A NEW CONCEPTUAL UNDERSTANDING OF THE BRAIN: WHAT IS WRONG WITH CLASSICAL CONCEPT AND HOW SHOULD IT BE CHANGED?

K.V. BAEV

St. Joseph's Hospital and Medical Center, Phoenix, Arizon, USA
kbaev@chw.edu

Dr. Baev holds the following degrees: M.D. in Psychiatry, M.Sc. in Physics and Electrical Engineering, Ph.D. in Neuroscience, Dr. Sci. in Neuroscience. I am an author of 131 publications including three books.

My scientific career started in 1969 when I joined the Ph.D. program in the Department of General Physiology at Bogomoletz Institute of Physiology, Ukrainian Academy of Sciences. The Department and the Institute were led by Prof. P.G. Kostyuk. This is where I learned how to use the microelectrode technique to study the brain.

After finishing the Ph.D. program in 1972, I became interested in neural control of inborn automatic motor behaviors like locomotion and scratching. Understanding the Central Pattern Generators (CPGs) for these movements was the primary goal of the research. I believed that solving the CPG problem would be a breakthrough which would help to explain not only inborn automatic behaviors, but learned automatic behaviors as well, which are components of any complex behavior in animals and humans.

At that time, and it is still mostly the case today, to understand a function of any neural center meant to superimpose two sets of knowledge, namely the knowledge of neuronal schematics and of the behavior of the corresponding neuronal elements during function. This approach is based on the classical conceptual understanding of the brain which can be found in any textbook of neurobiology. The microelectrode technique is the primary tool for this approach. In the 1970s, it was found that there was no plan of generator construction both in invertebrates and vertebrates, and my belief in classical approach to solve the CPG problem was completely lost.

At the beginning of 80s, I decided to concentrate my research on interaction between spinal CPGs and afferent flow with a hope that the CPG problem can be solved at a functional level. The approach was based on my earlier discovery (1978) that spinal CPGs for locomotion and scratching modulate primary afferent depolarization (PAD). Then I was the head of the Department of Physiology of the Spinal Cord and a deputy scientific director of Bogomoletz Institute and could dedicate more resources to this approach, which appeared to be a direct hit and to bring the solution of CPG problem by the end of 80s. It was found that CPG possesses a model of the behavior of its controlled object and uses this internal model to produce motor pattern after deafferentation.

After solving the CPG problem, my primary interests were completely within the realm of theoretical neuroscience. Theoretical explanation of functions of the highest brain levels, mechanisms of symptoms of Parkinson's disease, and a new brain concept are the results of my theoretical research. The research was conducted at St. Joseph's Hospital and Medical Center where I have been working since the beginning of 1992, first as a visiting professor and later as a senior staff scientist.

Introduction

The importance of conceptual views of the brain is hard to overestimate. They ultimately determine our understanding of the relationship between structure and function and, consequently, problem formulation, and how we should conduct research, both theoretical and experimental, including our strategies of analysis and synthesis.

Two major discoveries led to the new brain concept: (i) The discovery of the neural network computational principle. (ii) The solution of Central Pattern Generator (CPG) problem. Both happened in the 1980s, that is, quite a while ago.

The essence of the classical conceptual understanding of the brain consists in finding within the brain's neural network specific neural circuits that control corresponding forms of behavior. Search for specific reflex arcs or CPGs for automatic movements like locomotion or scratching are typical examples of implementation of the classical concept.

A rather simple and straightforward understanding of the relationship between structure and function which is applicable only to very simple controlling systems like thermostat or refrigerator lies at its core. It is a superposition of neuronal schematics and the behavior of the corresponding neuronal elements during function. Only by acquiring some special notions from neurocomputing and control theory we can change this understanding of the structure-function relationship in the brain. For a curious mind, these classical views create more questions than answers.

Several major questions are listed below:

• Is it correct to search for specific neural circuits within the brain's mega-network to explain behaviors?

• Why can the same neurons be part of different specific circuits?

