Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 145 (2018) 623–628

Procedia Computer Science

www.elsevier.com/locate/procedia

Postproceedings of the 9th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2018 (Ninth Annual Meeting of the BICA Society)

Cognitive architecture based on the functional systems theory

Vityaev E.E.*, Demin A.V.

Sobolev institute of mathematics SD RAS, Novosibirsk 630090, Russia

Abstract

In this paper the cognitive architecture based on the Functional Systems Theory (TFS) by P.K.Anokhin is presented. This architecture based on the main notions of this theory: goal, result, anticipation of the result. This theory is described on physiological and informational level. The logical structure of this theory was analyzed and used for the control system of the purposeful behavior development. This control system contains the hierarchy of functional systems that organize the purposeful behavior. The control system was used for the agents modeling that are solving the foraging task.

Keywords: Type your keywords here, separated by semicolons ;

1. Introduction

The theory of functional systems (TFS) developed by P.K. Anokhin and many other distinguished scientists of his school [1] is, at the moment, one of the few known theories in which the concepts of goal, purpose, result, and goal-directed activity are principal ones and which exposes the physiological mechanisms that implement these concepts.

A goal cannot be attained without having a criterion of its attainment; otherwise we can always assume that the goal has been already attained. The definition of goal is paradoxical (produces a goal paradox), since it does not imply knowledge of how, by what means, and when can it be attained. A definition of the goal allows us to define

* Corresponding author. Tel.: +7-913-945-4867.
E-mail address:vityaev@math.nsc.ru

10.1016/j.procs.2018.11.072

the result, attained through the goal, as that which meets some criterion. We drink water to quench thirst, eat to ease hunger, etc. Actions are always goal-directed. If an action does not have a goal, it is unclear when and how it should be terminated. The goal of an action is to change the current state and/or some external stimulus. Goal-directed activity consists in organizing the activities that will satisfy some of the organism's needs.

Desire is not passive. It makes no sense to desire if there is no possibility to get closer to satisfying the desire by some actions or activity. Desire is active, but meaningless without purposefulness – it causes the organism to be active and display some behavior in order to satisfy it. Thus the concept of goal emerges. Activity and actions are always goal-directed. If there is no goal for an action, it is unclear when (and by what means) it should be terminated. The goal cannot be attained without having a criterion of its attainment; otherwise we can always assume that it has already been attained.

Let us define the goal as such an activity/behavior that is aimed at satisfying certain criteria. Such a definition of goal allows us to define the result of attaining the goal as that, what we obtain by meeting the criterion and attaining the goal (fulfilling the desire). Between the concepts of goal and result, the following relationship holds: the result is obtained when the goal is attained and the criterion of its availability is "triggered". But when the goal is being set, we have the goal but not the result.

The definition of goal is paradoxical since the activity/behavior of satisfying some criteria does not essentially presuppose knowledge of how to attain the goal; you can set a goal without defining either how it can be attained, or by what means, or when. This paradoxicality of the goal concept we call the goal paradox. For the paradox solution we need an experience.

As will be seen latter on, in the framework of the theory of functional systems, brain activity during goal-directed behavior is seen as being constantly occupied by solving the goal paradox, and determining by what means, when, and how to attain goals.

Let us proceed to outline the theory of functional systems, in which the concepts of goal, result, and goal-directed activity are principal ones, and which analyzes the physiological mechanisms of these concepts.

2. The Theory of Functional Systems of Brain Function

The theory of functional systems (TFS) is a theory of systems, whose function is to attain goals (satisfy needs) by solving the goal paradox. Therefore, we will outline the theory of functional systems as a theory of solving goal paradoxes, and describe how the brain determines by what means, when, and how goals can be attained.

As achieving results consists in satisfying some criteria, this achievement should be registered in some way. In the physiological sense, what constitutes a criterion for registering the attainment of a result? According to P.K. Anokhin, this is physiologically realized by a "special receptor apparatus" [2]. The signaling of this receptor apparatus about obtaining a result (i.e., on the lack of deviation from the optimal level of metabolism) and attaining the goal is called reverse afferentation.

Now we can explain, within the framework of TFS, how goals are being physiologically set by the organism. An organism need constitutes a goal in TFS. The goal (and its attainment criterion): firstly, signals by means of reverse afferentation that there is a lack of some need; secondly, it sets a goal to wait for a signal, indicating that the results have been attained; and thirdly, it provides energy and actually forces the organism to attain the goal. Thus, the physiological mechanism of goal-setting in fact consists of the emergence of a need.

