On the Modelling an Artificial Cognitive System Based on the Human-Brain Architecture

Olga Chernavskaya, Dmitry Chernavskii and Yaroslav Rozhylo

Abstract The approach to modeling a cognitive system based on the human-brain architecture, called the Natural-Constructive Approach is presented. The key point of this approach is the following: an artificial cognitive system, being a complex multi-level combination of various-type neural processors, should be divided into two subsystems, by analogy with two cerebral hemispheres in a human brain. It is shown that one of them should necessarily contain a random element (*noise*) for generation of information (creativity); it is responsible for *learning*. The other one, being free of noise, is responsible for memorization and processing the *well-known* information. Emotions could be interpreted as the noise-amplitude variation and incorporated into the system by coupling the noise amplitude with the additional variable representing the aggregated value of neurotransmitter composition, which reflects the influence of subcortical brain structures. It is shown that the activity of both subsystems should be controlled by the noise-amplitude derivative.

Keywords Generation of information · Neuroprocessor · Emotions · Learning · Noise amplitude · Hemisphere · Stimulant · Inhibitor

1 Introduction

The scientific area of Artificial Intelligence (AI) covers various approaches to modeling the cognitive process: the Robotics [1, 2], Active Agent systems [3, 4], neuromorphic (neuron-based) models [5, 6], Brain Re-Engineering (BRE) [7, 8], etc. Let us stress that any neuromorphic model (though this approach seems to be the closest to the goal) inevitably meets with the "explanatory gap" between the brain and the mind [9]. This implies that we do have a lot of information on the

O. Chernavskaya (✉) · D. Chernavskii
P.N.Lebedev Physical Institute, Russian Academy of Sciences, Moscow, Russia
e-mail: olgadmitcher@gmail.com

Y. Rozhylo
NGO Ukrainian Center for Social Data, Kyev, Ukraine

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_10

single neuron structure and functioning [10, 11], as well as on manifestation of psychological reactions [12–14], but there is lack of idea on how the first could provide the second.

Another problem is connected with the Nature challenges: any human-like cognitive model should be able to answer the questions: why there are just two cerebral hemispheres? Why any human person is individual? Why there are men and women? Up to our knowledge, these problems are not in the focus of modern researches, but this is strange. Perhaps, this could be explained by the complexity of the problems of such sort. Indeed, it is very difficult to imagine and create an artificial system which being manufactured according to standard procedure, would be strictly individual.

In the papers [15–17], there was proposed and elaborated so called "Natural-Constructive Approach" (NCA) to modeling the cognitive system. This approach is based on neurophysiology data on the human-brain structure [10, 11], the Dynamical Theory of Information [18–20], and neural computing [21] (combined with the nonlinear differential equation technique). It was shown that the cognitive architecture designed under NCA makes it possible to answer the majority of the Nature challenges. Several aspects of applying this approach to AI systems were discussed recently [22, 23]. In this paper, the main points of the NCA-architecture are discussed and compared with the neurophysiology data on the human brain structure.

2 Theoretical Foundation of NCA

2.1 Human Brain Architecture

According to the neurophysiology data (e.g., [10]), the human brain consists of the cerebral cortex and the sub-cortical structures, as it is presented in Fig. 1.

The neocortex is responsible for the high-order cognitive processes, while the sub-cortical structures (thalamus, amygdala, bazal ganglies, etc.) produce, in particular, the emotional bursts and participate in the memory formation.

Neocortex The neocortex itself could be (conventionally) divided into zones ("lobes") which are responsible for the vision (occipital lobe), motor activity (parietal lobe), and auditory activity (temporal lobe), etc. Temporal zone embraces Wernicke's and Broca's areas that are responsible, respectively, for language *hearing* (word perception) and *reproducing* (word production), but not for the *speech* itself. The *speech* function, i.e., coherent and sensible transmission of information, relates to the frontal lobe that is associated with *abstract* thinking. An appropriate model of cognitive process should reproduce and explain the distribution of these functions.

