2009 Special Issue

Language and cognition

Leonid Perlovsky*

Harvard University, SEAS, Cambridge, USA
The AF Research Laboratory, Sensors Directorate, Hanscom AFB, USA

ABSTRACT

What is the role of language in cognition? Do we think with words, or do we use words to communicate made-up decisions? The paper briefly reviews ideas in this area since 1950s. Then we discuss mechanisms of cognition, recent neuroscience experiments, and corresponding mathematical models. These models are interpreted in terms of a biological drive for cognition. Based on the Grossberg–Levin theory of drives and emotions, we identify specific emotions associated with the need for cognition. We demonstrate an engineering application of the developed technique, which significantly improves detection of patterns in noise over the previous state-of-the-art. The developed mathematical models are extended toward language. Then we consider possible brain–mind mechanisms of interaction between language and cognition. A mathematical analysis imposes restrictions on possible mechanisms. The proposed model resolves some long-standing language–cognition issues: how the mind learns correct associations between words and objects among an astronomical number of possible associations; why kids can talk about almost everything, but cannot act like adults, what exactly are the brain–mind differences; why animals don’t talk and think like people. Recent brain imaging experiments indicate support for the proposed model. We discuss future theoretical and experimental research.

1. Nativism, cognitivism, evolutionism

Complex innate mechanisms of the mind were not appreciated in the first half of the last century. Thinking of mathematicians and intuitions of psychologists and linguists were dominated by logic. Considered mechanisms of logic were not much different for language or cognition; both were based on logical statements and rules. Even fundamental Gödelian theory (Gödel, 1931/1994) establishing the deficiency of logic did not move thinking about the mind away from logic.

Contemporary linguistic interests in the mind mechanisms of language were initiated in the 1950s by Chomsky (1965). He identified the first mysteries about language that science had to resolve. “Poverty of stimulus” addressed the fact that the tremendous amount of knowledge needed to speak and understand language is learned by every child around the world even in the absence of formal training. It has seemed obvious to Chomsky that surrounding language cultures do not carry enough information for a child to learn language, unless specific language learning mechanisms are inborn in the mind of every human being. This inborn mechanism should be specific enough for learning complex language grammars and still flexible enough so that a child of any ethnicity from any part of the world would learn whichever language is spoken around, even if he or she is raised on the other side of the globe. Chomsky called this inborn learning mechanism Universal Grammar and set out to discover its mechanisms. He emphasized the importance of syntax and thought that language learning is independent of cognition. This approach to language based on innate mechanisms, is called nativism.

Chomsky and his school initially used available mathematics of logical rules, similar to rule systems of artificial intelligence. In 1981, Chomsky (Chomsky, 1981) proposed a new mathematical paradigm in linguistics, rules and parameters. This was similar to model-based systems emerging in mathematical studies of cognition. Universal properties of language grammars were supposed to be modeled by parametric rules or models, and specific characteristics of grammar of a particular language were fixed by parameters, which every kid could learn from a limited exposure to the surrounding language. Another fundamental change of Chomsky’s ideas (Chomsky, 1995) was called the minimalist program. It aimed at simplifying the rule structure of the mind mechanism of language. Language was modeled in closer interactions to other mind mechanisms, closer to the meaning, but stopped at an interface between language and meaning. Chomsky’s linguistics still assumes that meanings appear independently from language. Logic is the main mathematical modeling mechanism.

Many linguists disagreed with separation between language and cognition in Chomsky’s theories. Cognitive linguistics emerged in the 1970s to unify language and cognition, and explain...
creation of meanings. Cognitive linguistics rejected Chomsky’s idea about a special module in the mind devoted to language. The knowledge of language is different from that of other domains, and is based on conceptual mechanisms. It is embodied and situated in the environment. Related research on construction grammar argues that language is not compositional, not all phrases are constructed from words using the same syntax rules and the reference therein. We have dozens of such sensors, mechanism described in usethewordinstincttodenoteasimpleinborn,non-adaptive upwithinstinctualbehaviorandothernotveryusefulideas.We literature; the reason is that the notion of instinct was mixed description will berefinedlater; see (Christiansen & Kirby, 2003; Hurford, 2008). Evolutionary linguistics by simulation of societies of communicating agents (Brighton, Smith, & Kirby, 2005) demonstrated the emergence of a compositional language.

2. Cognition, dynamic logic, and the knowledge instinct

Consider a seemingly simple experiment. Close your eyes and imagine an object in front of you. The imagined image is vague, not as crisp and clear as with opened eyes. As we open eyes, the object becomes crisp and clear. It seems to occur momentarily, but actually it takes 1/5th of a second. This is a very long time for neural brain mechanisms – hundreds of thousands of neural interactions. Let us also note: with opened eyes we are not conscious about initially vague imagination, we are not conscious about the entire 1/5th of a second, we are conscious only about the end of this process: crisp, clear object in front of our eyes. The explanation of this experiment has become simple after many years of research that have found out what goes on in the brain during these 1/5th of a second.

2.1. Instincts, emotions, concepts

Explaining this experiment requires us to consider mechanisms of concepts, instincts, and emotions. We perceive and understand the world around due to the mechanism of concepts. Concepts are like internal models of objects and situations; this analogy is quite literal, e.g., during visual perception of an object, a concept-model of the object stored in memory projects an image (top- down signals) onto the visual cortex, which is matched there to an image projected from the retina (bottom-up signal; this simplified description will be refined later; see Grossberg (1988)).

The mechanism of concepts evolved for instinct satisfaction. The word instinct is not used currently in the psychological literature; the reason is that the notion of instinct was mixed up with instinctual behavior and other not very useful ideas. We use the word instinct to denote a simple inborn, non-adaptive mechanism described in Grossberg and Levine (1987). Instinct is a mechanism of the internal “sensor”, which measures vital body parameters, such as blood pressure, and indicate to the brain when these parameters are out of safe range. This simplified description will be sufficient for our purposes, more details could be found in Gnadt and Grossberg (2008) and Grossberg and Seidman (2006) and the references therein. We have dozens of such sensors, measuring sugar level in blood, body temperature, pressure at various parts, etc.

