Modeling Cognitive Evolution: the Natural Way towards Artificial Intelligence*

Vladimir G. Red'ko
Institute of Optical and Neural Technologies, RAS
Vavilova street, 44/2, Moscow, 119333, Russia.
Telephone: +7 (095) 135-6331, E-mail: redko@iont.ru

Abstract. The paper discusses the questions: How could we analyze the evolution of animal cognitive abilities? What are main "intelligent properties" of animals? How could we model emergence of natural intelligence? What are conceptual schemes, which could help us to perform such modeling? Could such modeling be a scientific basis for artificial intelligence?

Key words: evolution, animal cognitive abilities, scientific basis for artificial intelligence

1. Introduction

Biological evolution was able to create complex, harmonic, and very effective control systems, which govern the animal behavior. But how do these control systems operate? How did they emerge? What kinds of information processing, memory structures are used in animal control systems? How did animal cognitive abilities evolve? What kinds of "internal models" of environment do emerge in the animal "minds"? How are these "models" used in animal behavior? What were transitional stages between animal cognitive abilities and human intelligence?

In order to investigate such a wide spectrum of problems, it is reasonable to analyze (by means of mathematical and computer models) the animal control systems and emergence of animal "intelligent" features step by step, considering the biological evolutionary process as underlying background. In order to do the first steps in this direction, it is natural to represent the whole picture of evolution of animal cognitive abilities and some conceptual schemes, which could help us to model the process of evolution of "intelligent" properties of animals. This paper tries to represent such picture of evolution of animal cognitive abilities and to describe such conceptual schemes. It is argued that modeling of evolution of cognition can be natural scientific basis for artificial intelligence (AI) researches.

2. Could we really approach to understanding human level intelligence?

In order to demonstrate additionally the importance of the investigations of cognitive evolution, let's consider some philosophical aspects.

There is a very interesting and profound epistemological problem: why is *human* intelligence applicable to cognition of *nature*?

To illustrate the problems, let's consider physics, the most fundamental natural science. The powerfulness of physics is due to extensive and effective use of mathematics. However, why is mathematics applicable to physics? Indeed, a mathematician creates his theories, using his intelligence,

* The work is supported by Russian Academy of Science (Program of Presidium RAS "Intelligent Computer Systems", Project 2-45 and OITVS-02, Project No 2.1).

quite independently from real physical world. The mathematician can work in a silence of his cabinet, resting on a sofa, in an isolated prison cell. Why are his results applicable to real nature?

Are we able to approach to solution of these questions? In author's opinion, the answer is "yes". We can analyze evolutionary roots of human intelligence and try to investigate why and how did high level intelligent cognitive abilities evolutionary emerge. In other words, we can follow evolutionary roots of animal and human cognitive abilities and represent a general picture of evolutionary emergence of human thinking, human intelligence. We can analyze, why and how did animal and human cognitive features emerge, how do these cognitive features operate, why are they applicable to cognition of nature.

Can we really proceed in this way? Can we find evolutionary roots of human intelligence in animal cognition properties? Yes, we can. Let's consider the elementary logic rule that is used by a mathematician in deductive inferences, *modus ponens*: "if A is present and B is a consequence of A, then B is present", or $\{A, A \text{ -> B} \} \Rightarrow B$.

Now let's go from the mathematician to a Pavlovian dog [1]. The dog is subjected to the experiment of classical conditioning. A neutral conditioned stimulus (CS), a sound is followed by a biologically significant unconditioned stimulus (US), a food. The unconditioned stimulus arouses salivation. After a number of presentations of the pair (CS, US), the causal relation CS -> US is stored in dog's memory. Using this relation at a new presentation of the CS, the dog is able to do elementary "inference": {CS , CS -> US } => US. Then expecting the US, the dog salivates.

