Emergence of Growth, Complexity Threshold and Structural Tendencies During Adaptive Evolution of System

Andrzej Gecow

Institute of Paleobiology Polish Academy of Science, 00-818 Warszawa Twarda 51/55 Poland gecow@twarda.pan.pl

Summary. Adaptive evolution of a functioning network (e.g. Kauffman network) may force growth of this network. We consider random addition and removing of nodes in a wide range of networks types, including scale-free BA networks. Growth of a network leads to reaching a certain complexity threshold which appears suddenly as phase transition to chaos. It can be observed in the distribution of damage size of system outputs after damage spreading initialised by a small modification of the system. Over this threshold the adaptive condition as condition of network growth is a source of structural tendencies which were observed e.g. in developmental biology as regularities of ontogeny evolution, but still were not explained. Using specific algorithm the simulation has shown mechanisms of these tendencies. Our model describes living and human designed systems. We remark that the widely used two equally probable variants of signal (Boolean network, Ising model, spin glasses) are in many cases (especially outside physics) not adequate. They incorrectly lead to extreme structural stability instead of chaos. Therefore more than two equally probable variants of signal should be used. We[1] define fitness on systems outputs, which allows us to omit the local extremes and the Kauffman's complexity catastrophe.

1 Introduction

This article is a short overview of some new approach to emergence of basic properties like growth, complexity threshold and structural tendencies in natural and artificial complex systems. This approach needs a much more detailed description which is currently in press [8]. Complexity threshold appears suddenly during system growth in distribution of change size of output signals vector. In this form it can be observed in nature. It is a phase transition to chaos caused by growing number of nodes in the network when other parameters are constant. It is connected with damage spreading initialised by a small network disturbance, e.g. random addition or removing of a node.

[1] Research financed by government grant.

This study was directed to explain some known but not popular biological regularities in ontogeny[2] evolution. We observe in the simulation of different network types including scale-free BA networks some structural tendencies and their mechanisms corresponding with these regularities as an effect of random adaptive evolution of system over the above mentioned complexity and chaos threshold. We estimate that the typical living or human designed system is chaotic and grows under adaptive condition therefore we should expect such structural tendencies in these systems.

This interpretative expectation of chaos and independently the basic principle of Shannon theory lead us to treat the widely used assumption of $s = 2$ - two equally probable variants of signal (typical for Boolean network or for popular Ising and spin glasses models) as not adequate and dangerous in lots of cases other than physical spin. This assumption is extreme and can incorrectly lead to Kauffman's 'structural stability' [14]. We show, (fig.1) that using $K > 2$ - more than two input links per node cannot replace assumption of $s > 2$ (more than two equally probable variants of signal).

For simulation we use our simplified algorithm which allows us to obtain one particular vector of output signals instead of a periodic attractor. It gives correct answers for our statistical questions and practically allows to investigate structural tendencies which were known only on an intuitive level.

2 Model of system evolving under adaptive condition

Mechanism of emergence of regularities in ontogeny evolution is the main concern of this investigation. Therefore, we concern random adaptive evolution as the main source of desired statistical properties. Adaptive condition defined in the most simple way describes Darwinian mechanism of natural selection. Not only living systems grow under adaptive condition - most of human designed system as well, therefore adaptively evolved systems which are considered here constitute the main part of interesting systems.

We define adaptive condition on the effects of system function visible outside of system. They can be only output signals of system. We interpret them as system properties and we compare them to some arbitrary defined ideal. Similarity of the system's output signal vector and the ideal vector is treated as fitness. Adaptive condition does not allow the fitness to decrease.

This implicates that our system function and as network of nodes transform some signals. We use known Kauffman networks [14, 13, 11] but we use $s \geq 2$, mainly $s > 2$ (s - number of equally probable variants of signals) unlike Kauffman and Iguchi *et al.* $s = 2$, therefore our networks are generally not Boolean networks any more. There are two variants of signal for Boolean network but they do not have to be equally probable. We prefer the useful

[2] Simplifying - sequence of stages of organism development from zygote to adult form.

assumption of equally probable signal variants therefore for two variants but not equally probable e.g. 1/4 and 3/4 we would use $s = 4$ and the more probable original variant is represented by 3 new variants. It is rather not typical that two alternatives have identical probability. It would be strange, if in a real Boolean system each signal carried full available information optimally. Computer is a good sample of complex and real Boolean network. Note, that typically using zip you get a smaller file. This is the basis of Shannon's information theory [17], which considers this difference between information available to transmission by channel and real exploitation of them. We cannot use $s = 2$ as simplification because this value of s, especially when $K = 2$, is an extreme which can lead to different effects - i.e. it can place the system outside of chaos. In area of chaos using $s > 2$ we obtain different behaviour of network than using only $K > 2$ with $s = 2$. It is shown in fig.1 where we compare average damage size for different types of autonomous networks and for $s = 2$ and 4, $K = 2$ and 3.

