Neural Mechanisms of the Mind, Aristotle, Zadeh, and fMRI

Leonid I. Perlovsky

Abstract – **Processes in the mind: perception, cognition, concepts, instincts, emotions, and higher cognitive abilities are considered here within a neural modeling fields paradigm. Its fundamental mathematical mechanism is a process "from vague-fuzzy to crisp," called dynamic logic. The paper discusses why this paradigm is necessary mathematically, and relates it to a psychological description of the mind. Surprisingly, the process from "vague to crisp" corresponds to Aristotelian understanding of mental functioning. Recent fMRI measurements confirmed this process in neural mechanisms of perception.**

I. COMPUTATIONAL INTELLIGENCE, LOGIC, AND THE MIND

Computational intelligence as well as engineering intelligent systems, it seems, imply an understanding of the nature of intelligence. However a theory of intelligence does not exist, and an understanding of what is intelligent is subjective. In absence of such a theory, many consider the human mind as the highest example of an intelligent system. Thus understanding and modeling the mind is an essential part of computational intelligence, as reflected in the pages of this journal. Does a goal of understanding and modeling the mind imply that engineers should know cognitive science, biology, psychology, neuro-psychology, both experimental and theoretical, as well as the philosophy of mind? Practically it is not possible for each engineer to know all these disciplines, and the idea itself seems foreign to many practicing engineers in the field of computational intelligence. Nevertheless, integrating diverse knowledge about the mind developed in various fields is held by many as an ideal goal. This paper is a step toward this goal. We analyze recent successes in computational intelligence, relate them to fundamental mechanisms of the mind, and analyze how computational intelligence can contribute to a number of emerging engineering fields modeling higher human mental abilities and human cultures.

For thousands of years the idea that the mind works logically was considered to be originated by Aristotle. Logical-rule artificial intelligence developed logic into a powerful computational technology. However, it could not model many aspects of the mind. Fuzzy logic was developed by Lotfi Zadeh and neural network paradigms were proposed by a number of scientists as an alternative and even an opposition to logic. That story is well known to most researchers in computational intelligence. This paper discusses unorthodox views, first, that Aristotle did not think that the mind works logically, and second, that most neural networks and fuzzy logic systems have to perform logical steps at some points in their operations, and this principally limits their operational capabilities.

We describe neural modeling fields (NMF) operated by dynamic logic DL. It differs from other types of logic in that it does not describe states, but processes; it is a process-logic, a process "from vague-to-crisp." An initial state of DL is fuzzy-vague and the DL-process evolves it into a logical or near-logical state. Detailed descriptions of NMF-DL are available in referenced literature, while sections II and III briefly summarize this technique, emphasizing the fundamental ideas. Section IV illustrates the DL-process "from vague-to-crisp" and demonstrates applications of NMF-DL to the detection of patterns in difficult conditions, 100 to 1000 times more difficult than was previously considered possible. We discuss why such a breakthrough became possible mathematically. Then we suggest that NMF-DL detects patterns using similar mechanisms that the mind uses for perception. Section V discusses emerging engineering applications, which require understanding and mathematical modeling of higher mental abilities of the human mind, and briefly discusses how NMF-DL has been applied to these applications and how it models the mental abilities. Here, we argue that relatively simple mathematics can go a long way toward modeling the mind, if we understand the basic principles of the mind operations. The difficulty here is that scientific intuitions are often biased by logical thinking. By analyzing mechanisms of the mind, we discuss why this difficulty persists, even for scientists intent on overcoming logical limitations. A good illustration would be the historically slow acceptance of algorithms and theories not relying on logic, for example some the ideas of Aristotle, Zadeh, Grossberg, and others. Section VI concentrates on psychological interpretations of NMF-DL. It demonstrates that properties of the mind, which seemed mysterious, can be understood with relatively simple mathematical ideas and used in engineering systems within NMF-DL framework. Section VII discusses relationships between theories of Aristotle, Godel, and Zadeh and takes them beyond what was considered "received truth," we argue that DL is close to Aristotelian theory of the mind. Section VIII relates these discussions to logic-based AI, it moves beyond what has been discussed for decades, and relates to urgent problems of contemporary neural network engineering. Section IX discusses recent advancements in experimental neuroimaging and results that have proven fundamental, uniquely NMF-DL mechanisms (not advocated by any other neural or other mathematical theory) as actually being used in the mind-brain. Section X discusses future research directions: theoretical ideas, mathematical and engineering development, emerging engineering areas. We advocate that

˙ Harvard University, Cambridge, MA; and the US Air Force Research Laboratory, Sensors Directorate, Hanscom, MA; leonid@seas.harvard.edu

mathematics, engineering, and psychology should work jointly toward future intelligent systems, and we discuss directions for neuro-imaging and psychological experiments.

II. NEURAL MODELING FIELDS

Neural Modeling Fields (NMF) is a multi-layer, hetero-hierarchical system [1]. The mind is not a strict hierarchy; there are multiple feedback connections among several adjacent layers, hence the term hetero-hierarchy. NMF mathematically implements mechanisms of the mind including perception, cognition, concepts, instincts, emotions, and higher cognitive abilities as described below.

This section describes a basic mechanism of interaction between two adjacent hierarchical layers of bottom-up and top-down signals (fields of neural activation), Fig. 1.
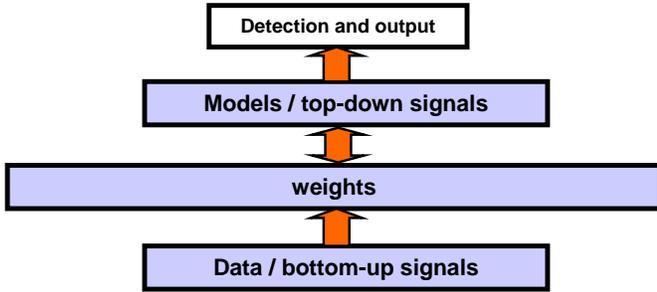

Fig. 1. Graphical representation of a single-layer of the NMF architecture. Bottom-up signals are unstructured data $\{\mathbf{X}(n)\}$ and output signals are recognized or activated concept-models $\{m\}$. Top-down, "priming" signals are produced by models, $\mathbf{M}_m(\mathbf{S}_m,n)$.

At the bottom of the mind hetero-hierarchy, input signals can be thought of as sensor signals, and outputs as activations of higher-levels models. These activation signals become inputs to the next processing layer. Sometimes, it will be more convenient to talk about these two signal-layers as an input to and output from a (single) processing-layer. At each layer, output signals are concepts recognized in (or formed from) input bottom-up signals. Input signals are associated with (or recognized, or grouped into) concepts according to the models (top-down signals) at this layer. This general structure of NMF corresponds to our general knowledge of neural structures in the brain; however, it is not mapped to specific neurons or synaptic connections. In the process of learning and understanding input bottom-up signals, models are adapted so that top-down signals generated by models better correspond to bottom-up signals.

At a particular hierarchical layer, we enumerate neurons by indices n = 1,... N. These neurons receive bottom-up input signals, $\mathbf{X}(n)$, from lower layers in the processing hierarchy. $\mathbf{X}(n)$ is a field of bottom-up neuronal synapse activations, coming from neurons at a lower layer. Each neuron has a number of synapses; for generality, we describe each neuron activation as a set of numbers, $\mathbf{X}(n) = \{X_d(n), d = 1,... D\}$. Top-down, or priming signals to these neurons are sent by concept-models, $\mathbf{M}_m(\mathbf{S}_m,n)$; we enumerate models by indices m = 1,... M. Each model is characterized by its parameters, $\mathbf{S}_m$; in the neuron structure of the brain they are encoded by strength of synaptic connections, mathematically, we describe them as a set of numbers, $\mathbf{S}_m = \{S_{a\,m}, a = 1... A\}$. Models represent signals in the following way. Say, signal $\mathbf{X}(n)$, is coming from sensory neurons activated by object m, characterized by parameters $\mathbf{S}_m$. These parameters may include position, orientation, or lighting of an object h. Model $\mathbf{M}_m(\mathbf{S}_m,n)$ predicts a value $\mathbf{X}(n)$ of a signal at neuron n. For example, during visual perception, a neuron n in the visual cortex receives a bottom-up signal $\mathbf{X}(n)$ from retina and a top-down priming signal $\mathbf{M}_m(\mathbf{S}_m,n)$ from an object-concept-model m. A neuron n is activated if both bottom-up signal and top-down priming signal are strong. Various models compete for evidence in the bottom-up signals, while adapting their parameters for better match as described below. The more top-down signals from the model m are matched to bottom-up signals, the higher is activation of the model m. This is a simplified description of perception. The most benign everyday visual perception uses many layers from retina to object perception. The NMF premise is that the same laws describe the basic interaction dynamics at each layer. Perception of minute features, or everyday objects, or cognition of complex abstract concepts is due to the same mechanism described below. Perception and cognition involve models and learning. In perception, models correspond to objects; in cognition models correspond to relationships and situations.

Learning is an essential part of perception and cognition. NMF learns motivated by internal "desire" to improve correspondence between top-down and bottom-up signals (a kind of reinforcement learning [2], or learning without a teacher). Psychologically, NMF learning is driven by the knowledge instinct (KI), which mathematically is described by increasing a similarity measure between the sets of models and signals, $L(\{X\},\{M\})$. Under certain conditions, L can be considered as an estimated likelihood function. The similarity measure is a function of model parameters and associations between the input bottom-up sensor signals and top-down, concept-model signals. For concreteness I refer here to an object perception using a simplified terminology, as if perception of objects in retinal signals occurs in a single layer. As mentioned, NMF is a hetero-hierarchical system comprised of many layers, which interaction is not strictly hierarchical; at higher levels input signals are activated (recognized) models at a lower level (or levels), but for simplicity I will mostly talk about recognition of objects in sensor signals).

