# On the Possibility to Resolve the Scientific Paradoxes in Artificial Cognitive System

## Natural-Constructive Approach to the Modelling A Cognitive Process

Olga Chernavskaya[1(✉)], Dmitry Chernavskii[1], and Yaroslav Rozhylo[2]

[1] P.N. Lebedev Physical Institute of RAS, Moscow, Russia
olgadmitcher@gmail.com, chernav26@mail.ru
[2] BICA Labs, Kyiv, Ukraine
yarikas@gmail.com

**Abstract.** The concept of scientific paradox and the possibility to reveal and resolve these paradoxes by means of artificial intelligence are discussed. The cognitive architecture designed under the Natural-Constructive Approach for modeling the cognitive process is presented. This approach is aimed to interpret and reproduce the human-like cognitive features including uncertainty, individuality, intuitive and logical thinking, and the role of emotions in cognitive process. It is shown that this architecture involves, in particular, the high-level symbolic information that could be associated with concept of "science". The scientific paradox is treated as impossibility to merge different representations of the same object. It is shown that these paradoxes could be resolved within the proposed architecture by decomposition of the high-level symbols into low-level of corresponding "images", with subsequent revision of the object's memorization procedure. This process should be accompanied by positive emotion manifestation (Eureka!).

**Keywords:** Neural processor · Image · Symbol · Memorization · Hemispheres

## 1 Introduction

According to Webster Dictionary [1], a *paradox* is defined as "a statement that is seemingly contradictory or opposed to common sense and yet is perhaps true". The Oxford Advanced Learner's Dictionary [2] suggests another definition: "A statement or proposition which, despite sound (or apparently sound) reasoning from acceptable premises, leads to a conclusion that seems logically unacceptable or self-contradictory". These definitions are similar, but not identical, thus, the logical problems start already at the stage of definition. Scientific paradoxes have no strict definition and are explained by examples. Is there any relationship between this problem and the Artificial Intelligence (AI)?

Modern AI systems ordinarily are aimed to solve certain set of problems *better* than human beings [3–6]. This implies that the processing speed is much higher, the reliability is better, the efficiency should be higher. However, these features do not cover


Y. Bi et al. (eds.), *Proceedings of SAI Intelligent Systems Conference (IntelliSys) 2016*, Lecture Notes in Networks and Systems 15, DOI 10.1007/978-3-319-56994-9_1

all the ability of human–level cognition. Initially, the main problem to be solved by human beings is the problem of survival, i.e., of adaptation to any possible situation. The adaptation process could not be predetermined since the new situations could not be predicted. Moreover, from an evolutionary viewpoint, the individuality of adaptation process is required to test various possible solutions.

The problem of revealing (formulating) and resolving the scientific paradoxes refers apparently to creativity work. Could an AI system be creative? Could it be able to resolve the paradox, i.e., make a *scientific discovery,* like some unordinary humans do? At the first view answer is "no". However, in principle this is possible for so called *human-level* AI.

Modeling the human-level cognition represents one of the modern and actual challenges. This trend embraces various approaches to the problem, such as Active Agent schemes, [7, 8], Brain Reversed Engineering [9, 10], Deep Learning [11], etc. In the papers [12, 13], there was proposed and employed so called Natural-Constructive Approach (NCA) to the cognitive system modeling, which is based on the Dynamical Theory of Information [14, 15], neurophysiology reasons [16], and the neural computing [17] (combined with the nonlinear differential equation technique). The cognitive architecture designed within this approach enables one to interpret and reproduce human-like features of the cognitive process, namely—uncertainty, individuality, participation of emotions, intuitive and logical thinking, etc. Several aspects of this architecture application were discussed recently [18–20]. In this work, the possibility of the AI with such architecture to formulate and resolve the scientific paradoxes is discussed.

## 2 On the Nature of Paradoxes

### 2.1 General Paradoxes

Generally, the paradox could be treated as a conflict derived from formally correct reasoning, which leads to mutually contradictory conclusions. Commonly, this conflict does not require any resolving; this refers rather to the humor, representing *unexpected* inference from common reasoning. As a typical example, one can remember the well-known sentence of Oscar Wilde: "I have heard so much scurrilous things about You that I am sure that You are worthy of respect". In [18], it was shown that the NCA architecture could reproduce the adequate reaction on this paradoxical information that could be interpreted as a laugh.

### 2.2 Scientific Paradoxes

Scientific paradoxes could be specified as a contradiction between two sets of axioms referring to the same real object. It should be stressed that this contradiction could appear only in our description of the object/effect/phenomena, not in Nature. The Nature never contradicts to itself, while any scientific description represents certain *idealization* of real phenomena, with some seemingly atypical peculiarities being neglected.

Apparently, the only way to resolve a scientific paradox is to revise the axiom sets and create certain new axiom which could assist to combine the original sets and thus, to eliminate the contradiction.