The interaction of different goals and results is organized in several ways according to TFS: by the "principle of the dominant", by the "hierarchy of results" and by "result models".

3. Central mechanisms of functional systems

"According to P.K. Anokhin, the central mechanisms of functional systems that support goal-directed behavioral acts have a similar structure" [2]. Let us examine in detail the architecture of goal-directed activity, as well as the physiological mechanisms of solving the goal paradox.

Afferent synthesis. The afferent synthesis, which includes the synthesis of motivational excitation, memory, contextual and triggering afferentation, constitutes the initial stage of a behavioral act of any complexity. ***Motivational Excitation***. As we know, the goal is set by an emerging need. But in case of goal-directed behavior, it

transforms into a motivational excitation. But a motivational stimulus does not consist in the excitation of receptors which stand 'on guard' for some physiological constant – it is rather an excitation of 'central brain structures' initiated by the arising need. It is the motivational stimulus that constitutes the goal set in the organism in case of goal-directed behavior. As in the case of needs, the motivational stimulus not only sets a goal but also energetically supports its attainment. ***Memory***. The whole sequence of stimuli that has led to goal attainment is recorded during reinforcement, starting with the motivational stimulus. Motivational stimulus extract from memory all previous sequences of actions which have led to attaining the result. ***Situational Afferentation***. While recording a memory trace, the situation in which the result is attained is also being recorded. This situation is registered, along with the motivation, as a necessary precondition for attaining the result. Thus, the motivational stimulus in this situation "extracts from memory" only those ways of attaining the goal that are possible in the given situation. ***Triggering Afferentation***. The fourth component of afferent synthesis is the triggering afferentation. It is essentially the same as the situational afferentation with the difference that it involves the time and place of attaining the result.

Consequently, the goal paradox is solved for the most part during afferent synthesis, as it's here that the what, how, and when of goal attainment are determined. Therefore, taking experience and environment into account, the motivational excitation as a goal automatically solves the goal paradox and determines by what means, how, and when can the goal be attained.

Decision-making. At the stage of afferent synthesis, motivational excitation can extract from memory several ways of attaining the goal. At the stage of decision making, only one of them is selected – thus forming the "program of actions".

Acceptor of Action Results. Suppose a program of actions is chosen. At that point, there is no guarantee yet that the final result will necessarily be attained, nor even intermediate ones. The goal can only be attained if each of the intermediate results of the current program of actions will be attained. Motivational excitation "extracts from memory" the entire sequence and the hierarchy of results that should be attained during the program of actions. This sequence and hierarchy of results are defined in TFS as the ***acceptor of action results***. Therefore, while being transformed into a particular goal, the motivational excitation extracts from the memory a particular criteria of this goal attainment. This consists of the whole sequence and the hierarchy of criteria of results which must be attained in the process of attaining the goal and performing the program of actions, i.e. the acceptor of action results. Thus, the acceptor of action results anticipates the particular criteria of attaining the goal.

Transforming motivational excitation as a goal into a particular goal, transforms the original paradoxical goal, for which it is not determined by what means, how, and when it can be attained, into a non-paradoxical particular goal, for which the final goal (and result) is divided into sub-goals (and sub-results), so that for each sub-goal it is already known by what means, how, and when it can be attained.

4. Formal model of TFS

Now let us assume that our model constitutes the control system of some animat that operates in discrete time $t = 0, 1, ...$ as it was done in [3]. Suppose the animat has a set of sensors $S_1, ..., S_n$, which characterize both the state of the animat itself and of external environment. Each sensor S_i has a set of possible indications VS_i. The animat also has a set of available actions in the environment $A = \{a_1, ..., a_m\}$. Any action that animat performs at a moment t_i may result at a moment $t_i + 1$ in some changes in the environment, and consequently, in his sensors indications. Since the animat «perceives» the world only through its sensors, then from its point of view the system's state at any given point in time can be written as a vector of all sensors indications $V(t) = (v_1, ..., v_n)$, where $v_i \in VS_i$ is the indications of the the *i*-th sensor at the moment *t*, and the states with same sensor indications are indistinguishable for it. The set of all possible states of the system is denoted by $S = (VS_1 \times VS_2 \times ... \times VS_n)$.

