Modelling Stigmergic Gene Transfer

Daniel Polani[1], Mikhail Prokopenko[2] and Matthew Chadwick[2]

[1]Department of Computer Science, University of Hertfordshire
Hatfield AL10 9AB, United Kingdom
[2]CSIRO Information and Communication Technology Centre
Locked bag 17, North Ryde, NSW 1670, Australia
Corresponding author: mikhail.prokopenko@csiro.au

Abstract

We consider an information-theoretic model studying the conditions when a separation 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". We separate the phenomena into a "waste" and a "loss" component; the "waste" measures "packaging" information which envelops the proto-cell's information, but itself does not contain any information of interest, the "loss" measures how much of the proto-symbolically encoded information is actually lost. We observe that transitions in the waste/loss functions correspond to the boundaries of the "error interval". Secondly, we study whether and how different proto-cells can share such information via a joint code, even if they have slightly different individual dynamics. Implications for the emergence of biological genetic code are discussed.

Introduction

It can be argued that "the capacity to represent nucleic acid sequences symbolically in terms of a (colinear) amino acid sequence" (Woese, 2004) did not exist at the very early evolutionary stages, and developed only in response to certain environmental conditions. The phase of nucleic acid life that did not use genetic coding is separated from the later evolutionary stages where such coding became beneficial, by the "coding threshold". In this paper, we consider a model for evolutionary dynamics in the vicinity of the "coding threshold". The model is an extension of the model introduced by Piraveenan et al. (2007) who identified conditions under which a separation between a proto-cell and its symbolic encoding becomes beneficial in terms of preserving the information within a noisy environment.

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". 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.

In this paper, we adopt an information-theoretic view that allows us to concentrate on generic processes common to a collection of primitive cells rather than on specific biochemical interactions within an environmental locality. Moreover, it 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.

It is conjectured that maximization of information transfer through selected channels is one of the main evolutionary pressures (Prokopenko et al., 2006; Klyubin et al., 2007; Piraveenan et al., 2007; Laughlin et al., 2000; Bialek et al., 2007): although the evolutionary process involves a larger number of drives and constraints, information preservation is a consistent motif throughout biology. 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. (2007), and focus on the information preservation property of evolution within a coupled dynamical system. Piraveenan et al. (2007) verified that the ability to symbolically encode nucleic acid sequences does not develop when environmental noise φ 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 separation between a proto-cell and its encoding.

Here we extend the model of Piraveenan et al. (2007) by identifying both encoding and translation that maximize the ability to recover as much original information as possible in the face of environmental noise and in presence of an imperfect internal processing. In doing so, we enhance the analysis by considering both the loss and the waste of the information. Finally, we study effects of co-evolution of multiple encodings entrapped by multiple ensembles using SGT.

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.

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 it 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.

The generic dynamical coupled system is described by the equations

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

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

where $X_{t,m}$ are the variables that describe multiple proto-cells, $1 \le m \le M$, 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. 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

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

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

The function φ_t describes the external (environment) noise that affects the proto-cells: it is the same for all cells, i.e, φ_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). 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. 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^*.

The entrapment mechanism that matches 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 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 ψ. 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.

Piraveenan et al. (2007) considered the case $m = 1$, equations (3)–(4), and function h being the identity (a single system). Here we consider a system with multiple proto-cells: $m \ge 1$, and contrast universality of the translation machinery: all functions h_m are identical, while $g_i \ne g_j$ for $i \ne j$, with universality of the proto-code: all proto-codes g_m are identical, while $h_i \ne h_j$ for $i \ne j$. We would like to point out that the system (1)–(2) is coupled not only due to the common environment noise φ, but also due to the shared translation machinery h or shared proto-code g. This coupling supports a simple information-theoretic model of HGT and specifically, SGT. As we are dealing only with the information content, the consideration of identical h_m's and/or identical g_m's allows us to study gene transfers without details of molecular (state-to-state) interactions.