### 2.3 Some Examples of Scientific Paradoxes

*Mechanics Versus Statistical Physics and Thermodynamics.* The most pronounced example of the scientific paradox, which has been formulated in XIXth and resolved the XXth century, is connected with contradictions between the mechanics and thermodynamics. This problem was formulated by L. Boltzmann (see [21, 22]) and resolved later in the works of N.S. Krylov [23] and Ya.G. Sinai [24]. The essence is as follows. According to Newton's laws, all the processes in a mechanical system should be strictly determined and *reversible* (in time); information on particle's position and velocity in the initial state should provide the possibility to reconstruct completely the subsequent trajectory. Thermodynamics generally describes the same mechanical particles, but the large amount of particles leads to *thermal equilibrium* and the concept of *increasing entropy* as a measure of disorder. In other words, those systems are predictable only at the macro level, while the micro-level information appears to be lost. At first, it seemed that the gap between these two extreme views on a mechanical system concerns the number of particles: at $n \gg 1$, the amount of micro-information exceeds the system's capacity. However, it was unclear what the state should be at $n \sim 1$, and what the mechanism of transition from classical mechanics to thermodynamics and statistical mechanics. Only after the works on Sinai's billiard [24], it became clear that the master condition for this transition is *instability* of the particle motion, which leads to essentially unpredictable and irreversible trajectories. In other words, it is the *instability* that provides the rising entropy and the transition to statistical mechanics. This problem was discussed in details in [14].

*Complementarity Principle in Quantum Mechanics.* This paradox is not actually resolved yet. Two basic principles of Quantum Mechanics actually contradict to each other. The first one, Schrodinger equation for the wave function $\Psi(x, t)$ [25], is based on presumption that the system has zero entropy and could be described by dynamical (*reversible*) equation. The second principle associates the module of $\Psi^2(x, t)$ with the *probability* to find a particle in the point (x, t), thus assuming that the system of particles could be described by stochastic (*irreversible*) equations. This apparent contradiction has been neglected, and, after Bohr [26], two basic principles were simply complemented,—this paradigm is called "Bohr's Complementarity principle". However, the problem of reconcilable representation of quantum mechanics still exists.

## 3 Basic Elements of NCA

### 3.1 Dynamical Theory of Information

*Definitions.* DTI is relatively new scientific discipline evolving since the middle of XX century. Thereby, there are several definitions for the concept of information in

literature, but none is generally accepted ultimately. The most constructive definition has been proposed by Quastler [27]: *information is the memorized choice of one version among a set of possible and similar ones*. This definition does not contradict to others, but gives an idea of how the information might emerge. The choice could be made as a result of two different processes, namely—

- ***reception*** of information is *superimposed* (forced) choice; it could be associated with the term "Supervised learning"
- ***generation*** of information is *free* (*random)* choice. This process could proceed only in the presence of chaotic (random) element commonly called the *noise*.

Depending on ***who*** makes the choice, there appear:

- ***objective*** information which represents the choice made by Nature, i.e., physical principles;
- ***conventional*** information is the choice made by certain *collective* as a result of mutual interaction. It is important that this choice should not be *the best* one, but it should be accepted by all the members of a given group. In certain sense, the self-organization of neural ensemble represents the choice made by this ensemble.

(a) *Definition of the cognitive process.* It is worth noting that in literature, there is a lack of unambiguous definition of the cognitive process. Within DTI, this process could be defined by means of listing those operation with an information, that should be performed during this process. Thus, cognitive process could be defined as *the self-organizing process of recording (perception), storing (memorization), encoding, processing, generating, and propagation of the "self" conventional information*.

(b) *Main Conclusion.* Since the generation and reception of information represent *dual* (complementary) processes requiring different conditions, they should proceed in two different subsystems. The generation subsystem should contain the random element (noise), the reception subsystem should be noise free.

(c) *Representation of Emotions.* Emotions should be divided into two types: *impulsive* ones (useful for generation of information) and *fixing* ones (effective for the reception). Since the generating process requires the noise, it seems natural to associate impulsive emotions (anxiety, nervousness) with the growth of noise amplitude. Vice-versa, fixing emotions could be associated with decreasing noise amplitude (relief, delight). By defining the goal of the living organism as a homeostasis, (i.e., calm, undisturbed, stable state), one may infer that, speaking very roughly, this classification could correlate with negative and positive emotions, respectively.

### 3.2 Neurophysiology Reasons

(a) *The Concept of Dynamical Formal Neuron.* The neuron is complex system that could not be reduced to simple two-state automate, as it was assumed under the formal neuron concept [28]. The most relevant model to describe the single-neuron activity is still the FitzHugh-Nagumo model [29, 30]. Within NCA, the continual representation for the *dynamical formal neuron* is used, which is a particular case of this model.

(b) *Enigma of two hemispheres.* Human brain is divided into two similar but not identical parts, or *hemispheres*. According to inference of practical psychologist E. Goldberg [31], there is functional specialization: the right hemisphere (**RH**) is responsible for processing of *new* information, i.e., *learning*, while the left one (**LH**) operates with the *well-known* information.

(c) *Emotions* are controlled by the level and composition of neural transmitters inside the human organism. This effect is accounted for by introduction of the aggregated variable $\mu(t)$, which represents the "effective" transmitter composition, i.e., stimulants minus inhibitors. Stable state (the *homeostasis*) corresponds to the equilibrium ($\mu_0 = 0$) [13].

### 3.3 Neural Computing

(a) *The neural processor* represents a plate populated by the dynamical formal neurons described by the nonlinear differential equations [12]. An information is stored in the trained connections between these neurons.

