New Species in Evolving Networks - stochastic theory and applications on the metaphoric level

Werner Ebeling, Rainer Feistel, Ingrid Hartmann-Sonntag, Lutz Schimansky-Geier

aInstitute of Physics, Humboldt University Berlin, Newtonstraße 15, D-12489 Berlin, Germany, e-mail: ebeling@physik.hu-berlin.de

Andrea Scharnhorst

bNetherlands Institute for Scientific Information Services (NIWI), Royal Netherlands Academy of Arts and Sciences, Joan Muyskenweg 25, Postbus 95110, 1090 HC Amsterdam, The Netherlands, email: Andrea.Scharnhorst@niwi.knaw.nl

Abstract

In this paper we develop a theory to describe stochastic influences on the fate of new species with nonlinear growth rates in evolutionary processes. We develop a theoretical framework based on a metaphoric use of notions as species, network, innovation, competition, survival and fitness. We introduce a stochastic picture describing the role of fluctuations for the survival of new species in non-linear systems. In particular we consider the fate of new species with nonlinear growth. As an application of the general model framework we consider the fate of ‘rare species’ in early biological evolution. Thus we move from the metaphoric level of the use of certain notions to a concrete context of interpretation. In general, we show that hypercycle systems do not represent the end of the evolutionary process as they may evolve further in small niches. This has implications for different types of applications ranging from biological and - on the metaphoric level - to socio-technological systems.

Key words: evolving networks, master equations, survival probability of mutations, evolutionary chance of hypercycles, biological metaphors

1 Introduction

Ray Paton’s publication record encompasses a wide range of different topics, reaching from modelling information processes in cells (Fisher et al., 1999) to macroeconomic behaviour (Hirano and Paton, 2000). Most impressively and parallel to his input in quite different disciplinary areas he has reflected about the mechanisms through which different disciplines can talk to each other. He identified visual representations as...
one way of breaking down barriers between disciplines (Paton and Neilson 1999). Another topic reappearing in many of his writings was the role of metaphors. In his most cited article, according to the Web of Knowledge, he wrote "new discoveries and insights have been made by scientists thinking metaphorically." (Paton, 1992, p. 279). For him a metaphor was a "trope which provides the context for making comparisons and describing differences between objects, and by which analogies and similes can be made" (Paton 1992, p. 281).

As he outlined himself in the article, he approached metaphors from a ‘realist understanding of science’ with the aim “to deal with the complexities of the real world” by abstraction and simplifying complexity. This is an approach which characterizes the theories of self-organization. However, one does not often come across actors in the field of system sciences who, whilst modeling, also reflect upon what they are doing. Paton is quite exceptional in this respect. His reflections about metaphors often start at the boundary between biology and computing (Paton, 1994).

In our article we will follow his line of reasoning by introducing a theoretical framework which is based on notions of species, networks, competition and selection in a quite general or, in other words, a metaphorical way. Based on the theory of self-organization processes we pay particular attention to the special problem of stochastic effects connected with the appearance of 'new species' and, what we call, 'sensitive networks'. In this way, we introduce the notion of 'innovation' in a broad sense as disturbance and instability of the system. With Paton we would say that 'innovation' is used as a 'systemic metaphor' (Paton, 1992). We will concentrate in this paper on the construction of one possible context of the metaphor in the field of biosystems. In this case we will discuss how new species can emerge and survive in a biological system. However, we also would like highlight the fact that the family of metaphors (Hellsten 2002) around 'survival', 'selection', and 'innovation' can also be transferred to describe 'innovations' in socio-economical and in technical systems (Feistel and Ebeling, 1989; Ebeling et al., 1990; Hartmann-Sonntag et al., 2004). The study of stochastic effects connected with the appearance of a new species is highly relevant for recent developments in evolutionary theory (Schuster and Sigmund, 1984; Nowak and Schuster, 1989; Reidys et al., 2001). Kaneko and Yomo (2002) have shown recently that the origin of heredity in replicating systems is closely connected with the control by minority molecules/species. Fluctuations in replication catalytic networks were very important in the early history of life (Kaneko and Yomo, 2002; Kaneko, 2003). The discreteness of replication played a very important role here. It has been shown that the discreteness may lead to stationary states, not seen in continuous systems (Kaneko, 2003; Togashi and Kaneko, 2003).

