Computational Foundations for Perceptual Symbol System

Leonid Perlovsky and Roman Ilin

Abstract

We describe computational foundations for Perceptual Symbol System (PSS). This requires new mathematical methods of dynamic logic (DL), which have overcome limitations of classical artificial intelligence and connectionist approaches. The paper discusses these past limitations, relates them to combinatorial complexity (exponential explosion) of algorithms in the past, and further to the static nature of classical logic. The new mathematical theory, DL, is a process-logic. We apply it to one aspect of PSS, learning of situations from object perceptions. We relate DL to essential PSS mechanisms of concepts, simulators, grounding. Experimental neuroimaging evidence for DL and PSS in brain imaging is discussed as well as future research directions.

1. Introduction. PSS, challenge of computational model

In a series of papers Barsalou developed a theory of perceptual symbol system (PSS), which grounds cognition in perception. “Grounded cognition… rejects the standard view that amodal symbols represent knowledge in semantic memory” (Barsalou 1999). This publication emphasized the roles of simulation in cognition. “Simulation is the reenactment of perceptual, motor, and introspective states acquired during experience with the world, body, and mind… when knowledge is needed to represent a category (e.g., chair), multimodal representations captured during experiences … are reactivated to simulate how the brain represented perception, action, and introspection associated with it.” Simulation is an essential computational mechanism in the brain. The best known case of these simulation mechanisms is mental imagery (e.g., Kosslyn 1980; 1994); other forms of grounded cognition include situated actions, social and environmental interaction (e.g., Barsalou 2003a; Barsalou et al. 2007; Yeh & Barsalou 2006). We would emphasize that imagery is a subset of simulation; it includes various sensory-motor and emotional signals, and its dynamic aspect in PSS is usually not available to consciousness. According to PSS cognition supports action. Simulation is a central mechanism of PSS, yet rarely, if ever, they recreate full experiences. Using the mechanism of simulators, which approximately correspond to concepts and types in amodal theories, PSS implements the standard symbolic functions of type-token binding, inference, productivity, recursion, and propositions. Using these mechanisms PSS retains the symbolic functionality. “Thus, PSS is a synthetic approach that integrates traditional theories with grounded theories.” (Barsalou 1999; 2005; 2007).

Grounded cognition includes cognitive linguistics theories (Lakoff & Johnson 1980; 1999; Boroditsky & Ramscar 2002; Langacker 1987; 1991; Talmy 1983; 1988; Fauconnier 1985; Tomasello 2003; and others); theories of situated action (Gibson 1979; Clark 1997; Breazeal 2002; Barsalou et al. 2007; etc.); grounding cognition, memories, actions, language, and symbols (Glenberg 1997; Glenberg & Kaschak 2002; Coventry & Garrod 2004; Pecher & Zwaan 2005);

and cognitive simulation theories, in particular PSS (Barsalou 1999), on which we concentrate in this paper.

According to Barsalou, during the Cognitive Revolution in the middle of the last century, cognitive scientists were inspired by new forms of representation "based on developments in logic, linguistics, statistics, and computer science." They adopted amodal representations, including feature lists, semantic networks, and frames (Barsalou & Hale 1993). Little empirical evidence supports amodal symbolic mechanisms (Barsalou 1999). It seems, amodal symbols were adopted largely because they promised to provide "elegant and powerful formalisms for representing knowledge, because they captured important intuitions about the symbolic character of cognition, and because they could be implemented in artificial intelligence." As we discuss in the next section, these promises were unfulfilled, they faced fundamental mathematical difficulties.

"Clearly an important goal for future theory is to implement and formalize" PSS. For initial attempts to formalize grounded theories mathematically see (Cangelosi et al. 2000; Joyce, Richards, Cangelosi, & Coventry 2003; Cangelosi & Riga 2006) and references therein. Computational models for PSS (Barsalou 1999; 2003a; 2007) require new mathematical methods different from traditional artificial intelligence, pattern recognition, or connectionist methods. The reason is that traditional methods encountered combinatorial complexity (CC), an irresolvable computational difficulty, when attempting to model complex systems, like mind, which required learning combinations of perceptual features and objects (Perlovsky 1994; 1997; 2000; 2002; 2006b; 2007a). Developing realistic and scalable formalization is a goal of this article. Although the developed mathematical formalism is quite general, here we concentrate on just one example of PSS mechanism: a mathematical description of models and simulators for forming and enacting representations of situations (higher level symbols) from perceptions of objects (lower level symbols).

2. Overcoming past mathematical difficulties

Simple object perception involves bottom-up signals from sensory organs and top-down signals from internal mind's representations (memories) of objects. During perception, the mind matches subsets of bottom-up signals corresponding to objects with representations of object in the mind. This produces object recognition; it activates brain signals leading to mental and behavioral responses (Grossberg 1982; Kosslyn 1994; Bar et al.. 2006; Schacter & Addis 2007).

2.1. Computational complexity since the 1950s

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 during the past fifty years. These difficulties were summarized under the notion of combinatorial complexity (CC) (Perlovsky 1998b). 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 computationally 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 of 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 training examples often was combinatorially large. Training had to include not only every object in its multiple variations, angles, etc., but also combinations of objects. Thus, self-learning approaches encountered *CC of learning requirements*.

Rule systems were proposed in the 1970’s to solve the problem of learning complexity (Minsky 1975; Winston 1984). Minsky suggested that learning was a premature step in artificial intelligence; Newton “learned” Newtonian laws, most of scientists read them in the books. Therefore Minsky has suggested, knowledge ought to be input in computers “ready made” for all situations and artificial intelligence would apply these known rules. Rules would capture the required knowledge and eliminate a need for learning. Chomsky’s original ideas concerning mechanisms of language grammar related to deep structure (Chomsky 1972) were also based on logical rules. Rule systems work well when all aspects of the problem can be predetermined. However in the presence of variability, the number of rules grew; rules became contingent on other rules and combinations of rules had to be considered. The rule systems encountered *CC of rules*.

In the 1980s, model systems were proposed to combine advantages of learning and rules-models by using adaptive models (Nevatia & Binford 1977; Bonnisone et al. 1991; Perlovsky 1987; 1988; 1991; 1994). Existing knowledge was to be encapsulated in models and unknown aspects of concrete situations were to be described by adaptive parameters. Along similar lines went the *principles and parameters* idea of Chomsky (1981). Fitting models to data (top-down to bottom-up signals) required selecting data subsets corresponding to various models. The number of subsets, however, is combinatorially large. A general popular algorithm for fitting models to the data, multiple hypothesis testing (Singer, Sea, & Housewright 1974) is known to face CC of computations. Model-based approaches encountered *computational CC* (N and NP complete algorithms).

2.2. Logic, CC, and amodal symbols.

In subsequent research, CC was related to logic underlying various algorithms and neural networks (Perlovsky 1996a; 2000; Marchal 2005). Logic serves as a foundation for many approaches to cognition and linguistics; it underlies most of computational algorithms. But its influence extends far beyond, affecting cognitive scientists, psychologists, and linguists, who do not use complex mathematical algorithms for modeling the mind. All of us operate under more than 2000 year old influence of logic, under a more or less conscious assumption that in the basis of the mind there are mechanisms of logic. As discussed in section 6, our minds are unconscious about its illogical foundations; we are mostly conscious about a small part of the minds mechanisms, which are approximately logical. Our intuitions, therefore, are unconsciously affected by bias toward logic. When laboratory data drive our thinking away from logical mechanisms, it is difficult to overcome the logical bias.

Relationships between logic, cognition, and language have been a source of longstanding controversy. Aristotle assumed a close relationship between logic and language. He emphasized that logical statements should not be formulated too strictly and language inherently contains the necessary degree of precision. According to Aristotle, logic serves to communicate already made decisions (Perlovsky 2007). The mechanism of the mind relating language, cognition, and the world Aristotle described as *forms*. Today we call similar mechanisms internal representations, or concepts, or simulators in the mind. Aristotelian forms are similar to Plato's ideas with a marked distinction, forms are *dynamic*: their initial states, before learning, are different from their final states of concepts (1995 / IV BCE). Aristotle emphasized that initial states of forms, forms-as-potentialities, are not logical (i.e. vague), but their final forms, forms-as-actualities, attained in the result of learning, are logical. This fundamental idea was lost during millennia of philosophical arguments. As discussed below this Aristotelian process of dynamic forms corresponds to Barsalou idea of PSS simulators, and in this paper we develop mathematics (dynamic logic) for this process.

The founders of formal logic emphasized a contradiction between logic with its law of excluded middle and language with its uncertainty. In the 19th century George Boole and great logicians following him, including Gottlob Frege, Georg Cantor, David Hilbert, and Bertrand Russell (see Davis 2000, and references therein) eliminated uncertainty of language from mathematics, and founded formal mathematical logic based on the "law of excluded middle." Hilbert developed an approach named formalism, which rejected intuition as a matter of scientific investigation and formally defined scientific objects in terms of axioms or rules. In 1900 he formulated famous Entscheidungsproblem: to define a set of logical rules sufficient to prove all past and future mathematical theorems. This entailed formalization of the entire human thinking and language. Formal logic ignored the dynamic nature of Aristotelian forms and rejected uncertainty of language. Hilbert was sure that his logical theory 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). However, Hilbert's vision of formalism explaining mysteries of the human mind came to end in the 1930s, when Gödel (1932/1994) proved internal inconsistency of formal logic. This is a reason why theories of cognition and language based on formal logic are inherently flawed.

There is a close relation between logic and CC. It turned out that combinatorial complexity of algorithms is a finite-system manifestation of the Gödel's theory (Perlovsky 1996a). If Gödelian theory is applied to finite systems, CC is the result, instead of the fundamental inconsistency. According to the law of excluded middle every statement is either true or false and nothing in between. Therefore, algorithms based on formal logic have to evaluate every variation in sensory signals or the mind's representations as a separate logical statement. A large number of combinations of these variations cause CC.

This general statement manifests in various types of algorithms in different ways. Rule systems are logical in a straightforward way, and the number of rules grows combinatorially. Pattern recognition algorithms and neural networks are related to logic in learning procedures: every training sample is treated as a logical statement ("this is a chair") resulting in CC of learning. Multivalued logic and fuzzy logic were proposed to overcome limitations related to the law of excluded middle (Zadeh 1965; Kecman 2001). Yet the mathematics of multivalued logic is no different in principle from formal logic with "excluded third" substituted by "excluded n+1." Fuzzy logic uses logic to set a degree of fuzziness. Correspondingly, it encounters a difficulty related to the degree of fuzziness: if too much fuzziness is specified, the solution does not achieve a needed accuracy, and if too little, it becomes similar to formal logic. If logic is used to find the appropriate fuzziness for every model at every processing step, then the result is CC. The mind has to make concrete decisions, for example one either enters a room or does not; this requires a computational procedure to move from a fuzzy state to a concrete one. But fuzzy logic does not have a formal procedure for this purpose; fuzzy systems treat this decision on an ad-hoc basis.