In constructing a mathematical description of the similarity measure, it is important to acknowledge two principles. First, the exact content of the visual field is unknown before perception occurs. Important information

could be contained in any bottom-up signal; therefore, the similarity measure is constructed so that it accounts for all input information, $\mathbf{X}(n)$,

$$L(\{\mathbf{X}\},\{\mathbf{M}\}) = \prod_{n \in N} \ell\,(\mathbf{X}(n)). \qquad (1)$$

This expression contains a product of partial similarities, $\ell(\mathbf{X}(n))$, over all bottom-up signals; therefore it forces the mind to account for every signal (even if one term in the product is zero, the product is zero, the similarity is low and the knowledge instinct is not satisfied); this is a reflection of the first principle. Second, before perception occurs, the mind does not know which retinal neuron corresponds to which object. Therefore a partial similarity measure is constructed so that it treats each model as an alternative (a sum over all models) for each input neuron signal:

$$\ell\,(\mathbf{X}(n)) = \sum_{m \in M} r(m)\,\ell\,(\mathbf{X}(n)\mid m). \qquad (2)$$

Here $\ell\,(\mathbf{X}(n)|m)$, or $\ell\,(n|m)$ for shortness, is a conditional similarity between signal $\mathbf{X}(n)$ and model $\mathbf{M}_m$. Combining eqs. (1) and (2), a similarity measure is constructed as follows [1]:

$$L(\{\mathbf{X}\},\{\mathbf{M}\}) = \prod_{n \in N} \sum_{m \in M} r(m)\,\ell\,(\mathbf{X}(n)\mid m). \qquad (3)$$

The conditional similarity $\ell(n|m)$ is function of a deviation of the model from data, or a modeling error $(\mathbf{X}(n) - \mathbf{M}(m))^2$; or more generally, $(\mathbf{X}(n) - \mathbf{M}(m))\,\mathbf{C}^{-1}\,(\mathbf{X}(n) - \mathbf{M}(m))^T$, where $\mathbf{C}^{-1}$ is an inverse covariance function of the errors, and superscript T denotes the transposed vector. The conditional similarity is convenient to define "conditional" on object m being present, therefore, when combining conditional similarities, they are multiplied by $r(m)$, which represents the measure of object m actually being present. In this case of conditional similarities depending on errors, it is natural to select ℓ as a Gaussian function [3], as in eq. (7).

The structure of (3) follows standard principles of the probability theory: a summation is taken over alternatives, m, and various pieces of evidence, n, are multiplied. This expression is not necessarily a probability, but it has a probabilistic structure. If conditional similarities are chosen to approximate conditional probability density functions, pdfs, the total similarity, eq.(3) approximates total likelihood and leads to near-optimal Bayesian decisions. Other choices of functions ℓ are sometimes appropriate, as discussed in [1]. If $\ell\,(n|m)$ approximates a conditional pdf, it is a conditional probabilistic measure that signal in neuron n originated from object m. Then L approximates total likelihood of observing signals $\{\mathbf{X}(n)\}$ coming from objects described by models $\{\mathbf{M}_m\}$. Coefficients $r(m)$, called priors in probability theory,

contain "prior" biases or expectations, expected objects m have relatively high $r(m)$ values; they are not necessarily known a priori, their true values are usually unknown and should be learned, like other parameters $\mathbf{S}_m$.

$$\qquad (1)$$

A. Notes on Structure of Similarity

In the probability theory, a product of probabilities usually assumes that evidence is independent. Expressions (2) and (3) contain products over n, but it does not assume independence among various signals $\mathbf{X}(n)$. There is a dependence among signals due to models: each model $\mathbf{M}_m(\mathbf{S}_m,n)$ predicts expected signal values in many neurons n. Statistically independent in this expression are deviations between models and signals, which usually are due to random errors and statistically independent.

Some signals may not fit into existing models. Even so a system can change the number and types of used models, still, some signals might come from sources, which are not of interest for detection and detailed recognition. Therefore, it is always useful to have at least one model describing these "extraneous" or clutter signals; in many cases, it is sufficient to have just one clutter model that is constant over all values of signals, $\mathbf{X}(n)$. With proper normalization its similarity can be written as $r(\text{clutter})/\text{volume}(\mathbf{X})$, where $\text{volume}(\mathbf{X})$ stands for the volume of the space $\mathbf{X}$. Then $r(\text{clutter})$ is the only constant-parameter of the clutter model, which should be learned from data (or estimated, in statistical terminology).

During the learning process, concept-models are constantly modified. From time to time a system forms a new concept, while retaining an old one as well; alternatively, old concepts are sometimes merged or eliminated. This mechanism works as follows. The initial system state contains many non-activated diverse models; parameters of non-activated models are not updated as described later in section III, except for their "strength" parameters $r(m)$. In interaction with bottom-up signals models are activated when a model strength $r(m)$ exceeds a predetermined threshold. The more diverse signals are coming, the more models are activated. Formation of new concepts and merging or elimination-forgetting of old ones require a modification of the similarity measure (3); the reason is that more models always result in a better fit between the models and data. This is a well known problem, it can be addressed by reducing similarity (3) using a "skeptical penalty function," $p(N,M)$ that grows with the number of models M, and this growth is steeper for a smaller amount of data N. For example, an asymptotically unbiased maximum likelihood estimation leads to multiplicative $p(N,M) = \exp(-N_{par}/2)$, where N_{par} is a total number of adaptive parameters in all models (this penalty function is known as Akaike Information Criterion, AIC; see [1] for further discussion and references).

AIC is theoretically expected to work well asymptotically, for large N. In many practical problems N is

not large and AIC is known to perform poorly. Often an empirical penalty function related to ridge regression works well, $p(N,M) = \exp(-a \sum_{m \in M} \mathbf{S}_m^2)$, where coefficient *a* is selected empirically. Alternatively, a penalty function can be defined following Statistical Learning Theory [4].

III. DYNAMIC LOGIC

The learning process in NMF consists in estimating model parameters **S** and associating signals, n, with concept-models, m, by maximizing similarity (3). Note, all possible combinations of signals and models are accounted for in expression (3). This can be seen by expanding a sum in (3), and multiplying all the terms; it would result in M^N items, a huge number. This is the number of combinations between all signals (N) and all models (M). Here is the source of combinatorial complexity (CC) of many algorithms used in the past and still in use now. For example, multiple hypothesis testing algorithms [5] attempt to maximize similarity L over model parameters and associations between signals and models, in two steps. First it takes one of the M^N items, which corresponds to one particular association between signals and models; and maximizes it over model parameters; this is performed over all items. Second, the largest item is selected (that is the best association for the best set of parameters). Such a program inevitably faces a wall of CC, the number of computations on the order of M^N.

CC was experienced in many algorithms and neural networks since the 1950s [6]. In learning pattern recognition algorithms and neural networks CC is experienced as complexity of *training* requirements; learning algorithms and neural networks should be trained to recognize every object one by one. Not only should every object be "shown" to the algorithm in multiplicity of its forms, angles, lightings, but also object combinations should be presented for training. The number of combinations is combinatorially large. Rule systems were initially proposed to overcome complexity of learning [7,8]. An initial idea was that rules would capture the required knowledge and eliminate a need for learning. However in presence of variability, the number of rules grew; rules became contingent on other rules; combinations of rules had to be considered; rule systems encountered *CC of rules*. Beginning in the 1980s, model-based systems were proposed. They used models which depended on adaptive parameters. The idea was to combine advantages of rules with learning-adaptivity by using adaptive models. The knowledge was encapsulated in models, whereas unknown aspects of particular situations were to be learned by fitting model parameters [5]. Fitting models to data required selecting data subsets corresponding to various models. The number of subsets, however, is combinatorially large. Model-based approaches encountered *computational CC* (N and NP complete algorithms). It turned out that CC is related

to Gödel theory: it is a manifestation of the inconsistency of logic in finite systems [9].

NMF overcomes this fundamental difficulty of many learning algorithms and solves the problem of concurrent model estimation and model-object association without CC by using dynamic logic [1,10]. An important aspect of dynamic logic is matching vagueness or fuzziness of similarity measures to the uncertainty of models. Initially, parameter values are not known, and uncertainty of models is high; so is the fuzziness of the similarity measures. In the process of learning, models become more accurate and the similarity measure more crisp, the value of the similarity increases. This process "from vague to crisp" is the essence of dynamic logic.

Mathematically it is described as follows. First, assign any values to unknown parameters, $\{\mathbf{S}_m\}$. Then, compute association variables f(m|n),

$$f(m|n) = r(m) \, \ell(\mathbf{X}(n)|m) \, / \sum_{m' \in M} r(m') \, \ell(\mathbf{X}(n)|m'). \qquad (4)$$

Eq.(4) looks like the Bayes formula for a posteriori probabilities. However, since the initial parameter values are not known, similarities do not correspond to any particular object or event and (4) describes an initial vague state of the dynamic logic process. If $\ell(n|m)$ in the result of learning approximate conditional likelihoods, f(m|n) approximate posteriori Bayesian probabilities for signal n originating from object m. The dynamic logic (DL) of the NMF is defined as follows,

$$d\mathbf{S}_m/dt = \sum_{n \in N} f(m|n)[\partial \ln \ell(n|m)/\partial \mathbf{M}_m]\partial \mathbf{M}_m/\partial \mathbf{S}_m ,$$

$$df(m|n)/dt = f(m|n) \sum_{m' \in M} \{[\delta_{mm'} - f(m'|n)] \bullet$$

$$[\partial \ln \ell(n|m')/\partial \mathbf{M}_{m'}])/\partial \mathbf{M}_{m'}/\partial \mathbf{S}_{m'} \bullet d\mathbf{S}_{m'}/dt,$$

$$\delta_{mm'} = 1 \text{ if } m=m', \ 0 \text{ otherwise.} \qquad (5)$$

Here, ln stands for natural logarithm, ∂ denotes partial derivatives, d(.)/dt denotes derivatives with respect to time t; this is the time of the internal dynamics of the NMF system (like a number of internal iterations). Eqs.(5) defines a system of linear differential equations of the first order with respect to time t; they describe the system internal dynamics, a process of DL. They can be solved by a standard differential equation solver.

The first of eqs.(5) is similar to standard gradient assent equations maximizing similarity eq.(3). The multiplier f(m|n) in this equation assigns to parameter $\mathbf{S}_m$ (model m) its "share" of the gradient. The second equation modifies association variables f(m|n) through time t, according to parameter changes defined by the first equation.

The following theorem was proven [1].

Theorem. Equations (5) define a convergent dynamic NMF system with stationary states defined by $\max_{\{Sm\}} L$. The basis for this proof is in that similarity eq.(1) monotonically increases along the DL process, eq.(5). In this way, similarity eq.(1) serves as a Lyapunov function for the system (5) assuring the convergence of the DL process [1].