Coupled logistic maps

The dynamical system employed is a logistic map $X_{t+1} = r X_t \left(1 - X_t\right)$, where r is a parameter, i.e. the function f_m is given by $f(x) = r_m x \left(1 - x\right)$. 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 $r = 3.5$ (for the single system), resulting in four states of the attractor of the logistic map (approximately $0.38, 0.50, 0.83, 0.87$). For multiple proto-cells, we used proto-cells with $r = 3.5$ as well as with $r = 3.46$ and $r = 3.48$. Each of these possesses four states of the respective attractor. The time $t = t_0$ is set after the logistic map settles into its attractors, having passed through a transient. The functions g and h are mappings from $[0,1]$ to $[0,1]$.

Coupled logistic maps have been extensively used in modelling of biological processes. One prominent study is the investigation of spatial heterogeneity in population dynamics (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 (1)–(2) 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).

Information preservation

In evolving the potential encoding system Y coupled with X via a suitable function g, we minimize Crutchfield's information distance (Crutchfield, 1990) between the initial X_{t_0} and recovered X_{t^*} states of the system:

$$d(X_{t_0}, X_{t^*}) = H(X_{t_0}|X_{t^*}) + H(X_{t^*}|X_{t_0}) \qquad (5)$$

The entropies are defined as

$$H(A) = -\sum_{a \in A} P(a)\, \log P(a)\,, \qquad (6)$$

$$H(A, B) = -\sum_{a \in A}\sum_{b \in B} P(a,b)\, \log P(a,b)\,, \qquad (7)$$

$$H(A|B) = H(A, B) - H(B) \qquad (8)$$

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

The distance $d(X_{t_0}, X_{t^*})$ measures the dissimilarity of two information sources X_{t_0} and X_{t^*}; 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. (2007), the information metric d 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 over the mutual information, the former leads to a more crisp recovery of structure in the phase-space.

The use of d also indicates the presence of two components of dissimilarity. The first is the loss of information, $H(X_{t_0}|X_{t^*})$, which measures how much uncertainty the final state has about the original state of the system. The second is the waste, $H(X_{t^*}|X_{t_0})$; as the system will aim to preserve as much information as possible about the state X_{t_0} (and only this information), any additional variability in X_{t^*} will be considered as "waste".

Minimization of the information distance (more precisely, maximization of $-d(X_{t_0}, X_{t^*})$) is achieved by employing a simple genetic algorithm (GA) (described in the Appendix).

In order to estimate the probability distribution of a random variable (X or Y) at a given time, we generate an initial random sample $(X_0) = (X_0^1, X_0^2, \ldots, X_0^K)$ 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 $[X] = [X_t^1, X_t^2, \ldots, X_t^K]$, 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 $(X_{t'}) = (X_{t'}^1, X_{t'}^2, \ldots, X_{t'}^K)$.

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 $(Y_{t_0}) = (Y_{t_0}^1, Y_{t_0}^2, \ldots, Y_{t_0}^K)$ for the variable Y. In the corresponding ensemble $[Y] = [Y_t^1, Y_t^2, \ldots, Y_t^K]$ each sample is identical to the the sample (Y_{t_0}).

Recapitulation of the Results for a Single System

We begin by revisiting the simple case $m = 1$ that was considered by Piraveenan et al. (2007): the function h is identity $h(y) = y$. The structure evolving in Y can be associated with "proto-symbols" ("codes") that help to retrieve at time t^* some (or most of the) information stored at t_0.

Figure 1 shows the ensemble $[X]$ at the time $t^* - 1$, i.e. right before the moment when the feedback from Y to X occurs. The environment noise φ ($u = 0.025$ and $l = 0.025$) disrupts the logistic map dynamics, and some information about the attractor of X and its four states is lost in the
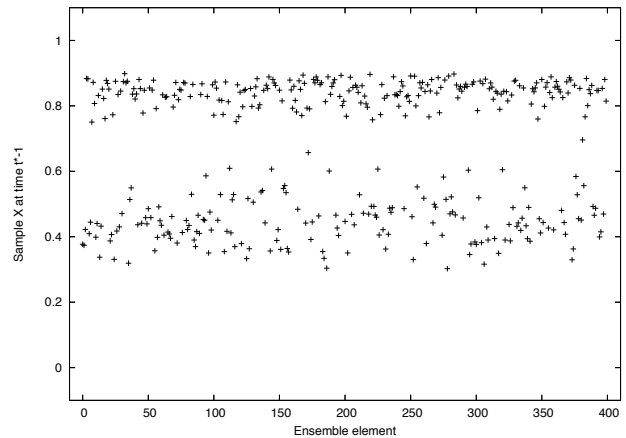