According to instinctual–emotional theory (Grossberg & Levine, 1987), communicating satisfaction or dissatisfaction of instinctual needs from instinctual parts of the brain to decision making parts of the brain is performed by emotional neural signals. The word emotion refers to several neural mechanisms in the brain (justin & Västfjäll, 2008); in this paper we always refer to the mechanism connecting conceptual and instinctual brain regions. Perception and understanding of concept-models corresponding to objects or situations that can potentially satisfy an instinctual need receive preferential attention and processing resources in the mind.

Projection of top-down signals from a model to the visual cortex primes or makes visual neurons to be more receptive to matching bottom-up signals. This projection produces imagination that we perceive with closed eyes, as in the closed–open eye experiment. Conscious perception occurs, as mentioned, after top- down and bottom-up signals match. The process of matching for a while presented difficulties to mathematical modeling, as discussed below.

2.2. Combinatorial complexity, logic, and dynamic logic

Perception and cognition abilities of computers still cannot compete with those of kids and animals. Most algorithms and neural networks suggested since 1950s for modeling perception and cognition, as discussed in Perlovsky (2006a), faced difficulty of combinatorial complexity (CC). Rule systems of artificial intelligence in the presence of variability has grown in complexity: rules have become contingent on other rules, and rule systems faced CC. Algorithms and neural networks designed for learning have to be trained to understand not only individual objects, but also combinations of objects, and thus faced CC of training. Fuzzy systems required a fuzziness level to be set appropriately in different parts of systems, also degrees of fuzziness vary in time, an attempt to select efficient levels of fuzziness would lead to CC.

These CC difficulties were related to Gödelian limitations of logic, they were manifestations of logic inconsistency in finite systems (Perlovsky, 2000). Even approaches designed specifically to overcome logic limitations, such as fuzzy logic and neural networks, encountered logical steps in their operations: neural networks are trained using logical procedures (e.g. “this is a chair”), and fuzzy systems required logical selection of the degree of fuzziness.

To overcome limitations of logic, dynamic logic was proposed (Perlovsky, 2000, 2006a; Perlovsky & McManus, 1991). In the next section we summarize the mathematical description of dynamic logic, here we describe it conceptually. Whereas logic works with statements (e.g. “this is a chair”), dynamic logic is a process from vague to crisp, from vague statement, decision, plan, to crisp ones. It could be viewed as fuzzy logic that automatically sets a degree of fuzziness corresponding to the accuracy of learning models.

Dynamic logic corresponds to the open–close eye experiment: initial states of models are vague. This experiment was recently performed with much more details using brain imaging. Bar et al. (2006) used functional Magnetic Resonance Imaging (fMRI) to obtain high spatial resolution of processes in the brain, which they combined with magneto-encephalography (MEG), measurements of the magnetic field next to the head, which provided high temporal resolution of the brain activity. Combining these two techniques the experimenters were able to receive high resolution of cognitive processes in space and time. Bar et al. concentrated on three brain areas: early visual cortex, object recognition area (fusiform gyrus), and object information semantic processing area (OFC). They demonstrated that OFC is activated 130 ms after the visual cortex, but 50 ms before object recognition area. This suggests that OFC represents the cortical source of top-down facilitation in visual object recognition. This top-down facilitation was unconscious. In addition they demonstrated that the imagined image generated by top-down signals facilitated from OFC to cortex is vague, similar to the closed–open eye experiment. Conscious perception of an object occurs when vague projections become crisp and match the crisp and clear image from the retina, and an object recognition area is activated.
2.3. The knowledge instinct and neural modeling field theory

The process of matching concept-models in memory to bottom-up signals coming from sensory organs is necessary for perception; otherwise an organism will not be able to perceive the surroundings and will not be able to survive. Therefore humans and higher animals have an inborn drive to fit top-down and bottom-up signals. We call this mechanism the instinct for knowledge (1991 [Perlovsky, 2006a]). This mechanism is similar to other instincts in that our mind has a sensor-like mechanism that measures a similarity between top-down and bottom-up signals, between concept-models and sensory percepts. Brain areas participating in the knowledge instinct were discussed in Levin and Perlovsky (2008). As discussed in that publication, biologists considered similar mechanisms since 1950s; without a mathematical formulation, however, its fundamental role in cognition was difficult to discern. All learning algorithms have some models of this instinct, maximizing correspondence between sensory input and an algorithm internal structure (knowledge in a wide sense). According to Grossberg and Levine (1987) instinct–emotion theory, satisfaction or dissatisfaction of every instinct is communicated to other brain areas by emotional neuronal signals. We feel these emotional signals as harmony or disharmony between our knowledge-models and the world. At lower layers of everyday object recognition these emotions are usually below the level of consciousness; at higher layers of abstract and general concepts this feeling of harmony or disharmony could be strong, as discussed in Perlovsky (2006b) it is a foundation of our higher mental abilities. We summarize now a mathematical theory combining the discussed mechanisms of cognition as interaction between top-down and bottom-up signals at a single layer in multi-layer heterarchical system following Perlovsky (2006a).

Neurons are enumerate by index \( n = 1, \ldots, N \). These neurons receive bottom-up input signals, \( X(n) \), from lower layers in the processing heterarchy. The word heterarchy is used by many neural and cognitive scientists to designate that the mind is organized in an approximate hierarchy; this hierarchy is not exact, cross-layer interactions are abundant (Grossberg, 1988); we would use hierarchy for simplicity, \( X(n) \) is a field of bottom-up neuronal synapse activations, coming from neurons at a lower layer. Top-down, or priming signals to these neurons are sent by concept-models, \( M_h \) \((S_h, n)\); we enumerate models by index \( h = 1, \ldots, H \). Each model is characterized by its parameters, \( S_h \). Models represent signals in the following sense. Say, signal \( X(n) \), is coming from sensory neurons activated by object \( h \), characterized by parameters \( S_h \). These parameters may include position, orientation, or lighting of an object \( h \). Model \( M_h \) \((S_h, n)\) predicts a value \( X(n) \) of a signal at neuron \( n \). For example, during visual perception, a neuron \( n \) in the visual cortex receives a signal \( X(n) \) from retina and a priming signal \( M_h \) \((S_h, n)\) from a concept-model \( h \). A neuron \( n \) is activated if both a bottom-up signal from lower layer input and a top-down priming signal are strong. Various models compete for evidence in the bottom-up signals, while adapting their parameters for better match as described below. This is a simplified description of perception. Models \( M_h \) specify a field of primed neurons \( n \), hence the name for this modeling architecture, modeling fields.

