

Emergence of Genetic Coding: an Information-theoretic Model

Mahendra Piraveenan¹, Daniel Polani² and Mikhail Prokopenko¹

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

²Department of Computer Science, University of Hertfordshire
Hatfield AL10 9AB, United Kingdom
Corresponding author: mikhail.prokopenko@csiro.au

Abstract. This paper¹ introduces a simple 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. The model evolves a dynamical system, where a conglomerate of primitive cells is coupled with its potential encoding, subjected to specific environmental noise and inaccurate internal processing. The separation between the conglomerate and the encoding is shown to become beneficial in terms of preserving the information within the noisy environment. This selection pressure is captured information-theoretically, as an increase in mutual information shared by the conglomerate across time. The emergence of structure and useful separation inside the coupled system is accompanied by self-organization of internal processing, i.e. an increase in complexity within the evolving system.

1 Introduction

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 [1]. 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 prokaryotic cells, susceptibility of aminoacyl-tRNA synthetases to horizontal gene transfer (HGT), etc. The simple assumptions that we make, following Woese [1], include the notion of primitive cells as loosely connected conglomerates existing during the “era of nucleic acid life” [2, 3],

¹ The Authors list is in alphabetical order.

and the conjecture that primitive cell organization was “largely horizontal” in nature [4, 5], 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” [1]. 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 [1] 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” [1].

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” [1, 6]. 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 [1, 6]. These observations can be formalised information-theoretically. More precisely, we intend to consider a communication channel between a conglomerate of primitive cells and itself at a future time point, and pose a question of the channel capacity constrained by the noise. By varying the nature and degree of this noise prevalent in the environment within which such conglomerates exist and evolve, we hope to identify conditions leading to a separation between the conglomerate *per se* and its encoding with “proto-symbols”. Specifically, we investigate conditions under which such separation is beneficial in terms of preserving the information within the noisy communication channel across time. In other words, the separation evolves to protect some information about the conglomerate in the encoding. A rudimentary translation then helps to recover the information that otherwise would have been lost due to the noise.

The adopted information-theoretic view allows us to concentrate on generic processes common to a collection of such conglomerates rather than on specific interactions within an environmental locality. Considering a collection of conglomerates is important because, as noted by Woese [1],

A sufficiently imprecise translation mechanism could produce “statistical proteins”, proteins whose sequences are only approximate translations of their respective genes [7]. 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” [3]. Looseness of the outcome is implied by an imprecise genome replication

comprising relatively few unique genes [5] — therefore, rather than trying to develop a dynamical system (conglomerate 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”.

In modeling the evolution, we adopt the view that maximization of information transfer through selected channels is one of the main evolutionary pressures [8–12]. Although the evolutionary process involves a larger number of drives and constraints, information fidelity (i.e. preservation) is a consistent motif throughout biology: e.g., modern evolution operates close to the error threshold [13], and biological sensorimotor equipment typically exhausts the available informative capacity (under given constraints) close to the limit [14]. Adami, in fact, argues that the evolutionary process extracts valuable information and stores it in the genes. Since this process is relatively slow [15], it is a selective advantage to preserve this information, once captured. In the following, we shall concentrate on the information preservation property of evolution. 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.

2 Modelling evolutionary dynamics

Our generic model for evolutionary dynamics involves a dynamical coupled system, where a conglomerate 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 conglomerate is represented as a dynamical system, the information about it can be captured generically via 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.

The dynamical coupled system is described by the equations

$$X_t = \begin{cases} f(X_{t-1}) + \varphi & \text{if } t \neq t^* \\ \alpha [f(X_{t-1}) + \varphi] + (1 - \alpha)h(Y_{t-1}) & \text{if } t = t^* \end{cases} \quad (1)$$

$$Y_t = \begin{cases} g(X_{t_0} + \psi) & \text{if } t = t_0 \\ Y_{t-1} & \text{if } t > t_0 \end{cases} \quad (2)$$

where X_t and Y_t are the variables that describe the conglomerate and its potential encoding respectively. Function f defines the dynamical system representing the conglomerate, and $\alpha \in [0, 1]$ sets the relative importance of translation h ; we used $\alpha = 1/2$. Noise functions φ and ψ are described in Section 2.1. The mechanism that extracts information from the conglomerate into its encoding at time t_0 is given by g .