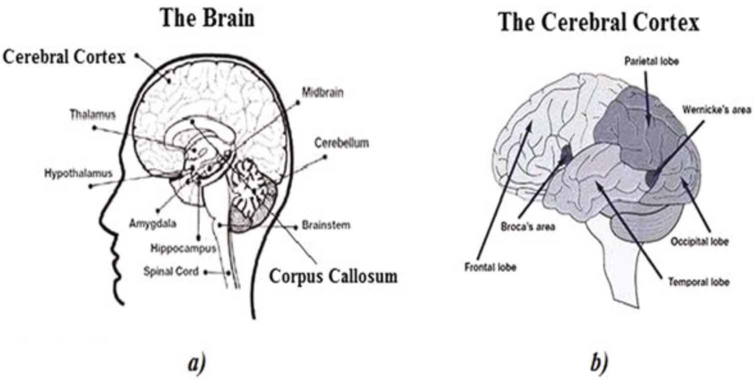

Fig. 1 Schematic representation of the human-brain structures: **a** whole view of the brain; **b** functional zones of the cerebral cortex (extracted from [24])

Sub-cortical Structures The role of sub-cortical structures was considered in [7, 8] and very interesting and completed models connecting the cognitive and emotional aspects of the learning and thinking processes were elaborated. In particular in [7], specific role of amygdala, basal ganglia, and thalamus in the learning process was demonstrated. However, there still remain unclear issues, in particular, the arrangement of the functional cortical zones. As it will be shown further, that arrangement meets quite natural explanation within the NCA architecture.

Elements of Experience Another aspect of experimental data concerns specific modification of those cortical neurons that are involved in the training process, which results in acquisition of a certain "skill" [11]. This effect is not reproduced under BRE approach [7, 8], but could be reproduced under NCA model.

Hemisphere Specialization This refers rather to neural psychology. There is a wide-spread opinion that the right hemisphere (RH) is responsible for intuition and the parallel processing of the information, while the left one (LH) provides logical thinking and sequential information processing [13, 14]. Another hypothesis that does not contradict but complements this one was proposed in [12]: RH is associated with learning new information, while LH does process the well-known one. Both these inferences (with minor revision) will be shown to be inherent for NCA architecture.

2.2 *Elements of Dynamical Theory of Information*

Definition of Information According to Quastler [24], *information is the memorized choice of one version among a set of possible and similar ones*. This definition gives an idea of *how* the information might emerge. The choice could be made as a result of two different processes, namely—*reception* (superimposed choice) and *generation* (free choice) of information. The process of generation of information could proceed only in the presence of chaotic (random) element, commonly called the *noise*.

Objective Versus Conventional Information Depending on *who* makes the choice, there appear: *objective* information (the choice made by Nature, i.e., physical principles) and *conventional* information, i.e., the choice made by certain *collective*. This choice should not be *the best*, but *individual* for a given group. In certain sense, the information that could arise in an ensemble of neurons represents the choice of this ensemble, that is, the very conventional information.

Definition of the Cognitive Process The self-organizing process of recording (perception), storing (memorization), encoding, processing, generating, and propagation of the "self" conventional information.

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

2.3 Neural Computing

Neural Processors: The Concept of Dynamical Formal Neuron The neural processor is treated as a *plate* populated by the *dynamical* formal neurons (n being the total number) described by the nonlinear differential equations (see [15, 17]), which represent a particular case of the FitzHugh-Nagumo model [25].

"Image" Information Information about ever met objects of any kind should be recorded and stored within the Hopfield-type [26] processor (*distributed* memory). It could be described by the following equations:

$$\frac{dH(t)}{dt} = \frac{1}{\tau^H}\left\{H - \beta\cdot(H^2-1) - H^3\right\} \equiv \frac{1}{\tau^H}\Im_H(H,\beta), \tag{1}$$

where H_i(t) represents the dynamical variable for Hopfield-type i-th neuron with stationary states being $H = +1$ (active) and $H = -1$ (passive); β_i is the parameter that control the activation threshold; τ is characteristic time of activation, i = 1… n. The image information is stored in the connections Ω(t), which are to be trained depending on the processor purpose.

Recording the information (learning) requires Hebbian training mechanism: initially weak connection become stronger ("blacker") in course of the learning process [27]:

$$\frac{d\Omega_{ij}^{Hebb}(t)}{dt} \propto \frac{\Omega_0}{4\tau^{\Omega}} \cdot [H_i(t) + 1] \cdot \left[H_j(t) + 1\right], \tag{2}$$

with Ω_0 and τ^{Ω} are the training-process parameters.

Storage and processing the well-known information (recognition, prognosis, etc.) require the training rule proposed by Hopfield himself [29]. This implies that all connections are initially equal and strong; during the training process the "irrelevant" ones are gradually *frozen out* according to the rule:

$$\frac{d\Omega_{ij}^{Hopf}(t)}{dt} \propto -\frac{\Omega_0}{2\tau^{\Omega}} \cdot \left[1 - H_i(t) \cdot H_j(t)\right], \tag{3}$$

what represents the principle of "redundant cut-off".