The knowledge instinct maximizes a similarity measure between top-down and bottom-up signals,

\[
L(X, M) = \prod_{n \in H} \prod_{n' \in H} r(h)(n|h). \tag{1}
\]

Here \( r(n|h) \) is a partial similarity of a bottom-up signal in pixel \( n \) given that it originated from concept-model \( h \); functional shape of \( r(n|h) \) often can be taken as a Gaussian function of \( X(n) \) with the mean \( M_h \) \((S_h, n)\). Partial similarities are normalized on objects (or concepts) \( h \) being definitely present, and coefficient \( r(h) \) estimate a probability of them actually being present. Similarity \( L \) accounts for all combinations of signals \( n \) coming from any model \( h \), hence the huge number of items \( H^N \) in Eq. (1); this is a basic reason for combinatorial complexity of most algorithms. From time to time a system forms a new concept-model, while retaining an old one as well; alternatively, old concepts are sometimes merged or eliminated. This requires a modification of the similarity measure (1); the reason is that more models always result in a better fit between the models and data. Therefore similarity (1) has to be reduced using a “skeptic penalty function”, \( p(N, M) \) that grows with the number of models \( M \), and this growth is steeper for a smaller amount of data \( N \).

The learning instinct demands maximizing the similarity \( L \) over model parameters \( S \). Dynamic logic maximizes similarity \( L \) while matching vagueness or fuzziness of similarity measures to the uncertainty of models. It starts with any unknown values of parameters \( S \) and defines association variables \( f(h|n) \),

\[
f(h|n) = r(h)(n|h) \sum_{h' \in H} r(h')(n|h'). \tag{2}
\]

Dynamic logic determining the Modeling Field (MF) dynamics is given by

\[
df(h|n)/dt = f(h|n) \sum_{h' \in H} \delta_{hh'} - f(h'|n) \cdot \\
\left( \sum_{n \in N} \frac{\partial \ln l(n|h')}{\partial M_{hh'}} \frac{\partial M_{hh'}}{\partial S_h} \cdot \frac{\partial S_h}{\partial t}, \tag{3}
\right.
\]

\[
dS_h/\partial t = \sum_{n \in N} f(h|n) \frac{\partial \ln l(n|h)}{\partial M_{h}} \frac{\partial M_{h}}{\partial S_h} \tag{4}
\]

where \( \delta_{hh'} \) is 1 if \( h = h' \), 0 otherwise.

Initially, parameter values are not known, and uncertainty of partial similarities is high (e.g., if \( l(n|h) \) is modeled by Gaussian functions, variances are high). So the fuzziness of the association variables is high. In the process of learning, models become more accurate, and association variables more crisp, as the value of the similarity increases. The number of models is determined in the learning process. The system always keeps a store of dormant models, which are vague, have low \( r(h) \), and do not participate in the parameter fitting; only their parameters \( r(h) \) are updated. When \( r(h) \) exceeds a threshold, a model is activated; correspondingly, an active model is deactivated when its \( r(h) \) falls below the threshold. MF organization is similar to ART (Carpenter & Grossberg, 1987) in that it models interaction between bottom-up and top-down signals. It is different in that it fits all models in parallel.

Dynamic logic process always converges (Perlovsky, 2000); it is proven by demonstrating that at each time step in Eqs. (3) and (4) (as long as the bottom-up signals remain constant), the knowledge instinct (1) increases; thus dynamic logic and the knowledge instinct are mathematically equivalent.

2.4. Perception example

Here we illustrate the developed technique with an example described in Perlovsky (2006a), which demonstrates that the described theory can find patterns below noise at about 100 times better in terms of signal-to-noise ratio, than previous state-of-the-art algorithms. The reason for choosing such an example is to demonstrate, in a relatively simple way, that engineering algorithms based on the mind cognitive mechanisms significantly exceed capabilities of ad hoc algorithms (Fig. 1).

As exact pattern shapes are not known and depend on unknown parameters, these parameters should be found by fitting the
pattern model to the data. At the same time it is not clear which subset of the data points should be selected for fitting. A previous state-of-the-art algorithm, multiple hypothesis testing (Singer, Sea, & Housewright, 1974) tries various subsets. In difficult cases, all combinations of subsets and models are exhaustively searched, leading to combinatorial complexity. In the current example the searched patterns are shown in Fig. 2(a) without noise, and in Fig. 2(b) with noise, as actually measured. Direct search through all combinations of models and data leads to complexity of \( M^{4000} \). A search in parameter space yields less complexity. Each pattern is characterized by a 3-parameter parabolic shape plus the 4th parameter for the average signal strength. The image size is 100 \( \times \) 100 points, and the true number of patterns is 3, which is not known. Therefore, at least 4 patterns should be fit to the data, to decide that 3 patterns fit best. Fitting 4 \( \times \) 4 = 16 parameters to 100 \( \times \) 100 grid by a brute-force testing would take about \( 10^{40} \) to \( 10^{42} \) operations, still a prohibitive computational complexity.

The models and conditional similarities for this case are described in detail in Linnehan et al. (2003): a uniform model for noise, Gaussian blobs for highly fuzzy, poorly resolved patterns, and parabolic models for the patterns of interest. The number of computer operations in this example was about \( 10^5 \). Thus, a problem that was not solvable due to CC becomes solvable using dynamic logic.

In this example dynamic logic performs better than the human visual system. This is understood due to the fact that the human visual system is optimized for different type of images, not for parabolic shapes in noise.