At time $t = t_0$, noise is introduced into the environment affecting dynamics of the conglomerate. Our null hypothesis is that the ability to symbolically encode nucleic acid sequences does not develop when (possibly anisotropic) environmental noise φ is outside a certain range. In other words, it is precisely a limited reduction in the information channel’s capacity, brought about by the environmental noise, that creates the

selection pressure for the separation between a conglomerate and its encoding. At the time $t = t_0$, information from the conglomerate is accessed by the system Y (encoding) via the internal processing function g . The accessed information is not entirely accurate because of possible limitations of the access mechanism, and the noise function ψ represents this inaccuracy.

We aim to identify a function g that maximizes the ability to recover, in the face of environmental noise and in presence of an imperfect internal processing, as much original information as possible — i.e., the ability to resist both the external noise φ and internal noise ψ . The feedback from the Y to X occurs at the time point t^* , i.e. the function h translates the input Y_{t^*-1} from the encoding back into the conglomerate.

2.1 External and internal noise

The function φ describes the external (environment) noise that affects the variable X after some information has been stored in Y . This noise represents a pressure to push the system X towards certain attractors. It is implemented as a random variable $\varphi \in [-l, u]$, 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 . In other words, positive values may be more sparsely distributed than the negative if u is larger than l .

The function ψ represents the internal noise associated with accessing information from X_t by the system Y_t at any given time t . In other words, it represents the inaccuracy within the internal communication channel. In addition, the noise ψ may be interpreted as inaccuracy of the environment's representation within the encoding Y which indirectly “perceives” the environment through the system X . This noise is modelled as uniform random noise $\psi \in [-b, b]$, where $0 < b \ll 1.0$.

2.2 State-space

The dynamical system employed is a logistic map $X_{t+1} = rX_t(1 - X_t)$, where r is a parameter, i.e. the function f is given by $f(x) = rx(1 - x)$. 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 attractors of the logistic map (approximately $0.38, 0.50, 0.83, 0.87$). The time $t = t_0$ is set after the logistic map settles into its attractors, having passed through a transient. The function g is a surjective mapping from $[0, 1]$ to $[0, 1]$. The function h is identity $h(y) = y$.

In order to estimate the probability distribution of a random variable (X or Y) at each time-point, we generate an initial random sample $(X_{t_0}) = (X_{t_0}^1, X_{t_0}^2, \dots, X_{t_0}^K)$ of length K . Each $X_{t_0}^i$, 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 $t_0 \leq t \leq T$, and T is a time horizon. Within the ensemble, each X_t^i time series may have a different initial value $X_{t_0}^i$. At any given time t' , we can obtain a sample $(X_{t'}) = (X_{t'}^1, X_{t'}^2, \dots, X_{t'}^K)$.

Given the initial sample (X_{t_0}) , and the mapping $Y_{t_0} = g(X_{t_0} + \psi)$, we can generate the initial 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 initial sample.

3 Genetic Algorithm

In evolving the potential encoding system Y coupled with X via a suitable function g , we maximize the mutual information between the initial X_{t_0} and recovered X_{t^*} states of the system, by employing a simple genetic algorithm (GA). The mutual information between A and B is defined as $I(A; B) = \sum_{a \in A} \sum_{b \in B} P(a, b) \log \frac{P(a, b)}{P(a)P(b)}$, where $P(a)$ is the probability that A is in the state a , and $P(a, b)$ is the joint probability.

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 determined by the mutual information between X_{t_0} and X_{t^*} , denoted $I_g(X_{t_0}; X_{t^*})$ and estimated via the mutual information between samples (X_{t_0}) and (X_{t^*}) .

Since the information from Y_{t^*-1} (different for each individual) is fed back into X_{t^*} , equation (1), the sample (X_{t^*}) is specific for each individual within the population. Therefore, it may be contrasted with the sample (X_{t_0}) which is identical across the population, producing distinct fitness values $I_g(X_{t_0}; X_{t^*})$ for each individual g . The experiments were repeated for different ensembles X_t .

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.

4 Results

Our main goal is to verify whether, as the evolution takes place in a noisy environment, some structure emerges in Y (e.g., attractors are observed in Y -space). The emergent

structure in Y can be associated with ‘‘proto-symbols’’ (‘‘codes’’) that help in retrieving at time t^* some (or most of the) information accessed at t_0 . This phenomenon must be supported by self-organization within the processing function g between t_0 and t^* .

