

# Stigmergic Gene Transfer and Emergence of Universal Coding

Mikhail Prokopenko,<sup>1</sup> Daniel Polani,<sup>2</sup> and Matthew Chadwick<sup>3</sup>

<sup>1</sup>CSIRO Information and Communication Technology Centre, Locked bag 17, North Ryde, NSW 1670, Australia

<sup>2</sup>Department of Computer Science, University of Hertfordshire, Hatfield AL10 9AB, United Kingdom

<sup>3</sup>address, UK

We consider a simple information-theoretic model for evolutionary dynamics approaching the “coding threshold”, where the capacity to symbolically represent nucleic acid sequences emerges in response to a change in environmental conditions. We study the conditions when a coupling between the dynamics of a ‘proto-cell’ and its proto-symbolic representation becomes beneficial in terms of preserving the proto-cell’s information in a noisy environment. In particular, we are interested in understanding the behaviour at the “error threshold” level which, in our case, turns out to be a whole “error interval”. The useful coupling is accompanied by self-organization of internal processing, i.e. an increase in complexity within the evolving system. Secondly, we study whether and how different proto-cells can stigmergically share such information via a joint encoding, even if they have slightly different individual dynamics. Implications for the emergence of biological genetic code are discussed.

## I. INTRODUCTION

### A. Understanding emergence of genotype-phenotype relationship

One of the most fundamental problems in biology and artificial life is the definition and understanding of “the gene”. As pointed out by Carl Woese, whose work provided a very strong motivation for this study, this problem continues to contribute to much debate between classical biologists who understand “the gene to be defined by the genotype-phenotype relationship, by gene expression as well as gene replication” and many molecular biologists who declared the problem to be solved when the Watson-Crick structure of DNA clearly revealed the mechanism of gene replication (Woese, 2004). Woese strongly argues against fundamentalist reductionism and presents the real problem of the gene as “how the genotype-phenotype relationship had come to be”. In other words, the main question is how the mechanism of translation evolved.

The evolution of the translation mechanism is a complicated process, and we may only intend to analyse its simplified models. However, in doing so we shall take a principled approach and consider a model of evolutionary dynamics in a generic information-theoretic way, without obscuring it with hypothetical aspects such as biochemical composition of “primordial soup”, structural properties of procaryotic cells, susceptibility of aminoacyl-tRNA synthetases to horizontal gene transfer (HGT), etc. The simple assumptions that we make, following Woese (2004), include the notion of primitive cells as loosely connected conglomerates existing during the “era of nucleic acid life” (Vetsigian et al., 2006; Woese, 1972), and the conjecture that primitive cell organization was “largely horizontal” in nature (Woese, 1998; Woese and Fox, 1977), making the simple cellular componentry open to HGT.

In taking the information-theoretic view, we focus on the “coding threshold” separating the phase of nucleic acid life from the evolutionary stage “where the capacity to represent nucleic acid sequence symbolically in terms of a (colinear) amino acid sequence developed” (Woese, 2004). More precisely, we hope to understand the pressures that forced such

a transition to “proto-symbols” encoding features of primitive cells in dedicated sequences and enabling a rudimentary translation. The analysis presented by Woese (Woese, 2004) sheds light not only on this transition, but also on saltations that have occurred at other times, e.g. advents of multicellularity and language. The common feature is “the emergence of higher levels of organization, which bring with them qualitatively new properties, properties that are describable in reductionist terms but that are neither predictable nor fully explainable therein” (Woese, 2004).

More importantly, the reason for the increase in complexity can be identified as *communication* within a complex, sophisticated network of interactions: “translationally produced proteins, multicellular organisms, and social structures are each the result of, emerge from, fields of interaction when the latter attain a certain degree of complexity and specificity” (Barbieri, 2003; Woese, 2004). The increase of complexity is also linked to adding new dimensions to the phase space within which the evolution occurs, i.e. expansion of the network of interacting elements that forms the medium within which the new level of organization (entities) comes into existence (Barbieri, 2003; Woese, 2004).

### B. Guiding self-organization

An increase of complexity is one of the landmarks of self-organization. The latter is usually defined as an increase in order within an open system, without an explicit external control. In addition, it is expected that the increased order, i.e. the more complex inner organization, exhibits both robustness and dynamics. Robustness is understood if the system continues to function in the face of perturbations (Wagner, 2005), while dynamics are interpreted via local interactions among subsystems or components of the system. Kauffman (Kauffman, 2000) suggests that the underlying principle of self-organization is the generation of constraints in the release of energy. According to this view, the constrained release allows for such energy to be controlled and channelled to perform some useful work. This work in turn can be used to build better and more efficient constraints for the release of

further energy and so on. As pointed out by Prokopenko et al. (2008), the lack of agreement of what is meant by complexity, constraints, etc. leaves any definition of self-organization somehow vague. A quantitative approach suggests to measure complexity precisely, and demands that the complexity of external influence into a self-organizing system should be strictly less than the gain in internal complexity (Prokopenko et al., 2008).

These observations can be formalized information-theoretically. More precisely, we intend to consider a communication channel between a proto-cell and itself at a future time point, and pose a question of the channel capacity constrained by the noise. According to this approach, polluting the channel with the noise corresponds to adding constraints on self-organization, guiding it in a specific way. By varying the nature and degree of the noise prevalent in the environment within which such proto-cells exist and evolve, we hope to identify conditions leading to self-organization of an efficient coupling between the proto-cell *per se* and its encoding with “proto-symbols”. Specifically, we investigate conditions under which such coupling is beneficial in terms of preserving the information within the noisy communication channel across time. We intend to demonstrate that the coupling evolves to protect some information about the proto-cell in the encoding. A rudimentary translation may help to recover the information that otherwise would have been lost due to the noise.

### C. Stigmergic Gene Transfer

It is important to realize two features of the early phase in cellular evolution that existed before the “coding threshold”. First of all, the “players are cell-like entities still in early stages of their evolution”, and that “the evolutionary dynamics... involves communal descent” (Vetsigian et al., 2006). That is, the cells are not yet well-formed entities that replicate completely, with an error-correcting mechanism. Rather, the proto-cells can be thought of as conglomerates of substrates, that exchange components with their neighbours freely — horizontally. The notion of vertical descent from one “generation” to the next is not yet well-defined. This means that the descent with variation from one “generation” to the next is not genealogically traceable but is a descent of a cellular community as a whole.