Is logic still possible after Gödel? The contemporary state of this field was reviewed in (Marchal 2005). It appears that logic after Gödel is much more complicated and much less logical than was assumed by founders of artificial intelligence. CC cannot be solved within logic. Penrose thought that Gödel's results entail incomputability of the mind processes and testify for a need for new physics (Penrose 1994). An opposite position in (Perlovsky 2000; 2006a;c; 2006b) is that incomputability of logic does not entail incomputability of the mind. These publications argue that logic is not the basic mechanism of the mind.

To summarize, various manifestations of CC are all related to formal logic and Gödel theory. Rule systems rely on formal logic in a most direct way. Even mathematical approaches specifically designed to counter limitations of logic, such as fuzzy logic and the second wave of neural networks (developed after the 1980s) rely on logic at some algorithmic steps. Self-learning algorithms and neural networks rely on logic in their training or learning procedures: every training example is treated as a separate logical statement. Fuzzy logic systems rely on logic for setting degrees of fuzziness. CC of mathematical approaches to the mind is related to the fundamental inconsistency of logic. Therefore logical inspirations, leading early cognitive scientists to amodal brain mechanisms, could not realize their hopes for mathematical models of the brain-mind.

Why did the outstanding mathematicians of the 19th and early 20th c. believed in logic to be the foundation of the mind? Even more surprising is the belief in logic after Gödel. Gödelian theory was long recognized among most fundamental mathematical results of the 20th c. How is it possible that outstanding minds, including founders of artificial intelligence, and many cognitive scientists and philosophers of mind insisted that logic and amodal symbols implementing logic in the mind are adequate and sufficient? The answer, in our opinion, might be in the "conscious bias." As we discuss in section 6, non-logical operations making up more than 99.9% of the mind functioning are not accessible to consciousness (Bar et al. 2006). However, our consciousness functions in a way that makes us unaware of this. In subjective consciousness we usually experience mind as logical. Our intuitions are “consciously biased.” This is why amodal logical symbols, which describe a tiny fraction of the mind mechanisms, have seemed to many the foundation of the mind.

Another aspect of logic relevant to PSS is that it lacks dynamics; it is about static statements such as “this is a chair.” Classical logic is good at modeling structured statements and relations, yet it misses the dynamics of the mind and faces CC. The essentially dynamic nature of the mind is not represented in mathematical foundations of logic. Dynamic logic discussed in the next section is a logic-process. It overcomes CC by automatically choosing the appropriate degree of fuzziness for every mind’s concept at every moment. DL combines advantages of logical structure and connectionist dynamics. This dynamics mathematically represents the learning process of Aristotelian forms (which are different from classical logic as discussed in this section) and serves as a foundation for PSS concepts and simulators.

2.3. Dynamic logic-process

DL describes perception as an interaction between bottom-up and top-down signals (Perlovsky 2000; 2006b; 2007c). This section concentrates on an initial relation of brain processes to mathematics of DL. To concentrate on this relationship, we much simplify discussion of the brain structures. We discuss visual recognition of objects as if retina and visual cortex is a single processing layer of neurons where recognition occurs (which is not true, detailed relationship of the DL process to brain is considered in given references). Perception consists in association-matching of bottom-up and top-down signals. Sources of top-down signals are representations-models; in perception processes models are modified-learned and new models are formed; since an object is never encountered exactly the same as previously, perception is always a learning process. The DL processes along with concept-representations are mathematical models of the PSS simulators. Bottom-up signals, in this simplified discussion, are a field of neuronal synapse activations in visual cortex. Sources of top-down signals are representation-concepts or, equivalently, model-simulators (for short, models). Each model projects a set of priming, top-down signals, representing the bottom-up signals expected from a particular object. Models depend on parameters. Parameters characterize object position, angles, lightings, etc. (In case of learning *situations* considered later, parameters characterize objects and relations making up a situation.) To summarize this highly simplified description of a visual system, the learning-perception process "matches" top-down and bottom-up activations by selecting "best" models and their parameters and the corresponding sets of bottom-up signals.

Mathematical measures of the "best" fit between bottom-up and top-down signals were given in (Perlovsky 2000; 2006b). They are similar to probabilistic or informatics measures. In the first case they represent probabilities that the given (observed) data come from representations-models of particular objects. In the second case they represent information contained in representations-models about the observed data. These similarities are maximized over the model parameters. Result can be interpreted correspondingly as a maximum probability (likelihood) that models-representations fit sensory signals, or as maximum information in models-representations about the signals. Both similarity measures account for all expected models and for *all combinations* of signals and models. Correspondingly, a similarity contains a large number of items, a total of M^N, where M is a number of models and N is a number of signals; this huge number is the cause for the combinatorial complexity discussed previously.

Maximization of a similarity measure is a mathematical model of an unconditional drive to improve the correspondence between bottom-up and top-down signals (representations-models). In biology and psychology it was discussed as curiosity, cognitive dissonance, or a need for knowledge since the 1950s (Harlow 1950; Festinger 1957; Cacioppo & Petty 1982). This process involves learning-related emotions evaluating satisfaction of this drive for knowledge (Grossberg & Levine 1987; Perlovsky 2006b; 2007c) In computational intelligence it is even more ubiquitous, every mathematical learning procedure, algorithm, or neural network maximizes some similarity measure. In the process of learning, 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 discarded.

The DL learning process, let us repeat, consists in estimating model parameters and associating subsets of signals with concepts by maximizing a similarity. Although a similarity contains combinatorially many items, DL maximizes it without combinatorial complexity (Perlovsky 1996b; 1997; 2000; 2007c) as follows. First, fuzzy association variables are defined, which give a measure of correspondence between each signal and each model. They are defined similarly to the a posteriori Bayes probabilities, they range between 0 and 1, and as a result of learning they converge to the probabilities, under certain conditions. Often they are close to bell-shapes.

The DL process is defined by a set of differential equations given in the above references; together with models it gives a mathematical description of the PSS simulators. To keep the paper self-consistent we summarize these equations in Appendix 1. As a model of perception-cognitive processes, DL models PSS and not amodal signals, DL also explains how illogical dynamic PSS gives rise of classical logic in the human mind, and what is the role of amodal symbols. This is discussed in details in section 5.

The DL model of simulator-processes avoids combinatorial complexity because there is no need to consider separately various combinations of bottom-up and top-down signals. Instead, all combinations are accounted for in the DL simulator-processes. Initially, models do not match data; association variables are wide and vague, they take near homogeneous values across the data, associating all representation-models (through simulator processes) with all input signals (Perlovsky 2000; 2006b). The DL simulator-processes improve matching, models better fit data, errors become smaller, bell-shapes concentrate around relevant patterns in the data (objects), and the association variables tend to 1 for correctly matched signal patterns and models, and 0 for others; thus certain representations get associated with certain subsets of signals (objects are recognized and concepts formed). This process "from vague-to-crisp" that matches bottom-up and top-down signals has been independently conceived and demonstrated in brain imaging research to take place in human visual system (Bar, Kassam, Ghuman, Boshyan, Schmid, Dale, Hamalainen, Marinkovic, Schacter, Rosen, & Halgren 2006).

Mathematical convergence of the DL process was proven in (Perlovsky 2000). It follows that the simulator-process of perception or cognition forms objects or concepts among bottom-up signals, which are most similar in terms of the similarity measure. Despite a combinatorially large number of items in the similarity, a computational complexity of DL is relatively low, it is linear in the number of signals, and therefore could be implemented by a physical system, like a computer or brain.

2.4. Example of DL, object perception in noise

The purpose of this section is to illustrate DL simulator-process described above; therefore we use a simple example, still unsolvable by other methods, (mathematical details are omitted, they could be found in Linnehan, Mutz, Perlovsky, Weijers, Schindler, & Brockett 2003). In this example, DL searches for patterns in noise. Finding patterns below noise can 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 the locations and orientations of patterns are not known, it is not clear which subset of the data points should be selected for fitting. A standard approach for solving this kind of problem, which has already been mentioned, is multiple hypothesis testing (Singer, Sea, & Housewright 1974); this algorithm exhaustively searches all combinations of subsets and models and faces combinatorial complexity. In the current example, we are looking for 'smile' and 'frown' patterns in noise shown in Fig.1a without noise, and in Fig.1b with noise, as actually measured (object signals are about 2-3 times below noise). The image size in this example is 100x100 points (N = 10,000 bottom-up signals, corresponding to the number of receptors in an eye retina), and the true number of models is 4 (3+noise), which is not known. Therefore, at least M = 5 models should be fit to the data, to decide that 4 fits best. This yields complexity of $M^N = 10^{5000}$; this number is much larger than the size of the Universe and the problem was considered unsolvable. Fig. 1 illustrates DL operations: (a) true ‘smile’ and ‘frown’ patterns without noise; (b) actual image available for recognition; (c) through (h) illustrates the DL process, it show improved models at various steps of solving DL eqs.(3), total of 22 steps. In terms of signal-to-noise ratio it is 10,000% improvement over the previous state-of-the-art. (In this example DL actually works better than human visual system; the reason is that human brain is not optimized for recognizing these types of patterns in noise). The main point of this example is that DL simulator-perception is a process “from vague-to-crisp,” similar to visual system processes demonstrated in (Bar et al. 2006).