The role of 'new species' is of basic importance for Darwin's theory. The problem of the fate of a small number of individuals of a new species in biological systems is to some ex-
tent mathematically equivalent to the fate of a new mode of behavior, a new technology, a new product or a new idea etc. (Bruckner et al., 1990, 1996). The appearance of new species can also be described from a network view. For specific applications to socio-economic networks we refer to our recent review (Hartmann-Sonntag et al., 2004). Most of the work published so far has been concentrated on problems of linear growth, but here we will concentrate on nonlinear growth processes, including saturation effects as well as hyperbolic growth.

2 New species in sensitive networks

2.1 The space of description and graphical representation

Let us assume that a biological species $i$ can be encoded by, at least formally, a sequence of $\nu$ letters $[A,C,G,T]$. We may think here of the correspondence to a DNA-code. Let us assume that the correspondence "species" — "sequence" is one to one. Then, we may introduce the Gödel number $i$ corresponding to the sequence as a numbering of the species. The set of states $i$ forms the sequence space (Eigen, 1971; Eigen and Schuster, 1977, 1978; Feistel and Ebeling, 1989). An evolutionary process corresponds to a path in this sequence space. Further we introduce an occupation number $N_i$ which denotes, how many representatives of the species $i$ are participating in the evolutionary competition. The time-dependent occupation numbers $N_i(t)$ are the basic theoretical quantities we are using for the description of the evolutionary process. Since the $N_i(t)$ are always non-negative integers, the description must necessarily be stochastic. Stochastic processes of this type were developed in physics in the framework of the theory of stochastic transitions between urns, established by Ehrenfest and Ehrenfest (1907) and in biomathematical context in the theory of the so-called birth- and death processes (Bartholomay, 1958a; Eigen, 1971; Schuster and Sigmund, 1984). The Ehrenfest theory describes the transitions of 'balls' between a numbered set of 'urns': In the new context the number of representatives $N_i$ of the species $i$, corresponds to the number of 'balls' in the urn with the number $i$. The number of species $s$ is finite and may be very large, this is a property of sequence spaces. We will assume that in any possible elementary process the occupation number changes only by $\pm 1$ (one-step-processes).

In the following, we consider in particular systems in which the total number of elements is constant.

$$\sum_{i=1}^{s} N_i(t) = N = \text{const.} \quad (1)$$

This condition is considered by Eigen as the most simple model of competition (Eigen, 1971). Eigen denotes this model as condition of constant overall organization. The condition of constant total particle number restricts all processes to transitions on a positive cone in the occupation number space (Feistel and Ebel-
In biological evolution, in difference with, e.g., technological one, individuals cannot change between species. However, mathematically a correspondence between transitions and the mutual occurrence of birth and death processes exists. Birth- and death processes with nearly constant $N$ may be approximated by transition processes which correspond to a symbolic traveling of balls between urns. We formulate our approach also by using a network picture. The system is composed of a large set of enumerable types (species). Each of these types is represented by a node. At a certain point in time, only a small part of these nodes are active. The pattern of interaction between them (e.g., mutual influence on self-reproduction) determines the dynamic composition of the system. This can be visualized by (active) links between the nodes. This active network will produce a dynamics which has a certain set of stable states (Ebeling and Sonntag, 1986). We assume that the activated part of the network is embedded in a much larger network of inactive nodes and links which represent future possibilities in the evolution. A mutation/innovation appears when an unoccupied node becomes occupied for the first time. With this first occupation, a set of links connecting the 'new' node with already occupied nodes also becomes activated. It is readily apparent that such an event changes the whole composition of the system. Accordingly, the stable state that the system might have reached already becomes, instable and the system searches for a new stable state. If we assume that the interaction between the nodes (types) is a competitive one, the stable state of a certain activated network can also include the deactivation of certain nodes. Types (nodes) which are selected out will transit to a non-occupied inactive status. With respect to structural (static) properties of such a relational networks we refer to earlier work (Hartmann-Sonntag et al., 2004). Here we concentrate on the evolution of a specific dynamic network composed by two nodes and in particular on stochastic effects connected with mutations. Let us note that the same model approach can be extended to a multidimensional case.

