbf{S}_0$  and converges  $n = t/\Delta t \rightarrow \infty$  to some other  $\mathbf{S}_\infty$ . Suppose the system is conceived at  $\mathbf{S}_0$ , and  $h$  represents (for example) reproduction when an appropriate phenotype  $\mathbf{S}_\infty$  is reached. Thus  $h(x)$  can have two values, i.e.,  $B_0$  not able to reproduce, and  $B_1$ , mature enough to reproduce. Then  $x = (\mathbf{S}_0, \mathbf{S}_{\Delta t}, \dots, \mathbf{S}_{n\Delta t}, \dots)$  until  $h(x) = B_1$ .

Structure is now subsumed *within the sequential grammar and syntax of the dual information source* rather than within the cross sectional internals of  $(w_{ij})$ -space, a simplifying shift in perspective.

This transformation carries heavy computational burdens, as well as providing deeper mathematical insight.

First, the fact that “viable” networks comprise a tiny fraction of all those possible emerges trivially from the spinglass formulation simply because of the “mechanical” limit that the number of paths from  $\mathbf{S}_0$  to a fixed  $\mathbf{S}_\infty$  will always be far smaller than the total number of possible paths, most of which simply do not end on the target configuration.

A similar result also comes easily using Kolmogorov complexity (KC). The central idea of KC is to *examine the properties of the message sent* rather than of the information source sending it. The KC measure of a message is the minimal length of a program fed to a Universal Turing Machine (UTM) that replicates it. As Bennett (1982) puts it,

A string is called “algorithmically random” if it is not expressible as the output of a program much shorter than the string itself. A simple counting argument shows that, for any length  $N$ , most  $N$ -bit strings are algorithmically random. [For example], there are only enough  $N - 10$  bit programs to describe at most  $1/1024$  of all the  $N$ -bit strings.

From the information source perspective, that inherently subsumes a far larger set of dynamical processes than possible in a spinglass model or the set of things replicable by some theoretical (but unconstructable) Turing machine, the result is what Khinchin (1957) characterizes as the “E-property” of a stationary, ergodic information source. This property is that, in the limit of infinitely long output, the classification of output strings into two sets is possible:

1. a very large collection of gibberish which does not conform to underlying (sequential) rules of grammar and syntax, in a large sense, and which has near-zero probability, and
2. a relatively small “meaningful” set, in conformity with underlying structural rules, having very high probability.

The essential content of the Shannon–McMillan Theorem is that, if  $N(n)$  is the number of meaningful strings of length  $n$ , then the uncertainty of an information source  $X$  can be defined as  $H[X] = \lim_{n \rightarrow \infty} \log[N(n)]/n$ , that can be expressed in terms of joint and conditional probabilities as in Eq. (2.3) above. Proving these results for general stationary, ergodic information sources requires considerable sophisticated mathematical machinery, and does not at all emerge in the trivial manner of the spinglass or KC analyses (Cover and Thomas 2006; Dembo and Zeitouni 1998; Khinchin 1957).

Thus the information source technique carries with it an inherent mathematical burden.

Second, information source uncertainty has an important heuristic interpretation that Ash (1990) describes as follows:

...[W]e may regard a portion of text in a particular language as being produced by an information source. The probabilities  $P[X_n = a_n | X_0 = a_0, \dots, X_{n-1} = a_{n-1}]$  may be estimated from the available data about the language; in this way we can estimate the uncertainty associated with the language. A large uncertainty means, by the [Shannon–McMillan Theorem], a large number of “meaningful” sequences. Thus given two languages with uncertainties  $H_1$  and  $H_2$  respectively, if  $H_1 > H_2$ , then in the absence of noise it is easier to communicate in the first language; more can be said in the same amount of time. On the other hand, it will be easier to reconstruct a scrambled portion of text in the second language, since fewer of the possible sequences of length  $n$  are meaningful.

This will prove important below.

Third, information source uncertainty is homologous with free energy density in a physical system, a matter having implications across a broad class of dynamical behaviors.

The free energy density of a physical system having volume  $V$  and partition function  $Z(K)$  derived from the system’s Hamiltonian—the energy function—at inverse temperature  $K$  is (e.g., Landau and Lifshitz 2007)

$$\begin{aligned} F[K] &= \lim_{V \rightarrow \infty} -\frac{1}{K} \frac{\log[Z(K, V)]}{V} \\ &= \lim_{V \rightarrow \infty} \frac{\log[\hat{Z}(K, V)]}{V} \end{aligned} \tag{2.4}$$

where  $\hat{Z} = Z^{-1/K}$ .

The partition function for a physical system is the normalizing sum in an equation having the form

$$P[E_i] = \frac{\exp[-E_i/kT]}{\sum_j \exp[-E_j/kT]} \quad (2.5)$$

where  $E_i$  is the energy of state  $i$ ,  $k$  a constant, and  $T$  the system temperature, and  $P[E_i]$  is the probability of state  $i$ .

Feynman (2000), following the classic arguments of Bennett (1982, 1988) that present idealized machines using information to do work, concludes *the information contained in a message is most simply measured by the free energy needed to erase it*. Bennett's arguments are clever indeed, and Feynman's treatment of them is well worth reading.

Thus, according to this argument, source uncertainty is homologous to free energy density as defined above, i.e., from the similarity with the relation  $H = \lim_{n \rightarrow \infty} \log[N(n)]/n$ .

Ash's comment above then has an important corollary: If, for a biological system,  $H_1 > H_2$ , then, other things being equal, source 1 will require more metabolic free energy than source 2.



Gene Expression and Its Discontents  
The Social Production of Chronic Disease  
Wallace, R.; Wallace, D.  
2016, XIII, 344 p. 52 illus., Hardcover  
ISBN: 978-3-319-48077-0