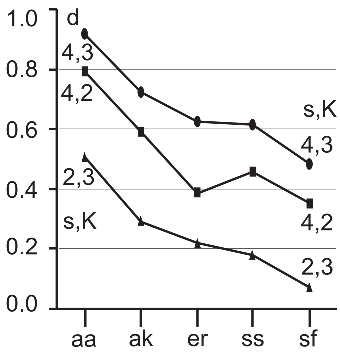

Fig. 1. Average damage size for different network types: *aa*-aggregate of automata $k = K$, Kauffman networks: *ak*-($k = K$), *er*-random Erdős-Rényi network, *ss*-single-scale and *sf*-scale-free; and different parameters s, K=4,3; 4,2 and 2,3. Note that for 4,2 and 2,3 the coefficient $w = 1.5$. We cannot limit ourselves to one of K, s and w parameters or one network type. The shown points have 3 decimal digits of precision. These networks are autonomous, they are built without adaptive condition. The *er* network is the only one with $k = 0$.

In the system with feedbacks there is no single output signal vector, but a cyclic attractor instead - a finite sequence of output signals vectors. It is not easy to compare such attractors, but for statistical questions considered in this approach we construct some simplified algorithm which answers this question correctly although it generates single particular vector of output signals after each change of system. This algorithm is dedicated for statistical simulation of damage spreading in synchronous mode. We perform calculations only for nodes with at least one changed input signal, not for entire two systems - changed A and undisturbed B - as in classic method [12]. We do not care what the remaining input signals are, they can be e.g. effects of feedback loops. A node is calculated only once [6]. Such an algorithm works fast and gives correct statistical effects.

We consider random additions and removals of nodes as changeability. Unbalance of them can lead to network growth. Only shown in fig.1. random Erdős-Rényi networks [5] does not grow. We denote it as *er*, in [11] the name RBN is used. Kauffman works using the *er* network [14, 13]. Patterns of connection of a new node define the remaining network types, which can be scale-free [3, 2] (sf, se) and single-scale [1] (ss) where number k of outgoing links is flexible or ak and aggregate of automata [6, 8] (aa, an, lx) where k is constant. We assume constant number K of node inputs (unlike in [11]).

The condition for adaptive evolution is that after each change the system's fitness (similarity of the system's output vector and the ideal output vector) must not decrease. We define tendencies as difference in changeability distribution between free process and adaptive evolution. These differences are generated by elimination - effect of adaptive condition when fitness is significantly higher than random. We use minimum $s = 4$ for simulation of adaptive evolution, for such s and 64 outputs of system we use 48 identical output signals as fitness level. Generally for any s (e.g. 8,16, 64) we use such fitness on a level of 3/4 of its maximum value. It is important to keep this level stable because otherwise the process quickly reaches the maximum fitness value. There is no local extreme in such an optimisation. In order to achieve this we change the ideal output vector each time fitness crosses an assumed level.