Secondly, genetic code that appears at the coding threshold is “not only a protocol for encoding amino acid sequences in the genome but also an innovation-sharing protocol” (Vetsigian et al., 2006), as it used not only as a part of the mechanism for cell replication, but also as a way to encode relevant information about the environment. Different proto-cells may come up with different innovations that make them more fit to the environment, and the “horizontal” exchange of such information may be assisted by an innovation-sharing protocol - a proto-code. With time, the proto-code develops into a universal genetic code.

Such innovation-sharing is perceived to have a price: it implies ambiguous translation where the assignment of codons

to amino acids is not unique but spread over related codons and amino acids. (Vetsigian et al., 2006). In other words, accepting innovations from neighbours requires that the receiving proto-cell is sufficiently flexible in translating the incoming fragments of the proto-code. Such a flexible translation mechanism, of course, would produce imprecise copies. However, a descent of the whole innovation-sharing community may be traceable: i.e., in a statistical sense, the next “generation” should be correlated with the previous one. As noted by Woese (2004),

A sufficiently imprecise translation mechanism could produce “statistical proteins”, proteins whose sequences are only approximate translations of their respective genes (Woese, 1965). While any individual protein of this kind is only a highly imprecise translation of the underlying gene, a consensus sequence for the various imprecise translations of that gene would closely approximate an exact translation of it.

In other words, a given gene can be translated not into a unique protein but instead into a family of related protein sequences: “early life did not require a refined level of tolerance” (Vetsigian et al., 2006). Looseness of the outcome is implied by an imprecise genome replication comprising relatively few unique genes (Woese and Fox, 1977) — therefore, rather than trying to develop a dynamical system (a proto-cell plus encoding) that fully preserves the information about the conglomerate, we only need to develop dynamics that corresponds to “statistical proteins”, preserving information in a “consensus sequence”. While any individual protein of this kind is only a highly imprecise translation of the underlying gene, a consensus sequence for the various imprecise translations of that gene would closely approximate an exact translation of it”. That is, the consensus sequence would capture the main information content of the innovation-sharing community.

Moreover, it can be argued that the universality of the code is a generic consequence of early communal evolution mediated by horizontal gene transfer (HGT), and that thus HGT enhances optimality of the code (Vetsigian et al., 2006):

HGT of protein coding regions and HGT of translational components ensures the emergence of clusters of similar codes and compatible translational machineries. Different clusters compete for niches, and because of the benefits of the communal evolution, the only stable solution of the cluster dynamics is universality.

The adopted information-theoretic view allows us to handle particular HGT scenarios where certain fragments necessary for cellular evolution begin to play the role of the proto-code. One scenario may assume that the proto-code is initially located within its proto-cell, and is functionally “separated” from the rest of the cell when such a split becomes beneficial. Another scenario suggests that the proto-code is present in an environmental locality, and subsequently entrapped by the proto-cells that benefit from such interactions. We believe

that the first scenario (“internal split”) is less likely to produce either universal code or universal translational machinery than the second scenario (“entrapment”). In general, it is quite possible that internal split and entrapment played complementary roles. Importantly, however, there was an indirect exchange of information among the cells via their local environment, which is indicative of stigmergy. Henceforth, we would like to refer to such gene transfer as *stigmergic gene transfer* (SGT): proto-cells find matching fragments, use them for coding, modify and evolve their translation machinery, and exchange certain fragments with each other via the local environment. SGT can be thought of as a sub-class of HGT, differing from the latter in that the fragments exchanged between two proto-cells may be modified during the transfer process by other cells in the locality.

As pointed out by Polani et al. (2006), information should not be considered simply as something that is transported from one point to another as a “bulk” quantity — instead, “looking at the intrinsic dynamics of information can provide insight into inner structure of information”. It is conjectured that maximization of information transfer through selected channels is one of the main evolutionary pressures (Bialek et al., 2006; Klyubin et al., 2007; Laughlin et al., 2000; Lizier et al., 2008; Piraveenan et al., 2007a,b, 2009; Prokopenko et al., 2006): although the evolutionary process involves a larger number of drives and constraints, information fidelity (i.e. preservation) is a consistent motif throughout biology. Modern evolution operates close to the error threshold (Adami, 1998), and biological sensorimotor equipment typically exhausts the available informatory capacity (under given constraints) close to the limit (Laughlin et al., 1998). Adami, for instance, argues that the evolutionary process extracts valuable information and stores it in the genes (Adami, 1998). Since this process is relatively slow (Bennett, 1990; Lloyd, 1990), it is a selective advantage to preserve this information, once captured.

In this paper, we follow the model of Piraveenan et al. (2007a) and Polani et al. (2008), focusing on the information preservation property of evolution within a coupled dynamical system. These previous studies verified that the ability to symbolically encode nucleic acid sequences does not develop when environmental noise  $\varphi$  is too large or too small. In other words, it is precisely a limited reduction in the information channel’s capacity, brought about by the environmental noise, that creates the appropriate selection pressure for the coupling between a proto-cell and its encoding. Here we extend these models by introducing co-evolution of multiple proto-cells entrapping a common encoding using SGT.

In the following, we shall concentrate on the information preservation property of evolution in the vicinity of the “coding threshold”. Everything else is modeled minimalistically: we encapsulate the influence of evolutionary constraints within a dynamical system, and represent the acquisition of valuable information by an explicit “injection” of information at the beginning of each trajectory.

## II. MODELLING EVOLUTIONARY DYNAMICS

Our generic model for evolutionary dynamics involves a dynamical coupled system, where a proto-cell is coupled with its potential encoding, evolving in a fitness landscape shaped by a selection pressure. The selection pressure rewards preservation of information in presence of both environmental noise and inaccuracy of internal coupling. When the proto-cell is represented as a dynamical system, the information about it may be captured generically via the structure of the phase-space (e.g., states and attractors) of the dynamical system. In particular, a loss of such information corresponds to a loss of structure in the phase-space, while informational recovery would correspond to recovery of the equivalent (e.g., isomorphic) structure in the phase-space. Importantly, the information about the attractors can be compactly encoded if there is a need for it.

For example, the states of the system may loosely correspond to dominant substrates (e.g., prototypical amino acids), used by the cell. The chosen representation does not have to deal with the precise dynamics of biochemical interactions within the cell, but rather focuses on structural questions of the cell’s behavior: does it have more than one attractor, are the attractors stable (periodic) or chaotic, how many states do the attractors cycle through, etc. Representing the dynamics in this way avoids the need to simulate the unknown cellular machinery, but allows us to analyze under which environmental conditions the SGT may have become beneficial. In particular, if the potential encoding develops to have a compact structure that matches the structure of the cell’s phase-space, then the encoding would be useful in recovering such structure, should the latter be affected by environmental noise. Information is understood in Shannon sense (reduction of uncertainty), and a loss of such information corresponds to a loss of structure in the phase-space. At the same time, informational recovery would correspond to recovery of some isomorphic structure in the phase-space.