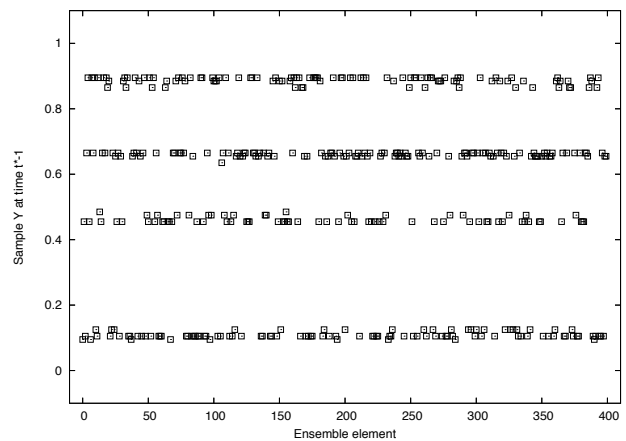Figure 1: Two remaining "clusters" in the sample (X_{t^*-1}).

Figure 2: Evolved g (noise $\varphi = \pm 0.025$; $\psi = \pm 0.015$) containing four clusters in the encoding (Y_{t^*-1}). Function h is identity.

course of time: the observed sample (X_{t^*-1}) does not contain four clear clusters.

Figure 2 shows the evolved encoding ensemble $[Y]$ at the time $t^* - 1$, while Figure 3 shows the recovered ensemble $[X]$ of the evolved coupled system 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 states of the attractor of X. The evolved encoding allows to recover the information within X, as evidenced by four clear clusters within the sample (X_{t^*}).

The clustering corresponds to the emergence of discrete "proto-symbols" in the encoding Y. The information reconstructed at time t^* is not 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". So far the recapitulation of the past results.
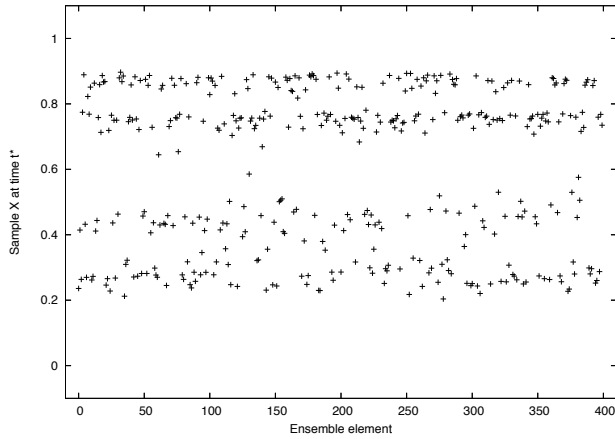
Figure 3: Four recovered clusters in sample (X_{t^*}). $d(X_{t_0}, X_{t^*}) \approx 1.5$ bits. Contrast with Figure 1.
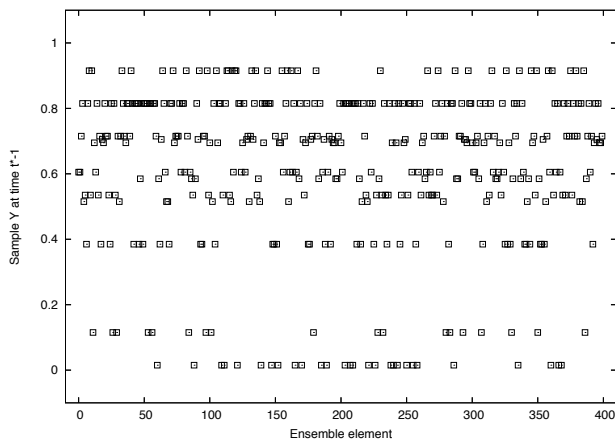