Encoding The conversion of an "*image*" (a set of M connected neurons) into a *symbol* (single neuron at the higher level of hierarchy) occurs Grossberg-type [28] processor with nonlinear competitive interactions that could be described by equations:

$$\begin{aligned} \frac{dG_k(t)}{dt} = \frac{1}{\tau_k^G} \cdot \{[-(\alpha_k - 1) \cdot G_i + \alpha_k \cdot G_k^2 - G_k^3] \\ - \sum_{l \neq k}^{n} \Gamma_{kl}(t) \cdot G_k \cdot G_l\} + Z(t)\xi_k(t); \end{aligned} \tag{4}$$

where G_k are variables for Grossberg-type dynamical formal neurons; k – 1…n; Z (t) being the noise amplitude, $0 < \xi_k(t) < 1$ is random function (obtained, e.g., by the Monte-Carlo method). Note that these equations are written to provide stationary states of neurons $G = +1$ (active) and $G = 0$ (passive). The parameters are: τ^G—characteristic activation time, α_k is the activation threshold (it controls the competitive ability of the k-th neuron). Competitive connections are trained according to:

$$\frac{d\Gamma_{kl}(t)}{dt} = -\frac{\Gamma_0}{\tau^{\Gamma}} \{G_k \cdot G_l(G_k - G_l)\}, \tag{5}$$

where τ^{Γ} being the characteristic time of the winner choosing, Γ_0 the model parameter. Analysis of this model has shown (see [15]) 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 [29] is realized: "Winner Take All". It should be stressed that it is impossible to predict in advance, *what exactly* neuron would be a winner for a given image; this choice should be made by the plate itself 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).

Formation of Generalized Images After the given *G*-type neuron became the *symbol*, it should be eliminated from the competitive struggle and acquires the possibility to cooperate with other neuron-symbols to form the *generalized* image ("image-of-symbols") by means of the cooperative connections training in analogy to Eq. (2). This effect could be provided by *parametric* modification of the neuron-symbol. Actually, at the time scale $t \gg \tau^{\Gamma}$, the neuron-symbols should behave as the *H*-type neurons, while "free" *G*-neurons could compete only.

Semantic Connections Very important role belongs to the inter-plate connections between the symbol and its image (including the generalized one). In the process of symbol formation, these connections are to be trained according to Hebbian principle, i.e., by analogy with Eq. (2):

$$\frac{d\Psi^0_{ik}(t)}{dt} \propto \frac{\Psi_0}{2\tau^{\Psi}} \cdot G^1_k \cdot [H_i(t) + 1]; \quad \frac{d\Psi^{\sigma-1}_{kl}(t)}{dt} \propto \frac{\Psi_0}{\tau^{\Psi}} \cdot G^{\sigma}_k \cdot G^{\sigma-1}_l, \tag{6}$$

where only active neurons participate in the training process; $\sigma = 2 \ldots N$ represents the number of the plate ("hierarchy level"), N is total number of levels; Ψ_0 and τ^{Ψ} are training parameters.

3 Cognitive Architecture Under NCA

3.1 Scheme of the Cognitive System

The scheme of the architecture designed under NCA in [15–17] is presented in Fig. 2. The whole system is divided into two (similar) "hemi-systems": RH (Right Hemi-system) containing the noise, and LH (Left Hemi-system) that is free of noise. The terms are chosen to correlate these "hemi"-systems with cerebral hemispheres, with the cross-hemi-system connections Λ being related with *corpus callosum* that is aimed to ebsure the dialog between subsystems. 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 Godberg [12]. This fact represents a pleasant surprise and indirect confirmation of NCA relevance.

