

## Chapter III

# Modeling Field Theory of Higher Cognitive Functions

Leonid Perlovsky, Air Force Research Center, USA

## Abstract

---

*The chapter discusses a mathematical theory of higher cognitive functions, including concepts, emotions, instincts, understanding, imagination and intuition. Mechanisms of the knowledge instinct are proposed, driving our understanding of the world. Aesthetic emotions and perception of beauty are related to “everyday” functioning of the mind. We briefly discuss neurobiological grounds as well as difficulties encountered by previous attempts at mathematical modeling of the mind encountered since the 1950s. The mathematical descriptions below are complemented with detailed conceptual discussions so the content of the chapter can be understood without necessarily following mathematical details. We relate mathematical results and computational examples to cognitive and philosophical discussions of the mind. Relating a mathematical theory to psychology, neurobiology and philosophy will improve our understanding of how the mind works.*

## Working of the Mind

---

How the mind works has been a subject of discussions for millennia; from Ancient Greek philosophers to mathematicians, to modern cognitive scientists. Words like *mind*, *thought*, *imagination*, *emotion* and *concept* present a challenge: People use these words in many ways colloquially, but in cognitive science and in mathematics of intelligence they have not been uniquely defined and their meaning is a subject of active research and ongoing debates (for the discussions and further references see: Grossberg (1988); Albus, and Meystel (2001); Perlovsky (2001). Standardized definitions come after completion of a theoretical development (for instance “force” was defined by Newton’s laws, following centuries of less precise usage). Whereas the mind theory is a developing science, this chapter adheres to the following guidelines regarding our proposals: (1) they must correspond to current discussions in the scientific and mathematical community, (2) they must correspond to philosophical discussions and general cultural usage, (3) they must be clear and mathematically tractable, and finally (4) deviations or discrepancies must be noted and discussed. A dictionary definition of the mind, which we take as a starting point, includes conscious and unconscious processes, thought, perception, emotion, will, memory and imagination, and it originates in brain (*The American Heritage College Dictionary*, 2000). These constituent notions will be discussed throughout the chapter. Specific neural mechanisms in the brain “implementing” various mind functions constitute the relationship between the mind and brain. We will discuss possible relationships of the proposed mathematical descriptions to neural structures in the brain.

The problem addressed in this chapter is developing a mathematical technique suitable to describe higher cognitive functions. Such a technique could serve two purposes. First, it would lead to the development of smart computers and intelligent robots. Second, it would help to unify and clarify complex issues in philosophy, psychology, neurobiology and cognitive science. I rely on an underlying methodological assumption that “the minds” are actually existing physical entities, and in various disciplines I am interested in contents related to this assumption of “physics of the mind.” Achieving the two purposes of intelligent computers and mind theory will require the collaboration of many people. Developing intelligent computers based partially on ideas described in this chapter is being pursued by several dozens of companies. Similarly, several university groups pursue research relating these ideas to specific disciplines (philosophy, psychology, neurobiology, evolution of languages, psycholinguistics and even musical theory). In this chapter the purpose is limited to describing the mathematical technique and to making a step toward relating it to a vast field of philosophy, psychology, neurobiology and cognitive science. Needless to say, not every point of view can be addressed, and not every reader can be satisfied. The aim of the chapter will be achieved if the reader gets a taste for and an interest in the unifying approach to this vast and fascinating field.

A broad range of opinions exists about the mathematical methods suitable for the description of the mind. Founders of artificial intelligence, including Allan Newell (1983) and Marvin Minsky (1988), thought that formal logic was sufficient and no specific mathematical techniques would be needed to describe the mind. An opposite view was advocated by Brian Josephson (1997) and Roger Penrose (1994), suggesting that the

mind cannot be understood within the current knowledge of physics; new unknown yet physical phenomena will have to be accounted for explaining the working of the mind. Some authors considered quantum computational processes that might take place in the brain (Penrose, 1994; Josephson, 1997; Hameroff, 1994). This chapter develops a point of view that there are few specific mathematical constructs, or “first principles” of the mind. Several researchers advocated this view. Grossberg (1988) suggested that the first principles include a resonant matching between bottom-up signals and top-down representations, as well as an emotional evaluation of conceptual contents (Grossberg & Levine, 1987). Zadeh (1997) developed the theory of granularity; Meystel (1995) developed the hierarchical multi-scale organization; Edelman suggested neuronal group selection (see Edelman & Tononi, 1995); and the author suggested the knowledge instinct, aesthetic emotions and dynamic logic among the first principles of the mind (Perlovsky, 2001; Perlovsky & McManus, 1991; Perlovsky, 1996).

This chapter presents modeling field theory (MFT), a mathematical “structure” that we propose is intrinsic to operations of the mind, and dynamic logic, governing its temporal evolution. It discusses specific difficulties encountered by previous attempts at mathematical modeling of the mind and how the new theory overcomes these difficulties. I show an example of solving a problem that was unsolvable in the past. We argue that the theory is related to an important mechanism behind workings of the mind, which we call “the knowledge instinct” as well as to other cognitive functions. I discuss neurobiological foundations, cognitive, psychological and philosophical connections, experimental verifications and outline emerging trends and future directions.

## **Logic and the Mind**

---

For long time, people believed that intelligence was equivalent to conceptual understanding and reasoning. A part of this belief was that the mind works according to logic. Although it is obvious that the mind is not logical, over the course of two millennia since Aristotle, many people came to identify the power of intelligence with logic. Founders of artificial intelligence in the 1950s and 60s believed that by relying on rules of logic they would soon develop computers with intelligence far exceeding the human mind.

The beginning of this story is usually attributed to Aristotle, the inventor of logic. He was proud of this invention and emphasized, “nothing in this area existed before us” (Aristotle, IV BCE, a). However, Aristotle did not think that the mind works logically; he invented logic as a supreme way of argument, not as a theory of the mind. This is clear from many Aristotelian writings, for example in “Rhetoric for Alexander” (Aristotle, IV BCE, b), he lists dozens of topics on which Alexander had to speak publicly. For each topic, Aristotle identified two opposing positions (e.g., making peace or declaring war; using or not using torture for extracting the truth, etc.). Aristotle gives logical arguments to support each of the opposing positions. Clearly, Aristotle saw logic as a tool to express decisions that were already made; he did not consider logic as the mechanism of the mind. Logic, if you wish, is a tool for politicians. (I would add that scientists should use logic to present their results, but not to arrive at these results). To explain the mind, Aristotle

developed a theory of Forms, which will be discussed later. During the centuries following Aristotle, the subtleties of his thoughts were not always understood. With the advent of science, the idea that intelligence is equivalent to logic was gaining ground. In the nineteenth century, mathematicians turned their attention to logic. George Boole noted what he thought was not completed in Aristotle's theory. The foundation of logic, since Aristotle (Aristotle, IV BCE, c), was the law of excluded middle (or excluded third): Every statement is either true or false, any middle alternative is excluded. But Aristotle also emphasized that logical statements should not be formulated too precisely (say, a measure of wheat should not be defined with an accuracy of a single grain), that language implies the adequate accuracy and everyone has his mind to decide what is reasonable.

Boole thought that the contradiction between exactness of the law of excluded middle and vagueness of language should be corrected, and a new branch of mathematics, formal logic, was born. Prominent mathematicians contributed to the development of formal logic in addition to Boole, including Gottlob Frege, Georg Cantor, Bertrand Russell, David Hilbert and Kurt Gödel. Logicians "threw away" uncertainty of language and founded formal mathematical logic based on the law of excluded middle. 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. 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" (see Hilbert, 1928). In 1900, he formulated the famous Entscheidungsproblem: to define a set of logical rules sufficient to prove all past and future mathematical theorems. This entailed 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*. 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 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 were trying to develop a self-consistent mathematical logic, free from paradoxes of this type. But in 1931, Gödel (see in Gödel, 1986) proved that it is not possible, formal logic was inexorably inconsistent and self-contradictory.

Belief in logic has deep psychological roots related to functioning of the human mind. A major part of any perception and cognition process is not accessible to consciousness directly. We are conscious about the "final states" of these processes, which are perceived by our minds as "concepts," approximately obeying formal logic. For this reason prominent mathematicians believed in logic. Even after the Gödelian proof, founders of artificial intelligence still insisted that logic is sufficient to explain how the mind works. This is examined in the next section; for now let us simply state that logic is not a mechanism of the mind, but rather the result of the mind's operation (in Section 5 we discuss mathematics of dynamic logic, which suggest a mathematical explanation of how logic appears from illogical states).

## **Perception, Complexity, and Logic**

---

Simple object perception involves signals from sensory organs and internal representations of objects. During perception, the mind associates subsets of signals corresponding to objects with object representations. This recognition activates brain signals leading to mental and behavioral responses, which are important for the phenomenon of understanding.

Developing mathematical descriptions of the very first *recognition* step in this seemingly simple association-recognition-understanding process has not been easy; a number of difficulties have been encountered over the last fifty years. These difficulties were summarized under the notion of combinatorial complexity CC (Perlovsky, 1998). “CC” refers to multiple combinations of various elements in a complex system; for example, recognition of a scene often requires concurrent recognition of its multiple elements that could be encountered in various combinations. CC is prohibitive because the number of combinations is very large. For example, consider 100 elements (not too large a number); the number of combinations of 100 elements is  $100^{100}$ , exceeding the number of all elementary particle events in life in the Universe. No computer would ever be able to compute that many combinations.

The problem was first identified in pattern recognition and classification research in the 1960s and was named “the curse of dimensionality” (Bellman, 1961). It seemed that adaptive self-learning algorithms and neural networks could learn solutions to any problem “on their own” if provided with a sufficient number of training examples. The following thirty years of developing adaptive statistical pattern recognition and neural network algorithms led to a conclusion that the required number of combinations often was combinatorially large. Self-learning approaches encountered *CC of learning requirements*. Rule-based systems were proposed to solve the problem of learning complexity. An initial idea was that rules would capture the required knowledge and eliminate a need for learning. However, in the presence of variability, the number of rules grew; rules depended on other rules, combinations of rules had to be considered and rule systems encountered *CC of rules*. Beginning in the 1980s, model-based systems were proposed. They used models that depended on adaptive parameters. The idea was to combine advantages of learning-adaptivity and rules by using adaptive models. The knowledge was encapsulated in models, whereas unknown aspects of particular situations were to be learned by fitting model parameters (see discussions in [1], and in Perlovsky, Webb, Bradley, & Hansen, 1998). Fitting models to data required selecting data subsets corresponding to various models. The number of subsets, however, was combinatorially large. A general popular algorithm for fitting models to data, multiple hypotheses testing (Singer, Sea & Housewright, 1974) is known to face CC of computations. Model-based approaches encountered *computational CC* (N and NP complete algorithms).

CC is related to the type of logic underlying various algorithms and neural networks (Perlovsky, 1998). Formal logic is based on the “law of excluded middle,” according to which every statement is either true or false and nothing in between. Therefore, algorithms based on formal logic have to evaluate every variation in data or models as a separate logical statement (hypothesis). A large number of combinations of these

variations result in combinatorial complexity. In fact, combinatorial complexity of algorithms based on logic was related to Gödel theory: It is a manifestation of the inconsistency of logic in finite systems (Perlovsky, 1996). Multivalued logic and fuzzy logic were proposed to overcome limitations related to the law of excluded middle (Kecman, 2001). Yet the mathematics of multivalued logic is no different in principle from formal logic; “excluded middle” is substituted by “excluded  $n+1$ .” Fuzzy logic encountered a difficulty related to the degree of fuzziness. If too much fuzziness is specified, the solution does not achieve the required accuracy, if too little, it becomes similar to formal logic. Complex systems require different degrees of fuzziness in various elements of system operations; searching for the appropriate degrees of fuzziness among combinations of elements again would lead to CC. Is logic still possible after Gödel? Bruno Marchal (2005) recently reviewed the contemporary state of this field; logic after Gödel is much more complicated and much less logical than was assumed by the founders of artificial intelligence. The problem of CC remains unresolved within logic.

Various manifestations of CC are all related to formal logic and Gödel theory. Rule systems relied on formal logic in a most direct way. Self-learning algorithms and neural networks relied on logic in their training or learning procedures, every training example was treated as a separate logical statement. Furthermore, fuzzy logic systems relied on logic for setting degrees of fuzziness. CC of mathematical approaches to theories of the mind are related to the fundamental inconsistency of logic.

## **Structure of the Mind**

---

In the 1950s and 60s, developers of artificial intelligence naïvely believed that they would soon create computers exceeding human intelligence, and that mathematics of logic was sufficient for this purpose. As we discussed, logic does not work, but the mind does. So let us turn to the mechanisms of the mind. Possibly, we will find inspiration for developing the mathematics needed for intelligent computers and decipher mechanisms of higher cognitive functions. Mechanisms of the mind, essential for the development of a mathematical theory of intelligence in this chapter include: instincts, concepts, emotions and behavior. Let us look briefly at their current definitions in cognitive science and psychology.

