1 Platonic model of mind as an approximation to neurodynamics

Włodzisław Duch

Department of Computer Methods, Nicholas Copernicus University, Grudziądzka 5, 87-100 Toruń, Poland.
E-mail: duch@phys.uni.torun.pl

Abstract. Hierarchy of approximations involved in simplification of microscopic theories, from sub-cellural to the whole brain level, is presented. A new approximation to neural dynamics is described, leading to a Platonic-like model of mind based on psychological spaces. Objects and events in these spaces correspond to quasi-stable states of brain dynamics and may be interpreted from psychological point of view. Platonic model bridges the gap between neurosciences and psychological sciences. Static and dynamic versions of this model are outlined and Feature Space Mapping, a neurofuzzy realization of the static version of Platonic model, described. Categorization experiments with human subjects are analyzed from the neurodynamical and Platonic model points of view.

1.1 Introduction

There is no doubt that higher cognitive functions result from the activity of the brain and thus should be implementable by neural networks. The present shortcomings of neural networks are connected with the lack of modularity and low complexity of the models rather then with the inherent limitations of the neural modeling itself. Computational cognitive neuroscience [1], recent descendant of neurodynamics, tries to unravel the details of neural processes responsible for brain functions. On the other end of the spectrum artificial intelligence aims at building intelligent systems starting from the processing of symbols [2]. Between neurochemicals and symbols a whole spectrum of theories at different levels of complexity exist. Reduction of simplified higher level theories to more detailed lower level theories may not be feasible but it is worthwhile to consider approximations that may at least justify the use of some higher level concepts.

Several books on application of the theory of dynamical systems to human development process appeared recently [3]. Although useful as a metaphoric language dynamical systems theory and connectionist modelling have not yet produced quantitative explanations of the brain functions. Computational Cognitive Neuroscience [1] allows for comparison of computer simulations with experimental investigations, at least for low-level cognition. Development of topographical feature maps, for example in the visual system (cf. Erwin *et al.* [4]), is a good example of theoretical and experimental interplay at the microscopic level. One obvious drawback of the whole information processing paradigm from the cognitive science perspective is its lack of overlap with psychology. Mind is rarely mentioned in computational cognitive neuroscience [1] because current approximations to neural dynamics lead only to a description of behavior. Cognitive science seems to be a collection of loosely related subjects without central theory [5]. A new language to speak about mind events, linking neurodynamics with cognitive psychology, is needed. The main goal of this paper is to show that a theory connecting neuroscience and psychology fits into the hierarchical scheme of approximations between simplified theories of neurodynamics and theories of behavior based on symbolic and finite-state approaches. I will call it "Platonic model" because it treats the mind and its ideas (objects) seriously. In the next section models and approximations to neural functions at various levels, from sub-cellural to the large brain structures, are discussed. Relations between different levels of modeling are of particular interest. In the third section basic ideas of the Platonic model of mind are introduced and in the next section Feature Space Mapping neurofuzzy network, a realization of the static version of the Platonic model, is presented. In the fifth section Platonic model is applied to two problems in category learning, as studied by psychology. A brief discussion on the future of the Platonic model closes this paper.

1.2 Hierarchical approximations to neurodynamics

Nervous system has distinct organization that may be investigated at many temporal and spatial scales or levels [1], starting from single molecules (spatial scale 10^{-10} m), through synapses, neurons, small networks, topographical maps, brain systems, up to the whole brain and central nervous system level. Each level of description has its specific theories, methods and goals. The very notion of "levels" is approximate since different levels cannot usually be decoupled. Main levels of approximations are presented [6] in Table 1.

1. Quantum level. Processes at the genetic and molecular level have direct influence on the states of the brain and contents of mind. Ira Black may be right [7]: information processing is ultimately done on the molecular level, states of the brain are concentrations of different neurochemicals. Understanding the qualia problem [8], including consciousness, may depend directly on the molecular level. Regarding consciousness as physical states of the brain endowed with structures of sufficient complexity, with specific control and feedback architecture, avoids the philosophical problems of the information processing paradigm.

2. Molecular and synaptic level. Computer simulations of synaptic processes gives results comparable in many details with the results of neurophysiological measurements [9]. Mechanisms of fast changes, leading to action potentials, are better understood than those responsible for slow, permanent changes, although the role of Long Term Potentiation (LTP) and Long Term Depression (LTD) in memory processes is slowly uncovered [1]. Permanent synaptic changes cannot be decoupled from the molecular and genetic level, while understanding of action potentials is possible at the bioelectrical level.

Cognitive phenomena	Levels and Models	Theories/Methods
Reasoning, problem solving, thinking, creativity	Brain, rational mind; knowledge-based systems, symbolic processing	Artificial Intelligence, Psychology of Thinking
Behavior, complex learning, immediate reactions ($t \approx 1$ s)	Large brain structures; probabilistic and finite state automata	Cognitive and Behavioristic Psychology, Machine Learning
Intuition, categorization, recognition	Computational maps, transcortical neural assemblies; Platonic models, feature spaces	Psychology, Neurobiology, Machine learning
Memory effects	Small networks; recurrent neural networks and dynamical systems	Computational Neuroscience, EEG, MEG, fMRI
Basic learning, internal representations	Neural assemblies; spiking neurons	Neurophysiology, single and multielectrode recordings
Conditioning, basic forms of learning	Single neurons, electric compartmental models, LTP, LTD	Biophysics, Neurophysiology
Moods, habits, addictions	Molecular and synaptic level, biophysics of ion channels, membrane properties	Neurochemistry, Genetics, Biophysics, Psychopharmacology
	Quantum level, small molecules	Neurochemistry, Biophysics

Table 1. Levels of modeling, from neurochemistry to psychology of thinking.

3. Single neuron level. Simulations reproduce experimental results with sufficient accuracy to give us confidence that the bioelectrical description of single neuron based on the Hodgkin-Huxley equations is essentially correct [9]. Anderson and van Essen [10] write that "the firing rates of cortical neurons encode an irregular, but complete, sampling of Probability Density Functions (PDF's) of multidimensional space". In particular they analyze multimodal topographical maps of the superior colliculus in terms of PDFs, maps integrating sensory and motoric components (saccadic eye movements) and discuss computation of PDFs by visual cortex. Bayesian analysis [11] of neuron responses allows to compute the posterior probability $P(s|r) = P(stimulus|response)$ for responses of many neurons $r = (r_1, r_2, ...r_N)$. Assuming that the estimation of $P(r_i|s)$ is known from multielectrode measurements:

$$P(s|r) = P(s|r_1, r_2, ...r_N) = \frac{P(s)\prod_{i=1}^{N} P(r_i|s)}{\sum_{s'} P(s')\prod_{i=1}^{N} P(r_i|s')} \tag{1.1}$$

where a uniform prior is usually assumed for $P(s)$. This method showed much faster convergence to correct stimuli than the population vector method [11]. Metric properties of trains of spikes are directly related to stimuli [12]. Both the absolute time and the intervals between the spikes carry information allowing for discrimination of different clusters of stimuli.

