

Computational physics of the mind.

Włodzisław Duch

Department of Computer Methods, Nicholas Copernicus University,

ul. Grudziądzka 5, 87-100 Toruń, Poland.

e-mail: duch@phys.uni.torun.pl

WWW: <http://www.phys.uni.torun.pl>

In the XIX century and earlier such physicists as Newton, Mayer, Hooke, Helmholtz and Mach were actively engaged in the research on psychophysics, trying to relate psychological sensations to intensities of physical stimuli. Computational physics allows to simulate complex neural processes giving a chance to answer not only the original psychophysical questions but also to create models of mind. In this paper several approaches relevant to modeling of mind are outlined. Since direct modeling of the brain functions is rather limited due to the complexity of such models a number of approximations is introduced. The path from the brain, or computational neurosciences, to the mind, or cognitive sciences, is sketched, with emphasis on higher cognitive functions such as memory and consciousness. No fundamental problems in understanding of the mind seem to arise. From computational point of view realistic models require massively parallel architectures.

I. PSYCHOPHYSICS AND COMPUTATIONAL MODELS OF THE MIND

Basic concepts of physics, such as energy, mass, time, temperature or strangeness are highly abstract metaphors useful in constructing models of reality. These models relate observations and measurements to other observations and measurements. In the early history of physics results of measurements were directly related to sensory experiences. In Galileo times confirmation of two independent senses was required to acknowledge a new phenomenon and to avoid self-deception (telescope, giving only optical measurements, was therefore highly suspect). Understanding the relation of objective measurements to psychological sensations was very important. Newton tried to model spectral hues by points on a circle, Helmholtz and later Schrödinger [1] by curved Riemannian manifolds. Psychological spaces for representation of pure tones, odors and tastes were also proposed.

Creation of good models to relate various features of sensory perception proved to be much more difficult than creation of models based on objective measurements of physical quantities. Methods of measuring the strength of psychological sensations in relation to the intensity of physical stimuli were developed by E.H. Weber (1834, 1846) and G.T. Fechner, whose classic book *Elements of psychophysics* was published in 1860. This book had strong influence on Ernst Mach, who developed measurement theory and wrote that “a psychophysical measurement formula assigns numbers to sensations in the same way a thermometer assigns the temperature to a state of heat.” Unfortunately it is not so simple.

Psychophysics has another important aspect, even more difficult than quantification and description of psychological sensations. “Psychophysical problem”, also known as the mind-body problem, concerns the very relations between the mental and the physical. Thus psychophysics should be placed on the crossroads of psychology, physics and philosophy. Problems raised in XIX century are still not resolved, as the recent review of the history of psychophysics has showed [2]. Psychophysics has been of marginal interest to physicists (with notable exception of acoustics and optics communities concerned with tone, speech and visual perception). This situation may change since it became recently clear that the way to understand the mind leads through modeling of neural processes at many levels, from biophysical to the systems level [3]. Computational physicists will undoubtedly play a major role in these modeling attempts. The final goal - understanding the brain and building artificial minds - encompasses much more than the original goals of psychophysics. In a sense it may prove to be the final goal of science as we know it.

I will present here a sketch of the path that leads from computational models of brain functions to models of the mind, a path from *physis* to *psyche*, something that Wolfgang Pauli always wanted to achieve. In 1952 he wrote [4]: “It would be most satisfactory if *physis* and *psyche* could be seen as complementary aspects of the same reality”. We are slowly reaching this point but there is no hope that one unified theory will be sufficient to explain cognitive phenomena at all levels. Understanding human linguistic competence at the level of complex neural networks is as difficult as

understanding hydrodynamics at the level of quantum mechanics. To understand cognition one should identify different levels of description, find the models appropriate to a given level and show how, at least in principle, higher levels may be reduced to lower levels.

Quantum mechanics, together with biochemistry and biophysics, is very useful in description of phenomena at molecular level ($1\text{ \AA} - 1\mu\text{m}$), at the level of neurotransmitters and neuromodulators, expression of genes, membrane and synaptic processes, molecular mechanisms of the long-term memory, such as the long-term potentiation (LTP) [5]. Models of a single neuron based on classical physics are sufficiently accurate to explain neurophysiological data. Neurons are represented by electrical circuits showing similar behavior, and described by “leaky integrator” equations. For groups of neurons quantum processes are important only as a source of noise in the system. There are quantum processes that can directly influence responses of single neurons, for example single photons may excite cells of the retina, but such processes seem to be rather exceptional and restricted to a few sensory subsystems. Computational biologically motivated models of neurons (for software references see [6]) describe quite accurately various details of neuronal responses but are usually too complex to use them in simulation of larger groups of neurons. Existence of stable solutions in a network of spiking neurons is determined by a single parameter (the mean firing rate) [7], explaining why associative networks based on the simplest two-states neurons have the same power of storing and retrieving information as the networks based on a very sophisticated models of neurons described by the Hodgkin-Huxley equations.

The best simplified neuron transfer functions that should be used in models of biological neural networks (spatial scales around $1\text{ mm} - 1\text{ cm}$) depend on the actual structure and functions that are modeled. At the level of computational maps (spatial extent about 1 cm) one can still use rather homogenous neural networks to model self-organized development of such structures as a result of interaction with the environment (unsupervised training). Higher level of more complex brain functions requires a number of neural modules cooperating together. It is only at this highest level that the mind-like properties of the system begin to show up and somehow the brain dynamics gives rise to the subjective mind. At this level the direction one should take in attempts to model these functions is unclear. Some experts claim that this problem may never be solved and no new ideas to solve it are in sight [8]. This is a necessary step in connecting (“reducing”) psychology to computational physics of the brain processes. The main purpose of this paper is to show how to make this final step on a path from brain to mind.

II. COMPUTATIONAL BRAIN

Psychophysics in a broad sense must be based on computational physics of brain processes. Mind is an emergent property of the brain, a very complex, modular dynamical system. Some physicists argue that incorporation of the mind or mental processes to natural sciences is possible only using quantum mechanics [9]. The long time scales of higher cognitive processes associated with conscious perception, requiring from tenth to several seconds, are in agreement with the typical times of cooperation of assemblies of noisy neurons via electrical excitations, slowed by the synaptic processes mediated by biochemical neurotransmitters. It is hard to imagine quantum processes that would be so slow.

Penrose [10] has argued that cognitive processes are non-computational in nature since formal systems are not able to answer some Gödel-type questions related to their own specification. These arguments have been discussed already by Gödel himself and repeated many times by Turing, Lucas and other philosophers (for a discussion see Penrose [10]). Human brain is too complex to contemplate any questions of Gödel type requiring full formal specification of neural machinery, therefore claims that humans are able to answer such questions about themselves are greatly exaggerated. These deliberations may only prove that it is not possible to create computational equivalent of God, a system that will have a perfect knowledge of everything, including itself. Humans do not possess such knowledge either. From the Gödel argument Penrose concludes that completely new physics is required to understand human mind, physics that should be based on noncomputable processes, but he fails to find any clues how such processes could look like. This is an example of extremely speculative approach to the problem of mind, an approach that is certainly not directed at an understanding of the human cognition.

Another issue is the computational power of the brain. With a total of about 40 billion neurons (including about 10 billion neurons in the neocortex) and the number of synapses on the order of

10^{14} , operating with a maximum speed of 100 operations per second and resolution of about 7 bits there is enough computing power and enough adaptive parameters to account for various aspects of human memory and cognition. The problem is not in the complexity or speed of information processing, as some authors looking for faster computational processes in cellular microtubules suggest [10], but with the overall organization. The brain contains dozens of large structures with rather different neuroanatomy and functions, even neocortex has highly modular structure.

Some proponents of the quantum mechanical approach to the mind [9] try to understand “thoughts” at a very vague level. Empirical Theory of Mind should be much more precise and ambitious. It should explain: basic facts about perception, e.g. stereoscopic vision and psychophysical data; dynamical optical illusions such as “color phi”, metacontrast, Stroop interference, tachistoscope results [11]; thousands of facts from cognitive psychology, such as the typing skills or the power law of learning [12]; stages of development, from infancy to adulthood, such as learning to walk, learning basic perceptual categories and knowledge structures [13]; various types of memory and amnesia; conscious and subconscious perception, relation of perception to brain events; qualia, mental content, meaning of symbols; states of consciousness (such as the dream states, daydreaming, hypnotic and other unusual states of mind); formation of ego, personality, Multiple Personality Disorder (MPD); intuition and immediate response behavior; linguistic competence, thinking and reasoning; psychiatric disorders (from anxiety and dyslexia to schizophrenia), blindsight, hysterical blindness; exceptional abilities, e.g.: “idiot savants” syndrome and many other cognitive phenomena. Great advances have been made recently in most of these areas, increasing our confidence in the soundness of computational approach to cognitive neuroscience.