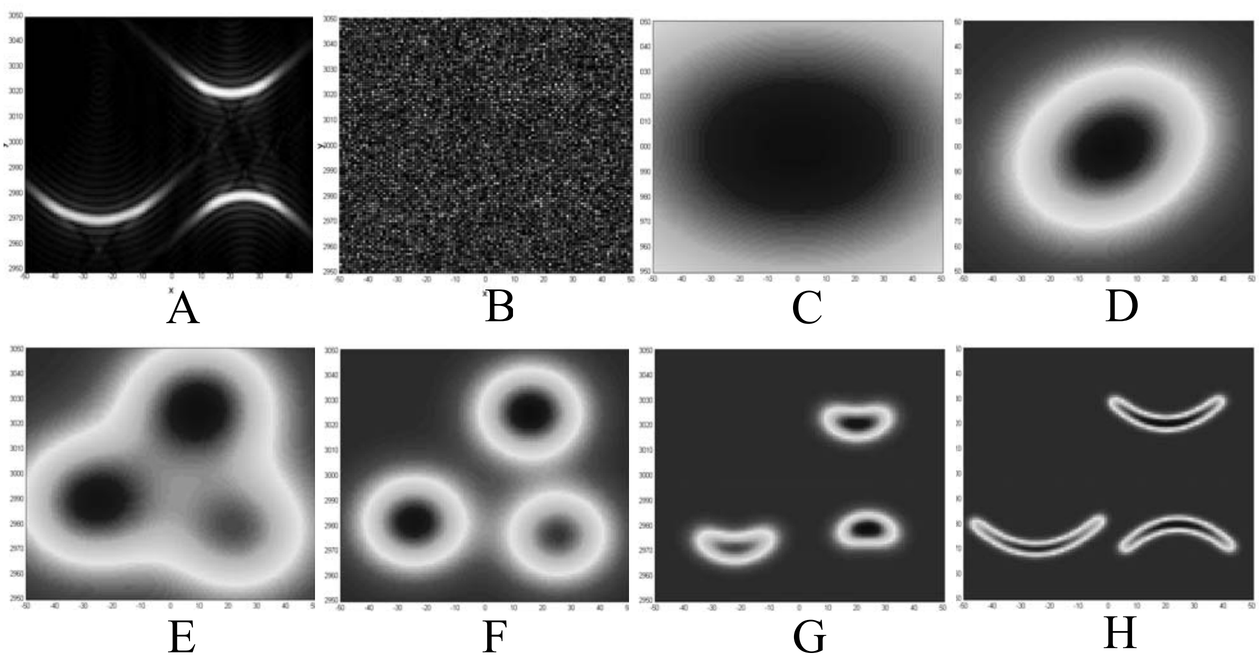

Fig.1. Finding 'smile' and 'frown' patterns in noise, an example of dynamic logic operation: (a) true 'smile' and 'frown' patterns are shown without noise; (b) actual image available for recognition (signals are below noise, signal-to-noise ratio is between ½ and ¼, 100 times lower than usually considered necessary); (c) an initial fuzzy blob-model, the vagueness corresponds to uncertainty of knowledge; (d) through (h) show improved models at various steps of DL (eq.(3) are solved in 22 steps). Between stages (d) and (e) the algorithm tried to fit the data with more than one model and decided, that it needs 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, 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 needs more complex parabolic models to describe the data. Iterations stopped at (h), when similarity (1) stopped increasing.

3. DL of PSS: perceptual cognition and simulators

Section 2.4 illustrated DL for recognition of simple objects in noise, a case complex and unsolvable for prior state-of-the-art algorithms, still too simple to be directly relevant for PSS. Here we consider a problem of situation learning, assuming that object recognition has been solved. In computational image recognition this is called "situational awareness" and it is a long-standing unsolved problem. The principled difficulty is that every situation includes many objects that are not essential to recognition of this specific situation; in fact there are many more "irrelevant" or "clutter" objects than relevant ones. Let us dwell on this for a bit. Objects are spatially-limited material things perceptible by senses. A situation is a collection of contextually related objects that tend to appear together and are perceived as meaningful, e.g., an office, a dining room. The requirement for contextual relations and *meanings* makes the problem mathematically difficult. Learning *contexts* comes along with learning situations; it reminds the problem of a chicken and egg. We subliminally perceive many objects, most of which are irrelevant, e.g. a tiny scratch on a wall, which we learn to ignore. Combinations of even a limited number of objects exceed what is possible to learn in a single lifetime as meaningful situations and contexts (e.g. books on a shelf) from random sets of irrelevant objects (e.g. a scratch on a wall, a book, and a pattern of tree branches in a window). Presence of hundreds irrelevant objects makes learning by a child of mundane situations a mathematical mystery. In addition, we constantly perceive large numbers of different objects and their combinations, which do not correspond to anything worth learning.

An essential part of learning-cognition is to learn which sets of objects are important for which situation (contexts). The key mathematical property of DL that made this solution possible, same as in the previous section, is a process "from vague-to-crisp." Concrete crisp models-representations of situations are formed from vague models in the process of learning (or cognition-perception). We illustrate here that complex symbols, situations, are formed by situation-simulators from simpler perceptions, objects, which are simpler perceptual symbols, being formed by simulators at "lower" levels of the mind, comparative to "higher" situation-simulators. Situation-simulators operate on PSS representations of situations, which are dynamic and vague assemblages of situations from imagery (and others) bits and pieces perceived at lower levels; this dynamic process of DL-PSS-simulation is mostly unconscious. We will discuss in details in section 5 that these are perceptual symbols as described in (Barsalou 1999). DL mathematically models PSS simulators (Barsalou 1999), processes that match bottom-up perceptions with top-down signals and form (assemble) symbols in cognition-perception. An essential mechanism of DL cognition-perception is a process of simulation of perceptual imagination-cognitions;

these situation-symbols are simulated from simpler perceptions-objects (imagination here is not limited to imagery, and is mostly unconscious). And the same mechanism can simulate plans and more complex abstract thoughts, although this is beyond the scope of the present paper. Thus we demonstrate that DL mathematically models PSS simulators.

3.1. DL formulation

In a simplified problem considered here, the task is for an intelligent agent (a child) to learn to recognize certain situations in the environment; while it is assumed that a child has learned to recognize objects. For example, situation "office" is characterized by the presence of a chair, a desk, a computer, a book, a book shelf. Situation "playground" is characterized by the presence of a slide, a sandbox, etc. The principal difficulty is that many irrelevant objects are present in every situation. (This child learning is no different mathematically from an adult recognition.)

In the example below, D_o is the total number of objects that the child can recognize in the world (it is a large number). In every situation he or she perceives D_p objects. This is a much smaller number compared to D_o. Each situation is also characterized by the presence of D_s objects essential for this situation ($D_s < D_p$). Normally nonessential objects are present and D_s is therefore less than D_p. The sets of essential objects for different situations may overlap, with some objects being essential to more than one situation. We assume that each object is encountered in the scene only once. This is a minor and nonessential simplification, e.g. we may consider a set of similar objects as a new object. For example, "book" is an object and "books" is another object referring to more than one book.

The real life learning is sequential as a child is exposed to situations one at a time. Again, DL can handle this, but in this paper we consider the data about all the situations available at the time of learning.

Following (Ilin & Perlovsky 2009) a situation can be mathematically represented as a vector in the space of all objects, $\mathbf{X}_n = (x_{n1}, \dots x_{ni}, \dots x_{nDo})$. If the value of x_{ni} is *one* the object i is present in the situation n and if x_{ni} is *zero*, the corresponding object is not present. Since D_o is a large number, $\mathbf{X}_n$ is a large binary vector with most of its elements equal to zero. A situation model is characterized by parameters, a vector of probabilities, $\mathbf{p}_m = (p_{m1},.. p_{mi},... p_{mDo})$. Here p_{mi} is the probability of object i being part of the situation m. Thus a situation model contains D_o unknown parameters. Estimating these parameters constitutes learning.

We model the elements of vector $\mathbf{p}_m$ as independent (this is not essential for learning, if presence of various objects in a situation actually is correlated, this would simplify learning, e.g. perfect correlation would make it trivial). Correspondingly, conditional probability of observing vector $\mathbf{X}_n$ in a situation *m* is then given by the standard formula (Jaynes 2003).

$$l(\mathbf{X}(n) \mid \mathbf{M}_m(n)) = \prod_{i=1}^{Do} p_{mi}^{x_{ni}}(1 - p_{mi})^{(1-x_{ni})}$$

Consider N perceptions a child was exposed to (N includes real “situations” and “irrelevant” random ones); among them most perceptions were “irrelevant” corresponding to observing random sets of objects, and M-1 “real” situations, in which D_s objects were repeatedly present.

All random observations we model by 1 model ("noise"); assuming that every object has an equal chance of being randomly observed in noise (which again is not essential) the probabilities for this noise model, m=1, are p_{1i}=0.5 for all *i*. Thus we define M possible sources for each of the N observed situations.

The total likelihood-similarity for our M models (M-1 "real" and 1 noise) is given by the same equation as similarity in the previous example (Perlovsky 2006b, also Appendix 1). And the same DL equations maximize it over the parameters, which in this case are the probabilities of objects constituting various situations.

For shortness, we did not discuss relations among objects. Spatial, temporal, or structural connections, such as "to the left," "on top," or "connected" can be easily added to the above DL formalism. Relations and corresponding markers (indicating which objects are related) are no different mathematically than objects, and can be considered as included in the above formulation. The formulation here assumes that all the objects have already been recognized, but the above formulation can be applied without any change to real, continuously working brain with multiplicity of concurrently running simulators at many levels, feeding each other. Also modality of objects (various sensorial or motor mechanisms) requires no modifications (emotions can be included as well, but emotions are not always reduced to representations and this requires a separate discussion beyond this paper). The bottom up signals do not have to be definitely recognized objects, these signals can be sent before objects are fully recognized, while object simulators are still running and object representations are vague; this would be represented by x_{ni} values between 0 and 1. The presented formalization therefore is a general mechanism.

3.2. Example of symbol-situation learning using DL

In this example we set the total number of recognizable objects equal to 1000 (D_o=1000). The total number of objects perceived in a situation is set to 50 (D_p=50). The number of essential objects is set to 10 (D_s=10). The number of situations to learn (M-1) is set to 10. Note that the true identities of the objects are not important in this simulation so we simply use object indexes varying from 1 to 1000. The situation names are also not important and we use situation indexes. We would emphasize that the use of numbers for objects and situation, while may seem consistent with amodal symbols, in fact is nothing but notations. The principled differences between PSS and amodal systems are mechanisms in the brain and their modeling, not mathematical notations. Among these mechanisms are simulators, mathematically described by DL. Let us repeat, amodal symbols are governed by classical logic, which is static, and faces CC. DL is a process and overcomes CC. DL operates on PSS representations (models $\mathbf{p}_m$), which are vague collections of objects (some of these objects could also be vague, not completely assembled yet representation, but this is beyond this paper). Another principled difference is interaction between perceptual-based bottom-up and top-down neural fields $\mathbf{X}_n$ and $\mathbf{M}_m$; indexes n and m are just mathematical shorthand for corresponding neural connections. In this paper we consider object perception and situation perception in different sections, but of course the real mind-brain operates continuously, "objects" in this section are neural signals sent to situation-recognition brain area (and corresponding simulators) by excited neuron fields corresponding to models of recognized-objects in section 2 (and as discussed, these signals are being sent before objects are fully recognized, while object simulators are still running).

The data for this example are generated by first randomly selecting D_s=10 specific objects for each of the 10 groups of objects, allowing some overlap between the groups (in terms of specific objects). This selection is accomplished by setting the corresponding probabilities p_{mi} = 1. Next we add 40 more randomly selected objects to each group (corresponding to D_p=50). We also generate 10 more random groups of 50 objects to model situations without specific objects (noise); this is of course equivalent to 1 group of 500 random objects. We generate N'=800 perceptions for each situation resulting in N=16,000 perceptions (data samples, n = 1… 16,000) each represented by 1,000-dimensional vector $\mathbf{X}_n$. These data are shown in Fig. 2 sorted by situations.