2.2 New species in early evolution

The evolution of life started with some rare species in the primordial soup. Underoccupied networks in the 'primordial soup' are supposed to contain primarily two kinds of chemical species, (a) those abundant due to the stability of their chemical equilibrium concentration, as e.g. inorganic salts dissolved in the ocean, and (b) those produced spontaneously far from equilibrium at rates exceeding their natural decay rate, as e.g. simple amino acids in Urey's experiments. But there is a third class of molecules of central interest here - the rare ones which appear randomly as few individuals, not even measurable in term of common concentration units (Ebeling and Feistel, 1982), and usually decay again soon later. In terms
of the network, these molecules like ‘flash lights’ occupy particular empty states only temporarily with a single copy. For this reason a stochastic treatment is a ’conditio sine qua non’. Under specific circumstances, these molecules may chemically react with the abundant species and thus migrate to neighbouring empty places of the net, or even trigger catalytic cascades propagating along the edges. The catalytic structures (occupied connected parts of the net) formed by such cascades are cycle-free with overwhelming probability, but only if they by chance contain a feedback loop to their own tree’s root they may become self-amplifying occupied clusters of the network. The governing equations for the ”rare” chemical species \( x_i \) typically have the form

\[
\frac{dx_i(t)}{dt} = \sum_j E_{ij} x_j - k_i x_i
\]  

(2)

We may formulate easily a corresponding stochastic master equation. The cross-catalytic \( N \times N \) rate matrix \( E \) is non-negative and only sparsely occupied by \( n \) nonzero elements. Potentially surviving species, i.e. cascades including a feedback loop, correspond to non-negative eigenvectors of this matrix, which follow from the theory of non-negative matrices (Gantmacher, 1953, Ebeling and Feistel, 1977, 1982). In a network with very small connectivity \( c = n/N \), the feedback probability decreases rapidly with the power \( c^l \), (\( N \)-number of nodes, \( n \)-number of edges, \( l \) - length of the feedback reaction chain) (Sonntag et al. 1981). Of the shortest loops one may imagine, simple autocatalysis seems unlikely to offer sufficient evolving potentials, but already a pair of mutual catalysts seems a promising candidate. RNA chains are preferred models for the spark that ignited life. Unfolded, they represent linear sequences of nucleotide ’symbols’, storing digital information as very first ’genes’, in folded form, they possess a ’phenotype’ by their catalytic activity. If such a pair has once appeared and multiplied sufficiently often, its survival probability comes close to unity, and its modifications can start linear competition and selection within this population, thus proceeding to eigenvectors (species) with successively higher eigenvalues of the catalytic matrix \( E \). However, as soon as these eigenvalues approach the order of magnitude of spontaneous formation of the abundant chemicals of kind (b), the latter species become increasingly exhausted and change into the category of rare compounds. Since the elements of \( E \) are proportional to ambient ’food levels’ providing the required free energy fueling the reactions, \( E_{ij} = F_{ijk} x_k \), the equations (2) for the extended set of rare species must be rewritten now as

\[
\frac{dx_i(t)}{dt} = \sum_{jk} F_{ijk} x_j x_k - k_i x_i
\]  

(3)

The transition from eq. (2) to non-linear equations of competition eq. (3) thus appears as inevitable consequence of the early evolution process itself. On the one hand it brings about qualitatively new features like ’once-for-ever’ selection, i.e. new species appearing in a system de-
scribed by eq. (3) have - in a deterministic game - almost no chance to survive against the dominating established ones. On the other hand, new evolution directions become both forced and enabled this way by the creation of spatial structures.