Understanding psychological interpretation of this theorem is necessary for building intelligent systems approaching human intelligence. Psychologically, DL satisfies the KI. As discussed in section VI, satisfaction or dissatisfaction of an instinct gives rise to emotional neural signals. Specific emotions related to knowledge are called aesthetic emotions. In section VI we discuss that aesthetic emotions serve as a foundation of higher mental abilities. For now we can state that the KI is satisfied during the DL process, NMF-DL enjoys learning. This statement with respect to eqs. (5) may seem a stretch, but when considering thousands of agents each governed by hierarchical structures of eqs.(5), communicating among each other, then possibly the word "enjoy" would not seem too much out of place. Engineering intelligent systems requires understanding these connections of mathematics and psychology; section VI is devoted to this topic, but for now we return to the mathematical properties of the DL process convergence.

It follows that the stationary states of an NMF system are the maximum similarity states satisfying the knowledge instinct. When partial similarities are specified as probability density functions (pdf), or likelihoods, the stationary values of parameters $\{S_m\}$ are asymptotically unbiased and efficient estimates of these parameters [11]. A computational complexity of the NMF method is linear in N.

In plain English, this means that dynamic logic is a convergent process. It converges to the maximum of similarity, and therefore satisfies the knowledge instinct. The similarity measure (3) is highly nonlinear in terms of model parameters, therefore convergence may occur to a local maximum. This local convergence usually occurs after few iterations (as illustrated in the following chapters). The local rather than global convergence sometimes presents an irresolvable difficulty in many applications. In DL this problem is resolved in three ways. First, the large initial standard deviations of conditional similarities smoothes local maxima. Second, according to the above description (section II, A), a large number of dormant models is used initially; many of them could be terminated and activated many times. Therefore if a particular pattern is not "captured" after few iterations, it will be captured at a later iteration, after a model re-activation. Third, some of the activated models converge to spurious events not corresponding to real patterns of interest; say they will come nearby only few data points. In these cases, their strength r(m), and their local log-likelihood ratio [12] will be low and spurious events are discarded. A detailed characterization of performance usually requires operating curves [13], plots of probability of detection vs. probability of false alarm, computed for various signal-to-clutter ratios, densities of objects, and other parameters. Such detailed characterization is beyond the scope of this paper. We would just add that the NMF-DL performance was proven to come close to the information-theoretic performance limits determined by Cramer-Rao Bounds in several cases, when we performed such an investigation [1]. Finally, in practical applications, detecting an object is only a part of the overall procedure. A detection procedure discussed in the following sections is performed periodically over time, as new data are acquired [12]. In this process spurious events are discarded, and patterns not detected initially are detected at a later time.

Note, in practical solutions of eqs.(5) by iterative steps, the first equation can be substituted by re-computation of f(m|n) at every iteration step according to its definition eq.(4). Also we should clarify that when discussing models, sometimes we refer to $\mathbf{M}_m$, and sometimes to conditional similarities $\ell(n|m)$; vagueness and fuzziness always refer to the state of similarities.

Relationship of DL to several other types of logic has been considered in [14].

A number of engineering applications of NMF-DL were developed, [1,10,12,15,16,17,18,19,20, 21] and many others. In many cases complexity of the problem was beyond capabilities of existing algorithms, e.g. an applications to multiple target tracking in strong clutter was presented in [12]. NMF-DL applications have often resulted in significant savings in complexity and in two to three orders of magnitude improvement in terms of signal-to-clutter ratio; as mentioned often it results in the best possible solution, which cannot be improved by any algorithm or neural network; several novel types applications were developed [18,20,20,22,23,24,25,26,27,28,29], which could not have been previously considered. Some of these applications are briefly summarized in section V, after section IV considers applications to object perception for difficult cases where solutions have not been available previously.

IV. EXAMPLES OF NMF OBJECT PERCEPTION

Here we illustrate operations of NMF-DL in the task of object detection (in engineering terms) or perception (in psychological and neural-network terms). We illustrate the DL process "from vague-to-crisp," estimation of parameters S_m, association variables f(m|n), and the number of objects. We consider difficult cases, when object signals are below noise or clutter, so that solutions were not previously available. The second example particularly well illustrates graphically the DL process "from vague to crisp." As discussed in following sections, this unique property of DL was proved in recent neuro-imaging experiments to be a part of neural mechanism of perception in the mind, it is fundamental for understanding higher cognitive functions, and for developing emerging application areas.

Finding patterns below clutter is an exceedingly complex problem. If an exact pattern shape is not known and depends on unknown parameters, these parameters should be found by fitting the pattern model to the data. However, when the locations and orientations of patterns are not known, it is not clear which subset of the data points should be selected for fitting. A standard approach for solving this kind of problem, which has already been discussed, is multiple hypothesis testing [5]. It faces combinatorial complexity, as discussed, since combinations of subsets and models are exhaustively searched; as a result, detection performance is limited not by information present in data, but by computational complexity and consequently by signal-to-clutter ratio.

In the first examples (preliminary results have been presented in [15]) an elongated object to be detected moves along an unknown path and rotates with unknown speed; exact shape of the object is unknown, and a signal strength is lower than clutter by about 5 times. A sequence of 25 images, 256 x 256 pixel each, have been available for processing. Problems of such complexity have not been previously considered.

To apply NMF-DL to this problem one needs to develop parametric adaptive models of expected patterns. As mentioned the models and conditional similarities for this case were described in details in [15]: a uniform model for clutter in 2 dimensions of $\mathbf{X}(n) = (X, Y)$, is given by

$$\ell(\mathbf{X}(n)|m = clutter) = 1/ (\Delta X * \Delta Y),$$
$$\Delta X = (Xmax\text{-}Xmin), \Delta Y = (Ymax\text{-}Ymin); \qquad (6)$$

Minimal and maximal values of coordinates were taken equal to limits of the available imagery data (1 and 256).

Gaussian blobs for highly-fuzzy, poorly resolved patterns, are given by

$$\ell(\mathbf{X}(n)|m=blobs) = (1/2\pi\sigma_m) \exp[-(\mathbf{X}(n)\text{-}\mathbf{M}_m)^2/(2\sigma_m^2)]; \quad (7)$$

Here model $\mathbf{M}_m$ is a mean of the Gaussian density,

$$\mathbf{M}_m(\mathbf{S}_m,n) = (X_m,Y_m), \qquad (8)$$

and σ_m is the standard deviation (σ_m^2 is variance). Unknown parameters of this model included in $\mathbf{S}_m$, are $(X_m, Y_m, \sigma_m, r_m)$, they are estimated according to the same eqs.(5). A model for a moving and rotating object is given by

$$\ell(\mathbf{X}(n)|m = moving, rotating) = (1/2\pi)\det\mathbf{C}^{-1/2}$$
$$* \exp[-(\mathbf{X}(n) - \mathbf{M}_m) \, \mathbf{C}_m^{-1} \, (\mathbf{X}(n) - \mathbf{M}_m)^T]; \qquad (9)$$

$$\mathbf{M}_m = (X_m + T*V_{mx}, Y_m + T*V_{my}); \qquad (10)$$

$$\mathbf{C}_m = diag(C1_m + C2_m \cos(T*\omega_m), C1_m + C2_m \sin(T*\omega_m)). \quad (11)$$

Here, $\mathbf{M}_m$ is a center of object m moving with velocity $\mathbf{V}_m = (V_{mx}, V_{my})$; $\mathbf{C}_m$ is a diagonal covariance determining a rotating elongated shape, set to $C2_m = 100*C1_m$; T is a time of actual object motion and rotation; ω_m is a frequency of the object rotation. Parameters of these models included in $\mathbf{S}_m$ are $(X_m, Y_m, V_{mx}, V_{my}, C1_m, \omega_m, r_m)$.

Fig. 2a shows one frame with moving and rotating object measured at close range, so that signal-to-clutter ratio (S/C) is about 300, which have been considered necessary for a reliable object detection; on the right is a signal strength-to-color mapping bar in arbitrary units. Fig. 2b shows similar image at realistic range of interest, S/C = 0.2.

Fig. 2c shows a NMF-DL iteration 10; there are 26 activated blob-models with relatively low strength (activation threshold was set at $r_m = 0.0001$ of the total signal strength) and 1 vague activated moving and rotated elongated object model (uniform clutter model is not shown). Intermediate frames with motion and rotation of the object are not shown.
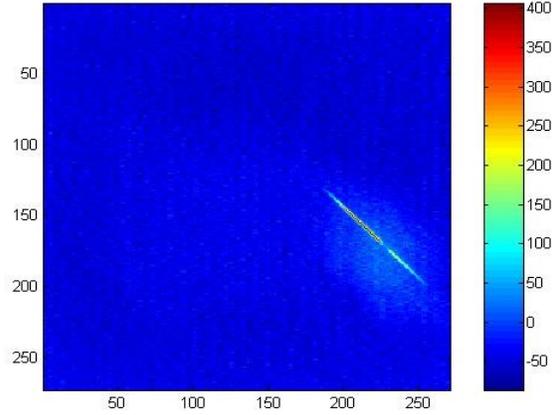

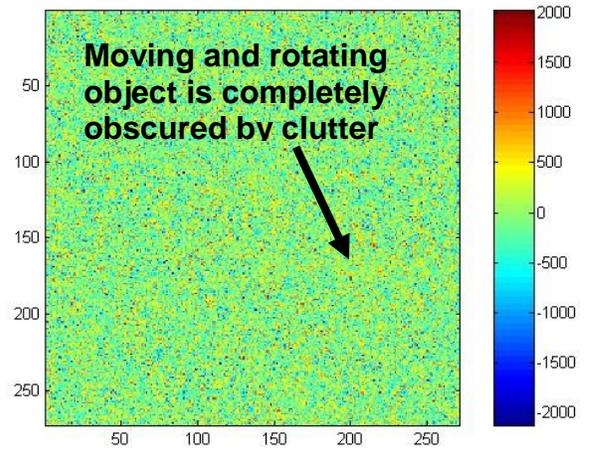Fig. 2a. One frame S/C = 300.
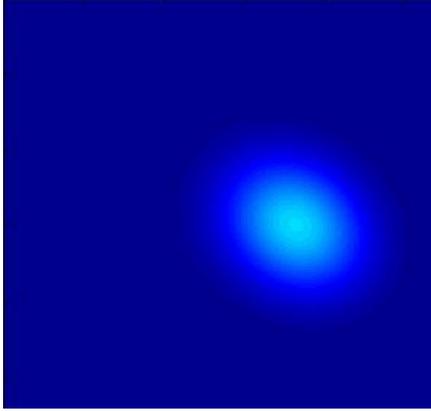

Fig. 2b. One frame S/C = 0.2.