4. Neural assemblies. Comparison of general trends observed in live slices of neural tissue or in some brain structures with computer simulations based on the simplified spiking neuron models are quite successful [13]. The noisy "integrate-and-fire" neurons are the only well-justified approximate models of real neurons. It is commonly believed that the activity of a neuron, proportional to the number of spikes per second, grows in sigmoidal fashion until saturation, corresponding to the maximum firing rate, is reached. In fact neurons are able to recognize specific combination of inputs. Using a small number of dendritic inputs and changing the phases and frequencies of the incoming spike trains quite complex output patterns may be observed (Duch and Ludwiczewski, in prep.). This leads to non-monotonic transfer functions. Associative memories based on neurons with such functions have larger capacities and are able to escape from local minima [14]. In the most common models of simplified neurons for a fixed norm of weights and input signals the activity $I = \mathbf{W} \cdot \mathbf{X} = I_{max} - d(\mathbf{W}, \mathbf{X})^2$ is a function of distance $d(\cdot)$ between inputs and weights. Large activation is obtained when positive weights (excitatory synapses) are matched with positive dendritic inputs (frequencies of spikes above the average), while negative weights (inhibitory synapses) are matched with negative inputs (frequencies below average). Therefore neural models based on localized transfer functions, where weights play the role of prototypes, are approximations to spiking models.

5. Small networks. Every representational scheme has to face a problem of combinatorial explosion: concepts are combined in an infinite number of ways creating new concepts [15]. To solve this problem Neural Cell Assemblies (NCA), groups of neurons that strongly activate each other, were proposed already in 1949 by Donald Hebb [16]. The cerebral cortex has a very modular structure [17] and cortical minicolumns [17] are natural candidates for NCAs. These vertically oriented cords of neurons contain in almost all neocortical areas about 110 cells. The diameter of such a column is about 30 microns and the number of minicolumns in the neocortex is about 500 millions. Hundreds of such minicolumns are gathered into *cortical columns* (maxicolumns). Connections between different columns are reciprocal. In addition the thin neocortex has a hierarchical, layered structure, with six principal layers. Minicolumns behave as oscillators and recurrent excitations of such oscillators leads to an entrainment and synchronization of neuronal pulses, called also "synfire chains" [18].

In computer simulations and experiments with real neurons Traub *et al.* [13] showed how groups of minicolumns synchronize and communicate over larger distances. 40 Hz gamma oscillations sustained by inhibitory neurons provide a local clock and spikes produced by excitatory piramidal cells appear about 5 ms after the main beats, forming a kind of "doublets" and binding the activity of widely separated groups of neurons. Several authors [19] stress the activity of whole populations of neurons. Modeling brain structures should take into account lamination of cortical layers, columnar structure, geometry of connections, modulation of densities of different type of cells, formation of topographic maps.

6. Transcortical neural networks. Recordings of the single-neuron activity in the infero-temporal cortex of monkeys performing delayed match-to-sample tasks show that the activity of a neural cell assembly has attractor dynamics [20]. Several stable patterns of local reverberations may form, each coding a specific perceptual or cognitive representation. Via axon collaterals of pyramidal cells extending at distances of several millimeters, each NCAs excites other NCAs coding related representations. Mental events are correlated with attractor dynamics of local and transcortical neural cell assemblies (TNCAs). The low level cognitive processes, realized mostly by various topographical maps, define features of internal representations: analog sensory signals, linguistic tokens, iconic images. In visual system local form, color and texture are passed to the infero-temporal (IT) cortex which implements the memory of the visual objects. Sensory information from visual, auditory and somatosensory areas converges in multiple areas in such subcortical structures as superior colliculus. These areas probably integrate also motor responses, creating “multisensory multimotor map”. Neurons responding to multimodal features act as higher order feature detectors. Mind objects are composed primarily of preprocessed sensory data, iconic representations, perception-action multidimensional objects and abstract symbols. Successful models of memory, such as the tracelink model of Murre [21], make good use of modular neocortex structure, postulating that each episodic memory is coded in a number of TNCAs (simultaneously activated memory traces). Their attractors compete to dominate the global dynamics of the brain, reinstating similar neural state as during the actual episode.

7. Large brain structures. Topographical maps are found in the neocortex, in the old cortex and in some subcortical nuclei [1]. Another type of computational map, called population coding, seems to be an ubiquitous mechanism of the brain information processing. Both information and its significance is present in the population coding mechanism. Arbitrary vector field corresponding to a complex information encoded by a population of neurons is a linear combination of elementary basis fields. Sensorimotor functions are described by non-euclidean coordinate transformations in frames of reference intrinsic to biological systems [22]. Statistical mechanics of neocortical interactions (SMNI) [23], a mesoscopic theory of neural tissue, averages neural activity at multiple spatial and time-scales and is able to explain the lifetimes and the magic number 7 ± 2 of short-term memories [2] as quasi-stable dynamical patters in the model brain with typical size and shape.

8. Symbols, reasoning and the rational mind. Cognitive processes are based on memory, internal representations and categorization of the data provided by environment. Symbols corresponding to perceptual or abstract categories correspond to quasi-stable attractor states of brain dynamics. The next step – rational mind – requires understanding of the long-term dynamics of transitions between the attractor states. Although the architecture of human mind seen at the highest, cognitive level has been considered mainly from the symbolic artificial intelligence perspective [2] symbols, logic and reasoning should result from an approximation to the dynamical systems behavior. A drastic, although quite natural simplification of the neurodynamics leads to a discrete finite automata (DFA), such as the hidden Markov models. Elman [24] writes that cognition (and language in particular) are best understood as the behavior of a dynamical system. In his view mind objects are not abstract symbols but rather regions of the state space, and rules and logics are embedded in the dynamics of the system. Grammatical constructions are represented by trajectories in the state space.

Understanding higher mental activity directly in terms of neural processes in the brain does not seem likely. Computational neuroscience may be our best approach to ultimate understanding of the brain and mind but chances that neural models are going to explain all aspect of cognition are small. Although chemistry has been reduced to physics long time ago the work of experimental chemists has hardly been affected by the development of quantum chemistry. Language of neuroscience and language of psychology is quite different. Understanding of the mind requires intermediate theories, between neural and mental, physical and symbolic. Perhaps specific theories at different levels of complexity and a few bridges between them is all we can hope for.

1.3 Platonic mind

Approximations discussed in the previous section describe behavior without referring to the mind. A model describing the stream of mind events – recognition of objects, categorizations, trains of thoughts, intuitive decisions and logical reasonings – is needed to place cognitive science on solid grounds. In this section a model based on a more subtle approximation to neurodynamics than finite state models is introduced. It is called "the Platonic model" since it treats the space in which mind events take place seriously [25] and represents concepts as idealized objects in this space. However, in contrast to what Plato believed in, the content of our minds is just a shadow of neurodynamics taking place in the brain, rather than being a shadow of some ideal world.