(b) *Imaginary information* should be recorded and stored within the Hopfield-type [32] processor (distributed memory) providing associative correlations.

(c) *Encoding.* The conversion of an image into a symbol is to proceed by means of the Grossberg-type [33] processor with nonlinear competitive interactions providing the localization effect. This implies the choice of single neuron (**symbol**) to represent all the information on a given image. Here, the Kohonen paradigm [34] "Winner Take All" should be realized. Under NCA, this process should be *unstable* to secure unpredictable symbol position. This feature secures the *individuality* of a given system.

## 4 Cognitive Architecture Within NCA

### 4.1 The Processes of Recording and Memorization of the Image Information

The model for Hopfield-type processor could be written as:

$$\frac{dH_i(t)}{dt} = \frac{1}{\tau_i^H}[H_i - \beta_i(H_i^2 - 1) - H_i^3] + \sum_{j \neq i}^{N} \Omega_{ij} H_j + Z_i(t)\xi(t), \quad (1)$$

where $H_i(t)$ are variables describing the state of $i$-th dynamical formal neuron, $\tau_i^H$—characteristic activation time, $\beta_i$—parameters that characterize the activation threshold; $\Omega_{ij}$—matrix of connections between neurons; $i, j = 1.....n$. Stationary states are $H_i = +1$ (active) and $H_i = -1$ (passive), that provides the effect of the neuron switching on/off under its neighbor's influence. The last term in Eq. (1) refers to random element (the *noise*), where $Z_i(t)$—amplitude, $0 < \xi(t) < 1$ is random function calculated by e.g., Monte-Carlo method. The presence of noise provides spontaneous activation of image chains, thus involving practically all the information recorded on the plate—actually, it corresponds to the *parallel* processing of information (see [35, 36]).

According to DTI, primary information processing,—that is, its perception (recording) and storage (memorization),—requires ***two*** plates of the distributed memory. The function of recording consists in the ***choice*** of the recording version, thus it could be done with necessary participation of a noise. The function of information storage requires ***selection***, that is, memorizing only relevant (actual) information and ignoring the unnecessary one. Thus, two problems—recording and storage of the information—should be solved by means of two different Hopfield-type plates (see Fig. 1).

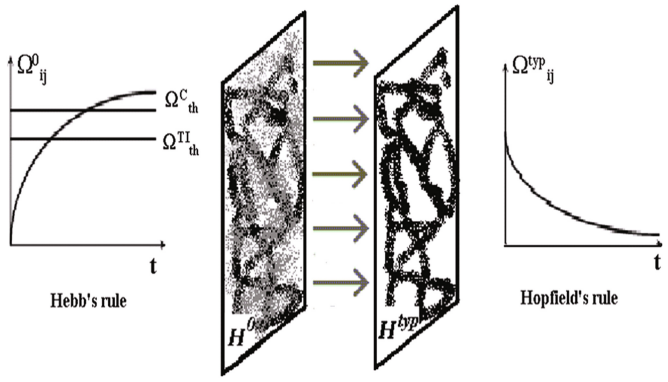


**Fig. 1.** Formation of the *typical image set* from the *fuzzy set* (center) and the time-dependence of the connection strength typical for corresponding sets (left and right sides)

The Hopfield-type plate used for perception and recording the information from various sensory systems is called the ***fuzzy set*** (or the *primary images* plate) $\boldsymbol{H^0}$. It provides recording of *any* information whenever presented to the system; the connections are to be trained according to Hebbian rule [37]: the strength of connections (initially weak) is increasing with the presentation activity (duration or recurrence) of a given object as:

$$\Omega_{ij}^{Hebb}(t) = \Omega_0 \cdot \frac{1}{4\tau_\Omega} \cdot \int_0^t \left[H_i(t') + 1\right] \cdot \left[H_j(t') + 1\right] dt' \cdot \zeta(t'), \qquad (2)$$

where $\Omega_0$, $\tau_\Omega$—are the training parameters, $\zeta(t)$ is monotonic integrable function to provide the saturation effect. Note that only active neurons are involved in training. This dynamics is presented in the left side of Fig. 1.

When the given chain of connections exceeds certain threshold value $\Omega_{th}^{typ}$ (below will be referred on as "black", i.e., *bold*, or *strong* connections), this "image" is to be transferred by the direct inter-plates connections to another plate for memorization and storage.

Selective memorizing with reduction of inessential information could be realized by means of another Hopfield-type plate which is called a ***set of typical images*** $\boldsymbol{H^{typ}}$. It should be trained just as it was proposed by J. Hopfield in [32]: all the connections are initially strong and equal. Then, the unnecessary (*waste*) connections which are ***not*** involved into the "black"-image chain diminish in the training process:

$$\Omega_{ij}^{Hopf}(t) = \Omega_0 \left\{ 1 - \frac{1}{2\tau_0} \int_0^t [1 - H_i(t')H_j(t')]\varsigma(t')dt' \right\}. \quad (3)$$