4.1 Emergence of structure in the encoding

We begin by analyzing the optimization constraint — the system X . Figure 1 (left) shows the ensemble $[X]$ at the time $t^* - 1$, i.e. right before the moment when the feedback from Y to X occurs. It can be observed that the environment noise φ ($u = 0.025$ and $l = 0.025$)² disrupts the logistic map dynamics, and some information about the attractors of X is lost in the course of time: the observed sample (X_{t^*-1}) does not contain four clear clusters.

A random processing function g results in an encoding Y that, despite some structure (not shown), is not able to ‘‘help’’ X in recovering the lost information. Figure 1 (right) shows the un-recovered ensemble $[X]$ at the time t^* that contains values that are much more diverse than the four attractors of the original logistic map. As a result, the mutual information $I_g(X_{t_0}; X_{t^*})$ is low (≈ 0.7 bits).

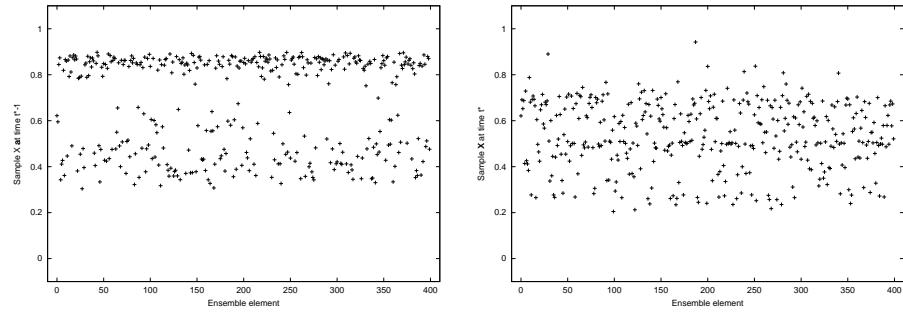


Fig. 1. Random g (noise $\varphi \pm 0.025$; $\psi = \pm 0.015$). Left: two remaining ‘‘clusters’’ in the sample (X_{t^*-1}). Right: the sample (X_{t^*}) does not recover full information about four attractors.

Let us evaluate now the evolved coupled system. Figure 2 shows the encoding ensemble $[Y]$ at the time $t^* - 1$, as well as the recovered ensemble $[X]$ at the time t^* . The sample (Y_{t^*-1}) settles into four clusters that can be easily represented by four ‘‘codes’’ corresponding to the four attractors of X . This emergent encoding allows to recover the information within X , as evidenced by four clear clusters within the sample (X_{t^*}). The corresponding increase in the information-based fitness function is shown in Figure 3.

The clustering corresponds to emergence of discrete ‘‘proto-symbols’’ in the encoding Y . In the simplest case, each non-empty bin in the range of g may be associated with such a cluster, creating a symbol in the encoding Y . Following this, the function h may reconstruct precise information about X_{t_0} by clustering X_{t^*-1} into the clusters encoded by Y_{t^*-1} , and then simply using the inverse mapping g^{-1} to retrieve the original attractor. Without the clustering, however, the information reconstructed at time

² Similar results were obtained with anisotropic noise.

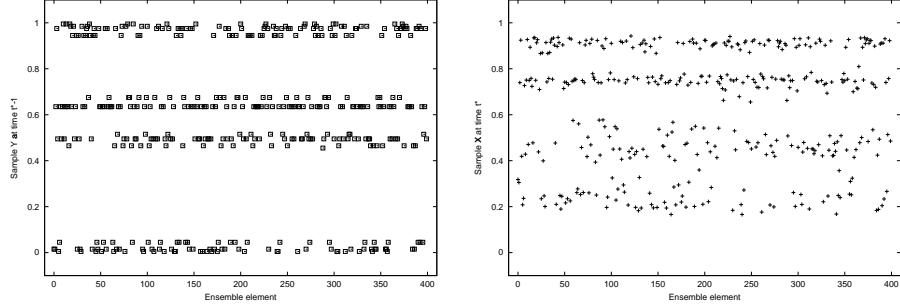


Fig. 2. Evolved g (noise $\varphi = \pm 0.025$; $\psi = \pm 0.015$). Left: four clusters in the encoding (Y_{t^*-1}). Right: Four recovered clusters in sample (X_{t^*}). $I(X_{t_0}; X_{t^*}) \approx 1.4$ bits. Contrast with Figure 1.