Large-scale neurodynamics responsible for behavior sometimes takes place in low-dimensional spaces. For example, Kelso [3] has observed that although the dynamics of movements of fingers is controlled by millions of neurons there are simple phase transitions which are described in a low-dimensional (in fact one-dimensional) spaces. His analysis is based on the enslaving principle of synergetics developed by Haken [26], stating that in some circumstances all modes of a complex dynamical system may be controlled (enslaved) by only a few control modes. Attractors of such systems lie on a low-dimensional hyperplane in the state space which has a huge number of dimensions. Recently Edelman and Intrator [27] proposed in context of perceptual problems that learning should be viewed as extraction of low-dimensional representations. In the Platonic model mind events, including learning, take place in relatively low dimensional feature spaces. Objects in these feature spaces correspond directly to the attractors of TNCA dynamics. In the olfactory system the dynamics is chaotic [1] and reaches an attractor only when a specific external input providing a proper combination of features is given as a cue. TNCAs recognize input vectors (cues) that fall into local basin of attractor and adopt patterns of activations that are interpreted by other TNCAs as multidimensional fuzzy internal representations (objects) containing output decisions. This process will be presented in feature spaces (Fig. 1.1).

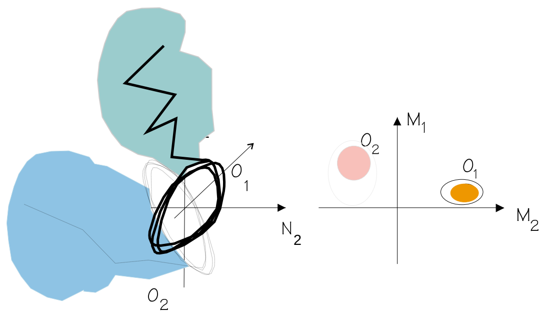

Fig. 1.1. Relation between attractors representing correlation of the spiking activity of a group of neurons (here just two) in the space of $N_i, i = 1..n$ and objects in the feature space $M_i, i = 1..m$, where $n \gg m$.

General description. A coordinate system based on features of mental representations obtained from lower-level modules, such as topographical maps or population coding computational maps, defines a multidimensional space [25], called here "the mind space", serving as an arena in which mind events take place. In this space a "mind function" is defined, describing fuzzy "mind objects" as regions of space where the mind function $M(X)$ has non-zero values. The size and shape of these mind objects (contours of $M(X)$=const) should correspond to basins of attractors of the underlying neurodynamical processes. High $M(X)$ value means that if a specific combination of features X is given as an input to the neurodynamical system, the time to reach an attractor will be short. The internal variables of neurodynamics may influence the topography of mind space but are not explicitly represented. The name "mind space" is replaced by more modest "feature space" for Platonic models using inputs of single modality, such as computational maps. Platonic models are constructed either as an approximation to neurodynamics (cf. next section) or using information about similarities, associations and response times from psychological experiments.

The model has **three time scales**. Very slow evolutionary processes determine the type of available information, creating basic framework for mind spaces and their topography. Learning, or changes in the topography of mind spaces, creation of new objects, forgetting or decay of some objects, changing their mutual relations, requires long time scale. Dynamics of activation of mental representations, trains of thoughts or perceptions is faster. At a given moment of time combination of input features activates one object. "Mind state", given by a fuzzy point (underlying neurodynamics is noisy), evolves in the mind space and "activates" different objects – a searchlight beam lighting successive objects is a good metaphor here [28]). Human conceptual space is better approximated by a number of linked low-dimensional spaces rather than one large mind space. Local feature spaces model complex feature extraction at the level of topographical maps, providing even more complex components building higher-level "hypothesis" feature spaces that integrate multimodal representations, and at the highest level creating the space in which mind events take place (Fig 1.2). Values provided by the emotional system change $M(X)$ gradients in the mind space, influencing evolution of the mind state. The idea of a "mind space" or "conceptual space" is metaphorical, but the same is true for such concepts of physics as space-time or a wavefunction. Some **definitions and properties** of the Platonic model are summarized below.

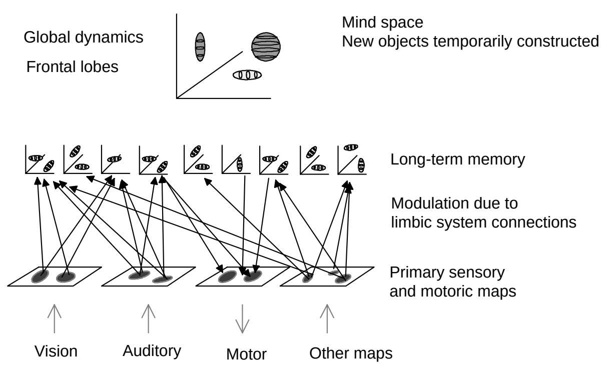

Fig. 1.2. Local feature spaces provide complex features for intermediate level, with competing local "hypothesis spaces" based on long-term memory, providing data for the highest-level "mind space".

– Mind objects, represent long-term memory traces, are defined as points, sets of points or local maxima of $M(X)$ proportional to the confidence of assigning the combination of features at the point X to certain object. In particular $M(X)$ may be a sum of all joint probability density functions for all objects contained in the mind space, with X corresponding to all relevant features, including symbolic names treated as one of the dimensions. Several prototypes may create a complex shape corresponding to category that cannot be defined by simple enumeration of features (Fig. 1.3). Some objects may continuously change into other objects (ex: color or emotions) in which case density will have maxima only in the subspace corresponding to labels (names) of the objects and change continuously in other dimensions.

– Knowledge is contained in the topography of mind spaces, while reasoning and logic are approximations to mind state dynamics. Learning is equivalent to creation of new objects and changing topographical relations among existing mind objects. Topography may also be partially fixed by *a priori* design (evolutionary processes) or deep-rooted knowledge. Recognition or categorization of unknown combination of features X is based on similarity and is done by finding local maxima of the mind function. Learning and recognition form static version of the Platonic model that should be sufficient to explain immediate ("intuitive") reactions in short-time frame experiments. "Intuition" is based on the topography of the mind space.

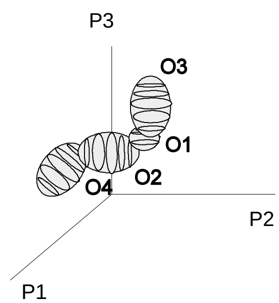

Fig. 1.3. Mind objects (contours of $M(X)$ =const are shown here) corresponding to categories are more than enumeration of features or sets of exemplars.

– Evolution of the mind state, including the dynamics of activation of mind objects, forms the dynamic part of the Platonic model, described here only very briefly [6]. In analogy to physics concepts like inertia (the tendency to stay inside areas of high $M(X)$, momentum of the mind state (proportional to the derivative $\dot{X}(t) = \partial X(t)/\partial t$ of the state vector), potential function $V(X) = -M(X)$ and energy to overcome barriers (proportional to the difference: $\max M(X) - \min M(X)$) may be defined. Transition probabilities $p(A \rightarrow B)$ from mind object A to mind object B are given by the conditional expectancies. An object represented by the density $O(X^{(i)})$ is "activated" when the mind state $X(t) \in O(X^{(i)})$ points at or belongs to it. In the hierarchical model each feature space (Fig. 1.2) has separate evolution providing information based on local mind states to spaces higher in hierarchy.