The dynamics of connection training is presented in the right side of Fig. 1. Let us point out that $\boldsymbol{H^{typ}}$ plate is formed *after and on the base* of $\boldsymbol{H^0}$ due to the system's self-organization on the principle of "connection blackening". In this process, the neurons that compose the "black" image get the status of ***typical attributes*** of the given image. Those neurons that have relatively weak (*grey*) connections with them do not participate in forming the typical image at the $\boldsymbol{H^{typ}}$ plate. They are presented in the fuzzy set $\boldsymbol{H^0}$ only as a "grey ***halo***" around the "black" image and could be called ***atypical*** (*inessential*) ***attributes.***

It should be stressed that the well-trained plate $\boldsymbol{H^{typ}}$ should also execute the image-recognition function (as was initially implied by J. Hopfield himself in [32]): due to effect of neighbors, each image presented to this plate (even damaged) is treated as the one already recorded there and appears to be *refined*, i.e., reduced to the initially-recorded form (that confirms the inability of this plate to record *new* images).

### 4.2 Symbolic Infrastructure

*Encoding* the image information, i.e., converting the image (a set of $M$ neurons) into a symbol (single neuron at a higher- level plate of hierarchy) requires the "Grossberg-type" processor [33] with localization effect providing by nonlinear (competitive) interaction. This processor could be described by the following equations:

$$\frac{dG_k(t)}{dt} = \frac{1}{\tau_k^G} \cdot \{-(\alpha_k - 1) \cdot G_i + \alpha_k \cdot G_k^2 - G_k^3\} - \sum_{l \neq k}^{n} \Gamma_{kl}(t) \cdot G_k \cdot G_l + Z_k(t)\xi(t), \quad (4)$$

where $G_k$ are variables for Grossberg-type dynamical formal neurons; $k$–1….$n$. For further analysis, these equations are written in the form providing stationary states of

neurons to be equal to: $G = +1$ (active) and $G = 0$ (passive). The parameters are: $\tau^G$—characteristic activation time, $\alpha_k$—activation threshold (controls the competitive ability of the $k$-th neuron). The last (noise) term in Eq. (4) is the same as in Eq. (1).

The process of a symbol formation could be presented as follows. At the first stage, the processor $\boldsymbol{G}$ is exposed to an image, i.e., the same set of $M$ neurons at $\boldsymbol{G}$ plate as at the image plate $\boldsymbol{H}^{typ}$ should be excited. The effect of choosing single neuron among $M$ active ones is provided by training the connections $\Gamma$ according to the rule:

$$\frac{d\Gamma_{kl}(t)}{dt} = -\frac{1}{\tau}\{G_k \cdot G_l(G_k - G_l)\}, \tag{5}$$

where $\tau^\Gamma$ being the characteristic time of the winner-choosing process. Analysis of this model [12] has shown that in the symmetrical case, $\alpha_k(t = 0) = \alpha$ and $\Gamma_{lk} = \Gamma_{kl} = \Gamma(t = 0) = \Gamma_0$, the process of choosing the symbol appears to be *unstable*. This implies that the slight casual advantage of one active neuron does provoke its expansion and suppression of the others (as a result of nonlinear interaction). Thereby, the paradigm of Kohonen [34] is realized: "Winner Take All". It should be stressed that it is impossible to predict in advance, *what concrete* neuron would appear to be a winner for a given image; this choice should be made by the plate of neurons themselves in the process of symbol formation. This very fact secures the *individuality* of an artificial system. Note that the process of symbol formation represents a typical example of appearance of the *conventional* information within a given system (collective of neurons).

After the given $G$-neuron had got a status of symbol and had formed the inter-plate connections with corresponding image neurons, it should leave a competitive struggle for the right to be a symbol of another image. This effect could be provided by *parametric* modification of the neuron-symbol: $\alpha_k \rightarrow \alpha_k(f(\{H_i\}))$. Actually, at the time scale $t \gg \tau^\Gamma$, the neuron-symbol stops its competitive interaction with neighbors, but acquires a possibility to participate in the *cooperative* interactions with other neuron-symbols ("free" $\boldsymbol{G}$-neurons could only compete).

Another very important point should be stressed. Encoding (i.e., symbol formation) means as well the *comprehension* of the image information received from outside. The very fact of symbol formation implies that the system had apprehended the given set of $M$ active neurons at the plate $\boldsymbol{H}$ as a representation of a single real object, and had awarded a proper symbol ("name") to it. Therefore below, the inter-plate connections between the symbol (on the G-type plate) and its image neurons (on the H-type plate) will be called as *semantic* ones.

### 4.3 Equations for Interaction of the Whole Neuron Ensemble

According to the DTI main conclusion, the whole system should be divided into two subsystems, which are called **RH** and **LH** in analogy with the cerebral hemispheres. The equations describing interactions between neurons of various type plates could be written in the form (see [12, 13]):

$$\frac{dG_k^{R,\sigma}}{dt} = \frac{1}{\tau_G}[\hat{Y}\{\alpha_k, G_k^{R,\sigma}, G_l^{R,(\sigma+\nu)}\}] + Z(t)\cdot\xi(t) - \Lambda(t)\cdot G_k^{L,\sigma}, \tag{6}$$