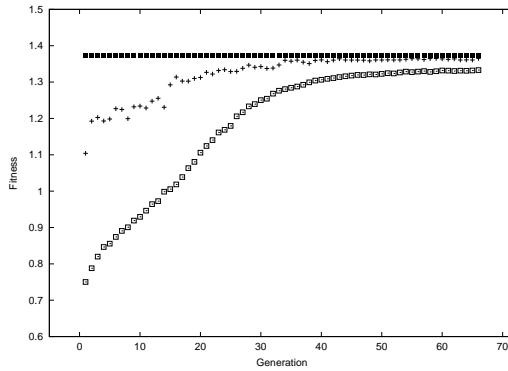


Fig. 3. Mutual information (fitness) in bits. Squares indicate the average fitness, '+'s show fitness of the best individual in each generation. Black squares trace possible maximum fitness: the mutual information when (X_{t^*}) is as informative as (X_{t_0}) , i.e., $I(X_{t_0}; X_{t^*}) = H(X_{t_0})$ (entropy).

t^* is not precise, and rather than having four crisp attractors, X can be described as an individual with an imprecise translation of the underlying gene within a “consensus sequence” [1], analogous to a “statistical protein”.

This experiment demonstrated that noise within the environment affects the emergence of structure in the encoding, allowing to recover the information within X . The null hypothesis, however, is that the coding threshold is not approached when environmental noise φ is outside a certain range. To verify the null hypothesis let us consider an evolution of the coupled system with a) larger noise φ ; and b) without noise φ .

Figure 4 shows the ensemble $[X]$ at the time $t^* - 1$, as well as the un-recovered ensemble $[X]$ the time t^* . These results indicate that in this case there is too much noise in the environment to make the encoding useful.

In a noise-free environment ($\varphi = 0.0$), we expect a lesser or no pressure on the coupled system. In fact, there are four unobscured clusters in the sample (X_{t^*-1}), as shown in Figure 5 (left). However, the evolved encoding, being still affected by internal processing noise ψ , loses some of this information and hinders the recovery at t^* . Figure 5 (right) shows results of translation of ill-structured encoding: with zero environmental noise there is no pressure for emergence of a more precise structure in the encoding.

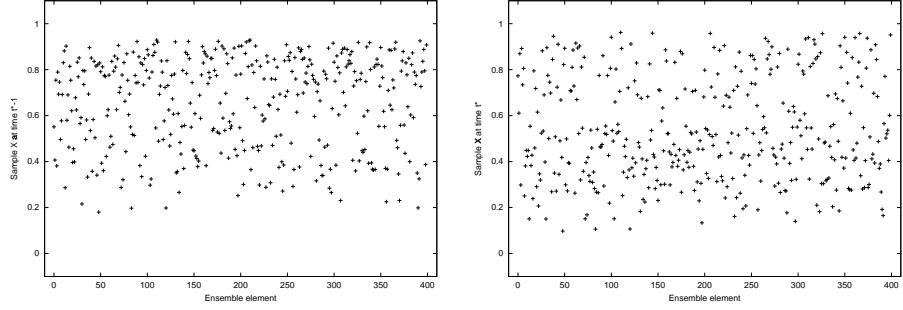


Fig. 4. Evolved g ($\varphi = \pm 0.06$; $\psi = \pm 0.015$). Left: the sample (X_{t^*-1}) . Right: the sample (X_{t^*}) does not recover the information about four attractors. Mutual information $I(X_{t_0}; X_{t^*}) \approx 1.1$ bits. Contrast with Figure 2.

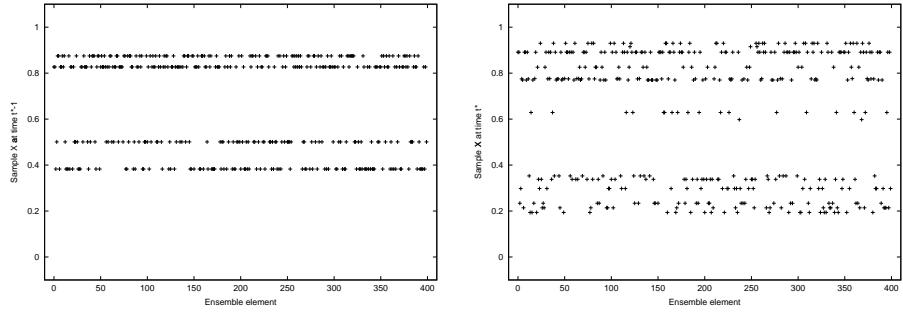


Fig. 5. Evolved g ($\varphi = 0.0$; $\psi = \pm 0.015$). Left: the sample (X_{t^*-1}) . Right: the sample (X_{t^*}) obscures the four attractors. Mutual information $I(X_{t_0}; X_{t^*}) \approx 1.4$ bits. Contrast with Figure 2.