### A. A model with direct fusion of system and translation dynamics

The generic dynamical coupled system (Polani et al., 2008) is described by the equations

$$X_{t,m} = \begin{cases} f_m(X_{t-1,m}) + \varphi_t & t \neq t^* \\ \alpha [f_m(X_{t-1,m}) + \varphi_t] + (1 - \alpha)h_m(Y_{t-1,m} + \psi_{t,m}) & t = t^* \end{cases} \quad (1)$$

$$Y_{t,m} = \begin{cases} g_m(X_{t,m} + \psi_{t,m}) & t = t_0 \\ Y_{t-1,m} & t > t_0 \end{cases} \quad (2)$$

where  $X_{t,m}$  are the variables that describe multiple proto-cells,  $1 \leq m \leq M$ , and and  $Y_{t,m}$  their potential encodings at time  $t$ , respectively. Function  $f_m$  defines the dynamical system representing the dynamic for proto-cell  $m$ . The function  $g$  is a mapping from  $[0, 1]$  to  $[0, 1]$ . Parameter  $\alpha \in [0, 1]$  sets the relative importance of the translation  $h$  from symbols

(e.g., proto-codons) into the proto-cell state (e.g., proto amino acids).

In the simplest case,  $m = 1$  (one cell), and  $\alpha = 1/2$ , the system reduces to

$$X_t = \begin{cases} f(X_{t-1}) + \varphi_t & t \neq t^* \\ \frac{1}{2} [f(X_{t-1}) + \varphi_t] + \frac{1}{2} h(Y_{t-1} + \psi_t) & t = t^* \end{cases} \quad (3)$$

$$Y_t = \begin{cases} g(X_t + \psi_t) & t = t_0 \\ Y_{t-1} & t > t_0 \end{cases} \quad (4)$$

The function  $\varphi_t$  describes the external (environment) noise that affects the proto-cells: it is the same for all cells, i.e.  $\varphi_t$  is independent of  $m$ . This noise represents a pressure to push the system  $X$  towards certain attractors (implementation is described in Appendix VII.A).

The function  $\psi_{t,m}$  represents both the matching noise associated with accessing information from  $X_{t_0,m}$  by  $Y_{t_0,m}$  at time  $t_0$ , and the noise of ambiguous back-translation (applied only at  $t^*$ ). In other words, it represents the inaccuracy within the internal encoding/translation channel. In addition, the noise  $\psi$  may be interpreted as inaccuracy of the environment's representation within the encoding  $Y$  which indirectly "perceives" the environment through the system  $X$  (implementation is described in Appendix VII.A).

## B. Entrapment and SGT

The entrapment mechanism that matches the information from the proto-cell with its encoding (i.e. which encodes its information) at time  $t_0$  is given by  $g_m$ . At time  $t = t_0$ , noise is introduced into the environment affecting dynamics of the proto-cell. At the same time  $t = t_0$ , information from the proto-cell  $X_{t_0,m}$  is accessed by the system  $Y_{t_0,m}$  (encoding) via the matching function  $g_m$ . This process is affected by the noise  $\psi$ . The feedback from  $Y$  to  $X$  (henceforth we drop subscripts when the meaning is clear) occurs at time  $t^*$ , i.e. the function  $h_m$  translates the input  $Y_{t^*-1,m}$  from the encoding back into the proto-cell. This internal translation is subjected to internal noise as well.

In evolving the potential encoding system  $Y$  coupled with  $X$  via a suitable function  $g$ , one attempts to preserve information between the initial  $X_{t_0}$  and recovered  $X_{t^*}$  states of the system, as described in Section III.

Piraveenan et al. (2007a) considered the case  $m = 1$ , equations (3)–(4), and function  $h$  being the identity (a single system). Polani et al. (2008) considered a system with multiple proto-cells:  $m \geq 1$ , and contrasted universality of the translation machinery: all functions  $h_m$  are identical, while  $g_i \neq g_j$  for  $i \neq j$ , with universality of the proto-code: all proto-codes  $g_m$  are identical, while  $h_i \neq h_j$  for  $i \neq j$ . The system (1)–(2) is coupled not only due to the common environment noise  $\varphi$ , but also due to the shared translation machinery  $h$  or shared proto-code  $g$ . This coupling supported a simple information-theoretic model of HGT and specifically, SGT. As only the information content is dealt with, the consideration of identical  $h_m$ 's and/or identical  $g_m$ 's allowed to study gene transfers without details of molecular (state-to-state) interactions.

However, the numeric nature of the composition

$$\alpha [f_m(X_{t-1,m}) + \varphi_t] + (1 - \alpha)h_m(Y_{t-1,m} + \psi_{t,m})$$

in equation (1) obscures the pure information-theoretic view. More precisely, the linear combination (direct fusion) of original dynamics and back-translation of the encoding places a bias on possible encodings. In the next section we describe an improved model which is free of this shortcoming.

## C. An SGT model with indirect fusion

As mentioned above, the objective of modeling is to preserve information between the initial  $X_{t_0}$  and recovered  $X_{t^*}$  states of the system. One may then pose a question whether it is necessary to explicitly model a recovered system  $X_{t^*}$ . After all, the information-theoretic framework allows us to formulate this question in terms of computing a difference between the initial system  $X_{t_0}$  and the joint system  $(X_{t^*-1}, Y_{t^*-1})$ , where the latter is sampled *before* any possible back-translation is applied. In other words, one demands that the system  $(X_{t^*-1}, Y_{t^*-1})$  jointly preserves information about the original system. This formalization becomes possible because the information accessible by any back-translation is contained within the joint system  $(X_{t^*-1}, Y_{t^*-1})$  anyway.

Removing back-translation from the description simplifies the model as follows:

$$X_{t,m} = f_m(X_{t-1,m}) + \varphi_t \quad (5)$$

$$Y_{t,m} = \begin{cases} g_m(X_{t,m} + \psi_{t,m}) & t = t_0 \\ Y_{t-1,m} & t > t_0 \end{cases} \quad (6)$$

Capturing difference between the initial system  $X_{t_0}$  and the joint system  $(X_{t^*-1}, Y_{t^*-1})$ , where  $t^* > t_0$ , information-theoretically will formalize our objective function: information preservation. When dealing with multiple systems, one needs to consider respective differences between the initial systems  $X_{t_0,m}$  and the joint systems  $(X_{t^*-1,m}, Y_{t^*-1,m})$ .