Figure 4: Evolved g (noise $\varphi = \pm 0.025$; $\psi = \pm 0.015$), with variable function h (see Figure 5).
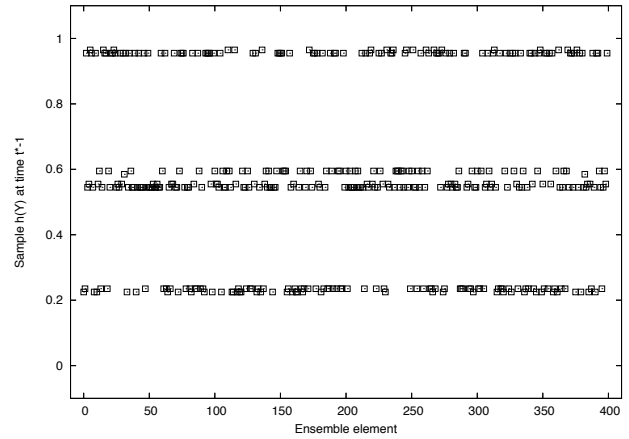

Figure 5: Evolved h (noise $\varphi = \pm 0.025$; $\psi = \pm 0.015$), complementing the encoding g (see Figure 4).

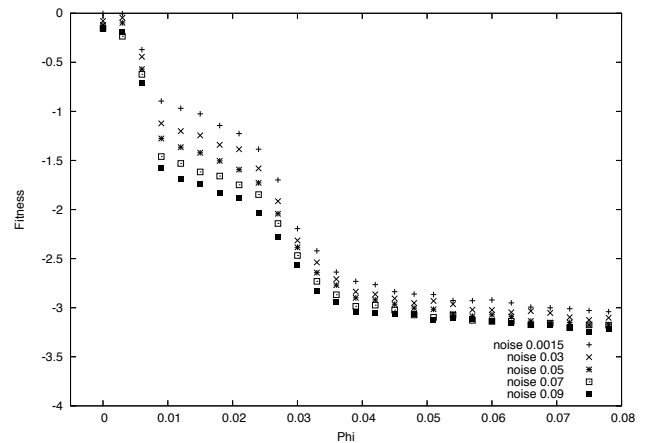Figure 6: Fitness, i.e. $-d(X_{t_0}, X_{t^*})$, over noise φ, for different noise levels ψ.

Optimizing the Recovery Function h

Now we consider the extended case where the translation function h is subject to optimization as well. This time, the evolved encoding ensemble $[Y]$ at the time $t^* - 1$ (Figure 4), does not have four clear clusters. However, this lack of adequate encoding is complemented by a more refined translation that evolved in parallel, as evidenced by Figure 5. The end result (not shown) is analogous to the one presented by Figure 3.

Figure 6 traces fitness, $-d(X_{t_0}, X_{t^*})$ (for the best individual), over the external noise φ, for different internal noise levels ψ. We can observe a steady decrease in fitness punctuated by two sharper transitions, that form three plateaus. As conjectured by Piraveenan et al. (2007), the encoding is not beneficial when the environmental noise φ is outside a certain range. The middle plateau is precisely the region specifying this range, i.e. the "error interval". It is also evident that within this plateau, sensitivity to internal noise ψ is the highest.