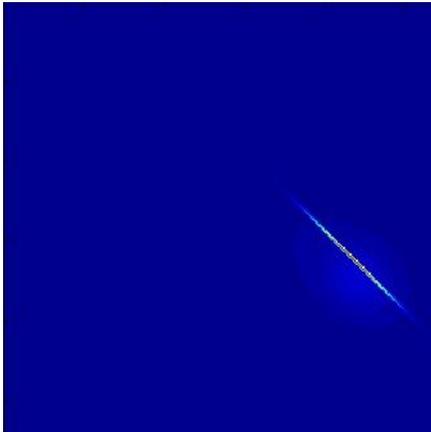Fig. 2c. DL processing of Fig. 2b data, iteration 10.

Fig.2d. DL convergence results, iteration 600.

Fig. 2d shows a NMF-DL convergence results at iteration 600. (Possibly the number of iterations could be significantly reduced, no efforts were devoted to this). The model of the elongated object as well as surrounding clutter blobs are estimated closely to the image acquired at high image-to-clutter ratio, Fig. 2a, which previously was not considered possible. The S/C improvement is about 1,500 (to make this point certain, 150,000% improvement).

In the second example, NMF-DL is looking for 'smile' and 'frown' patterns in noise shown in Fig.3a without clutter, and in Fig.3b with clutter, as actually measured. This example also is beyond capabilities of previously existing techniques (preliminary results have been presented in [17]). Each pattern is characterized by a parabolic shape. The image size in this example is 100x100 points, and the true number of patterns is 3, which is not known. In this example it is relatively easy to estimate the algorithmic complexity of various algorithms. Using a multiple hypothesis testing brute-force approach will take about $M^N = 10^{6,000}$ operations Alternatively, the complexity of computation can be estimated by trying various parameter values. Most algorithms before deciding that 3 patterns fit data best, would have to try more than 3 patterns, say, at least 4; as discussed below each model is characterized by 5 parameter; fitting 5x4=20 parameters to 100x100 grid by testing of parameter values would take about 10^{43} to 10^{45} operations, a prohibitive computational complexity in both cases. DL computational complexity equals N*M*it*ops, where *it* is the number of steps (iterations) in solving eqs.(5), and *ops* is a number of operations per step. The DL complexity in the example Fig.3 turned out equal 10^9, so that a problem previously unsolvable due to CC have been solved using NMF-DL, improvement over the previous state of the art is 7,000% in S/C.
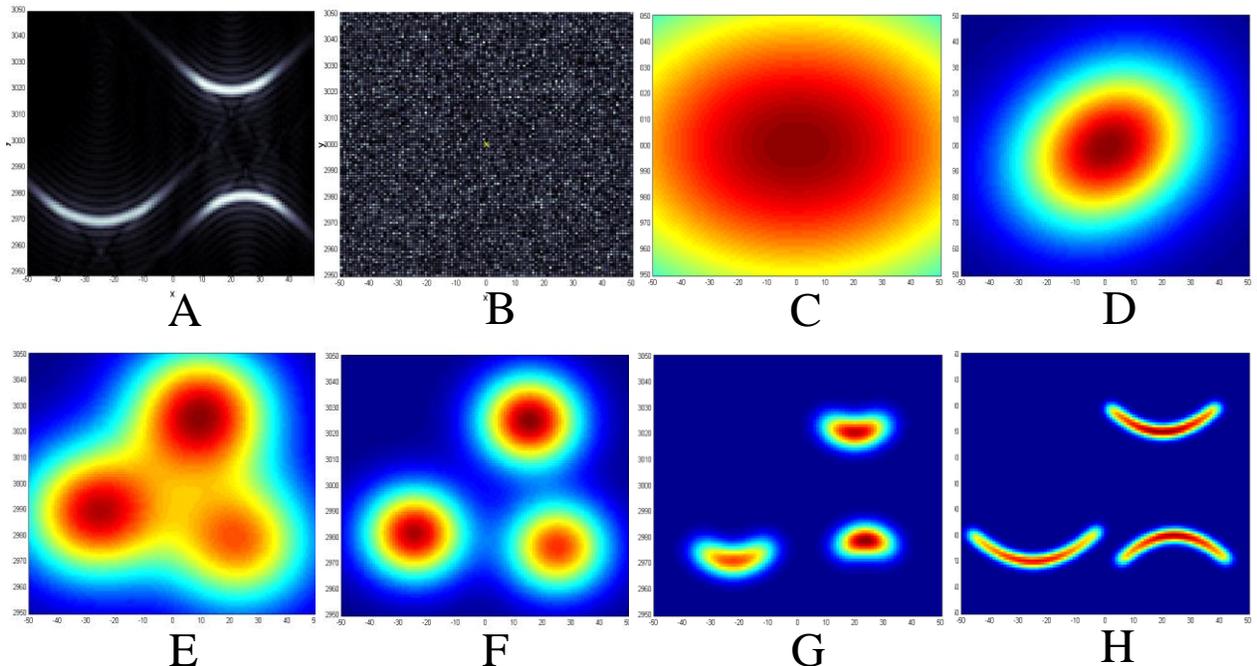
Fig.3. An example of NMF-DL perception of 'smile' and 'frown' objects in clutter in 2-dimensions of X and Y: (a) true 'smile' and 'frown' patterns are shown without clutter; (b) actual image available for recognition (signal is below clutter, S/C ~ 0.5); (c) an initial fuzzy blob-model, the fuzziness corresponds to uncertainty of knowledge; (d) through (h) show improved models at various iteration stages (total of 22 iterations). The improvement over the previous state of the art is 7,000% in S/C; this example is discussed in more details in [17].

To apply NMF-DL to this problem one needs to develop parametric adaptive models of expected patterns. The models and conditional partial similarities for this case were described in details in [17]; a uniform model for clutter and Gaussian blob models are similar to eq.(6), (7). Parabolic models for 'smiles' and 'frowns' use the same functional form for conditional similarities as (7) with parabolic-shaped models,

$$\mathbf{M}_m = (X_m + x, Y_m + ax^2); \tag{12}$$

here a is a parameter determining the curvature of parabolic shapes, X_m, Y_m give the apex of the parabola, and x is a horizontal coordinate deviation from the apex (a coordinate, not a parameter).

The number of computer operations in this example was about 10^9. Thus, a problem that was not solvable due to CC becomes solvable using dynamic logic.

In both examples, at the beginning of adaptation process, parameters are inaccurate, the models do not match data patterns, and variances (covariances) are large; correspondingly, conditional similarities are vague-fuzzy. During an adaptation process, initial uncertain models and fuzzy similarities are associated with structures in the input signals, model parameters become more accurate, models better match patterns in the data, and variances are reduced with successive iterations (similarities become less vague and crisper). The type, shape, and number of models are selected so that the internal representation within the system is similar to input signals: the NMF-DL concept-models represent structure-objects in the signals. In the second example, Fig. 3(b) shows actual image available for recognition, signal is below clutter, approximately S/C = 0.5; 3(c) is an initial vague-fuzzy model, a large fuzziness corresponds to uncertainty of knowledge; (d) through (h) show improved models at various iteration stages (total of 22 iterations). Between iterations (d) and (e) the NMF-DL system activated three Gaussian models. There are several types of models: one uniform model describing clutter (it is not shown) and a variable number of blob models and parabolic models, which number, location, and curvature are estimated from the data. Until about stage (g) NMF-DL used simple blob models, at (g) and beyond, NMF-DL decided that it needs more complex parabolic models to describe the data. Iterations stopped at (h), when similarity stopped increasing. S/C improvement in this case against best previously available algorithms is two orders of magnitude (about 7,000%)

V. A BRIEF REVIEW OF SEVERAL NMF-DL APPLICATIONS

NMF-DL has improved solutions in several classical engineering application areas by orders of magnitude, solved several types of problems that have been considered unsolvable, and as discussed in the following section, explains some properties of the mind that have not been explained before. These strong claims require equally strong justifications, which is one of the goals of this section. Another goal is to summarize here several NMF-DL applications relatively inaccessible to this journal readers; some of these applications address classical engineering problems, in which significant improvement was obtained over all previously available algorithms, like the one considered in the previous section; the solved problems have been previously considered unsolvable. Other applications are related to higher functioning of the mind, to modeling language and cultures, to developing next generations of search engines, to modeling and diagnosing cultures, and other application areas that are emerging as novel and important areas of engineering, and our goal is to emphasize that NMF-DL brings these areas within modeling of the mind, and thus, neural networks.

We begin with an important application area of clustering [30]. Relating NMF-DL to this mature area helps understanding more complicated applications considered in the previous section and even more complex emerging applications considered later. NMF-DL clustering applications have been described in [1,10] (see also references and open source publications in [31]). When Gaussian functions are used for $\ell(n|m)$ as in eq. (7), and $\mathbf{X}(n)$, $\mathbf{M}_m$ are points in a multi-dimensional feature space, eqs.(5) lead to a Gaussian Mixture (GM) clustering. To account for non-spherical cluster shapes, diagonal or full covariance matrixes can be used [1,10]. Although GM clustering has been considered in open literature [32] prior to NMF-DL open publications, GM clustering has not been considered practically useful, because of problems with local convergence and for other reasons [30,33]; According to Fukunaga [34], NMF-DL demonstrated that GM can be practically useful. Any mixture model can be used within NMF-DL formalism. If in addition to sources of interest, random sources of signals of no interest are also present, using clutter model eq.(6) would greatly improve result, especially if clutter is dense [1].

Clustering usually is used when there is no knowledge about expected structures in data. However, usually there *is* some knowledge, or at least intuition about expected data structures, and NMF-DL enables to transform these vague knowledge or intuitions into mathematical formulation and significantly improve clustering according to subjective criteria of the scientist. This publication emphasizes detection, recognition, perception (and many other applications) where some prior knowledge exists about data structures. In these cases, as discussed, conditional similarities describe errors (deviations) between models and data; in these cases using Gaussian functions is advisable, unless specific knowledge guides to using different types of

functions (see discussions and examples in [1]).