The whole system represents complex multi-level block-hierarchical combination of different-type processors. The lowest level is presented by Hopfield-type processors H^0 and H^{typ}, the other levels embrace the Grossberg-type processors,

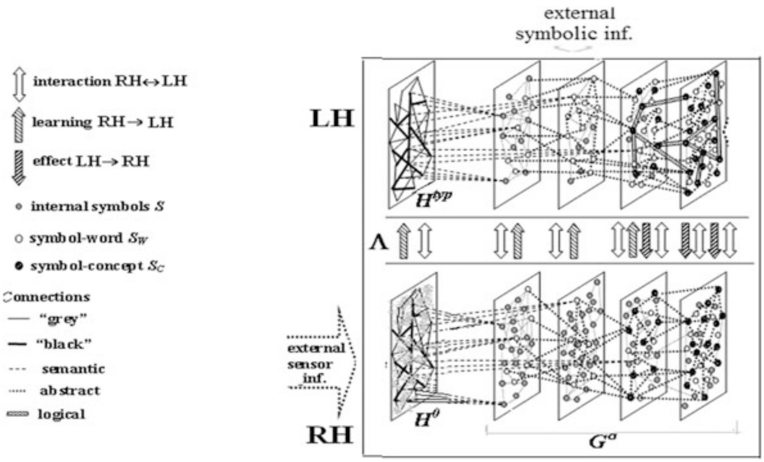

Fig. 2 The scheme of cognitive architecture

which, however, could produce the "generalized images" ("image-of-symbols") by the same mechanism of the cooperative-connection training. The number of G-type plates is neither limited, nor fixed: they appear "as required" in course of the system evolution.

This structure evolves by itself due to the self-organizing principle of "connection blackening" (see below). This implies that all the connections in RH are training according to the Hebb's principle of "connection amplification", as in Eq. (2). In LH, on the contrary, all connections are training according to original version proposed in [26], i.e. the principle of "redundant cut-off", as in Eq. (3). Thus, RH provides the *choice*, while LH performs the *selection,* with RH being the Supervisor for LH.

3.2 Equations for Neuron Ensemble Interaction

The equations describing interactions between neurons of various-type processors could be written in the form (see [15–17]):

$$\begin{aligned}\frac{dG_k^{R,\sigma}}{dt} = \frac{1}{\tau_G}&[\Im_G(G_k^{R,\sigma}, \alpha_k^\sigma(\{\Psi_{ki}^{R,(\sigma-1)}\}, G_{\{k\}}^{R,(\sigma+\nu)}))\\ &+\hat{\Upsilon}\{G_k^{R,\sigma}, G_l^{R,(\sigma+\nu)}, \Omega_{kl}^{R,\sigma}, \Psi_{ki}^{R,(\sigma-1)}, \Psi_{ki}^{R,(\sigma+1)}\}]\\ &+Z(t)\cdot\xi(t) - \Lambda(t)\cdot G_k^{L,\sigma},\end{aligned} \tag{7}$$

$$\frac{dG_k^{L,\sigma}}{dt} = \frac{1}{\tau_G}[\Im_G(G_k^{L,\sigma}, \alpha_k^\sigma(\{\Psi_{ki}^{L,(\sigma-1)}\}, G_{\{k\}}^{L,(\sigma+\nu)})) + \hat{\Upsilon}\{G_k^{L,\sigma}, G_l^{L,(\sigma+\nu)}, \Omega_{kl}^{L,\sigma}, \Psi_{ki}^{L,(\sigma-1)}, \Psi_{ki}^{L,(\sigma+1)}\}] + \Lambda(t)\cdot G_k^{R,\sigma} \quad (8)$$

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 hierarchy level (for the sake of brevity, the image plate H is treated as zero-level plate G^0). The functional $\Im(G, \alpha)$ describes the internal dynamics of a single neuron, the functional $\Im\{\alpha_k, G_k^\sigma, G_k^{\sigma+\nu}, \Omega^\sigma, \Psi^\sigma\}$ describes the intra- and inter-plate interactions between neurons (for details, see [16]); α_k and τ_G are model parameters. Here, as in Eq. (4), the term Z(t)ξ(t) in (7) corresponds to the random component ("*noise*"), with Z(t) being the noise amplitude. It is presented in RH only.

It is important to stress that the *parametric modification* of those neurons that participated in forming any "information item" (image, symbol, generalized image, etc) $\alpha_k^\sigma \to \alpha_k^\sigma(\{\Psi_{ki}^{R,(\sigma-1)}\}, G_{\{k\}}^{R,(\sigma+\nu)})$ *does explain* the experimental effect [11] on morphological changes of the neurons which have acquired certain "skill".

The last term in Eqs. (7), (8) refers to the cross-subsystem connections Λ(t) which control the activity of each subsystem thus providing their "dialog". Here and below, it is accepted: $\Lambda = +\Lambda_0 = \Lambda^{R\to L}$, and vice-versa, $\Lambda = -\Lambda_0 = \Lambda^{L\to R}$. These connections are not trained, but should switch depending on the stage of the problem solving. The mechanism of their switching should be specified (see below).