The definitions of instincts, concepts and emotions, as mentioned, are the subject of research and debate, while theories of life and intelligence are in development. Let me summarize few related definitions (The American Heritage College Dictionary, 2000; Catholic Encyclopedia, 2005; Wikipedia, 2005) as a starting point for further elaboration. Instincts are innate capabilities, aptitudes or behavior, which are not learned, complex and normally adaptive. Instincts are different from reflexes, a word used for more simple immediate mechanisms. In humans and higher animals, instincts are related to emotions. Psychoanalysts equated instincts with human motivational forces (such as sex and aggression); today these are referred to as instinctual drives. Motivation is based on emotions, on the search for positive emotional experiences and the avoidance of negative ones.

We will use a word “concept” to designate a common thread among words like concept, idea, understanding, thought or notion. Different authors use these words with subtle differences. A common thread among these words is an abstract, universal psychical entity that serves to designate a category or class of entities, events or relations. A concept is the element of a proposition rather in the way that a word is the element of a sentence. Concepts are abstract in that they omit the differences of the things in their extension, treating them as if they were identical. Concepts are universal in that they apply equally to everything in their extension. Plato and Aristotle called them ideas, or forms, and considered them the basis for how the mind understands of the world. Similarly, Kant considered them a foundation for the ability to understand, the contents of pure reason. According to Jung, conscious concepts of the mind are learned on the basis of inborn unconscious psychic structures, archetypes. Contemporary science often equates the mechanism of concepts with internal representations of objects, their relationships, situations, etc.

Ray Jackendoff (2002) considers the term *representation* or *symbol* as too loaded with the “thorny philosophical problem of intentionality,” and uses the word *model*. I do not think we should be afraid of intentionality; John Searle’s (1980, 1983) emphasis on intentionality as “aboutness” is too narrow.<sup>2</sup> All brain mechanisms and mental functions are intentional; in fact everything within a living being is a result of long evolution and has evolved with a certain intent, or better put, a purpose. We are purposeful beings, and I will return to this discussion later. But I agree with Jackendoff, that the word model is most appropriate for concept or representation.

Emotions refer to both expressive communications and to internal states related to feelings. Love, hate, courage, fear, joy, sadness, pleasure and disgust can all be described in both psychological and physiological terms. Emotion is the realm where thought and physiology are inextricably entwined, and where the self is inseparable from individual perceptions of value and judgment. Emotions are sometimes regarded as the antithesis of reason, as suggested by phrases such as “appeal to emotion” or “don’t let your emotions take over.” A distinctive and challenging fact about human beings is a potential for both opposition and entanglement between will, emotion and reason. It has also been suggested that there is no empirical support for any generalization suggesting the antithesis between reason and emotion, indeed, anger or fear can often be thought of as a systematic response to observed facts. What should be noted, however, is that the human psyche possesses many possible reactions and perspectives in regard to the internal and external world—often lying on a continuum—some of which may involve the extreme of pure intellectual logic (often called “cold”), other the extreme of pure emotion unresponsive to logical argument (“the heat of passion”). In any case, it should be clear that the relation between logic and argument on the one hand, and emotion on the other, merits careful study. Many have noted that passion, emotion or feeling can add backing to an argument, even one based primarily on reason—particularly in regard to religion or ideology, areas of human thought which frequently demand an all-or-nothing rejection or acceptance, that is, the adoption of a comprehensive worldview partly backed by empirical argument and partly by feeling and passion. Moreover, several researchers have suggested that typically there is no “pure” decision or thought, that is, no thought based “purely” on intellectual logic or “purely” on emotion—most decisions and cognitions are founded on a mixture of both.

An essential role of emotions in the working of the mind was analyzed by many researchers, from various perspectives: philosophical—Rene Descartes (1646)<sup>3</sup>, Immanuel Kant (1790) and Jean Paul Sartre (1948); analytical psychology—Carl Jung (1921); psychological and neural—Stephen Grossberg and Daniel Levine (1987), Andrew Ortony (1990) and Joseph Ledoux (1998); philosophical-linguistic—P. Griffiths (1998); neuro-physiological—Antonio Damasio (1995); and from the learning and cognition perspective by the author (Perlovsky, 1999). Descartes attempted a scientific explanation of passions. He rationalized emotions, explained them as objects and related them to physiological processes. According to Kant, emotions are closely related to judgments, about which individual experiences and perceptions correspond to which general concepts and vice versa. The ability for judgment is a foundation of all higher spiritual abilities, including the beautiful and sublime. Kant's aesthetics has been a foundation of aesthetic theories to this very day (we will continue this discussion later). Sartre equated emotions, to a significant extent, with unconscious contents of psyche; today this does not seem to be adequate. Jung analyzed conscious and unconscious aspects of emotions. He emphasized undifferentiated status of primitive, fused emotion-concept-behavior psychic states in everyday functioning and their role in psychoses. He also emphasized the rational aspect of conscious, differentiated emotions. Ortony explains emotions in terms of knowledge representations and emphasizes abductive logic as a mechanism of inferring other people's emotions. Ledoux analyses neural structures and pathways involved in emotional processing, especially fear. Griffiths considers basic emotions and their evolutionary development within social interactions. According to Damasio, emotions are primarily bodily perceptions, and feelings of emotions in the brain invoke "bodily markers." Grossberg and Levine consider emotions as neural signals that relate instinctual and conceptual brain centers. In processes of perception and cognition, emotions evaluate concept-models of objects and situations for satisfaction or dissatisfaction of instinctual needs. In Section 6, I discuss relationships of these various theories of emotions to mathematical descriptions; here I will just mention that this mathematical description closely corresponds to ideas of Kant, Jung, Grossberg and Levine. Ideas of Sartre and Damasio I did not find detailed enough for mathematical elaboration.

Behavior is comprised of many mechanisms. It is controlled by the endocrine and nervous systems. The complexity of an organism's behavior is related to the complexity of its nervous system. In this chapter I refer only to neurally controlled behavior; it involves mechanisms of negative feedback (e.g., when reaching an object with a hand) and positive feedback (e.g., when making a decision). The first does not reach consciousness, a second, is potentially available to consciousness (Grossberg, 1988).

Even this cursory review of basic notions used for describing the mind illustrates that they are far from being crystal clear; some notions may seem to contradict others. Below I summarize and simplify this discussion of basic mechanisms of the mind and relate them to a mathematical discussion in the next section. Some readers may question my way of summarization and simplification of the huge body of ongoing discussions; therefore, let me repeat that I draw my inspiration in trying to find unifying themes in commonsense understanding and technical discussions from ancient philosophers to today's research in multiple disciplines. The volume of this chapter does not allow for detailed discussions of all points of views. Presented here are summaries and references with few discussions,

and the reader will judge to what extent I succeeded in unifying and simplifying this complex and diverse field.

Explaining basic mind mechanisms, let me repeat, requires no mysterious assumptions, and mathematical descriptions can be developed. Among the mind cognitive mechanisms, the most directly accessible to consciousness are concepts. Concepts are like internal models of the objects and situations in the world. This analogy is quite literal, e.g., during visual perception of an object, a concept-model in our memory projects an image onto the visual cortex, where it is matched to an image projected from the retina (this simplified description will be refined later).

Concepts serve for satisfaction of the basic instincts, which emerged as survival mechanisms long before concepts. We have briefly mentioned current debates on the roles of instincts, reflexes, motivational forces and drives. Inborn, unconscious, less adaptive and more automatic functioning often is referred to as instinctual. This lumping together of various mechanisms is inappropriate for the development of a mathematical description of the mind's mechanisms. I follow proposals (see Grossberg & Levine, 1987, for further references and discussions) to separate instincts as internal sensor mechanisms indicating the basic needs, from "instinctual behavior," which should be described by appropriate mechanisms. Accordingly, I use the word "instincts" to describe mechanisms of internal sensors: For example, when the sugar level in blood goes below a certain level an instinct "tells us" to eat. Such separation of instinct as "internal sensor" from "instinctual behavior" is only a step toward identifying all the details of relevant biological mechanisms.

How do we know about instinctual needs? We do not hear instinctual pronouncements or read dials of instinctual sensors. Instincts are connected to cognition and behavior by emotions. Whereas, in colloquial usage emotions are often understood as facial expressions, higher voice pitch and exaggerated gesticulation, these are outward signs of emotions, serving for communication. A more fundamental role of emotions within the mind's system is that emotional signals evaluate concepts for the purpose of instinct satisfaction. This evaluation is not according to rules or concepts (like rule-systems of artificial intelligence), but according to a different instinctual-emotional mechanism, described first by Grossberg and Levine (1987), and described below for higher cognitive functions. The emotional mechanism is crucial for breaking out of the "vicious circle" of combinatorial complexity.

A mathematical theory described in the next section leads to an inevitable conclusion: Humans and higher animals have a special instinct responsible for cognition. Let me emphasize, this is not an abstract mathematical theorem, but a conclusion from the basic knowledge of the mind's operations as described in thousands of publications. Clearly, humans and animals engage in exploratory behavior, even when basic bodily needs, like eating, are satisfied. Biologists and psychologists discuss curiosity in this regard (Berlyne, 1960; 1973). However, it is not mentioned among "basic instincts" on a par with those for food and procreation. The reasons were that it was difficult to define, and that its fundamental nature was not obvious. The fundamental nature of this mechanism is related to the fact that our knowledge always has to be modified to fit the current situation. One rarely sees exactly the same object: Illumination, angles and surrounding objects are usually different; therefore, adaptation-learning is required. A mathematical formulation of the mind's mechanisms makes obvious the fundamental nature of our desire for

knowledge. In fact, virtually all learning and adaptive algorithms (tens of thousands of publications) maximize correspondence between the algorithm's internal structure (knowledge in a wide sense) and objects of recognition. As discussed in the next section, concept-models that our mind uses for understanding the world are in constant need of adaptation. Knowledge is not just a static state; it involves a process of adaptation and learning. Without adaptation of concept-models, we will not be able to understand the ever-changing surrounding world. We will not be able to orient ourselves or satisfy any of the bodily needs. Therefore, we have an inborn need, a drive, an instinct to improve our knowledge. I call it *the knowledge instinct*. Mathematically, it is described as a maximization of a similarity measure between concept-models and the world (as it is sensed by sensory organs; but I would add that the very sensing is usually adapted and shaped by the knowledge instinct).

Emotions evaluating satisfaction or dissatisfaction of the knowledge instinct are not directly related to bodily needs. Therefore, they are "spiritual," or aesthetic emotions. I would like to emphasize that aesthetic emotions are not peculiar to the perception of art, they are inseparable from every act of perception and cognition. Conceptual-emotional understanding of the world results in actions in the outside world or within the mind. In this chapter we only discuss an internal behavior within the mind, the behavior of learning and understanding the world. In the next section we describe a mathematical theory of conceptual-emotional recognition and understanding. In addition to concepts and emotions, the theory involves mechanisms of intuition, imagination, the conscious and unconscious. This process is intimately connected to an ability of the mind to think and to operate with symbols and signs. The mind involves a hierarchy of multiple layers of concept-models, from simple perceptual elements (like edges or moving dots), to concept-models of objects, to relationships among objects, to complex scenes and upwards along a hierarchy toward the concept-models of the meaning of life and purpose of our existence. Hence, the tremendous complexity of the mind; still, relatively few basic principles go a long way in explaining this system.

I would like to mention that Ortony and Turner (1990) summarized views of fourteen authors on basic emotions; three authors mentioned emotions that I consider aesthetic (Frijda, Izard and McDougall mentioned interest and wonder). One reason for scientific community being slow in adopting these results is the already mentioned cultural bias against emotions as a part of thinking processes. Plato and Aristotle thought that emotions are "bad" for intelligence, and this is a part of our cultural heritage ("you have to be cool to be smart"), and the founders of artificial intelligence repeated this truism (Newell, 1983). Yet, as discussed in the next section, combining conceptual understanding with emotional evaluations is crucial for overcoming the combinatorial complexity and to understanding how the mind works.

I'd like to add a side comment. In neural, cognitive and psychological literature about the mind and brain, one often encounters a statement that the brain is a "kludge," a non-elegant, non-optimal design, a concoction of modules that appeared in evolution first for one purpose, and were then used for a different purpose, etc. (Clark, 1987; Minsky, 1995; Pinker, 1995; Chomsky, 2000). These statements are made usually by non-mathematicians, whose ideas about mathematical optimality and elegance are at best naive (we mention that in this line of research, many considered formal logic as the peak of optimality and elegance, even after Gödel proved its mathematical inconsistency).

Mathematical analysis of evolution demonstrates just the opposite (Perlovsky, 2002); there was more than enough information for evolution to attain optimality. The mind is often optimal (Charness & Levin, 2003). Among those preaching non-optimality of the brain and mind, no one produced a computer program working better or more optimal than the mind. Therefore, it is reasonable to consider mathematically optimal methods for modeling the mind.

## **Modeling Field Theory (MFT)**

---

Modeling field theory is a multi-layer, hetero-hierarchical system (Perlovsky, 2001). The mind is not a strict hierarchy; there are multiple feedback connections among several adjacent layers, hence the term hetero-hierarchy. MFT mathematically implements mechanisms of the mind discussed above. At each layer there are concept-models encapsulating the mind's knowledge; they generate top-down signals, interacting with input, or bottom-up signals. These interactions are governed by the knowledge instinct, which drives concept-model learning, adaptation and formation of new concept-models for better correspondence to the input signals.

