# ON SIMPLIFYING BRAIN FUNCTIONS

**Włodzisław Duch**[1]
Department of Computer Methods,
Nicholas Copernicus University,
Grudziadzka 5, 87-100 Torun, Poland


**A new approximation to the dynamics of neocortex, preserving essential functions but not the microstructure, is presented. Local feature spaces are used to replace neurodynamics of neural cell assemblies in categorization, approximation and generalization processes. A global workspace (mind space), resulting from cooperation among local feature spaces, facilitates memory-based reasoning. Implementation of this idea leads to a modular, neurofuzzy system as a model of mind. This approach provides a link between neurodynamics and cognitive science. The status and further directions of this project are reviewed.**

## 1. INTRODUCTION

Two quite distinct paradigms of understanding the human intelligence and the human mind are widely accepted. First, artificial intelligence (AI) starts from the higher cognition perspective aiming at intelligent systems based on symbol processing [1]. There are serious problems at the very foundation of such an approach, starting with the famous mind-body problem (how can the non-material mind interact with matter), the symbol grounding problem (how can the meaning be defined in a self-referential symbolic system) or the frame problem (catastrophic breakdowns of intelligent behavior for "obvious" tasks) [2]. Second, neurodynamics investigates models of neural networks inspired by the neural structure of the brain. Such neural models seem to be suited best for the low-level cognitive tasks, such as vision or auditory processes, or for simple classification tasks, while they are somehow restricted in their abilities to realize predefined knowledge structures and in using such structures in sequential reasoning processes. Since intelligence in Nature is exhibited only by biological brains one of the most important tasks in science is to find an approximation to brain's functions leading to a theory of mind.

The present shortcomings of neural networks in modeling higher cognitive functions are connected with the lack of modularity and low complexity of the models rather then with the inherent limitations of the neural modeling itself. Much is know about the details of neural processes responsible for brain functions. Neurodynamics [3] and the computational cognitive neurosciences are thriving fields [4] but understanding of the higher mental activity directly in terms of neural processes in the brain does not seem likely. Macroscopical theories are reducible only in principle to microscopical descriptions. Phenomenological concepts in

---

[1] e-mail: duch@phys.uni.torun.pl, WWW: http://www.phys.uni.torun.pl,
archive ftp.phys.uni..torun.pl/pub/kmk/papers

chemistry, physics and other branches of science are not easily reducible to fundamental interactions. Concepts of neuroscience and concepts of psychology are quite different yet there must be a way of obtaining psychological concepts as an approximation to the neurodynamics of the brain circuits. In this paper a sketch of such theory is presented. The main goals of this theory are:

1) Introduce approximations to the neurodynamics, in agreement with the neurobiological facts, leading to precise concepts for description of cognitive states (mental events).

2) Use these concepts as a language to build a theory of cognitive systems.

3) Apply this theory to explain features of human cognitive processes, such as identification, association, generalization, reasoning, empirical facts related to consciousness.

4) Construct adaptive systems according to specifications, systems processing the incoming signals, categorizing, learning from examples and from general laws, self-organizing, reasoning and performing other cognitive functions.

In the spirit of Allan Newell's "Unified theories of cognition" [1] the theory should be supported by software allowing for verification of its premises and models. The SOAR system developed by Newell and his collaborators is based on production rules and has nothing to do with neurobiology. It may be useful in modeling some cognitive functions but will not help us to understand how they arise from the brain dynamics. Feature Space Mapping (FSM) neurofuzzy system [5] is inspired by the brain and is able to provide not only models but also understanding of cognitive processes and their relation to the brain's dynamics. The mind-like properties of cognitive systems are seen in this theory as an approximation to the brain functions.

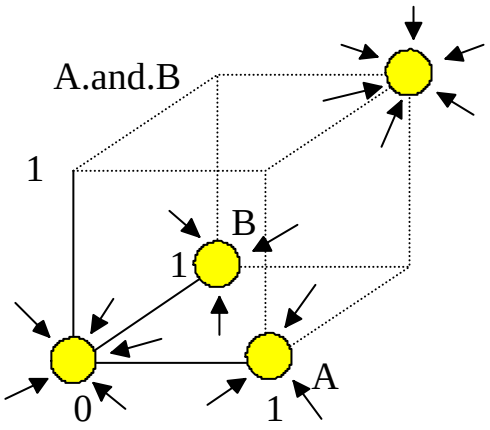
## 2. FROM NEURONS TO SYMBOLS.