The enhanced model (indirect fusion of system and translation dynamics) can be further extended by reintroducing the back-translation function  $Z(t^*) = h(X_{t^*-1}, Y_{t^*-1})$ , and measuring a difference between the initial system  $X_{t_0}$  and the recovered system  $Z(t^*)$ . This will be a subject of future research.

## D. Coupled logistic maps

Each dynamical system is a logistic map  $X_{t+1} = rX_t(1 - X_t) + \delta_m$ , where  $r$  is a parameter, and  $\delta_m$  is an additive constant (see Appendix VII.B), used to differentiate between multiple systems  $f_m$ . We used  $r = 3.5$ , resulting in four states of the attractor of the logistic map for each of the multiple systems  $X_m$ .

Coupled logistic maps have been extensively used in modelling of biological processes. One prominent study is the investigation of spatial heterogeneity in population dynamics

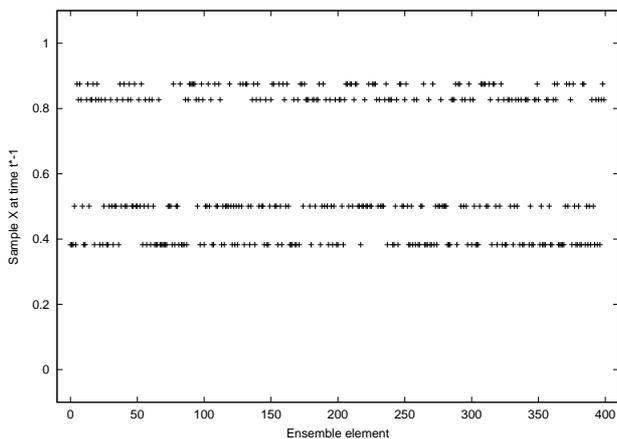


FIG. 1 Initial sample ( $X_{t_0}$ ). Noise  $\varphi = 0.025$ .

(Lloyd, 1995) who examined the dynamic behaviour of the model using numerical methods and observed a wide range of behaviours. For instance, the coupling was shown to stabilize individually chaotic populations as well as cause individually stable periodic populations to undergo more complex behaviour. Importantly, a single logistic map can only have one attracting periodic orbit, but multiple attractors were shown by Lloyd (1995) for coupled logistic maps.

Logistic maps were chosen to model the system (5)–(6) mostly due to their simplicity, well-understood behaviour in the vicinity of chaotic regimes (e.g., bifurcations and symmetry breaking), the possibility of multiple attractors in coupled maps, as well as their ability to capture both reproduction and starvation effects (that are important for studying the structure in the phase-space).

Let us consider an example, with a single system. Original information is represented by four clear clusters observed in Figure 1. Figure 2 shows the ensemble  $[X]$  at the time  $t^* - 1$ . The environment noise  $\varphi$  disrupts the logistic map dynamics, and some information about the attractor of  $X$  and its four states is lost in the course of time: the observed sample ( $X_{t^*-1}$ ) does not contain four clear clusters.

The encoding structure that is to be evolved in  $Y$  can be associated with “proto-symbols” (“codes”) that help to complement at time  $t^* - 1$  the remaining information contained in the “polluted” system ( $X_{t^*-1}$ ).

### III. INFORMATION PRESERVATION

Information Theory was originally developed by Shannon (1948) for reliable transmission of information from a source  $A$  to a receiver  $B$  over noisy communication channels. Put simply, it addresses the question of “how can we achieve perfect communication over an imperfect, noisy communication channel?” (MacKay, 2003). When dealing with outcomes of imperfect probabilistic processes, it is useful to define the information content of an outcome  $x$  which has the probability  $P(a)$ , as  $\log_2 \frac{1}{P(a)}$  (it is measured in bits): improbable outcomes convey more information than probable outcomes.

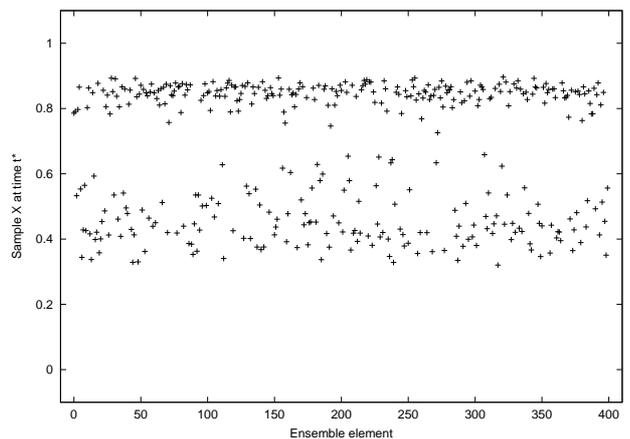


FIG. 2 Two remaining “clusters” in the sample ( $X_{t^*-1}$ ). Noise  $\varphi = 0.025$ .

Given a probability distribution  $P$  over the outcomes  $a \in \mathcal{A}$  (a discrete random variable  $a$  representing the process, and defined by the probabilities  $P(a) \equiv P(A = a)$  given for all  $a \in \mathcal{A}$ ), the average Shannon information content of an outcome is determined by

$$H(A) = \sum_{a \in \mathcal{A}} P(a) \frac{1}{\log P(a)} = - \sum_{a \in \mathcal{A}} P(a) \log P(a) \quad (7)$$

Henceforth we omit the logarithm base 2. This quantity is known as (*information*) *entropy*. Intuitively, it measures, also in bits, the amount of freedom of choice (or the degree of randomness) contained in the process — a process with many possible outcomes has high entropy. This measure has some unique properties that make it specifically suitable for measuring “how much “choice” is involved in the selection of the event or of how uncertain we are of the outcome?” (Shannon, 1948).

In evolving the potential encoding system  $Y$  coupled with  $X$  via a suitable function  $g$ , we minimize difference between the initial system  $X_{t_0}$  and the joint system ( $X_{t^*-1}, Y_{t^*-1}$ ) at some time  $t^* - 1$ . This difference is captured information-theoretically via Crutchfield’s information distance (Crutchfield, 1990) between two components:

$$d(A, B) = H(A|B) + H(B|A) \quad (8)$$

The entropies are defined as

$$H(A|B) = H(A, B) - H(B) \quad (9)$$

$$H(A, B) = - \sum_{a \in \mathcal{A}} \sum_{b \in \mathcal{B}} P(a, b) \log P(a, b), \quad (10)$$

where  $P(a)$  is the probability that  $A$  is in the state  $a$ , and  $P(a, b)$  is the joint probability.