This section describes a basic mechanism of interaction between two adjacent hierarchical layers of bottom-up and top-down signals (fields of neural activation); 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, input signals are associated with (or recognized as, or grouped into) concepts according to the models and the knowledge instinct at this layer. These recognized concepts become output signals for the next layer. This general structure of MFT corresponds to our knowledge of neural structures in the brain. This is true about mathematical description in the following sub-sections; however, it is not mapped to specific neurons or synaptic connections. How actual brain neurons "implement" models and the knowledge instinct is a subject for future research. The knowledge instinct is described mathematically as maximization of a similarity measure. In the process of learning and understanding input signals, models are adapted for better representation of the input signals so that similarity between the models and signals increases. This increase in similarity satisfies the knowledge instinct and is felt as aesthetic emotions.

### **The Knowledge Instinct**

---

At a particular hierarchical layer, we index neurons by  $n = 1, \dots, N$ . 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,  $\mathbf{X}(n) = \{X_d(n), d = 1, \dots, D\}$ . Top-down, or priming, signals to these neurons are sent by concept-models,  $\mathbf{M}_h(\mathbf{S}_h, n)$ , indexed by  $h = 1, \dots, H$ . Each model is characterized by its parameters,  $\mathbf{S}_h$ ; in the neuron structure of the brain they are

encoded by strength of synaptic connections; mathematically, we describe them as a set of numbers,  $\mathbf{S}_h = \{S_h^a, a = 1, \dots, A\}$ . Models *represent* signals in the following way. Consider signal  $\mathbf{X}(n)$  coming from sensory neurons activated by object  $h$ , characterized by parameters  $\mathbf{S}_h$ . These parameters may include position, orientation or lighting of an object  $h$ . Model  $\mathbf{M}_h(\mathbf{S}_h, n)$  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 signal  $\mathbf{X}(n)$  from the retina and a priming signal  $\mathbf{M}_h(\mathbf{S}_h, n)$  from an object-concept-model  $h$ . A neuron  $n$  is activated if both the bottom-up signal from lower-layer-input and the top-down priming signal are strong. Various models compete for evidence in the bottom-up signals, while adapting their parameters for a better match as described below. This is a simplified description of perception. The most benign everyday visual perception uses many layers from retina to object perception. The MFT 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. Input signals, models and other parts of the learning mechanisms at a single processing layer described below are illustrated in Figure 1.

Learning is an essential part of perception and cognition, and is driven by the knowledge instinct. It increases a similarity measure between the sets of models and signals,  $L(\{\mathbf{X}\}, \{\mathbf{M}\})$ . The similarity measure is a function of model parameters and associations between the input bottom-up 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.

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

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

This expression contains a product of partial similarities,  $l(\mathbf{X}(n))$ , over all bottom-up signals; therefore it forces the mind to account for every signal (even if only one term in the product is zero, the product is zero, and thus 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 models) for each input neuron signal. Its constituent elements are conditional partial similarities between signal  $\mathbf{X}(n)$  and model  $\mathbf{M}_h$ ,  $l(\mathbf{X}(n)|h)$ . This measure is “conditional” on object  $h$  being present,<sup>4</sup> therefore, when combining these quantities into the overall similarity measure,  $L$ , they are multiplied by  $r(h)$ , which represents the measure of object  $h$  actually being present. Combining these elements with the two principles noted above, a similarity measure is constructed as follows [<sup>5</sup>]:

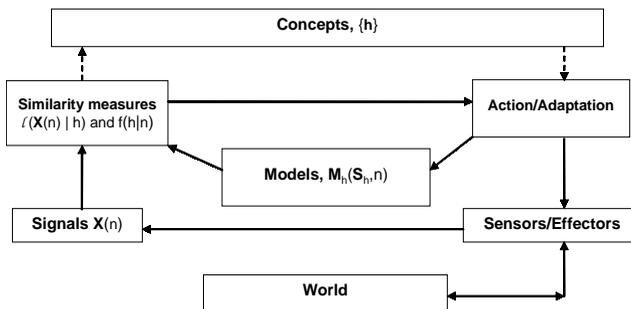
$$L(\{\mathbf{X}\}, \{\mathbf{M}\}) = \prod_{n \in N} \sum_{h \in H} r(h) l(\mathbf{X}(n) | h). \quad (2)$$

Figure 1: For a single layer of MFT, bottom-up input signals are unstructured data  $\{\mathbf{X}(n)\}$  and output signals are recognized or formed concepts  $\{h\}$  with high values of similarity measures. Top-down, “priming” signals are models,  $\mathbf{M}_h(\mathbf{S}_h, n)$ . Conditional similarity measures  $l(\mathbf{X}(n)|h)$  and association variables  $f(n|h)$  (Equation (3)), associate data and models. They initiate adaptation, Equations (4) and (5), and concept recognition (see Equation (14) and discussion there). The adaptation-learning cycle defined by this structure and Equations (3), (4) and (5) maximizes similarity measure (1). Psychologically, it satisfies the knowledge instinct; changes in similarity (1) correspond to aesthetic emotions. New data coming from sensors, if they do not match exactly existing models, reduce similarity value, do not satisfy the knowledge instinct and produce negative aesthetic emotions. This stimulates the constant renewal of adaptation-learning cycles.

The structure of (2) follows standard principles of the probability theory: A summation is taken over alternatives  $h$  and various pieces of evidence  $n$  are multiplied. This expression is not necessarily a probability, but it has a probabilistic structure. If learning is successful, it approximates probabilistic description and leads to near-optimal Bayesian decisions. The name “conditional partial similarity” for  $l(\mathbf{X}(n)|h)$ , or simply  $l(n|h)$ , follows the probabilistic terminology. If learning is successful,  $l(n|h)$  becomes a conditional probability density function, a probabilistic measure that signals in neuron  $n$  originated from object  $h$ . Then  $L$  is the total likelihood of observing signals  $\{\mathbf{X}(n)\}$  coming from objects described by models  $\{\mathbf{M}_h\}$ . Coefficients  $r(h)$ , called priors in probability theory, contain preliminary biases or expectations, expected objects  $h$  have relatively high  $r(h)$  values; their true values are usually unknown and should be learned, like other parameters  $\mathbf{S}_h$ .

Note. In the probability theory, a product of probabilities usually assumes that evidence is independent. Expression (2) contains a product 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  $\mathbf{M}_h(\mathbf{S}_h, n)$  predicts expected signal values in many neurons  $n$ .

Figure 1. Learning mechanisms of a single processing layer



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. In this chapter we consider a case when functional forms of models  $M_h(S_h, n)$  are all fixed, and learning-adaptation involves only model parameters  $S_h$ . More complicated structural learning of models is considered in Perlovsky (2004; 2006). Formation of new concepts and merging or elimination-forgetting of old ones require a modification of the similarity measure (2); 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 (2) using a “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  $p(N, M) = \exp(-N_{\text{par}}/2)$ , where  $N_{\text{par}}$  is a total number of adaptive parameters in all models (this penalty function is known as Akaike Information Criterion, see Perlovsky (1996) for further discussion and references).

## Dynamic Logic

---

The learning process consists in estimating model parameters  $S$  and associating signals with concepts by maximizing the similarity (2). Note, all possible combinations of signals and models are accounted for in expression (2). This can be seen by expanding a sum in (2) and multiplying all the terms; it would result in  $H^N$  items, a huge number. This is the number of combinations between all signals ( $N$ ) and all models ( $H$ ). Here is the source of CC of many algorithms used in the past. For example, multiple hypothesis testing algorithms attempt to maximize similarity  $L$  over model parameters and associations between signals and models in two steps. First they take one of the  $H^N$  items, that is one particular association between signals and models, and maximize it over model parameters. 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  $H^N$ .

Modeling field theory solves this problem by using dynamic logic (Perlovsky, 1996; 2001). 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 is the mechanism of dynamic logic.

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

$$f(h|n) = r(h) l(\mathbf{X}(n)|h) / \sum_{h' \in H} r(h') l(\mathbf{X}(n)|h'). \quad (3)$$

Equation (3) looks like the Bayes formula for *a posteriori* probabilities; if  $l(n|h)$  in the result of learning become conditional likelihoods,  $f(h|n)$  become Bayesian probabilities

for signal  $n$  originating from object  $h$ . The dynamic logic of the Modeling Fields (MF) is defined as follows:

$$\begin{aligned} df(h|n)/dt = f(h|n) \sum_{h' \in H} \{ [\delta_{hh'} - f(h'|n)] \cdot \\ [\partial \ln l(n|h') / \partial \mathbf{M}_{h'}] \partial \mathbf{M}_{h'} / \partial \mathbf{S}_h \cdot d\mathbf{S}_h / dt, \end{aligned} \quad (4)$$

$$d\mathbf{S}_h / dt = \sum_{n \in N} f(h|n) [\partial \ln l(n|h) / \partial \mathbf{M}_h] \partial \mathbf{M}_h / \partial \mathbf{S}_h, \quad (5)$$

here,

$$\delta_{hh'} \text{ is 1 if } h=h', 0 \text{ otherwise.} \quad (6)$$

Parameter  $t$  is the time of the internal dynamics of the MF system (like a number of internal iterations). A more specific form of (5) can be written when Gaussian-shape functions are used for conditional partial similarities:

$$l(n|h) = G(\mathbf{X}(n) | \mathbf{M}_h(\mathbf{S}_h, n), \mathbf{C}_h). \quad (7)$$

Here  $G$  is a Gaussian function with mean  $\mathbf{M}_h$  and covariance matrix  $\mathbf{C}_h$ . Note, a ‘‘Gaussian assumption’’ is often used in statistics; it assumes that signal distribution is Gaussian. This is not the case in (7): Here signal is not assumed to be Gaussian. Equation (7) is valid if *deviations* between the model  $\mathbf{M}$  and signal  $\mathbf{X}$  are Gaussian; these deviations usually are Gaussian. If they are not Gaussian, (7) is still not a limiting assumption: A weighted sum of Gaussians in (2) can approximate any positive function, like similarity. Now the dynamic logic of the MF can be defined as follows:

$$d\mathbf{S}_h^a / dt = [\mathbf{Y}_h^{-1}]^{ab} \mathbf{Z}_h^b, \quad (8)$$

$$\mathbf{Y}_h^{ab} = \sum_{n \in N} f(h|n) [\mathbf{M}_h^{:a} \mathbf{C}_h^{-1} \mathbf{M}_h^{:b}], \quad (9)$$

$$\mathbf{Z}_h^b = \sum_{n \in N} f(h|n) [\mathbf{M}_h^{:b} \mathbf{C}_h^{-1} \mathbf{D}_{nh}], \quad (10)$$

$$d\mathbf{C}_h / dt = -0.5 \mathbf{C}_h^{-2} \sum_{n \in N} f(h|n) [\mathbf{C}_h - \mathbf{D}_{nh} \mathbf{D}_{nh}^T]; \quad (11)$$

$$\mathbf{D}_{nh} = (\mathbf{X}(n) - \mathbf{M}_h), \quad (12)$$

Here, superscript T denotes a transposed row-vector; summation is assumed over repeated indexes a, b; and (;) denotes partial derivatives with respect to parameters S with corresponding indexes:

$$\mathbf{M}_h^{;b} = \partial \mathbf{M}_h / \partial \mathbf{S}_h^b. \quad (13)$$

The following theorem was proven (Perlovsky, 2001):

*Theorem.* Equations (3) through (6) (or (3) and (8 through 12)) define a convergent dynamic MF system with stationary states defined by  $\max\{\mathbf{S}_h\}L$ .

It follows that the stationary states of an MF 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  $\{\mathbf{S}_h\}$  are asymptotically unbiased and efficient estimates of these parameters (Cramer, 1946). A computational complexity of the MF 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. Several aspects of MFT convergence are discussed below (in sections “Example of Dynamic Logic Operations,” “MFT hierarchical organization” and “MFT Dynamics”). If likelihood is used as similarity, parameter values are estimated efficiently (that is, in most cases, parameters cannot be better learned using any other procedure). Moreover, as a part of the above theorem, it is proven that the similarity measure increases at each iteration. The psychological interpretation is that the knowledge instinct is satisfied at each step: A modeling field system with dynamic logic *enjoys* learning.

## **Example of Dynamic Logic Operations**

Finding patterns below noise could be 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 location and orientation 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 these kinds of problems as discussed, multiple hypothesis testing (Singer, Sea, & Housewright, 1974), tries all combinations of subsets and models, faces combinatorial complexity. In this example, we are looking for “smile” and “frown” patterns in noise shown in Figure 2a without noise, and in Figure 2b with noise, as actually measured. Each pattern is characterized by a three-parameter parabolic shape. The image size in this example is 100x100 points, and the true number of patterns is three, which is not known. Therefore, at least four patterns should be fit to the data, to decide that three patterns fit best. Fitting 4x3=12 parameters to a 100x100 grid by brute-force testing would take about  $10^{32}$  to  $10^{40}$  operations, a prohibitive computational complexity.

To apply MFT and dynamic logic to this problem, we need to develop parametric adaptive models of expected patterns. We use a uniform model for noise, Gaussian blobs for highly-fuzzy, poorly resolved patterns and parabolic models for “smiles” and “frowns.” The parabolic models and conditional partial similarities for this case are described in detail in Linnehan et al. (2003). The number of computer operations in this example (without optimization) was about  $10^{10}$ . Thus, a problem that was not solvable due to CC becomes solvable using dynamic logic.