On a set of states of the system $S = (VS_1 \cup VS_2 \cup ... \cup VS_n)$ we define a set of predicates $PS = \{P_1, ..., P_k\}$, each of which is calculated on the basis of sensors indications. Each state of the system can thus be written as a vector $s = (p_1, ..., p_k),\ p_i \in \{0,1\}$ of predicates values from PS, where 1 means validity of a predicate, and 0 – its falsity. The state may be described by a subset of predicates $s = (p_{i_1}^e, ..., p_{i_e}^e),\ p_{i_1}^e, ..., p_{i_e}^e \subseteq p_1, ..., p_k$. The animat's task is to

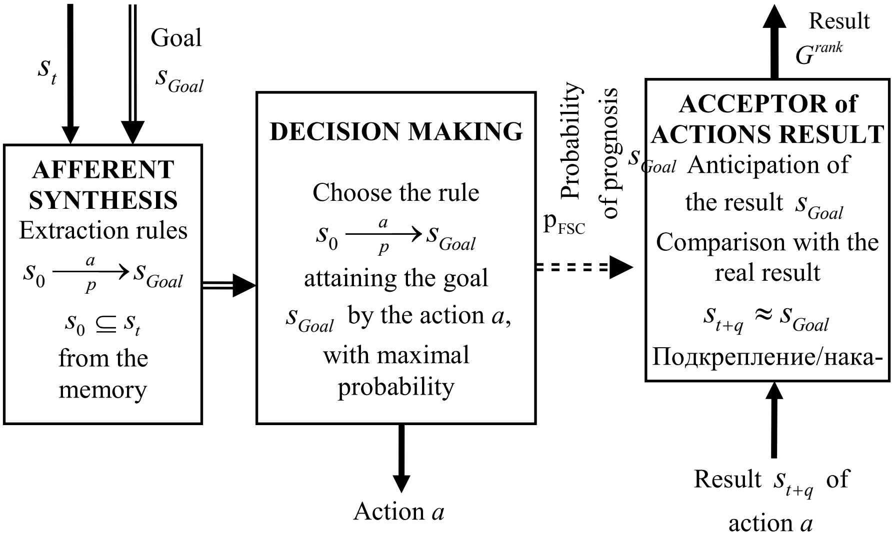

Fig. 1. functional system that realize sensory corrections.

attain a certain goal. Let us define a goal *Goal* as a state of the system $s_{Goal} = (p_{i_1}^{goal}, ..., p_{i_{goal}}^{goal})$, which it is required to attain. A notation $(p_{i_1}^{goal}, ..., p_{i_{goal}}^{goal})$ means that predicates $p_{i_1}^{goal}, ..., p_{i_{goal}}^{goal}$ should be true when the goal is attained.

Let us clarify concepts of event and history. By an event $e = (s_0, s_e, a)$ we will understand a singular fact of transferring the system from the state $s_0 = (p_1^0, ..., p_k^0)$ into a state $s_e = (p_1^e, ..., p_k^e)$ as a result of an action a, and by a history of events – a set of pairs (e_t, t), where $e_t = (s_t, s_{t+1}, a)$ is an event and *t* is a moment in time when this event has occurred.

Let us define a rule *R* that predicting a change of a state(s) after the execution of an action *a*, as a transformation $R = (s_0 \xrightarrow[p]{a} s_e)$, where: s_0 – is the initial state of the system $(p_{i_1}^0, ..., p_{i_0}^0)$; s_e – is the final state of the system $(p_{i_1}^e, ..., p_{i_e}^e)$; a – is an action that transforms the initial state into the final one; p – is the probability with which the action transforms the initial state into the final one.

Let us first define a functional system $FSC = (s_{Goal}, R_1, ..., R_n, p_{FSC})$ that realizes one action. Functional system FSC performs transformations $s_0 \xrightarrow[p_{FSC}]{R_1, ..., R_n} s_{Goal}$, where $s_{Goal} = (p_{i_1}^{goal}, ..., p_{i_{goal}}^{goal})$ – is the target state of the functional system, $R_1, ..., R_n$ – are rules of the form $s_0 \xrightarrow[p]{a} s_{Goal}$, using which from various initial states s_0 and some action *a* the system can get to the target state s_{Goal} (fig. 1). An estimate of the probability of attaining a goal by a functional system can be calculated based on the statistics of attaining goals: if *n* is the number of cases in which a request to attain a goal s_{Goal} was received and *m* is the number of cases in which the selected rules and sequences/hierarchies of actions led to attaining the goal s_{Goal}, then $p_{FSC} = m/n$.