Equations (2) are invariant against transformations $x_i = N_i/V$ from concentrations to particle numbers in a reaction volume $V$ considered, while eq. (3) leads to formally enhanced reaction rates in small volumes,

$$\frac{dN_i(t)}{dt} = \frac{1}{V} \sum_{jk} F_{ijk} N_j N_k - k_i N_i \quad (4)$$

For an adequate stochastic description of survival probabilities in this situation, the classical master equations have to be replaced by multivariate or functional master equations (Feistel and Ebeling 1989), taking into account the spatially random distribution of particles with finite chemical reaction radii. Like in classical phase transitions, nucleation becomes important for qualitative transitions of distributed systems; new species can become germs with overcritical diameters as local fluctuations, and afterwards conquer the entire system. This behaviour is well studied especially for bistable reactions (Ebeling and Feistel, 1982), where the species occupy spatially isolated regions and their competition becomes expressed by the speed and direction of interface propagation. If this way the species appear locally as droplets (coacervates, microreactors), the selection process can be formulated in terms of their droplet growth speed. Independent of the nonlinear character of internal chemical reactions, the dynamical equations for droplet competition appear as linear equations of type (2) again (Feistel et al., 1980; Feistel, 1983; 1990), where the growth rates (selective values) are derived from nonlinear eigenvalue problems like (3). As a conclusion, linear molecular competition of primordial species begins with short feedback loops and their parasitic 'cascades'. Self-accelerated reaction velocities cause the exhaustion of originally ambient food species and turn the equations into nonlinear ones, which favour nucleation and the formation of spatially separated species. Evolution based on linear, Darwinian competition laws becomes re-established again at the qualitatively new individual level after the related, inherent breaking of spatial symmetry (Ebeling and Feistel, 1992).

2.3 Old and new species in a simplified nonlinear model

Let us study now a special situation of competition on the network: We have a node or more general a connected subgraph representing a coupled system of particles of the type of an hypercycle (Eigen and Schuster, 1977, 1978) which is in competition with a similar configuration with different growth parameters. Without loss of generality let us call species 1 the "master" and 2 the "newcomer".
We assume the special kinetics of competition with "constant overall organization"

\[ \dot{x}_i = a_i x_i + b_i x_i^2 \varphi(t) x_i ; \]

\[ i = 1, 2 \]  

(5)

The function \( \varphi(t) \) follows from

\[ x_1 + x_2 = \frac{N}{V} = C = \text{const}. \]

Equations of this form follow with negative \( b_i \) from saturations effects; positive \( b_i \) have been derived for so-called hypercyclic systems (Eigen and Schuster, 1978). The deterministic behaviour of hypercycles is well understood (Eigen and Schuster, 1978). As a result of positive quadratic terms in the growth rates the phase space may be splitted into two regions separated by a separatrix \( S_i \) (Fig. 1). A certain species \( i \) can

\[ x_i = 0 \quad S_i \quad x_i = C \]

Fig. 1. Deterministic phase space only win \( (x_i = C \text{ for } t \to \infty) \), if the initial condition places the system beyond the point of the separatrix:

\[ x_i(0) > S_i ; \quad S_i = \frac{b_j - (b_j + b_i)}{b_i + b_j} \]

(6)

This way it is possible that a new species \( i \), even when having a greater growth rate, will not win because at the beginning the number of representatives is too small.

\[ x_i(t = 0) < S_i \]

(7)

Such a situation has been called once-forever selection or hyperselection. In the case of macromolecular evolution this feature was used to explain the uniqueness of the genetic code. For technological evolution we have argued elsewhere that hyperselection is an alternative explanation for so-called lock-in phenomena of technologies (Bruckner et al., 1996). The situation changes completely if we look at the stochastic picture.