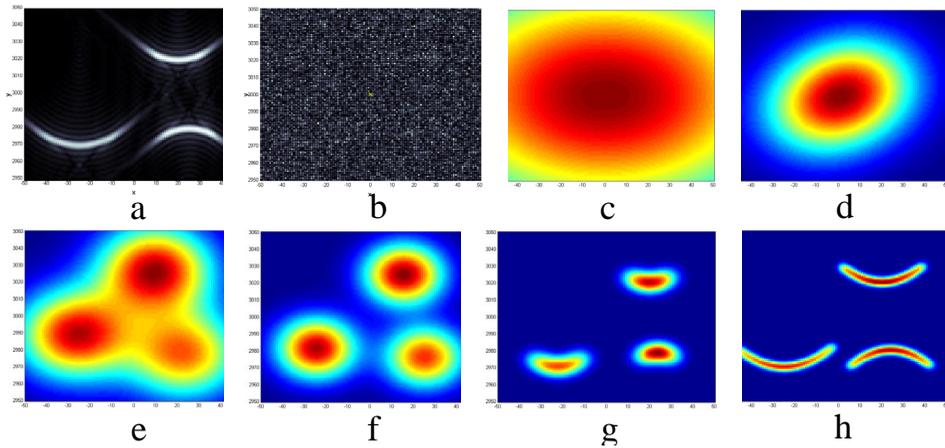
During an adaptation process, initial fuzzy and uncertain models are associated with structures in the input signals, making the fuzzy models more definite and crisp. The type, shape and number of models are selected so that the internal representation within the system is similar to input signals: The MF concept-models represent structure-objects in the signals. The figure below illustrates operations of dynamic logic. In Figure 2 (a) true “smile” and “frown” patterns are shown without noise; (b) actual image available for recognition (signal is below noise, signal-to-noise ratio is between  $-2\text{dB}$  and  $-0.7\text{dB}$ ); (c) an initial fuzzy model, a large fuzziness corresponds to uncertainty of knowledge; (d) through (h) show improved models at various iteration stages (a total of 22 iterations). Every five iterations the algorithm tried to increase or decrease the number of pattern-models. Between iterations (d) and (e), the algorithm decided that it needed three Gaussian models for the “best” fit. There are several types of models: One uniform model describing noise (it is not shown) and a variable number of blob models and parabolic models, in which number, location and curvature are estimated from the data. Until about stage (g), the algorithm used simple blob models; at (g) and beyond, the algorithm decided that it needed more complex parabolic models to describe the data. Iterations stopped at (h), when similarity stopped increasing.

In Figure 2: (a) true “smile” and “frown” patterns are shown without noise; (b) actual image available for recognition (signal is below noise, signal-to-noise ratio is between  $-2\text{dB}$  and  $-0.7\text{dB}$ ); (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). Between stages (d) and (e) the algorithm tried to fit the data with more than one model and decided that it needed three blob-models to “understand” the content of the data. There are several types of models: One uniform model describing noise (it is not shown) and a variable number of blob-models and parabolic models, in which number, location and curvature are estimated from the data. Until about stage (g), the algorithm “thought” in terms of simple blob models; at (g) and beyond, the algorithm decided that it needed more complex parabolic models to describe the data. Iterations stopped at (h), when similarity (2) stopped increasing. This example is discussed in more detail in Linnehan, Mutz, Perlovsky, Weijers, Schindler, and Brockett (2003).

## **MFT Hierarchical Organization**

The previous sub-sections described a single processing layer in a hierarchical MFT system. At each layer of a hierarchy there are input signals from lower layers, models, similarity measures (2), emotions, which are changes in similarity (2), and actions; actions include adaptation, behavior satisfying the knowledge instinct — maximization of similarity, Equations (3) through (6) (or (3) and (8 through 12)). An input to each layer

Figure 2. Finding “smile” and “frown” patterns in noise, an example of dynamic logic operation



is a set of signals  $\mathbf{X}(n)$ , or in neural terminology, an input field of neuronal activations. The result of signal processing at a given layer is activated models, or concepts  $h$  recognized in the input signals  $n$ ; these models, along with the corresponding instinctual signals and emotions, may activate behavioral models and generate behavior at this layer.

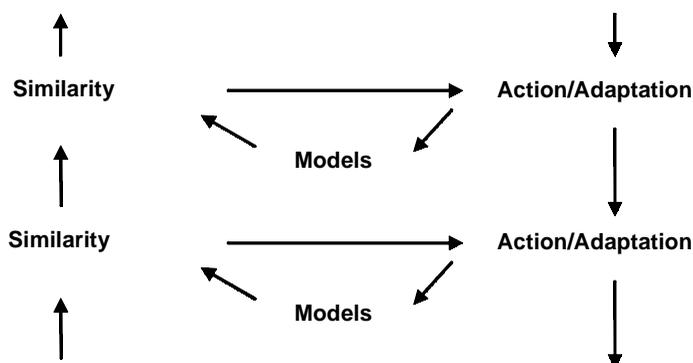
The activated models initiate other actions. They serve as input signals to the next processing layer, where more general concept-models are recognized or created. Output signals from a given layer, serving as input to the next layer, could be model activation signals,  $a_h$ , defined as:

$$a_h = f(h|n). \quad (14)$$

Alternatively, output signals may include model parameters. The Hierarchical MF system is illustrated in Figure 3. Within the hierarchy of the mind, each concept-model finds its “mental” meaning and purpose at a higher layer (in addition to other purposes). For example, consider a concept-model “chair.” It has a “behavioral” purpose of initiating sitting behavior (if sitting is required by the body), this is the “bodily” purpose at the same hierarchical layer. In addition, it has a “purely mental” purpose at a higher layer in the hierarchy, a purpose of helping to recognize a more general concept, say, of a “concert hall,” which model contains rows of chairs.

Models at higher layers in the hierarchy are more general than models at lower layers. For example, at the very bottom of the hierarchy, if we consider vision system, models correspond (roughly speaking) to retinal ganglion cells and perform similar functions; they detect simple features in the visual field; at higher layers, models correspond to functions performed at V1 and higher in the visual cortex, that is, detection of more complex features, such as contrast edges, their directions, elementary moves, etc. Visual

*Figure 3. Hierarchical MF system: At each layer of a hierarchy there are models, similarity measures and actions (including adaptation, i.e., maximizing the knowledge instinct-similarity); high levels of partial similarity measures correspond to concepts recognized at a given layer; concept activations are output signals at this layer and they become input signals to the next layer, propagating knowledge up the hierarchy.*



hierarchical structure and models are studied in detail (Grossberg, 1988; Zeki, 1993); these models can be used in MFT. At still higher cognitive layers, models correspond to objects, to relationships among objects, to situations and relationships among situations, etc. (Perlovsky, 2001). Still higher are even more general models of complex cultural notions and relationships, like family, love, friendship and abstract concepts, like law, rationality, etc. Contents of these models correspond to a cultural wealth of knowledge, including the writings of Shakespeare and Tolstoy; detailed mechanisms of the development of these models are beyond the scope of this chapter (they are addressed in (Perlovsky, 2004; 2006). At the top of the hierarchy of the mind, according to Kantian analysis<sup>6</sup>, are models of the meaning and purpose of our existence, unifying our knowledge, and the corresponding behavioral models aimed at achieving this meaning.

From time to time, as discussed, a system forms a new concept or eliminates an old one. This mechanism works as follows. At every layer, the system always keeps a reserve of fuzzy inactive concept-models (with large covariance,  $C$ , Equation 7). They are inactive in that their parameters are not adapted to the data; therefore their similarities to signals are low. Yet, because of a large fuzziness (covariance) the similarities are not exactly zero. When a new signal does not fit well into any of the active models, its similarities to inactive models automatically increase (because first, every piece of data is accounted for [see footnote 4 for further discussion], and second, inactive models are vague-fuzzy and potentially can “grab” every signal that does not fit into more specific, less fuzzy, active models). As activation signal  $a_n$ , Equation (14), for an inactive model exceeds a certain threshold, the model is activated. Similarly, when an activation signal for a particular model falls below a threshold, the model is deactivated. Thresholds for activation and deactivation are set usually by mechanisms at a higher hierarchical layer based on prior information, system resources, numbers of activated models of various

types, etc. Activation signals for active models at a particular layer  $\{ a_n \}$  form a “neuronal field,” which provides input signals to the next layer, where more abstract and more general concepts are formed, and so on along the hierarchy toward higher models of meaning and purpose.

## Higher Cognitive Functions

---

This section relates in more detail the above mathematical descriptions to higher cognitive functions, illustrating that MF theory is rich enough to describe the mind on one hand without mysticism, and on the other hand, without reductionism, in general agreement with cognitive science, psychology and philosophy. A fundamental role in our higher cognitive functions is played by the knowledge instinct, which we described mathematically as maximization of similarity between concept-models and the world. The mathematical description of higher cognitive functions opens perspectives to better understanding of the mind functioning, and to solving previously unsolved problems associated with higher mind functions, including consciousness, feelings of the sublime, and beauty.

### MFT Dynamics

---

Dynamical equations (3) and (4-6), or (7-12), describe an elementary process of perception or cognition maximizing similarity between models and the world, in which a large number of model-concepts compete for incoming signals; model-concepts are modified and new ones are formed, and eventually, connections are established among signal subsets on the one hand, and model-concepts on the other. Perception refers to processes in which the input signals come from sensory organs and model-concepts correspond to objects in the surrounding world. Cognition refers to higher layers in the hierarchy where the input signals are activation signals from concepts activated at lower layers, whereas model-concepts are more complex, abstract and correspond to situations and relationships among lower-layer concepts.

This process is described by dynamic logic. Its salient mathematical property is a correspondence between uncertainty in models and fuzziness in associations  $f(h|n)$ . During perception, as long as model parameters do not correspond to actual objects, there is no match between models and signals; many models poorly match many objects, and associations remain fuzzy. This can be described more specifically, if Gaussian functions are used for  $l(\mathbf{X}|h)$ : For poorly matched models, the covariances,  $C_h$ , are large (that is, model uncertainties are large). Uncertainty in models prevents  $f(h|n)$  from attaining definite (0,1) values. Eventually, one model ( $h'$ ) wins the competition for a subset  $\{n'\}$  of input signals  $\mathbf{X}(n)$ , when parameter values match object properties.  $C_{h'}$  becomes smaller than other  $C_h$ , and  $f(h'|n)$  values become close to 1 for  $n \in \{n'\}$  and 0 for  $n \notin \{n'\}$ . Upon the convergence, the entire set of input signals  $\{n\}$  is approximately divided into subsets, each associated with one model-object;  $C_{h'}$  becomes small, and fuzzy  $a$

*priori* concepts become crisp concepts. Cognition is different from perception in that models are more general, more abstract and input signals are the activation signals from concepts identified (cognized) at a lower hierarchical layer. The general mathematical laws of cognition and perception are similar in MFT. Let us discuss relationships between the MFT theory and concepts of the mind originated in psychology, philosophy, linguistics, aesthetics, neuro-physiology, neural networks, artificial intelligence, pattern recognition and intelligent systems.

## Elementary Thought-Process

---

Thought-process, or thinking, involves a number of sub-processes and attributes, including internal representations and their manipulation, attention, memory, concept formation, knowledge, generalization, recognition, understanding, meaning, prediction, imagination, intuition, emotion, decisions, reasoning, goals and behavior, conscious and unconscious (Perlovsky, 2001; Grossberg, 1988; Meystel, 1995). A “minimal” subset of these processes has to involve mechanisms for afferent and efferent signals (Grossberg, 1988), in other words, bottom-up and top-down signals coming from outside (external sensor signals) and from inside (internal representation signals). According to Carpenter and Grossberg (1987), every recognition and concept formation process involves a “resonance” between these two types of signals. In MFT, at every layer in a hierarchy, the afferent signals are represented by the input signal field  $\mathbf{X}$ , and the efferent signals are represented by the modeling field signals  $\mathbf{M}h$ ; resonances correspond to high similarity measures  $l(n|h)$  for some subsets of  $\{n\}$  that are “recognized” as concepts (or objects)  $h$ . The mechanism leading to the resonances is given by (3-6), or (7-12), and we call it an elementary thought-process. In this process, subsets of signals corresponding to objects or situations are understood as concepts and signals acquire meaning.

Kant’s three volumes on the theory of the mind, Critique of Pure Reason, Critique of Judgment and Critique of Practical Reason (Kant, 1781; 1790; 1788) describe the structure of the mind similar to MFT. Pure reason, or the faculty of understanding, contains concept-models. The faculty of judgment, or emotions, establishes correspondences between models and data about the world acquired by sensory organs (in Kant’s terminology, between general concepts and individual events). Practical reason contains models of behavior. Kant was first to recognize that emotions are an inseparable part of cognition. The only missing link in Kantian theory is the knowledge instinct. Kant underappreciated a pervading need for concept adaptation; he considered concepts as given *a priori*.