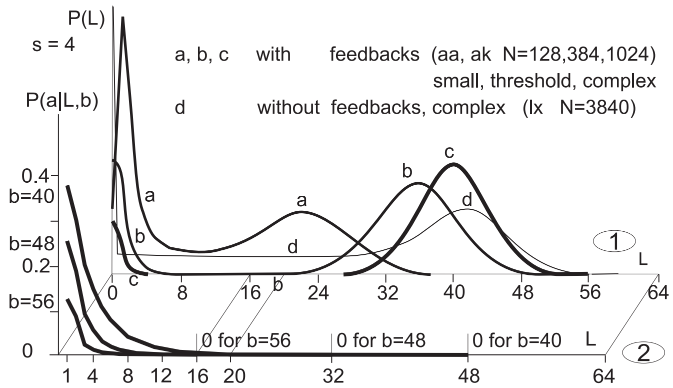

Fig. 2. Complexity threshold (1) and small change tendency (2) for 64 output signals and $s = 4$. Over the complexity threshold the main right peak is out of range of acceptable cases for higher fitness b, therefore all which can help for damage to fadeout in the first few steps, creates tendencies. For simulation of adaptive growth the fitness was fixed on $b = 48$ level by changes of ideal vector. 1- $P(L)$ distribution from simulations as the effect of random changes in the networks using aa networks as an example. Curves for ak are identical, for other networks are very similar, they differ with regard to the size of first peak. The base curve c - for complex network with feedbacks. It has two peaks and exact zero in-between. The same network at complexity threshold - curve b, and below it, when it was still small - curve a. For network lx like aa but strongly ordered and without feedbacks - curve d, space between peaks has a small but nonzero value. 2 - Theoretically calculated distribution of probability $P(a|L, b)$ of acceptance by adaptive condition, for three higher values of fitness b. For interesting cases of higher fitness the probability of acceptance is significantly different from zero only for very small change size ($L < 20$ for $b = 40$). To the right of indicated points the probability is equal exactly zero.

The main tendency, a simple but important one, is the tendency to prefer only small changes in adaptive evolution. Size of change L can be measured as number of changed output signals. This tendency is shown in fig.2.1 for three values of fitness b.

3 Growth and Complexity Threshold

Typically growth of a network is assumed as e.g. in the pattern of preferential attachment for Barabási-Albert scale-free network which concerns only addition of new node to the network like is used also in [11]. We concern random addition and removing of nodes and using equal probability of them before imposing the adaptive condition, we obtain total predominance of addition in set of changes accepted by adaptive condition. This is one of the structural tendencies. However, the mechanism of this growth does not always work. It needs some range of parameters of system to work. They are not yet introduced, we come back to growth tendency later. For a chaotic system the growth mechanism works better like other structural tendencies mechanisms.

Complexity threshold emerges during system growth. Complexity has as many descriptions and definitions [16, 9] as different aspects and meanings. If similar states of system create very different effects, we must know much more to predict these effects. We connect such a feature with the complexity of system but this is exactly chaos. A good phenomenon to observe this feature is damage spreading [12, 14]. For big chaotic systems damage typically explodes but for ordered or small systems it fades out. Using k and s we define coefficient of damage propagation: $w = k(s-1)/s$ which means an average number of changed output signals of node if one input signal is changed. It is a good, intuitive indicator of damage explosion expectation, and 'chaotic' feature of system. However, only big systems can become chaotic (how big - that depends on the network type). We have investigated this dependency and we have found some properties of damage size distribution e.g. $P(L)$, which we define as complexity threshold.

During system growth the distribution $P(L)$ of damage size evolves: It starts as one peak, similar to distribution for an ordered system, when damage quickly fades out. Next, second peak appears. It shifts (during system growth) to the right and stops at a certain position of damage equilibrium. For networks with feedbacks, before the peak stops, between peaks there appears a large period of practically exactly zero frequency. This process is shown in fig.2.2 where curves a, b and c are for consecutive stages of network growth. We define complexity threshold as reaching a stable position (80% of maximum value of L which can be calculated [14]) by the right peak and appearance of zero in-between peaks. This phenomenon can be easily observed and measured in the reality and in the simulation. It is also strongly connected with different mechanisms of structural tendencies observed in adaptive evolution of complex and functioning systems.

4 Terminal Modification and Terminal Predominance of Addition Tendencies

To discuss structural tendencies (except growth) some parameter indicating place in the network is needed. We introduce depth D - in the simple case it may be the shorter distance to the outputs of system. It is a structural, sequential, approximated measure of functional order of signal flow through the network. 'Terminal' means: near output, and 'deep' - far from output.

The first structural tendency is 'terminal modification tendency'. It means that most of changes constituting adaptive evolution take place near assessed outputs of an evolving system. This is a known regularity of ontogeny evolution introduced by Naef in 1917 [15]. It has a few explanations [4, 14, 18]. The mechanism observed in simulation in details is a quantitative confirmation of the Darwinian suggestion that changes in early stages of ontogeny usually yield monstrosities which are typically lethal. De Beer explanation, still used in biology, is extremely poor and naive. This tendency is observed and known on an intuitive level in human activity, but has not been formulated or investigated so far. It is easy to observe in computer programme development.