3.3 Elementary Learning Act: The Principle of Connection Blackening

The elementary act of the new symbol formation is presented in Fig. 3.

This process corresponds to the self-organizing principle of "connection blackening" and proceeds in two steps. At the first step (Fig. 3a), an image formed by the previous-level (σ − 1) symbols in RH, after its cooperative connections Ω^R become strong ("black") enough, is delivered by the direct inter-plate connections ψ to the symbolic plate G^σ. Simultaneously, by the inter-subsystem connections $\Lambda^{R\to L}$, it is transferred to the "image" plate $G^{\sigma-1}$ in LH. At the second step (see Fig. 3b), new symbol is formed due to the winner-choosing process described above by Eqs. (4), (5). Semantic inter-plate connections Ψ^R are trained according to Eq. (6) up to sufficiently strong ("black") state which is defined by the condition:

$$\frac{1}{M_k^\sigma}\cdot\sum_i^{M_k^\sigma}\Psi_{ik}^{R\sigma} - \Psi^{\text{thr}} \geq 0; \leftrightarrow i \in \{M_k^\sigma\}, \quad (9)$$

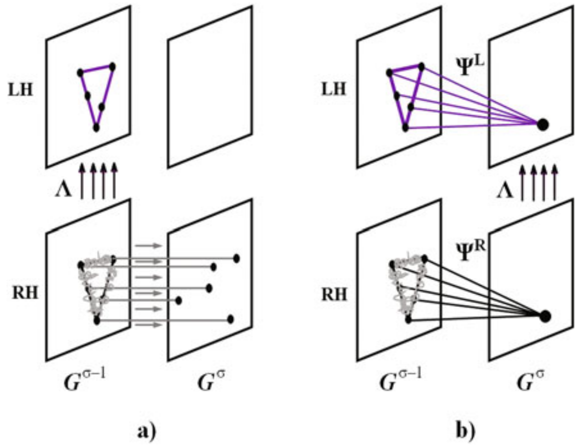

Fig. 3 Elementary learning act: **a** the "image" is created in RH at the level σ and transferred to LH; **b** its "symbol" is formed in RH (with its semantic inter-plate connections Ψ^{R}) and transferred to LH

where $\Psi_{ik}^{R\sigma}$ are semantic connections between *k*-th symbol at *σ*-th level and its "image" neurons at the level $\sigma - 1$, M_{k}^{σ} is the number of connections, Ψ^{thr} being the connection's threshold value (considered as the model parameter); summation proceeds over each *i*-th neuron that belong to the given image. After the condition (7) is fulfilled, the process of symbol formation is completed. Then the inter-subsystem connections Λ^{RL} should switch on to transfer it to the LH, where semantic connections $\Psi_{ik}^{L\sigma}$ are trained by the Hopfield-type rule, as in Eq. (3). Thus, the elementary learning act (or the "cog" according to terminology of Anokhin [5]) is accomplished.

This elementary act is reproduced ("replicated") at every level of the architecture, being a "*cell*" of the whole architecture presented in Fig. 2. In physics, this type of organization is called *scaling,* and the whole structure is described by the term *fractal.*

3.4 Functions of Processors at Different Levels of Hierarchy

The hierarchy levels of the architecture presented in Fig. 2 perform different functions.

Image ("Visual") Information The lowest level σ = 0 is represented by the Hopfield-type plates containing the *image* information. The plate H^{0} in RH should carry the *whole* image information received by the "receptors" of a given system. This plate is responsible for *recording new images* (learning). The plate H^{typ} in LH contains images selected for storage (memorization); they are called *typical images* and play main role in recognition of already known objects.

Symbolic (Semantic) Information The next level $\sigma = 1$ is occupied by the *symbols of typical images*, which are formed in RH and transferred to LH. Each symbol does carry a *semantic* content which consists in the *decomposition* into its image. At the same level, the symbols could cooperate and create the *generalized images* ("image-of-symbols") in RH, which acquire their own symbols at the next level $\sigma + 1$. After formation of "black" semantic connections this symbol are to be transferred to LH.

Standard Symbol (Words) At the middle levels $\sigma > 1$, internal symbols (image's "name") in RH are to correlate with external standard conventional names (*words*) S_W, which are obtained as *external symbolic information* (see Fig. 2) directly by LH. After establishing the correspondence between internal and external meaning of the words, the system could express, as well as understand the external symbolic information. Thus, this very part of the scheme should embrace the Broca's area (expressing words) and Wernicke's area (understanding words).