In general case functional systems are sequences and the hierarchies of the functional systems FSC. A functional system $FS = (s_{Goal}, \langle FSC_1, ..., FSC_n \rangle, p_{FS})$ that combines a sequence of functional systems of the form FSC is defined as:

$$FS = s_0 \xrightarrow[\to s_1 \to s_2 \to ... \to s_{goal} \quad p_{FS} = p_{FSC_1} \bullet ... \bullet p_{FSC_n}]{FSC_1, ..., FSC_n} s_{goal}, \text{ where}$$

$$FSC_1 = (s_0 \xrightarrow[p_{FSC_1}]{R_1^1, ..., R_{v_1}^1} s_1),\ FSC_2 = (s_1 \xrightarrow[p_{FSC_2}]{R_1^2, ..., R_{v_2}^2} s_2)\ ...\ FSC_n = (s_{n-1} \xrightarrow[p_{FSC_n}]{R_1^n, ..., R_{v_n}^n} s_{goal})$$

are functional systems of the type FSC. The goal of the functional system FS is to successively attain goals $s_1 \to s_2 \to ... \to s_{goal}$ using functional systems $FSC_1, ..., FSC_n$ with a resulting probability $p_{FS} = p_{FSC_1} \bullet ... \bullet p_{FSC_n}$.

Functional systems FS that attain some goal stated by motivation include different sequences and hierarchy of

functional systems FSC. These sequences and hierarchy may be described by sequences of functional systems FS including both functional systems FS, FSC recursively. Then functional system $FS = (s_{Goal}, \{\langle FS_1^1, ..., FS_n^1 \rangle\}, p_{FS})$ is a set of sequences/hierarchies of functional systems that realize transformations

$$FS = s_0 \left\{ \xrightarrow[\to s_1 \to s_2 \to ... \to s_{goal} \quad p_{FS} = p_{FS_1^1} \bullet ... \bullet p_{FS_n^1}]{FS_1^1, ..., FS_n^1} \right\} s_{goal},$$

where FS_i^1 is either FS or FSC. For example, if $FS_i^1, FS_j^1 \in \langle FS_1^1, ..., FS_n^1 \rangle, i < j$ realizes transformations

$$FS_i^1 = s_{i-1} \left\{ \xrightarrow[\to s_1^i \to s_2^i \to ... \to s_{goal}^i \quad p_{FS_i^1}]{FS(i)_1^2, ..., FS(i)_{n_i}^2} \right\} s_i, \quad FS_j^1 = s_{j-1} \left\{ \xrightarrow[\to s_1^j \to s_2^j \to ... \to s_{goal}^j \quad p_{FS_j^1}]{FS(j)_1^2, ..., FS(j)_{n_j}^2} \right\} s_j,$$

then functional systems $FS(i)_1^2, ..., FS(i)_{n_i}^2, FS(j)_1^2, ..., FS(j)_{n_j}^2$ are of the level 2 and the transformation realized by the functional system FS has the form

$$FS = s_0 \left\{ \xrightarrow[\to s_1 \to s_2 \to ... \to s_{i-1} \{\langle \to s_1^i \to s_2^i \to ... \to s_{goal}^i \rangle\} ... \to s_{j-1} \{\langle \to s_1^j \to s_2^j \to ... \to s_{goal}^j \rangle\} ... \to s_{goal} \quad p_{FS}]{FS_1^1, ... FS_{i-1}^1 \{\langle FS(i)_1^2, ..., FS(i)_{n_i}^2 \rangle\}, ..., FS_{j-1}^1 [FS(j)_1^2, ..., FS(j)_{n_j}^2], ..., FS_n^1} \right\} s_{goal}.$$

In accordance with the theory of the movement organization by N.A. Bernstein (1997), the leading level of the organization of movements is the top level $FS = (s_{Goal}, \{\langle FS_1^1, ..., FS_n^1 \rangle\}, p_{FS})$ of rank 1, which corresponds to the domonant motivation. Functional systems of the lower levels are activated by the functional systems of the upper levels.