– Short-term memory (STM) mechanisms are identified with the highest "mind space" level in the hierarchical Platonic model. Here new objects appear and decay after a short time. Primary objects are composed of preprocessed sensory data and are stored permanently after several repetitions in one or more feature spaces. Secondary objects (for example mathematical concepts) are defined as combinations of more abstract features. Mind space plays similar role to frames in AI – once a partial frame is build feature spaces provide more details and the competition among them is won by those that fit in the STM space (cf. Baars theory [29]). Since STM capacity is limited creation of complex object relies on the "chunking" mechanism, which is the basic mechanism of creating higher levels of abstraction [2]. Symbols or linguistic labels are particularly effective (facilitating faster recognition) in identifying complex objects, therefore chunking mechanism relies primarily on symbols. Inductive learning and problem solving requires structured mind objects at the STM level, using something like the Evolving Transformation Systems [30], which has quite natural geometrical interpretation, including parametrized distance functions.

It is interesting to note that common English language expressions: to put in mind, to have in mind, to keep in mind, to make up one's mind, be of one mind ... (space) are quite natural in this model. Platonic model allows to discuss mental events using concepts that are justified by neuroscience. As a simple illustration consider the masking phenomenon [8] in which a disk is briefly shown on the screen. The actual mind state is pushed by the external inputs in the direction of the object representing perception of such disk (Fig 1.4), but because of the inertia it takes about 500 ms to activate this object and send the information to the higher level mind space. The effect of priming may shorten this time by placing the state of mind closer to objects that will be activated in near future. The duration of the perceptual input has to be sufficiently long to force the change of momentum of the mind state to reach and activate the object invoking the perception of a disk. If it is too short, or if after brief 30 ms exposure another object is shown the mind state will be pushed in direction of the new object, without activation of the first one. This is analogous to the scattering phenomena in physics and leads to masking of the first stimuli by the second. The same analysis may be done at the level of neurodynamics but connections with mental events are not so obvious as in the case of Platonic model.

Mathematical realization. Function $M(X(t))$ describing topography of the feature spaces has natural realization in form of a modular neural network, such as the mixture density network modeling joint probability distributions of the inputs (cf. next section). Transition probabilities in psychological experiments with similarity judgments are not symmetric or transitive. This has been a problem for Shepard [31], who proposed that the likelihood of invoking the same response by two stimuli should be proportional to the proximity of these stimuli in a psychological representation space. Symmetry is broken if local coordinate systems for mind objects are used. The distance from A to B should be measured using local metric at A, since activation of A represented by local NCA spreads to the NCA

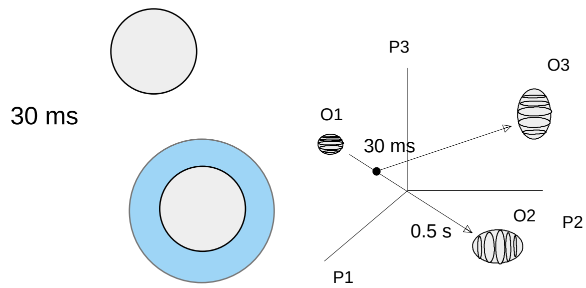

Fig. 1.4. Illustration of the masking phenomenon.

representing object B, and the time it takes to activate B should be proportional to the distance in the feature space. Metric functions are defined only at maxima of $M(X)$. Second approach is based on Finsler geometry [32]. If time is used as a measure of distance (as is frequently the case on the mountain paths) than the distance between points $A,\ B$ connected via a curve $X(t)$ parametrized by t should depend not only on the intermediate positions $X(t+dt)$ but also on the derivative $s(A,B)=\int_A^B L(X(t),\dot{X}(t))dt$, where $L()$ is the metric function (Lagrangian in physics). The distance between A and B may than be taken as a minimum of $s(A,B)$ over all curves connecting the two points, which leads to a variational problem (this integral is called "action" and all fundamental laws of physics may be presented in such variational form). Finsler metric function is defined by the mind function:

$$s(A,B)=\int_A^B \exp\left[\alpha(\nabla M(X(t))^2-\beta M(X(t))\right]dt \qquad (1.2)$$

where α and β are constants and the gradient is calculated along the $X(t)$ curve. For flat densities (constant $M(X)$) $s(A,B)$ is proportional to Euclidean distance, for linearly increasing $M(X(t))$ it grows exponentially with t and for linearly decreasing $M(X(t))$ it grows slowly like $1-e^{-Ct}$. Finsler metric may be obtained from psychophysical distance measures between different stimuli, which are derived from similarity matrices applying multidimensional scaling (MDS) in low-dimensional spaces. A metric for feature space may also be derived directly from the comparison of trains of impulses [12], population vector analysis [1], or Bayesian analysis (Eq. 1.1) of stimuli. In the last case the matrix $P(s|r)$ may be subjected to MDS analysis in the same way as the matrix of psychophysical responses.

Neural dynamics should be replaced by simpler dynamics in feature spaces. Symbolic approach to dynamics [33] and the cell mapping method of Hsu [34] provide simplified trajectories preserving sufficient information about basins of attractors and transition probabilities. For some stimuli (input states) $X\in O(A)$ belonging to the basin of attractor A averaged responses Y of the system (for example, calculated using Bayesian analysis Eq. 1.1) are collected. Transients T (number of iteration steps to reach an attractor, or response time in psychological experiments) from a given initial value of X should be measured and averaged over different internal states. Local maximum of $M(X,Y)$ covers all points X corresponding to $T \approx 0$ and leading to the response Y. The distance (in Finsler metric sense) between the point (X,Y) and the area representing the attractor should be equal to the value of the transient. For simple dynamical systems we have defined the $M(X,Y)$ values in such a way that the number of steps in a simple gradient dynamics searching for maximum of $M(X,Y)$ by moving a fixed step size in the direction of gradient is equal to the number of iteration steps needed to reach an attractor.

Related work and possible extensions. Anderson and van Essen [10] regard the firing rates as sampling of probability density functions, and objects in feature spaces are essentially such densities. In the perception-related similarity experiments discussed by Shepard [31] psychological representation spaces are always low-dimensional. In rotation-invariant object recognition experiments Edelman, Intrator and Poggio [35] obtained quite good results using nearest-neighbor similarity based computations, while a small number of principal component basis functions was sufficient for high accuracy of shape recognition. In vision problems although the measurement space (number of receptive fields) is huge internal representation of shapes seems to be low dimensional [27]. Processing of spatial information by the sensory and motoric systems is basically three-dimensional, although the number of internal dimensions is very large – many muscles work synchronously in simplest tasks. In philosophy the idea of conceptual spaces is pursued by Peter Gärdenfors and his collaborators (see http://lucs.fil.lu.se/Projects/Conceptual.Spaces/). Van Loocke presented a connectionist approach to the dynamics of concepts [36]. It would be interesting to repeat some of his results using the Platonic mind model.