Abstract Information At the higher levels $\sigma \gg 1$, the *abstract information* emerges. This means the infrastructure of symbols and their connections, which are not mediated by real images, i.e., the neuron-progenitors of Hopfield-type plates. Here, the generalized images are converted into the *symbol-concepts* $\boldsymbol{S_C}$ that could not be related to any concrete object (e.g., *conscience*, *infinity*, *beauty*, *number*, etc.). This information appears in the already trained system as a result of interactions of all the plates and present not the "perceptible", but "deduced" knowledge.

Thus, the system as a whole does *grow up* from the lower *image* information levels, over *semantic* information (understandable for a given individual system only), to higher levels of *abstract* information which could be verbalized and *propagated* within given society. Note that the same evolution process is typical for human beings in course of its ontogenesis.

It is important to stress that at each stage of new level formation, some information is not delivered to LH, but remains in RH as the *latent* (*hidden*) individual information for a given system. This very information could be interpreted as an *intuition* (see [15, 17]).

Note that the plates (processors) in Fig. 2 could be arranged not in parallel, but sequentially along some surface. Then the "functional zones" of NCA-architecture represents the "mirror reflection" of that presented in Fig. 1b.

4 Representation of Emotions

4.1 The Problem of Formalization of Emotions in AI

Incorporating the emotions into artificial cognitive system represents really the challenge (see [30] and refs. therein), since there is the same “explanatory gap”. From the mind viewpoint (psychology), they represent *subjective self-appraisal* of the current/future state. From the “brain viewpoint”, emotions are associated with objective and experimentally measured *composition of neural transmitters*, which is controlled by the sub-cortical structures (thalamus, basal ganglia, *amygdala*, etc., see Fig. 1) [7, 8]. All the variety of neurotransmitters can be sorted into two groups: the *stimulants* (like *adrenalin*, *caffeine*, etc.) and the *inhibitors* (*opiates*, *endorphins*, etc.). Under NCA, this factor could be accounted for by incorporating additional *aggregated variable* $\mu(t)$ to be the effective difference between the stimulants and inhibitors.

In psychology, the self-appraisal (emotion) is ordinarily associated with achieving a certain *goal*. Commonly, they are classified as positive and negative ones, with increasing probability of the goal attainment leads to positive emotions, and vice versa. Furthermore, any *new* (*unexpected*) thing/situation calls up *negative* emotions [12], since it requires additional efforts to hit the new goal.

According to DTI, emotions could be classified as *impulsive* (useful for generating information) and *fixing* (effective for reception). Since the process of generation requires the noise, it seems natural to associate impulsive emotions (*anxiety, nervousness*) with the *growth of noise amplitude.* Vise versa, fixing emotions could be associated with *decreasing* noise amplitude (*relief*, *delight*). Defining the living-organism goal as a *homeostasis* (calm, undisturbed, stable state), one may infer that, speaking roughly, this classification could correlate with negative and positive emotions, respectively.

4.2 Main Hypothesis on Emotion Representation in AI

Based on these reasons, one could propose the following set of hypothesis:

Proposition 1 *The influence of neurotransmitters could be accounted for by the system of equations that link the noise amplitude $Z(t)$ with the aggregated variable $\mu(t)$ that represents virtual composition of neurotransmitters (stimulants minus inhibitors).*

Proposition 2 *The emotional reaction of human beings could be interpreted in AI systems as the time derivative of the noise amplitude, i.e., $dZ(t)/dt$. The absolute value of derivative dZ/dt corresponds to the degree of emotional manifestation:*

drastic change (jump) in Z(t) imitates either panic (dZ/dt > 0), or euphoria (dZ/dt < 0), and so on. Note that this value could be either positive, or negative that could be (very roughly) related to negative and positive emotions, respectively.