Of course, the application of modus ponens rule (purely deductive) by the mathematician and the inductive "inference" of the dog are obviously different. However, can we think about evolutionary roots of logical rules used in mathematics? Yes, we certainly can. The logical conclusion of the mathematician and the "inductive inference" of the dog are qualitatively similar.

Moreover, we can go further. We can imagine that there is a semantic network in the dog's memory. This network is a set of notions and links between notions. For example, we can imagine that the dog has notions "food", "danger", "dog of the opposite sex" – these notions correspond to main animal needs: energy, safety and reproduction. Further, the notion "food" can have semantic links to notions "meat", "bread" and so on. We can also imagine that a semantic link between a CS and a US is generated in dog's memory at classical conditioning. So, we can imagine generation and development of different useful semantic networks during dog's life. These networks reflect dog's experience and stored in dog's memory. To some extent, these semantic networks are similar to semantic networks that are studied in AI researches [2].

Thus, we can think about evolutionary roots of inference rules and logical conclusions and we can really investigate evolutionary roots of human intelligence. Let's consider the main stages of evolution of animal "intelligent" features.

3. Intelligent inventions of biological evolution

In this section we will begin from the very beginning – from the simplest forms of live – and will try to extract the levels of the "intelligent inventions" of biological evolution. We will mention key examples of "inventions" and outline the corresponding mathematical models, which have been already developed.

First level. An organism perceives different states of the external environment; the information about these states is memorized in the organism genome and is inhered. The organism adaptively uses the information about these states by changing its behavior in accordance with changes of the environment states.

An example of this level is the *regulation of enzyme synthesis* in bacteria in accordance with the classical scheme by F. Jacob and J. Monod [3]. This scheme of regulation can be outlined as follows. Bacterium *E. Coli* uses glucose as its main nourishment. However, if glucose is absent, but another substrate, lactose is present in the environment, *E. Coli* turns on the synthesis of special enzymes, which transform lactose into the usual nourishment, glucose. When bacterium returns into a glucose-rich environment, the synthesis of transforming enzymes is turned off. This scheme of regulation can be considered as the unconditional reflex at the molecular-genetic level. It can also be considered as a scheme of primordial control system.

The mathematical model of such scheme of regulation, "Adaptive syser" was created and analyzed by V.G. Red'ko [4]. The model represents a possible scheme of the origin of primeval control system at prebiological level.

Second level. An organism individually stores information about situations in the external environment in its short-term memory. This memorizing ensures the acquired adaptation of the organism to events in the environment.

An example of this level is the *habituation* of infusoria, demonstrated by W. Kinastowski [5]. If an infusorium is subjected many times to a neutral stimulus, e.g. drop of water, its reaction (twitching) on the stimulus is initially large, but in further course of the experiment the reaction is decreased. This form of adaptation is of the short-term type. According to the experiments of W. Kinastowski, the habituation of infusoria is formed during 10-30 minutes; it is maintained during 1-3 hours.

Tsetlin's automata are well-developed mathematical models that correspond approximately to the "intelligence level" of habituation [6]. Tsetlin's automata illustrate the simple acquired properties of biological organisms and simple adaptive behavior in changing external environment.

Third level. An organism individually stores the causal relations between the events in the external environment. The causal relations are stored in long-term memory.

An example of this level is *classical conditioning*. In well-known experiments of I.P. Pavlov [1] on a dog, a neutral conditioned stimulus, CS was followed by a biologically significant unconditioned stimulus, US. The unconditioned stimulus aroused a certain unconditioned response. After a number of presentations of the pair (CS, US), the CS alone became able to arouse the same (conditioned) response.

The classical conditioning has several non-trivial particularities. There are three stages of learning procedure in classical conditioning: pre-generalization, generalization, and specialization [7]. During the pre-generalization, the conditioned response is still absent, but there is the increase of electrical activity in different areas of the animal brain. During the generalization, both the CS and other (differential) stimuli, which are similar to the CS, arouse the conditioned response. The generalization is followed by the specialization, at which the response to differential stimuli is gradually vanished, whereas the response to the CS is retained.