The distance  $d(A, B)$  measures the dissimilarity of two information sources  $A$  and  $B$ ; it is a true metric in the sense that it fulfils the axioms of metrics, including the triangle

inequality. In addition, as opposed to the mutual information used in (Piraveenan et al., 2007a), the information metric  $d(X_{t_0}, (X_{t^*-1}, Y_{t^*-1}))$  is sensitive also to the case when one information source is contained within another. While the results do not radically depend on the choice of distance  $d(X_{t_0}, (X_{t^*-1}, Y_{t^*-1}))$  over the mutual information, the former leads to a more crisp recovery of structure in the phase-space.

In addition, we wish to reward crispness in the encoding function  $g$ , in other words, express a preference toward more concise proto-codes. This preference can be simply captured by minimization of entropy,  $H(g)$ , of function  $g$ , mapping from  $[0, 1]$  to  $[0, 1]$ . This places another constrain guiding self-organization.

Combining the distance  $d(X_{t_0}, (X_{t^*-1}, Y_{t^*-1}))$  and entropy  $H(g)$  produces our information-theoretic objective function for a single system:

$$F^1 = -[d(X_{t_0}, (X_{t^*-1}, Y_{t^*-1})) + H(g)] \quad (11)$$

For a multiple system ( $M$  ensembles), the fitness function generalizes as

$$F^M = -\left[\frac{1}{M} \sum_m d(X_{t_0,m}, (X_{t^*-1,m}, Y_{t^*-1,m})) + H(g)\right] \quad (12)$$

In this case, the challenge is to produce such a universal encoding  $g$  that the corresponding systems  $Y_m = g(X_m + \psi_m)$ , using the same mapping  $g$ , are complementary to their polluted counterparts  $X_m$ . In other words, the joint systems  $(X_m, Y_m)$  preserve at a later time  $t^*$  as much information as possible about the respective initial systems  $X_m$  at time  $t_0$ , and in doing so use the same, universal, encoding  $g$ . This is an enhancement of the original SGT model with direct fusion (Polani et al., 2008) where either multiple  $h_m$ 's or multiple  $g_m$ 's were allowed.

Maximization of the fitness function is achieved by employing a simple genetic algorithm (GA) (described in the Appendix).

#### IV. RESULTS FOR A SINGLE SYSTEM

The increase in the information-theoretic fitness function observed over a number of generations is shown in Figure 3.

Let us, at this stage, analyze self-organization of structure (increase of order) within the processing function  $g$ . Figures 4 and 5 contrast a randomly selected function  $g$  at the start of the evolution (noise  $\varphi$  is set to  $\varphi = 0.025$ , while noise  $\psi = 0.015$ ), with the best individual function  $g$  after 512 generations. The important difference is in the way of mapping states of  $X$  into the encoding  $Y$ . The selection pressure resulted in a more “condensed” mapping. This is achieved by an increase in organization as well as robustness in  $g$ : a small shift from  $x$  to  $x \pm \psi$  results now only in a small difference between  $g(x)$  and  $g(x \pm \psi)$ .

The self-organization of  $g$  counters the effect of internal processing noise  $\psi$ , given the noise  $\varphi$  in the environment. In general, following (Haken, 1983; Prigogine, 1980), we may

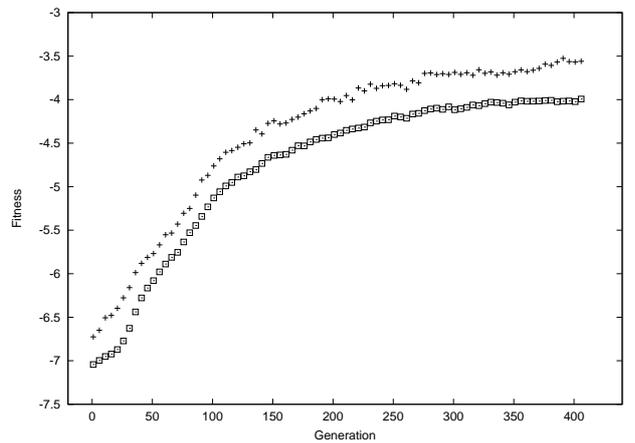


FIG. 3 Fitness in bits. Squares indicate the average fitness, '+'s show fitness of the best individual in each generation.

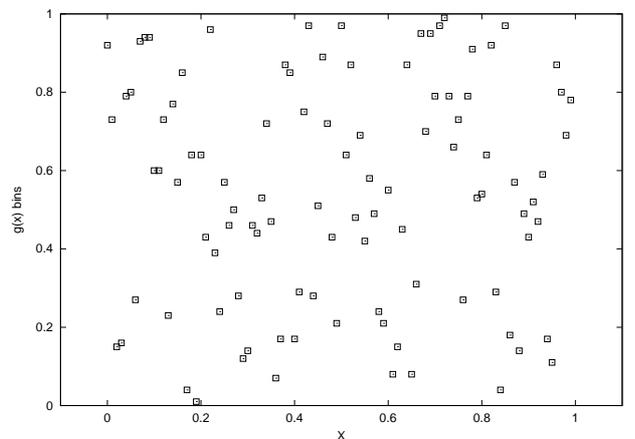


FIG. 4 One ensemble: a random function  $g$  at the start of evolution: no structure is observed.

say that self-organization results from fluctuations, that is, internal information processing has self-organized in response to environmental “pollution”. This self-organization helps  $Y$  to maintain the structure of the space  $X$  (namely, the information about the attractor’s structure): the four “proto-codes” correspond to the four states of the attractor of  $X$ .

The experiment also demonstrated that noise within the environment affects the self-organization within the encoding. Figure 6 traces fitness  $F^1$  (for the best individual), over the external noise  $\varphi$ , for different internal noise levels  $\psi$ . We can observe a steady decrease in fitness punctuated by two transitions, that form three plateaus. As conjectured by Piraveenan et al. (2007a), the encoding is not beneficial when the environmental noise  $\varphi$  is outside a certain range (in this instance,  $0.008 < \varphi < 0.03$ ). The middle plateau is precisely the region specifying this range, i.e. the “error interval”.