Cognitive processes performed by the brain allow for construction of an internal model of reality from the sensory data. A natural approach to mind should therefore start with models inspired by the brain, models capable of learning, categorization and internal representation of the sensory data. The task may be roughly divided into two parts: low-level cognition, or preliminary analysis and preprocessing of the incoming sensory signals in the sensory reception, and the higher-level cognition, where the internal representations are used during perception, thinking and problem solving. The low-level processing of sensory data is modeled by self-organizing, unsupervised neural networks, by computational maps and by the population coding mechanisms [5]. The central problem remains: how to bridge the gap between the mind and the brain? How to link the mental and the physical? In the following sections I will sketch the solution to this problem. Short introduction to the cognitive modeling will be given first, followed by a section that introduces the main idea allowing to build a connection between the brain and the mind. A section outlining the mind space approach to the higher cognition is followed by some remarks on the problem of consciousness and final conclusions. There is no standard view or general agreement on how to model the mind. The views presented below at least offer a well defined direction and seem to solve fundamental problems of cognitive sciences, bridging the gap between the brain and the mind, justifying some approaches to the higher cognition as well as offering new fruitful practical models.

III. NEURAL AND COGNITIVE MODELING

An artificial neural network may be defined as “a processing device, either an algorithm, or actual hardware, whose design was motivated by the design and functioning of human brains and components thereof” [6]. Since neural networks are popular and almost every approximation or classification algorithm may be presented in a network form there is a tendency to add the adjective “neural” in cases where no biological motivation is justifiable. In such cases a name “adaptive system” should be preferable to “neural network”. Adaptive system A_W is a system with internal adjustable parameters W performing vector mappings from the space of inputs X to the space of outputs $Y = A_W(X)$. Neural networks may not be the best adaptive systems for all approximation problems [14].

One of the first attempts to model psychophysics of perception at a neural level was done by Rashevsky in 1938. His book [15] was republished in 1960 and pioneered the field of continuous neural models based on dynamical systems (differential equations), known as neurodynamics. The paper of McCulloch and Pitts in 1943 (reprinted in Vol 2 of [16]) was very influential and Rashevsky came to the conclusion that “the proper mathematical tool for representing the observed discontinuous interaction between neurons was not differential equation but the Boolean Algebra of Logical Calculus”

[15]. Soon it became apparent that the relation of reaction times to stimulus intensities modeled by differential equations are not easily reproducible by logical calculus. Rashevsky thought that his differential equations describe average activity of a very large number of neurons. He developed a number of highly specific models for psychophysical and neurophysiological phenomena and this line of research is still continued [17].

A. Neural cell assemblies

Although it is not quite clear how to divide the gray matter into functional units or how to average neuronal activity over these units the idea of neural cell assemblies (NCAs), advocated in the classical book of Hebb [18], is very attractive. Some neural modelers argue that the microcolumns of neocortex are the required functional units [19]. These microcolumns, distinguishable using neuroanatomical techniques, contain between $10^4 - 10^5$ neurons in a $1 - 2$ mm high columns spanning six layers of neocortex, within the cortical area of a fraction of mm^2 . Vertical connections inside a column are excitatory and their density is of an order of magnitude higher than the density of connections with neurons outside of the column. Axons of some NCA neurons spread horizontally on several millimeters enabling mutual excitation of different NCAs. Small (about 100 neurons) functional groups of neurons with inhibitory connections were also considered [19]. Although NCAs should play an important role in models of brain functions their description requires rather complex dynamical models. Neurons integrate the incoming signals and, if the total electric potential on their body integrated in a short time exceeds a threshold value, they send series of spikes. To simplify mathematical models average firing frequency of the neuron is taken as a measure of its activity. The output from a given neuron is determined by computing its activation as a weighted sum of the incoming signals (average firing frequencies):

$$I(t) = \sum_i W_i x_i(t) \quad (1)$$

where the coefficients W_i represent couplings with other neurons (due to synaptic conductivities) and are positive for excitatory and negative for inhibitory connections. If this total activation is larger than some threshold value the neuron outputs a signal with strength $f(I(t))$ determined by a monotonic function f of $I(t)$. Assuming that in an assembly of stochastic neurons the distribution of the thresholds for firing is normal (Gaussian) with some mean θ the probability of firing is described by a sigmoidal function, i.e. a function growing sharply above the threshold and reaching saturation for large values of the argument. The most common type of function with the sigmoidal shape is:

$$\sigma(I) = (1 + e^{-(I-\theta)/T})^{-1} \quad (2)$$

The constant T determines the slope of the sigmoidal function around the linear part and θ is the inflection point. It should be stressed that the use of such neuron transfer function is based on rather unrealistic assumptions and neural models useful in modeling neurophysiological phenomena on a single neuron level are based on very complex models of spiking neurons provided by biophysicists.

For neural networks that use sigmoidal transfer functions (usually multilayered perceptrons, MLP) powerful mathematical results exist showing that if there is enough data for training a universal approximator may be built with only one single layer of processing elements [20]. Sigmoidal functions have non-local behavior, i.e. their value is different from zero in an infinite domain. If a neural network composed from such elements is trained to recognize a specific class $\{X_k\} \in C$ using a set of training data the decision regions in the input space of all possible X vectors are formed by cutting the input space with hyperplanes (combinations of sigmoidal functions). There are a few disadvantages of such classification. A unique class is determined for all input values X , the system “pretends” that it knows everything, which is obviously false. Especially in the regions of the input space that are far from the training data hyperplanes, extending to infinity, enforce arbitrary classifications. If the network is complex and the training data set is small positions of the hyperplanes are to a large extent undetermined, depending on the initial state of the network. The accuracy of approximation grows with the number of adaptive parameters (weights W_{ij} in neural networks), but if the training data set is not sufficiently large the network may change into a look-up table and may not generalize smoothly on the test data set. On the other hand if the number of adaptive parameters is too small the network will not learn

much. This effect, known as the bias-variance dilemma, is analogous to the “overfitting” phenomenon in function approximation.

Another class of powerful functions used in the approximation theory is called the radial basis functions (RBFs). Some of these functions are non-local while some, such as the Gaussian functions, are localized. Networks using RBF transfer functions, called RBF networks, are also universal approximators [21]. There is no particular reason why the functions processed by the nodes of artificial neural network should be radial and in fact we have obtained better convergence with localized non-radial functions [20]. One may argue that the transfer functions localized in the input space are biologically plausible as models of single neurons as well as groups of neurons. Some neurons act in a very selective way as feature detectors. In the network of spiking neurons not only the value of the signals, but also the synchronization of the phases of the incoming trains of impulses is important. Therefore the activation $I(t)$ of neurons is high only for very specific combinations of the incoming signals. In simplified network models activations (averaged values of the incoming signals) and transfer functions are used instead of the synchronized spikes (phases and amplitudes). To account for the synchronization effects transfer functions of model neurons should be localized [20]. Neurons in the V4 visual cortex area respond selectively to such features as color, form and texture; neurons in the infero-temporal (IT) cortex react selectively to more complex combination of visual features characterizing basic natural objects such as hands or faces (frontal and profile views); neurons in the superior temporal sulcus are sensitive to facial gestures; neurons in Medio-Temporal (MT) react to the position of an object, etc [5]. It should be possible to obtain such specific reactions as a consequence of the structure of connections based on spiking neurons using neurodynamical approach, but in more complex cases it would be very difficult. Networks based on the sigmoidal transfer functions are also rather difficult to construct, train and analyze in this case. It is much easier to model directly the input/output relationships of groups of neurons reacting selectively to the features in the input data if localized transfer functions are used. After all we want to model the function rather than the precise structure of the brain.