The causal relation between the CS and the US is stored in the long-term memory: a conditional reflex is conserved during several weeks for low-level vertebrates and up to several years (and may be during the whole life) for high-level animals. The characteristic feature of classical conditioning is the spontaneous recovery: the renewal of a conditioned response, which takes place several hours after extinguishing of a conditional reflex [8]. The conditional learning depends strongly on animal motivations; e.g., in experiments of I.P. Pavlov the dog was hungry. The biological meaning of classical conditioning is prediction of future events in environment and adaptive use of this prediction [9].

There are a number of mathematical and cybernetic models of conditional reflex, created and investigated by A.A. Lyapunov [10], S. Grossberg [11], A. Barto & R. Sutton [12] and others. But, in author's opinion, some significant aspects of classical conditioning have not been mathematically described yet. This concerns mainly the feature of the spontaneous recovery, the role of a motivation in conditional learning, and the biological meaning of the classical conditioning.

There are several levels of "intelligent inventions" between classical conditioning and human intelligence. We only mention some of them here.

Instrumental conditioning is similar to classical conditioning, but it is more complex: an animal has to discover adequate new conditioned response (that is not known to it in advance), in order to obtain a reinforcement after presentations of a conditional stimulus.

Chains of conditioning is a sequence of conditioning responses that is formed on the base of old conditioning relations, which have been already stored in animal memory.

High-level animals use the non-trivial *models of external environment* in their adaptive behavior. Certainly, some forms of "behavioral logic" are used in such modeling to predict future situations and to reach a goal. Examples of such "intelligent" behavior are well-known experiments of W. Köhler on apes [13]. Apes were able to use several instruments (sticks, boxes) on order to get over several difficulties and solve a complex task of reaching food. Obviously, apes use certain models and certain logic during solving these tasks.

The *final level* we consider is *human logic*. The mathematical theories of our logic are well developed. There is the propositional calculus, there is the predicate calculus [14], and there are theories of mathematical inference [15]. Theories of inductive and fuzzy logic were intensively developed in the last decades [16, 17].

Thus, it is possible to extract the several key "intelligent inventions" and consider the sequence of achievements of biological evolution (Fig.1). The abilities to cognize the natural phenomena is gradually increased in this sequence.

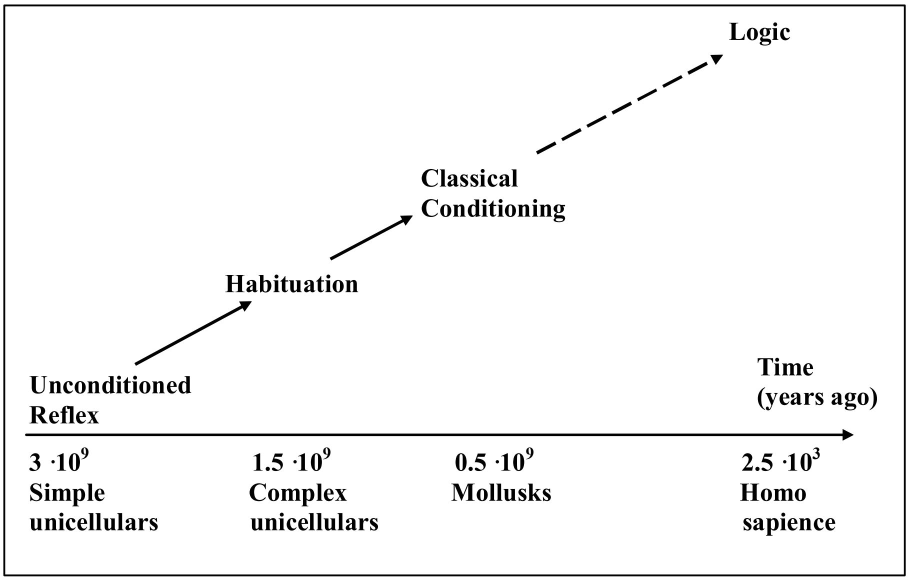

Fig.1. Intelligent inventions of biological evolution.