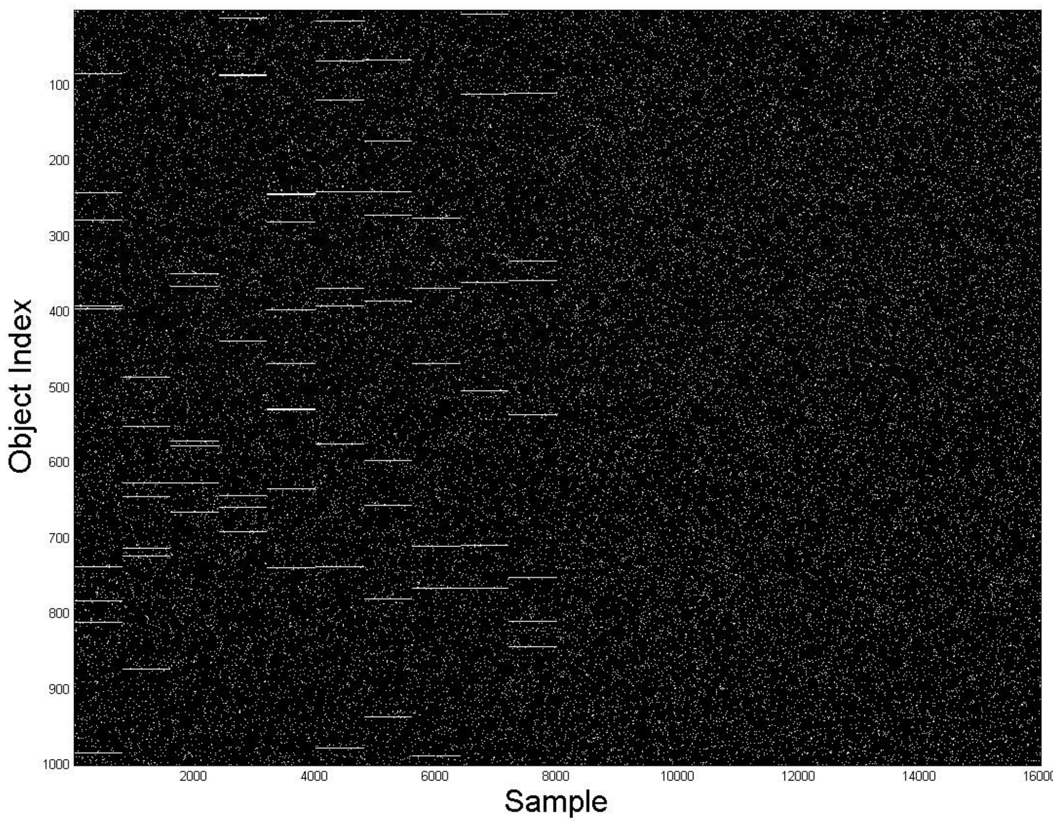

Fig. 2. Generated data; object index is along vertical axes and situation index is horizontal. The perceptions (data samples) are sorted by situation index (horizontal axis); this makes visible the horizontal lines for repeated objects.

The samples are randomly permuted, according to randomness of real life perceptual situations, in Fig. 3. The horizontal lines disappear; the identification of repeated objects becomes nontrivial. An attempt to learn groups-situations (the horizontal lines) by inspecting various horizontal sortings (until horizontal lines would become detectable) would require $M^N = 10^{16000}$ inspections, which is of course impossible.

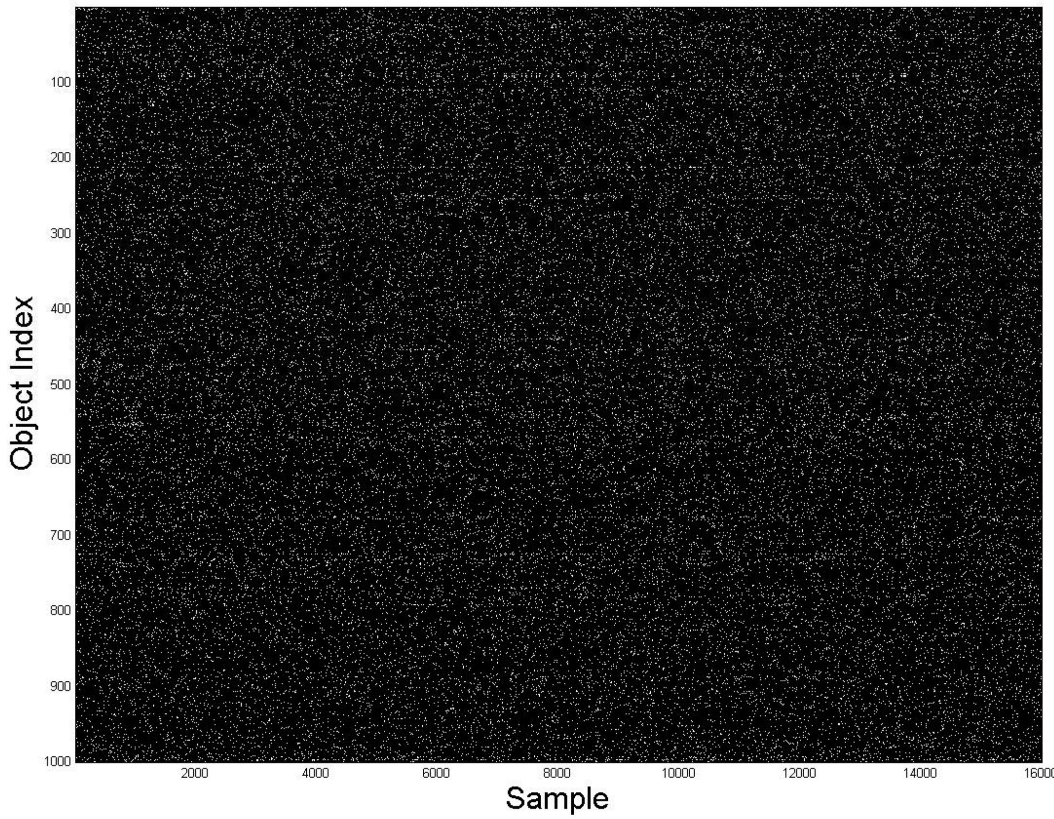

Fig. 3. Generated data, same as Fig. 2, randomly sorted by situation index (horizontal axis), as available to the DL algorithm for learning.

DL algorithm is initiated similarly to section 2 by defining 20 situational models (an arbitrary selection, given actual 10 situations) and one random noise model to give a total of M=21 models (in section 2.4, Fig.1 models were automatically added by DL as required; here we have not done this (mostly, because it would be too cumbersome to present results). The models are initialized by assigning random probability values to the elements of the models. These are the initial vague perceptual models which assign all objects to all situations.

Fig. 4 illustrates the initialization and the iterations of the DL algorithm (the first 3 steps of solving DL equations. Each subfigure displays the probability vector $\mathbf{p_m}$ for each of the 20 models. The vectors have 1000 elements corresponding to objects (vertical axes). The values of each vector element are shown in gray scale. The initial models assign nearly uniformly distributed probabilities to all objects. The horizontal axes are the model index changing from 1 to 20. The noise model is not shown. As the algorithm progresses, situation grouping improves, and only the elements corresponding to repeating objects in "real" situations keep their high values, the other elements take low values. By the third iteration the 10 situations are identified by their corresponding models. The other 10 models converge to more or less random low-probability vectors.

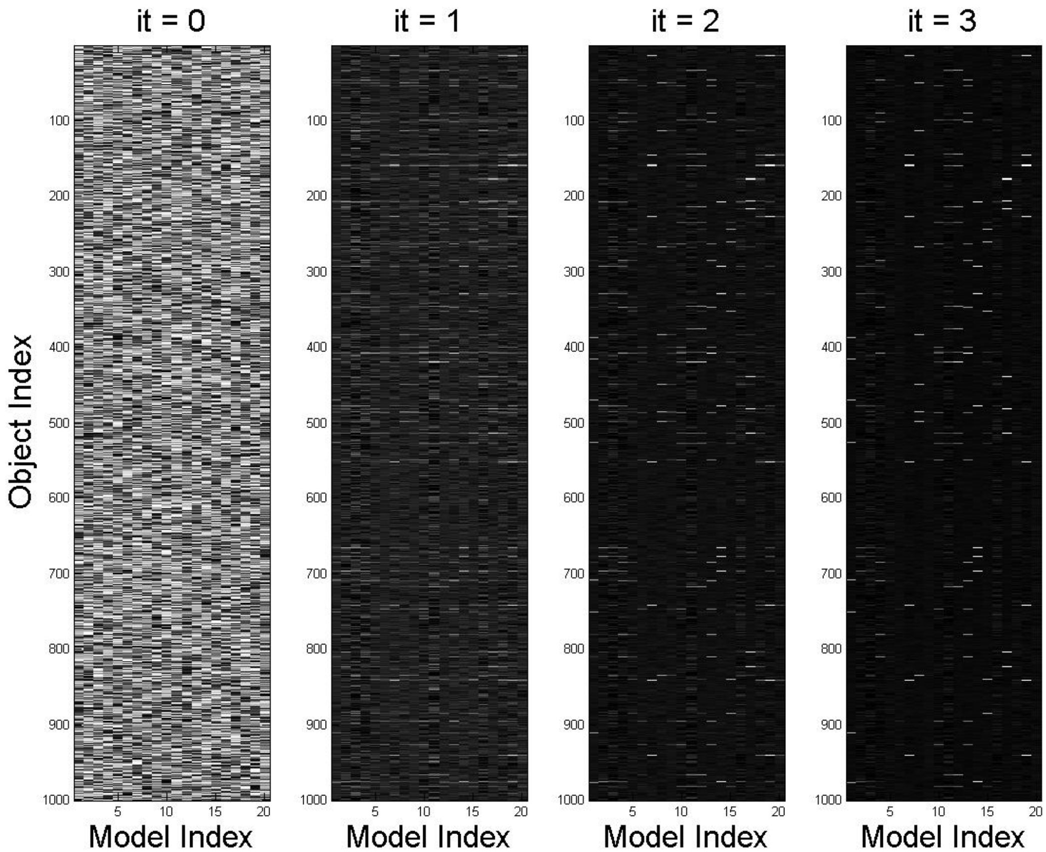

Fig. 4. DL situation learning. Situation-model parameters converge close to true values in 3 steps.