B. Neurodynamics

Neurodynamical models pioneered by Rashevsky had random and recursive connections (cf. review article on the early models [22]). Models with excitatory connections (positive weights) only tend to the maximal or minimal values of activity but models with excitatory and inhibitory connections show a rich and interesting stable behavior. Another style of neural modeling based on stochastic approach to neurons was inspired by statistical mechanics [23] and nonequilibrium thermodynamics [24] instead of classical dynamical systems. This line explores the fruitful connections with the Ising and with the spin glass models [25] and leads to a number of interesting applications in modeling brain functions [26]. In the real brain random organization in the small scale is combined with highly specific organization of groups of neurons. Many groups of randomly connected cells, called netlets, were used for simulations showing interesting cooperative effects, including cyclic attractors [27]. Deterministic models try to get rid of the randomness by some kind of averaging procedures. However, there is experimental evidence that some groups of neurons behave in a chaotic way. For example in the olfactory bulb [36] chaotic EEG behavior is observed in the absence of stimuli and synchronized behavior arises when an odorant is present.

One of the most interesting early attempts to create computational theory of brain functions was made by Caianiello [28]. His guiding principle was the conviction that dynamical laws obeyed by the brain concern large neuronal assemblies and should not be very complicated. Caianiello proposed to divide the dynamics of the brain’s neural network according to the time scale. Fast dynamics, related to the retrieval of information, is described by the neuron equations. Slow dynamics, related to the synaptic plasticity and learning, is described by the mnemonic equations. This “adiabatic” approximation is well justified for the long-term memory (although fast learning processes, such as LTP [29], are probably involved in the long-term memory). The neuron equations may be written as:

$$a_i(t + \tau) = \Theta \left[\sum_{k,j} W_{ij}^{(k)} a_j(t - k\tau) - \theta_i \right] \quad (3)$$

where Θ is a step function (neurons are either active $a_i = 1$ or nonactive $a_i = 0$), τ is the time step, W_{ij} is the strength of synaptic connection between neurons i and j ; θ_i is the threshold of excitation

of the neuron i and k numbers previous times steps that can influence new activity $a_i(t + \tau)$. In the absence of learning the dynamics of this system, identified with the “thought processes”, has stable states of activity, described by the vector $a = (a_i)$ determined by the W_{ij} matrix.

The mnemonic equations used by Caianiello are more complicated:

$$\begin{aligned} \frac{dW_{ij}^{(k)}(t)}{dt} = & \quad (4) \\ & \left[\alpha^{(k)} a_i(t) a_j(t - \tau) - \beta^{(k)} \Theta \left(W_{ij}^{(k)}(t) - W_{ij}^{(k)}(0) \right) \right] \\ & \times W_{ij}^{(k)}(t) \Theta \left(A_{ij}^{(k)} - W_{ij}^{(k)}(t) \right) + \text{inhibition} \end{aligned}$$

The inhibitory terms are analogous to the excitatory ones. The first term in these equations is of the Hebbian type [18], i.e. it is proportional to the product of the pre- and post-synaptic activities. Second term, with $\beta \ll \alpha$ describes forgetting, while the last term restricts the connection strength to maximum values preventing their unbounded growth. Networks of processing elements operating in accordance with the neuron and mnemonic equations were used by Caianiello to study learning, forgetting, conditioning, analysis and spontaneous formation of patterns of reverberations. Logic plays a role of constraints on the type of behavior of the dynamical system. One may expect all kinds of effects in such complex system, including chaotic and quasi-periodic attractors and nonlinear resonances. Characterization of this system requires determination of spontaneous modes of reverberation from neuron equations. Short reverberations appear with the frequency of 10 Hz (assuming realistic time quantization connected with the average firing rate of biological neurons), in agreement with the observed EEG recordings. In the brain stable reverberations of a few neurons were observed lasting for minutes [28]. Epileptic seizures are one possible form of catastrophic instabilities in the network. Analyzing the mnemonic equations Caianiello points out that more realistic description of the brain should contain at least two additional structures: reticular activation system necessary for attention and thalamic structures controlling emotions.

Many other models of neural networks have been developed, for example perceptrons and the multilayered versions of perceptrons that are so popular in applications [16], but these models are not interesting from the cognitive modeling point of view. In fact the model of Caianiello, although quite successful for qualitative explanations, is not specific enough to explain quantitatively experimental data (Caianiello admitted in his later years that his model explains everything, therefore it does not explain anything). The book by D.S. Levine [30] reviewing various cognitive models does not even mention his model – more specific models of associative learning, sensory representation, lateral inhibition, competitive learning, conditioning, attention, reinforcement, coding and categorization, control, optimization and knowledge representation are discussed instead.

C. Computational maps

As the first step towards model of the mind one should consider the formation of features that the subjective experience is composed from. There are two basic mechanisms that the brain is using to analyze the sensory signals and to extract useful features from them: self-organizing maps and population coding.

Experimental data on neural mechanisms leading to a formation of orientation and ocular dominance maps in the primary visual cortex are quite detailed. More than ten computational models have been proposed to account for these data [31]. Competitive Hebbian models describe the development of visual system on the mesoscopic level close to the resolution of neurobiological experimental data. In orientation and ocular dominance maps these models predict global disorder and anisotropies of the localized responses of the neural cell assemblies, singularities and fractures, and are able to simulate development of feature detectors under the exposure to a restricted set of oriented visual features (for example, when only stripes are shown, or when monocular deprivation is enforced). Correlations between these two types of maps are also well reproduced. Very successful models are based on the self-organizing feature maps [32]. Response properties of cortical cell groups located at position \mathbf{r} in the visual neocortex involve the retinal location $(x(\mathbf{r}), y(\mathbf{r}))$, the degree of preference for orientation

$(q(\mathbf{r})\sin(2\phi(\mathbf{r})), q(\mathbf{r})\cos(2\phi(\mathbf{r}))$ (orientation maps code for 180 degree periodic orientation), and the ocular dominance $z(\mathbf{r})$. Feature vector $\Phi_t(\mathbf{r})$ composed from these five features evolves according to:

$$\Phi_{t+1}(\mathbf{r}) = \Phi_t(\mathbf{r}) + \alpha h_S(\mathbf{r}, \mathbf{r}')[\mathbf{V}_{t+1} - \Phi_t(\mathbf{r})] \quad (5)$$

with $0 < \alpha < 1$ and the stimulus \mathbf{V}_{t+1} chosen at random using some probability distribution. The local neighborhood function h_S is defined as:

$$h_S(\mathbf{r}, \mathbf{r}') = \exp(-\|\mathbf{r} - \mathbf{r}'\|/2\sigma^2) \quad (6)$$

$$\mathbf{r}'(\mathbf{V}, \Phi(\mathbf{r})) = \min_{\mathbf{r}'} \|\mathbf{V} - \Phi(\mathbf{r}')\| \quad (7)$$

Each presentation of a stimulus V leads to a change of the feature vector $\Phi(\mathbf{r}')$ of neurons in the neighborhood of \mathbf{r}' , i.e. to a change of features coded by the group of neurons that are already most sensitive to this particular stimulus.

Another feature detection mechanism that seems to be ubiquitous in the brain is known as population coding [33]. Initially it has been discovered in the motoric areas of the brain, but now it has also been found in premotor, parietal, temporal, occipital and cerebellar cortex. Here a whole population of neurons reacts simultaneously to the incoming stimuli. A model of this neural response assigns a unitary vector u_i to each cell in the population, oriented in the direction of movement corresponding to the maximum cell activity. A population vector is formed summing the unitary vectors u_i with weights equal to the discharge frequency $A(u_i, V_e)$ (activity) of each cell. This frequency depends on the feature vector V_e related to the event e , such as a presentation of some stimulus or motoric action. For example, the population vector may code an implicit representation of the direction of movement. Experimental verification of this model involved recordings from the motor cortex of monkeys performing mental rotation tasks. In this case the cell activity for a rotation vector V was computed from:

$$A(u_i, V) = u_i \cdot V = \alpha \cos(u_i, V) + \beta \quad (8)$$

Experiments show that not only the distribution of activities of neurons in the population represents the direction of movement in space but also the population vector rotates during the mental rotation tasks and the norm of the population vector determines when the actual movement will take place. Both information and the significance of this information is represented in the brain via the population coding mechanism (dual coding principle).