The analysis of mentioned models of "intelligent inventions" (as well as other models of intelligent adaptive behavior) demonstrates that we are very far from full-scale theory of evolution of cognition. The models developed can be considered only as first steps towards such a theory. These models have obviously fragmentary character; there are no models, which could describe the transition stages between the intelligent inventions of different evolutionary levels.

Thus, modeling of "intelligent inventions" of biological evolution is at initial stages of development. So, it is reasonable to consider ideas and methodological schemes, which could help to model these "inventions". In the next section we will outline some of methodological approaches: the metasystem transition theory by V.F. Turchin [18] and the biologically-inspired framework for modeling of intelligent adaptive behavior, which is based on the concept of functional system by P.K. Anokhin [9].

4. Methodological approaches

4.1. Metasystem transition theory by V.F. Turchin

In the book "The Phenomenon of Science. A Cybernetic Approach to Human Evolution", V.F. Turchin outlined the evolution of cybernetic properties of biological organisms and considered the evolution of scientific cognition as a continuation of biocybernetic evolution [18]. In order to interpret the increase of complexity of cybernetic systems during evolution, V.F. Turchin proposed the *metasystem transition theory*. This theory introduced a general cybernetic scheme of evolutionary transitions between different levels of biological organization.

In simple words, the metasystem transition theory can be outlined as follows. A transition from a lower level of system hierarchy to a next higher level is a symbiosis of a number of systems $\boldsymbol{S}_i$ of low level into the combined set $\Sigma_i \boldsymbol{S}_i$; the symbiosis is supplemented by emergence of the additional system $\boldsymbol{C}$, which controls the behavior of the combined set. This metasystem transition results in creation of the system $\boldsymbol{S'}$ of new level ($\boldsymbol{S'} = \boldsymbol{C} + \Sigma_i \boldsymbol{S}_i$), which can be included as a subsystem into the next metasystem transition.

V.F. Turchin notes the following examples of metasystem transitions:

control of position = movement
control of movement = irritability (simple reflex)
control of irritability = (complex) reflex
control of reflex = associating (conditional reflex)
control of associating = human thinking
control of human thinking = culture

V.F. Turchin describes the metasystem transition as certain cybernetic analog of the physical phase transition. He pays special attention to quantitative accumulation of progressive traits in subsystems $\boldsymbol{S}_i$ just before a metasystem transition and to multiplication and developments of subsystems of the penultimate level of the hierarchy after the metasystem transition.

The metasystem transition theory provides us with an interpretation of general processes of evolutionary increase of complexity. The more intimate processes of intelligent adaptive behavior can be analyzed on the base of the theory of functional systems, which was proposed and developed in the 1930-1970s by Russian physiologists P.K. Anokhin [9].

4.2. Theory of functional systems by P.K. Anokhin

Anokhin's functional system is a neurophysiological system that is aimed at achievement of an organism's vital needful result. The main mechanisms of the functional system operation are (Fig.2):

1) afferent synthesis,
2) decision making,
3) generation of an acceptor of an action result,
4) generation of the action (efferent synthesis),
5) the complex action,
6) an achievement of a result,
7) backward afferentation about parameters of the result, comparison of the result with its model that was generated in the acceptor of the action result.