• What is the role of divergence and convergence, and what is the purpose of parallel circuits?

• What is the role of dendritic trees?

• Is it adequate to use diagrams in which a functionally similar group of neurons is represented as one element to explain functions or computer simulations?

• Why are biological neural networks built in the way they are? Classical neuroscience does not address this most fundamental question. There is not even room for this question within the framework of classical conceptual brain views.

The classical views did not produce the desired results. The application of classical strategy to solve the CPG problem failed. It is unclear how to apply the strategy of selecting specific circuits to the highest brain levels? Should we choose a specific circuit for each specific behavior? The neural network computational principle shows why classical brain concept is inadequate to analyze network controlling systems like the brain.

The neural network computational principle and its corollaries

The understanding that biological neural networks perform computations came from neurocomputing in the 1980s. Finally, it became clear that the converging and diverging connections within a network play a crucial role in the computations. Let us consider the so-called three-layer Kolmogorov's unidirectional neural network (Fig. 1). This mathematical construction includes three layers. The processing elements (***n***) of the first layer are fan-out units that distribute input signals, the input vector components (**x**), to the processing elements of the second hidden layer. The processing elements of the hidden layer neither directly receive inputs from nor provide direct outputs to the external world. The transfer function (the rule that transforms input signals into output signals) of these units is similar to a linear-weighted sum. The output processing elements (***m***) of the third layer send signals to the external world, i.e., output vector (**y**). The transfer function of output units is highly nonlinear. A theorem was proven by

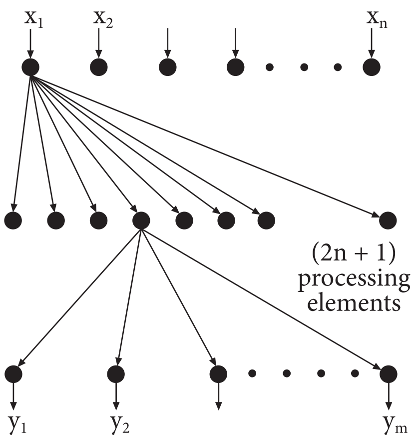

Fig. 1. Architecture of Kolmogorov's neural network. (See more explanation in the text)

Kolmogorov that such a three-layer neural network can implement any continuous mapping function, if its synaptic weights are adjusted properly (Kolmogorov, 1957). The analogy between Kolmogorov's construction and biological neural networks is clear. For example, input, hidden layer, and output elements correspond to sensory neurons, interneurons, and motoneurons, respectively. This network performs computations on input signals and generates corresponding motor control output. Kolmogorov's theorem is an existence theorem. Later numerous other theorems were proven which showed how to construct such a network for practical needs.

The neural network computational principle demonstrates how the same network can implement different functions. The principle has numerous corollaries. The most important of them are listed below:

• Selecting a specific circuit from the biological neural network; an approach used by classical neuroscience to explain behaviors like reflexes, program controls, etc., is an unjustifiable oversimplification which does not provide an adequate explanation of the corresponding behavior.

• Computational models based on classical serial circuit diagrams are also inadequate.

• The distribution of synaptic weights across any specific network has never been addressed by classical neurobiology.

• The network computational principle changes the understanding of the structure-function relationship. As mentioned earlier, the classical understanding of the structure-function relationship calls for the synthesis of detailed knowledge of connections between interacting neurons, their properties, and their corresponding behavior. As we can see, the result of such synthesis has very little to do with explaining the function computed by the network under study.

• The network computational principle is necessary but not sufficient to solve the problem of the structure-function relationship for the brain. At first glance, the network computational principle is the necessary missing link. One might assume that now, when we finally understand what biological neural networks *are doing*, we will be able to successfully unveil their functions. However, further analysis shows that the network computational principle is necessary but not sufficient to solve the problem of the structure — function relationship for the brain. The network computational principle taken alone "makes" the problem of the structure-function relationship even more complex. It does not show an alternative to the classical conceptual understanding of the brain and, consequently, of the structure-function relationship in it. The problem is complicated by the fact that the same function can be computed by different networks and that the same network can compute different functions. In addition, dendritic processes during behavior remain unobservable or incompletely observable. At the same time, most of the information processing, especially in vertebrates, takes place at the level of the dendrites where the predominant number of synapses are located. The way to complement the network computational principle by using a functional approach, understanding structure through function, is described below.