Mussa-Ivaldi [34] has suggested that an arbitrary vector field corresponding to a complex information encoded by a population of neurons may be reproduced by a linear combination of elementary basis fields. In particular force fields generated by the premotor circuits in a frog's spinal cord are a combination of only four basis fields. Motor behavior is obtained as a superposition of time-dependent basis fields, or pattern generators:

$$F(x, t) \propto \sum_i W_i \phi_i(x, t) \quad (9)$$

This may be the nature's solution to the function approximation problem. Dual population coding of more abstract multidimensional features provides a model for representation of complex information facilitating also the use of this information by other mental processes [35]. Consider a feature space (for example, a color, a phoneme or a physical space) and an object represented by a feature vector V_e (vector of attributes) in this space. A population of neurons may code such objects if each cell in the population responds to a particular combination of features in a localized way, for example by changing its activity according to

$$A(u_i, V) = P_e e^{-\frac{\|u_i - V\|^2}{\sigma^2}} \quad (10)$$

where P_e is a scale factor depending on the relevance of the event e and the σ parameter depends on the selectivity of neural cells. The metric in the feature space is defined by the norm $\|\cdot\|$. This model has been applied to the local motion detection by the mediotemporal cortex area as well as to the formation of categories of basic objects of perception by NCAs in the infero-temporal cortex [35]. The distribution of activities in the whole population encodes reaction to the event-related feature vector and the total (summed) activation may be regarded as the pertinence of this vector. Both feature vector

and its pertinence are coded by the same activity of neural cell assemblies and context dependent sensitivity of the pertinence may explain the interaction between mental representations and operation of selective attention.

Population coding and self-organized maps provide mechanisms for analysis of the sensory input data into a combination of features. Assemblies of neurons react selectively to some of these combinations. At the brain level more and more details are unraveled. The basic question is: how to proceed from the brain to the mind events?

IV. FROM BRAIN TO MIND

It is impossible to talk about mind events in terms of computational processes of neural hardware, as it is impossible to talk about properties of the liquid water in terms of properties of the hydrogen and oxygen gases or even properties of the individual molecules of water. Cooperative properties in interacting systems are always qualitatively different from properties of the constituents. It is not surprising that specific interactions of a large number of neurons give rise to qualitatively new properties. The emergence of higher cognitive functions from the complexity of the brain is not more strange than the emergence of wetness from a mixture of two gases. At each level of description an appropriate language and theory should be used. The flow of water in a river is described by hydrodynamics, the theory that uses specific concepts that cannot be directly derived from quantum mechanics. It is impossible to speak about mental events using directly the language of neural events as it is impossible to speak about turbulence in quantum mechanical terms. An appropriate language should be created for cognitive sciences.

The question “what is the mind” begs for a model, a language to speak of the mind. Is it possible to answer such question? Ultimately we may only explain what is the structure of the mind, as we can explain what is the structure of the matter, even though we cannot define what the matter is. In physics we use with great success concepts such as energy, space or time, concepts that are not reducible to simpler concepts. What kind of an answer can we expect to the question: “what is the mind” ? Will an answer based on physical concepts, such as “the mind is a result of zero point vibration vacuum fluctuation interactions with the coherent states of microtubules” satisfy us? Or an answer claiming that the mind is a result of logical properties of some mathematical structures? Such answers are irrelevant to the understanding of cognitive processes and to the cognitive sciences in general. A meaningful language to describe mind events is needed. It should allow for more precise description of mind events and should lead to models of mind that can be tested in computer simulations. The language should be extendable but at the initial stage it should cover only the basic features of the mind, in a similar way as the mathematical language of classical physics allowed to describe planetary motions and other simple physical systems.

In the Platonic world of mind there are only abstract concepts, represented by symbols. Two-year old children may not yet use symbols and yet their minds are already quite complex. What are the real mind events made from? They are composed primarily from the preprocessed sensory data and are mostly recognitions, recollections and memorizations. Direct sensory perceptions and iconic memories of past experiences form primary mind objects, while secondary, more abstract objects acquire their meaning through their relations with the primary objects. Once a primary object, an experience involving bodily sensations as well as some internal mind states, is committed to the long-term memory it can be recalled back, reinstating the brain/body states, thus repeating the experience (mental training, for example in music, really works).

The ambitious model of Caianiello has not influenced the mainstream of neural models of cognition because it lacked the modularity and specificity of different structures of the brain. Some of the insights offered by this model may ultimately prove to be true. It is clear that stable reverberations in the brain are connected with thoughts and perceptions. Direct observation of neural activity during such cognitive tasks as smelling [36], hearing words and meaningless sounds [37] or watching a series of pictures by monkeys [38] shows that global reverberations, interpreted as synchronized activity of a number of neural cell assemblies, correspond to perceptions and thoughts. Synchronization of oscillations of groups of neurons in the gamma band of EEG has been observed in many areas of neocortex as a result of visual stimulation [39]. Attractor character of the neural dynamics [26] has been demonstrated already in the experiments performed on cats by John *et.al.* [40]. Cats were trained

to react to two different frequencies of pulsating light. Intermediate frequencies were leading to one of the two dynamics of the visual neurons and to the corresponding behavior of animals.

Unfortunately neural systems showing interesting behavior are complex and difficult to analyze. Dynamical systems in physics are usually analyzed in a low dimensional space, rarely higher than five dimensional. Networks showing interesting properties may have hundreds or thousands of neurons and such large number of parameters makes them difficult to understand and control. Yet large groups of neurons cooperate and behave as if govern by a low-dimensional parameter space dynamics [41]. Biological systems are remarkably stable and resistant to noise in the input data as well as to the noise in the nervous system itself (stochastic nature of neural activations due to the fluctuations in ionic concentrations, neurotransmitter release, synchronization problems). Large artificial neural networks with many adaptive parameters tend to work as look-up tables if the number of training examples is not sufficiently large. Adding noise to the training data is equivalent to regularization of the error function, smoothing the approximation performed by the network [42]. Adding small amplitude noise to dynamical system reduces sensitivity to the initial conditions leading to more regular basins of attractors.

Since neurodynamical models are too difficult to use them for modeling complex functions of the brain approximations are necessary. Very interesting results have been obtained in recent years using symbolic dynamics [43]. The whole phase space is divided into hypercubes, each with a symbolic name, and a trajectory passing through these hypercubes is replaced by a string of symbolic names. In effect one obtains a low-resolution picture of the phase space. Another way to simplify the description of dynamical systems is to observe that the most important information about dynamical (brain) states is in the structure of attractors and repellers, and in the probabilities of transitions between different attractor states resulting from external perturbations or internal noise.

How are these attractors formed in the brain? Stable reverberations of neural cell assemblies, identified with determination of categories, thoughts and perceptions, arise because slow learning processes (“mnemonic equations” of Caianiello) make synaptic changes creating appropriate conditions for these reverberations. Due to the evolutionary selection processes the brain has learned basic categorization of lines, colors, shapes, visual objects, sounds, smells and other sensations. The low-level processing described in the previous section provides the stimuli for the higher-level cell assemblies (frontal and parietal cortex areas) and some of these NCAs respond by reaching attractor states while other groups of neurons stay in a chaotic state. In psychological terms this means that in the space of internal features of representation a particular combination of these features, corresponding to a particular category of perception, has been recognized and is kept active for some time. Such feature spaces (also known as mental spaces, conceptual spaces or mind spaces) were popular in psychology since a long time [44]. It is possible to analyze mental events in feature spaces. The correspondence of objects (concepts, categories) in feature spaces and attractors in the dynamics of the brain spans a bridge between the mental and the physical [45] (cf. Fig. 1).

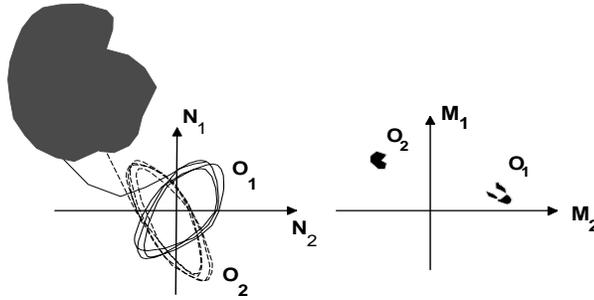


FIG. 1. Correspondence between the brain events described in the phase space of the brain function model based on a dynamical system and the mind events taking place in the feature space. Chaotic regions of the phase space correspond to random walks in the feature space.

In physics simple models and methods are of greatest value. Consider a non-linear oscillator with many attractor states, such as the Chua oscillator [46]. Such systems consist of a linear inductor L , two capacitors C_1 and C_2 , a resistor R and a nonlinear element called Chua diode, which acts as a voltage-controlled resistor. Attractor states code some information and may serve as a model for memory traces, as in the case of the model of olfactory cortex [36]. We are really interested only in the classification of stable states of such systems. One way to perform this classification is by dividing the phase space into the basins of attractors. For simple dynamical systems like the Chua oscillator the phase space is equivalent to the feature space. However, one should remember that the phase space for neuronal cell assemblies is very large while the feature space may be quite small (large groups of neurons code such low-dimensional features as colors or phoneme recognition).