An ability of dynamic logic to extract signals from strong noise and clutter was used in many applications; we would mention here an application to EEG signals (Kozma, Deming, Perlovsky, Levine, & Perlovsky, 2007). Potentially, EEG signals contain information about brain cognitive events; detecting these signals and estimating their parameters could be utilized to allow quadriplegics to move a computer cursor or steer their wheelchairs with their thoughts; or those playing computer games could control actions on the screen with their thoughts. The difficulty is that EEG signals are notoriously noisy. The referenced article describes a dynamic logic algorithm for extracting cognitively related events from EEG.

### 3. Extension to language

All linguistic theories, as reviewed at the beginning of the paper, are formulated as logical systems, and face combinatorial complexity. This is possibly why computers do not understand human language, and in particular, Google, Yahoo, and other search engines, while being immensely useful, cause so much frustrations to their users. Extension of dynamic logic to language promises to remedy the situation. Here we briefly summarize this extension following Perlovsky (2006c). The challenge in extending dynamic logic to language has been in substituting derivatives in Eqs. (3) and (4) with equivalent procedures suitable for linguistic constructs that are essentially discrete, non-differentiable structures. For example, consider a phrase “Leonid sits in a chair”. A language learning procedure should be able to figure out that the gist of this phrase, its most essential part is {sit, chair}; “Leonid” can be substituted by many other nouns, and “in,” “a” are even more dispensable. The main idea of dynamic logic is learning sentence structures not by trying all possible combinations of words, but by taking a “derivative” of a phrase with respect to constituent words. But of course standard mathematical definition of a derivative is not applicable to this situation in principle. Language constructs are essentially discrete and non-differentiable.

A suitable “derivative-like” procedure was described in Perlovsky (2006c). Here we summarize it for a word–phrase layer; where bottom-up signals are comprised of words, top-down models are phrases, and these phrase-models are learned without combinatorial complexity. The bottom-up input data, \( X(n) \), in this “phrase-layer” MF system, are word strings, for simplicity, of a fixed length, \( S, X(n) = \{ w_{n+1}, w_{n+2}, \ldots w_{n+5} \} \). Here \( w_n \) are words from a given dictionary of size \( K, W = \{ w_1, w_2, \ldots w_K \} \), and \( n \) is the word position in a body of texts. A simple phrase-model often used in computational linguistics is “a bag of words”, that is, a model is a subset of words from a dictionary, without any order or rules of grammar,

\[
M^L_h (S_h, n) = \{ w_{h1}, \ldots, w_{h3} \}.
\]

A superscript \( L \) here denotes a language model, the parameters of this model are its words, \( M^L_h (S_h, n) = S_h = \{ w_{h1}, \ldots, w_{h3} \} \).
The language learning (traditionally called language acquisition in Chomskyan linguistics) consists in defining models—concepts—phrases best characterizing the given body of texts in terms of a similarity measure.

Fig. 2. Hierarchical integrated language–cognition MF system. “Heterarchy” refers to cross-layer connections, not shown, and to the consequence that the hierarchical structure is not logically strict as may appear from the figure. At each layer in a hierarchy there are integrated language and cognition models (thick arrow). Similarities are integrated as products of language and cognition similarities. Initial models are fuzzy placeholders, so integration of language and cognition is sub-conscious. Association variables depend on both language and cognitive models and signals. Therefore language model learning helps cognitive model learning and v.v. High-layer abstract cognitive concepts are grounded in abstract language concept similarities.

\[ p(h|s) = \frac{(2\pi \sigma_h^2)^{-1/2} \exp[-0.5\delta(h, s)/\sigma_h^2]}{\sum_{h'} \exp[-0.5\delta(h', s)/\sigma_h^2]} \]  

(10)

Here, \( \delta \) is an average distance of the word \( w_{h,s} \) from its phrase-model, \( h \). A non-normalized measure of belonging of word \( s \) to phrase \( h \) is given by \( p(h|s) \). It is used to define probabilistic-like vague phrase contents, \( \phi(s|h) \).

This is now used to define the dynamics of the word contents of the phrase-models in the dynamic logic process as follows. Consider a simple example, let us limit the problem to learning phrases of a fixed length, say, we would like to learn 5-word phrases. Start with a large value of \( S > 5 \) (e.g. \( S = 50 \)), and with arbitrary word contents of phrases (6). In each iteration, compute Eq. (7) through (10). Reduce \( S \) by 1: in each phrase-model eliminate 1 word with the minimal \( \phi(s|h) \),

\[ w_{h,s'} \] is the least “probable” word in model \( h \) and it is eliminated on this iteration. \( S \) is changed to \( S-1 \). Continue iterations until \( S \) reaches the desired value 5.

The dynamic logic procedure defined in this way results in learning phrase-models and accomplishes the goal of language acquisition without combinatorial complexity. The computational complexity is moderate, \( \sim N^H K^S \). This overcoming of CC is the major accomplishment of this procedure.

The procedure outlined in this section is general in that it is applicable to all higher layers in the mind hierarchy and to cognitive as well as language models. Similarly to learning here phrase-models composed of words, at higher layers, composition of larger chunks of text from smaller chunks of texts can be learned. Similarly, cognitive models of higher layers, say images composed of objects can be learned. This procedure provides a mathematical foundation for perceptual symbol systems described in Barsalou (1999). In particular the mathematical procedure for his “simulators”, which, as he suggested, support abilities for abstract concepts, propositions, and productivity that is an ability to produce limitless structures from finite amount of basic components.

Lower layer models, may require continuous parametric models, like laryngeal models of phonemes (Lieberman, 2002). These can be learned from language sounds using procedures similar to Section 2.

An important property of the procedure described in this section, let me repeat, is productivity. This word is used to describe the ability of the human mind to use finite means (words, concepts)
to produce virtually infinite (combinatorially large) number of more complex structures. In linguistics, traditionally, productivity has been associated with recursion (Barsalou, 1999; Chomsky, 1995; Hauser, Chomsky, & Fitch, 2002; Pinker, 1994), while non-combinatorial mathematical procedures for acquisition-learning of this ability have not been demonstrated. Here we demonstrated a mathematical procedure for non-combinatorial learning, which results in combinatorially powerful productivity. Recursion is implemented through the hierarchy. This section considered in detail two-layer hierarchy, words–phrases; it could be extended in a straightforward way down to morphology in linguistics or feature-object representation in perception, as well as up to higher levels of language and cognition.