The solution of central pattern generator problem and a notion of a generic neural optimal control system

The solution of CPG problem came from functional approach, from studies of interaction of spinal CPGs with afferent flow (Baev et al. 1991a; Baev et al. 1991b; Baev and Shimansky, 1992). Experiments have demonstrated that hindlimb CPGs for locomotion and scratching in cats possess a model of object behavior. From the control theory point of view, this means that the controlling system is an optimal one and the internal model is the result of its optimality. Hindlimb CPGs for locomotion and scratching share the same controlled object and spinal circuitry. Therefore, a CPG is a regime of work of a neural optimal control system (NOCS).

The following logical conclusions are obvious:

• The spinal NOCS, which controls the hindlimb locomotor and scratching movements, also controls all of its other movements, that is, it is multifunctional.

• Other body parts are also controlled by NOCSs.

• The brain is a hierarchical system of interacting NOCSs. Hierarchically higher control levels are also NOCSs. Higher NOCSs also contain models of their controlled objects behavior, which are lower than NOCSs.

• Non-motor neuronal automatisms are also controlled by NOCSs.

A generic NOCS includes two major functional units, namely a controller and an internal model of the dynamics of the controlled object (CO) (Fig. 2). The controller utilizes information about the current state of the CO to compute the control signal that will move the CO from its initial point to its destination point along the optimal trajectory. The internal model is a predictive mechanism. At any given moment, it predicts the next most probable state of the CO after the CO receives the controlling signal from the controller. This internal source of afferent information, expected or model afferent flow, interacts with actual afferent flow in a rather unusual way, and this interaction is far from trivial. The model afferent flow is treated by the NOCS and hence by the CPG as a component of actual afferent flow (Baev et al. 1991a, 1991b). This type of interaction between model and actual afferent flows was defined as a parity interaction. Both flows produce primary afferent depolarization in the spinal cord. This mechanism enables the NOCS to pay the highest attention to the most active informational channels. Silent or low active channels are discarded by this mechanism and hence are considered as nonreliable sources of information by the controlling system. This explains the holographic properties of the afferent system, that is, why after partial or complete limb deafferentation, the CPG retains the ability to generate rhythm. After complete limb deafferentation, the control system relies solely on the information provided by the internal model. Therefore, the generation of a rhythmic

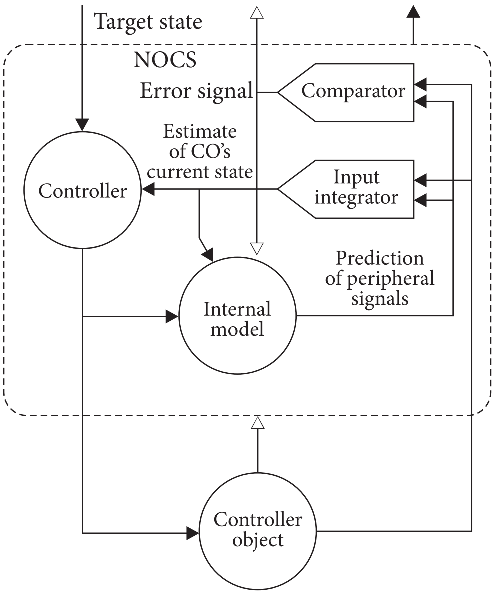

Fig. 2. Major functional blocks of an NOCS. Informational and initiating signals are shown by black and white arrows, respectively. See text for further detail

motor pattern by a CPG in the absence of peripheral afferent feedback is a reverberation not between flexor and extensor half-centers as suggested by classical views, but rather within the loop which includes the internal model (Fig. 2).