Multi-Layer Perceptron (MLP) and RBF neural networks may learn the borders of basins of attractors thus giving a proper categorization of stable states of the dynamical system. I will introduce here an approximation to the behavior of a dynamical system by a simple gradient dynamics in the feature space. The phase space may contain many internal variables while the feature space only the input, or control variables. In the two-dimensional cases points in the phase space are sometimes color-coded (for example, in the representation of iterative maps) to show how quickly starting from these initial conditions the system will reach an attractor. One may analyze the transients and the attractors in a number of ways, for example by looking at Lyapunov exponents or using symbolic dynamics [43]. Since we do aim at rough characterization of a dynamical system rather than at a precise description a coarse resolution in the symbolic dynamics is usually sufficient. The distance from an attractor may be visualized by the “equitime lines” (or hypersurfaces). Each basin of an attractor has a central part and a peripheral part, where the dynamical system needs a long time to reach the attractor. The system is more resistant to external perturbations or internal noise when it is in the inner basin region and more sensitive when it is in the peripheral region.

This behavior may be modeled in several ways. The simplest is to introduce in the feature space a density $\rho(X)$ proportional to the time it takes to reach the attractor starting from X . This density should be negative for basins of attractors, corresponding to the minima in the $(X, \rho(X))$ space and positive for the repelling areas of the feature space, corresponding to the maxima. External influences

or the internal noise induce transitions between various attractors of the dynamical system. Introducing stochastic gradient dynamics on the $(X, \rho(X))$ surface one may try to reproduce the behavior of the dynamical system under perturbation.

At the first glance this approach may seem similar to the Haken's synergetic computers approach [47], but this is not the case. Synergetic approach to pattern recognition is based on the point attractors and simplified analysis of stability. Such analysis is hard to perform in general case. Consider the simplest case of a dynamical system that should learn to perform the XOR operation, i.e. the mapping (for convenience -1 is used instead of 0)

$$(x, y) \rightarrow z; \quad (-1, -1) \rightarrow -1; (-1, +1) \rightarrow +1; (+1, -1) \rightarrow +1; (+1, +1) \rightarrow -1$$

One can find a dynamical system $W(t) = (x(t), y(t), z(t)) = (x_i(t)); \dot{W} = F(W(t))$ such that the XOR values are point attractors of the system. For example, assuming that

$$\dot{W} = -\frac{\partial V}{\partial x_i} \quad (11)$$

$$V(x_1, x_2, x_3) = \sum_{ij} \lambda_{ij} x_i x_j + \sum_{ijk} \lambda_{ijk} x_i x_j x_k + \sum_{ijkl} \lambda_{ijkl} x_i x_j x_k x_l \quad (12)$$

one can train this 117 parameter system using the XOR training data and come to the following set of equations (additional requirements minimizing the number of non-zero λ parameters were used below):

$$V(x, y, z) = 3xyz + \frac{1}{2} (x^2 + y^2 + z^2)^2 \quad (13)$$

$$\dot{x} = -\frac{\partial V}{\partial x} = -3yz - (x^2 + y^2 + z^2)x$$

$$\dot{y} = -\frac{\partial V}{\partial y} = -3xz - (x^2 + y^2 + z^2)y \quad (14)$$

$$\dot{z} = -\frac{\partial V}{\partial z} = -3xy - (x^2 + y^2 + z^2)z$$

We can approximate the behavior of this dynamical system quite well with a combination of four localized functions (for example with Gaussian functions) centered in the (x, y, z) corners of a cube (Fig. 2, where local gradient field is marked with arrows).

$$\begin{aligned} M(x, y, z) &= \sum_{k=1}^4 G(x_k, y_k, z_k, \sigma) \\ &= G(-1, -1, -1, \sigma) + G(-1, +1, +1, \sigma) + G(+1, -1, +1, \sigma) + G(+1, +1, -1, \sigma) \end{aligned} \quad (15)$$

The dispersion of these functions may either be related to the resolution of input data (what do we still consider to be -1 and what $+1$) or to the basins of attractors of the dynamical system. The number of parameters (12 Gaussian centers and σ) is in this case much lower and the system is quite easy to analyze. An alternative approximation is based on the local coordinate systems, each defining an independent metric $\|X\|_k$ describing the distance between a given point X and the attractor k (in the sense of the number of iterations or time steps to reach this attractor). This is a generalization of the nearest neighbor classification method known from pattern recognition (cf. [32]), in which case all metric functions are identical and one global metric may be introduced.

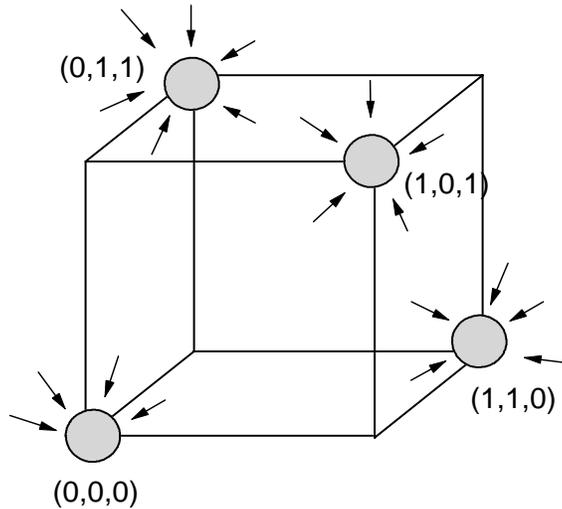


FIG. 2. Direct representation of the attractors in the XOR problems. The density of localized function at a given point depends on the time to reach an attractor from such initial conditions. Arrows show local gradients of this density.

Models of the brain functions at the level of dynamical systems are difficult. Approximations based on the direct modeling of the input/output relationships are easier and lead to the feature spaces, introducing an arena for mind events. Due to a high dimensionality of the phase space and relatively low dimensionality of the feature space such approximations must have statistical nature. The dynamical system should be perturbed and left in some internal state until it stabilizes; then the input features are presented and the evolution of these features to a new state followed. Combinations of features leading to the same type of dynamical behavior form clusters $\rho(x)$ in the feature space. Vector quantization techniques based on noisy prototypes are used to represent these clusters. The state of the system is represented in the feature space by the input vector of feature values, evolving in a stochastic way (due to a noise in the system) from its initial value to maxima of these $\rho(x)$ clusters. Instead of the neural network dynamics in the phase space vector quantization with stochastic gradient dynamics in the feature space is used. Approximation to the transition probabilities between different attractors of a perturbed dynamical system may require additional dimensions (internal features) in the feature space.

Consider an example of the olfactory cortex. It is composed from a three layers of neurons, with the top layer of inhibitory interneurons connected to the olfactory bulb (input), the middle excitatory layer of pyramidal neurons and the lower layer of inhibitory “feedback neurons” (Fig. 3). The inputs $I_i(t)$ come from the chemoreceptor cells in the nose. Skarda and Freeman [36] have analyzed this system using biologically plausible spiking neurons. A simplified model of the olfactory cortex is similar to the Hopfield network and other recursive networks [16]. The time evolution of neural activities $A_i(t)$ is given by the following equation [48]:

$$\frac{dA_i}{dt} = -\frac{A_i}{\tau_i} + \sum_{j \neq i} W_{ij} g_j(A_j(t - \Delta_{ij})) I_i(t) + \xi(t) \quad (16)$$

$$g_j(A_j) = C Q_j \left\{ 1 - \exp \left[-\frac{\exp(A_j) - 1}{Q_j} \right] \right\} \quad (17)$$

where τ_i is a characteristic time constant, W_{ij} are connection weights, Δ_{ij} are time delays between the two i and j neurons, and $\xi(t)$ is the noise. The transfer function g_j has been determined from the experimental work [36] and is of the sigmoidal type, with the gain parameter Q_i (representing arousal by neuromodulators) and a normalization factor C . In computer simulations with 100 units in each layer this dynamical system shows interesting behavior: depending on the value of Q it may have a point attractor, a limit cycle, a strange attractor, long transients similar to chaos and truly chaotic