$$\frac{dG_k^{L,\sigma}}{dt} = \frac{1}{\tau_G}[\hat{Y}\{\alpha_k, G_k^{L,\sigma}, G_l^{L,(\sigma+\nu)}\}] + \Lambda(t)\cdot G_k^{R,\sigma}, \tag{7}$$

where $G_k^{R,\sigma}, G_k^{L,\sigma}$ are dynamical variables referring to the **RH** and **LH,** respectively; σ is the number of symbol's level (for the sake of brevity, the image plate ***H*** is treated as zero-level plate ***G*$^{0}$***). The functional $Y\{\alpha_k, G_k^{\sigma}, G_k^{\sigma+\nu}\}$ describes the intra- and inter-plate interactions between neurons (for details, see [12]); $\alpha_k$ and $\tau_G$ are model parameters. Here, as in Eq. (1), the term Z(t)ξ(t) in Eq. (6) corresponds to the random component (*"noise"*): Z(t) is the noise amplitude, $0 < \xi(t) < 1$ is random function (obtained, e.g., by the Monte-Carlo method). It is presented in **RH** only, thus securing the ability to generate information.

Connections Λ(t) between those subsystems play the role of *corpus callosum* and provide the "dialog" between the subsystems. They should not be trained, but have to switch on depending on the current goals. At the final stage of learning, $\Lambda^{R\to L}$ have to switch on accordingly to the "connection blackening" principle. At the stage of solving the problems, the role and mechanism of Λ(t) are to be specified.

Within NCA, there is another block of equations, which specifies the mutual influence of "cognitive" and "emotional" variables and controls the mechanism of switching the cross-subsystem connection Λ(t):

$$\begin{aligned}\frac{dZ(t)}{dt} = \frac{1}{\tau^Z}\cdot\{&a_{Z\mu}\cdot\mu + a_{ZZ}\cdot(Z - Z_0) + F_Z(\mu, Z) + \\ &X\{\mu, G_k^{R,o}\} + [\chi(\mu)\cdot D - \eta(\mu)\cdot\delta(t - t_{D=0})]\},\end{aligned} \tag{8}$$

$$\frac{d\mu}{dt} = \frac{1}{\tau^\mu}\cdot\{a_{\mu\mu}\cdot\mu + a_{\mu Z}\cdot(Z - Z_0) + F_\mu(\mu, Z)\}, \tag{9}$$

$$\Lambda(t) = -\Lambda_0\cdot th\left(\gamma\cdot\frac{dZ(t)}{dt}\right), \tag{10}$$

where $a$, χ, η, τ are model parameters, the value $Z_0$ corresponds to the rest-state level of noise, the functional $X\{\mu, G_k^{R,\sigma}\}$ refers to the process of new symbol formation (which decreases *Z(t)* value, see details in [13]). The linear in Z and μ part in Eqs. (9)–(10) provides the system's homeostasis, i.e., stationary stable state corresponding to $\{Z = Z_0, \mu = 0\}$. The functions $F_Z(\mu, Z)$ and $F_\mu(\mu, Z)$ are written in order to take into account possible nonlinear effects (see [18]). The last term in Eq. (9) refers to the processing of incoming information; *D* stays for the *discrepancy* between the *incoming* and *internal* (learned and stored) information which provokes *Z increasing*. This very situation refers to the "effect of unexpectedness" that should give rise to human-like "negative" emotions. Vise versa, finding the solution to the problem ($D = 0$) results in

momentary decrease of Z, which corresponds to "positive" emotional splash. Thus, the model (9)–(10) seems quite reasonable.

Finally, Eq. (10) specifies the relation between the cross-subsystem connections Λ(t) and the derivative of the noise amplitude; γ is the model parameter regulating the switching rate. Here, it is accepted that $\Lambda(t) = \Lambda_0 \equiv \Lambda^{R\to L}$, and vice versa, $\Lambda(t) = -\Lambda_0 \equiv \Lambda^{L\to R}$, thus "positive" emotions are associated with *Z* decreasing and activity of **LH**, while "negative" emotions (increasing Z) require **RH** activation. Lack of emotions (*dZ/dt* equals to zero) doesn't require any activity transfer. Thus, the system of Eqs. (1)–(10) appears to be completed and self-consistent.

### 4.4 The Scheme of the Cognitive System Within NCA

The particular version of NCA architecture presented in Fig. 2 has been worked out in the papers [12, 13]. The main constructive feature of this architecture consists in splitting the whole system into two (similar) subsystems: **RH** (Right Hemi-system) containing the noise, and **LH** (Left Hemi-system) free of noise. The terms are chosen to correlate conventionally with cerebral hemispheres. The noise in **RH** provides generation process, i.e. production of *new* information and *learning*. **LH** is responsible for reception and processing the already known (learned) information. This specialization, being the theoretical result of DTI principles only, surprisingly coincides with inference of practicing psychologist Elkhonon Goldberg [31]. This fact represents a pleasant surprise as well as indirect confirmation of NCA relevance.

All the connections in RH are training according to the Hebbian rule [37]: being initially weak, some connections become stronger ("blacker") during the learning process up to certain threshold value. Then the learned image is transferred to LH. In LH, on the contrary, all connections are trained according to original version of J. Hopfield [32] "redundant cut-off". Thus, RH provides the *choice*, while LH performs the *selection.*

The whole system represents complex multi-level structure that does evolve by itself (in Fig. 2—from the left to the right) due to the self-organizing principle of "connection blackening". This implies that at each level, the elementary act of the image processing in **RH** and transferring to **LH** is repeated. In physics, there is special term "scaling" for such principle of organization, and the result is called a *fractal.* The system contains four basic elements.