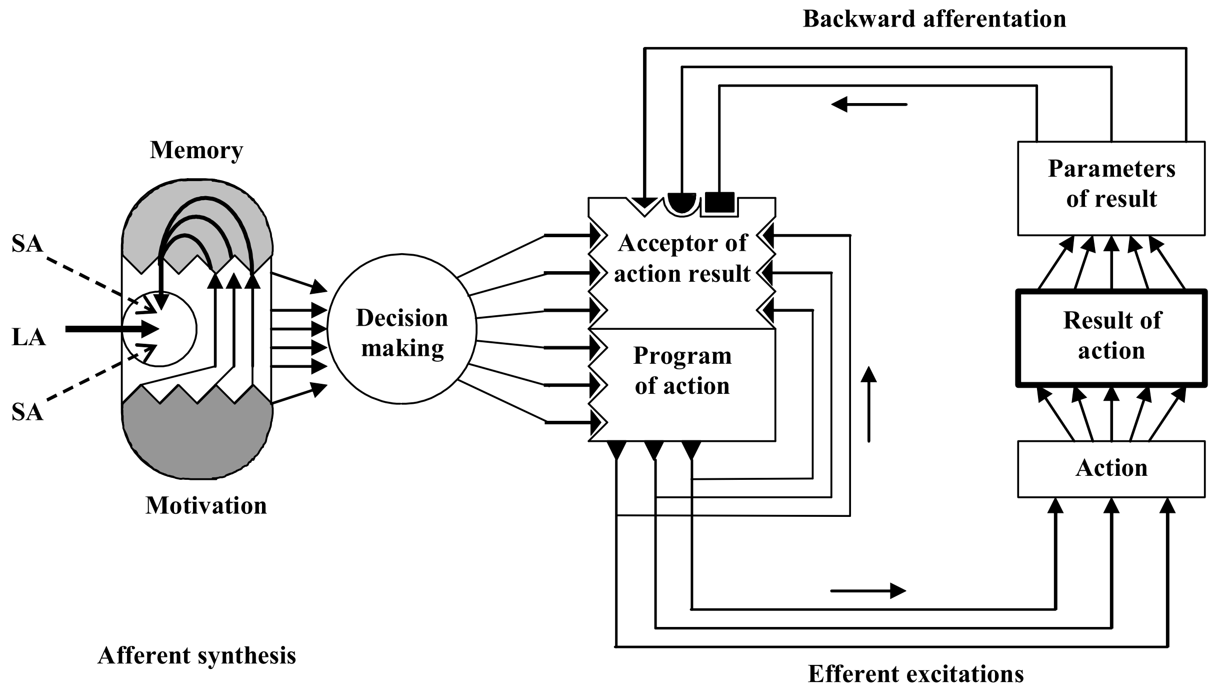

Fig. 2. General architecture of a functional system. LA is launching afferentation, SA is situational afferentation.

Operation of a functional system can be described as follows.
An afferent synthesis involves synthesis of neural excitations that are due to 1) dominating motivation, 2) launching and situational afferentation and 3) inherited and acquired memory.

The afferent synthesis is followed by decision making, which means a reduction of degree of freedom for an efferent synthesis and selection of a particular action in accordance with dominating animal need and other constituents of the afferent synthesis.

Next step of the operation is the generation of the acceptor of the action result. This step is the formation of a prognosis of the result. The prognosis includes forming of particular parameters of the foreseeing result.

The efferent synthesis is a preparation for the effectory action. The efferent synthesis implies generation of some neural excitations before generation of an action command.

All stages of result achievement are permanently estimated by means of backward afferentation. If parameters of an actual result are different from parameters of the acceptor of action result, then the action is interrupted and new afferent synthesis takes place. In this case, all operations of the functional system are repeated until the final needful result is achieved.

Thus, operation of the functional system has a cyclic (with backward afferent links) self-regulatory organization.

The most important particularity of Anokhin's theory is orientation of operation of any functional system to achievement of a final needful result.

The next particularity is dynamism, temporality. At each behavioral action, different neural and other regulatory structures of an organism are mobilized into a functional system.

In addition, an important concept of the functional system theory is systemogenesis. The essence of systemogenesis is that organism functional systems – needed for adaptive behavior of animals and men – are formed at both pre-natal period and ontogenesis.