Models of the brain function require a series of approximations. In the first step biochemical and bioelectrical processes are drastically simplified by introducing model neurons treated as electrical devices [6]. This approximation averages over many types of neurotransmitters and neuromodulators hiding all couplings with the autonomous nervous system in a few parameters, such as the membrane time constants or electrotonic space constants. Various cell shapes (especially dendrites) are taken into account by the compartmental models of neurons. Actual synapses are found on dendritic spines and probably support non-linear sigma-pi computations. Activity, measured in real neurons using the frequency of spiking discharges, is not the only important parameter. Recent experiments [7] show that temporal coincidence of spikes with accuracy better than 1 ms is preserved in live slices of neuronal tissue. Temporal coincidence is also crucial for development of somatotopic representations. All this complexity is usually simplified into a point-like, single compartment model neurons characterized by additive excitation thresholds and synaptic weights. While networks composed from such simplified elements show interesting computational properties such approximation is clearly quite unrealistic.

Single neurons are important in many perceptual processes (cf. Barlow, p. 415 in [6]) but basic cognitive functions require cooperation of small groups of neurons. The concept of a neural cell assembly (NCA) was introduced in 1949 by Donald Hebb in his seminal book [9]. The cerebral cortex has indeed a very modular structure [ 6,11]. Macrocolumns,

distinguishable using neuroanatomical techniques, contain between $10^4$-$10^5$ neurons in a 1-2 mm high column spanning six layers of the neocortex, within the cortical area of a fraction of $mm^2$. Axons of some NCA neurons spread horizontally on several millimeters enabling entrainment of cooperating NCAs. Within the macrocolumn one may distinguish minicolumns, much smaller functional groups of neurons with inhibitory connections. They have only 110 neurons each in a column of 30 μm diameter. These minicolumns behave as oscillators and recurrent excitations of such oscillators lead to entrainment and synchronization of neuronal pulses (Singer in [6], p. 960). Vertical connections inside these minicolumns are largely excitatory and the density of these connections is of an order of magnitude higher than of the connections with neurons outside of the column. Pyramidal cells dominate at the surface of the neocortex, clustering their axon terminals at about 0.5 mm, in the neighboring macrocolumns.

These basic neuroanatomical facts force us to use models of neocortex based on interacting modules or groups of neurons. Attractor neural networks [8] and other neurodynamical approaches are suitable as models of single modules (cortical columns). Although dynamics of such systems may be rather complex the number of attractors or stable, synchronized and entrained patterns of excitations, is rather limited. Quasi-discreet states of cortical minicolumns are the basic building blocks of the global dynamics of the brain. Consider a dynamical system $\dot{\mathbf{X}}(t) = F(\mathbf{X}(t))$ with a large number of internal degrees of freedom $\mathbf{X}_{int}$ and a small number of inputs $\mathbf{X}_{inp}$. Attractors are activated by specific inputs $\mathbf{X}_{inp}$ dividing the input space into basins of attractors. For example, a cortical minicolumn may learn to solve the A.AND.B or any other logical problem establishing 4 attractors. In the input space (feature space) the corners of the cube will represent the shortest transients of the phase space trajectories and the basins of attractors will belong to the neighborhood of these corners. Complex neural dynamics is replaced by simple gradient feature space dynamics introducing the density of feature space objects proportional to the transients of the neural dynamics. This approximation leads to a symbolic interpretation of brain events. Conscious experience of thoughts and perceptions requires activity in the frontal lobe areas [4]. Axons of neurons reaching frontal lobes from NCAs of other neocortex areas carry only an approximate information about their activity. This information is sufficient to distinguish between feature space objects but not the details of neural dynamics. States of mind are composed from discretized information identified in the brain with symbolic categories and continuos information from sensory signals. Continuous states define fields in feature spaces, for example color perception may be analyzed in the 4-dimensional feature space despite the fact that there are millions of neurons in the visual cortex devoted to the analysis of color.