- *Primary images* ***I*** at the plate $\boldsymbol{H^0}$ include any available imaginary information: all signals from receptors are written as the chains of activated neurons forming the ***images***. The inter-plate connections between neurons are modified from weak (grey) to strong (black) ones upon presentation of the objects. This level carries out the function of *recording* the "sensory" information and refers to **RH**.
- *Typical images* ***TI*** are presented at the plate $\boldsymbol{H}^{typ}$ in **LH,** which perceives only the images recorded by strong enough (black) connections. Their functions are: to store useful information and filter out unnecessary one, and to recognize already learned images.

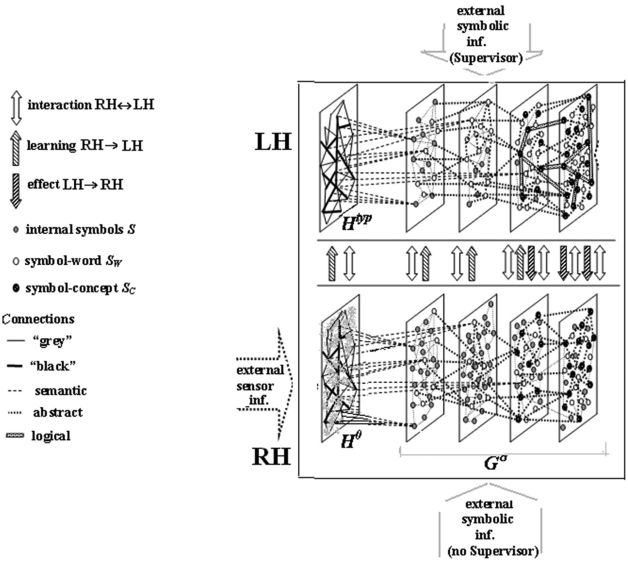


**Fig. 2.** Schematic representation of the cognitive architecture

- *Symbolic* (*semantic*) *information—symbols* ***S*** correspond to *typical* images and are formed in **RH** (with the noise participation). Each symbol possesses a semantic content, i.e., awareness of the fact that the chain of active image neurons describes a real particular object. At the same level one can find a ***standard symbol*** (symbol-word $\boldsymbol{S_W}$ that are presented in **LH** mainly) to indicate the same specific object. Symbols provide the interaction between the plates, i.e., processing of the sensible information.
- *Abstract information*—the whole infrastructure of symbols ***S***, standard symbols $\boldsymbol{S_W}$, and their interconnections. Those items are not connected with the neurons-progenitors on the plates ***H***, and thus, are not associated with any real object, but appear in the well-trained system due to interaction of all the plates (the "deduced knowledge"). Their function is to implement a communication with other systems ("*to explain by words*") and comprehend the symbolic information. The highest hierarchy levels are occupied by the *generalized symbols*, or *symbol-concepts* $\mathbf{S_C}$, such as "*conscience*", "*beauty*", "*infinity*", etc. These symbols have no corresponding special material object, but do have sense for the given system.

The emergence of each subsequent level is accompanied by a *reduction* of information. So, primary images recorded by weak (grey) connections are not transferred to $\boldsymbol{H^{typ}}$ level, and thus, could not be associated with any symbol—this information turns out to be neither conscious, nor controlled by the system itself. Such chains could be activated by the noise only, that corresponds to an *inspiration* (the "aha moment"). The whole set of "grey images" could be treated as the *sub-consciousness*.

The lower levels of the architecture represent the *latent* (hidden) individual information of the system, the "thing-in-itself". Only the higher levels, the *abstract*

*verbalized* information, make sense in the common meaning of consciousness ("*to bring on the level of consciousness*").

Note that the *latent* (hidden*)* information has its own "levels of depth", with the bulk being stored in **RH**. This very information could be interpreted as the basis for *intuition*. The *logical* thinking should be related to *verbalized concepts* and *abstract relations*, but those that are accepted in a given society. It refers to **LH** only.

### 4.5 Appearance and Resolution of Symbolic Paradoxes

As it could be seen in Fig. 2, the symbols at high hierarchy levels are connected with a lot of symbols at lower levels, and, ultimately, with a lot of images. Those symbols represent the concepts (in particular, the concept of "science") which are far from each other and not always could be integrated (merged).

The learning process as a whole could be presented as follows. The images (including the generalized images, "image-of-symbol") are processed in **RH** up to the state of strong (black) connections, turn to be *typical images* and should then be delivered to corresponding level in **LH** for storage and conversion into the higher-level symbols. In this process, the image neurons acquire the status of "*typical attributes*". The neighboring "*halo*"-neurons (having only *grey* connections with the core image ones) are "eliminated" (not delivered to the next level), being treated as *atypical (inessential) attributes.* Thus, **RH** plays the role of Supervisor, with **LH** receiving only a *part* of information stored in **RH**.