A dynamic aspect of the working of the mind, as given by MFT and dynamic logic, was first given by Aristotle (IV BCE). He described thinking as a learning process in which an *a priori* form-as-potentiality (fuzzy model) meets matter (sensory signals) and becomes a form-as-actuality (a concept). He pointed out an important aspect of dynamic logic: reduction of fuzziness during learning; forms-potentialities are fuzzy (do not obey logic), whereas forms-actualities are logical.

History preserved for us evidence of Aristotle’s foresight. When Alexander the Great, an Aristotelian pupil, was fighting in Persia, he wrote to his teacher: “Aristotle, I heard

you are writing books now. Are you going to make our secret knowledge public?" In a reply letter Aristotle wrote: "Alexander, do not worry: nobody will understand" (Plutarch, II AD).

## Understanding

---

In the elementary thought process, subsets in the incoming signals are associated with recognized model-objects, creating *phenomena* (in the MFT-mind) which are *understood* as objects. In other words, *signal subsets* acquire *meaning*; for example, a subset of retinal signals acquires the meaning of a chair. There are several aspects to understanding and meaning. First, object-models are connected by emotional signals (Perlovsky, 2001; Grossberg & Levine, 1987) to instincts that they might satisfy, and also to behavioral models that can make use of them for instinct satisfaction. Second, an object is understood in the context of a more general situation in the next hierarchical layer, consisting of more general concept-models, which accepts as input-signals the results of object recognition. That is, each recognized object-model (phenomenon) sends (in neural terminology, "activates") an output signal; a set of these signals comprises input signals for the next layer models, which "cognize" more general concept-models, like relations and situations. This process continues up the hierarchy of the mind toward the most general models a system could come up with, such as models of the universe (scientific theories), models of self (psychological concepts), models of the meaning of existence (philosophical concepts) and models of *a priori* transcendent intelligent subject (theological concepts).

## Conscious and Unconscious

---

Why is there consciousness? Why would a feature like consciousness appear in the process of evolution? The answer to this question seems clear: Consciousness directs the will and results in a better adaptation for survival. In simple situations, when only minimal adaptation is required, instinct alone is sufficient, and unconscious processes can efficiently allocate resources and will. However, in complex situations, when adaptation is complicated, various instincts might contradict one another. Undifferentiated unconscious psychic functions result in ambivalence and ambitendency; every position entails its own negation, leading to an inhibition. This inhibition cannot be resolved by an unconscious that does not differentiate among alternatives. Direction is impossible without differentiation. Consciousness is needed to resolve an instinctual impasse by suppressing some processes and allocating power to others. By differentiating alternatives, consciousness can direct a psychological function to a goal.

Totality and undividedness of consciousness are the most important adaptive properties needed to concentrate power on the most important goal at every moment. This is illustrated, for example, by clinical cases of divided consciousness and multiple personalities, resulting in maladaptation up to a complete loss of functionality. Simple consciousness needs only to operate with relatively few concepts. One needs more and more differentiation for selecting more and more specific goals. The scientific quest is to

explain the emergence of consciousness from the unconscious in the process of evolution. Consciousness has emerged, driven by unconscious urges for improved adaptation, by the knowledge instinct. And, among goals of consciousness is improvement of understanding of what is not conscious, inside and outside of the psyche. Thus, the cause and the end of consciousness are unconscious; hence the limitations of consciousness, its causal mechanisms and goals are in the unconscious.

Most of our organismic functioning, like breathing, digestion, etc. are unconscious. In the process of evolution, only gradually have psychic processes separated from other organismic functions. In psychic functioning, our evolutionary and personal goals are to increase consciousness. But, this is largely unconscious, because our direct knowledge of ourselves is limited to consciousness. This fact creates a lot of confusion about consciousness. So, what is consciousness?

Consciousness is an awareness or perception of inward psychological facts, a subjective experience of sensing, feelings or thoughts. This definition is taken from the Webster's *Dictionary*. But a more detailed, scientific analysis of consciousness has proven to be difficult. For a long time it seemed obvious that consciousness completely pervades our entire mental life, or at least its main aspects. Now, we know that this idea is wrong, and the main reason for this misconception has been analyzed and understood: We are conscious only about what we are conscious of, and it is extremely difficult to notice anything else.

Popular misconceptions about consciousness noted by Jaynes (1976) include: Consciousness is nothing but a property of matter, or a property of living things or a property of neural systems. These three "explanations" attempted to dismiss consciousness as an epiphenomenon, an unimportant quality of something else. They are useless because the problem is in *explaining* the relationships of consciousness to matter, to life and to neural systems. These dismissals of consciousness are not very different from saying that there is no consciousness; but, of course, this statement refutes itself (if somebody makes such a statement unconsciously, there is no point of discussing it). A dualistic position is that consciousness belongs to the world of ideas and has nothing to do with the world of matter. But the scientific problem *is* in explaining the consciousness as a natural-science phenomenon; that is, to relate consciousness and the material world. Searle (1992) suggested that any explanation of consciousness has to account for it being real and based on physical mechanisms in the brain. Among properties of consciousness requiring explanation, he listed unity and intentionality (we perceive our consciousness as being *unified* in the space of our perceptions and in the time of our life; consciousness is about something; this "about" points to its *intentionality*).

Searle (1997) reviews recent attempts to explain consciousness, and comes to a conclusion that little progress was made during the 1990s. Penrose (1994) suggested that consciousness cannot be explained by known physical laws of matter. His arguments descend from the Gödel's proofs of inconsistency and incompleteness of logic. This, however, only proves (Perlovsky, 1996) that the mind is not a system of logical rules, which we have already discussed in previous sections.

Knowledge of consciousness is primarily of introspective origin. Understanding of consciousness requires differentiating conscious and unconscious psychic processes, so we need to understand what is psychic, what is unconscious and what is conscious-

ness. Our experiences can be divided into somatic and psychic. A will modifying instinctual reflexes indicates a presence of psyche, but not necessarily consciousness. Often, we associate consciousness with a subjective perception of free will. Consciousness about somatic experiences is limited by the unknown in the outer world. Similarly, consciousness about psychic experiences is limited by *the unknown in the psyche, or unconscious*. Roughly speaking, there are three conscious/unconscious levels of psychic contents: (1) contents that can be recalled and made conscious voluntarily (memories); (2) contents that are not under voluntary control; we know about them because they spontaneously irrupt into consciousness; and (3) contents inaccessible to consciousness. We know about the latter through scientific deductions.

Consciousness is not a simple phenomenon, but a complicated differentiated process. Jung (1921) differentiated four types of consciousness related to experiences of feelings, thoughts, sensations and intuitions. In addition to these four psychic functions, consciousness is characterized by attitude: introverted, concentrated mainly on the inner experience, or extroverted, concentrated mainly on the outer experience. Interplay of various conscious and unconscious levels of psychic functions and attitudes results in a number of types of consciousness; interactions of these types with individual memories and experiences make consciousness dependent on the entire individual experience producing variability among individuals.

*Intentionality* is a property of referring to something else, and consciousness is about something. This “aboutness” many philosophers refer to as intentionality. In everyday life, when we hear an opinion, we do not just collate it in our memory and relate to other opinions (like a pseudo-scientist in a comedy), this would not lead very far. We wish to know what are the aims and intentions associated with this opinion. Often, we perceive the intent of what is said better than specific words, even if the words are chosen to disguise the intent behind causal reasoning. The desire to know and the ability to perceive the goal indicate that in psyche, *final standpoint or purpose* is more important than the *causal* one. This intentionality of psyche was already emphasized by Aristotle (VI BCE) in his discussions of the end cause of forms of the mind. Intentionality of consciousness is more fundamental than “aboutness,” it is *purposiveness*.<sup>7</sup>

The intentional property of consciousness led many philosophers during the last decades to believe that intentionality is a unique and most important characteristic of consciousness: According to Searle, only conscious beings could be intentional. But, we feel that this view is not adequate. Intentionality is a fundamental property of life; even the simplest living being is a result of long evolution, and its every component, say a gene or a protein, has a purpose and intent. In particular, every model-concept has evolved with an intent or purpose to recognize a particular type of signal (event, message or concept) and to act accordingly (e.g., send recognition messages to other parts of the brain and to behavioral models). Aristotle was the first to explain the intentionality of the mind this way; he argued that intentionality should be explained through the *a priori* contents of the mind.<sup>8</sup>

Even in an artificial intelligent systems, every part is intentional; they were designed and built with intent to accomplish something. Every concept-model is intentional; the intent is to recognize an object or a situation. As discussed previously, objects that we see around us belong not to the outer world of matter, but to the world of concepts, the realm of interaction between the mind and matter. Thus, every object is an intentional concept.

It is important to differentiate this statement from a philosophical position of *pan-psychism*, which assumes that matter itself, in all its forms, has a degree of psyche or spirit as its fundamental property. Pan-psychism does not really explain matter or psyche. This is why Descartes “exorcised” spirit from the world of matter. To a significant degree, pan-psychism is a result of a failure to differentiate between the world of matter and the world of concepts.

This analysis of intentionality works for cultural concepts as well. Every cultural concept and every man-made object are intentional because they emerged, or were created (consciously or otherwise), with a specific intent (or purpose). The intentionality, I repeat, is therefore the same property that Kant called purposiveness.<sup>9</sup> There are two aspects of purposiveness or intentionality: Higher intellectual intention of a concept is to correspond to the world and thus to satisfy the knowledge instinct; and lower bodily intention is to be used for appropriate utilitarian or bodily-instinctive purposes. (For example, a table in my kitchen is not just a thing-in-itself, but an intentional concept-object; its higher intellectual intention is to recognize the table-as-a-part-of-material-world and use it for building a coherent picture of the world in my mind, and its lower bodily intention is to use the table-object appropriately for sitting and eating, etc. In this regard, some philosophers (Freeman, 2000) talk about the table as an *external representation* of the concept “table”).

Is there any specific relationship between consciousness and intentionality? If so, it is just the opposite of Searle’s hypothesis of intentionality implying consciousness. Affective, subconscious lower-bodily-level emotional responses are concerned with immediate survival, utilitarian goals, and therefore are *intentional in the most* straightforward way. A higher-intellectual-level consciousness is not concerned with immediate survival, but with the overall understanding of the world, with knowledge and beauty; it can afford to be impartial, abstract and less immediately-intentional than the rest of the psyche. Its intentions might be directed toward meanings and purposes of life. The highest creative aspect of individual consciousness and the abilities of perceiving the beautiful and sublime are intentional without any specific, lower-level utilitarian goal. They are intentional toward self-realization, toward future-self beyond current-self.

*Unity of consciousness* refers to conscious mental states being parts of a unified sequence, and simultaneous conscious events are perceived as unified into a coherent picture. Searle’s unity is close to what Kant called “the transcendental unity of apperception.” In MFT, this internal perception is explained, as all perceptions, due to a property of the special model involved in consciousness, called Ego by psychologists. The properties of the Ego-model explain the properties of consciousness. When certain properties of consciousness seems difficult to explain, we should follow the example of Kant; we should turn the question around and ask: Which properties of Ego model would explain the phenomenological properties of consciousness?

Let us begin the analysis of the structures of the Ego-model and the process of its adaptation to the constantly changing world, from evolutionary-preceding simpler forms. What is the initial state of consciousness? Is it an undifferentiated unity or a “booming, buzzing confusion”? Or, let us take a step back in the evolutionary development and ask, What is the initial state of pre-conscious psyche? Or, let us move back even farther toward evolution of sensory systems and perception. When building a robot for a factory floor, why provide it with a sensor? Obviously, such an expensive thing as

a sensor is needed to achieve specific goals: to sense the environment with the purpose to accomplish specific tasks. Providing a robot with a sensor goes together with an ability to utilize sensory data. (Why have sensors otherwise?)

Similarly, in the process of evolution, sensory abilities emerge together with perception abilities. A natural evolution of sensory abilities can not result in a “booming, buzzing confusion,” but must result in evolutionary advantageous abilities to avoid danger, attain food, etc. Initial perception abilities are *limited* to a few types of concept-objects (light-dark, warm-cold, edible-nonedible, dangerous-attractive, etc.) and are directly “wired” to proper actions. When perception functions evolve further, beyond immediate actions, it is through the development of complex internal model-concepts, which unify simpler object-models into a unified and flexible model of the world. Only at this point of possessing relatively complicated differentiated concept-models composed of a large number of sub-models, can an intelligent system experience a “booming, buzzing confusion,” if it faces a new type of environment. A primitive system is simply incapable of perceiving confusion: It perceives only those “things” for which it has concept-models, and if its perceptions do not correspond to reality, it just does not survive without experiencing confusion. When a baby is born, it undergoes a tremendous change of environment, most likely without much conscious confusion. The original state of consciousness is undifferentiated unity. It possesses a single modality of primordial undifferentiated Self-World.

The initial unity of psyche limited abilities of the mind, and further development proceeded through differentiation of psychic functions or modalities (concepts, emotions and behavior); they were further differentiated into multiple concept-models, etc. This accelerated adaptation. Differentiation of consciousness is a relatively recent process (Jaynes, 1976; Jung, 1921).