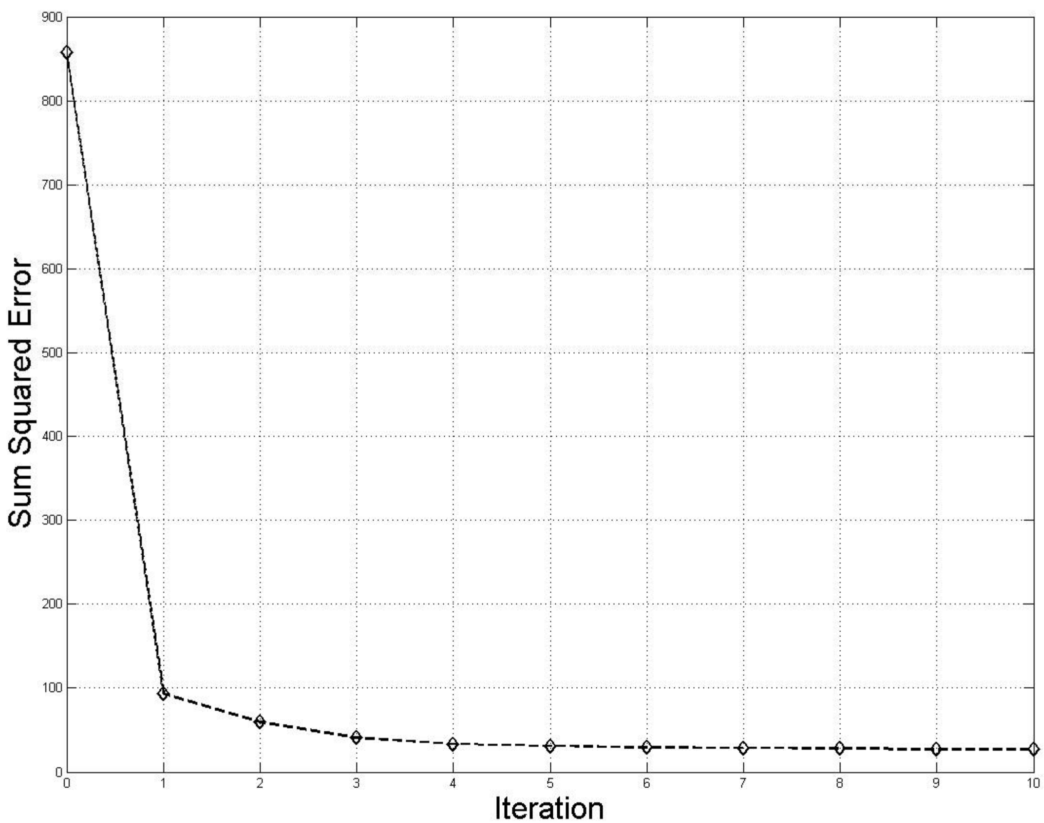

Fig. 5. Errors of DL learning are quickly reduced in 3-4 steps, iterations continue until average error reached low value of 0.05 (10 steps).

This fast and accurate convergence can be seen from Figs. 5 and 6. We measure the fitness of the models to the data by computing the sum squared error, using the following equation.

$$E = \sum_{m \in \{B\}} \sum_{i=1}^{Do} (p_{mi} - p_{mi}^{True})^2$$

In this equation the first summation is over the subset {B} containing top 10 models that provide the lowest error (and correspondingly, the best fit to the 10 true models). In the real brain, of course, the best models would be added as needed, and the random samples would accumulate in the noise model automatically; as mentioned, DL can model this process and the reason we did not model it, is that it would be too cumbersome to present results. Fig. 5 shows how the sum squared error changes over the iterations of the DL algorithm. It takes only a few iterations for the DL algorithm to converge. Each of the best models contains 10 large and 990 low probabilities. Iterations stop, when average error of probabilities reached a low value of 0.05 resulting in the final error E(10) =1000*(0.05^2)*10 = 25.

Fig. 6 shows average associations, A(m,m') among true (m) and computed models (m'); this is an 11x11 matrix according to the true number of different models (it is computed using association variables between models and data, f(m|n))

$$A(m,m') = (1/N') \sum_{n=1}^{N} f(m|n)* f(m'|n),\ m' \in \{B\},$$

$$A(m,11) = (1/10*N') \sum_{m' \notin \{B\}} \sum_{n=1}^{N} f(m|n)* f(m'|n),\ m' \notin \{B\}$$

Here, f(m|n) for true 10 models m is either 1 (for N' data samples from this model) or 0 (for others), f(m'|n) are computed associations, in the second line all 10 computed noise models are averaged together, corresponding to one true (random) noise model. The correct associations on the main diagonal in Fig. 6 are 1 (except noise model, which is spread among 10 computed noise models, and therefore equals 0.1) and off-diagonal elements are near 0 (incorrect associations, corresponding to small errors shown in Fig 5.)

Again, as in section 2, learning of perceptual situation-symbols has been accomplished due to DL process-simulator, which simulated internal model-representations of situations, **M**, to match patterns in bottom-up signals **X** (sets of lower-level perceptual object-symbols).

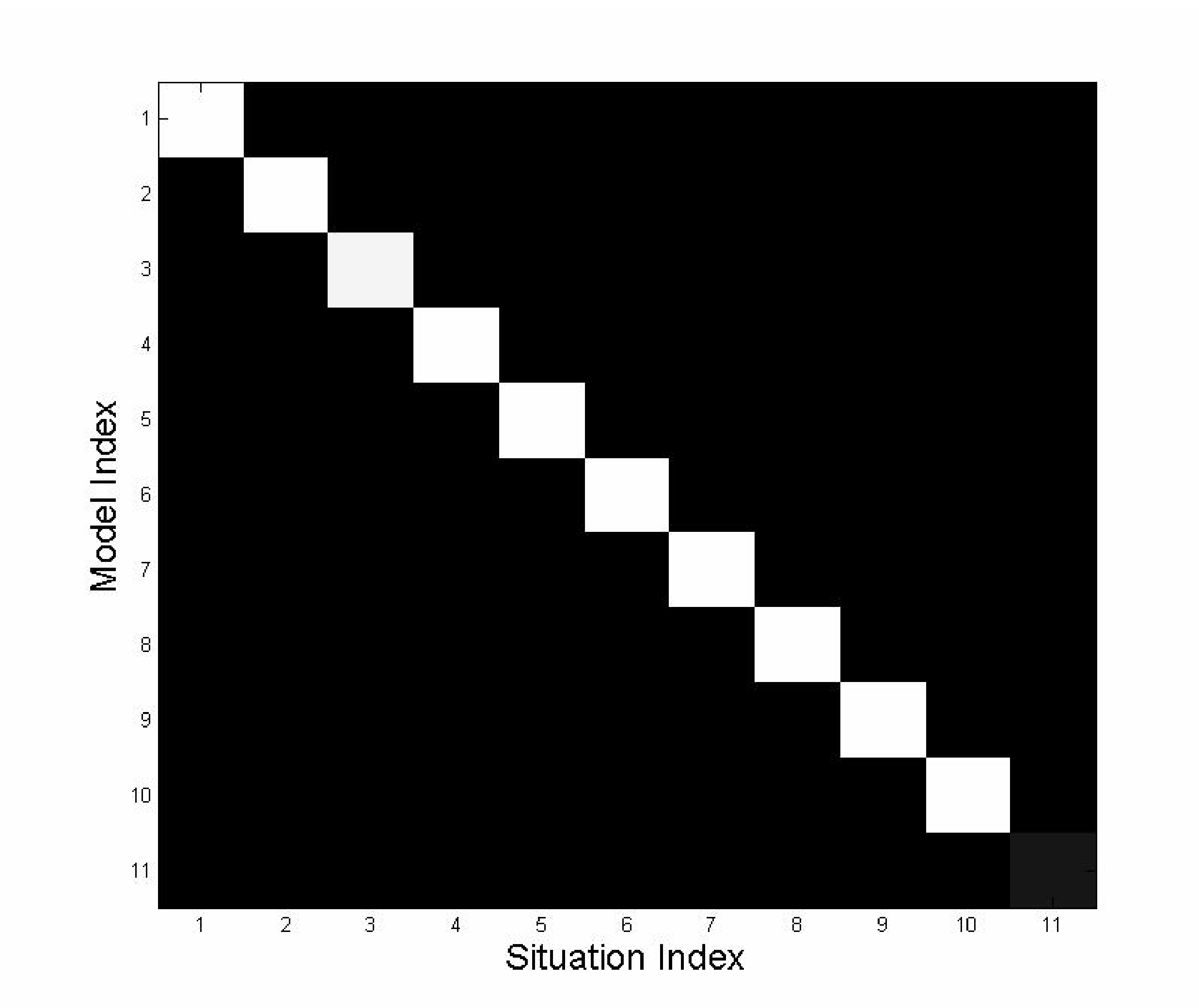

Fig. 6. Correct associations are near 1 (diagonal, except noise) and incorrect associations are near 0 (off-diagonal).

4. Simulators, concepts, grounding, binding, and DL

As described previously Barsalou argued for PSS, which grounded perception, cognition, and high-level symbol operation in modal symbols, which are ultimately grounded in the corresponding brain systems. Previous section provides an initial "first step" toward developing formal mathematical description suitable for PSS. We have considered just one subsystem of PSS, a mechanism of learning, formation, and recognition of situations from objects making up the situations. The mind's representations of situations are symbol-concepts of a higher level of abstractness than symbol-objects making them up. The proposed mathematical formalism can be advanced straightforwardly to "higher" levels of more and more abstract concepts. We would add a word of caution: such application to more abstract ideas may require an additional grounding in language (Perlovsky 2007b; 2009a); however this discussion is beyond the scope of the paper. Similarly, the proposed mathematical formalism can be applied at a lower level of recognizing objects as constructed from their parts; mathematical techniques of sections 2 and 3 can be combined to implement this PSS object recognition idea as described in (Barsalou 1999). The proposed theory mathematically models *productivity* of the mind concept-simulator system.

Modeling situations in PSS as a step toward general solution of the binding problem is discussed in (Edelman & Breen 1999). DL provides a general approach to the binding problem similar to the "corkboard" approach described in (Edelman & Intrator 2001). This publication also discusses the role of context similar to the DL scene modeling.

Below we discuss relationships between mathematical DL procedures of previous sections and fundamental ideas of PSS. Section 2 concentrated on the principal mathematical difficulty experienced by all previous attempts to solve the problem of complex symbol formation from less complex symbols, combinatorial complexity (CC). CC was resolved by using DL, a mathematical theory, in which learning begins with vague (non-specific) symbol-concepts, and in the process of learning symbol-concepts are becoming concrete and specific. Learning could refer to a child's learning, which might take days or months or an everyday perception and cognition, taking 1/6th of a second (in the later case learning refers to the fact that every specific realization of a concept in the world is different in some respects from any previous occurrences, therefore learning-adaptation is always required). In the considered case of learning situations as compositions of objects, the initial vague state of each situation-symbol is a nearly random and vague collection of objects, while the final learned situation consists of a crisp collection of few objects specific to this situation. In the learning process random irrelevant objects are "filtered out," their probability of belonging to a concept-situation is reduced to zero, while probabilities of relevant objects, making up a specific situation is increased to a value characteristic of this object being actually present in this situation. Relation of this DL process to PSS is now considered.

First we address concepts and their development in the brain. According to (Barsalou 2007),

> "The central innovation of PSS theory is its ability to implement concepts and their interpretative functions using image content as basic building blocks."