behavior. The system has been analyzed from the point of view of its internal dynamics and the trajectories presented in [48] compare the activity of two arbitrarily selected neurons (Fig. 3). Such models are very interesting because they allow to some degree comparison with neurophysiological data. However, from this point of view it is hard to understand what happens in the mind of a sniffing animal. The odor feature space (early descriptions of this space tried to represent it as a prism [49]) is based on the input dimensions I_i only, characterizing arousals of different chemoreceptors. Prototype odors are stored in this space as fuzzy objects (Fig. 3). The input signals $I_i(t)$ and the noise $\xi(t)$ push the mind state vector $X(t)$ randomly around the whole space, the state vector roughly following the input signals. If the input signal stabilizes for a while (an odor is presented to the system) the gradient dynamics brings the mind state vector $X(t)$ to the $I_i(t)$ region of the feature space. The size of the steps in gradient dynamics corresponds to the degree of arousal (parameter Q in the dynamic model). If the arousal is very high the mind state vector jumps all over the space in a chaotic way – panic states of mind correspond to chaos in the dynamical system. If the arousal is very low the system state vector $X(t)$ needs a long time to reach the $I_i(t)$ region and the noise in the system makes its trajectory similar to random walk. If after a short time no object is found in the input region a new memory trace starts to form. Memory traces are formed and strengthen any time the system abides in some region of the feature space.

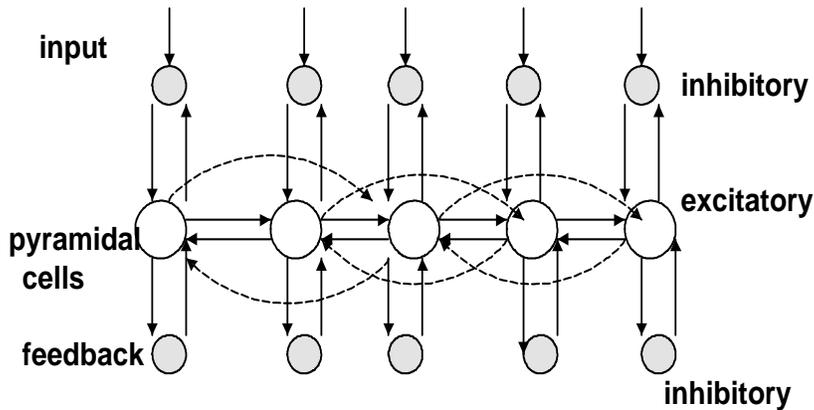


FIG. 3. Model of the olfactory cortex. The top layer of inhibitory interneurons receives the inputs from the olfactory bulb, the bottom layer contains feedback inhibitory interneurons and the middle layer excitatory neurons with lateral connections. The model shows rich behavior including limit cycles and chaotic regions of the phase space.

V. HIGHER COGNITION

The approach described here provides a mechanism to reduce mind events to the brain events. The model of mind is composed from a number of linked feature spaces, each corresponding to a certain brain structure. At the highest level feature detectors contribute to the global dynamics of the brain. In the dynamical model a single complex oscillator or a small recurrent neural network represents a localized brain function, such as those realized by the motoric, sensory or one of the emotional areas. A few such subsystems representing different modalities and feature detectors are coupled together and provide parameters for the global, distributed system, responsible for the multimodal information processing. Subjective states of the total system are identified with the global dynamical states – in case of humans they involve large parts of the neocortex (the frontal lobes, parietal areas) and subcortical structures. The highest level dynamics corresponds to a feature space called further the mind space. Combinations of all input features in the mind space define mind objects.

Some features of the internal mind spaces objects (representations) are of binary type: existing-nonexisting. Other features have various degrees, for example numerical values or sizes. A “horse”

has size, shape, head, tail, mane, makes sounds, likes to run, eats grass and is defined by all these and many other properties. Each of these features is to some degree fuzzy since they are approximations to a noisy dynamics of the brain NCAs. The symbolic name “horse”, as well as a few features of the horse mentioned above easily “activates” the corresponding object in our mind, i.e. the mind state vector $X(t)$ moves to the region of the mind space where this object is stored. This particular combination of features forms a unique category. The meaning of the object is grounded in the combination of all relevant features of inner representation, some of them related to analog sensory data. More abstract information is represented as objects in form of kinesthetic image schemes and time-related information is converted to spatial relations [38].

Mind space is thus defined by the coordinate system based on all features of internal representation. Objects in the mind space are defined by the “mind function” $M(X)$ for all relevant features X_i . Nonzero values of the mind function define these objects as a fuzzy regions in the mind space. Topographical relations of objects in this space are very difficult to imagine because of the large number of dimensions involved. Problems with understanding how the mind works are to a large degree connected with the difficulties in understanding relations in multidimensional spaces. The mind function, defined in the mind space, represents all fuzzy objects (prototypes blurred by noisy dynamics) that such system is able to recognize, i.e. correctly classify using partial description or distorted inputs. Cognitive system is able to modify the contents of the mind space by adding more objects (learning and remembering), modifying existing objects or learning new associations (changing topographical relations between existing objects). However, the adiabatic learning hypothesis still holds: topography of the mind space defined by the mind function $M(X)$ changes slowly in comparison to the changes of the mind state $X(t)$.

At present we have considered several ways to approach the problem of simplified description of feature spaces. First, the functional approach leads to a generalization of the Radial Basis Function model [21], especially in its constructivist form, where the network grows with the incoming data while the clusters in the feature space are constructed. This approach has been used in the Feature Space Mapping (FSM) system [20], where learning is presented from the geometrical point of view as finding the best possible description of the “mind objects” in the feature spaces. From this point of view the best neural processing functions should have small number of parameters and should allow for a great flexibility in defining regions of non-vanishing density in the feature space. Recognition is identified with the local maximum of density in which information about the object is stored. The input is composed from distorted or partially known vector of features pointing to a region in the feature space where the search for the object (local maximum of density) is made. An application of the functional approach to the “naive physics” or qualitative physics problem, i.e. mind models of the basic concepts of electric circuits, has already been presented [20].

Another way of describing the feature space, based also on a functional approach, is to use the Hilbert space formalism defining transition probabilities by overlaps of the densities representing different attractors, i.e. different dynamical states or categories. Field computation in the brain based on a continua of microfeatures were developed into a model of a *simulacrum*, a continuous representation of information [50]. Other connections worth exploring are deterministic finite automata (DFA) and their stochastic versions, which may be used to model transitions between different attractors in dynamical systems and vice versa [51]. A theory of conceptual mental spaces has been developed in cognitive linguistics [52].

This approach solves the neurons vs. symbols debate [53] defining the language useful in description of cognitive states, mind events and relation of these events to the dynamics of the brain. The problem of mind modeling is reduced from the problem of neural dynamics to a problem of finding the best approximation for this dynamics, for example by a geometrical representation of objects in the feature spaces. Models of cognitive systems based on the feature space concept are realized in a natural way by the neurofuzzy systems, combining neural optimization of adaptive parameters (learning) with symbolic representation based on fuzzy logics [20]. Learning and categorization is presented in such systems as a problem of constructing an appropriate geometrical model of the data in feature spaces of much lower dimension than the number of parameters necessary to describe the dynamics of the underlying network.

Artificial intelligence aims predominantly at higher cognition: language, thinking, reasoning, problem solving, expert knowledge processing. Most of the research in this field is based on the assumption that higher cognition is based on symbol processing (cf. the book “Unified theories of cognition” by Allen Newell [12]). Computational models of mind should lead from Artificial Intelligence to models

of Artificial Minds. Concepts of the mind space, mind function and mind states, the role of fuzzy logic and the dynamics of mind states, associations as coupling of attractors, adaptive resonances and descriptions of the mind space based on analogies with quantum mechanical Hilbert space formalism seem to be fruitful in establishing the bridge between neurosciences and cognitive psychology. What is needed is a systematic investigation of different levels of approximations to the brain's dynamics.

It is clear that the mind space approach applied to the higher cognition leads to a memory based models of reasoning. Objects are associated with each other, or are judged to be similar, if they are close in the mind space. MBR, or memory-based reasoning, seems to be superior to neural network models in such complex tasks as learning pronunciations of English words using a small (700 words) training sample, predicting secondary structure of proteins, classification of free-text examples of arbitrary length in the real world applications and in many other tasks [54]. MBR uses various metric functions to make associations based on the memorized objects but it does not make them fuzzy and does not use local learning techniques to optimize their shapes, as does FSM [20]. Newell [12] based his SOAR theory on the assumption that experts use 50-100 thousands of "chunks" of knowledge, memorized in form of production rules. We have already shown how the feature space may replace production rules in reasoning [20]. One may also justify probabilistic reasoning by introducing an appropriate metric in feature spaces.