Be reminded that the information distance $d(X_{t_0}, X_{t^*})$ consists of two components: the loss $H(X_{t_0}|X_{t^*})$, and the waste $H(X_{t^*}|X_{t_0})$. 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 much of the proto-symbolically encoded information is actually lost. Figure 7 plots fitness over noise φ, for specific ψ, and shows loss and waste for the best individual. At the first plateau (very small noise), $d(X_{t_0}, X_{t^*}) = 0$, and both loss and waste are zero. At the medium plateau, the recovered system cannot get any closer to X, because the waste cannot be avoided, while the loss is still zero or minimal. At the last plateau ($\varphi > 0.025$), the loss begins to increase for the first time. So not only is there a waste, but the recovered system loses some information. The loss reaches 0.5, waste reaches 2.5, and $d(X_{t_0}, X_{t^*})$ reaches 3.0 (twice as large as the distance at the medium plateau). So the cascade of plateaus is explained by: (i) everything is recoverable (the first plateau); (ii) waste appears (the medium plateau); (iii) loss appears (the last plateau).
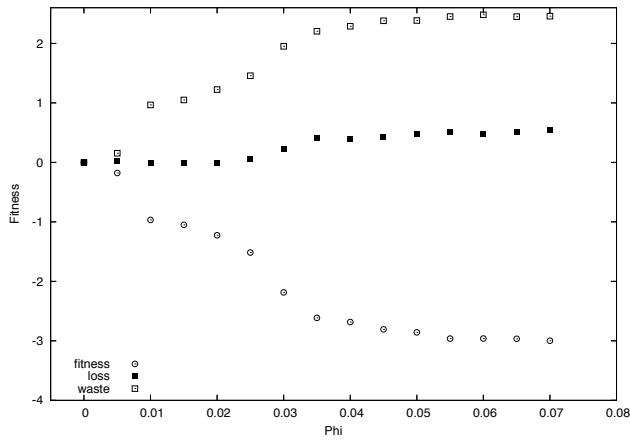
Figure 7: Fitness $-d(X_{t_0}, X_{t^*})$, loss $H(X_{t_0}|X_{t^*})$, and waste $H(X_{t^*}|X_{t_0})$, over noise φ, for specific $\psi = 0.015$.
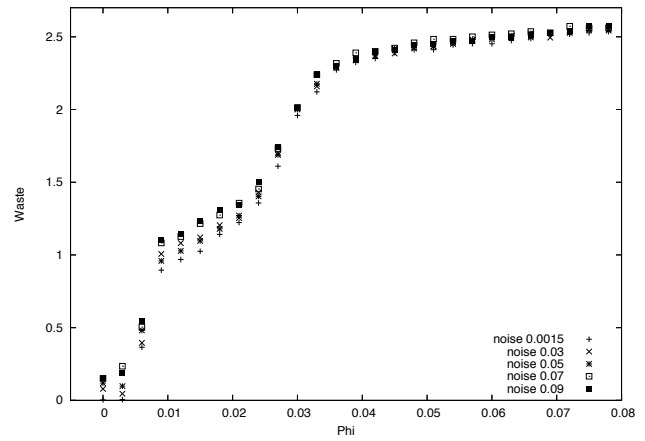

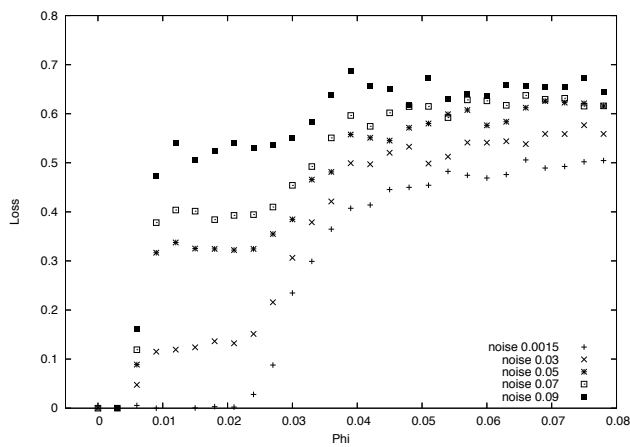Figure 9: Waste $H(X_{t^*}|X_{t_0})$ over noise φ, for different noise levels ψ.

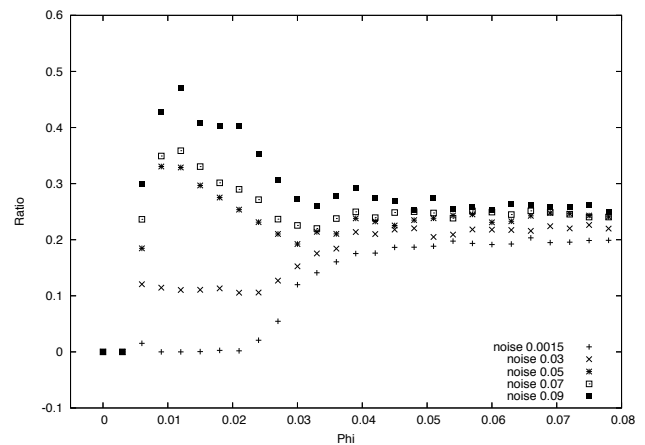Figure 8: Loss $H(X_{t_0}|X_{t^*})$ over noise φ, for different noise levels ψ.

Figure 10: Loss/waste ratio over noise φ, for different noise levels ψ.