Consciousness is about aspects of concept-models (of the environment, self, past, present, future plans and alternatives) and emotions (evaluative feelings)<sup>10</sup> to which we can direct our attention. As already mentioned, MFT explains consciousness as a specialized Ego-model. Within this model, consciousness can direct attention at will. This conscious control of will is called the free will. A subjective feeling of free will is a most cherished property of our psyche. Most of us feel that this is what makes us different from inanimate objects and simple forms of life. And this property is a most difficult one to explain rationally or to describe mathematically. But, let us see how far we can go towards understanding this phenomenon. We know that raw percepts are often not conscious. For example, in the visual system, we are conscious about the final processing stage and the integrated crisp model but unconscious about intermediate processing. We are unconscious about eye receptive fields: about details of visual perception of motion and color as far as it takes place in our brain separately from the main visual cortex, etc. These unconscious perceptions are illustrated in blindsight: A visual perception occurs, but a person is not conscious of it (Zeki, 1993). In most cases, we are conscious only about the integrated scene, crisp objects, etc.

These properties of consciousness follow from properties of concept-models; they have conscious (crisp) and unconscious (fuzzy) parts, which are accessible and inaccessible to consciousness, that is to the Ego-model. In pre-scientific literature about mechanisms of the mind, there was a popular idea of homunculus, a little mind inside our mind, which perceived our perceptions and made them available to our mind. This naive view is

amazingly close to actual scientific explanation. The fundamental difference is that the scientific explanation does not need an infinite chain of homunculi inside homunculi. Instead, there are hierarchy of the mind models with their conscious and unconscious aspects. The higher in the hierarchy, the less is the conscious differentiated aspect of the models, until at the top of the hierarchy there are mostly unconscious models of the meaning of our existence (which we discuss later).

Our internal perceptions of consciousness, let me repeat, are due to the Ego-model “perceiving” crisp conscious parts of other models similar to models of perception “perceiving” objects in the world. The properties of consciousness as we perceive them, such as continuity and identity of consciousness, are due to properties of the Ego-model. What is known about this “consciousness” model? Since Freud, a certain complex of psychological functions was called Ego. Jung considered Ego to be based on a more general model or archetype of Self. Jungian archetypes are psychic structures (models) of a primordial origin, which are mostly inaccessible to consciousness, but determine the structure of our psyche. In this way, archetypes are similar to other models, e.g., receptive fields of the retina are not consciously perceived, but determine the structure of visual perception. The Self archetype determines our phenomenological subjective perception of ourselves and, in addition, structures our psyche in many different ways which are far from being completely understood. An important phenomenological property of Self is the perception of uniqueness and indivisibility (hence, the word *individual*).

Consciousness, to a significant extent, coincides with the conscious part of the archetype-model of Self. A conscious part of Self belongs to Ego. Not everything within Ego (as defined by Freud) is conscious. Individuality, as a total character distinguishing an individual from others, is a main characteristic of Ego. Not all aspects of individuality are conscious, so the relationships among the discussed models can be summarized to some extent, as:

$$\text{Consciousness} \in \text{Individuality} \in \text{Ego} \in \text{Self} \in \text{Psyche}.$$

The sign “ $\in$ ” here means “is a part of.” Consciousness-model is a subject of free will; it possesses, controls and directs free will. Free will is limited by laws of nature in the outer world and in the inner world by the unconscious aspects of Self. Free will belongs to consciousness, but not to the conscious and unconscious totality of the psyche.

Many contemporary philosophers consider subjective nature of consciousness to be an impenetrable barrier to scientific investigation. Chalmers differentiated hard and easy questions about consciousness as follows. Easy questions, that will be answered better and better, are concerned with brain mechanisms: Which brain structures are responsible for consciousness? Hard questions, that no progress can be expected for, are concerned with the subjective nature of consciousness and *qualia*, subjective feelings associated with every conscious perception. Nagel (1974) described it dramatically with a question: “What is it like to be a bat?” But I disagree. I don’t think these questions are hard. These questions are not mysteries; they are just wrong questions for a scientific theory. Newton, while describing the laws of planet motion, did not ask: “What is it like to be a planet?” (even so, something like this feeling *is* a part of scientific intuition). The subjective nature of consciousness is not a mystery. It is explained due to the subjective

nature of the concept-models that we are conscious of. The subjectivity is the result of combined apriority and adaptivity of the consciousness-model, the unique genetic *a priori* structures of psyche together with our unique individual experiences. I consider the only hard questions about consciousness to be *free will and the nature of creativity*.

Let us summarize. Most of the mind's operations are not accessible to consciousness. We definitely know that neural firings and connections cannot be perceived consciously. In the foundations of the mind there are material processes in the brain inaccessible to consciousness. Jung suggested that conscious concepts are developed by the mind based on genetically inherited structures or archetypes, which are inaccessible to consciousness (Jung, 1921; 1934). Grossberg (1988) suggested that only signals and models attaining a resonant state (that is signals matching models) can reach consciousness. This was further detailed by Taylor (2005); he related consciousness to the mind being a control mechanism of the mind and body. A part of this mechanism is a prediction model. When this model's predictions differ from sensory observations, the difference may reach a resonant state, which we are consciousness about. To summarize the above analyses, the mind mechanisms, described in MFT by dynamic logic and fuzzy models, are not accessible to consciousness. Final results of dynamic logic processes, resonant states characterized by crisp models and corresponding signals are accessible to consciousness.

## **Imagination**

---

Imagination involves excitation of a neural pattern in a sensory cortex in the absence of an actual sensory stimulation. For example, visual imagination involves excitation of visual cortex, say with closed eyes (Grossberg, 1988; Zeki, 1993). Imagination was long considered a part of thinking processes; Kant (1790) 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 the recent past would not know how to relate to this (Newell, 1983; Minsky, 1988; Carpenter and Grossberg, 1987) resonance model and the MFT dynamics, 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 more readily excited). In MFT, the imagined neural patterns are given by models  $M_h$ .

Visual imagination, as mentioned, can be “internally perceived” with closed eyes. The same process can be mathematically modeled at higher cognitive layers, where it involves models of complex situations or plans. Similarly, models of behavior at higher layers of the hierarchy can be activated without actually propagating their output signals down to actual muscle movements and 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. Such a mathematical procedure based on MFT and dynamic logic was implemented for several applications (some of them were described in Perlovsky, 2001; other applications for Internet search engines, military applications and financial predictions remain unpublished). It is our suggestion that this mathematical description corresponds to actual workings of the mind. At this point, I would only say that this

suggestion does not contradict known facts from psychology and neuro-biology. Current research relating this mathematical description to psychology of high cognitive behavior, to its evolution, and to brain regions supporting this functioning is going on in collaboration with other researchers (D. Levine and F. Fontanari). Sometimes, imagination involves detailed alternative courses of actions considered and evaluated consciously. Sometimes, imagination may involve fuzzy or vague, barely conscious models, which rich 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 one possible; conscious evaluation cannot involve all possible courses of action, it would lead to combinatorial complexity and impasse. It remains to be proven in brain studies, which will identify the exact brain regions and neural mechanisms involved.

MFT (in agreement with neural data) just adds details to Kantian description: Thinking is a play of top-down *higher-hierarchical-layer* imagination and bottom-up *lower-layer* understanding. Kant identified this “play” [described by (3-6) or (7-12)] as a source of aesthetic emotion. Kant used the word “play,” when he was uncertain about the exact mechanism; this mechanism, according to our suggestion, is the knowledge instinct and dynamic logic.

## **Instincts and Emotions**

---

Functioning of the mind and brain cannot be understood in isolation from the system’s “bodily needs.” For example, a biological system (and any autonomous system) needs to replenish its energy resources (eat). This and other fundamental unconditional needs are indicated to the system by instincts. As we discussed, scientific terminology in this area is still evolving; for our purpose of making a step toward uncovering neural mechanisms of the mind, we describe instincts mathematically as internal sensors indicating unconditional needs of an organism. Emotional signals, generated, say, by instinct for food are perceived by psyche as “hunger,” and they activate behavioral models related to food searching and eating. In this chapter we are concerned primarily with the behavior of understanding the surrounding world, with acquiring knowledge. The knowledge instinct demands that we improve our knowledge; the corresponding “internal sensor” is a measure of similarity between the knowledge (internal models of the mind) and the world around us (that we sense through sensory organs). Bodily instinctual influences on understanding modify the object-perception process (3) - (6) in such a way that desired objects get enhanced recognition. This is the reason a hungry person “sees food all around.” In MFT it can be accomplished by modifying priors,  $r(h)$  in Equations (2) and (3), according to the degree to which an object of type  $h$  can satisfy a particular instinct. Details of these mechanisms are not considered here.

## **Aesthetic Emotions and the Instinct for Knowledge**

---

Recognizing objects in the environment and understanding their meaning is so important for survival that a special instinct evolved for this purpose. This instinct for learning and

improving concept-models I call the instinct for knowledge. In MFT it is described by maximization of similarity between the models and the world, Equation (2). Emotions related to satisfaction-dissatisfaction of this instinct are perceived by us as harmony-disharmony (between our understanding of how things ought to be and how they actually are in the surrounding world). According to Kant (1790), these are aesthetic emotions (emotions that are not related directly to satisfaction or dissatisfaction of bodily needs).

The instinct for knowledge makes little kids, cubs and piglets jump around and play fight; their inborn models of behavior must adapt to their body weights, objects and animals around them long before the instincts of hunger and fear will use the models for direct aims of survival. Kiddy behavior just makes the work of the knowledge instinct more observable; to varying degrees, this instinct continues acting all of our life. All the time we are bringing our internal models into correspondence with the world; in adult life, when our perception and understanding of the surrounding world is adequate, aesthetic emotions are barely perceptible: The mind just does its job. Similarly, we do not usually notice adequate performance of our breathing muscles and satisfaction of the breathing instinct. However, if breathing is difficult, negative emotions immediately reach consciousness. The same is true about the knowledge instinct and aesthetic emotions: If we do not understand the surroundings, if objects around do not correspond to our expectations, negative emotions immediately reach consciousness. We perceive these emotions as disharmony between our knowledge and the world. Thriller movies exploit the instinct for knowledge: They are mainly based on violating our expectations; their personages are shown in situations where knowledge of the world is inadequate for survival.

Let me emphasize again, aesthetic emotions are not peculiar to art and artists, they are inseparable from every act of perception and cognition. In everyday life we usually do not notice them. Aesthetic emotions become noticeable at higher cognitive layers in the mind's hierarchy, when cognition is not automatic, but requires conscious effort. Damasio's (1995) view of emotions defined by visceral mechanisms, as far as discussing higher cognitive functions, seems erroneous in taking secondary effects for the primary mechanisms. People often devote their spare time to increasing their knowledge, even if it is not related to their job and a possibility of promotion. Pragmatic interests could be involved: Knowledge makes us more attractive to friends and could help find sexual partners. Still, there is a remainder, a pure joy of knowledge: aesthetic emotions satisfying the knowledge instinct.

## **Beautiful and Sublime**

---

Contemporary cognitive science is at a complete loss when trying to explain the highest human abilities, the most important and cherished abilities to create and perceive beautiful and sublime experiences. Their role in the working of the mind is not understood. MFT explains that 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 models at "higher" levels of the mind hierarchy. The highest forms of aesthetic emotion are related to the most general and most important

models near the top of the hierarchy. According to Kantian analysis (1790; 1798), among the highest models are those of the meaning of our existence, of our purposiveness or intentionality; beauty is related to improving these models.

Models of our purposiveness are largely fuzzy and unconscious. Some people, at some points in their life, may believe that their life purpose to be finite and concrete. Say, 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 achieved a finite goal of making money or raising good children knows that this is not the end of his or her aspirations. The reason is that everyone has an ineffable feeling of partaking in the infinite, while at the same time knowing that our 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 fuzzy and partly unconscious.

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 book. These issues are not new; philosophers and theologians have expounded upon them from time immemorial. The 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. The MFT explanation of the nature of beautiful resolves a number of mysteries and contradictions in contemporary aesthetics (Perlovsky, 2002; 2006).

The feeling of spiritual sublimity is similar and different from the beautiful. Whereas the beautiful is related to improvement of the models of *cognition*, the sublime is related to improvement of the models of *behavior* realizing the highest meaning in our life. The beautiful and sublime are not finite. MFT tells us that, mathematically, improvement of complex models is related to choices from an infinite number of possibilities. A mathematician may consider  $100^{100}$ , or million power million as a finite number. But for a physicist, a number that exceeds all elementary events in the life of the Universe is infinite. A choice from infinity is infinitely complex and contains infinite information. Therefore, choices of the beautiful and sublime contain infinite information. This is not a metaphor, but exact mathematical fact. Beauty is at once objective and subjective. It really exists; cultures and individuals cannot exist without an ability to appreciate beauty, and still, it cannot be described by any finite algorithm or a set of rules.

Beauty of a physical theory, discussed sometimes by physicists, is similar in its infinity to beauty in an artwork. For a physicist, beauty of a physical theory is related to a need to improve the models of the meaning in our understanding of the universe. This satisfies a scientist's quest for purpose, which he identifies with purpose in the world.