4. Cognition and language

Do we use phrases to label situations that we already have understood, or the other way around, do we just talk without understanding any cognitive meanings? It is obvious that different people have different cognitive and linguistic abilities and may tend to different poles in cognitive-language continuum, while most people are somewhere in the middle in using cognition to help with language, and vice versa. What are the neural mechanisms that enable this flexibility? How do we learn which words and objects come together? If there is no specific language module, as assumed by cognitivist linguists, why kids learn language by 5 or 7, but do not think like adults?

Virtually nothing is known about neural mechanisms integrating language and cognition. Here we propose a computational model that potentially can answer the above questions, and that is computationally tractable, it does not lead to combinatorial complexity. It is experimentally testable. Also it implies relatively simple neural mechanisms, and explains why human language and human cognition are inextricably linked. It suggests that human language and cognition have evolved jointly.

4.1. Dual model

We propose that integration of language and cognition is accomplished by a dual model. Every model in the human mind is not separately cognitive or linguistic, and still cognitive and linguistic contents are separate to a significant extent. Every concept-model \( M_h \) has two parts, linguistic \( M^l_h \) and cognitive \( M^c_h \):

\[
M_h = \{ M^l_h, M^c_h \}.
\]

A sensor data stream constantly comes into the mind from all sensory perceptions; every part of this data stream is constantly evaluated and associated with models (12) according to the mechanisms of dynamic logic described above. In a newborn mind both types of models are vague mostly empty placeholders for future cognitive and language contents. The neural connections between the two types of models are inborn; the mind never has to learn which word goes with which object. As models acquire specific contents in the process of growing up and learning, linguistic and cognitive contents are always staying properly connected.

During the first year, infants learn some objects and situations in the surrounding world, this means that cognitive parts of some models at the layer of objects and situations become less vague and acquire a degree of specificity. Language models at the layer of objects and above remain vague. After one year of age, language model adaptation speeds up; language models become less vague and more specific much faster than the corresponding cognitive models. This is especially true about contents of abstract models, which cannot be directly perceived by the senses, such as “law”, “state”, “rationality”. This explains how it is possible that kids by the age of five can talk about most of the contents of the surrounding culture but cannot function like adults: language models are acquired ready-made from the surrounding language, but cognitive models remain vague and gradually acquire concrete contents throughout life. This is the neural mechanism of what is colloquially called “acquiring experience”.

Human learning of cognitive models continues through the lifetime and is guided by language models. The knowledge instinct drives the human mind to develop more specific and concrete cognitive models by accumulating experience throughout life in correspondence with language models.

4.2. Experimental evidence, answers and questions

As mentioned, experimental evidence for the dual model is almost nil. The first experimental indication has appeared in Franklin et al. (2008). Those researchers demonstrated that categorical perception of color in prelinguistic infants is based in the right brain hemisphere. As language is acquired and access to lexical color codes becomes more automatic, categorical perception of color moves to the left hemisphere (between two and five years) and adult’s categorical perception of color is based in the left hemisphere (where language mechanisms are located).

These experiments have provided evidence for neural connections between perception and language, a foundation of the dual model. Possibly it confirms another aspect of the dual model: the crisp and conscious language part of the model hides from our consciousness vaguer cognitive part of the model. This is similar to what we observed in the closed–open eye experiment: with opened eyes we are not conscious about vague imagination–priming signals.

So, we can answer some of the questions posed at the beginning of the section. Language and cognition are separate and closely related mechanisms of the mind. They evolve jointly in ontological development and learning, and possibly these abilities evolved jointly in evolution—this we address in more detail in the next section. This joint evolution of dual models from vague to more crisp content resolves the puzzle of associationism: there is no need to learn correct associations among combinatorially large number of possible associations, words and objects are associated all the time while their concrete contents emerge in the mind.

Perception of objects that can be directly perceived by sensing might be to some extent independent from language, nevertheless, as above experimental data testify, even in these cases language affects what we perceive. In more complex cognition of abstract ideas, which cannot be directly perceived by the senses, we conclude that language parts of models are more crisp and conscious; language models guide the development of the content of cognitive models. Language models also tend to hide cognitive contents from consciousness. It follows that in everyday life most thinking is accomplished by using language models, possibly with little engagement of cognitive contents.

We know that thinking by using cognitive contents is possible, for example when playing chess. Mathematics is another example, but not necessarily a good one, because mathematics uses its own “language” of mathematical notations. Do mathematical notations play a similar role to language in shaping cognitive contents and thinking? My guess would be that this is the case for a C student of mathematics, but creative thinking in mathematics and in any other endeavor engages cognitive models. Needless to say, this requires special abilities and significant effort. Possibly fusiform gyrus plays a role in cognition shadowed by language. More detailed discussion of possible brain regions involved in the knowledge instinct are discussed in Levin and Perlovsky (2008). This is a vast field for experimental psychological and neuro-imaging investigations.
4.3. Dual hierarchy

Previous sections focused mostly on processes at a single layer in the mind hierarchy. Contents of lower layers in the hierarchy are perceptual elements, objects; higher up are relationships among objects, situations, more and more abstract and general model-concepts etc., and near the top are the most general concepts of the purpose and meaning of life (Levin & Perlovsky, 2008; Perlovsky, 2006a, 2006b). At every layer there is a similarity measure defining the knowledge instinct, models, emotions, and actions, including adaptation of models. An input to each layer is a set of signals $X(n)$. The result of signal processing at a given layer are 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 or lower layers. The activated models also send their activation signals to the next higher layer.

The dual model implies two parallel heterarchies of language and cognition, as illustrated in Fig. 2. This architecture along with dynamic logic equations in previous sections solve an amazing mystery of the human mind, which we are so used to that it almost never has been even formulated as requiring an explanation.