Figures 8 and 9 "zoom" into the dynamics of loss and waste for different levels of internal noise ψ, and show that the loss also appears if the internal noise $\psi > 0.015$. It is also evident that the loss is more sensitive to internal noise than waste. The waste, on the other hand, simply follows the cascade of plateaus. The difference between loss and waste is highlighted in Figure 10 that traces the ratio $\frac{H(X_{t_0}|X_{t^*})}{H(X_{t^*}|X_{t_0})}$. This ratio is most turbulent at the medium plateau, supporting the hypothesis of its special role. We note that the transitions in the waste/loss functions correspond to the boundaries of the medium plateau, marking the "error interval".

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 $m = 3$ ($r = 3.5$, $r = 3.46$, and $r = 3.48$), and contrast the universality of the translation machinery with the universality of the proto-code.

Single g and multiple h
Let us assume that all available proto-codes g_m are identical (universal code), but $h_i \neq h_j$ for $i \neq j$. In this case, the system achieves recovery comparable with the single system for each of the logistic maps, $d(X_{t_0}, X_{t^*}) \approx 1.5$, but the structure of the code is slightly different. As shown in Figure 11, for the logistic map $r = 3.5$ fewer clusters evolve than for the singular system (shown in Figure 4). However, the translation machinery depicted in Figure 12 is as structured as that of the singular system (shown in Figure 5). This supports a conjecture that multiple systems exert some pressure for proto-code's universality.

Multiple g and single h
Here we consider the opposite case: abundance of available proto-codes: $g_i \neq g_j$ for $i \neq j$, but translation machinery is universal: all functions h_m are identical. Again, the system achieves the recovery of the singular system for each of the logistic maps, $d(X_{t_0}, X_{t^*}) \approx 1.5$, but the structure of both the code and translation machinery is more compact, as
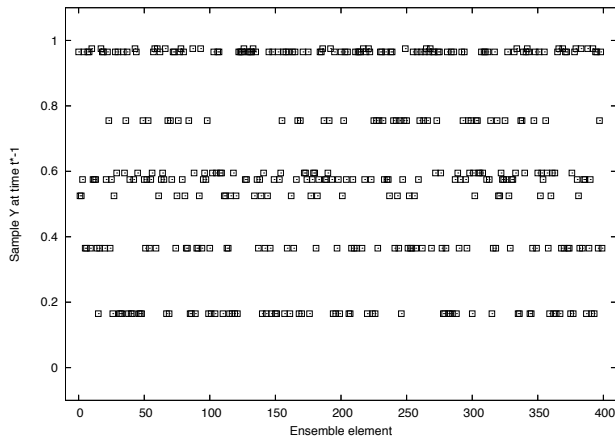
Figure 11: Single g. Evolved g (noise $\varphi = \pm 0.025$; $\psi = \pm 0.015$), for three ensembles with variable function h (see Figure 12). Shown for ensemble with $r = 3.5$.
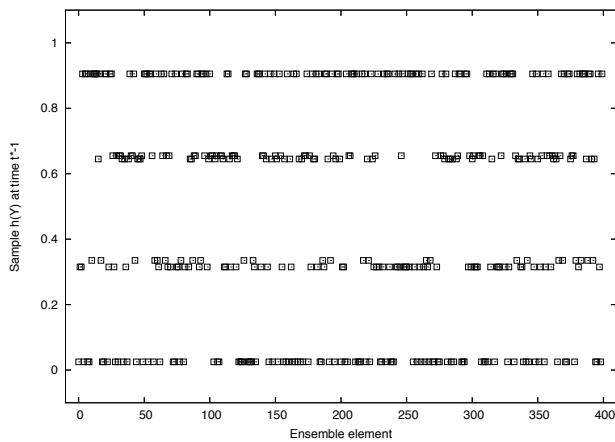

Figure 13: Single h. Evolved g (noise $\varphi = \pm 0.025$; $\psi = \pm 0.015$), for ensemble with $r = 3.5$, co-evolved with three ensembles; complementing the translation h (see Figure 14).

Figure 12: Single g. Evolved h (noise $\varphi = \pm 0.025$; $\psi = \pm 0.015$), for ensemble with $r = 3.5$, co-evolved with three ensembles; complementing the encoding g (see Figure 11).

Figure 14: Single h. Evolved h (noise $\varphi = \pm 0.025$; $\psi = \pm 0.015$), for three ensembles with variable function g (see Figure 13).

shown in Figures 13 and 14. This supports a conjecture that co-evolution of multiple systems may yield not only universality of proto-code, but also uniform translation machinery.