3 Stochastic theory of survival of the mutations in binary competition

3.1 Master equations

First, we formulate a master equation for a binary evolutionary process replacing an old species for a new one observing the condition:

\[ N = N_1 + N_2 . \]  

(8)

Let us assume that 1 is the master species and 2 is the new species. We assume that during elementary processes the occupation number only changes by \( \pm 1 \). This is the so called one-step-process. During transition processes at most two occupation numbers can change in such a way:

\[ \frac{N_1}{N_2} \longrightarrow \frac{N_1 - 1}{N_2 + 1} \]

(9)

Due to the conservation condition we can eliminate the master species and consider probabilities for the newcomer only. Introducing transition probabilities for increase/decrease of the new individuals follows:
\[
\frac{\partial}{\partial t} P(N_2; t) = W_{N_2-1}^+(N_2 | N_2 - 1) P(N_2 - 1; t) + W_{N_2+1}^-(N_2 | N_2 + 1) \times P(N_2 + 1; t) - W_{N_2}^+(N_2 + 1 | N_2) P(N_2; t) - W_{N_2}^-(N_2 - 1 | N_2) P(N_2; t)
\] (10)

In the deterministic case a new species with higher growth rates is always successful. This is the case of pure Darwinian selection where the fittest and only the fittest survives at the end of the process.

What we observe in biological systems is usually a variety of species. This can be either explained due to the action of mutation processes and error reproduction or due to the presence of stochastic processes. As we will show, in the stochastic situation even a new better species will only survive with a certain probability. Empirical studies of growth processes in ensembles of species might shed light on which growth mechanism is present in a certain system.

### 3.2 Survival of new species

In the following, we consider the general case where the growth rates of a certain new species contain both linear as well as non-linear terms. Special versions of this very difficult problem were already considered in (Ebeling et al., 1981; Bruckner et al., 1996). The transition probabilities for this case are:

\[
W^-(N_2) = \frac{a_1}{N} N_1 N_2 + \frac{b_1}{NV} N_2 N_1^2,
\] (11)

\[
W^+(N_2) = \frac{a_2}{N} N_1 N_2 + \frac{b_2}{NV} N_1 N_2^2.
\] (12)

Again we assume that we have absorber states. Further, the number of elements in the system is constant: \(N_1 + N_2 = N = \text{const.}\). Introducing (11) and (12) into equation (10) we obtain the master equation for this case. The deterministic counterpart to the stochastic dynamics we get by multiplying the resulting master equation with \(N_k/V\) and summing over all occupation numbers. Then, we receive after factorization of the mean values the deterministic equation of hypercycle type formulated in the second section eq.(5) with \(a_i = E_i\). As mentioned above equa-
tions of the same form have been derived for so-called hypercyclic system to describe the evolution of two coupled cycles of macromolecules (Eigen and Schuster, 1978). However a dynamics of this type is much more general and describes many situations of competition between old and new. We have shown that in the deterministic case the new species 2 only can win \((x_2 = C\) for \(t \to \infty\)), if the initial condition places the system beyond the separatrix. We will show that the situation changes drastically if we look at the stochastic picture.

For the general case of linear and quadratic rates the survival probability for a new species \(\sigma_{N_2(0),N}\) can be calculated:

\[
\sigma_{N_2(0),N} = \frac{1}{1 + \sum_{j=1}^{N_2(0)-1} \prod_{i=1}^{j} \frac{a_1 + b_1 \frac{N-i}{V}}{a_2 + b_2 \frac{i}{V}}}
\]

\[
1 + \sum_{j=1}^{N-1} \prod_{i=1}^{j} \frac{a_1 + b_1 \frac{N-i}{V}}{a_2 + b_2 \frac{i}{V}}
\]

As can be seen from the formula above, the survival probability \(\sigma_{N_2(0),N}\) for a new species not only depends on the selection advantage (the relation of the parameters \(a_i, b_i\)), but also on the size of the system \(N\) (the overall number of individuals) and the initial number of individuals of the new species.