This aspect of PSS theory is implemented in DL in a most straightforward way. Concept-situations in DL are collections of objects (symbol-models at lower levels, which are neurally connected to neural fields of object-images). Considering objects to be perceptual entities-symbols in the brain, concept-situations are collections of perceptual symbols. In this way situations are perceptual symbols of a higher order complexity than object-symbols, they are grounded in perceptual object-symbols (images), and in addition, their learning is grounded in perception of images of situations.

A PSS mathematical formalization of abstract concepts (Barsalou 2003b), not grounded in direct perceptions, will be considered in subsequent publications. Here we just mention that the proposed model is applicable to higher levels, "beyond" object-situations; it is also applicable to language, including syntax, and to language-cognition interaction (see also Perlovsky 2004; 2006d; 2009a; Fontanari & Perlovsky 2007; 2008a;b; Fontanari, Tikhanoff, Cangelosi, Ilin, & Perlovsky 2009). Also, in section 3 we did not consider *relations* among objects specifically; nevertheless our DL formalization does not exclude relations from a list of objects. A complete consideration in future will include relations and markers, indicating which objects and in which way are related; these relations, markers and their learning are not different mathematically from objects (in case of language, relations and markers would address syntax).

Barsalou (2008) has described development of concepts in the brain as forming collections of *correlated features.* This is explicitly implemented in the DL process described in section 3. This mathematical representation corresponds to multimodal and distributed representation in the brain. A mathematical set or collection is implemented in the brain by a population of conjunctive neurons (Simmons & Barsalou 2003). Specific relations of various parts of neural fields in this paper and brain systems will be addressed in future research along with modeling brain operations on more comprehensive scale.

DL learning and perception-cognition processes are mathematical models of PSS simulators. DL symbol-situations are not static collections of objects but dynamic processes. In the process of learning they "interpret individuals as tokens of the type" (Barsalou 2008). They can model multi-modal distributed representations (including motor programs) as described in the reference.

The same DL mathematical procedure can apply to perception of a real situation in the world as well as an imagined situation in the mind. This is the essence of imagination. Models of situations (probabilities of various objects belonging to a situation, and objects attributes, such their locations) can depend on time, in this way they are parts of simulators accomplishing cognition of situations evolving in time. If "situations" and "time" are invoked "with closed eyes" and pertain to the mind's imaginations, the simulators implement imagination-thinking process, or planning.

Usually we perceive-understand a surrounding situation, while at the same time thinking and planning future actions and imagine consequences. This corresponds to running multiple simulators in parallel. Some simulators support perception-cognition of the surrounding situations as well as ongoing actions, they are mathematically modeled by DL processes that converged to matching internal representations (types) to specific subsets in external sensor signals (tokens). Other simulators simulate imagined situations and actions related to perceptions, cognitions, and actions, produce plans, etc.

Developed here DL modeling of PSS models mathematically what Barsalou (2003b) called dynamic interpretation of PSS (DIPSS). DIPSS is fundamental to modeling abstraction processes in PSS. Three central properties of these abstractions are type–token interpretation; structured representation; and dynamic realization. Traditional theories of representation based on logic model interpretation and structure well but are not sufficiently dynamical. Conversely, connectionist theories are dynamic but are inadequate at modeling structure. PSS addresses all three properties. Similarly, the DL mathematical process developed here addresses all three properties. In type–token relations "propositions are abstractions for properties, objects, events, relations and so forth. After a concept has been abstracted from experience, its summary representation supports the later interpretation of experience." Correspondingly in the developed mathematical approach, DL models a situation as a loose collection of objects. Its summary representation (the initial model) evolves-simulates representation of a concrete situation in the process of perception of this concrete situation. A loose collection of property and relation simulators evolves according to DL to represent abstractions. This DL process involves structure (initial model) and dynamics (the DL process). Examples of DL-PSS for symbolic predication, conceptual combinations, and abstract concepts will be addressed in future publications.

5. Perceptual vs. amodal symbols

Since any mathematical notation may look like an amodal symbol, in this section we discuss the roles of amodal vs. perceptual systems in DL. This would require clarification of the word *symbol*. We touch on related philosophical and semiotic discussions and relate them to mathematics of DL and to PSS. For the sake of brevity within this paper we limit discussions of general interest, emphasizing connections between DL, perceptual, and amodal symbols; extended discussions of symbols can be found in (Perlovsky 2006b;d). We also summarize here related discussions scattered throughout the paper.

"Symbol is the most misused word in our culture" (Deacon, 1998). Why the word "symbol" is used in such a different way: to denote trivial objects, like traffic signs or mathematical notations, and also to denote objects affecting entire cultures over millennia, like Magen David, Swastika, Cross, or Crescent? Let us compare in this regard opinions of two founders of contemporary semiotics, Charles Peirce (Peirce 1897; 1903) and Ferdinand De Saussure (1916). Peirce classified signs into symbols, indexes, and icons. Icons have meanings due to resemblance to the signified (objects, situations, etc.), indexes have meanings by direct connection to the signified, and *symbols* have meaning due to arbitrary conventional agreements. Saussure used *different* terminology, he emphasized that *signs* receive meanings due to arbitrary conventions, whereas *symbol* implies motivation. It was important for him that motivation contradicted arbitrariness. Peirce concentrated on the process of sign interpretation, which he conceived as a triadic relationship of sign, object, and interpretant. Interpretant is similar to what we call today a representation of the object in the mind. However, this emphasis on interpretation was lost in the following generation of scientists. This process of "interpretation" seems close to DL processes and PSS simulators. We therefore follow Saussurean designation of symbol as a motivated process.

Motivationally-loaded interpretation of symbols was also proposed by Jung (1921). He considered symbols as processes bringing unconscious contents to consciousness. Similar are roles of PSS simulators and DL processes.

In the development of scientific understanding of symbols and semiotics, the two functions, *understanding language and understanding world*, have often been perceived as identical. This tendency was strengthened by considering logic to be the mechanism of both, language and cognition. According to Russell (1919), language is equivalent to axiomatic logic, "[a word-name] merely to indicate what we are speaking about; [it] is no part of the fact asserted… it is merely part of the symbolism by which we express our thought". Hilbert (1928) was sure that his logical theory also describes 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." Similarly, logical positivism centered on "the elimination of metaphysics through the logical analysis of language" – according to Carnap (1959) logic was sufficient for the analysis of language. As discussed in section 2.2, this belief in logic is related to functioning of human mind, which is *conscious* about the final states of DL processes and PSS simulators, which are perceived by our minds as approximately logical amodal symbols. Therefore we identify amodal symbols with these final static logical states, signs.

DL and PSS explain how the mind constructs symbols, which have psychological values and are not reducible to arbitrary logical amodal signs, yet are intimately related to them. In section 4 we have considered objects as learned and fixed. This way of modeling objects indeed is amenable to interpreting them as amodal symbols. Yet, we have to remember that these are but final states of previous simulator processes, perceptual symbols. Every perceptual symbol-simulator has a finite dynamic life, and then it becomes a static symbol-sign. It could be stored in memory, or participate in initiating new dynamical perceptual symbols-simulators. This infinite ongoing dynamics of the mind-brain ties together static signs and dynamic symbols. It grounds symbol processes in perceptual signals that originate them; in turn, when symbol-processes rich their finite static states-signs, these become perceptually grounded in symbols that created them. We could become consciously aware of static sign-states, express them in language and operate with them logically. Then, outside of the mind-brain dynamics, they could be transformed into amodal logical signs, like marks on a paper. Dynamic processes – symbols-simulators are usually not available to consciousness. These PSS processes involving static and dynamic states are mathematically modeled by DL in section 4.

To summarize, DL does not model amodal symbols, which are governed by classical logic that leads to combinatorial complexity. DL operates on different type of PSS representations, which are vague combinations of lower-level representations. These lower-level representations could include memory states as well as vague dynamic states from concurrently running simulators – DL processes of the on-going perception-cognition. (We do not discuss in this paper the nature of memory, which may involve various working memories as well as permanent memories with their own simulators and dynamics). To the extent that the mind-brain is not a strict hierarchy, the same-level and higher-level representations could be involved along with lower levels. Thus DL models processes-simulators, which operate on PSS representations. These representations are vague and DL processes are assembling and concretizing these representations. As described in several references by Barsalou, bits and pieces from which these representations are assembled, could include mental imagery as well as other components, including multiple sensor, motory, and emotional modalities (these details are beyond this paper); these bits and pieces all are mostly inaccessible to consciousness during the process dynamics. DL also explains how logic and ability to operate amodal symbols originate from illogical operations of PSS.

The developed DL formalization of PSS, suggests using a word *signs* for amodal static logical constructs outside of the mind, including mathematical notations; and to reserve *symbols* for perceptually grounded motivational cognitive processes in the mind-brain. Memory states, to the extent they are understood as static entities, are signs in this terminology. Logical statements and mathematical signs are perceived and cognized due to PSS simulator processes and after being understood become signs. Perceptual symbols, through simulator processes, tie together static and dynamic states in the mind. Dynamic states are mostly outside of consciousness, while static states might be available to consciousness.

6. Experimental evidence

Bar et al. (2006) demonstrated in neuroimaging experiments that visual perception proceeds according to DL simulating crisp perceptions from initial vague representations. Experimental procedures in this reference 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*, (the authors in this publication refer to low spatial-frequency content images). 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.

The brain continuously extracts rudimentary information from early sensory data and simulates predictions, which facilitate perception and cognition in the relevant context by pre-sensitizing relevant representations. This includes predictions of complex information, such as situations and social interactions (Bar 2007; 2009). Predictions are initiated by gist information rapidly extracted from sensory data. At the "lower"-object level this gist information is a vague image of an object (low spatial frequency, Bar et al. 2006). At higher levels "the representation of gist information" is yet to be defined." The model developed here defines this higher-level gist information as vague collections of vague objects, with relevant objects for a specific situation having just slightly higher probabilities than irrelevant ones. The developed model is also consistent with the hypothesis in (Bar 2007) that perception and cognition at higher levels relies on mental simulations. Mathematical predictions in this paper suggest specific properties of these higher-level simulators, which could be verified experimentally.

7. Future research

Future research will address the DL mathematical description of PSS throughout the mind hierarchy; from features and objects "below situations" in the hierarchy to abstract models and simulators at higher levels "above situations." Modeling across the mind modalities will be addressed including diverse modalities, symbolic functions, conceptual combinations, predication. Modeling features and objects would have to account for suggestions that perception of features are partly inborn (Barsalou 1999); this development therefore might require new experimental data on which feature aspects are inborn (Edelman & Newell, 1998). The developed DL formalization of PSS corresponds to observations in (Wu & Barsalou 2009) and it will be used for generating detailed experimentally verifiable predictions. The DL formalism developed here has 2 hierarchical levels, objects and situations, and it demonstrates bindings within these two levels. In future hierarchical extension of the DL binding will be related to hierarchy as a general mathematical principle (Edelman 2003). Similarly, the *recursive* property of cognition and language will be modeled as a mathematical hierarchy.