4.2 Self-organization within the processing function

At this stage, we analyse self-organization within the processing function g . Figure 6 contrasts a randomly selected function g at the start of the evolution (noise φ is defined as $\varphi = 0.025$, while noise $\psi = \pm 0.03$ is increased to better illustrate the phenomenon), with the best individual function g after 100 generations. The important difference is in the way of mapping attractors of X (and their neighbourhoods) into the encoding Y . Let us consider, for example, the ψ -neighbourhood of the attractor $x \approx 0.50$. Figure 6 (left) shows that this neighbourhood has values between 0.07 and 0.91. This wide distribution is explained by the internal processing noise ψ and absence of any organization in g , so that even a slight shift from x to $x \pm \psi$ may result in a large difference between $g(x)$ and $g(x \pm \psi)$. Figure 6 (right) shows the best individual function g evolved after 500 generations. Here, despite the same level of noise ψ , we can observe that the ψ -neighbourhood of a given attractor in X (e.g. the neighbourhood of $x \approx 0.50$) is mapped by the evolved function g to a smaller interval (e.g., it has values between 0.93 and 0.96). In other words, the selection pressure resulted in a more “condensed” mapping. This is achieved by an increase in organization (stability) in g : a small shift from x to $x \pm \psi$ results now only in a small difference between $g(x)$ and $g(x \pm \psi)$. Moreover, only the attractors’ neighbourhoods are generally handled in this way — no two other similar values of the domain are mapped to points close to each other in the range.

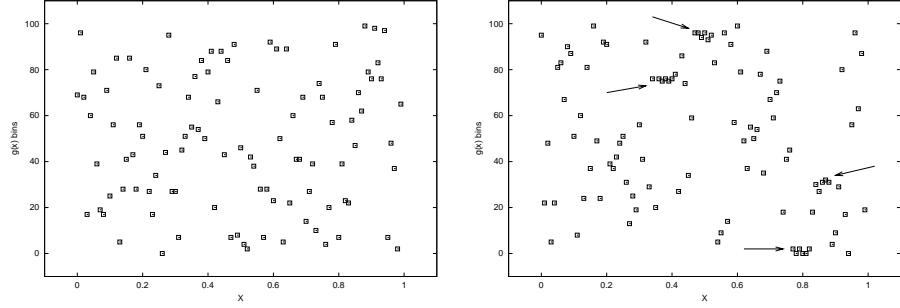


Fig. 6. Left: a random function g at the start of evolution, shown as a sample (Y_{t_0}). No structure is observed. Right: evolved g , shown as a sample (Y_{t_0}). Arrows point to regions of stability.

The self-organization of g counters the effect of internal processing noise ψ , given the noise φ in the environment³. This self-organization helps Y to maintain the structure of the space X (namely, the information that it had a certain number of attractors). The function h translates this encoding back into X .

5 Discussion and Conclusions

We considered a model for evolutionary dynamics in the vicinity of the “coding threshold”, and identified conditions under which a separation between a conglomerate of primitive cells and its symbolic encoding becomes beneficial in terms of preserving the information within a noisy environment. The model evolves a dynamical system, where a conglomerate is coupled with its potential encoding, affected by environmental noise and inaccurate internal processing. The experiments supported the hypothesis that the ability to symbolically encode nucleic acid sequences does not develop in the absence of environmental noise or when the noise is too high (reminiscent of the U-shaped curve relating adaptive pressure to perception accuracy [18]). Instead, a (possibly anisotropic) limited noise constraining the channel’s capacity creates the selection pressure for the separation between the conglomerate and its encoding. Formally, we captured this selection pressure as an increase in mutual information shared by the conglomerate across time. The experiments provided evidence that the emergence of structure and useful separation inside the coupled system is accompanied by self-organization of internal processing, i.e. an increase in complexity within the evolving system.

Following this path suggests a few intriguing possibilities for coupling systems in different ways, e.g. considering systems where a) the encoding evolves in parallel with the conglomerate, being subjected to different noise; b) the encoding may be used by other co-evolving conglomerates; c) two coupled systems similar to the one analyzed in this work exchange the respective encodings. These directions are under investigation, aiming at understanding of emergence and universality of genetic code. Woese observed that “statistical proteins form the basis of a powerful strategy for searching protein phase space, finding novel proteins” [1]. We believe that further modelling of the evolutionary dynamics in such a space may explain mechanisms resolving Eigen’s paradox [19] and leading to convergence on “the lingua franca of genetic commerce” [1].