Another important classical application is tracking. From the NMF-DL point of view, the difference between tracking and clustering is in the models. Whereas in clustering the models $\mathbf{M}_m$ are points in multidimensional feature spaces, in NMF-DL tracking models describe tracks in 2 or 3 dimensional geometric coordinate spaces. This view on tracking as clustering has been revolutionary, when first published in 1991 (see references in [1]). It led to breakthrough improvements for tracking in clutter, to maximum likelihood tracking in strong clutter, and enabled derivation of Cramer-Rao Bounds (CRB) for tracking in clutter, all of these have been previously considered impossible (see references in [1]). Algorithms in this area have been continuously improving since the WWII, nevertheless, popular algorithms for tracking in clutter grossly under-perform the CRB. When tracking in clutter, tracking (estimating track parameters) and association (deciding which data points belong to which track, or to clutter) have to be performed jointly, so called "track-before-detect." This problem is often considered NP-complete and therefore unsolvable [35]. NMF-DL tracker [12] improved practical performance by two orders of magnitude (9,000% in S/C) and achieved the information-theoretic limit of the CRB. Multiple hypotheses tracking and other combinatorially complex algorithms such as particle filters (which consider tracking and association as separate parts of the problem) are more complex than NMF-DL in implementation and inferior in performance by orders of magnitude. Still they are much more known and used than NMF-DL tracker. So judging from this practice, NMF-DL tracking is still a revolution in waiting, 17 years after open literature publication.

Why are some important mathematical discoveries immediately recognized and adopted by engineering community, such as e.g. Aristotelian logic, and logic-based AI, whereas other immensely important discoveries, such as Aristotelian theory of the mind (summarized in section VII), the Gödelian theory (recognized overnight, but implications are still ignored), Zadeh's fuzzy logic, and others, including NMF-DL, remain misunderstood and unaccepted for years? Engineering community used to relegate these questions to "philosophy," unneeded to engineers, or belonging at best to marketing-engineering. I suggest that our community is ready to consider this question as an important area of science and engineering. The proposed here novel research direction considers acceptance (or not) of scientific ideas as based on processes in the mind-brain, and therefore being a subject for study, particularly by the TNN community. Specific ideas and research directions are considered in section VIII.

Another classical important engineering area addressed by NMF-DL is fusion of signals from multiple sensors, platforms, or data bases. In dense (or strong) clutter, detecting relevant signals or information may not be possible in a single sensor image, or in a single data base. The problem is similar to tracking; detection, tracking, and fusion have to be performed concurrently, sometimes it is called "fuse-before-track" or "fuse-before-detect," these problems are usually considered unsolvable because of CC. Similar situation exists in data mining; when mining multiple data bases, how would the algorithm know that a word or phrase in one data base is related to a telephone call in another data base, unless say, a keyword allows the algorithm to connect the relevant pieces of information. For fusion, the NMF-DL equations in the previous sections require no modifications; the data have now an additional index, indicating a sensor (or data source) and correspondingly the models have to be developed for each sensor. Data and models may include only geometric measurements, or classification features as well; the latter case is called feature-added fusion. In [18] NMF-DL has been applied to a complicated case of feature-added fusion using sensors on three platforms; the S/C ratio was inadequate to detect objects in a single frame, or even to track and detect using a single sensor. Therefore a joint feature-added detection-tracking-classification had to be performed. An additional difficulty was that GPS and other navigation tools were not sufficient for accurate coordinate registration, so all of the above had to be performed jointly with the relative coordinate estimation of the three platforms. Problems of this level of difficulty have never been previously considered, and there is no other algorithm or neural network capable of solving them.

Emerging area of engineering, design of Internet search engines, have been considered in [19,26]. Everyone is familiar with frustrations of using Yahoo or Google, because they do not understand what a user really wants. These references consider how to model language understanding (and learning). The inability so far to engineer natural language understanding, after more than 30 years of efforts, is related in these papers to CC of the problem, and an extension of DL to language learning is developed.

Another emerging area of engineering is modeling cultures and their evolution. Misunderstanding among cultures is possibly the most significant problem facing the humankind in the 21st century. In [21,23,25,28,29,29] NMF-DL has been extended to modeling cultures; it has been demonstrated that differences in language emotionalities is an important mechanism of different cultural evolutionary paths, and a joint psycholinguistic and mathematical problem was outlined along with approximate solutions. Preliminary experimental confirmations of these ideas have been published in [36]. Several neuro-imaging laboratories are working on more detailed verification of this theory.

A next step to higher intelligence involves integrating language with cognition; the above references provide a mathematical basis for its solution, and a preliminary experimental neuroimaging support was recently published [37]. Currently computers do not understand contents of verbal communications; while attempting to fuse sensor data, contents of verbal communication is left for human use. Understanding contents of verbal communication and fusing

it with sensor data using NMF-DL was described in [20]. Integration of language and cognition is also a necessary step for emerging engineering area of collaborative systems, where computers would learn from people, communicate with users in natural language and also with understanding of how words and phrases relate to object and situations.

Even more intelligent human-computer communication area emerges. Future computer systems would be able to communicate with humans emotionally as well as conceptually. Current "emotional" toys and robots simulate emotional look-alike without having any mechanisms resembling human (or animal) emotions. The mechanisms of emotions and their role in cognition is discussed in [22,24,25,27,29]. As discussed in section III specific emotions related to satisfaction of the knowledge instinct are called aesthetic emotions, they serve as foundations of our higher mental abilities. These references also discuss the role of the beautiful in cognition and consciousness, and the role of music in evolution of cultures. These references also explain why high-level cognition cannot be developed separately from language; both abilities have to be developed jointly [38,39,40].

Engineering applications of NMF-DL briefly summarized in this section addressed a number of classical problems that could not have been previously solved. We also summarized emerging application areas related to high cognitive abilities, which did not previously exist, and which will be the basis for developing future systems with higher level intelligence. This overview gives a foundation for psychological interpretation of NMF-DL in the next section.

VI. PSYCHOLOGICAL INTERPRETATION OF NMF-DL

This section describes relationships between NMF-DL and high-level mental abilities: computational intelligence is closer to the mind that many believe. A better understanding of the mind is offered here, as well as a direction to proceed in order to verify this understanding using neural experiments. This understanding is necessary for engineering the next-level intelligent systems, including a collaborative human-computer system, understanding language and using emotions as part of their cognitive mechanisms. Some of these systems are already under development, as discussed in the previous section. Similarly, some psychological contents of this section are either known in neuro-psychology, or are the subject of ongoing experimental verifications. In addition, we outline future neuro-psychological experimental programs. It turns out that by unifying and combining psychological facts with mathematical facts, much is gained for deeper understanding of the mind-brain as well as for engineering of intelligent systems.

A. NMF-DL and the Knowledge Instinct

NMF-DL matches top-down model signals $M_m(S_m,n)$ to bottom-up signals $X(n)$, ultimately coming from sensor organs (eyes). This process is necessary to understand what's going on around us. To satisfy any instinctual need – for food, survival, or procreation – first and foremost we need to understand the world around. Therefore, understanding the world is an instinctual need. Understanding is achieved by improving models, which contain knowledge about the world. Therefore we call this instinct the knowledge instinct (KI). It is an inborn mechanism in our minds, an instinctual drive for cognition which compels us to constantly improve our knowledge of the world. Mathematically it is given by maximizing the similarity measure between the knowledge-models and the world, L, eq.(3),

Biologists and psychologists have discussed various aspects of this mechanism, a need for positive stimulations, curiosity, a motive to reduce cognitive dissonance, a need for cognition [41,42,43,44,45,46]. Until recently, however, this drive was not mentioned among 'basic instincts' on a par with instincts for food and procreation.

The fundamental nature of this mechanism became clear during mathematical modeling of workings of the mind. Our knowledge always has to be modified to fit the current situations. We don't usually see exactly the same objects as in the past: angles, illumination, and surrounding contexts are different. Therefore, our internal representations have to be modified; adaptation-learning is required [47,48,49].

Virtually all learning and adaptive algorithms maximize correspondence between the algorithm internal structure (knowledge in a wide sense) and objects of recognition; the psychological interpretation of this mechanism is KI. The mind-brain mechanisms of KI are discussed in [46]. As we discuss below, it is a foundation of our higher cognitive abilities, and it defines the evolution of consciousness and cultures. NMF-DL mathematically implements the KI, and basic mechanisms of the mind identified by many authors [1,47,48,49,50]. These include concepts, instincts, intuition, imagination, emotions, and aesthetic emotions of the beautiful.

B. Mechanism of Concepts.

We use "concept" to designate a common thread among usages of words like concept, idea, understanding, thought, or notion. Concepts are abstract in that they treat individual entities as if they were identical. Emphasizing this property, Middle Age philosophers used the term "universals." Plato and Aristotle called them ideas or forms, and considered them the basis for the mind understanding of the world. Similarly, Kant considered them a foundation for the ability for understanding, the contents of pure reason [51]. According to Jung, conscious concepts of the mind are learned on the basis of inborn unconscious psychic structures, or archetypes [52]. Contemporary science often equates the mechanism of concepts with internal

representations of objects, their relationships, situations, etc. In NMF concepts are described by models, $\mathbf{M}_m$. The essential mechanism of dynamic logic, as discussed, is the process "from vague to crisp," models stored in memories are vague, fuzzy, uncertain; during perception and cognition they generate initial top-down signals; interacting with bottom-up signals, models become concrete, certain, and crisp.

C. Mechanism of Instincts.

The functioning of the mind and brain cannot be understood in isolation from the system's "bodily needs." A biological system needs to replenish its energy resources (to eat). This and other fundamental unconditional needs are indicated to the system by instincts. Scientific terminology in this area is still evolving; in NMF, as a step toward uncovering neural mechanisms of the mind, we describe instincts mathematically as internal sensors, which measurements directly indicate unconditional needs of an organism. For example, instinct for food measures sugar level in blood. Our bodies have many internal sensors measuring body states essential for survival, such as blood pressure, temperature, etc. In this paper we discussed in details mechanisms of KI, which are described mathematically as maximization of similarity measure, L, eq.(3), between bottom-up signals and model-generated top-down signals.

D. Mechanism of Emotions.

How do instinctual measurements affect our thinking and behavior? Clearly, we do not consciously "read" instinctual sensor "dials." Instinctual needs are made available to decision-making parts of our minds by emotional neural signals [50]. In this way emotional signals affect processes of perception and cognition. Objects satisfying instinctual needs receive priority in perception and recognition. For example, when the sugar level in the blood gets low, we feel the corresponding emotional signals as hunger, and recognition of food objects receives priority over other objects.

E. Mechanism of Aesthetic Emotions.

KI constantly generates emotional signals, which we perceive as feelings of harmony or disharmony between our knowledge and the world [22,24,27]; these emotions drive us to improve our mind's models-concepts for better correspondence to surrounding objects and events.