A special case of this formula is obtained for the case of purely quadratic growth \(a_i = 0\). Then we get for the case \(N_2(0) = 1\) (only one individual of species 2 occurs at \(t = 0\)):

\[
\sigma_2 = \frac{1}{1 + \frac{b_1}{b_2} N^{-1}}
\]

\(\sigma_2\) is the probability, that the new technology wins the competition process.

Fig. 3. Survival probability of a new species of a new species with purely quadratic growth rate for infinite number of generations. We study systems of different size (fat line \(N = 10\), dashed line \(N = 40\), dotted line \(N = 70\), fine line below \(N = 100\)).

We see that in the stochastic case, the separatrix \(S\) is penetrated with a certain probability. The once-forever behaviour disappears. The better species can win the selection process with certain probability even if it starts only with one individual at the beginning. This probability increases rapidly for smaller systems, as follows from the equation and as it can be seen in Fig. 3 (Ebeling and Sonntag, 1986). In the stochastic description, the presence of fluctuations helps new species to cross the barrier for entry into the system. The introduced model is composed in a modular way, to allow applications on the
metaphoric level. Different processes related to the change in the occupation number space have been introduced. We called these processes ‘elementary processes’. In principle the introduction of different further processes is possible within the model framework. The task consists of the definition of processes which can be observed empirically and finding out the corresponding transition probabilities. If we link the model to its deterministic counterpart, the instrumentarium of dynamic systems becomes available. This way, at least, we can hope to get some insights in the analytic structure of the model and possible stationary states as well as their stability behaviour. Using the stochastic model we can obtain statements about the probabilistic systems behaviour. Interesting investigations can be made by variation of the parameters. Different kinds of connections can be analyzed. Beyond the biological interpretation chosen for the model, the framework can be applied on the metaphoric level in quite different contexts. In any of these new application areas types, elements and the network of connections have to be interpreted in an appropriate new way. Examples are given in (Feistel and Ebeling, 1989; Bruckner et al., 1990).

4 Conclusions

This paper is devoted to the study of birth and death processes connected with new species in evolving networks. In particular, we investigate the influence of stochastic effects on evolution. We develop our own approach of sensitive networks relevant to the description of new species. We call a network sensitive because evolution processes described in terms of the removal or appearance of a node might change the dynamic behaviour of the system dramatically. We use a discrete representation of the system in terms of a space of occupation numbers. Then, the growth or death of new species can be described as a hopping process on a lattice inside the positive cone of the occupation number space. Further, we introduce the notion of an under-occupied system. This way we implement a set of possible future paths of developments in our modelling.

In the centre of the paper we derive descriptions for a network of two interacting species as simplest case of a sensitive network. We use a stochastic description of such a basic unit of an evolving network and base this on the theory of birth and death processes. We define master equations and non-linear differential equations as their deterministic equivalents. Contrary to deterministic models, the stochastic description offers the advantage that, at finite times, new species can arise or ’die out’. Further, the emergence of new species can be treated as a singular stepwise event implying the transition from an under-occupied to an occupied state.

In special cases the master equation can be solved analytically and a stationary survival probability for an innovation can be derived. We show that the stochastic dynamics differ
essentially from the deterministic. Separatrices, which decompose the phase space, cannot be intersected in the deterministic case. In the stochastic case they can be crossed. This way, a 'once-for-ever' or hyperselection that is known in evolutionary theory as lock-in phenomena (dead ends) for species can be avoided.

The effects of survival in nonlinear systems of hypercycle type are strongly size dependent. In smaller systems the probability for survival of new species even in a hyperselection situation increases.

Although we use biological evolution as main reference point, we also point to the fact that the modular structure of the model also allows - on the metaphoric level - for its application in quite different fields. First applications in fields as population theory, technological evolution and evolution of science have been presented in earlier work (Feistel and Ebeling, 1989; Bruckner et al., 1990). Further applications are in progress.

References


coacervates. Biofiz. 25, 882-887 [in Russian].