Be reminded that the information distance  $d(A, B)$  consists of two components: the loss  $H(A|B)$ , and the waste  $H(B|A)$ . The waste measures packaging information which envelops the proto-cell’s information, but itself does not contain any information of interest, while the loss measures how

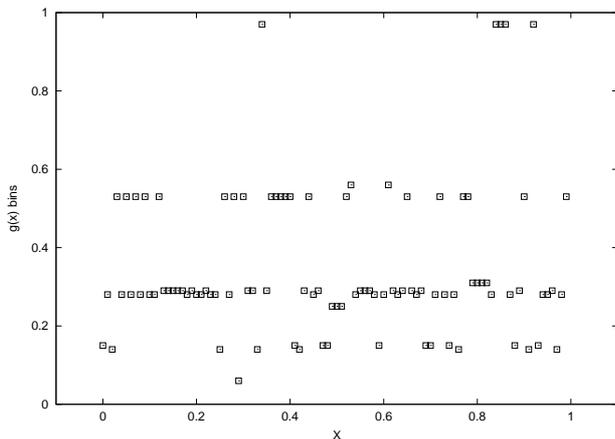


FIG. 5 One ensemble: evolved  $g$ : a self-organized encoding is observed. Contrast with Figure 4.

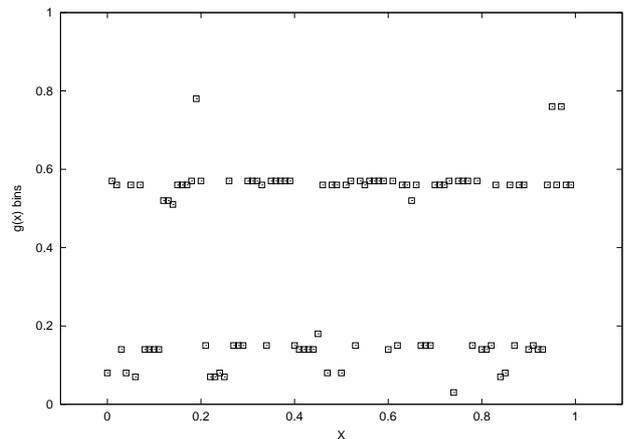


FIG. 7 Multiple ensembles: evolved  $g$ : a self-organized encoding is observed. Compare with Figure 5.

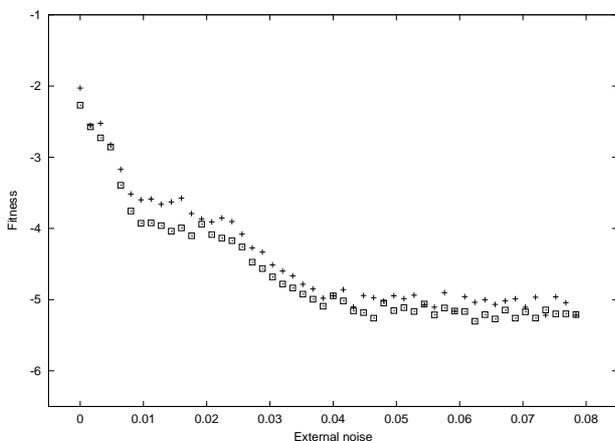


FIG. 6 One ensemble: fitness  $F^1$ , over external noise  $\varphi$ . Internal noise  $\psi = 0.015$  ('+'s), and  $\psi = 0.045$  ('squares').

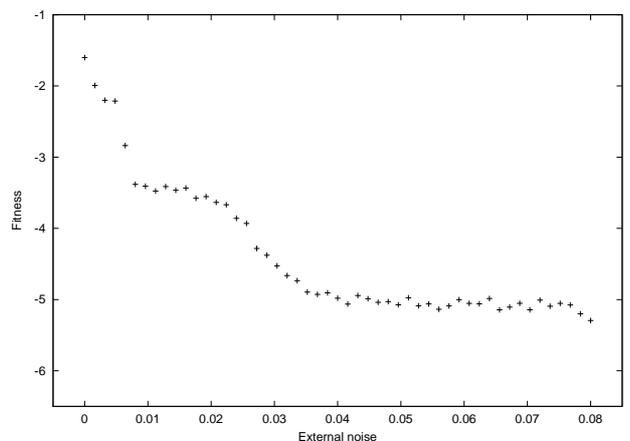


FIG. 8 Multiple ensembles: fitness  $F^M$ , over noise  $\varphi$ , for noise level  $\psi = 0.015$ . Compare with Figure 6.

much of the proto-symbolically encoded information is actually lost. Polani et al. (2008) explained the cascade of plateaus as follows: (i) everything is recoverable (the first plateau); (ii) waste appears (the medium plateau); (iii) loss appears (the last plateau).

The clustering in the encoding  $Y$  corresponds to self-organization of discrete “proto-symbols” in the encoding. The information reconstructed at time  $t^*$  will not be precise, and rather than having four crisp states,  $X$  can be described as an individual with an imprecise translation of the underlying gene within a “consensus sequence” (Woese, 2004), analogous to a “statistical protein”.

## V. RESULTS FOR MULTIPLE SYSTEMS

In this section, we now focus on a system with multiple proto-cells which share the coding channel. Concretely, we consider  $M = 4$ ,  $r = 3.5$ , for various  $\delta_m$ , and attempt to evolve a universal proto-code.

To reiterate, the challenge is to produce a universal encod-

ing  $g$  such that the corresponding systems  $Y_m = g(X_m + \psi_m)$ , using the same mapping  $g$ , complement their polluted counterparts  $X_m$  in preserving as much information as possible about the respective initial systems. As shown in Figure 7, this challenge is successfully met: a self-organized encoding is observed, and can be compared with the encoding evolved for the single system.

This supports the conjecture that multiple systems exert joint pressure on proto-code’s universality, allowing for SGT: the very same “codes” are successfully used by multiple ensembles in preserving information unique to each ensemble.

Figure 8 traces fitness  $F^M$  (for the best individual), over the external noise  $\varphi$ . We can again observe a steady decrease in fitness punctuated by two transitions and forming three plateaus.

## VI. CONCLUSION AND FUTURE WORK

We considered an information-theoretic model based on dynamical systems for self-organization of a universal encoding

able to preserve information over time, when the main system is suffering from perturbations. While doing so, we extended previous work, by employing a purely information-theoretic fitness function aimed at capturing the indirect fusion of (untranslated) encoding and system dynamics. Furthermore, we studied the effects on a small population of systems sharing an encoding, and verified the conjecture that SGT is possible.

It is striking that the pressure to develop a distinctive “symbolic” encoding does only develop if the noise in the original system is in a particular range, not too small and not too large.

Scanning through different noise levels, we observe several plateaus of the fitness corresponding to qualitative jumps in the way not only the initial state is encoded but how the system dynamics is affected by the noise. The middle plateau which is most relevant for the self-organization of distinct symbols turns out to be the most sensitive for the precise level of noise. **CHECK**

The multiple system scenario shows that universal encoding can be successfully used by several systems which differ slightly. However, at this point, we did not yet model the explicit recovery, by re-introducing back-translation function, and comparing the initial system dynamics with the dynamics recovered. This will be addressed in future work.