Mathematically aesthetic emotions are given by changes in similarity measure dL/dt. When new data are coming, which do not correspond to existing models, the similarity change dL/dt is negative, understanding is low, and aesthetic emotions are negative, indicating dissatisfaction of the learning instinct. This stimulates learning. In the process of learning, dL/dt is positive, dynamic logic NMF emotionally enjoys learning. It might seem as an exaggeration, when we refer to a simple algorithm "enjoying" learning of simple patterns. However, when thousands of DL-NMF agents would understand the world (or Internet), while communicating among themselves and human users, the words "emotions" and "enjoy" would be more easy to accept as accurate description and similar to mechanisms of the human mind.

At lower levels of perception we usually do not notice these emotions; they are below the level of consciousness as long as our perception is adequate. However, they reach conscious level fast, when perception does not correspond to events; thriller movies exploit this mechanism of aesthetic emotions and KI.

Emotions related to knowledge have been called aesthetic since Kant. KI and related emotions refer to processes in the brain, and in this way they are more "spiritual" than bodily instincts of hunger or procreation. Aesthetic here does not refer to specifically artistic experiences or perceptions of art. I would like to emphasize that aesthetic emotions are an inseparable part of any perception or cognition. These most everyday emotions are related in the next paragraph to perception of the beautiful.

D. Mechanism of Emotions of the Beautiful.

Cognitive science is at a complete loss when trying to explain the highest human abilities, the most important and cherished ability to create and perceive the beautiful. Its role in the working of the mind was not understood. Aesthetic emotions discussed above are often below the level of consciousness at lower levels of the mind hierarchy. Simple harmony is an elementary aesthetic emotion related to improvement of object-models. Higher aesthetic emotions are related to the development and improvement of more complex "higher" models at higher levels of the mind hierarchy. At higher levels, when understanding important concepts, aesthetic emotions reach consciousness.

Models at higher levels of the mind hierarchy are more general than lower-level models; they unify knowledge accumulated at lower levels. The highest forms of aesthetic emotions are related to the most general and most important models near the top of the mind hierarchy. According to Kantian analysis [53,54] among the highest models are models of the meaning of our existence, of our purposiveness or intentionality. KI drives us to develop these models, because in addition to detailed models of objects and events required at every hierarchical level, another aspect of knowledge is a more general and unified understanding of lower-level models at higher levels. The most general models at the top of the hierarchy unify all our knowledge and are

perceived as the models of meaning and purpose of existence. In this way KI corresponds to Kantian analysis.

Everyday life gives us little evidence to develop models of meaning and purposiveness of our existence. People are dying every day and often from random causes. Nevertheless, life itself demands belief in one's purpose; without such a belief it is easier to get drunk or take drugs than to read this article.

Models of our purposiveness are vague and unconscious. Some people, at some points in their life, may believe that their life purpose is finite and concrete, for example to make a lot of money, or build a loving family and bring up good children. These models are aimed at satisfying powerful instincts, but not the knowledge instinct, and they do not reflect the highest human aspirations. Everyone who has achieved a finite goal of making money or raising good children knows that this is not the end of his or her aspirations. The psychological reason is that everyone has an ineffable feeling of partaking in the infinite, while at the same time knowing that one's material existence is finite. This contradiction cannot be resolved. For this reason models of our purpose and meaning cannot be made crisp and conscious, they will forever remain vague, fuzzy, and mostly unconscious. The feel of emotions of beautiful is related to improving these highest models.

These issues are not new; philosophers and theologians expounded them from time immemorial. The NMF-DL and knowledge instinct theory gives us a scientific approach to the eternal quest for the meaning. We perceive an object or a situation as beautiful, when it stimulates improvement of the highest models of meaning. Beautiful is what "reminds" us of our purposiveness. This is true about perception of beauty in a flower or in an art object. Just an example, R. Buckminster Fuller, an architect, best known for inventing the geodesic dome wrote: "When I'm working on a problem, I never think about beauty. I think only how to solve the problem. But when I have finished, if the solution is not beautiful, I know it is wrong" [55]. Similar things were told about scientific theories by Einstein and Poincare. The NMF explanation of the nature of the beautiful helps understanding an exact meaning of these statements and resolves a number of mysteries and contradictions in contemporary aesthetics [22,24,27,28].

E. Mechanisms of Imagination.

Imagination involves excitation of a neural pattern in a sensory cortex in absence of an actual sensory stimulation. For example, visual imagination involves excitation of visual cortex, say, with closed eyes [47,48,49]. Imagination was long considered a part of thinking processes; Kant [53] emphasized the role of imagination in the thought process, he called thinking "a play of cognitive functions of imagination and understanding." Whereas pattern recognition and artificial intelligence algorithms of recent past would not

know how to relate to this [56,57], Carpenter and Grossberg's adaptive resonance model [58] and NMF both describe imagination as an inseparable part of thinking. Imagined patterns are top-down signals that prime the perception cortex areas (priming is a neural terminology for making neurons to be more readily excited). In NMF, the imagined neural patterns are given by models $\mathbf{M}_m$.

Visual imagination, as mentioned, can be "internally perceived" with closed eyes. The same process can be mathematically modeled at higher cognitive levels, where it involves models of complex situations or plans. Similarly, models of behavior at higher levels of the hierarchy can be activated without actually propagating their output signals down to actual muscle movements and to actual acts in the world. In other words, behavior can be imagined, along with its consequences, it can be evaluated, and this is the essence of plans. Sometimes, imagination involves detailed alternative courses of actions considered and evaluated consciously. Sometimes, imagination may involve fuzzy or vague, barely conscious models in the process of adaptation, which reach consciousness only after they converge to a "reasonable" course of action, which can be consciously evaluated. From a mathematical standpoint, this latter mechanism is the only possible, conscious evaluation cannot involve all possible courses of action; it would lead to combinatorial complexity and impasse.

In agreement with neural data, NMF adds details to Kantian description: thinking is a play of top-down higher-hierarchical-level imagination and bottom-up lower-level understanding. Kant identified this "play" as a source of aesthetic emotion. Kant used the word "play," when he was uncertain about the exact mechanism; this mechanism, according to this paper, is KI and dynamic logic.

F. Mechanism of Intuition.

Intuitions include inner perceptions of models, imaginations produced by them, and their relationships with objects in the world. Their mathematical-psychological status might be similar to Figs 3d through 3g; but the whole process in Fig.3 is fast, it takes about 180 ms, and usually it does not reach consciousness until 3h when it becomes conscious perception. What is subjectively perceived as intuition includes higher-level models of relationships among simpler models, the models in the process of their development, when it takes long time. Intuitions involve vague-fuzzy unconscious concept-models, which are in a state of being formed, learned, and being adapted toward crisp and conscious models (say, a theory). Conceptual contents of vague models are undifferentiated and partly unconscious. Similarly, conceptual and emotional contents of these vague mind states are undifferentiated; vague concepts and emotions are mixed up. Vague mind states may satisfy or dissatisfy the knowledge instinct in varying degrees before they become differentiated and accessible to consciousness,

hence the vague complex emotional-cognitive feel of an intuition. Contents of intuitive states differ among people, but the main mechanism of intuition is the same among artists and scientists. Composers' intuitions are mostly about sounds and their relationships to psyche. Painters' intuitions are mostly about colors and shapes and their relationships to psyche. Writers' intuitions are about words, or more generally, about language and its relationships to psyche. Mathematicians' intuitions are about structure and consistency within a theory, and about relationships between the theory and a priori content of psyche. Physicists' intuitions are about the real world, first principles of its organization, and mathematics describing it. Let me repeat that contents of this subsection as well as of the entire section is a summary of many publications referenced in the previous section.

VII. DYNAMIC LOGIC, ZADEH, GÖDEL, AND ARISTOTLE

Initial state of models in NMF-DL are vague, fuzzy; they do not satisfy rules of logic, but are more similar to the fuzzy logic introduced by Lotfi Zadeh [59] for describing mathematically the imprecision of the mind's reasoning. Fuzzy logic emphasizes that every statement is a matter of degree. This is widely believed to be a sharp break with traditions of classical Aristotelian logic. It is interesting therefore to note that the Aristotelian way of thinking is closer to the fuzzy logic of Zadeh and to dynamic logic, than is usually appreciated. Aristotle closely tied logic to language. He emphasized that logical statements should not be formulated too specifically, otherwise meaning might be lost. He argued, "language contains necessary means for appropriate formulation of logical statements" and "common sense must be used to do it" [60]. However, Aristotle also formulated the "law of excluded middle", which contradicted the uncertainty of language. For more than two thousand years, the legacy of Aristotle has contained this unresolved contradiction.

The story of formal logic, logical AI, and neural networks is widely known. Here I will tell this story emphasizing the uniquely novel side, opposite to what is written in many textbooks. Why David Hilbert believed in logic before Gödel? Why Marvin Minsky and John McCarthy believed in logic even after Gödel? Have we finally understood Gödel and expelled logic from computational intelligence, or is it still lurking behind the corner in a fundamental way and why? So let us look at the old story once more.

The contradiction was noted in the 19th century by George Boole, who thought that logic could be improved by excluding any uncertainty which is a part of causal language. A great school of logic formalization emerged, promising in the eyes of many to completely and forever formalize scientific discourse. Prominent mathematicians contributed to the development of formal logic, including George Boole,

Gottlob Frege, Georg Cantor, Bertrand Russell, David Hilbert, and Kurt Gödel. Logicians cast aside the uncertainty of language and founded formal mathematical logic based upon the law of excluded middle. Most physicists today agree that the exactness of mathematics is an inseparable part of physics, but formal logicians went beyond this. Hilbert developed an approach named formalism, which rejected intuition as a part of scientific investigation and thought to define scientific objects formally in terms of axioms or rules. The physical reality of the world, he thought, could be equally represented by any set of axioms that did not contradict physical data.

This and the following excerpts from the history of formal logic, which might seem well known to some researchers, are repeated here with novel and unique emphases with a goal to understand the recent and future development of the fields of neural network and computational intelligence. In particular, in this and following sections we investigate answers to questions, why so many smart people, including Hilbert and founders of logic-based artificial intelligence believed that logic is sufficient to understand workings of the mind.

Hilbert was sure that his logical theory also described mechanisms of the mind: "The fundamental idea of my proof theory is none other than to describe the activity of our understanding, to make a protocol of the rules according to which our thinking actually proceeds" [61]. In the 1900s he formulated his famous Entscheidungsproblem: to define a set of logical rules sufficient to prove all past and future mathematical theorems [62]. This entailed the formalization of scientific creativity and the entire human thinking.