In developmental psychology in recent years the language of computational models and dynamical systems [13,41] becomes a new foundation for the whole field. Ultimately large and complex models of brain structures will have to be simulated using neural hardware. Several large-scale projects, some based on custom neural chips, are in the design phase (for a review see [55]). Brain-size neuro-computer (about 10^{10} neurons), built using current technology chips, organized according to a "fractal architecture", would fill the room 32 by 10 by 10 meters (about the size of Mark I, one of the first computers), and should have about 25 TBytes of RAM [55]. Such a large size is due to the essentially two-dimensional structure of integrated circuits. Artificial mind models based on the mind space idea approximating the dynamical models should be much simpler and smaller.

VI. MIND, CONSCIOUSNESS AND PHILOSOPHICAL ISSUES.

The model presented here seems to solve the fundamental problems of cognitive sciences. Mind is a function of the brain and the environment. The role of large parts of the brain is to provide various features necessary to set up the mind space for an individual mind. Environment provides the training data to set up the objects and relations between them in the mind space. The mind is an emergent property of the brain, unique and subjective, depending on the individual experiences and history. It is embodied in neural hardware. Although we may talk about mind events taking place in an abstract mind space activation of mind objects leads to physical actions. Linguistic symbols are parts of multidimensional mind objects, although due to the complexity of linguistic subsystems and their coupling with the global dynamics of the brain several quite complex mind spaces are necessary [52].

The mind-body problem does not arise since the mind and the brain are just two sides of the same coin. Symbols are grounded in real world experiences. Qualitative character of experiences is related to subjectivity arising from the uniqueness of each organism. At the dynamical level of description different internal states of the brain (due to the sensory inputs or internal dynamics) are specific to a given organism, depend on its history and conditioning. Consider a rat conditioned to associate color green with tasty food and color red with electric shock. Representing the level of stress on one axis (measured by the arousal of the amygdala and lateral hypothalamus structures) and the perceived color on the other axis we may visualize the mind objects that were created due to the conditioning (Fig. 4). The qualitative character of rat's experience is quite obvious.

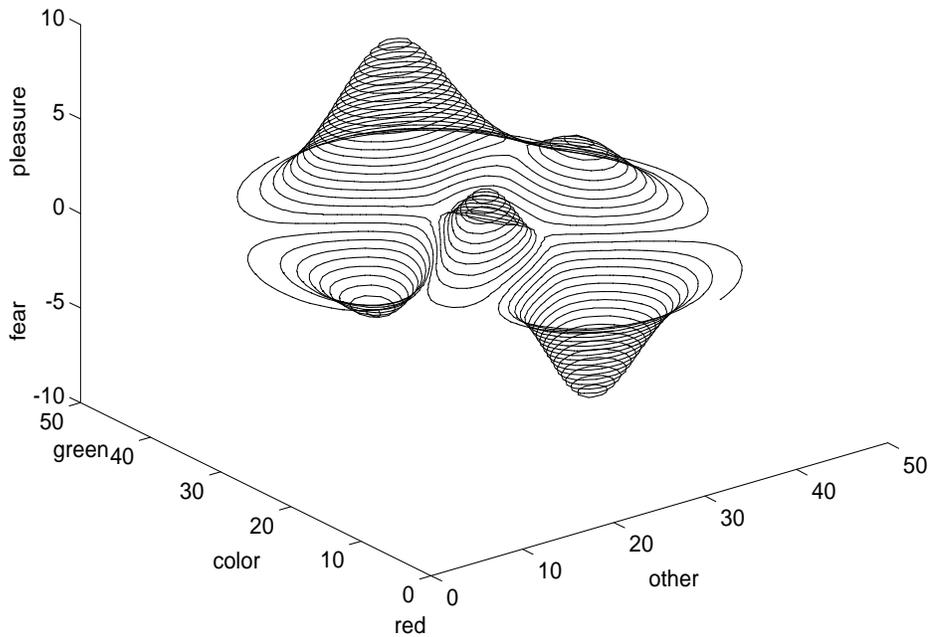


FIG. 4. Mind space of a rat conditioned to respond with fear to a red light and joy to a green light. The density of fuzzy mind objects created by this conditioning in the mind space of a rat is represented by the density of lines in the drawing.

From the subjective, experiential point of view the language of abstract mind events is more relevant than the language describing the underlying brain dynamics. One may say that if two mind objects in the mind space are close the probability of association between them is high and they will be labeled as similar. In this case high transition probability will be found in the brain dynamics between the two attractor states representing such objects. One may also argue for the Platonic view of the mind: potentially all mind objects are in the universal mind space, although only something as sophisticated as the brain can to some degree recognize and thus copy them to a local mind space. The objects and events of mind are to a large degree realized in a similar way in different brains of the same species. Mind space as an arena of mind events is a useful metaphor, leading to a specific language to speak about mind events, a language that is much more natural and easier to use than specification of brain states.

It should be clear that in this view consciousness or awareness is a subtle reaction of the mind/brain/body system that requires a very complex mind space with a large dimensionality. In his cognitive theory of consciousness Baars [56] has introduced the model of consciousness seen as a global workspace. The idea of a mind space as an approximation to the dynamics of the brain gives support to this model from quite different perspective. In the long run ideas presented here should lead beyond metaphoric use of concepts from nonlinear dynamics to psychodynamics and psychoanalysis [57], although elucidation of the mind space topography at such a level of sophistication as needed in psychodynamics may be very difficult.

VII. SUMMARY

In a broad sense psychophysics is still being born. It should be one of the core sciences exploring the relations between the brain and the mind. Many branches of science contribute to the emerging identity of cognitive science as a unique science aimed at understanding the information processing capabilities

of the brain, including mental phenomena. Neurosciences, cognitive psychology, linguistics, computer science (artificial intelligence) and even philosophy have contributed to cognitive science. Contribution of physics is indirect, via the experimental techniques in the brain research, via biophysical models at the level of single cell, statistical methods applied to the recurrent neural networks [26] and various dynamical models of the brain processes [16,28]. Contributions of physics to cognitive science will not be recognized until psychophysics, the branch of physics devoted to understanding the relations of the brain and mental processes, will not establish itself within physics first. Most of the computational work on the brain is done under the “computational cognitive neuroscience” heading by experts coming from various backgrounds. At present computational physicists and computational mathematicians are the best qualified scientists to contribute to this new, exciting field.

In this paper I have outlined a new approach to modeling the mind functions. I have argued that a reasonable view of the mind, including such elusive concepts as consciousness, is possible, and that all we can hope for is to create many theories, at different levels of description, building bridges between them. The concepts of neural cell assemblies and their attractor states, identification of the global brain dynamics with conscious perception, and the approximation of biologically motivated dynamical, modular neural networks by feature spaces and mind spaces, replacing complex dynamics by activation of mind objects, seem to be such a bridge between the brain and the mind, or neuroscience and psychology. If physics is understood as an attempt to understand Nature than understanding of the mind is its greatest challenge and it is a job for computational physicists.

ACKNOWLEDGMENTS

Support by the Polish Committee for Scientific Research, grant 8T11F 00308, is gratefully acknowledged. This article has been prepared for the European Summer School on Computing Techniques in Physics, Skalský Dvůr, 5-14.09.1995. I want to thank the organizing committee for the invitation to lecture there on such a strange subject as “Psychophysics”.