In linguistics the idea of mental spaces, called also conceptual spaces or feature spaces, is well established. The problem of deep semantics across heterogeneous sources has become especially acute in the Digital Library Initiative (see http://www.grainger.uiuc.edu/dli/), a large-scale project aimed at building digital repositories of scientific and engineering literature. Searching for information in related fields the meaning of keywords is captured using concept spaces, based on co-ocurrence analysis of many terms. In 1995 two symposia on semantic spaces were organized (in Montreal and Pittsburg, see http://locutus.ucr.edu/hds.html). Semantic spaces are used for modeling of lexical ambiguity, typicality effects, synonym judgments, lexical/semantic access, priming effects and other linguistic phenomena. Conceptual spaces are used to solve the co-reference problems in linguistics [37]. In an experiment by Ripps *et al.* [38] subjects rated similarities between pairs of words that were names of 13 birds and 12 mammals. We (Duch and Naud, unpublished) have used 30 verbal descriptions (features) of the same animals creating mind objects in the feature space and comparing the multidimensional scaling (MDS) maps obtained from experimental similarity data and from MDS projections of mind objects. The two maps are almost identical and show higher order category formation (domestic birds, etc.).

Platonic model may also include a subspace for emotions. In a paper by Yanaru *et al.* [39] an "engineering approach" to emotions has been presented. 8 primary emotions (joy, anger, expectation, hate, sadness, fear, surprise, acceptance) are used to define 68 linear combinations corresponding to “mixed emotions”. The goal of this approach is to be able to predict the change of emotional state when prompted by external inputs (words, sentences, advertisements). Despite some shortcomings representation of emotional space and analysis of emotion dynamics by following the trajectories in such space may at least partially be feasible. Moreover, it seems possible to connect results of psychological experiments with EEG measurements allowing to differentiate between emotional states of subjects.

All applications mentioned above are only special cases of the static version of the Platonic model.

1.4 Feature Space Mapping network.

Feature Space Mapping (FSM) network [40,41] is an implementation of the static version of the Platonic model. FSM may be viewed as a neural network based on estimation of joint probability densities, as a fuzzy expert system based on representation of knowledge by fuzzy sets or as a generalization of memory-based approaches in which exemplar and prototype theories of categorization [42] find natural extension. Rough description of the topography of the mind space is created using clustering techniques based on dendrograms or simplified decision trees [43]. This method of initialization is much more effective [43] than adding one neuron at a time in the constructive RBF methods [44] and sometimes FSM network *before learning* gives results of similar quality to the final results obtained by other approaches.

Preliminaries: given a set of training examples $\mathcal{D} = \{X^k, Y^k\}$ create $M(X, Y; P)$ in a network form (P are parameters), giving outputs approximating the landscape of the joint probability density $p(X, Y|\mathcal{D})$. Function $M(X, Y; P)$ should be neither equal to, or proportional to, this density; all that is required is that the maxima of conditional probabilities $Y_p(X) = \max_Y p(Y|X, \mathcal{D})$ agree with the corresponding maxima of $M(X, Y; P)$ obtained by calculation of $Y_M(X) = \max_Y M(X, Y; P)$. This task is simpler than the full estimation of the joint or conditional probabilities.

Transfer functions: FSM uses separable transfer functions for description of the feature space objects. Separable transfer functions $s(\mathbf{X}; P) = \prod_{i=1}^{N} s_i(X_i; P_i)$ are chosen because: 1) calculation of projections of N-dimensional objects on arbitrary subspaces is easy, facilitating learning and recognition in lower-dimensional subspaces; 2) fuzzy logic interpretation is straightforward, with $s_i(X_i; P_i)$ being local, or context-dependent, membership functions; 3) separable functions are more flexible than radial basis functions; we have compared the convergence of various neural networks on classification and approximation problems and found products of sigmoids (biradial functions) to be superior [45] to Gaussian and other functions. Increasing the slopes these functions change into a window-type rectangular functions $L(X_i; t_i, t'_i) = \sigma(X_i + t_i)(1 - \sigma(X_i + t'_i)) \rightarrow \Theta(X_i + t_i)(1 - \Theta(X_i + t'_i))$, where Θ is a step function, Smooth change from the convex to cuboidal density contours corresponds to change from fuzzy to crisp logic interpretation of network nodes.

Rotation of densities. An important drawback of RBF and other density networks is their inability to provide simple description of skewed distributions in high-dimensional spaces using rotated densities. In practice for high-dimensional inputs covariance matrices in Mahalanobis distance calculations are always diagonal. Rotations are used in FSM in two ways. In the initialization phase [43] rotation matrix is replaced with a series of 2×2 transformations which are fixed during learning. If this is not sufficient vector $\mathbf{K}$ of adaptive parameters is used during learning. In N-dimensional space a "slice" of density perpendicular to $\mathbf{K}$ is cut:

$$S(\mathbf{X}; \mathbf{t}, \mathbf{t}', \mathbf{K}) = L(\mathbf{K} \cdot \mathbf{X}, t_N, t'_N) \prod_{i=1}^{N-1} L(X_i, t_i, t'_i) \tag{1.3}$$

Learning: a robust constructive algorithm is used to build the FSM network. Feature selection is performed by adding penalty term encouraging dispersions to grow:

$$E(P) = E_0(P) + \lambda \sum_{i}^{N} 1/(1 + \sigma_i) \tag{1.4}$$

where $E_0(P)$ is the standard quadratic error function and P represents all adaptive parameters, such as positions t_i and dispersions $\sigma_i = |t_i - t'_i|$ of localized units. The sum runs over all inputs for the most active node. If σ_i includes the whole range of input data the connection to X_i is deleted. An alternative approach to feature selection is to expand dispersions after several training epochs as much as possible without increasing the classification error. Details of the learning algorithm are described in [6,40]. New nodes are added using similar criteria to RAN networks [44], and removed by checking the quality Q_k (estimated by dividing the number of correctly classified vectors through the number of all vectors handled by this node). Low quality units are removed from the network, allowing it to grow more "healthy" nodes. FSM network may contain a number of recognition, or identification, modules and some modules implementing logical operations [40]. This design corresponds to a number of separate mind spaces, each with its own coordinate system based on the unique input features. There is no division between supervised or unsupervised learning since any partial input is used to improve the mind space topography in the subspace in which it is defined. Completion of partially known inputs or inverse problems are treated in the same way as simple classifications, i.e. by finding the local maxima of $M(X,Y)$ function first in the subspace of known inputs and then in the remaining dimensions.