Proposition 3 *The same derivative should control the "dialog" between hemi-systems: increasing Z(t) (negative emotions) corresponds to activation of RH, while decreasing Z(t) (positive emotions) switches on LH activity.*

4.3 Equations for Coupling the Activity of Cortex and Subcortical Structures

The equations representing these propositions could be written in the form:

$$\frac{dZ(t)}{dt} = \frac{1}{\tau^Z} \cdot \{a_{Z\mu} \cdot \mu + a_{ZZ} \cdot (\mathrm{Z} - \mathrm{Z}_0) + F_Z(\mu, Z) - \hat{X}\{\mu, G_k^{R,\sigma}, \Psi\} + [\chi(\mu) \cdot D - \eta(\mu) \cdot \delta(t - t_{D=0})]\}, \tag{10}$$

$$\frac{d\mu}{dt} = \frac{1}{\tau^\mu} \cdot \{a_{\mu\mu} \cdot \mu + a_{\mu Z} \cdot (\mathrm{Z} - \mathrm{Z}_0) + F_\mu(\mu, Z)\}, \tag{11}$$

$$\Lambda(t) = -\Lambda_0 \cdot th(\gamma \cdot \frac{dZ}{dt}), \tag{12}$$

where φ, a, χ, η, τ are model parameters, the functional $\mathrm{X}\{\mu, \mathrm{G}_\mathrm{k}^{\mathrm{R},\sigma}, \Psi\}$ refers to the process of new symbol formation presented in Fig. 3 (which should decrease $Z(t)$ value). Linear in Z and μ part in Eqs. (10), (11) provides the system's homeostasis: stationary stable state corresponds to $\{\mathrm{Z} = \mathrm{Z}_0, \mu = 0\}$. The functions $F_Z(\mu, \mathrm{Z})$ in (10) and $F_\mu(\mu, \mathrm{Z})$ in (11) are written to account for possible nonlinear effects (see [16, 22]).

The last term in (8) refers to processing the incoming information. Here, *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's negative emotions and, according to Eq. (12), leads to activation of RH: $\Lambda(\mathrm{t}) = -\Lambda_0 \equiv \Lambda^{\mathrm{L}\rightarrow\mathrm{R}}$. Vice versa, finding the solution to the problem ($D = 0$) causes rapid *decline* of Z, which corresponds to positive emotional splash and LH activation ($\Lambda(\mathrm{t}) = +\Lambda_0 \equiv \Lambda^{\mathrm{R}\rightarrow\mathrm{L}}$); then RH gets the possibility to be "at rest". Thus, the model (10)–(12) seems quite reasonable and self-consistent.

5 Conclusion

Thus, it is shown that the cognitive architecture designed in analogy to the human-brain structure really provides the possibility to interpret and reproduce the peculiarities of human cognitive process under NCA. The key point of NCA is the following: an artificial cognitive system, being a complex multi-level combination of various-type neural processors, should be divided into *two subsystems*, like human brain does (two cerebral hemispheres). It is shown that one of them should necessary contain a random element (*noise*) for generation of information (creativity); it is responsible for *learning* (an analogy to the right hemisphere). Another one should be responsible for memorization and processing the *well-known* information (after learning), by analogy with the left hemisphere; this subsystem should be free of noise. It is shown that the noise-reach subsystem could provide an *intuition*, while the noise-free one is associated with *logical* thinking. Both subsystems should be linked by the cross-subsystem connections (by analogy with *corpus callosum*). Those connections should switch on depending on the stage of solving a problem.

It is shown that the human emotions, being a subjective appraisal of the current/future state, and, simultaneously, the product of the neural transmitters, are inherently embedded into the NCA-architecture. Emotions are interpreted as *dynamical variations of the noise amplitude*, and these very variations should *control the activity of two* sibsystems. Accounting for the neurophysiology, this interpretation requires including an additional variable which corresponds to the “effective” composition of neural transmitters. The system of coupled equations on the noise amplitude and the neural-transmitter variable is proposed, which provides reasonable correlation between switching the subsystem activity and the necessity of additional stimulant (or inhibitor) production.

Under NCA, the emotions are classified as *impellent* (stimulating the *generation* of information) and *fixing* (stimulating the *memorization* process) ones. Very roughly, this classification corresponds to common *negative* and *positive* emotions, respectively. Since the generation of information requires rather high level of noise, thus the increase of the noise amplitude does simulate impellent (“negative”) emotions. Vice versa, a termination of the learning act, i.e., ertain “*skill*” acquiring, causes the decrease of noise that corresponds to the fixing (“positive”) emotions. Thus, this idea seems reasonable and deserves further elaboration.

It is shown that the experimentally observed modifications on the gene-expression level of those neurons that are involved into skill-acquiring process [11] could be reproduced under NCA by *parametric modification* of the “trained” neurons. The latter is possible only within the concept of *dynamical* formal neuron.