Woese observed that “statistical proteins form the basis of a powerful strategy for searching protein phase space, finding novel proteins” (Woese, 2004). We believe that further modelling of the evolutionary dynamics in such a space may explain mechanisms resolving Eigen’s paradox (Eigen, 1971) and leading to convergence on “the lingua franca of genetic commerce” (Woese, 2004).

## VII. APPENDIX: METHODS

### A. External and internal noise

The function  $\varphi_t$  describes the external (environment) noise that affects the proto-cells: it is the same for all cells, i.e.  $\varphi_t$  is independent of  $m$ . It is implemented as a random variable  $\varphi_t \in [-l, u]$ , where  $u > 0$  and  $l > 0$ , which is uniformly distributed, with probability 1/2, between 0 and  $l$ , and with probability 1/2 between 0 and  $u$  (sampled at each time step). In other words, positive values may be more sparsely distributed than the negative if  $u$  is larger than  $l$ .

The function  $\psi_{t,m}$  represents both the matching noise associated with accessing information from  $X_{t_0,m}$  by  $Y_{t_0,m}$  at time  $t_0$ , and the noise of ambiguous back-translation (applied only at  $t^*$ ). This noise is modelled as uniform random noise  $\psi_{t,m} \in [-b_m, b_m]$ , where  $0 < b_m \ll 1.0$ , and is used only at  $t_0$  and  $t^*$ .

### B. Coupled logistic maps

A logistic map  $X_{t+1} = rX_t(1 - X_t) + \delta_m$  is defined by parameters  $r$  and  $\delta_m$ ,  $0 \leq \delta_m < 1.0$ , used to differentiate between multiple systems  $f_m$ . That is, the function  $f_m$  is given by  $f(x) = rx(1 - x) + \delta_m$ ; if the right-hand side expression

is above zero,  $f(x)$  is set to  $f(x) - 1.0$ . The logistic map  $f$  is initialized with a value between 0.0 and 1.0, and stays within this range if the value of  $r$  is within the range  $[0, 4.0]$ . We used  $r = 3.5$ , resulting in four states of the attractor of the logistic map (if  $\delta_m = 0$  then the states are approximately 0.38, 0.50, 0.83, 0.87). Each of the multiple systems with different  $\delta_m$  possesses four states of the respective attractor. The time  $t = t_0$  is set after each logistic map settles into its attractor cycle, having passed through a transient.

### C. State-space

In order to estimate the probability distribution of a random variable ( $X$  or  $Y$ ) at a given time, we generate an initial random sample  $(X_0) = (X_0^1, X_0^2, \dots, X_0^K)$  of size  $K$ . Each  $X_0^i$ , where  $1 \leq i \leq K$ , is chosen from a uniform random distribution within  $[0.0, 1.0]$ . The mapping  $X_{t+1}^i = f(X_t^i)$  produces an ensemble of  $K$  corresponding time series,  $1 \leq i \leq K$ , denoted as  $[X] = [X_t^1, X_t^2, \dots, X_t^K]$ , where  $0 \leq t \leq T$ , and  $T$  is a time horizon. Within the ensemble, each time series  $X_t^i$  may have a different initial value  $X_0^i$ . At any given time  $t'$ , we can obtain a sample  $(X_{t'}) = (X_{t'}^1, X_{t'}^2, \dots, X_{t'}^K)$ .

Given the sample  $(X_{t_0})$  at the time  $t = t_0$ , and the mapping  $Y_{t_0} = g(X_{t_0} + \psi)$ , we can generate the sample  $(Y_{t_0}) = (Y_{t_0}^1, Y_{t_0}^2, \dots, Y_{t_0}^K)$  for the variable  $Y$ . In the corresponding ensemble  $[Y] = [Y_t^1, Y_t^2, \dots, Y_t^K]$  each sample is identical to the sample  $(Y_{t_0})$ .

We generate an ensemble of  $X_t$  time series, each series governed by equation (1). The ensemble  $[X]$  provides a fixed constraint on the optimization. For *each* function  $g$ , an ensemble  $[Y]$  is then generated, using equation (2) — i.e., the values of the series  $Y_t$  depend on the choice of function  $g$ . The ensemble  $[X]$  is kept unchanged while we evolve the population of functions  $g$ , being an optimization constraint, but the ensemble  $[Y]$  differs for each individual within the population. The fitness of each function  $g$  is defined by equation (11) or (12), and estimated via the respective entropies.

The experiments were repeated for different ensembles  $X_t$ .

### D. Genetic Algorithm

We generate a population of  $g$  functions (the size of the population is fixed at 400). In order to implement the mapping  $g$ , the domain of  $g$  is divided into  $n$  consecutive bins  $x_i$  such that  $x_i = [(i - 1)/n, i/n)$  for  $1 \leq i < n$ , where  $[a,b)$  denotes an interval open on the right, and  $x_n = [(n - 1)/n, 1]$ . The range of  $g$  is divided into  $m$  consecutive bins  $y_j$  such that  $y_j = [(j - 1)/m, j/m)$  for  $1 \leq j < m$ , and  $y_m = [(m - 1)/m, 1]$ . Then each bin  $x_i$  in the domain is mapped to a bin  $y_j$  in the range:  $G : x_i \rightarrow y_j$ , where  $G$  represents the discretized mapping. Formally, any  $x \in x_i$  is mapped to  $g(x) \equiv \overline{G(x_i)}$ , where  $\overline{G(x_i)}$  is the median value of the bin  $G(x_i)$ . For example, if  $n = 100$ ,  $m = 10$ , and  $y_7 = G(x_{30})$ , that is, the bin  $x_{30} = [0.29, 0.30)$  is mapped to the bin  $y_7 = [0.6, 0.7)$ , then for any  $x \in x_{30}$  (e.g.,  $x = 0.292$ ), the function  $g(x)$  would return  $0.65 = \overline{y_7}$ .

Therefore, in the GA, each function  $g$  can be encoded as an array of  $n$  integers, ranging from 1 to  $m$ , so that the  $i$ -th element of the array (the  $i$ -th digit) represents the mapping  $y_j = G(x_i)$ , where  $1 \leq j \leq m$ .

We have chosen a *generation gap* replacement strategy. In our experiments, we set the generation gap parameter 0.3. In other words, the entire old population is sorted according to fitness, and we choose the best 30% for direct replication in the next generation, employing an elitist selection mechanism. The rest of selection functionality is moved into the (uniform) crossover. Mutation is implemented as additive creeping or random mutation, depending on the number of “digits” in the genome. If the number of digits is greater than 10, then additive creeping is used: a digit can be mutated within  $[-5\%, +5\%]$  of its current value. If the number of digits is less than 10, the random mutation is used with the mutation rate of 0.01.