FSM as fuzzy expert system. Representation of data by fuzzy regions of high density in the mind space make the FSM model equivalent to a fuzzy expert system. The rules of such systems are of the following type:
IF $(x_1 \in X_1 \wedge x_2 \in X_2 \wedge ...x_N \in X_N)$ THEN $(y_1 \in Y_1 \wedge y_2 \in Y_2 \wedge ...y_M \in Y_N)$
These and more general rules may be directly programmed in the FSM network. A number of logical problems have been solved using FSM density estimation as a

heuristic in the search process. Representing discrete changes of variables for any additive $A = B+C$, multiplicative $A = B \cdot C$ or inverse additive $A^{-1} = B^{-1}+C^{-1}$ law as $\Delta A = +$ for increase, $\Delta A = -$ for decrease and $\Delta A = 0$ for no change 13 out of 27 facts in 3-dimensional feature space $(\Delta A, \Delta B, \Delta C)$ are true ($\Delta A = 0$ if both $\Delta B = \Delta C = 0$, etc.). If many such laws are applicable for N variables out of 3^N possible solutions only a few are in agreement with all laws. For example, if $A_1 = A_2 \oplus A_3$; $A_2 = A_3 \oplus A_4$ ... $A_{N-2} = A_{N-1} \oplus A_N$ only $4N+1$ solutions exist. FSM can create the mind space corresponding to product of densities generated by all $A = B \oplus C$ laws and use it effectively in the input completion task. An application to the analysis of a simple electric circuit using network that knows Ohm and Kirchoff laws [40] shows how such symbolic knowledge helps to find a solution to a problem that for most neural networks is too difficult.

Two-dimensional maps of the mind space objects centered around the identified object help to visualize the multidimensional relations among mind objects. These maps are obtained by minimization of the measure of topographical distortion:

$$0 \leq D_n(r) = \sum_{i>j}^{N} (R_{ij} - r_{ij})^2 \Bigg/ \left(\sum_{i>j}^{N} r_{ij}^2 + \sum_{i>j}^{N} R_{ij}^2 \right) \leq 1 \qquad (1.5)$$

Here r are distances in the target space (minimization parameters) and R are distances in the input space. This measure estimates the loss of information when low-dimensional representations replace high-dimensional.

To summarize, FSM is an ontogenic density network realization of the static part of the Platonic model. It creates mind objects using training data and laws constraining possible values of inputs. It enables symbolic interpretation of the objects represented by densities in feature spaces. Initialization is based on clusterization, associations are based either on the distance between mind objects (between local maxima) or on the overlaps of the densities representing mind objects. Learning is done by a combination of supervised and unsupervised techniques, adding and removing nodes of the network. Generalization is controlled by the adaptive parameters such as dispersions and by the degree of complexity of the FSM network. Implementation of typical expert system production rules is straightforward. Memory-based reasoning and input completion tasks are easily solved by identifying the relevant nodes and focusing on a single unknown variable each time. Formation of categories and metaconcepts for groups of objects is possible by investigating their projections on various subspaces. In input completion problems a projection of mind objects on a subspace of known inputs is made. Separable functions allow to drop the unknown factors and quickly find the relevant nodes selecting the most active ones. FSM may also answer questions of the type: find all objects similar to X and evaluate their similarity. This network has been applied to a number of classification problems, logical rule extraction, task completion problems and logical reasoning problems [41,46] with very good results.

1.5 Categorization in psychology: a challenge

Categorization, or creation of concepts, is one of the most important cognitive processes. Current research on category learning and concept formation ignores constraints coming from neural plausibility but there is enough experimental data from psychology, brain imaging and recordings of neural activity in monkeys [20] to make the understanding of category learning from several perspectives a feasible task for cognitive sciences. Several models of categorization of perceptual tasks have been compared by Cohen and Massaro [47], including Fuzzy Logical Model of Perception (FLMP), Gaussian Multidimensional Scaling Model (GMM), Theory of Signal Detection (TSD), Feedforward Connectionist Model (FCM) and Interactive Activation and Competition Model (IAC). All these models predict probabilities of responses in a prototypical two and four-response situations in an almost equivalent way. To show that these models contain some truth one should try to justify them as approximations to neural dynamics. These categorization models [47] may be derived as static approximations to the Platonic model described here.

A classic category learning task experiment has been performed by Shepard *et al.* in 1961 and replicated by Nosofsky *et al.* [48]. Subject were tested on six types of classification problems for which results were determined by logical rules of growing complexity. For example, Type II problems had the XOR structure (i.e. XOR combination of two features determines which category to select). Although the exact equations governing brain dynamics in category learning are unknown they may be simplified to a prototype dynamics, with two inputs and one output. In the XOR case:

$$V(x_1, x_2, x_3) = 3x_1x_2x_3 + \frac{1}{4}\left(x_1^2 + x_2^2 + x_3^2\right)^2; \quad \dot{x_i} = -\frac{\partial V}{\partial x_i} \tag{1.6}$$

Taking a general quartic polynomial form of $V(x_1, x_2, x_3; \lambda_{ij}, \lambda_{ijk}, \lambda_{ijkl})$ one may adapt such 117 parameter system to training data. Equations given above may be treated as a canonical or prototype dynamics for all tasks in which decisions are based on the XOR rule. Starting from examples of patterns serving as point attractors it is possible to construct dynamical system using a set of frequency locking nonlinear oscillators [26] modeling cortical minicolumns. The system Eq. 1.6 has attractors at $(0, 0, 0)$, $(-1, -1, -1)$, $(1, 1, -1)$, $(-1, 1, 1)$, $(1, -1, 1)$; the first attractor is of the saddle point type and defines a separatrix for the basins of the other four.

In case of Shepard experiments [48] the feature space contains axes for shape, color and size. In the feature space (Fig. 1.5) the four vertices of the cube represent the shortest transients of the trajectories. Eq. 1.6 were solved for large number of points in a cube containing the attractors. For each initial point the number of iterations in the Runge-Kutta procedure needed for convergence to the point attractor were recorded in a $T(x_1, x_2, x_3)$ matrix. These values were fitted to several functions $M(x_1, x_2, x_3)$, with best results obtained using hyperbolic tangent basis functions (accuracy within a few percent). Original dynamics based on differential

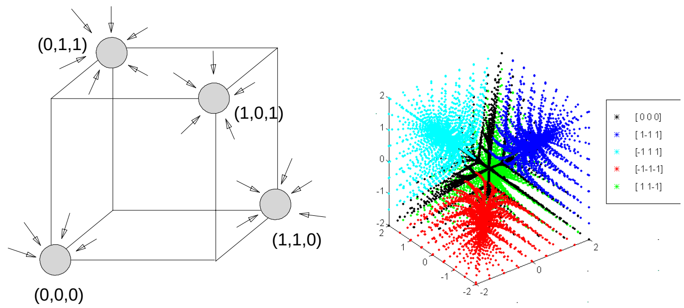

Fig. 1.5. Direct representation of the attractors in the Type II problems. The density of localized function at a given point depends on the time to reach an attractor from such initial conditions. Arrows show local gradients of this density.

equations was than replaced by gradient dynamics, with most trajectories looking very similar to the original ones (Duch and Kisielewski, unpublished).