Experimental research (Bar et al. 2006; Bar 2007) can address specific properties of higher level simulators predicted here. Among these is a prediction that early predictive stages of situation simulations are vague. Whereas vague predictions of objects resemble low-spatial frequency of object imagery (Bar et al. 2006), "the representation of gist information on higher levels of analysis is yet to be defined" (Bar 2007). According to the developed model, vague predictions of situations should contain many less-relevant (and likely vague) objects with lower probabilities. Since the mathematical model proposed here is applicable to higher levels ("above" object-situations), this hypothesis should be relevant to the nature of information of higher level gists.

The present model can be expanded to address another topic discussed in (Bar 2007), "how the brain integrates and holds simultaneously information from multiple points in time." Two different mechanisms should be explored: first, explicit incorporation of time into models (so that model parameters-probabilities depend on time), and second, categorized temporal relations, such as "before," "after" can be included as any other relations-objects into models. A joint mathematical-experimental approach might be fruitful in this area.

Future research will address interaction between language and cognition. Since language is acquired from surrounding language, rather than from direct experience, language is a candidate system in the mind that is closer aligned with amodal symbols than with perceptual symbols. Kids at 5 years of age can talk about much of cultural content of the surrounding language, including highly abstract contents; yet, clearly kids do not have necessary experience to understand highly abstract concepts, as perceptual symbols, and to relate them to the world. A mathematical model of language-cognition interaction was proposed in (Perlovsky 2009a). It suggested that higher abstract concepts could be stronger grounded in language than in perception; not only kids, but also adults may operate with abstract concepts as with amodal symbols, and therefore have limited understanding grounded in experience of how abstract concepts relate to the world. It is possible that higher level concepts may be less grounded in perception and experience than in language. Future experimental research should address this hypothesis.

As mentioned, the developed model is applicable to language; correspondingly, it should bear on language evolution, and future research should address combining a language extension of this model with models of language evolution (Brighton, Smith, & Kirby 2005; Perlovsky & Fontanari 2006; Fontanari & Perlovsky 2007; 2008a;b; Fontanari, Tikhanoff, Cangelosi, Ilin, & Perlovsky 2009; Perlovsky 2009b).

The role of emotions in perception was addressed in (Barrett & Bar 2009). There are several mechanisms of emotions and future research should extend this paper formalism to modeling emotions. Also future research would explore the role of emotions in language-cognition interaction (Perlovsky 2009b), and the role of emotions in symbol grounding.

REFERENCES

Aristotle. (1995). *The complete works*. The revised Oxford translation, ed. J. Barnes, Princeton, NJ: Princeton Univ. Press. (Original work IV BCE)

Bar, M. (2009). Predictions: A universal principle in the operation of the human brain (Introduction). Theme issue: Predictions in the brain: Using our past to generate a future (M. Bar Ed.) *Philosophical Transactions R. Soc. B 364*, 1181–1182.

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences, 11(7),* 281-289.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, et al.. (2006). Top-down facilitation of visual recognition. USA: *Proceedings of the National Academy of Sciences, 103,* 449-54.

Barrett, L. F., & Bar, M. (2009). See it with feeling: Affective predictions in the human brain. *Royal Society Phil Trans B, 364*, 1325-1334.

Barsalou, L. W. (1999). Perceptual Symbol Systems, *Behavioral and Brain Sciences, 22*, 577–660.

Barsalou, L. W. (2003a). Situated simulation in the human conceptual system. *Language and Cognitive Processes, 18*, 513–562.

Barsalou, L. W. (2003b). Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society of London: Biological Sciences, 358*, 1177–1187.

Barsalou L. W. (2005). Abstraction as dynamic interpretation in perceptual symbol systems. In *Building Object Categories*, Carnegie Sympos. Ser., ed. L Gershkoff-Stowe, D Rakison, pp. 389–431. Mahwah, NJ: Erlbaum

Barsalou L. W. (2007). Grounding symbolic operations in the brain's modal systems. In *Embodied Grounding: Social, Cognitive, Affective, and Neuroscientific approaches*, ed. GR Semin, ER Smith. New York: Cambridge Univ. Press.

Barsalou, L. W. (2008). Grounded Cognition, *Annu. Rev. Psychol*. 59:617–45

Barsalou L. W, Breazeal C., & Smith LB. (2007). Cognition as coordinated noncognition. *Cogn. Process*. 8:79–91

Barsalou, L. W. & Hale, C. R. (1993). Components of conceptual representation: From feature lists to recursive frames. In: *Categories and concepts: Theoretical views and inductive data analysis*, ed. I. Van Mechelen, J. Hampton, R. Michalski & P. Theuns. Academic Press.

Bellman, R.E. (1961). *Adaptive Control Processes.* Princeton University Press, Princeton, NJ.

Bonnisone, P.P., Henrion, M., Kanal, L.N., & Lemmer, J.F. (1991). *Uncertainty in Artificial Intelligence 6*. North Holland, Amsterdam, The Netherlands.

Boroditsky L, Ramscar M. (2002). The roles of body and mind in abstract thought. Psychol. Sci. 13:185–88.

Breazeal C. (2002). Designing Sociable Robots. Cambridge, MA: MIT Press

Brighton, H., Smith, K., & Kirby, S. (2005). Language as an evolutionary system. *Physics of Life Reviews*, *2(3),* 177-226.

Cacioppo, J. T. & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42,* 116-131.

Cangelosi, A,, Greco, A., & Harnad S. (2000). From robotic toil to symbolic theft: grounding transfer from entry-level to higher-level categories. *Connect. Sci.* 12:143–62.

Cangelosi, A. & Riga T. (2006). An embodied model for sensorimotor grounding and grounding transfer: experiments with epigenetic robots. *Cogn. Sci.* 30:673–89.

Carnap, R. (1959). *The Logical Syntax of Language.* Totowa, NJ: Littlefield, Adams & Co.

Chomsky, N. (1957). *Syntactic Structures.* Haag, Netherlands: Mouton.

Chomsky, N. (1972). *Language and Mind.* Harcourt Brace Javanovich, New York, NY.

Chomsky, N. (1981). Principles and Parameters in Syntactic Theory. In N.Hornstein and D. Lightfoot, eds., Explanation in Linguistics. The Logical Problem of Language Acquisition, Longman, London.

Clark, A. (1997). *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT Press.

Coventry, K.R. & Garrod, S.C. (2004). *Saying, seeing and acting: The Psychological Semantics of Spatial Prepositions.* Psychology Press: Hove.

Deacon, T. W. (1998). *The Symbolic Species: The Co-Evolution of Language and the Brain.* New York: Norton.

Edelman, S. & Newell, F. (1998). On the representation of object structure in human vision: evidence from differential priming of shape and location. Research paper University of Sussex.

Edelman, S. & Breen, E. M. (1999). On the virtues of going all the way. A commentary on Barsalou's BBS article, Perceptual Symbol Systems, *Behavioral and Brain Sciences, 22*, 614.

Edelman, S. & Intrator, N. (2001). Towards structural systematicity in distributed, statically bound visual representations. *Cognitive Science, 27,* 73-110.

Edelman, S. (2003). But will it scale up? Not without representation. *Adaptive Behavior, 11(4),* 273-275.

Fauconnier, G. (1985). *Mental spaces: Aspects of meaning construction in natural language.* Cambridge, MA: MIT Press.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Evanston, IL: Row, Peterson.

Fiske, J. (1987). Film, TV and the Popular, ed. Bell and Hanet, Continuum: The Australian Journal of Media & Culture, **12**, pp.4-5.

Fontanari, J.F. & Perlovsky, L.I. (2007). Evolving Compositionality in Evolutionary Language Games. *IEEE Transactions on Evolutionary Computations., 11(6*), 758-769; on-line doi:10.1109/TEVC.2007.892763

Fontanari, J.F. & Perlovsky, L.I. (2008a). How language can help discrimination in the Neural Modeling Fields framework. *Neural Networks, 21(2-3),* 250–256.

Fontanari, J.F. & Perlovsky, L.I. (2008b). A game theoretical approach to the evolution of structured communication codes, *Theory in Biosciences, 127*, 205-214.

Fontanari, J. F., Tikhanoff, V., Cangelosi, A., Ilin, R., & Perlovsky, L.I. (2009). Cross-situational learning of object–word mapping using Neural Modeling Fields. *Neural Networks, 22*, in press. http://dx.doi.org/10.1007/s12064-008-0024-1.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Glenberg, A. M. (1997). What memory is for. *Behavioral & Brain Sciences, 20*, 1-55.

Glenberg A. M. & Kaschak, M. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9(3)*, 558-565.

Gödel, Kurt, Collected Works, Vol. I, "Publications 1929-1936," edited by S. Feferman, J. W. Dawson, Jr., S. C. Kleene, Oxford Univ. Press, New York, NY. See also discussions in Penrose, R. (1994*). Shadows of the Mind.* Oxford University Press, Oxford, England.

Grossberg, S. (1982). Studies of Mind and Brain. D.Reidel Publishing Co., Dordrecht, Holland.

Grossberg, S. & Levine, D.S. (1987). Neural dynamics of attentionally modulated Pavlovian conditioning: blocking, inter-stimulus interval, and secondary reinforcement. *Psychobiology*, **15**(3), pp.195-240.

Harlow, Harry. 1953. "Mice, Monkeys, Men, and Motives." *Psychological Review* 60:23-32.

Hilbert, David. (1928/1967). The foundations of mathematics. In J. van Heijenoort, Ed., From Frege to Gödel. Cambridge, MA: Harvard University Press, p.475.

Ilin, R. & Perlovsky, L. I. (2009) Cognitively inspired neural network for recognition of situations. *International Journal of Natural Computing Research*, 1, in print.

Joyce D., Richards L., Cangelosi, A., & Coventry K.R. (2003). On the foundations of perceptual symbol systems: Specifying embodied representations via connectionism. In: F. Detje, D. Dörner, H. Schaub (Eds.), T*he Logic of Cognitive Systems. Proceedings of the Fifth International Conference on Cognitive Modeling* (pp. 147-152), Universitätsverlag Bamberg.

Jung, C.G. (1921). *Psychological Types.* In the Collected Works, v.6, Bollingen Series XX, 1971. Princeton, NJ: Princeton University Press.

Kecman, V. (2001). Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models (Complex Adaptive Systems). The MIT Press, Cambridge, MA.

Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.

Kosslyn, S. M. (1994). *Image and Brain*. Cambridge, MA: MIT Press.

Lakoff, G. 1987. Women, Fire, and Dangerous Things. What Categories Reveal about the Mind. Chicago: University of Chicago Press.