If coefficient $w > 1$ (system is chaotic, we assume it is big), then on average on each time step of damage spreading the damage initiated by a network change should grow. This change has much higher probability of acceptance by adaptive condition when it is very small (small change tendency, fig.2.1). If damage does not fade out in the first few steps, then it explodes and achieves equilibrium level - right peak in damage size distribution (fig.2.2). All cases forming this peak have no chance for acceptance (compare fig.2.1 and fig.2.2) therefore all which help damage to fade-out create structural tendencies. One way of such 'help' occurs at the output of system where a node does not send damaged signal to another node of system but ends its propagation. Therefore changes of network near the output of system have a higher probability of acceptance - this is terminal modification tendency.

Simulation effects of terminal modification tendency checked for different networks are shown in fig.3 in upper row. There are three simulation series summarized: aggregate of automata ($K = k = 2$, each outgoing link has its own signal which may be different than other output signal of this node) fig.3.1 with feedbacks (*aa* network), fig.3.3 without feedbacks (*an*) and fig.3.5 single-scale (*ss*) and modified scale-free networks (*se*). For *aa* and *an* networks the depth D has a more complex definition [6] than for *ss* and *se* where simply the shorter way to output is used. This tendency needs a higher s especially for the case without cost (described below) and networks *ss* and *se*.

Another structural tendency is 'terminal predominance of addition over removal' and complementary: 'simplification of deeper part of system'. These tendencies correspond to Weismann's 'terminal addition' regularity [19] and effect in old biological dispute on classic basic problem, nowadays forgotten

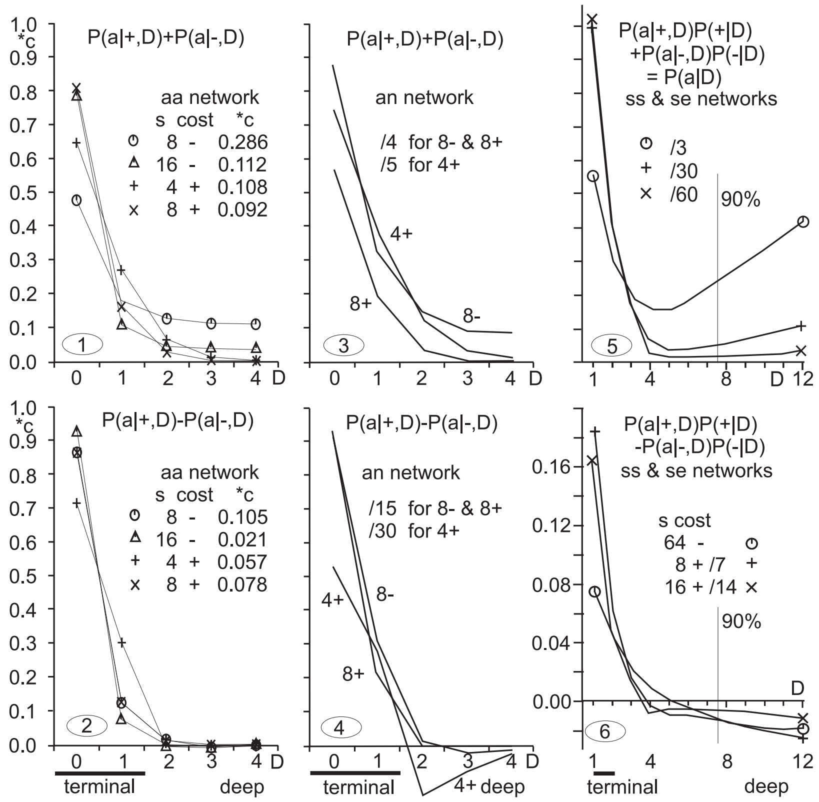

Fig. 3. The main result of simulations, (scaled by c for comparison). On the left (1,2) for *aa*, in the middle (3,4) for aggregate without feedbacks '*an*' and on the right (5,6) for Kauffman networks of variable node degree k - scale-free *se* and single-scale *ss*. The results for *ss* and *se* are so similar, that one curve for both of them is used. Dependency on depth D as structural measure of functional order. Definition of D for 1-4 is more complex [6], for 5 and 6 it is the shortest way to outputs, here the boundaries 90% of nodes are shown. On the top (1,3,5): Terminal modifications and conservation of early area tendency. On the bottom (2,4,6): Balance of addition and removing: terminal predominance of additions (over removing) tendency and tendency of simplification (predominance of removing over addition) of early parts.

and especially not popular [10, 20] - recapitulation of phylogeny[3] in the ontogeny. This tendency did not have any explanation and was therefore negated.