## **Intuition**

---

Intuitions include inner perceptions of object-models, imaginations produced by them, and their relationships with objects in the world. They include also higher-level models

of relationships among simpler models. Intuitions involve fuzzy unconscious concept-models, which are in a state of being formed, learned and adapted toward crisp and conscious models (say, a theory). Conceptual contents of fuzzy models are undifferentiated and partly unconscious. Similarly, conceptual and emotional contents of these fuzzy mind states are undifferentiated; concepts and emotions are mixed up. Fuzzy mind states may satisfy or dissatisfy the knowledge instinct in varying degrees before they become differentiated and accessible to consciousness, hence the vague and 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. Composer's intuitions are mostly about sounds and their relationships to psyche. Painter's intuitions are mostly about colors and shapes and their relationships to psyche. Writer's intuitions are about words, or more generally, about language and its relationships to psyche. Mathematical intuition is about structure and consistency within a theory, and about relationships between the theory and *a priori* content of psyche. Physical intuition is about the real world, first principles of its organization and mathematics describing it.

## **Creativity, Differentiation and Synthesis**

---

Creativity is an ability to improve and create new model-concepts. In a small degree it is present in everyday perception and cognition. Usually the words "creativity," "creative" or "discovery" are applied to improving or creating new model-concepts at higher cognitive levels, concepts that are important for the entire society or culture. Making one's discoveries well-known to the entire society, recognized and available to the entire culture is a separate process. We know, for example, that skillful and expensive advertising could be important, but this is not a part of the original discovery or creation. Certain discoveries and works of art became famous centuries after their creation.

MFT explains that a completely crisp model could only match a very specific content; therefore, it cannot lead to the creation of new contents. Creativity and discovery, according to MFT, involve vague, fuzzy models, which are made more crisp and clear. It occurs, therefore, at the border between consciousness and unconscious. A similar nature of creative process, involving consciousness and unconscious, was discussed by Jung (1921). Creativity usually involves intuition, as discussed above: fuzzy undifferentiated feelings-concepts.

Two main mechanisms of creativity are differentiation and synthesis. Differentiation is a process of creating new, more specific and more detailed concept-models from simpler, less differentiated and less conscious models. Mathematical mechanisms of differentiation were discussed in earlier.

Synthesis is a process of connecting detailed, crisp concept-models to unconscious instincts and emotions. The need for synthesis comes from the fact that most of our concept-models are acquired from language. The entire conceptual content of the culture is transmitted from generation to generation through language; cognitive concept-models can not be transmitted directly from brain to brain. Therefore, concepts acquired from language have to be transformed by individual minds into cognitive concepts. Mathematical mechanisms of integrating cognition and language require the extension of MFT considered in Perlovsky (2004). It explains that purely-language concepts could

be detailed and conscious, but not necessarily connected to cognitive concept-models, emotions and to the knowledge instinct. Every child acquires language between one and seven years, but it takes the rest of their life to connect abstract language models to their life's needs, to cognitive concept-models, to emotions and instincts. This is the process of synthesis; it integrates language and cognition, concepts and emotions, conscious and unconscious, instinctual and learned into a unified whole.

Differentiated concepts acquire meaning in connections with instinctual and unconscious. In the evolution of the mind, differentiation speeds up the development of consciousness, but may bring about a split between conscious and unconscious and between emotional and conceptual. If the split affects collective psyche, it leads to a loss of the creative potential of a nation. This was the mechanism of death of great ancient civilizations. The development of culture, the very interest of life, requires *combining differentiation and synthesis*. Evolution of the mind and cultures is determined by this complex, non-linear interaction: One factor prevails, then another. Evolution of cultures as determined by the interaction of mechanisms of differentiation and synthesis is considered in more detail in Perlovsky (2006).

## **Mind, Brain and MFT Testing**

---

Historically, the mind is described in psychological and philosophical terms, whereas the brain is described in terms of neurobiology and medicine. Within scientific exploration, the mind and brain are different description levels of the same system. Establishing relationships between these descriptions is of great scientific interest. Today we approach solutions to this challenge (Grossberg, 2000), which eluded Newton in his attempt to establish a physics of “spiritual substance” (Westfall, 1983). Detailed discussion of established relationships between the mind and brain is beyond the scope of this chapter, as well as relating in detail MFT to brain mechanisms is a subject of ongoing and future research. In this section we briefly mention the main known and unknown facts and give references for future reading. General neural mechanisms of the elementary thought process which are similar in MFT and ART (Carpenter & Grossberg, 1987), have been confirmed by neural and psychological experiments. These include neural mechanisms for bottom-up (sensory) signals, top-down imagination model-signals and the resonant matching between the two (Grossberg, 1975). Adaptive modeling abilities are well studied with adaptive parameters identified with synaptic connections (Koch & Segev, 1998; Hebb, 1949); instinctual learning mechanisms have been studied in psychology and linguistics (Piaget, 2000; Chomsky, 1981; Jackendoff, 2002; Deacon, 1998).

Ongoing and future research will confirm, disprove or suggest modifications to specific mechanisms considered: model parameterization and parameter adaptation, reduction of fuzziness during learning, similarity measure (2) as a foundation of the knowledge instinct and aesthetic emotion, relationships between psychological and neural mechanisms of learning on the one hand and, on the other, aesthetic feelings of harmony and emotion of beautiful. Specific neural systems will have to be related to mathematical descriptions on the one hand, and on the other, to psychological descriptions in terms of subjective

experiences and observable behavior. Ongoing joint research with Fontanari addresses the evolution of models jointly with that of language (Fontanari & Perlovsky, 2005a;b; 2004); joint research with Levine addresses relationships of MFT and the knowledge instinct to issues of behavioral psychology and to specific brain areas involved in emotional reward and punishment during learning. We will have to develop differentiated forms of the knowledge instinct, which are the mechanisms of synthesis, and of the infinite variety emotions perceived in music (Perlovsky, 2006). Future experimental research needs to study in detail the nature of hierarchical interactions: To what extent is the hierarchy “hardwired” versus adaptively emerging? What is a hierarchy of learning instinct? We will have to develop further interactions between cognitive hierarchy and language hierarchy (Perlovsky, 2004).

## **Teleology, Causality and the Knowledge Instinct**

---

Teleology explains the universe in terms of purposes. In many religious teachings it is a basic argument for the existence of God: If there is purpose, an ultimate designer must exist. Therefore, teleology is a hot point of debates between creationists and evolutionists: Is there a purpose in the world? Evolutionists assume that the only explanation is causal. Newton’s laws gave a perfect causal explanation for the motion of planets: A planet moves from moment to moment under the influence of a gravitational force. Similarly, today’s science explains motions of all particles and fields according to causal laws, and there are exact mathematical expressions for fields, forces and their motions. Causality explains what happens in the next moment as a result of forces acting in the previous moment. Scientists accept this causal explanation and oppose teleological explanations in terms of purposes. The very basis of science, it seems, is on the side of causality, and religion is on the side of teleology.

This assumption, however, is wrong. The contradiction between causality and teleology does not exist at the very basic level of fundamental physics. The laws of physics, from classical Newtonian laws to quantum superstrings, can be formulated equally as causal or as teleological. An example of teleological principle in physics is energy minimization, i.e., particles moving so that energy is minimized. It is as if particles in each moment know their purpose: to minimize the energy. The most general physical laws are formulated as minimization of action. Action is a more general physical entity than energy; it is an intuitive name for a mathematical expression called Lagrangian. Causal dynamics, motions of particles, quantum strings and superstrings are determined by minimizing Lagrangian-action (Feynman & Hibbs, 1965). A particle under force moves from point to point as if it knows its final purpose, to minimize Lagrangian-action. Causal dynamics and teleology are two sides of the same coin.

The knowledge instinct is similar to these most general physical laws: Evolution of the mind is guided by maximization of knowledge. A mathematical structure of similarity (2), or its continuous version, is similar to Lagrangian and plays a similar role: It bridges causal dynamic logic of cognition and teleological principle of maximum knowledge. Similar to fundamental physics, dynamics and teleology are equivalent. Dynamic logic follows from maximization of knowledge and vice versa. Ideas and concept-models change under the “force” of dynamic logic, as if they know the purpose: maximum

knowledge. One does not have to choose between scientific explanation and teleological purpose: Causal dynamics and teleology are equivalent.

## Acknowledgments

---

It is my pleasure to thank people whose thoughts, ideas, encouragement and support shaped this chapter: R. Brockett, G. Carpenter, A. Chernyakov, R. Deming, V. Dmitriev, W. Freeman, K. Fukunaga, L. Garvin, I. Ginzburg, R. Gudwin, S. Greineder, S. Grossberg, M. Karpovsky, S. Kashin, L. Levitin, R. Linnehan, T. Luginbuhl, A. Meystel, S. Middleton, K. Moore, C. Mutz, A. Ovsich, R. Payne, N. Pechersky, I. Perlovsky, V. Petrov, C. Plum, J. Rumer, E. Schakhnovich, W. Schoendorf, N. Shokhirev, J. Sjogren, D. Skatrud, R. Streit, E. Taborsky, I. Ternovsky, T. Ternovsky, E. Tichovolsky, B. Weijers, D. Vinkovetsky, Y. Vinkovetsky, P. Werbos, M. Xiarhos, L. Zadeh and G. Zainiev. Anonymous reviewers made many valuable suggestions.

## References

---

- Albus, J.S., & Meystel, A.M. (2001). *Engineering of mind: An introduction to the science of intelligent systems*. New York: Wiley.
- Aristotle. (1995). *Complete Works of Aristotle* (IV BCE, a), J. Barnes (Ed.), Princeton, NJ: Princeton University Press.
- Aristotle. (1995). Rhetoric for Alexander, in *Complete Works of Aristotle* (IV BCE, b), J. Barnes (Ed.), Princeton, NJ: Princeton University Press.
- Aristotle. (1995). Organon, in *Complete Works of Aristotle* (IV BCE, c), J. Barnes (Ed.), 18a28-19b4; 1011b24-1012a28, Princeton, NJ: Princeton University Press..
- Aristotle. (1995). Metaphysics, in *Complete Works of Aristotle* (IV BCE), J. Barnes (Ed.), W.D. Ross (Trans.), Princeton, NJ: Princeton University Press.
- Bellman, R.E. (1961). *Adaptive control processes*. Princeton, NJ: Princeton University Press.
- Berlyne, D. E. (1960). *Conflict, arousal, and curiosity*. New York, NY: McGraw-Hill.
- Berlyne, D. E. (1973). *Pleasure, reward, preference: Their nature, determinants, and role in behavior*. New York, NY: Academic Press.
- Carpenter, G.A., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics and Image Processing*, 37, 54-115.
- Catholic Encyclopedia. (<http://www.newadvent.org/cathen/08050b.htm>)
- Chalmers, D.J. (1997). *The conscious mind: In search of a fundamental theory*. Oxford University Press.

- Charness, G., & Levin, D. (2003). *When optimal choices feel wrong: A laboratory study of Bayesian updating, complexity, and affect*. (Paper 9-03, <http://repositories.cdlib.org/ucsbecon/dwp/9-03>). Departmental Working Papers, Department of Economics, UCSB.
- Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge, UK: Cambridge University Press.
- Chomsky, N. (1981). *Principles and parameters in syntactic theory*. In N. Hornstein & D. Lightfoot (Eds.), *Explanation in linguistics: The logical problem of language acquisition*. London: Longman.
- Clark, A. (1987). The kludge in the machine. *Mind and Language*, 2, 277-300.
- Cramer, H. (1946). *Mathematical methods of statistics*. Princeton, NJ: Princeton University Press.
- Damasio, A.R. (1995). *Descartes' error: Emotion, reason, and the human brain*. New York, NY: Avon.
- Deacon, T.W. (1998). *The symbolic species: The co-evolution of language and the brain*. New York: W.W. Norton & Company.
- Descartes, R. (1646/1989). *The passions of the soul: Les passions de lame*. Indianapolis IN: Hackett Publishing Company, 1989.
- Edelman, G. M., & Tononi, G. (1995). *A universe of consciousness: How matter becomes imagination*. New York: Basic Books.
- Feynman, R.P. & Hibbs, A. R. (1965). *Quantum mechanics and path integrals*. New York: McGraw-Hill.
- Fontanari, J.F., & Perlovsky, L.I. (2004). Solvable null model for the distribution of word frequencies. *IE 70*, 04290.
- Fontanari, J.F., & Perlovsky, L.I. (2005). *Evolution of communication in a community of simple-minded agents*. IEEE International Conference. On Integration of Knowledge Intensive Multi-Agent Sys., Waltham, MA.
- Fontanari, J.F., & Perlovsky, L.I. (2005). *Meaning creation and modeling field theory*. IEEE Int. Conf. On Integration of Knowledge Intensive Multi-Agent Sys., Waltham, MA.
- Freeman, W. J. (2000). *Neurodynamics: An exploration in mesoscopic brain dynamic*. New York: Springer.
- Freeman, W.J. *Mass action in the nervous system*. New York: Academic Press.
- Gödel, K. (1986). *Kurt Gödel collected works, I*. S. Feferman, et al. (Eds.), Oxford: Oxford University Press.
- Griffiths, P. E. (1998). *What emotions really are: The problem of psychological categories*. Chicago, IL: University Of Chicago Press.
- Grossberg, S., & Levine, D.S. (1987). Neural dynamics of attentionally modulated Pavlovian conditioning: Blocking, inter-stimulus interval, and secondary reinforcement. *Psychobiology*, 15(3), 195-240.
- Grossberg, S. (1975). *Neural networks and natural intelligence*. Cambridge, MA: MIT Press.