When functional system FS receives a request to attain a goal s_{Goal} in the current state $s_t = (p_1^t, ..., p_k^t)$ it:

1) extracts from the memory all sequences/hierarchies of the functional systems FSC that are applicable in the current situation $s_0 \subseteq s_t$, where s_0 - is the first state of the sequence/hierarchy (afferent synthesis);
2) from all these sequences/hierarchies it choose one that attain the goal s_{Goal} with the maximum probability p_{FS} (decision making);
3) generate a «specific goal» (the highest motivation) as a sequence and hierarchy of all goals of functional subsystems contained in the chosen sequence/hierarchy. For example, for the above mentioned functional system, it would be a sequence

$$s_0 \to s_1 \to s_2 \to ... \to [\to s_1^i \to s_2^i \to ... \to s_i] ... \to [\to s_1^j \to s_2^j \to ... \to s_j] ... \to s_{goal};$$

3) predicts that the goal s_{Goal} will be attained with probability p_{FS};
4) anticipates (using acceptor of action results) the achievement of the entire sequence/hierarchy of goals for all its functional subsystems after the corresponding actions;
5) successively activates execution of actions in functional subsystems FSC;
6) if the goal s_{Goal} is not attained by some functional subsystem and this fact is registered by the acceptor of action results, then orienting-investigative reaction occurs, which selects another sequences/hierarchies of the functional systems from the set $\{\langle FS_1^1, ..., FS_n^1 \rangle\}$ or selects another functional system FS to attain the goal s_{Goal}. Activated rule of the corresponding functional subsystem is penalized;
7) the goal s_{Goal} is attained and results for each functional subsystem is registered by the acceptor of action results and all activated rules for each functional subsystem are rewarded.

Let us describe all elements of the architecture of functional systems using definitions introduced above.

Afferent synthesis involves the synthesis of motivational excitation, memory, situational and triggering afferentation, as well as reverse afferentation on performed actions. All these afferentation are recorded by a set of

sensors $S_1, ..., S_n$. Motivational excitation also sets a goal $Goal = (p_{i_1}^{goal}, ..., p_{i_{goal}}^{goal})$.

Memory. Each goal can be attained by various sequences of actions. Therefore motivation extracts from the memory all sequence/hierarchies $\langle FS_1^1, ..., FS_n^1 \rangle$ of functional subsystems of the functional subsystem $FS = (s_{Goal}, \{\langle FS_1^1, ..., FS_n^1 \rangle\}, p_{FS})$ that attain the goal.

Situational and launching afferentation specifies the current state of the system $s_t = (p_1, ..., p_k)$ at each instant of time *t*. The initial states $s_0 = (p_{i_1}^0, ..., p_{i_0}^0)$ of the first functional systems of each sequence/hierarchy $\langle FS_1^1, ..., FS_n^1 \rangle$ should correspond to the current state $s_0 \subseteq s_t$.

«Extracting» from the memory the whole experience, transforms motivational excitation as a goal into a «specific goal» of «higher motivation» that determines how it may be attained. A «specific goal» for a functional system $FS = (s_{Goal}, \{\langle FS_1^1, ..., FS_n^1 \rangle\}, p_{FS})$ is the entire set of sequence/hierarchies of functional systems $\{\langle FS_1^1, ..., FS_n^1 \rangle\}$.

Decision making. At the previous stage of the afferent synthesis a set of sequences/hierarchies $\{\langle FS_1^1, ..., FS_n^1 \rangle\}$ of functional systems that attain the goal s_{Goal} is extracted from the memory. At the stage of decision making only one of these is chosen as a "program of actions". The decision making is carried out by a switching function of emotions [4-5].

Acceptor of action results. For the "program of actions" motivational excitation extracts from the memory a "particular criteria" of attaining the goal – i.e. *acceptor of action results*, which consists a full totality of criteria of attaining the whole sequence/hierarchy of goals, for example, the sequence/hierarchy of goals

$$s_0 \to s_1 \to s_2 \to ... \to [\to s_1^i \to s_2^i \to ... \to s_i]... \to [\to s_1^j \to s_2^j \to ... \to s_j]... \to s_{goal}.$$

Described architecture was used for the adaptive control system development that successfully solved the foraging task [3].

Acknowledgements

This work is supported by the Russian Science Foundation grant №17-11-01176.

References

[1] Anokhin, P.K. (1974) "Biology and neurophysiology of the conditioned reflex and its role in adaptive behaviour", Oxford: Pergamon press.

[2] Sudakov K.V. (1984): "The general theory of functional systems". Moscow: Medicine. (in Russian).

[3] Muhortov V.V., Khlebnikov S.V., Vityaev E.E. (2012) Improved algorithm of semantic probabilistic inference in task of 2-dimention animat *Neuroinformatics*, **6** (**1**): 50-62. (in Russian).

[4] Simonov P.V. (1981) "Emotional Brain". Moscow: Science. (in Russian).

[5] Simonov P.V. (1975) Higher nervous activity (motivational and emotional aspects). Moscow: Science. (in Russian).