³ In general, following [16, 17], we may say that self-organization results from fluctuations, that is, internal information processing has self-organized in response to environmental “pollution”.

References

1. Woese, C.R.: A new biology for a new century. *Microbiology and Molecular Biology Reviews* **68**(2) (2004) 173–186
2. Woese, C.R.: The emergence of genetic organization. In Ponnamperuma, C., ed.: *Exobiology*. North-Holland Publishing Co., Amsterdam, The Netherlands (1972) 301341
3. Vetsigian, K., Woese, C.R., Goldenfeld, N.: Collective evolution of the genetic code. *Proc. Natl. Acad. Sci. USA* **103** (2006) 10696–10701
4. Woese, C.R.: The universal ancestor. *Proc. Natl. Acad. Sci. USA* **95** (1998) 6854–6859
5. Woese, C.R., Fox, G.E.: The concept of cellular evolution. *J. Mol. Evol.* **10** (1977) 1–6
6. Barbieri, M.: *The organic codes: an introduction to semantic biology*. Cambridge University Press, Cambridge, United Kingdom (2003)
7. Woese, C.R.: On the evolution of the genetic code. *Proc. Natl. Acad. Sci. USA* **54** (1965) 1546–1552
8. Klyubin, A.S., Polani, D., Nehaniv, C.L.: All else being equal be empowered. In Capcarrère, M., Freitas, A., Bentley, P., Johnson, C., Timmis, J., eds.: *Advances in Artificial Life*, 8th European Conference, ECAL, 2005, Canterbury, UK, September 5-9, 2005, Proceedings. Volume 3630 of Lecture Notes in Computer Science. Springer (2005) 744–753
9. Prokopenko, M., Wang, P., Valencia, P., Price, D., Foreman, M., Farmer, A.: Self-organizing hierarchies in sensor and communication networks. *Artificial Life* **11** (2005) 407–426
10. Prokopenko, M., Gerasimov, V., Tanev, I.: Measuring spatiotemporal coordination in a modular robotic system. In Rocha, L., Yaeger, L., Bedau, M., Floreano, D., Goldstone, R., Vespignani, A., eds.: *Artificial Life X: Proceedings of The 10th International Conference on the Simulation and Synthesis of Living Systems*, Bloomington IN, USA (2006) 185–191
11. Prokopenko, M., Gerasimov, V., Tanev, I.: Evolving spatiotemporal coordination in a modular robotic system. In Nolfi, S., Baldassarre, G., Calabretta, R., Hallam, J., Marocco, D., Meyer, J.A., Parisi, D., eds.: *From Animals to Animats 9: 9th International Conference on the Simulation of Adaptive Behavior (SAB 2006)*, Rome, Italy, September 25-29 2006. Volume 4095 of Lecture Notes in Computer Science., Springer Verlag (2006) 558–569
12. Klyubin, A.S., Polani, D., Nehaniv, C.L.: Representations of space and time in the maximization of information flow in the perception-action loop. *Neural Computation* (2007) accepted 17 October 2006, in press
13. Adami, C.: *Introduction to Artificial Life*. Springer (1998)
14. Laughlin, S.B., de Ruyter van Steveninck, R.R., Anderson, J.C.: The metabolic cost of neural information. *Nature Neuroscience* **1**(1) (1998) 36–41
15. Zurek, W.H., ed.: *Valuable Information*. In Zurek, W.H., ed.: *Complexity, Entropy and the Physics of Information*. Santa Fe Studies in the Sciences of Complexity, Reading, Mass., Addison-Wesley (1990)
16. Prigogine, I.: *From Being to Becoming: Time and Complexity in the Physical Sciences*. W. H. Freeman & Co., San Francisco (1980)
17. Haken, H.: *Synergetics, an Introduction: Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry, and Biology*. Springer-Verlag, 3rd ed., New York (1983)
18. Todd, P.M., Miller, G.F.: Exploring adaptive agency ii: Simulating the evolution of associative learning. In Meyer, J.A., Wilson, S.W., eds.: *From animals to animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior*, Cambridge, MA: MIT Press/Bradford Books (1991) 306–315
19. Eigen, M.: Self-organization of matter and evolution of biological macromolecules. *Naturwissenschaften* **58**(10) (1971) 465–523