The sets of possibilities for random addition and removing of node is not the same. Removing can only be drawn from nodes currently present in the network, but addition has a much larger set of abilities. This difference creates difference of probability of acceptance in different areas of network, which differ on modification tempo because of terminal modification tendency. In

[3] Simplifying - sequence of adult form during evolution.

a more constant area between two accepted changes it is possible to check all nodes that they cannot be removed, but in more often changed area new changes will change checked situation on time and create new possibilities to remove. Such phenomena have influence on balance of addition and removing in different areas and in the total system growth.

Fig.3 in lower row shows these tendencies observed in the same simulation series as the tendency of terminal modifications.

One of typical removing cases occurs when a ‘transparent’ node is added, one which does not change signals of remaining nodes. Some of interpretations of such a case suggest forbidding transparent addition. We introduce such forbidding using strict inequality in the adaptive condition for additions, and weak inequality for removals. This is equivalent to cost function for additions of new automata. Cost gives much stronger tendencies but also gives some limitation for network growth [6] which are moved away using more states for fitness for *aa* and *an* networks. For *ss* and *se* networks cost forces only 1% of removing because for more removing networks do not grow (fig.3.6 and fig.3.5). For these networks using more states for fitness also helps, but still does not allow equal probability for addition and removing.

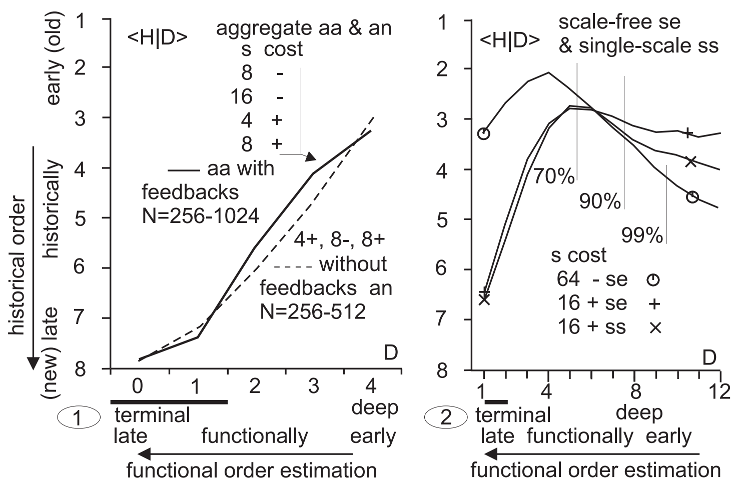

Fig. 4. Similarity of historical H and functional D order. H is a sequence of addition of given node to the network. The same set of simulations as in fig. 3. On the left (1) for aggregate with feedbacks (*aa*) - continuous lines, and without feedbacks (*an*) - dashed lines. On the right (2) for Kauffman networks with feedbacks (*se* and *ss*), containing $k = 0$ and $k = 1$. Here D is the shortest distance to outputs and can be large, but with small probabilities, therefore the boundaries 90% of nodes are shown.

Similarity of functional (measured using depth D) and historical sequence (sequence of addition of currently present nodes in the network) is an effect of these two main tendencies (terminal modifications and terminal predominance of additions). This similarity observed in simulations is shown in fig.4. These two main tendencies are clearly observed in network with fixed number k of outgoing links for each node in network. If we allow k to be flexible, like in BA scale-free network, then in the adaptive evolution of network we observe much more different mechanisms creating some tendencies. One of them, competitive to terminal modification tendency, is the tendency to fade out change (damage) in deep areas, where nodes with $k < 2$ are concentrated. On a node with $k = 1$ damage cannot multiply but if $k = 0$ damage fades out obligatorily. In a deep area the coefficient of damage propagation can be therefore locally $w < 1$ like for ordered systems. Deep fade out effect is visible in fig.3.5 especially for the case without cost.