Simplification of the brain dynamics requires classification of the existing attractor states, transition probabilities between them

and formation of new attractor states during learning. Feature spaces allow for a very useful approximation of this dynamics. More drastic approximations lead to a discreet finite automata (DFA), such as the hidden Markov models (HMM), semantic nets and probabilistic reasoning used in AI. Feature spaces are used in cognitive psychology to discuss mental representations [11]. Therefore the correspondence between feature spaces representing mind events and attractors of the global dynamics of the brain (primarily in the frontal lobes) is a natural link between the neural and psychological sciences. The low level cognitive processes, realized mostly by various topographical maps and population coding mechanisms, define features of internal representations. These features are related to many types of data: analog sensory signals, linguistic variables, numbers, visual images. Real mind objects are composed primarily of preprocessed sensory data, iconic representations and perception-action objects.

## 3. DYNAMICAL FEATURE SPACE MAPPING

Inspiration from the analysis of the brain structures leads to a hierarchical, modular model of the cognitive system. Adiabatic learning hypothesis allows to separate three relatively independent types of learning. The incoming data $\mathbf{X}_{inp}$ is pre-processed to define basic features of internal representation. Development of the best feature detectors corresponds to the slowest learning processes due to the genetic evolution. Each feature contributes one dimension to the local feature space, defining a coordinate system. This space contains "mental objects ", such as categories related to perception, concepts and memories. The input vector $\mathbf{X}_{inp}$ points to a particular position in the feature space. The underlying dynamical system is constantly evolving. In the absence of external signals the activity of the input neurons $\mathbf{X}_{inp}$ changes in a random way. Such chaotic dynamics has indeed been observed in the brain, the most detailed observations coming from the olfactory system [4]. Disregarding the time delays - they are important only in modeling reaction times in psychology - the state of the feature space is described by the vector

$$\mathbf{S}(t) = \mathbf{X}_{inp}(t) + \beta \nabla_{\mathbf{S}} M(\mathbf{S}; t) + \eta(t)$$

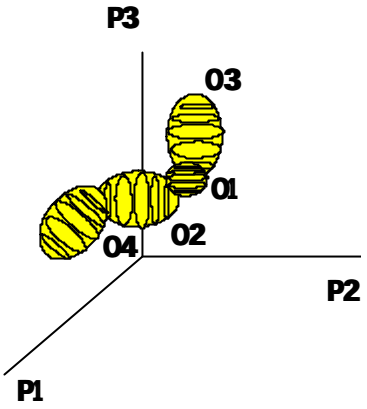
where $\beta$ is a step size constant, $M(\mathbf{S};t)$ is a memory function describing objects as densities in the feature space and $\eta$ is a Gaussian noise with changeable width. This equation describes fast dynamics related to the object recognition. Memory traces $M(\mathbf{S};t)$ change according to slower dynamics, allowing the system to learn new things and forget those that were not memorized beyond certain level:

$$M(\mathbf{S}; t+1) = M(\mathbf{S}; t) + \alpha \cdot \rho(\mathbf{S})(M_c - M(\mathbf{S}; t))$$

$$M(\mathbf{S}; t+1) = M(\mathbf{S}; t)\left(1 - \Theta(M_c' - M(\mathbf{S}; t))\left(1 - e^{-t/\tau}\right)\right)$$

The first equation describes formation of new memory objects and bounded growth of the old ones. Random walk of the state vector $\mathbf{S}$ leaves a weak memory trace. If a new input $\mathbf{X}_{inp}$ appears the state vector stays around $\mathbf{X}_{inp}$ long enough to leave more stable memory trace. The iterative growth of the objects in the feature space proceeds by adding to the local density $M(\mathbf{S};t)$ an additional density term $\rho(\mathbf{S})$ proportional to the difference between the $M_c$ (maximum value) and the actual density. The iterative decay of densities lower than the threshold $M_c' < M_c$ is achieved using the step function $\Theta$ and prevents permanent formation