Conclusion and Future Work

We considered an information-theoretical model based on dynamical systems for the emergence of protected informational channels able to preserve information in a system over time when the main channel is suffering from perturbations. Doing so, we extended previous work, by not only introducing the optimization of a backtranslation mechanism, but also the consideration of the information metric and the more refined analysis able to resolve loss as well as waste in the resulting encoding. Furthermore we studied the effects on a small population of systems sharing an encoding.

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
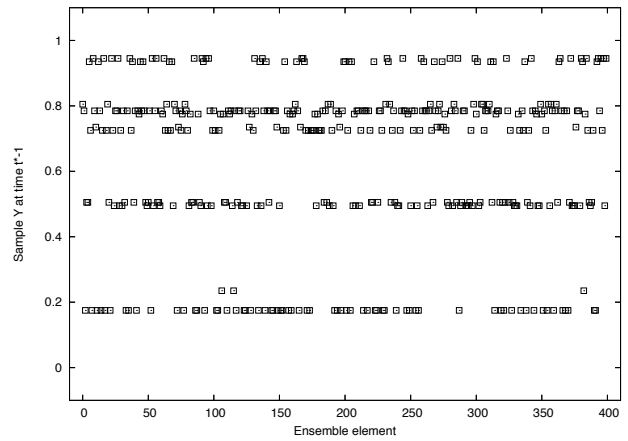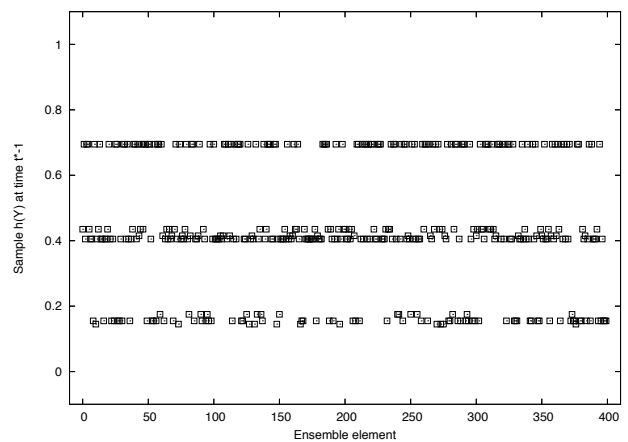
not too large.

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 emergence of distinct symbols turns out to be the most sensitive for the precise level of noise.

The waste/loss analysis shows that with increasing noise, first the waste grows away from 0, at first without any loss. Only at higher noise levels the loss begins its growth. These transitions correspond closely to the plateau transitions.

The multiple system scenario shows that joint translation "machineries" can be successfully used by several systems which differ slightly. However, at this point, we did not yet model the competition between different translation and information exchange models. This will be addressed in future work.

Appendix

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 (and function h). The ensemble $[X]$ is kept unchanged while we evolve the population of functions g (and h), being an optimization constraint, but the ensemble $[Y]$ differs for each individual within the population. The fitness of each function g (and h) is determined by the negative distance between X_{t_0} and X_{t^*}, denoted $d(X_{t_0}; X_{t^*})$, defined by equation (5), and estimated via the respective conditional entropies 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 (and h) 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 $1 \leq j \leq m$. Function h is coded analogously.

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.

References

Adami, C. (1998). *Introduction to Artificial Life*. Springer, New York.

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. (2007). Efficient representation as a design principle for neural coding and computation. In preparation.

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.

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.

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.

Piraveenan, M., Polani, D., and Prokopenko, M. (2007). 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.

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 Animats 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.

Vetsigian, K., Woese, C., and Goldenfeld, N. (2006). Collective evolution and the genetic code. *PNAS*, 103(28):10696–10701.

Woese, C. R. (1965). On the evolution of the genetic code. *Proc. Natl. Acad. Sci. USA*, 54:1546–1552.

Woese, C. R. (2004). A new biology for a new century. *Microbiology and Molecular Biology Reviews*, 68(2):173–186.

Zurek, W. H., editor (1990). *Complexity, Entropy and the Physics of Information*, Santa Fe Studies in the Sciences of Complexity, Reading, Mass. Addison-Wesley.