Any NOCS receives two types of afferent signals: initiating signals and signals that contain the current informational contex*t* in which the system as a whole finds itself. There is only one difference between these two types of afferent signals. An NOCS attempts to minimize initiating signals, more exactly, to minimize the integral measure of initiating signals by using informational signals to compute proper output. Initiating signals are analogous to "energetic" signals. They activate NOCSs resulting in the realization of any NOCS's corresponding automatism. An automatism is a program of action or sequential actions that is stored in the NOCS and, during execution, leads to a minimization of its corresponding initiating signal. Withdrawal or escape reactions and locomotion are typical examples of automatisms that serve to avoid danger, i.e., to minimize the initiating signal that signifies danger. Reflexes, being automatisms, usually serve to minimize certain initiating signals. However, often the final goal of an automatism is not minimization but maximization of the initiating signal, for example, during pleasurable reactions, when the corresponding signals are maximized. This situation poses no contradiction, because the nervous system can easily invert the sign of the signal so that the latter becomes maximal instead of minimal. It should be noted that appropriate segregation of afferent flow into these two types of signals can be only made at the level of the recipient NOCS.

The brain is a hierarchical neural optimal control system

A simple algorithm can be used to analyze any neural control system based on the proposed conceptual views. For each hierarchical level, both functional subsystems — controller and model — and the controlled object must be identified, and the various types of afferent signals — informational and initiating — must be determined. Special attention should be paid to error distribution systems. This algorithm was successfully used to reach several important conclusions about the functions of the cerebellum, and the highest levels like cortico — basal ganglia thalamocortical loops. Surprisingly, these conclusions were reached without detailed knowledge of the computational abilities of underlying neural networks (see Baev, 1997, 1998; Baev, 2009).

Conclusions

Obviously, the brain physiology and the pathophysiology of brain disorders should be created anew, if NOCSs are considered functional blocks of the brain. Pathology in any functional block of NOCS may lead to the manifestation of brain pathological symptoms. The new brain concept has been successfully applied to explain Parkinson's disease, that is, how the lack of dopaminergic neurons is translated into motor symptoms (Baev, 1995, 1997, 1998, 2009).

Therefore, the new brain concept qualifies for a unifying brain theory capable of explaining brain functions of various complexities, from the simplest ones to functions of the highest brain levels. Within the framework of the new brain concept, a neuronal structure can be understood only through its function.

Finally, a conclusion can be made that the concept is applicable to other network systems such as molecular, ecosystems, and even social ones.

It will not be easy to change the classical brain views that have been in circulation for generations. Extraordinary educational measures are required to accelerate broad acceptance of the new brain concept. Read my recent article (Baev, 2009) or visit http://www.neurallaws.com for additional information.

REFERENCES

Baev, KV, 1995. Disturbances of learning processes in the basal ganglia in the pathogenesis of Parkinson's disease: A novel theory. Neurol Res. 17, 38-48.

Baev, KV, 1997. Highest level automatisms in the nervous system: A theory of functional principles underlying the highest forms of brain function. Progr Neurobiol. 51, 129-166.

Baev KV, 1998. Biological neural networks: Hierarchical concept of brain function, Birkhauser, Boston.

Baev KV, 2009. A new conceptual understanding of brain function: Basic mechanisms of brain-initiated normal and pathological behaviors. Crit Rev Neurobiol 19(2-3): 119-202.

Baev KV, Esipenko VB, Shimansky YuP, 1991a. Afferent control of central pattern generators: Experimental analysis of locomotion in the decerebrate cat. Neurosci. 43, 237-247.

Baev KV, Esipenko VB, Shimansky YuP, 1991b. Afferent control of central pattern generators: Experimental analysis of scratching in the decerebrate cat. Neurosci. 40, 239-256.

Baev KV, Shimansky YuP, 1992. Principles of organization of neural systems controlling automatic movements in animals. Progr Neurobiol. 39, 45-112.

Kolmogorov AN, 1957. On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. Dokl. Akad. Nauk USSR (in Russian). 114, 953-956.