The architecture proposed possesses an essential distinctive feature, namely—so called "fuzzy set", i.e. the Hopfield-type plate in **RH** for storing all the image information whenever (even occasionally) perceived by the system and recorded by relatively weak ("grey") connections. These connections represent the *latent* information for a given individual system, since those links appear to be lost at the stage of transferring from **RH** to **LH**.

The contradictory information arises in the very process of eliminating the inessential attributes since here, the typical images (together with corresponding symbols) turn out to be *too simplified,* with the associative connections between images being *depleted*. This very mechanism results in the fact that the high-hierarchy symbols corresponding to complex multifunctional processes (e.g., "scientific-direction" symbols) could not be integrated, i.e., there are no proper connections between them. This very effect is called the "*scientific paradox*". In order to resolve this paradox the system should decompose these symbols down to the image level and activate the *fuzzy set* in **RH** to reveal the lost (*grey*, i.e., weak) associative connections between corresponding images. This revision process should result in formation of the *new symbol* representing a resolution to the given paradox. According to NCA representation of emotions (see above), this moment should be accompanied by the bright splash of positive emotions ("*Eureka!*!").

Returning to the paradox between Classical and Statistical Mechanics, one can infer that the *unstable processes* were initially treated as atypical (inessential) ones for Classical Mechanics. From the other hand, the possibility to obtain irreversible trajectory for even one particle moving in especial boundary condition such as the Sinai's

billiard, seemed inessential (atypical) for the Statistical Mechanics. Thus, the very important feature that actually resolves the paradox had been lost in both scientific branches.

The mechanism described refers to the individual AI system. However, the concept of "science" in the public cognition appears due to the same principle of neglecting the unimportant (seemingly) features of the given phenomena. Revealing the associative connections between the high-level symbols which traditionally were not taken into account as "atypical" ones could be treated as a *scientific discovery*. Note that AI system could provide this process even better than humans: the "inessential" attributes are already recorded in the fuzzy set, and there is no need to find them. However, this is true for "rich" enough systems with large and rich (abundant) fuzzy set. This feature quite correlates with the human term "wise person".

## 5 Discussion and Conclusions

Thus, it is shown that the cognitive architecture designed within NCA represents the self-organized system that evolves according to the "connection blackening" principle from low (image) level up to the higher levels of abstract (conceptual) information. The resulting architecture represents a complex multi-level construction of neural processors of different types, capable to perform the functions of recording, memorization, coding, processing and generation of information. The emotions are treated as the derivative of the noise amplitude dZ(t)/dt that should help to activate or, vice-versa, set at rest the cognitive process; they are inherently embedded into the system from the very beginning.

It should be stressed that the approach presented contains *three principle presumptions* that distinguish it from other ones. First, the main principles of DTI being applied to a cognitive process result in inference that the whole system *should by divided into two* coupled subsystems for generation and reception of information, an analogy to cerebral hemispheres. Second, the process of generation of information requires *participation of the random element* (*noise*). Note that within NCA, the noise is treated not as annoying but unavoidable disturbance (as it is in radio-physics, etc.), but as required and "full member" of a cognitive process. Third, the learning principles in those subsystems *should differ*: generation of information requires Hebbian training rule, while the reception requires Hopfield-type training. These very features enable us to interpret and imitate the human-level cognitive features, namely—uncertainty, individuality, capability of intuitive and logical thinking, the impact of emotions effect on cognitive process, etc.,—in the AI designed under NCA.

The integration (merging up) of the high-level symbols relating to different scientific directions, thus resolving the scientific paradoxes, represents actually the creative work that such AI system really could do. It is shown that these problems should be solved at the *lowest level* of the architecture—at so called "*fuzzy set*" $\boldsymbol{H^0}$ in the subsystem **RH.** This plate plays an especial role and could be associated with the human *sub-consciousness.* Nevertheless, this mechanism could work only in the case of *rich repertoire* stored in the fuzzy set. The same is true for human beings: scientific discovery requires deep understanding, great erudition, and large experience to be done.