Almost as soon as Hilbert formulated his formalization program, the first hole appeared. In 1902 Russell exposed an inconsistency of formal procedures by introducing a set R as follows: R is a set of all sets, which are not members of themselves [63]. Is R a member of R? If it is not, then it should belong to R according to the definition, but if R is a member of R, this contradicts the definition. Thus, either way we get a contradiction. This became known as the Russell's paradox. Its jovial formulation is as follows: A barber shaves everybody who does not shave himself. Does the barber shave himself? Either answer to this question (yes or no) leads to a contradiction. This barber, like Russell's set, can be logically defined but cannot exist. For the next 25 years mathematicians where trying to develop a self-consistent mathematical logic, free from the paradoxes of this type. But in 1931 Gödel proved that it is not possible [64]; formal logic was inconsistent, and self-contradictory.

Then, why 25 years after Gödel where founders of artificial intelligence sure that logic is sufficient? Today we know that logic is not a fundamental mechanism of the mind. Still, as discussed, most algorithms and neural networks use logic at some fundamental step and this limits their functioning. Why this has to be so? This paper answers this question, analyzes "why" and answers "how" of the logical biases in algorithms, neural networks, and scientists' minds.

The paper explains that logical or approximately logical reasoning is not a fundamental mechanism, but a result of dynamic logic process, "from vague to crisp."

Returning to Aristotle, we note that he considered logic as a way to correctly argue for conclusions which have been already obtained. This is clearly seen, for example, from "Rhetoric for Alexander" [65], where he lists logical arguments that should be used in public speeches, arguing both sides of various political issues. Such issues might include declaring war or making piece, signing treaty or refusing it, trusting or mistrusting a witness, whether or not to use torture to obtain trustworthy evidence, etc. Aristotle provided exact logical ways to argue both for and against any issue. Never had he given the impression that logic was a mechanism of obtaining truth. Logic, to him, was a tool of politics and not of science, not a primary mechanism of the mind. I would extend Aristotelian arguments for scientists: use logic when writing a paper, but not when solving a new problem, and not when developing neural networks or algorithms for solving new problems.

When Aristotle was seeking an explanation of workings of the mind, he developed a theory of Forms [60]. The main tenets of this theory are that perception and cognition are processes in which "a priori Forms meet matter," or in contemporary scientific language, top-down signals interact with bottom-up signals. This process is the foundation for all our experience, and it creates concepts with which our mind thinks and perceives individual objects and situations. Before "meeting matter" a priori Forms exist in our minds as "potentialities." After "meeting matter," they turn into "actualities." He emphasized that potentialities do not obey the rule of excluded middle, and therefore are not logical, but actualities obey logic [66]. Thus, the process of "Forms meeting matter" corresponds to interaction between top-down and bottom up signals, as described by dynamic logic, "from vague to crisp."

VIII. EARLY ARTIFICIAL INTELLIGENCE AND LOGIC

For a long time people believed that intelligence is equivalent to conceptual understanding and reasoning, and that the mind works according to logic. Although it is obvious that the mind is not logical, over the course of the two millennia since Aristotle, the power of intelligence has been identified with logic. Founders of artificial intelligence in the 1950s and 60s believed that by relying on rules of logic they would soon develop computers more intelligent than the human mind. However, this did not happen.

One may wonder why, despite the Gödel's theory developed in the 1930s and immediately recognized as a fundamental result, mathematicians still relied on formal logic when developing artificial intelligence in the 1950s and 60s, and many still rely today?

The reason is related to mechanisms of the mind, which were understood recently [47,48,49]. Most of the mind's mechanisms are usually inaccessible to consciousness, e.g., we are not conscious about individual neural firings and intermediate signal processing steps. Only the "final results" of perception and cognition, clear crisp logic-like perceptions and thoughts, are available to consciousness. These "final results" approximately obey rules of logic. The mind creates logic out of illogical mechanisms according to dynamic logic. Logical conscious states are like islands in the ocean of unconscious. But in our consciousness there are only crisp logical results, and consciousness works so that we subjectively feel as if we smoothly flow from one conscious logical state to the next. Our intuitions about the mind, including scientific intuitions are strongly biased toward logic. This is why, most algorithms and neural networks (even if ostensibly designed to oppose logic) use logic in a fundamental way, as discussed in section III.

This "logical bias" of conscious thinking also answers another question posed in section V. Why are some mathematical discoveries immediately adopted by engineering community, such as logic-based AI, whereas other immensely important discoveries are accepted only after many years? Among theories waiting decades for acceptance are Zadeh's fuzzy logic, Kanehman-Tversky's theory [67] (2003 Nobel Prize, after Tversky's death), Grossberg's theories, and others, including NMF-DL. The conclusion from the above analysis is that theories of illogical mechanisms remain misunderstood and unaccepted for years because of logical bias in scientific thinking.

IX. EXPERIMENTAL VALIDATION OF DYNAMIC LOGIC

Neural processes of perception involved in dynamic logic are complex and only recently understood [47,48,49] Using this understanding, experimental validation of dynamic logic can be obtained by everyone in 3 seconds. Just close you eyes and imagine a familiar object that you observed in front of you just a second ago. Your imagination is vague-fuzzy, not as crisp as perception of the object with open eyes. As discussed earlier, imagination is produced in the visual cortex by top-down signals from models in your memory. This proves that in the initial stages of perception memories-models producing top-down signals are vague, as in dynamic logic. This is a unique property of DL, no other theory emphasized the fundamental role of vagueness of initial top-down projections.

Detailed neurological and fMRI neuroimaging studies [68,69,70] confirmed that conscious perceptions are preceded by activation of cortex areas, where top-down signals originate; initial top-down projections are *vague*. These experiments confirmed the unique property of DL, a process "from vague to crisp."

X. FUTURE DIRECTIONS

NMF-DL eqs.(4) describe a single layer interaction of top-down and bottom-up signals. It seems clear how to combine layers into a hierarchy; detailed mathematical and simulation studies of multi-layer hierarchical NMF-DL still is one of the future research directions. What changes are necessary for self-expanding hierarchy? What is optimal hierarchy? Would the highest models of meaning and purpose appear in a single-agent NMF-DL system, under the drive from KI? Or would it be necessary to consider multi-agent systems with communicating agents, competing for various resources, and using their knowledge and hierarchical organization in this competition, before importance of the highest models of meaning would be observed? What would be the differences between these models? What would we learn from such models about the meanings of our own lives? This research program outlined above encompasses an ambitious goal of modeling the mind, human societies, cultures, and their evolutions.

NMF-DL can be related to chaotic neurodynamics [71]; this reference suggests that DL might be "implemented" in the brain as a phase transition from high-dimensional chaotic state to a low dimensional chaos; to research on spiking neurons [72] possibly implying that DL might be "implemented" in the brain as an increased correlation of spiking trains; to research on consciousness [73,74].

A hypothesis that algorithms and scientific theories advocating logic as their base are accepted much faster than those that use logic to uncover illogical bases can be verified in history of science and psychological experiments.

Human mind is not driven entirely by KI. A different mechanism of decision-making is aimed at discarding "too much" knowledge; there is a well established psychological principle of "effort minimization," including cognitive effort. It would be necessary to develop more complicated models, which will take into account both principles [46].

NMF-DL should be extended to modeling mind's ability for language, and interactions between cognition and language. Initial results [19,21] indicate that these processes define evolution of languages and cultures. The next step should simulate multi-agent systems, each agent possessing NMF-DL mind, and through simulations connect results of this paper with the simplified theory presented in [21].

Using experimental studies such as [69,69,70] mathematical mechanisms in this paper should be related to specific mind's modules and circuits. Further experimental studies should extend detailed mathematical description of a single mind layer to the entire hierarchy of the mind, to high cognitive functions, to cognition of abstract concepts and emotions of the beautiful.

REFERENCES

[1] Perlovsky, L.I. *Neural Networks and Intellect: Using model based concepts*. New York: Oxford University Press, 2001.

[2] A. G. Barto, R. S. Sutton and P. S. Brouwer. "Associative search network: A reinforcement learning associative memory." *IEEE Transactions on Systems, Man, and Cybernetics*, 40: 201-211, 1981.

[3] M. H. DeGroot, and M. J. Schervish. *Probability and Statistics*. Reading, MA: Addison Wesley, 2001.

[4] V.N. Vapnik. *Statistical Learning Theory*. New York: Wiley-Interscience, 1998.

[5] R. A. Singer, R. G. Sea, and R. B. Housewright. "Derivation and Evaluation of Improved Tracking Filters for Use in Dense Multitarget Environments." *IEEE Transactions on Information Theory*, IT-20, 423-32, 1974.

[6] L.I. Perlovsky. "Conundrum of Combinatorial Complexity." IEEE Trans. PAMI, 20(6) pp. 666-670, 1998.

[7] M.L. Minsky. "A Framework for Representing Knowlege." In *The Psychology of Computer Vision*, ed. P. H. Whinston, McGraw-Hill Book, New York, 1975.

[8] P. H. Winston. *Artificial Intelligence*. Reading, MA: Addison-Wesley, 1984.

[9] Perlovsky, L.I. (1996). "Gödel Theorem and Semiotics." *Proceedings of the Conference on Intelligent Systems and Semiotics '96*. Gaithersburg, MD: v. 2, pp. 14-18.

[10] L.I Perlovsky and M. M. McManus. "Maximum Likelihood Neural Networks for Sensor Fusion and Adaptive Classification." *Neural Networks* 4 (1), pp. 89-102, 1991.

[11] Cramer, H. (1946). *Mathematical Methods of Statistics*, Princeton NJ: Princeton University Press.

[12] L.I. Perlovsky and R.W. Deming. "Neural Networks for Improved Tracking." *IEEE Trans. Neural Networks,* 18(6), pp. 1854-1857, 2007.

[13] Wikipedia. June, 2007. http://en.wikipedia.org/wiki/Receiver_operating_characteristic.

[14] B. Kovalerchuk and L. I. Perlovsky. "Dynamic Logic of Phenomena and Cognition." *International Joint Conference on Neural Networks*, Atlanta, 2009, accepted for publication; Submitted for journal publication.