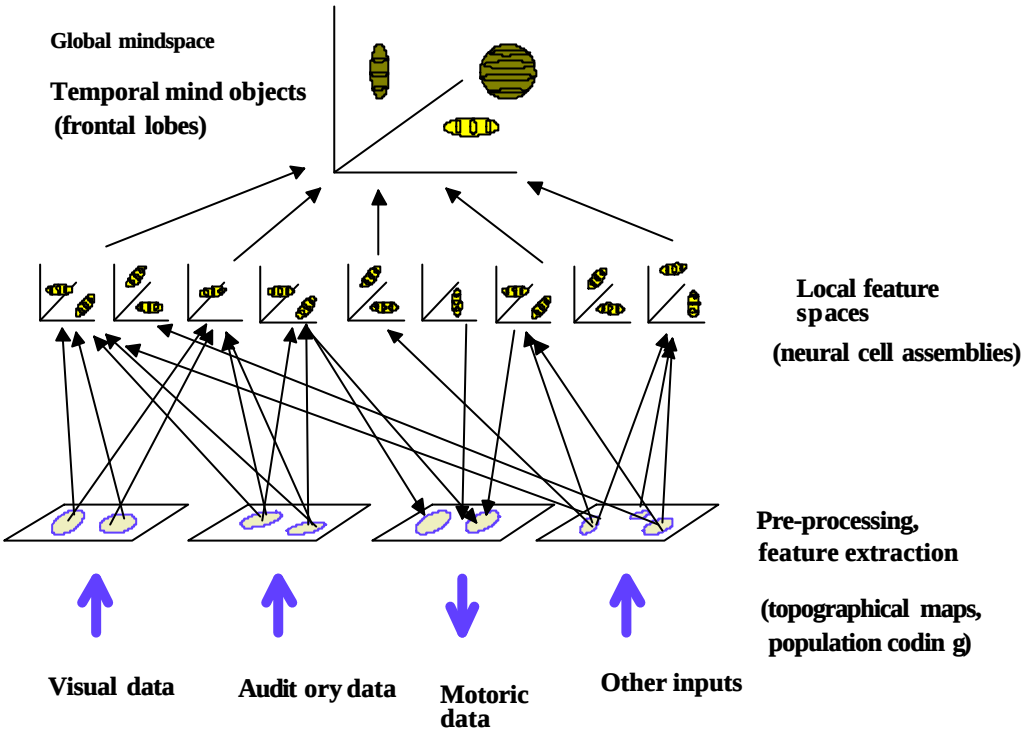
of weak memory traces. Additive noise $h$ added to the state $\mathbf{S}$ blurs each object making it fuzzy and enabling generalization in a controlled fashion. Four data vectors shown in the figure above correspond to objects defined by P1-P3 properties. They form one object in the feature space due to the noisy dynamics of the system. In cognitive psychology the problem of category learning is still a current issue [12]. Formation of categories results from blurring the prototype objects by noise. Categories cannot be defined by lists of features since objects in the feature spaces may have complex shapes.




Learning in this approach is reduced to geometrical problem of creation of objects in the feature spaces. Although it looks quite different from the conventional neural approach in practice an approximation to the learning and retrieval dynamics described above is afforded by the growing RBF type of network, such as the RAN [12] or FSM [5] networks modified to include decaying nodes. There is no particular reason why these functions should be radial as in RBF. The problem is to obtain the maximum flexibility with the smallest number of parameters. More complex problems are solved by "divide and conquer" principle: a large set of local feature spaces processes the incoming signals in parallel, some of them discover that the input vectors match the existing memory traces and activate them, and the results are copied to a global workspace called "the mind space" where various mind objects are temporarily created. An alternative to the density-based model of memory traces is given by the local adaptive coordinate systems (in preparation).