Models at higher layers are composed of lower layer models (say scenes are composed of objects). In parallel, language is used to describe scenes linguistically with sentences composed of words. Words–object relations at lower layers are preserved at higher layers of phrase–scene relations. This holds true across tremendous number of the phrase–situation layer models, using various combinations of the same words from the lower layer. This amazing property of our mind seems so obvious, that nontrivial complexity of the required mechanism has only been noticed once (Deacon, 1997).

Deacon also suggested that the hierarchy sets the human mind apart from the animal world. Every human culture possesses both abilities; and there is no species that has either language or cognition at the human level. Here we discuss mathematical reasons why hierarchy can only exist as a joint dual hierarchy of language and cognition. This dual hierarchy architecture gives a mathematical reason for this fact. Only at the lower layers in the hierarchy cognitive models can be learned by direct perception of the world. Learning is grounded in “real” objects. At higher levels, however, learning of cognitive models has no ground. In artificial intelligence it was long recognized that learning without grounding could easily go wrong, learned or invented models may correspond to nothing real or useful (Meytel & Albus, 2001).

The mechanism of the dual model sets the human mind apart from the rest of the animal world. Consider an example of a dog learning to bring shoes to a human master on a verbal command. A dog, it seems, can jointly learn language and cognition (a word “shoes” and an object shoes). Does that mean that a dog possesses a dual model? No. The dog can do it, because it perceives an object (shoes) in the world. Learning a word “shoes” is grounded in direct perception of object-sound “shoes”. Such a direct grounding in sensory signals exists only at the very bottom of the mind hierarchy. At higher layers, cognitive concepts are grounded in language concepts due to the dual models, and language models, as discussed, are grounded in talking with other people and in mutual understanding. Using the dual models, the knowledge instinct drives the mind to acquire cognitive models corresponding to language models (colloquially, “experience”).

The fact that the cognitive hierarchy cannot be learned without language hierarchy is so fundamental and under-appreciated that I would give another explanation for this reason in different words. Consider learning situations on top of already learned object perception. When deciding which set of objects constitutes a concept-model that should be learned and remembered, one would encounter a situation such as follows: entering a room one sees a desk, a chair, books, shelves...and a barely seen scratch on a wall. Is this scratch on a wall as important as other objects? Is the fact that a red-color book is on the left from a blue-color one important and should be learned as a separate situation-model? In fact there are much more insignificant objects and infinity of their combinations, such as scratches, dust, relative positions, etc., than important objects and their combinations of significance for any situation. No one would have enough experience in a lifetime to learn which objects and in which combinations are important and which are not for each situation. Only from language do we learn what is typical and important for understanding various situations. Even more complicated is learning of abstract concepts, which cannot be perceived by the senses directly. Many linguists beginning from Chomsky have under-appreciated this fact and developed linguistic theories assuming that theories of cognitive meaning should be handed down by professors from cognitive departments.

Language hierarchy is acquired by the human mind from the surrounding ready-made language. Learning language hierarchy at all layers is grounded in communication with other people around; people talk to and understand each other. This provides grounding for language learning. Try to teach a dog to understand the word “rational”, or any abstract concept, whose meaning is based on several hierarchical layers; this is not possible. It is known that the smartest chimps after long training can barely understand few concepts at the second layer (Savage-Rumbaugh & Lewine, 1994).

4.4. Cognitive linguistics and dynamic logic

The proposed theory shares many assumptions with cognitive linguistics, including the embodied nature of cognition and language. As emphasized by Barsalou (1999), this idea belongs to a tradition that goes back more than two millennia. Section 3 describes just a first step toward mathematics of cognitive linguistics, which should be extended toward evolution of semantics and grammar in terms of experiential constructs. This development would utilize accounts developed by cognitive linguistics, and would provide mathematical foundations for these developments.

The interaction between cognition and language proposed here, while being just the first step, resolves long-standing problems of cognitive linguistics. Jackendoff (1983) suggested that

“the meaning of a word can be exhaustively decomposed into finite set of conditions...necessary and sufficient...”.

The very language used in this quote exposes the logical way of thinking, which leads to wrong conclusions. Meanings of words do not reside in other words, but in words relations to real world situations. According to the mechanism suggested in this paper, meanings reside in cognitive parts of the dual models.

Gradually, cognitive linguistics moved away from a strictly compositional view of language. Lakoff (1988) emphasized that abstract concepts used by the mind for understanding the world have metaphorical structure. Metaphors are not just poetic tools, but a mind mechanism for creating new abstract meanings. Lakoff’s analysis brought this cultural knowledge of the role of metaphorical thinking within the mainstream of science. There was still a big gap between Lakoff’s analysis of metaphors on the one hand and neural and mathematical mechanisms on the other. The “Metaphors we live by” is a metaphorical book (the pun is intended) in that it begs the question: Who is that homunculus in the mind, interpreting the metaphorical theater of the mind? What are the mechanisms of metaphorical thinking? According to the current paper, a metaphor extends an old understanding to the new meaning by extending a vague cognitive model to a crisp
language model and making it vaguer; this is followed by dynamic logic creation of several more specific new language and cognitive models.

In works of Jackendoff (1983), Langacker (1988), Talmay (1988) and other cognitive linguists it was recognized that dichotomies of meanings (semantic–pragmatic) and dichotomies of hierarchical structures (superordinate–subordinate) were limiting scientific discourse and have to be overcome. Consider the following opinions on meaning creation:

“in a hierarchical structure of meaning determination the superordinate concept is a necessary condition for the subordinate one...COLOR is a necessary condition for determining the meaning of RED” (Jackendoff, 1983).

“The base of predication is nothing more than...domains which the prediction actually invokes and requires” (Langacker, 1988).

These judgments illustrate difficulties encountered when attempting to overcome old dichotomies. Logical intuitions guide these judgments and limit their usefulness. Attempts to implement mathematically mechanisms assumed by these examples would lead to combinatorial complexity. Problems of meaning and hierarchy still remind the old question about the chicken and the egg, what came first? If superordinate concepts come before subordinate ones, where do they come from? Are we born with the concept COLOR in our minds? If predications invoke domains, where do domains come from? These complex questions with millenarian pedigrees are answered mathematically in this paper: Hierarchy and meaning are emerging jointly with cognition and language. In evolution and individual learning, superordinate concepts (COLOR) are vaguer, less specific, and less conscious than subordinate ones (RED). RED can be vividly perceived, but COLOR cannot be perceived. RED can be perceived by animals. But, the concept COLOR can only emerge in the human mind, due to joint operation of language and cognition.