[15] R. Deming, S. Higbee, D. Dwyer, M. Welser, L. Perlovsky, and P. Pellegrini. "Robust detection and spectrum estimation of multiple sources from rotating-prism spectrometer images." *Proceedings of SPIE Remote Sensing Europe*, Stockholm, Sweden, 2006.

[16] R. Linnehan, J. Schindler, D. Brady, R. Kozma, R. Deming, L. Perlovsky. "Dynamic Logic Applied to

SAR Data for Parameter Estimation Behind Walls. *Proc. IEEE Radar Conference.* Piscataway, NJ: IEEE Press, pp. 850-855, 2007.

[17] R. Linnehan, C. Mutz, L. I. Perlovsky, B. Weijers, J. Schindler, R. Brockett. "Detection of Patterns Below Clutter in Images." *Int. Conf. on Integration of Knowledge Intensive Multi-Agent Systems*, Cambridge, MA, 2003.

[18] R.W. Deming and L.I. Perlovsky. "Concurrent multi-target localization, data association, and navigation for a swarm of flying sensors, *Information Fusion*, 8, pp.316-330, 2007.

[19] L.I. Perlovsky. "Integrating Language and Cognition." *IEEE Connections*, Feature Article, 2(2), pp. 8-12, 2004

[20] L.I. Perlovsky. "Cognitive High Level Information Fusion." *Information Sciences*, 177, pp. 2099-2118, 2007.

[21] L.I. Perlovsky. "Evolution of Languages, Consciousness, and Cultures. *IEEE Computational Intelligence Magazine,* 2(3), pp.25-39, 2007.

[22] L.I. Perlovsky. "Aesthetics and Mathematical Theories of Intellect." *Iskusstvoznanie*, 2(02), pp.558-594, Moscow, (Russian), 2002.

[23] L.I. Perlovsky. "Evolution of consciousness and music." *Zvezda*, 2005(8), 192-223, St. Petersburg (Russian), 2005.

[24] L.I. Perlovsky. "Modeling field theory of higher cognitive functions." In A. Loula, R. Gudwin, J. Queiroz, Eds. *Artificial cognition systems*. Hershey, PA: Idea Group, pp. 64-105, 2007.

[25] L.I. Perlovsky. "Music–the first principles." *Musical Theater*, 2006. <http://www.ceo.spb.ru/libretto/kon_lan/ogl.shtml>

[26] L.I. Perlovsky. "Symbols: Integrated cognition and language. In R. Gudwin, J. Queiroz (Eds.). *Semiotics and intelligent systems development.*" Hershey, PA: Idea Group, pp.121-151, 2007.

[27] L. I. Perlovsky. "Toward physics of the mind: Concepts, emotions, consciousness, and symbols." Physics of Life Reviews, 3(1), 22-55, 2006.

[28] L. I. Perlovsky. "Joint evolution of cognition, consciousness, and music." *Lectures in Musicology, School of Music*, Columbus, OH: University of Ohio. 2006.

[29] L. I. Perlovsky. "Music and consciousness." *Leonardo, Journal of Arts, Sciences and Technology*, 41(4), pp. 420-421, 2008.

[30] R. Xu and D. C. Wunsch II. "Survey Of Clustering Algorithms." *IEEE Trans. on Neural Networks*, Vol. 16(3), pp. 645-678, 2001.

[31] L. I. Perlovsky, web page, http://www.leonid-perlovsky.com

[32] D.M. Titterington, A.F.M. Smith, U.E. Makov. *Statistical Analysis of Finite Mixture Distributions.* Chichester, UK: John Wiley and Sons, 1985.

[33] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. (First Edition). New York: Academic Press, 1972.

[34] K. Fukunaga, private communication, 1986.

[35] I. O. Sebe, S. You, U. Neumann. "Globally optimum multiple object tracking." In M. K. Masten, L. A. Stockum, Eds. *SPIE Proceedings*, Vol. 5810, Acquisition, Tracking, and Pointing XIX, pp.82-93, 2005.

[36] C. L. Harris, A. Ayçiçegi, and J. B. Gleason. "Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language." *Applied Psycholinguistics*, 24: 561-579, 2003.

[37] A. Franklin, G. V. Drivonikou, A. Clifford, P. Kay, T. Regier, and I. R. L. Davies. "Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults." *Proceedings of the National Academy of Sciences,* 105: 3221–3225, 2008.

[38] J.F. Fontanari and L.I. Perlovsky. "How language can help discrimination in the Neural Modeling Fields framework." *Neural Networks*, 21(2-3), pp. 250–256, 2008.

[39] J. F. Fontanari, V. Tikhanoff, A. Cangelosi and L. I. Perlovsky. "A cross-situational algorithm for learning a lexicon using Neural Modeling Fields." *International Joint Conference on Neural Networks*, Atlanta, accepted, 2009.

[40] L. I. Perlovsky, "Language and Cognition." *Neural Networks*, 22(3), in print, 2009.

[41] H.F. Harlow and C. Mears. *The Human Model: Primate Perspectives*, Washington, DC: V. H. Winston and Sons, 1979.

[42] D. E. Berlyne. *Conflict, Arousal, and Curiosity.* New York, NY: McGraw-Hill, 1960.

[43] D. E. Berlyne. *Pleasure, Reward, Preference: Their Nature, Determinants, and Role in Behavior.* New York, NY: Academic Press, 1973.

[44] Festinger, L. (1957). A Theory of Cognitive Dissonance, Stanford, CA: Stanford University Press.

[45] J.T. Cacioppo, R.E. Petty, J.A. Feinstein, and W. B. G. Jarvis. "Dispositional Differences in Cognitive Motivation: The Life and Times of Individuals Varying in Need for Cognition." *Psychological Bulletin*, 119(2), pp. 197-253, 1996.

[46] D. S. Levine and L. I. Perlovsky. "Neuroscientific Insights on Biblical Myths: Simplifying Heuristics versus Careful Thinking: Scientific Analysis of Millennial Spiritual Issues." *Zygon, Journal of Science and Religion*, 43(4), 797-821, 2008.

[47] S. Grossberg. *Neural Networks and Natural Intelligence*. Cambridge, MA: MIT Press, 1988.

[48] S. Zeki. *A Vision of the Brain*. Oxford, England: Blackwell, 1993.

[49] S. M. Kosslyn, G. Ganis, W. L. Thompson. "Neural foundations of imagery." *Nature Reviews Neuroscience*, 2, pp. 635-643, 2001.

[50] S. Grossberg and D. S. Levine. "Neural dynamics of attentionally modulated Pavlovian conditioning: blocking, inter-stimulus interval, and secondary reinforcement." *Applied Optics,* 26, pp. 5015-5030, 1987.

[51] I. Kant. *Critique of Pure Reason.* Tr. J.M.D. Meiklejohn,. New York, NY: Willey Book, 1781/1943.

[52] C. G. Jung., *Psychological Types.* In the Collected Works, Princeton, NJ: Princeton University Press, 1921/1971.

[53] I. Kant. *Critique of Judgment,* tr. J.H.Bernard, London: Macmillan & Co., 1790/1914.

[54] Kant, I. *Anthropology from a Pragmatic Point of View.* Tr. M.J. Gregor. Boston, MA: Kluwer Academic Pub., 1798/1974.

[55] http://www.quotationspage.com/quote/26209.html

[56] M. Minsky. *The Society of Mind*. Cambridge, MA: MIT Press, 1988.

[57] R. Penrose. *Shadows of the Mind*. Oxford: Oxford University Press, 1994.

[58] G.A. Carpenter and S. Grossberg. "A massively parallel architecture for a self-organizing neural pattern recognition machine." *Computer Vision, Graphics and Image Processing*, 37, pp. 54-115, 1987.

[59] L. A. Zadeh. "Fuzzy sets." *Information and Control*, 8, pp. 338-353, 1965.

[60] Aristotle. "Metaphysics," In T*he Complete Works of Aristotle*, Ed. J. Barnes, Princeton, NJ: Princeton Univ. Press, IV BCE / 1995.

[61] D. Hilbert. "The foundations of mathematics." In J. van Heijenoort, Ed., *From Frege to Gödel*. Cambridge, MA: Harvard University Press, p.475. 1928/1967

[62] D. Hilbert. "Hilbert's Problems and Their Solvers." In *The Honors Class*, Ben Yandell, Natick, MA: AK Peters Ltd, 1900/2003.

[63] B. Russel and A. N. Whitehead. *Principia Mathematica.* Oxford, England: Cambridge Univ. Press, 1908/1962.

[64] K. Gödel. *Kurt Gödel collected works, I.* Ed. S. Feferman at al. Oxford University Press, 1986.

[65] Aristotle. "Rhetoric to Alexander." In *The Complete Works of Aristotle,* Ed. J.Barnes, Princeton, NJ: Princeton Univ. Press, IV BCE/1995.

[66] Aristotle, (IV BCE). "Topics." In *The Complete Works of Aristotle,* Ed. J.Barnes, Princeton, NJ: Princeton Univ. Press, IV BCE/1995.

[67] A. Tversky, and D. Kahneman. "The Framing of Decisions and the Rationality of Choice." *Science* 211: 453-458, 1981.

[68] D.L. Schacter, I.G. Dobbins, and D.M. Schnyer. "Specificity of priming: A cognitive neuroscience perspective." *Nature Reviews Neuroscience*, 5, 853-862, 2004.

[69] M. Bar, K. S. Kassam, A. S. Ghuman, , J. Boshyan, A. M. Schmid, A. M. Dale, M. S. Hamalainen, K. Marinkovic, D.L. Schacter, , B.R. Rosen, and E. Halgren. "Top-down facilitation of visual recognition." *Proceedings of the National Academy of Sciences USA*, 103, 449-54, 2006.

[70] D. L. Schacter and D. R. Addis. "The ghosts of past and future." *Nature*, 445, 27, 2007.

[71] L. I. Perlovsky and R. Kozma. Eds. *Neurodynamics of Higher-Level Cognition and Consciousness*. Heidelberg, Germany: Springer-Verlag, 2007.

[72] N. A. Bomberger, A. M. Waxman, B. J. Rhodes, and N. A. Sheldon. "A new approach to higher-level information fusion using associative learning in semantic networks of spiking neurons." Information *Fusion,* 8(3), pp. 227-251, 2007.

[73] G. Edelman and G. Tononi. *A Universe of Consciousness. How Matter Becomes Imagination*. Basic Books, 2001.

[74] D. J. Chalmers. *The Conscious Mind: In Search of a Fundamental Theory.* New York: Oxford University Press, 1997