It should be underlined that the theory of functional systems was proposed and developed in order to interpret a number of neurophysiological data. The theory was formulated in rather general and intuitive terms. The formalization of the functional system theory by means of mathematical/computer models is only beginning [19,20]. In particular, the model of the control system of an animat (natural or artificial organism) has been proposed in [20]. The model implies that the animat has several natural needs (the need of energy, the need of reproduction, etc.). In accordance with these needs, the animat has the hierarchical system of goals and sub-goals. The control system consists of a set of hierarchically linked functional systems and provides predictive and goal-directed behavior. The design of functional systems is based on neural networks, which are learned by means of the back-propagation method [21].

Though formal powerful models of the functional system theory have not yet created, this theory is based on numerous experimental data and provides us with important conceptual approach to understanding of brain operation. The theory could help us to understand neurophysiological aspects of prognosis, prediction, creation of causal relation between situations in animal minds. The functional system theory can be considered as rather universal scheme of intelligent adaptive behavior. So, the conception of functional systems could provide a general framework for development of mathematical models of "intelligent inventions" of different evolutionary levels.

Basing on Anokhin's theory of functional systems, and trying to model intelligent adaptive behavior of artificial entities, we can propose a very general scheme of control system for purposeful adaptive behavior. Such engineering interpretation of the functional system is shown in the Fig. 3. We can interpret "organism" at this scheme as "animal", "robot", "enterprise", or even "state" and "humankind".

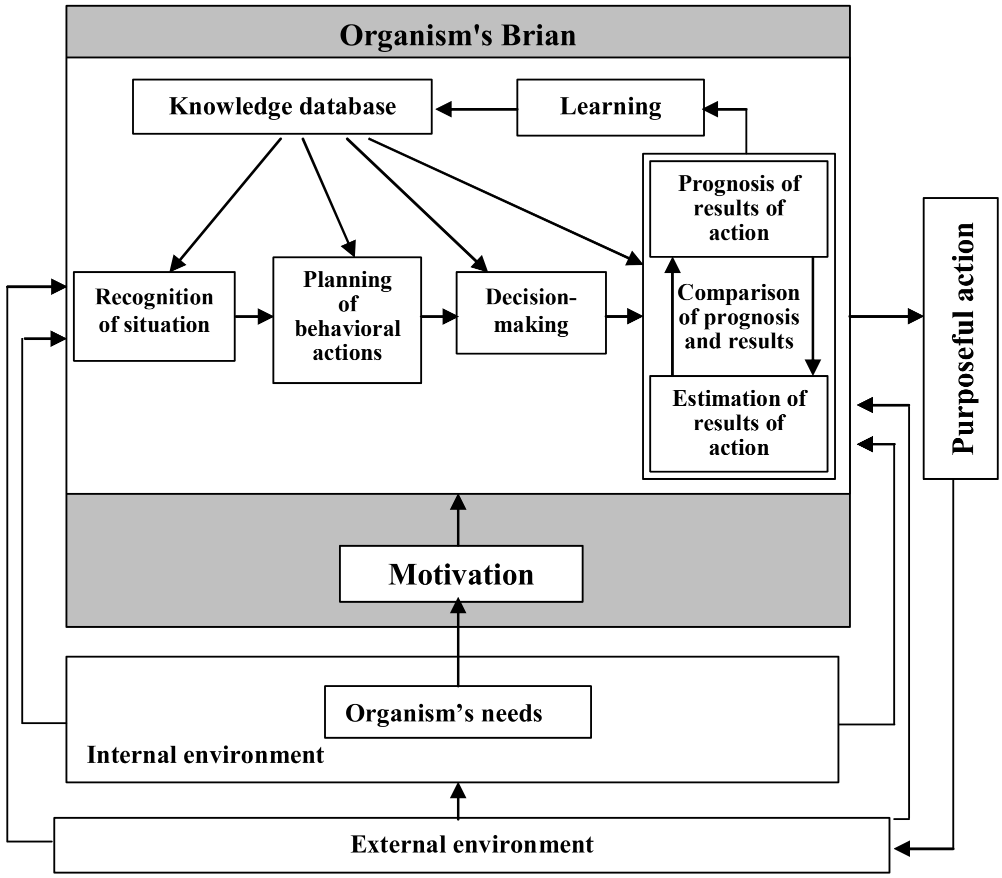

Fig. 3. Engineering interpretation of the functional system.