Jackendoff in his recent research (2002) concentrated on unifying language and cognition. He developed detailed models for such unification; however, his logical structures face combinatorial complexity.

Lakoff and Johnson (1999) brought within the realm of linguistics an emphasis on embodiment of the mind. The implication that the philosophical tradition will have to be reassessed, however, seems exaggerated. Recent synthesis of computational, cognitive, neural, and philosophical theories of the mind demonstrated the opposite (Perlovsky, 2000). Plato, Aristotle, and Kant, even in specific details about the mind mechanisms, were closer to contemporary computational theories, than the 20th century philosophers and mathematicians developing logical formalism and positivism.

Talmay (2000) introduced a notion of open and closed classes of linguistic forms. Open class includes most words, which could be added to language as needed, say, by borrowing from other languages. Closed class includes most grammatical structures, which are fixed for generations and cannot be easily borrowed from other languages. This point to an important aspect of interaction between language and cognition. Forms of the closed class interact with cognitive concepts, which emerged over thousands of years of cultural and language evolution. Thus, for each individual mind and for entire generations, which operate within constraints of existing grammar, many cognitive concepts are predetermined by language. Talmay identified cognitive concepts affected by closed forms. These forms are more basic for cognition than words and unconsciously influence entire cultures.

Kay (2002) proposed construction grammar (a direction closely associated with cognitive linguistics) to accommodate metaphorical and idiomatic linguistic constructions. Within generative-Chomsky-type grammar these could only be understood by rejecting word–phrase distinction and adopting a word–phrase continuum, which is of course impossible as it would lead to combinatorial complexity. These constructions, he suggested, cannot exist in a meaning free grammar; their construction and understanding require a combination of semantic and linguistic knowledge. Still, existing proposals for construction grammar formalism are logical and combinatorially complex. The dual model, instead, provides a necessary mechanism—cognitive and linguistic models act jointly.

Another proposal to combine cognition and language is given by Fauconnier and Turner (2008). They proposed what they called a double-scope blending, without however a specific mechanism to accomplish this. Its function is similar to the dual hierarchy proposed in this paper.

Dual hierarchy supports the cognitive linguistic idea that syntax is not a separate inborn “box” in the mind, but is a conceptual mechanism. Specifics of syntax according to this paper are encoded in concept–model contents at higher layers of phrases and sentences. It is possible that given a mechanism of the dual model, the hierarchy would evolve and syntax would be learned from surrounding language. To which extent syntax reflects structures in the world that could be directly learned along with language and encoded in cognitive and language models? In addition to the dual model, what other linguistic knowledge must be inborn? What is the role of dynamic logic and the dual model in morphology? Is it possible that the dual model is the only mechanism that is required to enable language and to set us aside from animals. Demonstrating this in simulations is a challenge for future research. If this could be demonstrated, Occam’s Razor would favor the dual model.

4.5. Evolutionary linguistics and dynamic logic

Evolutionary linguistics emphasizes that language properties evolved in the process of cultural evolution of languages (Christiansen & Kirby, 2003). Only those properties of languages survive that can be passed from generation to generation. Christiansen and Chater (2008) discussed how various linguistic phenomena are explained within the evolutionary framework. Brighton et al. (2005) demonstrated in mathematical simulation that evolutionary approach can explain the emergence of language compositionality. Compositionality, the language ability to construct words of sounds, phrases of words etc., is a fundamental language universal, unique to human languages. Brighton et al.’s research is especially elegant in that most simple assumptions were required, first an ability to learn statistically from limited examples which sounds go with which cognitive meanings, and second, the fact that training samples are insufficient and agents have to guess sounds for new meanings; so meanings that are similar in a certain way to the old ones are designated by sounds similar in some ways to the old sounds. This has led to compositionality.

Unfortunately, most work in evolutionary linguistics so far assumed that cognitive meanings already exist. Another detrimental aspect of existing work is logical computational basis, leading to combinatorial complexity. Dynamic logic, if applied to Brighton at al’s formulation, leads to non-combinatorial complexity of learning and production. If combined with the dual model, it does not require an assumption that the meanings have already existed; a result is the joint learning of combinatorial language and cognitive meanings (Fontanari, Tikhanoff, Cangelosi, & Perlovsky, 2009). The mathematical formulation in this paper leads to a joint evolution of language and cognitive meanings.

4.6. Contents of language faculty

Hauser et al. (2002) emphasized that “language is, fundamentally, a system of sound–meaning connections”. This connection is accomplished by a language faculty, which generates internal representations and maps them into the sensory–motor interface, and into the conceptual–intentional interface. In this way sound and meaning are connected. They emphasized that the most important property of this mechanism is recursion. However, they did not propose specific mechanisms how recursion creates represen-
tations, nor how it maps representations into the sensory–motor or conceptual–intentional interfaces.

A conclusion of the current paper is that it might not be necessary to postulate recursion as a fundamental property of a language faculty. In terms of the mathematical model of modeling fields and dynamic logic proposed here, recursion is accomplished by the hierarchy; a higher layer generates lower layer models, which accomplish recursive functions. We have demonstrated that the dual model is a necessary condition for the hierarchy of cognitive representations. It also might be a sufficient one. It is expected that the hierarchy is not a separate inborn mechanism; the hierarchy might emerge in operations of the dual model and dynamic logic in a society of interacting agents with intergenerational communications (along the lines of Brighton et al. (2005)). What exactly are the inborn precursors necessary for the hierarchy's ontological emergence, if any, is a challenge for the ongoing research. Anyway, reformulating the property of recursion in terms of a hierarchy, along with demonstrating that a hierarchy requires the dual model, the paper has suggested a new explanation that a single neurally simple mechanism is unique for human language and cognitive abilities. Initial experimental evidence indicates a support for the dual model; still further experiments elucidating properties of the dual model are needed.