-
- [1] E. Schrödinger, *Ann. Phys.* 63 (1920) 481
 - [2] D.J. Murray, *Behavioral and Brain Sci.* 16 (1993) 115-186
 - [3] P.S. Churchland, T.J. Sejnowski, *The computational brain.* (MIT, Bradford Book 1992)
 - [4] W. Pauli, *Der Einfluss archetypischer Vorstellungen und die Bildung naturwissenschaftlicher Theorien bei Kepler*, in: *Naturerklärung und Psyche* (Rascher, Zurich 1952), pp. 109-194
 - [5] M. S. Gazzaniga, ed. *The Cognitive Neurosciences.* (MIT, Bradford Book 1995)
 - [6] Frequently Asked Questions URL is <http://www.wipd.ira.uka.de/~prechelt/FAQ/neural-net-faq.html> or it may be obtained via ftp from [rtfm.mit.edu](ftp://rtfm.mit.edu), catalog <pub/usenet/news.answers> file [neural-net-faq](#)
 - [7] W. Gerstner and J. Leo van Hemmen, *Biol. Cybern.* 67 (1992) 195
 - [8] S.S. Rakover, *Precise of Metapsychology: Missing Links in Behavior, Mind, and Science.* PSYCOLOQUY 4 (1993) metapsychology.1.rakover.
 - [9] H.P. Stapp, *Mind, matter and quantum mechanics* (Springer Verlag, Heidelberg 1993)
 - [10] R. Penrose, *The Emperor’s new mind* (Oxford Univ. Press 1989); *In the Shadow of the Mind* (Oxford Univ. Press 1994)
 - [11] D.C. Dennett, *Consciousness explained* (Little Brown, Boston 1991)
 - [12] A. Newell, *Unified theories of cognition.* (Harvard Univ. Press, Cambridge, MA 1990)
 - [13] E. Thelen, L.B. Smith, *A Dynamic Systems Approach to the Development of Cognition and Action* (MIT Press 1994)
 - [14] W. Duch, G.H.F. Diercksen, *Neural networks as tools to solve problems in physics and chemistry* *Comp. Physics Communic.* 82 (1994) 91-103
 - [15] N. Rashevsky, *Mathematical Biophysics* (Dover, NY 1960)
 - [16] J.A. Anderson, E. Rosenfeld (Eds). *Neurocomputing: Foundations of Research* (MIT Press, MA 1988); J.A. Anderson, A. Pellionisz, E. Rosenfeld (Eds). *Neurocomputing 2: Directions for Research* (MIT Press, MA. 1990)
 - [17] S. Grossberg, *The Adaptive Brain* (North Holland 1987)
 - [18] D. Hebb, *The Organization of Behavior* (J. Wiley, NY 1949)

- [19] J. Murre, CALM, Categorization and Learning Modules (Erlbaum 1992)
- [20] W. Duch, G.H.F. Diercksen, *Feature Space Mapping as a universal adaptive system*. Comp. Phys. Commun. (1995); Duch W (1996) Systems Analysis-Modeling-Simulation (submitted), *From cognitive models to neurofuzzy systems - the mind space approach*.
- [21] T. Poggio and F. Girosi, Proc. of the IEEE 78 (1990) 1481; J. Platt, Neural Comput. 3 (1991) 213; V. Kadiramanathan, M. Niranjani, Neural Comput. 5 (1993) 954
- [22] L.D. Harmon and E.R. Lewis, Neural modeling. Physiological Reviews 46 (1968) 513-591
- [23] J.D. Cowan, A statistical mechanics of nervous activity. Lectures on Math. in Life Sciences 2 (1970) 1-57, ed. by M. Gerstenhaber (Am. Math. Soc, Providence RI)
- [24] W.J. Freeman, Mass Action in the Nervous system (Academic Press, NY 1975)
- [25] J.J. Hopfield, Proc. Nat. Acad. Sci. 79 (1982) 2554; *ibid* 81 (1984) 3088
- [26] D.J. Amit, Modeling brain function. The world of attractor neural networks. (Cambridge Univ. Press 1989); T.L.H. Watkin, A. Rau, M. Biehl, Rev. Modern Phys. 65 (1993) 499
- [27] P.A. Annios, Mathematical models of memory traces and forgetfulness. Kybernetik 10 (1972) 165-167; 11 (1972) 5-14
- [28] E.R. Caianiello, Outline of a theory of thought processes and thinking machines. Journal of Theor. Biology 2 (1961) 204-235; E.R. Caianiello, A theory of neural networks. In: Neural Computing Architectures, ed. I. Aleksander (MIT Press, MA 1989)
- [29] T.H. Brown, P.F. Chapman, E.W. Kairiss, C.L. Keenan, Long-term synaptic potentiation. Science **242** (1988) 724-728
- [30] D.S. Levine, Introduction to neural and cognitive modeling (L. Erlbaum, London 1991)
- [31] E. Erwin, K. Obermayer, K. Schulten, Models of Orientation and Ocular Dominance Columns in the Visual Cortex: A Critical Comparison. Neural Computation 7 (1995) 425-468
- [32] T. Kohonen, Self-Organizing Maps (Springer Series in Information Sciences, Vol. 30, 1995)
- [33] A.P. Georgopoulos, M. Taira, A. Lukashin, Cognitive neurophysiology of the motor cortex. Science 260 (1993) 47-52
- [34] F.A. Mussa-Ivaldi, From basis functions to basis fields. Vector field approximation from sparse data. *Biol. Cybernetics* 67 (1992) 479-489
- [35] E. Koechlin and Y. Burnod, Dual population coding in the neocortex: a model of interaction between representation and attention in the visual cortex. (preprint, Inst. des Neurosciences, Université Pierre et Marie Curie, Paris 1995)
- [36] C. Skarda, W.J. Freeman, How brains make chaos to make sense of the world. The Behavioral and Brain Sci. 10 (1987) 161-195
- [37] Pulvermueller F, Preissl H, Eulitz C, Pantev C, Lutzenberger W, Elbert T and Birbaumer N. (1994) PSY-COLOQUY 5(48) brain-rhythms.1.pulvermueller
- [38] M. Griniasty, M. Tsodyks, D. Amit, Conversion of temporal correlations between stimuli to spatial correlations between attractors. Neural Comput. 5 (1993) 1-17
- [39] A.K. Engel, P. König, A.K. Kreiter, T.B. Schillen, W. Singer. Temporal coding in the neocortex: new vistas on integration in the nervous system. Trends in Neurosc. 15 (1992) 218-226
- [40] E.R. John, M. Shimokochi, F.P. Bartlett, Neural readout from memory during generalization. Science **164** (1969) 1534-1536
- [41] J.A. Scott Kelso, Dynamic Patterns (Bredford Book, MIT Press 1995)
- [42] C.M. Bishop, Training with noise is equivalent to Tikhonov regularization. Neural Computation 7 (1995) 108-116
- [43] R. Clark, Dynamical systems. Stability, symbolic dynamics and chaos (CRC Press, London 1995)
- [44] R.E. Mayer, Thinking, problem solving, cognition. (WH Freeman 1992)
- [45] W. Duch (1994) UMK-KMK-TR 1/94 report, *A solution to the fundamental problems of cognitive sciences*, available from ftp.phys.uni.torun.pl/pub/papers/kmk
- [46] P. Deregai, Chua's oscillator: a zoo of attractors, J. of Circuits, Systems and Computers 3 (1993) 309-359; L.O. Chua, C.A. Desoer, E.S. Kuh, Linear and nonlinear circuits (McGraw-Hill, New York 1987); M.P. Kennedy, Three Steps to Chaos, IEEE Trans. Circuits and Systems 40 (1993) 640-674
- [47] H. Haken, Synergetic computers and cognition (Springer, 1991)
- [48] H. Lilienström, Modeling the dynamics of olfactory cortex using simplified network units and realistic architecture", Int. J. Neural Systems 2 (1991) 1-15; X. Wu, H. Lilienström, Regulating the nonlinear dynamics of olfactory cortex, Network 5 (1994) 47-60
- [49] F.A. Geldard, The human senses (Wiley, New York 1962)
- [50] B. MacLennan, Field computation in the brain, CS-92-174 (Univ. of Tennessee, Knoxville, TN 37996)
- [51] M.P. Casey, Computation in Discrete-Time Dynamical Systems (UCSD 1995, available in neuroprose).
- [52] G. Fauconniere, Mental Spaces (Cambridge Univ. Press 1994)

- [53] I. Aleksander, H. Morton, *Neurons and symbols* (Chapman and Hall 1993)
- [54] D. Waltz, Memory-based reasoning, in: M.A. Arbib, *The Handbook of Brain Theory and Neural Networks* (MIT Press 1995), pp. 568-570
- [55] J.N.H. Heemskerk, J.M.J. Murre, Brain-size neurocomputers. (1995, draft); J.N.H. Heemskerk, Overview of neural hardware (1995, draft), available from <ftp.mrc-apu.cam.ac.uk/pub/nn>;
- [56] B.J. Baars, *A cognitive theory of consciousness* (Cambridge Univ. Press 1988)
- [57] F. Verhulst, Metaphors for psychoanalysis. *Nonlinear Science Today* 4 (1994) 1-6