5. Animat approach to artificial intelligence

In the early 1990s, the animat approach to AI researches was proposed. This research methodology implies understanding intelligence through simulation of artificial animals ("animats") in progressively more challenging environments [22,23]. Let's illustrate the animat approach by characterizing viewpoint of investigators of two laboratories, which work in this research direction, the AnimatLab (Paris) and the AI laboratory of University of Zurich.

According to the AnimatLab [24], the goals of the animat approach are as follows:
"On the short term, the objectives of the animat approach are first to understand the mechanisms that afford animals the possibility of adapting and surviving, and then to import such mechanisms within artifacts < e.g., autonomous robots> capable of adapting themselves and of fulfilling their mission within more or less changing and unpredictable environments...
On the long term, the objective of the animat approach is to contribute to the advancement of cognitive sciences through the study of how human intelligence is rooted in simpler adaptive behaviors inherited from animals, in a bottom-up, evolutionary and situated perspective."

AI laboratory of University of Zurich uses "…a synthetic methodology that can be characterized as "understanding by building". It consists of three steps: (1) modeling spects of a biological system, (2) abstracting and exploring general principles of intelligence, and (3) using these principles in the design of artifacts" [25].

The animat researches are highly interdisciplinary; they are at the interface of neurosciences, cognitive science, ethology and ecology, on the one hand, of computer science and robotics, on the other hand.

6. Modeling cognitive evolution is the natural way towards artificial intelligence

AI is an area of applied researches. Experience demonstrates that an area of applied researches is successful, when there is a powerful scientific base for the area. For example, solid state physics was the scientific base for microelectronics in the second part of 20-th century. It should be noted that solid state physics is very interesting for physicists from scientific point of view. Therefore, physicists made a lot in scientific basis of microelectronics, independently of possible applications of their results. And results of microelectronics are colossal. Microelectronics is everywhere now.

What could be a scientific base of AI (analogously to the scientific base of microelectronics)? We can consider this problem in the following manner. Natural human intelligence was emerged through biological evolution. It is very interesting from scientific point of view to study evolutionary processes that resulted in human intelligence, to study cognitive evolution, evolution of cognitive animal abilities. Moreover, investigations of cognitive evolution are very important from epistemological point of view: such investigations could clarify the very profound epistemological problem (see the section 2): why is human intelligence applicable to cognition of nature? Therefore, we can conclude, that investigation of cognitive evolution can be the natural scientific base of AI developments.

What could be a relation between academic investigations of cognitive evolution and applied AI researches? In author's opinion, it is natural
1) to develop mathematical/computer models of "intelligent inventions" of biological evolution,
2) to represent by means of such models a general picture of cognitive evolution, and
3) to use these models as a scientific background for AI researches.

This point of view is similar to the animat approach, described in the section 5. However, development of the general theory of cognitive evolution, outlined here, would make this approach more fundamental and powerful.

References:

1. Pavlov, I. P. (1927), "Conditioned Reflexes". Oxford, London.
2. Lehmann, F. (Ed.) (1992), "Semantic Networks in Artificial Intelligence", Pergamon Press, Oxford.
3. Jacob, F. and Monod, J. (1961), "Genetic regulatory mechanisms in the synthesis of proteins". *J. Mol. Biol.* Vol. 3, pp. 318-356.
4. Red'ko, V.G. (1990), "Adaptive syser". *Biofizika* (*Biophysics*). Vol. 35. N.6, pp. 1007-1011 (In Russian). See also short on-line version of this paper: http://pespmc1.vub.ac.be/ADAPSYS.html
5. Kinastowski, W. (1963), "Der Einfluss der mechanischen Reise auf die Kontraktilitat von Spirostomum ambguum Ehrbg". *Acta Protozool.* V.1. Fasc. 23, pp. 201-222.