## 4. APPLICATIONS IN COGNITIVE SCIENCE AND SUMMARY

The model of cognitive system presented above may be implemented in a number of ways and can serve as a testbed for theories about human cognition. It is more suitable for that purpose than the production-rule based SOAR system of Newell and collaborators [1]. It is useful to differentiate between cognitive processes requiring simple associations and thinking or reasoning. Associative functions are based on knowledge that is readily available, intuitive, used in recognition and immediate evaluation. Thinking and reasoning are necessary for problem solving. Experimental techniques of cognitive psychology, such as probing the immediate associations between concepts and measuring the response times should give enough information to place basic objects corresponding to selected concepts or perceptions in the local feature spaces. Associations among mind objects, corresponding to the transition probabilities between different attractors of the underlying neurodynamics, should take into account not only the features of representations but also the spatio/temporal correlations. In the simplest model human reaction times for associations should be proportional to the distances of the corresponding objects in feature spaces. "Intuitive" responses should be based on the topography of the mind space. Logical and rule-




based reasoning is only an approximation to the dynamics of the state of mind, approximated here by the activation of a series of mind objects.

Adding hidden dimensions (corresponding to internal features that influence the dynamics but are not accessible through inputs or outputs of the system) allows to model arbitrary transition probabilities (associations of mind objects). Such problem solving tasks as playing chess seem to be based on a large memory (large number of mind objects) and on a memory-based reasoning with a rather limited exploration of the search space. Memory-based reasoning is related to probabilistic neural networks and outperforms in many cases other learning methods, including neural networks [15]. We have already shown how the FSM system [5] may be used in simple memory-based reasoning. We are also considering application of FSM in natural language analysis. The idea of mental spaces [16] has proved to be very fruitful in the study of reference problems in linguistics, although mental spaces were so far constructed as ordered sets and relations rather than feature spaces. More technical applications are considered elsewhere in this volume.

Only one type of intelligent devices exist in Nature - brains. Artificial Intelligence based on the symbolic paradigm has not been successful as a model of general intelligence. Homogenous neural networks are useful for classification or approximation, but they are not good models of the architecture of the brain either. Intelligence requires modularity. In this paper radically different approximation to the brain has been proposed, preserving function but simplifying structure.

**Acknowledgment**

Support by the Polish Committee for Scientific Research, grant 8T11F 00308, is gratefully acknowledged.

**REFERENCES**

[1] A. Newell, *Unified theories of cognition*. (Harvard Univ. Press, Cambridge, MA 1990)

[2] Harnad, S. (1990) *The symbol grounding problem*. Physica D 42: 335-346; Harnad, S. (1993) *Problems, problems: the frame problem as a symptom of the symbol grounding problem.* PSYCOLOQUY 4 (34) frame-problem.11

[3] N. Rashevsky, *Mathematical Biophysics* (Dover, NY 1960)

[4] M. S. Gazzaniga, ed. *The Cognitive Neurosciences* (MIT, Bradford Book 1995)

[5] W. Duch, G.H.F. Diercksen, *Feature Space Mapping as a universal adaptive system*. Computer Physics Communications **87** (1995) 341-371; W. Duch, Neural Network World **4** (1994) 645-654; W. Duch, R. Adamczak, N. Jankowski and A. Naud, Proc. of Engineering Applications of Neural Networks, Helsinki 1995, pp. 221-224

[6] M. Arbib, Ed. *The Handbook of Brain Theory and Neural Networks* (MIT Press 1995)

[7] P.R. Montague and T. Sejnowski, *The predictive brain: temporal coincidence and temporal order in synaptic learning mechanisms*. Learning and Memory 1 (1994) 1-33

[8] D.J. Amit, *Modeling brain function. The world of attractor neural networks* (Cambridge Univ. Press 1989)

[9] D. Hebb, *The Organization of Behavior* (J. Wiley, NY 1949)

[10] Y. Burnod, *An adaptive neural network: the cerebral cortex* (Prentice Hall 1990)

[11] I. Roth, V. Bruce, *Perception and Representation*, (Open University Press, 1995)

[12] T. Poggio and F. Girosi, *Networks for approximation and learning*. Proc. of the IEEE 78 (1990) 1481; J. Platt, *A resource-allocating network for function interpolation.* Neural Comput. 3 (1991) 213

[13] D.L. Waltz, *Memory-based reasoning*, in [6], pp. 568

[14] G. Fauconniere, *Mental Spaces* (Cambridge Univ. Press 1994)