Another conclusion of this paper is that the mechanism mapping between linguistic and cognitive representations is accomplished by the dual models. In previous sections we considered mathematical modeling of the “conceptual–intentional interface” for intentionality given by the knowledge and language instincts; in other words we considered only intentionalities related to language and knowledge. It would not be principally difficult to add other types of intentional drives following Grossberg and Levine (1987). The current paper has not considered the “sensory–motor interface”, which is of course essential for language production and hearing. This can be accomplished by the same mechanism of the dual model, with addition of behavioral and sensorial models. This task is not trivial; still it does not present principal mathematical difficulties.

We would also like to challenge an established view that specific vocalization is “arbitrary in terms of its association with a particular context”. In animals, vocalization is not voluntary (if at all) as humans'. Evolution of language required separation of conceptual, semantic contents from emotional ones, and from involuntary vocalizations. As language was evolving to less emotional, more semantic ability that differentiated psyche, another part of primordial vocalization was evolving toward more emotional, less semantic ability that maintained a unity of psyche. This ability evolved into music (Perlovsky, 2008). Cultures with semantically rich languages also evolved emotionally rich music (Perlovsky, 2006d, 2007). Songs affect us by unifying semantic contents of lyrics with emotional contents of music, which are perceived by ancient emotional brain centers. The same mechanism still exists to a lesser degree in languages; all existing languages retain some degree of emotionality in their sounds. Therefore languages, while evolving amodal symbol abilities, still retain their vocal modalities, and otherwise they will not be able to support the process of cognitive and cultural evolution. Symbolic structures of the mind are mathematically described by dynamic logic; the mind symbols are not static categories, but the dynamic logic processes. Due to the dual model they combine cognition and language, and can approximate amodal properties. The dual model suggests experimentally testable neural mechanisms combining two modal (perceptual, voice) symbol systems with the amodal ability.

6. Future research

The proposed mechanism of the dual model implies a relatively minimal neural change from the animal to the human mind. It possibly has emerged through combined cultural and genetic evolution and this cultural evolution most likely continues today. Dynamic logic resolves a long-standing mystery of how human language, thinking, and culture could have evolved in a seemingly single big step, too large for an evolutionary mutation, too fast and involving too many advances in language, thinking, and culture, happening almost momentarily around 50,000 years ago (Deacon, 1997; Mithen, 1998). Dynamic logic along with the dual model explains how changes, which seem to involve improbable steps according to logical intuition, actually occur through continuous dynamics. The proposed theory provides a mathematical basis for concurrent emergence of hierarchical human language and cognition. The dual model is directly testable experimentally. The next steps would further develop this theory in multi-agent simulations, leading to more specific and experimentally testable predictions.

Classic ideas of frames and schema were demonstrated to face combinatorial complexity, and this is the fundamental reason for their failures despite decades of development. The proposed theory produces conceptual relations, binding, and recursion through the mechanisms of the hierarchy; in addition it explains how language acquires meanings in connecting with cognition due to the mechanism of the dual model. These predictions are experimentally testable and several groups of psychologists and neuroscientists are working in these directions.

Evolutionary linguistics and cognitive science have to face a challenge of studying and documenting how primordial fused
model differentiated into several significantly independent mechanisms. In animal minds emotion–motivation, conceptual understanding, and behavior-voicing have been undifferentiated unity, their differentiation is a hallmark of human evolution. Was this a single step, or could evolutionary anthropology document several steps, when different parts of the model differentiated from the primordial whole?

The paper solved several principled mathematical problems, which involved combinatorial complexity, when using previously considered mechanisms inspired by logical intuition. Initial neuro-imaging evidence supports dynamic logic mechanism discussed in this paper, still much remains unknown, a wide field is ready for theoretical and experimental research.

Dynamic logic was experimentally demonstrated for perception of a single object; these experiments should be extended to perception of multiple objects in complex context, as well as for higher level cognition. Possibly experiments could probe more details of top-down and bottom-up signal interactions.

The paper challenges the idea that recursion is the main fundamental mechanism setting human language apart from animal abilities. It is proposed instead that recursion is accomplished by the hierarchy. The fundamental mechanism enabling the hierarchy, recursion, and connection of language and cognition is the dual model.

Detailed experimental evidence is needed for the dual model, especially the proposed development of cognitive models at higher layers throughout life under the guidance of language models.

The paper also challenges the idea of arbitrariness of vocalization. It is suggested that a significant degree of arbitrariness in current languages is a distal result of millennia of language evolution in presence of the dual model. Instead of assuming arbitrariness as fundamental, future research should concentrate on its emergence from primordial fusion of sound and emotion. This paper suggests that origins of language are contemporaneous with origins of music; primordial undifferentiated vocalizations split into two types: first, more semantic and less emotional mechanisms of language, and second, less semantic and more emotional mechanisms of music. Whereas language differentiates consciousness, music unifies conscious and unconscious into a whole self. It is my hope that a wide field of experimental research directions opened by this discussion will be explored (certain experimental programs are already being planned).

Ontological emergence of the hierarchy should be modeled mathematically and demonstrated experimentally; this could be done in conventional psychological experiments and in neuro-imaging.

Lower hierarchical layers, below words and objects, should be developed theoretically, or better to say, dynamic logic should be integrated with ongoing development in this area (see Guenther (2006)).

Mathematical simulations of the proposed mechanisms should be extended to engineering developments of Internet search engines with elements of language understanding. The next step would be developing interactive environments, where computers will interact among themselves and with people, gradually evolving human language and cognitive abilities. Developing intelligent computers and understanding the mind would continue to enrich each other.

To conclude, I would emphasize that the proposed approach of neural modeling fields, dynamic logic, and the dual model is a step toward unification of basic approaches to knowledge, its evolution, and learning in modern cognitive and neural sciences. These include classic representational approaches based on logic, rule systems, and amodal symbols; statistical pattern recognition and learning; embodied approaches of classic empiricism, situated actions, and cognitive linguistics; evolutionary linguistics; connectionism and neural networks. This step toward unification outlines a wide field for future theoretical and experimental research.

References