6. Tsetlin, M. L. (1973), "Automaton Theory and Modeling of Biological Systems". Academic Press, New York.
7. Kotlyar, B.I, Shulgovsky, V.V. (1979) "Physiology of Central Nervous System". Moscow State University Press, Moscow. (In Russian).
8. Voronin, L.G. (1977), "Evolution of Higher Nervous Activity". Nauka, Moscow. (In Russian).
9. Anokhin, P.K. (1974), "Biology and Neurophysiology of the Conditioned Reflex and Its Role in Adaptive Behavior". Pergamon, Oxford.
10. Lyapunov, A.A. (1958), "On some general problems of cybernetics" In: Problems of Cybernetics. Physmathgiz, Moscow. N.1, pp. 5-22. (In Russian).
11. Grossberg, S. (1974), "Classical and instrumental learning by neural networks". *Progress in Theoretical Biology*. Vol.3, pp. 51-141.
12. Barto, A.G., Sutton, R.S. (1982), "Simulation of anticipatory responses in classical conditioning by neuron-like adaptive element". *Behavioral Brain Research.* Vol.4, pp. 221-235.
13. Köhler, W. (1925), "The Mentality of Apes", Humanities Press, New York.
14. Kleene, S.C. (1967), "Mathematical Logic". J. Wiley & Sons, New York, London, Sydney.
15. Gentzen, G. (1935,1936), "Untersuchungen über das logische Schlieβen". *Mathematische Zeitschrift*, 1935, B. 39, pp. 176-210, 405-443. "Die Wiedersprunchsfreiheit der reinen Zahlentheorie". *Math. Ann*. 1936, Vol. 112, N. 4, pp.493-565. See also: Szabo, M.E. (editor) 1969. "The Collected Papers of Gerhard Gentzen". Studies in Logic and the Foundations of Mathematics. North Holland, Amsterdam.
16. Angluin, D., Smith, C.H. (1983), "Inductive inference: theory and methods". *Comp. Surveys*. Vol.15, N.3, pp. 237-269.
17. Zadeh, L.A., Klir, G.J. (Ed.), Yuan, B. (Ed) (1996), "Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems : Selected Papers by Lotfi A. Zadeh" (Advances in Fuzzy Systems - Applications and Theory , Vol. 6). 1996, World Scientific, Singapore.
18. Turchin, V.F. (1977), "The Phenomenon of Science. A Cybernetic Approach to Human Evolution". Columbia University Press, New York.
19. Umryukhin E.A. (1999), "Brain mechanisms: informational model and optimization of learning". Moscow (In Russian)
20. Anokhin, K.V., Burtsev, M.S., Zarayskaya, I.Yu., Lukashev, A.O., Red'ko, V.G. (2002), "Project "Animat Brain": designing model of adaptive behavior on the base of functional system theory". *Eights national conference on artificial intelligence. Proceedings*. Moscow, Fizmatlit, 2002. Vol. 2. pp. 781-789.
21. Rumelhart D.E., Hinton G.E., Williams R.G. 1986Learning representation by back-propagating error // *Nature*. Vol.323, N.6088, pp. 533-536.
22. Meyer, J.-A., Wilson, S.W. (Eds.) (1991), *From Animals to Animats: Proceedings of The First International Conference on Simulation of Adaptive Behavior*., Cambridge, MA: The MIT Press/Bradford Books.
23. Wilson, S.W. (1990), "The Animat Path to AI". In [22]. pp. 15-21.
24. http://animatlab.lip6.fr/index.en.html
25. http://www.ifi.unizh.ch/groups/ailab/