## References


1. http://www.merriam-webster.com/dictionary/paradox
2. Oxford Advanced Learner's Dictionary. http://www.oxforddictionaries.com/definition/english/paradox
3. Russell, S.J., Norvig, P.: Artificial Intelligence: A Modern Approach, 2nd edn. Prentice Hall, Upper Saddle River (2003)
4. Samsonovich, A.V., Ascoli, G.A., De Jong, K.A.: Human-level psychometrics for cognitive architectures. In: Smith, L., Sporns, O., Yu, C., Gasser, M., Breazeal, C., Deak, G., Weng, J. (eds.) Proceedings of the Fifth International Conference on Development and Learning (2006)
5. Vityaev, E.E., Perlovsky, L.I., Kovalerchuk, B.Y., Speransky, S.J.: Probabilistic dynamic logic of cognition. Biol. Inspired Cogn. Archit. **6**, 159–168 (2013). (Special issue: papers from Fourth Annual Meeting of the BICA Society)
6. Chen, D.L., Kim, J., Mooney, R.J.: Training a multilingual sportscaster: using perceptual context to learn language. J. Artif. Intell. Res. **37**, 397–435 (2010)
7. Laird, J.E.: The Soar Cognitive Architecture. MIT Press, Cambridge (2012)
8. Samsonovich, A.: Emotional biologically inspired cognitive architecture. Biol. Inspired Cogn. Archit. **6**, 109–125 (2013)
9. Doya, K.: Complementary roles of basal ganglia and cerebellum in learning and motor control. Curr. Opin. Neurobiol. **10**, 732–739 (2000)
10. Koziol, L.F., Budding, D.E.: Subcortical Structures and Cognition: Implications for Neurophysiological Assessment. Springer, Berlin (2009)
11. Adcock, R.A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., Gabrieli, J.D.E.: Reward-motivated learning: mesolimbic activation precedes memory formation. Neuron **50**, 507–517 (2006)
12. Chernavskaya, O.D., Chernavskii, D.S., Karp, V.P., Nikitin, A.P., Shchepetov, D.S.: An architecture of thinking system within the dynamical theory of information. Biol. Inspired Cogn. Archit. **6**, 147–158 (2013)
13. Chernavskaya, O.D., Chernavskii, D.S., Karp, V.P., Nikitin, A.P., Shchepetov, D.S., Rozhylo, Y.A.: An architecture of cognitive system with account for the emotional component. Biol. Inspired Cogn. Archit. **12**, 144–154 (2015)
14. Chernavskii, D.S.: The origin of life and thinking from the viewpoint of modern physics. Phys. Uspekhi **43**, 151–176 (2000). Synergetics and Information. Dynamical Theory of Information. Moscow, URSS, 2004 (in Russian)
15. Haken, H.: Information and Self-Organization: A Macroscopic Approach to Complex Systems. Springer, Berlin (2000)
16. Richter-Levin, G., Akirav, I.: Emotional tagging of memory formation—in the search for neural mechanisms. Brain Res. Rev. **43**(3), 247–256 (2003)
17. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, Berlin (2007)
18. Chernavskaya, O.D., Rozhylo, Y.A.: On possibility to imitate emotions and a "Sense of Humor" in an artificial cognitive system. In: Proceedings of the Eighth International Conference on Advanced Cognitive Technologies and Applications COGNITIVE 2016, March 20–24, pp. 42–47 (2016)
19. Chernavskaya, O.D., Rozhylo, Y.A.: On the modelling an Artificial Intelligence with Integrated Intuition and Emotions. In: Proceedings of 25th International Joint Conference on Artificial Intelligence (IJCAI-16), 9–16 July, New York, USA, (2016 in press)
20. Chernavskaya, O.D.: The cognitive architecture within the natural-constructive approach. In: Proceedings of FIRCES on BICA, pp. 1–7 (2016)

21. Boltzmann, L.: Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. Sitzungsberichte Akademie der Wissenschaften vol. 66, pp. 275–370, 1872; Further Studies on the Thermal Equilibrium of Gas Molecules. The Kinetic Theory of Gases. History of Modern Physical Sciences, vol. 1. pp. 262–349 (1872)
22. Ehrenfest, P., Ehrenfest-Afanassjewa, T.: The Conceptual Foundations of the Statistical Approach in Mechanics. Cornell University Press, Ithaca (1959)
23. Krylov, N.S.: Works on the Foundations of Statistical Physics (2014). https://www.google.com.ua/search?hl=ru&tbo=p&tbm=bks&q=isbn:1400854741. initially published in Russian in 1950
24. Sinai, Y.G.: On the foundation of ergodic hypothesis for a dynamical system of statistical mechanics. Sov. Math. Dokl. **4**, 1818–1822 (1963)
25. Schroedinger, E.: Abhandlungen zur Wellenmechanik, Leipzig, 1927. Collected papers on Wave Mechanics, Glasgow (1928)
26. Bohr, N.: Foundation of quantum physics. In: Kalckar, J. (ed.) Niels Bohr Collected Works, vol. 6. Elsevier, Amsterdam (2008)
27. Quastler, H.: The Emergence of Biological Organization. Yale University Press, New Haven (1964)
28. McCulloch, W.S., Pitts, W.: A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. **5**, 115 (1943)
29. FitzHugh, R.: Impulses and physiological states in theoretical models of nerve membrane. Biophys. J. **1**, 445 (1961)
30. Nagumo, J., Arimoto, S., Yashizawa, S.: An active pulse trans-mission line simulating nerve axon. Proc. IRE **50**, 2062 (1962)
31. Goldberg, E.: The Wisdom Paradox: How Your Mind Can Grow Stronger as Your Brain Grows Older. Gotham, New York (2006)
32. Hopfield, J.J.: Neural networks and physical systems with emergent collective computational abilities. PNAS **79**, 2554 (1982)
33. Grossberg, S.: Studies of Mind and Brain. Riedel, Boston (1982)
34. Kohonen, T.: Self-Organizing Maps. Springer, Heidelberg (1995)
35. Haken, H.: Information and Self-Organization: A Macro-scopic Approach to Complex Systems. Springer, Berlin (2000)
36. Bianki, V.L.: Parallel and sequential information processing in animals as a function of different hemispheres. Neurosci. Behav. Physiol. **14**(6), 497–501 (1984)
37. Hebb, D.O.: The Organization of Behavior. Wiley, London (1949)