- Grossberg, S. (2000). Linking mind to brain: The mathematics of biological intelligence. *Notices of the American Mathematical Society*, 471361-1372.
- Hameroff, S. (1994). *Toward a scientific basis for consciousness*. Cambridge, MA: MIT Press.
- Hebb, D. (1949). *Organization of behavior*. New York: J. Wiley & Sons.
- Hilbert, D. (1928/1967). The foundations of mathematics. In J. van Heijenoort (Ed.), *From Frege to Gödel* (p. 475), Cambridge, MA: Harvard University Press.
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. New York: Oxford University Press.
- James, W. (1890). *The principles of psychology*. New York: Dover Books,
- Jaynes, J. (1976). *The origin of consciousness in the breakdown of the bicameral mind* (2<sup>nd</sup> ed.), Boston, MA: Houghton Mifflin Co.
- Josephson, B. (1997). *An integrated theory of nervous system functioning embracing nativism and constructivism*. Nashua, NH: International Complex Systems Conference.
- Jung, C.G. (1921/1971). Psychological types. In *Collected Works* (v.6, Bollingen Series XX), Princeton, NJ: Princeton University Press.
- Jung, C.G. (1934/1969). Archetypes of the collective unconscious. In *Collected Works* (v.9, II, Bollingen Series XX), Princeton, NJ: Princeton University Press.
- Kant, I. (1781/1943). *Critique of pure reason*. J.M.D. Meiklejohn (Trans.), New York: John Wiley.
- Kant, I. (1788/1986). *Critique of practical reason*. J.H Bernard (Trans.) 1986, Hafner.
- Kant, I. (1790/1914). *Critique of judgment*. J.H. Bernard (Trans.), London: Macmillan & Co.
- Kant, I. (1798/1974). *Anthropology from a pragmatic point of view*. M.J. Gregor (Trans.), Boston, MA: Kluwer Academic Pub.
- Kecman, V. (2001). *Learning and soft computing: Support vector machines, neural networks, and fuzzy logic models (complex adaptive systems)*. Cambridge, MA: The MIT Press.
- Koch, C., & Segev, I. (Eds.). (1998). *Methods in neuronal modeling: From ions to networks*. Cambridge, MA: MIT Press.
- Ledoux, J. (1998) *The emotional brain: The mysterious underpinnings of emotional life*. New York: Simon & Schuster.
- Linnehan, R., Mutz, Perlovsky, L.I., Weijers, B., Schindler, J., & Brockett, R. (2003). *Detection of patterns below clutter in images*. Int. Conf. On Integration of Knowledge Intensive Multi-Agent Systems, Cambridge, MA, Oct. 1-3, 2003.
- Marchal, B. (2005). Theoretical computer science & the natural sciences. *Physics of Life Reviews*, 2(3), 1-38.
- Meystel, A. (1995). *Semiotic modeling and situational analysis*. Bala Cynwyd, PA: AdRem.

- Meystel, A.M., & Albus, J.S. (2001). *Intelligent systems: Architecture, design, and control*. New York: Wiley.
- Minsky, M. (1988). *The society of mind*. Cambridge, MA: MIT Press.
- Minsky, M. (1995). Smart Machines. In *The third culture* (152-166). Brockman (Ed.). New York: Simon & Shuster. 152-166.
- Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, 11, 207-212.
- Newell, A. (1983). Intellectual issues in the history of artificial intelligence. In *The Study of Information*. F.Machlup & U.Mansfield (Eds.). New York: John Wiley.
- Ortony, A., & Turner, T.J. (1990). What's basic about basic emotions? *Psychological Review*, 97, 315-331.
- Ortony, A., Clore, G.L., & Collins, A. (1990). *The cognitive structure of emotions*. Cambridge University Press.
- Penrose, R. (1994). *Shadows of the mind*. Oxford, England: Oxford University Press.
- Perlovsky, L.I. (1996). Mathematical concepts of intellect. In *Proceedings, World Congress on Neural Networks*. San Diego, CA.: L. Erlbaum Assoc., pp. 1013-16.
- Perlovsky, L.I. (1996). Gödel theorem and semiotics. In *Proceedings of the Conference on Intelligent Systems and Semiotics '96*. Gaithersburg, MD, v.2, pp. 14-18.
- Perlovsky, L.I. (1997). Physical concepts of intellect. *Proceedings of the Russian Academy of Sciences*, 354(3), 320-323.
- Perlovsky, L.I. (1998). Conundrum of combinatorial complexity. *IEEE Trans. PAMI*, 20(6), 666-70.
- Perlovsky, L.I. (1998). Cyberaesthetics: aesthetics, learning, and control. *STIS '98*, Gaithersberg, MD.
- Perlovsky, L.I. (1999). Emotions, learning, and control. Proceedings of the *International Symposium on Intelligent Control, Intelligent Systems & Semiotics*, Cambridge MA, pp. 131-137
- Perlovsky, L.I. (2001). *Neural networks and intellect: Using model based concepts*. New York: Oxford University Press.
- Perlovsky, L.I. (2002). Statistical limitations on molecular evolution. *Journal of Biomolecular Structure & Dynamics*, 19(6), 1031-43.
- Perlovsky, L.I. (2002). Aesthetics and mathematical theories of intellect. (Russian). *Iskusstvoznanie*, 2(02), 558-594.
- Perlovsky, L.I. (2004). Integrating Language and Cognition. *IEEE Connections*, Feature Article, 2(2), 8-12.
- Perlovsky, L.I. (2006). Symbols: Integrated cognition and language. In *Computational semiotics*. A. Loula, & R. Gudwin (Eds.). (In press.).
- Perlovsky, L.I. (2006). *The knowledge instinct*. Basic Books. (In press.).
- Perlovsky, L.I., & McManus, M.M. (1991). Maximum likelihood neural networks for sensor fusion and adaptive classification. *Neural Networks*, 4(1), 89-102.
- Perlovsky, L.I., Webb, V.H., Bradley, S.R., & Hansen, C.A. (1998). Improved ROTH detection and tracking using MLANS. *AGU Radio Science*, 33(4), 1034-44.

- Piaget, J. (1981/2000). *The psychology of the child*. H. Weaver (Trans.). Basic Books.
- Pinker, S. (1995). Smart machines. In *The third culture* (223-238). Brockman (Ed.). New York: Simon & Shuster.
- Plutarch. (2 AD/2001). *Lives*. Modern Library.
- Sartre, J.P. (1948/1984). *Emotions*. See also J. P. Sartre, *Existentialism and human emotions*. Citadel Press, Reissue edition, 1984.
- Searle, J.R. (1980). Minds, brains, and programs. *The Behavioral and Brain Sciences*, 3, Cambridge University Press.
- Searle, J.R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge University Press.
- Searle, J.R. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press.
- Searle, J.R. (1997). *The mystery of consciousness*. New York: New York Review of Books.
- Singer, R.A., Sea, R.G. & Housewright, R.B. (1974). Derivation and evaluation of improved tracking filters for use in dense multitarget environments. *IEEE Transactions on Information Theory*, IT-20, 423-432.
- Taylor, J. G. (2005). Mind and consciousness: Towards a final answer? *Physics of Life Reviews*, 2(1), 57.
- The American Heritage College Dictionary* (3<sup>rd</sup> ed.). (2000). Boston: Houghton Mifflin. I would emphasize, these general dictionary formulations could only be taken as a starting point.
- Westfall, R.S. (1983). *Never at rest: A biography of Isaac Newton*. Cambridge: Cambridge University Press.
- Wikipedia. <http://en.wikipedia.org/wiki/Instinct>.
- Zadeh, L.A. (1997). Information granulation and its centrality in human and machine intelligence. Proceedings of the *Conference on Intelligent Systems and Semiotics '97*. Gaithersburg, MD, pp. 26-30.
- Zeki, A. (1993). *A vision of the brain*. Oxford, England: Blackwell.

---

## Endnotes

- <sup>1</sup> A simple example of an adaptive model is linear regression: The knowledge is encapsulated in the choice of variables, the uncertainty and adaptivity is in the unknown parameters, fitted to the data. Whereas linear regression uses one model, model-based systems used a large number of models. For example, a scene is described using geometric models of many objects. Parameters may include: size, orientation angles, color, illumination conditions, etc. A simple, still nontrivial, problem causing difficulties in applications till today is tracking multiple objects in the presence of clutter (Singer, Sea, & Housewright, 1974; Perlovsky, Webb, Bradley, & Hansen, 1998).<sup>2</sup> Searle's views of intentionality are seen by many researchers as too narrow. (For example, see Workshop on Neurodynamics and

Dynamics of Intentional Systems, 2005, International Joint Conference on Neural Networks, Montreal, Canada). It is not possible to give a complete treatment of such issues as intentionality and purposiveness in this chapter, which differ in cognitive science, artificial intelligence, classical philosophy and theory of action (D. Davidson, *Essays on Actions and Events in Philosophical Essays of Donald Davidson*, Oxford University Press, 2002; M. Bratman, *Faces of Intention: Selected Essays on Intention and Agency*, Cambridge University Press, 1999). Nevertheless, it is impossible to avoid these issues, because intentionality and purposiveness, as discussed later, are fundamental to living beings and to higher brain functions. Therefore, my approach in this chapter here and below is to use commonsense understanding of terms whenever possible, while noticing discrepancies among various understandings and to give corresponding references for further reading. I would like to emphasize again that the choice among various understandings of many philosophical, psychological and cognitive terms used in this chapter is driven by four principles specified in the first section and is consistent with the mathematical theory presented next. Never was a single narrow technical definition selected to fit the mathematical structure. On the other hand, the mathematical structure turned out to be compatible with general understanding of these terms gradually developed since time of Socrates. The mutual compatibility of knowledge among Socrates, Plato, Aristotle, Kant, Jung and contemporary research is emphasized and discrepancies are noticed, whenever relevant.

<sup>3</sup> Descartes, Rene (1646). *Passions of the Soul*. Descartes deeply penetrated into the interaction between emotions and consciousness: “those who are the most excited by their passions are not those who know them best and their passions are... confused and obscure.” He did not differentiate between thoughts and emotions consistently: “of all the various kinds of thoughts... these passions,...” “...the action and the passion are thus always one and the same thing.” Descartes showed a bias toward unconscious and fused perception of emotions characteristic of a thinking psychological type.

<sup>4</sup> Mathematically, the condition that the object  $h$  is present with absolute certainty, is expressed by normalization condition:  $\int l(X|h) dX = 1$ . We should also mention another normalization condition:  $\int l(X(n)) dX(n) = 1$ , which expresses the fact that, if a signal is received, some object or objects are present with 100% certainty.

<sup>5</sup> This construction is not unique. Expression (2) is inspired by the notion of likelihood between models and signals. Another general type of similarity measure suitable for the knowledge instinct (Perlovsky, 2001) is inspired by the notion of mutual information in the models about the signals. Here we would like to mention a modification of (2) for a specific case. Sometimes a set of observations,  $N$ , is more convenient to describe mathematically as a continuous flow of signals, for example, a flow of visual stimuli in time and space; then, it is convenient instead of Equation

(1) to consider its continuous version,  $L = \exp +N \ln \left( \sum_{h \in H} r(h) l(X(n) | h) \right)$ , where

$N$  is a continuum of bottom-up signals, such as in time-space. <sup>6</sup> See (Kant, 1790). Some of my colleagues think that these are all Kantian ideas, other think that it is all my invention and Kant never said anything that specific. I agree with both. I learned these ideas from Kant, but it is also true that other Kant readers did not

understand Kant the same way. Every year there are dozens of papers published interpreting Kantian views, so clearly, the matter cannot be definitely settled within this chapter. I would just repeat that my inspiration for understanding Kant, Aristotle and many other philosophers and scientists is driven by a desire for a unified view of the mechanisms of the mind. I would emphasize that I am skeptical about the value of critical approaches to understanding old texts: When I read that Kant or Aristotle did not understand this or that, usually I feel that the authors of these statements do not understand Kant or Aristotle. Similarly, I wonder what Freud or Jung would think about their contemporary followers. In one movie comedy, Kurt Vonnegut wrote an essay for a college course about his own writing; this essay received “C-.” This joke reminds us to be modest about how well we can understand other people. I maintain that the only proof of correct understanding of any idea comes when it is positively integrated within the development of science.

- 7 Let me repeat, it is not possible to review all relevant points of view here<sup>2</sup>.
- 8 I’ll repeat that Aristotle called intentionality the “end causes of Forms” and he called the a priori contents of the mind the “Formal causes” that are the mechanisms of Forms (Metaphysics).
- 9 According to Kant, purposiveness is the a priori principle of the ability for judgment (Kant, 1790). It was mathematically described as the knowledge instinct: It drives us to find the best reason (knowledge) for any judgment or action in this chapter.
- 10 Consciousness of emotions and mechanisms of this type of consciousness is a separate, important and interesting topic considered elsewhere (L. Perlovsky, 2006, *The Knowledge Instinct*, Basic Books); it is beyond the scope of this chapter.

# Section II

---

## Methodological Issues