Inverse base rate effects. People learn relative frequencies (base rates) of categories and use this knowledge for classification - this is known as “the base rate effect”. Frequently repeated stimuli create large densities of feature space objects (deep basins of attractors) and the size of these objects depends on the inherent noise and variability of the stimuli. In some cases predictions contrary to the base rates are made [49]. Names of two diseases, C (for Common) and R (for Rare), are presented to participants, the first linked to symptoms I and PC, and the second I and PR. Here PC and PR are perfect predictors of the disease C and R. Associations (I,PC) $\rightarrow$ C are presented 3 times more often than (I,PR) $\rightarrow$ R. After a period of learning participants are asked to predict which disease corresponds to a combination of symptoms. For a single symptom I or combination of symptoms PC+I+PR most (about 80% and 60%, respectively) predict C, in agreement with the base rates. However, about 60% of participants associate symptoms PR+PC with the disease R, contrary to the base rate expectations. For many years this effect has eluded explanation until Kruschke and Erickson [50] have introduced a model integrating six psychological principles of human category learning: error-driven association learning, rapid shift of attention, base rate learning, short term memory effects, strategic guessing and exemplar-based representations. Except for strategic guessing all other principles may be absorbed in construction of representations of categories rather than in processes acting on these representations.

In such experiments the answers are determined by the sizes of basins of attractors represented by the shapes of objects in the feature space. The memory function describing these objects may be fitted to obtain observed probabilities of answers, as is usually done in psychological modeling [47]. Unfortunately they are defined in 4-dimensional space and are therefore hard to visualize. The C basin is larger, extends between I and PC+I vertices, forcing the R basin to be flatter and be closer to the PR+PC vertex than the C basin is, leading to the inverse base rate effect. Neurodynamical/Platonic model interpretation is compared with psychological point of view in Table 2. Our simulations of the inverse base rate effect predict a weak dependence of the probability of categorization on the order (PC, PR) or (PR, PC), and on the timing of presentation, due to the “trapping” of the mind state by different attractors. Psychophysical experiments confirm the idea that object recognition is affected not only by their similarity but also by the order in which images are presented [51].

Neurodynamical point of view	Psychological point of view
Learning I+PC more frequent → stronger synaptic connections, larger and deeper basins of attractors.	Learning Symptoms I, PC typical for C because they appear more often.
To avoid attractor around I+PC leading to C deeper, localized attractor around PR is created.	Rare disease R - symptom I is misleading, attention shifted to PR associated with R.
Probing Activation by I leads to C because longer training on I+PC creates larger common basin than I+PR.	Probing I → C in agreement with base rates, more frequent stimuli I+PC are recalled more often.
Activation by I+PC+PR leads more frequently to C because I+PC puts the system in the middle of C basin.	I+PC+PR → C because all symptoms are present and C is more frequent (base rates again).
Activation by PR and PC leads more frequently to R because the basin of attractor for R is deeper.	PC+PR → R because R is distinct symptom, although PC is more common.

Table 2. Comparison of neurodynamical and psychological points of view in the inverse base rate problem.

1.6 Summary

Understanding human mind and its relation to information processing by the brain is the greatest challenge that science has ever faced. A fruitful approach to this problem is to identify a hierarchy of models describing the brain processes at different levels of details and to look for approximations allowing to justify, at least in principle, simplified higher level theories as approximations to more detailed lower level models. Neural dynamics of large groups of neurons is usually approximated by finite state automata, leading to description of behavior but not mind. Internal representations, or objects of mind, supported by quasi-stable (attractor) neurodynamics, may be identified with objects defined in relatively low dimensional feature spaces endowed with Finsler metrics. This leads to a Platonic model.

Why should the Platonic model make the theory of mind easier than psychological theories that we have today? First, it may be (at least in principle, and sometimes in practice) derived as an approximation to neurodynamics. Second, psychologists already use the concept of “psychological (feature, conceptual) spaces” [42]. It is easier to discuss mental phenomena using the language of feature spaces than to talk about neural states. Third, psychology lacks the concept of space understood as an arena of mind events. It was only after concepts of space and time were established that physics started to develop. Fourth, data from neural responses as well as behavioral experiments may be combined to determine topography of the mind space. Categorization problems in psychology are one area of research where models of higher cognition may meet computational neuroscience on the middle ground of the Platonic models. Fifth, Platonic model has already inspired models of neurofuzzy systems [40,41] useful for technical applications and cognitive modeling. The static version of the Platonic model has been implemented and applied in a number of problems.

Platonic model of mind, geometric model treating these feature spaces as an arena for mental events, has a great potential to bridge the gap between neuroscience and psychology. It may integrate several trends in cognitive science, in particular linguistics, vision research (object recognition) and psychology (categorization, emotions), providing a useful language for analysis of mental phenomena. It may benefit from modern geometry, theory of dynamical systems, probability and estimation theory, neural networks, pattern recognition and inductive methods of artificial intelligence. Although Platonic model is in the initial stage of development in future it may play an important role in our understanding of the brain information processing capabilities.

Acknowledgements: Support by the Polish Committee for Scientific Research, grant 8T11F 00308, is gratefully acknowledged. It is a pleasure to thank prof. Jerzy Korczak for his hospitality at the Louis Pasteur Université in Strasbourg, France, where this paper was finished, and to Rafał Adamczak and Norbert Jankowski, who were involved in development and applications of the FSM system.

References

1. Gazzaniga M. S, ed. *The Cognitive Neurosciences* (MIT, Bradford Book 1995); Churchland P.S, Sejnowski T.J, *The computational brain.* (Bradford Book 1992)
2. Newell A, *Unified theories of cognition.* (Harvard Univ. Press, MA 1990)
3. Thelen E, Smith L.B, *A Dynamic Systems Approach to the Development of Cognition and Action* (MIT Press 1994); Kelso J.A.S, *Dynamic Patterns*, Bradford Book, MIT Press 1995
4. Erwin E, Obermayer K, Schulten K, *Models of Orientation and Ocular Dominance Columns in the Visual Cortex: A Critical Comparison*. Neural Comp. **7** (1995) 425-468
5. Stillings N.A., Feinstein M.H, Garfield J.L, Rissland E.L, Rosenbaum D.A, Wiesler S.E, Baker-Ward L. *Cognitive Science* (MIT Press 1987)
6. At http://www.phys.uni.torun.pl/kmk/publications.html address a longer version of this paper is available with numerous references.
7. Black I, *Information in the Brain.* (Bradford Book 1994)
8. Dennett D.C, Consciousness explained (Little Brown, Boston 1991)
9. Bower J.M, Beeman D, *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System.* (Springer 1994)