In these investigations there appear a few problems and additional assumptions are made to solve them. One of such problem was a big part of 'transparent' changes which do not change signals. They cloud other interesting effects and are interpretatively questionable - we introduce cost to remove such additions. Other problem is that nodes with $k = 0$ (without outgoing links) appear during removing of nodes for single-scale and scale-free networks. For scale-free network sf such nodes cannot come back into play and their number can grow creating a dummy system. We add drawing a node to the pattern of addition where only links were drawn and we use proportionality to $k + 1$ instead to k. Such network we named *se*. For application to description of processes like ontogeny an interpretation of the feedbacks is discussable. We have checked the main conclusion for *aa* network without feedbacks (*an* network) - terminal modifications and terminal predominance of additions as well as similarity of functional and historical orders do occur. In the simulation we observe many more interesting parameters and phenomena which will be described in more detail [8].

Summarise: adaptive condition as condition of network growth is a source of many important properties of complex systems.

5 Conclusion

Thanks to a specific algorithm these investigations open a new, large and still unexplored area of important phenomena (structural tendencies) waiting for mathematical description. In addition to structural tendencies we observe the emergence of complexity threshold upon which the mechanisms of these tendencies work. Structural tendencies are effects of adaptive condition as condition of network growth like preferential attachment for BA scale-free networks. Living and most of human designed networks grow under adaptive condition. Main application is, however, to show the mechanism of controversial recapitulation phylogeny in the ontogeny in developmental biology.

In this paper we only show the basic results (more in [8]) of a significant extension of an investigation published earlier [6, 7]. We check the conclusion for a wide range of network types which increases the importance of this conclusion. An explanation of the complexity threshold is proposed. We also argue why more than two equally probable signal variants should be used.

References

1. R. Albert, A.-L. Barabási: Statistical mechanics of complex networks. Rev. Mod. Phys., **Vol.74, No.1**, (2002) pp 47–97
2. A.-L. Barabási, E. Bonabeau: Scale-Free Networks. Scientific American, www.sciam.com (2003) pp 50–59
3. A.-L. Barabási, R. Albert, H. Jeong: Mean-field theory for scale-free random networks. Physica A **272**, (1999) pp 173–187
4. G. de Beer *Embryos and Ancestors* Oxford University Press,(1940)
5. P. Erdős and A. Rényi: Random graphs. Publication of the Mathematical Institute of the Hungarian Academy of Science, **5**, (1960) 17–61
6. A. Gecow, M. Nowostawski, M. Purvis: Structural tendencies in complex systems development and their implication for software systems. Journal of Universal Computer Science, **11** (2005) pp 327–356
   $http://www.jucs.org/jucs\_11\_2/structural\_tendencies\_in\_complex$
7. A. Gecow: From a "Fossil" Problem of Recapitulation Existence to Computer Simulation and Answer. Neural Network World. **3/2005**, pp 189–201
   $http://www.cs.cas.cz/nnw/contents2005/number3.shtml$
8. A. Gecow: Structural Tendencies - Effects of Adaptive Evolution of Complex (Chaotic) Systems. Int.J.Mod.Phys.C, in press.
9. M. Gell-Mann *What Is Complexity?* (John Wiley and Sons, Inc. 1995)
10. S. J. Gould *Ontogeny and phylogeny* (Harvard University Press, Cambridge, Massachusetts 1977)
11. K. Iguchi, SI. Kinoshita, H. Yamada, Boolean dynamics of Kauffman models with a scale-free network. J. Theor. Biol. **247**, pp 138–151, (2007)
12. N. Jan, L. de Arcangelis: Computational Aspects of Damage Spreading. In: *Annual Reviews of Computational Physics I*, ed by D. Stauffer (World Scientific, Singapore 1994) pp 1–16
13. S. A. Kauffman: Metabolic stability and epigenesis in randomly constructed genetic nets. J. Theor. Biol. **22**, pp 437–467 (1969)
14. S. A. Kauffman *The Origins of Order: Self-Organization and Selection in Evolution* (Oxford University Press, New York 1993)
15. A. Naef *Die individuelle Entwicklung organischen Formen als Urkunde ihrer Stammesgeschichte* (Jena 1917)
16. L. Peliti, A. Vulpiani: Measures of Complexity. Lecture Notes in Physics, **314** (1988)
17. C.E. Shannon, W. Weaver *The Mathematical Theory of Communication* (University of Illinois Press 1949)
18. I.I. Schmalhausen *Organism as a whole in individual and historical development* (in Russian) Nauka, Moscow (1982)
19. A. Weismann *The Evolution Theory* 2 vols. (London, 1904)
20. A. S. Wilkins *The evolution of developmental pathways* (Sinauer Associates, Inc. Sunderland, Massachusetts 2002) pp 19–22