Let us point out that NCA actually provides the “bridge” over the gap between physiological (“brain”) and psychological (“mind”) approaches to cognitive process. The concept of *conventional information*, being based on the material connections between neurons, represents the *free choice* of the neuron ensemble as a whole.

References

1. Chella, A., Mazotti, R.: Machine consciousness: a manifesto for robotics. Int. J. Mac. Conscious. **1**(1), 33–51. ISSN 1793-8430 (2009)
2. Kushiro, K., Harada, Y., Takeno, J.: Robot uses emotions to detect and learn the unknown. Biol. Inspired Cogn. Archit. **4**, 69–78 (2013)
3. Laird, J.E.: The Soar Cognitive Architecture. MIT Press (2012)
4. Samsonovich, A.: Bringing consciousness to cognitive neuroscience: a computational perspective. J. Integr. Des. Process Sci. **11**, 19–30 (2007)
5. Anokhin, K.V.: Cognitom: theory of realized degrees of freedom in brain. In: The Report Given at Fifth International Conference on Cognitive Science, Kenigsberg, June 2012. http://vk.com/Kognitivnaja_Nauka (in Russian)
6. Yakhno, V.G.: Basic models of hierarchy neuron-like systems and ways to analysis of some their complex reactions. Opt. Mem. Neuron NetW. **4**(2), 141 (1995)
7. Koziol, L.F., Budding, D.E.: Subcortical Structures and Cognition. Implications for Neurophysiological Assessment. Springer, Berlin (2009)
8. 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)
9. Levine, J.: Materialism and qualia: the explanatory gap. Pac. Phil. Quart. **64**(4), 354–361 (1983)
10. Stirling, J., Eliott, R.: Introducing Neuropsychology, 2nd Edition (Psychology Focus). Psychology Press (2010)
11. Svarnik, O.E., Anokhin, K.V., Aleksandrov, Y.I.: Experience of a First "Whisker-Dependent Skill Affects the Induction of c-Fos Expression in Somatosensory Cortex Barrel Field Neurons in Rats on Training the Second Skill". Neurosci. Behav. Physiol. (2015). doi:10.1007/s11055-015-0135-3
12. Goldberg, E.: The New Executive Brain. Oxford University Press (2009)
13. Solso, R.: Cognitive Psychology, 6th ed. Allyn & Bacon (2006)
14. Bianki, V.L.: Parallel and sequential information processing in animals as a function of different hemispheres. Neurosci. Behav. Physiol. **14**(6), 497–501 (1984)
15. 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. Arc. **6**, 147–158 (2013)
16. Chernavskaya, O.D., Chernavskii, D.S., Karp, V.P., Nikitin, A.P., Shchepetov, D.S., Rozhylo, Y.A.: An architecture of the cognitive system with account for emotional component. Biol. Inspired Cogn. Arc. **12**, 144–154 (2015)
17. Chernavskaya, O.D., Chernavskii, D.S.: A natural-constructive approach to modeling the cognitive process. J. Biophy. **61**(1), 155–169 (2016)
18. Haken, H.: Information and Self-Organization: A macroscopic approach to complex systems. Springer, Berlin (2000)
19. Progogine, I., Stengers, I.: The End of Certainty. Free Press (1997). ISBN 978-0-684-83705-5
20. Chernavskii, D.S.: The origin of life and thinking from the viewpoint of modern physics. Phys. Usp. **43**, 151–176 (2000)
21. Haykin, S.S.: Neural Networks and Learning Machines. Prentice Hall (2009)
22. Cernavskaya, 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, 20–24 Mar 2016, Rome, Italy (in press)
23. Cernavskaya, O.D.: The Cognitive Architecture within Natural-Constructive Approach. FIERCES on BICA 2016. Springer (in press)
24. Quastler, H.: The Emergence of Biological Organization. Yale University Press, New Haven (1964)

25. FitzHugh, R.: Impulses and physiological states in theoretical models of nerve membrane. Biophys. J. **1**, 445 (1961)
26. Hopfield, J.J.: Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. **79**, 2554 (1982)
27. Hebb, D.O.: The Organization of Behavior. Wiley (1949)
28. Grossberg, S.: Studies of Mind and Brain. Riedel, Boston (1982)
29. Kohonen, T.: Self-Organizing Maps. Physica-Verlag, Heidelberg (1995)
30. Hudlicka, E.: Affective BICA: Challenges and open questions. BICA **7**, 98–125 (2014)