Lakoff, G. & Johnson, M. (1980). Metaphors We Live By. Chicago: University of Chicago Press.

Lakoff, G. & Johnson, M. (1999). Philosophy in the flesh: The embodied mind and its challenge to western thought. New York: HarperCollins Publishers.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar Vol. 1: Theoretical Prerequisites.* Stanford: Stanford University Press.

Langacker, R. W. (1991). *Language, Image and Symbol.* The Hague: Mouton de Gruyter.

Lieberman, P. (2000). Human Language and our Reptilian Brain: The Subcortical Bases of Speech, Syntax, and Thought. Harvard University Press, Cambridge, MA.

Linnehan, R., Mutz, Perlovsky, L.I., C., Weijers, B., Schindler, J., Brockett, R. (2003). *Detection of Patterns Below Clutter in Images*. Int. Conf. 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), pp.1-38.

Minsky, M.L. (1975). A Framework for Representing Knowlege. In The Psychology of Computer Vision, ed. P. H. Whinston, McGraw-Hill Book, New York.

Nevatia R. & Binford, T.O. (1977). Description and recognition of curved objects. *Artificial Intelligence*, **8**(1), pp.77-98.

Pecher, D. & Zwaan, R.A., (Eds.) (2005). Grounding cognition: T*he role of perception and action in memory, language, and thinking.* Cambridge, UK: Cambridge University Press.

Peirce, C.S. (1897, 1903). Logic as Semiotic: The Theory of Signs. In Buchler ed., The Philosophical Writing of Peirce (1955), Dover, NY; Collected Papers V. II, Elements of Logic, Hartshorne and Weiss, eds. Belknap, Cambridge, MA.

Penrose, R. (1994). Shadows of the Mind. Oxford University Press, Oxford, England.

Perlovsky, L.I. (1987). Multiple Sensor Fusion and Neural Networks. DARPA Neural Network Study, MIT/Lincoln Laboratory, Lexington, MA.

Perlovsky, L.I. (1988). *Neural Networks for Sensor Fusion and Adaptive Classification.* Proc. 1st Annual Int. Neural Network Society Meeting, Boston, MA, p.42.

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

Perlovsky, L.I. (1994). Computational Concepts in Classification: Neural Networks, Statistical Pattern Recognition, and Model Based Vision. *Journal of Mathematical Imaging and Vision, 4 (1)*, 81-110.

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

Perlovsky, L.I. (1996b). *Mathematical Concepts of Intellect.* Proc. World Congress on Neural Networks, San Diego, CA; Lawrence Erlbaum Associates, NJ, pp.1013-16.

Perlovsky, L.I. (1997). *Physical Concepts of Intellect.* Proc. Russian Academy of Sciences, **354**(3), pp. 320-323.

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

Perlovsky, L.I. (2000). *Neural Networks and Intellect.* Oxford University Press, New York, NY.

Perlovsky, L.I. (2002). Computational Linguistics, Semiotics and the Mind. Computational Linguistics Colloquium, University of Trier, 27 June 2002, Trier, Germany.

Perlovsky, L.I. (2004). Integrating Language and Cognition. *IEEE Connections*, 2(2), 8-12.

Perlovsky, L.I. (2006a). Modeling field theory for higher cognitive functions. In *Artificial Cognition Systems*, Eds. A. Loula, R. Gudwin, J. Queiroz. Idea Group, Hershey, PA, pp.64-105.

Perlovsky, L. I. (2006b). Toward physics of the mind: concepts, emotions, consciousness, and symbols. *Physics of Life Reviews 3*, 23-55.

Perlovsky, L.I. (2006c). Fuzzy Dynamic Logic. New Math. and Natural Computation, 2(1), pp.43-55.

Perlovsky, L.I. (2006d). Symbols: Integrated cognition and language. In R. Gudwin, J. Queiroz (Eds.). S*emiotics and intelligent systems development*. Hershey, PA: Idea Group , pp.121-151.

Perlovsky, L.I. (2007a). The Mind vs. Logic: Aristotle and Zadeh. C*ritical Review*, 1(1), 30-33.

Perlovsky, L.I. (2007b). Neural Networks, Fuzzy Models and Dynamic Logic. *In Aspects of Automatic Text Analysis (Festschrift in Honor of Burghard Rieger)*, Eds. R. Köhler and A. Mehler, Springer, Germany, pp.363-386.

Perlovsky, L.I. (2007c). Neural Dynamic Logic of Consciousness: the Knowledge Instinct. In *Neurodynamics of Higher-Level Cognition and Consciousness*, Eds. Perlovsky, L.I., Kozma, R. ISBN 978-3-540-73266-2, Heidelberg, Germany: Springer Verlag,.

Perlovsky, L. I. (2009a). Language and Cognition. *Neural Network 22(3),* 247-257. doi:10.1016/j.neunet.2009.03.007

Perlovsky, L. I. (2009b). Language and Emotions: Emotional Sapir-Whorf Hypothesis. *Neural Networks*; doi:10.1016/j.neunet.2009.06.034.

Perlovsky, L.I. & Fontanari, J.F. (2006). How Language Can Guide Intelligence. *EVOLANG'06,* Rome Italy.

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

Russell, B. (1919). *Introduction to Mathematical Philosophy.* London: George Allen and Unwin, p.175.

Saussure, F. De. (1916/1965). *Course In General Linguistics*. McGraw-Hill, New York, NY, p.98.

Schacter, D. L. & Addis, D. R. (2007). The ghosts of past and future. *Nature*, *445*, 27.

Simmons, W. K. & Barsalou, L. W. (2003). The similarity-in-topography principle: reconciling theories of conceptual deficits. *Cogn. Neuropsychol*. 20:451–86.

Singer, R.A., Sea, R.G. and 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.

Talmy, L. (1983). *How language structures space*. New York: Plenum Press

Talmy, L. (1988). The relation of grammar to cognition. In Rudzka-Ostyn ed., *Topics in Cognitive Linguistics.* Amsterdam: Benjamins, pp.165-205

Tomasello, M. (2003). *Constructing a Language, A Usage-Based Theory of Language Acquisition.* Harvard University Press, Cambridge, MA.

Winston, P. H. (1984). Artificial Intelligence. *Addison-Wesley.* Reading, MA.

Wu L. & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation, *Acta Psychologica*, in print.

Yeh, W. & Barsalou, L. W. (2006). The situated nature of concepts. *Am. J. Psychol.* 119, 349–84.

Zadeh, L.A. (1965). Fuzzy Sets. *Information and Control, 8*, 338-352.

Appendix 1.

Bottom-up signals {$\mathbf{X}$(n)} in this simplified discussion, is a field of neuronal synapse activations in visual cortex. Here and below curve brackets {…} denote multiple signals, a field. Index n = 1,… N, enumerates neurons and $\mathbf{X}$(n) are the activation levels. Sources of top-down signals are representations or concept-models {$\mathbf{M}_m$(n)} indexed by m = 1,… M. Each model *m* $\mathbf{M}_m$(n) projects a set of priming, top-down signals, representing the bottom-up signals $\mathbf{X}$(n) expected from a particular object, *m*. Models depend on parameters {$\mathbf{S}_m$}, $\mathbf{M}_m(\mathbf{S}_m,n)$. Parameters characterize object position, angles, lightings, etc. (In case of learning *situations* considered in section 3, parameters characterize objects and relations making up a situation.) To summarize this highly simplified description of a visual system, *n* enumerates the visual cortex neurons, $\mathbf{X}$(n) are the "bottom-up" activation levels of these neurons coming from the retina, and $\mathbf{M}_m$(n) are the "top-down" activation levels (priming) of the visual cortex neurons. The learning-perception process "matches" these top-down and bottom-up activations by selecting "best" models and their parameters and the corresponding sets of bottom-up signals.

Let us concentrate on defining a mathematical measure of the "best" fit between bottom-up and top-down signals. It is constructed in such a way that any of object-models can be recognized. Correspondingly, a similarity measure is designed so that it treats each object-model as a potential alternative for each subset of signals (Perlovsky 2000, 2006),

$$L(\{\mathbf{X}\},\{\mathbf{M}\}) = \prod_{n\in N} \sum_{m\in M} r(m)\, l(\mathbf{X}(n) \mid \mathbf{M}_m(n)); \tag{A1}$$

Here, $l(\mathbf{X}(n)|\mathbf{M}_m(n))$ (or simply $l(n|m)$) is called a conditional similarity between one signal $\mathbf{X}(n)$ and one model $\mathbf{M}_m(n)$. Parameters $r(m)$ are proportional to the number of objects described by the model m. Expression (1) accounts for *all combinations* of signals and models in the following way. Sum $\sum$ ensures that any of the object-models can be considered (by the mind) as a source of signal $\mathbf{X}(n)$. Product $\prod$ ensures that all signals have a chance to be considered (even if one signal is not considered, the entire product is zero, and similarity L is 0; so for good similarity all signals have to be accounted for. This does not assume exorbitant amount of attention to each minute detail: among models there is a vague simple model for "everything else"). In a simple case, when all objects are perfectly recognized and separated from each other, there is just one object-model corresponding to each signal (other $l(n|m) = 0$). In this simple case expression (1) contains just 1 item, a product of all non-zero $l(n|m)$. In the general case, before objects are recognized, L contains a large number of *combinations* of models and signals; a product over N signals is taken of the sums over M models; this results in a total of M^N items; this huge number is the cause for the combinatorial complexity discussed previously.

The DL learning process consists in estimating model parameters $\mathbf{S}_m$ and associating subsets of signals with concepts by maximizing the similarity (1). Although (1) contains combinatorially many items, DL maximizes it without combinatorial complexity (Perlovsky, 1996b; 1997; 2000). First, fuzzy association variables f(m|n) are defined,

$$f(m|n) = r(m)\, l(n|m) \,/ \sum_{m' \in M} r(m')\, l(n|m'). \qquad \text{(A2)}$$

These variables give a measure of correspondence between signal $\mathbf{X}(n)$ and model $\mathbf{M}_m$ relative to all other models, m'. They are defined similarly to the a posteriori Bayes probabilities, they range between 0 and 1, and as a result of learning they converge to the probabilities under certain conditions.

DL process is defined by the following set of differential equations,

$$df(m|n)/dt = f(m|n) \sum_{m' \in M} \{[\delta_{mm'} - f(m'|n)]\, [\partial \ln l(n|m')/\partial \mathbf{M}_{m'}]\, (\partial \mathbf{M}_{m'}/\partial \mathbf{S}_{m'})\, d\mathbf{S}_{m'}/dt,$$

$$d\mathbf{S}_m/dt = \sum_{n \in N} f(m|n)[\partial \ln l(n|m)/\partial \mathbf{M}_m]\partial \mathbf{M}_m/\partial \mathbf{S}_m, \;\; \delta_{mm'} = 1 \text{ if } m=m', 0 \text{ otherwise.} \qquad \text{(A3)}$$