10. Anderson C, van Essen D, *Neurobiological computational systems*, in: *Computational intelligence imitating life*, ed. Żurada J.M, Marks R.J, Robinson C.J (IEEE Press, NY 1994)
11. Földíak P, *The 'Ideal homunculus': statistical inferences from neural population responses.* In: Eeckman F.H, Bower J.M (Eds.), Computation and neural systems (Kluver 1993), pp. 55-60
12. Victor J.D, Purpura K.P, *Metric-space analysis of spike trains*. Network 8 (1997) 127-164
13. Traub R.D, Whittington M.A, Stanford I.M, Jefferys J.G.R, *A mechanism for generation of long-range synchronous fast oscillations in the cortex*. Nature 382 (1996) 621-624
14. Yanai H-f, Amari S-i, Auto-associative memory with two-stage dynamics of non-monotonic neurons, IEEE Trans. Neural Networks 7 (1996) 803-815
15. Palm G, *Cell assemblies as a guidline for brain research,* Concepts in Neuroscience, **1** (1990) 133-147
16. Hebb D, *The Organization of Behavior* (J. Wiley, NY 1949)
17. Szentagothai J, *The 'module-concept' in the cerebral cortex architecture.* Brain Research 95 (1975) 475-496
18. Abeles M, *Corticotronics* (Cambridge Univ. Press, NY 1991)
19. Freeman W.J, *Mass Action in the Nervous system* (Academic Press, NY 1975); Cowan J.D, *A statistical mechanics of nervous activity.* Lect. on Math. in Life Sciences 2 (1970) 1-57, ed. by Gerstenhaber M. (Am. Math. Soc, Providence RI); Amari S-i, *Field theory of self-organizing neural nets.* IEEE Trans. on Systems, Man and Cybernetics 13 (1983) 741-748
20. Amit D.J, *The Hebbian paradigm reintegrated: local reverberations as internal representations*. Brain and Behavioral Science 18 (1995) 617-657
21. Murre J, *TraceLink: A model of amnesia and consolidation of memory.* Hippocampus 6 (1996) 675-684
22. Pellionisz A, Llinás R, *Tensorial approach to the geometry of brain function: cerebellar coordination via metric tensor*. Neuroscience 5 (1980) 1125-1136
23. Ingber L, *Statistical mechanics of multiple scales of neocortical interactions.* In: *Neocortical dynamics and Human EEG Rhythms*, ed. Nunez PL (Oxford University Press 1995), p. 628-681
24. Elman J.L, *Language as a dynamical system*, in: R.F. Port, T. van Gelder, Eds, *Mind as motion: explorations in the dynamics of cognition* (Cambridge, MA, MIT Press 1995), pp. 195-223
25. Duch W, *A solution to the fundamental problems of cognitive sciences*, UMK - KMK - TR 1/94 report (1994), available from the International Philosophical Preprints Exchange and ftp.phys.uni.torun.pl/pub/papers/kmk
26. Haken H, *Synergetic Computers and Cognition* (Springer 1991)
27. Edelman S, Intrator N, *Learning as extraction of low-dimensional representations*. In: Medin D, Goldstone R, Schyns P (Eds.), *Mechanism of Perceptual Learning* (Academic Press, in print)
28. Crick F, *The Astonishing hypothesis. The scientific search for the soul* (Charles Scribner's Sons: New York 1994)
29. Baars B.J, *A Cognitive Theory of Consciousness* (Cambridge University Press, Cambridge, MA 1988); Newman J, Baars B.J, *Neural Global Workspace Model*. Concepts in Neuroscience 4 (1993) 255-290
30. Goldfarb L, Abela J, Bhavsar V.C, Kamat V.N, *Can a vector space based learning algorithm discover inductive class generalization in symbolic environment?* Pattern Recognition Lett. 16 (1995) 719-726

31. Shepard R.N, *Toward a universal law of generalization for psychological science*. Science 237 (1987) 1317-1323
32. P.L. Antonelli, R.S. Ingarden, M. Matsumoto, *The Theory of Sprays and Finsler Spaces with Applications in Physics and Biology*. (Kluver 1993)
33. T. Bedford, M. Keane and C. Series, *Ergodic theory, symbolic dynamics and hyperbolic spaces* (Oxford University Press 1991)
34. Hsu C.S, *Global analysis by cell mapping*, J. of Bifurcation and Chaos 2 (1994) 727-771
35. Edelman S, Intrator N, Poggio T, *Complex Cells and Object Recognition* (submitted to NIPS'97)
36. Van Loocke P, *The Dynamics of Concepts. A connectionist model.* Lecture Notes in Artificial Intelligence, Vol. 766 (Springer Verlag 1994)
37. G. Fauconniere, *Mental Spaces* (Cambridge Univ. Press 1994)
38. Ripps L.J, Shoben E.J, Smith E.E, *Semantic distance and the verification of semantic relations*. J. of Verbal Learning and Verbal Behavior 12 (1973) 1-20
39. Yanaru T, Hirotja T, Kimura N, *An emotion-processing system based on fuzzy inference and its subjective observations.* Int. J. Approximate Reasoning 10 (1994) 99-122
40. Duch W, Diercksen G.H.F, *Feature Space Mapping as a universal adaptive system* Comp. Phys. Communic. **87** (1995) 341-371; Duch W, Adamczak R, Jankowski N, *New developments in the Feature Space Mapping model*, 3rd Conf. on Neural Networks and Their Applications, Kule, October 1997 (in print)
41. Duch W, Adamczak R, Jankowski N, Naud A, *Feature Space Mapping: a neurofuzzy network for system identification*, Engineering Applications of Neural Networks, Helsinki 1995, pp. 221–224
42. Roth I, Bruce V, *Perception and Representation,* (Open University, 1995)
43. Duch W, Adamczak R, Jankowski N, *Initialization of adaptive parameters in density networks*, Third Conference on Neural Networks and Their Applications, Kule, Poland (in print)
44. Platt J, *A resource-allocating network for function interpolation.* Neural Comp. 3 (1991) 213-225
45. Duch W, Jankowski N, *New neural transfer functions*. Applied Mathematics and Computer Science (in print, 1997)
46. Duch W, Adamczak R, Grąbczewski K, *Extraction of crisp logical rules using constrained backpropagation networks.* International Conference on Artificial Neural Networks (ICNN'97), Houston, TX, 9-12.6.1997, pp. 2384-2389
47. Cohen M.M, Massaro D.W, *On the similarity of categorization models*, In: F.G. Ashby, ed. *Multidimensional models of perception and cognition* (LEA, Hillsdale, NJ 1992), chapter 15.
48. Nosofsky R.M, Gluck M.A, Palmeri T.J, McKinley S.C, Glauthier P, *Comparing models of rule-based classification learning: a replication and extension of Shepard, Hovland and Jenkins (1961).* Memory and Cognition 22 (1994) 352-369
49. Medin D.L, Edelson S.M, *Problem structure and the use of base-rate information from experience.* J. of Exp. Psych: General 117 (1988) 68-85
50. Kruschke J. K, Erickson M.A, *Five principles for models of category learning.* In: Z. Dienes (ed.), *Connectionism and Human Learning* (Oxford, England: Oxford University Press 1996)
51. Wallis G, *Presentation order affects human object recognition learning*, Technical Report, Max-Planck Inst. of Biological Cybernetics, Aug. 1996