## VIII. ACKNOWLEDGMENTS

The authors are grateful to Joseph Lizier for open and motivating discussions; and to Mahendra Piraveenan for his exceptionally valuable prior contribution to this effort.

## References

- Adami, C. (1998). *Introduction to Artificial Life*. Springer, New York.
- Barbieri, M. (2003). *The organic codes: an introduction to semantic biology*. Cambridge University Press, Cambridge, United Kingdom.
- Bennett, C. H. (1990). How to define complexity in physics, and why. In (Zurek, 1990), pages 137–148.
- Bialek, W., de Ruyter van Steveninck, R. R., and Tishby, N. (2006). Efficient representation as a design principle for neural coding and computation. In *2006 IEEE International Symposium on Information Theory*, pages 659–663. IEEE.
- Crutchfield, J. P. (1990). Information and its Metric. In Lam, L. and Morris, H. C., editors, *Nonlinear Structures in Physical Systems – Pattern Formation, Chaos and Waves*, pages 119–130. Springer Verlag.
- Eigen, M. (1971). Self-organization of matter and evolution of biological macromolecules. *Naturwissenschaften*, 58(10):465–523.
- Haken, H. (1983). *Synergetics, an Introduction: Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry, and Biology*. Springer-Verlag, 3rd ed., New York.
- Kauffman, S. A. (2000). *Investigations*. Oxford University Press, Oxford.
- Klyubin, A., Polani, D., and Nehaniv, C. (2007). Representations of space and time in the maximization of information flow in the perception-action loop. *Neural Computation*, 19(9):2387–2432.
- Laughlin, S. B., Anderson, J. C., Carroll, D. C., and de Ruyter van Steveninck, R. R. (2000). Coding efficiency and the metabolic cost of sensory and neural information. In Baddeley, R., Hancock, P., and Földiák, P., editors, *Information Theory and the Brain*, pages 41–61. Cambridge University Press.
- Laughlin, S. B., de Ruyter van Steveninck, R. R., and Anderson, J. C. (1998). The metabolic cost of neural information. *Nature Neuroscience*, 1(1):36–41.
- Lizier, J. T., Prokopenko, M., and Zomaya, A. Y. (2008). Local information transfer as a spatiotemporal filter for complex systems. *Physical Review E*, 77(2):026110.
- Lloyd, A. (1995). The coupled logistic map: A simple model for the effects of spatial heterogeneity on population dynamics. *J. Theor. Biol.*, 173:217–230.
- Lloyd, S. (1990). Valuable information. In (Zurek, 1990), pages 193–197.
- MacKay, D. J. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, Cambridge.
- Piraveenan, M., Polani, D., and Prokopenko, M. (2007a). Emergence of genetic coding: an information-theoretic model. In Almeida e Costa, F., Rocha, L., Costa, E., Harvey, I., and Coutinho, A., editors, *Advances in Artificial Life: 9th European Conference on Artificial Life (ECAL-2007), Lisbon, Portugal, September 10-14*, volume 4648 of *Lecture Notes in Artificial Intelligence*, pages 42–52. Springer.
- Piraveenan, M., Prokopenko, M., and Zomaya, A. (2007b). Information-cloning of scale-free networks. In e Costa, F. A., Rocha, L. M., Costa, E., and I. Harvey, A. C., editors, *Advances in Artificial Life: 9th European Conference on Artificial Life (ECAL-2007), Lisbon, Portugal*, volume 4648 of *Lecture Notes in Artificial Intelligence*. Springer.
- Piraveenan, M., Prokopenko, M., and Zomaya, A. Y. (2009). Assortativeness and information in scale-free networks. *European Physical Journal B*, 67:291Ú300.
- Polani, D., M. Prokopenko, M., and Chadwick, M. (2008). Modelling stigmergic gene transfer. In Bullock, S., Noble, J., Watson, R., and Bedau, M. A., editors, *Artificial Life XI - Proceedings of the Eleventh International Conference on the Simulation and Synthesis of Living Systems*, pages 490–497. MIT Press.
- Polani, D., Nehaniv, C., Martinetz, T., and Kim, J. T. (2006). Relevant information in optimized persistence vs. progeny strategies. In Rocha, L., Yaeger, L., Bedau, M., Floreano, D., Goldstone, R., and Vespignani, A., editors, *Artificial Life X: Proceedings of The 10th International Conference on the Simulation and Synthesis of Living Systems*, Bloomington IN, USA.
- Prigogine, I. (1980). *From Being to Becoming: Time and Complexity in the Physical Sciences*. W. H. Freeman & Co., San Francisco.
- Prokopenko, M., Boschetti, F., and Ryan, A. (2008). An information-theoretic primer on complexity, self-organisation and emergence. *Complexity*, (DOI: 10.1002/cplx.20249).
- Prokopenko, M., Gerasimov, V., and Tanev, I. (2006). Evolving spatiotemporal coordination in a modular robotic system. In Nolfi, S., Baldassarre, G., Calabretta, R., Hallam, J., Marocco, D., Meyer, J.-A., and Parisi, D., editors, *From Animals to Animats 9: 9th International Conference on the Simulation of Adaptive Behavior (SAB 2006)*, volume 4095 of *Lecture Notes in Computer Science*, pages 558–569. Springer.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, July, October.
- Vetsigian, K., Woese, C., and Goldenfeld, N. (2006). Collective evolution and the genetic code. *PNAS*, 103(28):10696–10701.
- Wagner, A. (2005). *Robustness and Evolvability in Living Systems*. Princeton University Press, Princeton, NJ.
- Woese, C. R. (1965). On the evolution of the genetic code. *Proc. Natl. Acad. Sci. USA*, 54:1546–1552.
- Woese, C. R. (1972). The emergence of genetic organization. In Ponnampertuma, C., editor, *Exobiology*, page 301Ú341. North-Holland Publishing Co., Amsterdam, The Netherlands.
- Woese, C. R. (1998). The universal ancestor. *Proc. Natl. Acad. Sci. USA*, 95:6854–Ú6859.
- Woese, C. R. (2004). A new biology for a new century. *Microbiology*

*and Molecular Biology Reviews*, 68(2):173–186.  
Woese, C. R. and Fox, G. E. (1977). The concept of cellular evolution. *J. Mol. Evol.*, 10:1–6.  
Zurek, W. H., editor (1990). *Complexity, Entropy and the Physics*

*of Information*, Santa Fe Studies in the Sciences of Complexity, Reading, Mass. Addison-Wesley.