



2009 Special Issue

Language and emotions: Emotional Sapir–Whorf hypothesis

Leonid Perlovsky*

Harvard University, SEAS, Cambridge, USA

AF Research Laboratory, Sensors Directorate, Hanscom AFB, USA

ARTICLE INFO

Article history:

Received 12 May 2009

Received in revised form 10 June 2009

Accepted 25 June 2009

Keywords:

Language

Cognition

Emotions

Knowledge instinct

Dynamic logic

Mind

Heterarchy

Hierarchy

ABSTRACT

An emotional version of Sapir–Whorf hypothesis suggests that differences in language emotionalities influence differences among cultures no less than conceptual differences. Conceptual contents of languages and cultures to significant extent are determined by words and their semantic differences; these could be borrowed among languages and exchanged among cultures. Emotional differences, as suggested in the paper, are related to grammar and mostly cannot be borrowed. The paper considers conceptual and emotional mechanisms of language along with their role in the mind and cultural evolution. Language evolution from primordial undifferentiated animal cries is discussed: while conceptual contents increase, emotional reduced. Neural mechanisms of these processes are suggested as well as their mathematical models: the knowledge instinct, the dual model connecting language and cognition, neural modeling fields. Mathematical results are related to cognitive science, linguistics, and psychology. Experimental evidence and theoretical arguments are discussed. Dynamics of the hierarchy–heterarchy of human minds and cultures is formulated using mean-field approach and approximate equations are obtained. The knowledge instinct operating in the mind heterarchy leads to mechanisms of differentiation and synthesis determining ontological development and cultural evolution. These mathematical models identify three types of cultures: “conceptual” pragmatic cultures in which emotionality of language is reduced and differentiation overtakes synthesis resulting in fast evolution at the price of uncertainty of values, self doubts, and internal crises; “traditional–emotional” cultures where differentiation lags behind synthesis, resulting in cultural stability at the price of stagnation; and “multi-cultural” societies combining fast cultural evolution and stability. Unsolved problems and future theoretical and experimental directions are discussed.

Published by Elsevier Ltd

1. Emotional Sapir–Whorf hypothesis

Benjamin Whorf (Whorf, 1956) and Edward Sapir (Sapir, 1985) in a series of publications in the 1930s researched an idea that the way people think is influenced by the language they speak. Although there was a long predating linguistic and philosophical tradition, which emphasized the influence of language on cognition (Bhartrihari, 1971; Humboldt, 1836/1967; Nietzsche, 1876/1983), this is often referenced as Sapir–Whorf hypothesis (SWH). Linguistic evidence in support of this hypothesis concentrated on conceptual contents of languages. For example, words for colors influence color perception (Roberson, Davidoff, & Braisby, 1999; Winawer et al., 2007). The idea of language influencing cognition and culture has been criticized and “fell out of favor” in the 1960s (Wikipedia, 2009a) due to a prevalent influence of Chomsky’s ideas emphasizing language and cognition to be

separate abilities of the mind (Chomsky, 1965). Recently SWH again attracted much academic attention, including experimental confirmations (see the previous references) and theoretical skepticism (Pinker, 2007). Interactions between language and cognition have been confirmed in fMRI experiments (Simmons, Stephan, Carla, Xiaoping, & Barsalou, 2008). Brain imaging experiments by Franklin et al. (2008) demonstrated that learning a word “rewires” cognitive circuits in the brain, learning a color name moves perception from right to left hemisphere. These recent data address, in particular, an old line of critique of SWH: whether the relationships between cultures and languages are causal or correlational and if causal, what is the cause and what is the effect. Franklin et al. (2008) experiments have demonstrated that language affects thinking. This discussion will be continued later but first I would like to emphasize that all arguments and experiments referenced above concentrate on conceptual effects of language.

Emotional effects might be no less important (Guttfreund, 1990; Harris, Ayçiçeği, & Gleason, 2003). In particular indicative are results of Guttfreund (1990): Spanish–English bilinguals manifested more intense emotions in psychological interviews conducted in Spanish than in English, irrespective of whether

* Corresponding address: Harvard University, SEAS, Cambridge, USA. Tel.: +1 781 377 1728.

E-mail address: leonid@seas.harvard.edu.

their first language was English or Spanish. Still, experimental evidence suggesting interaction between the emotional contents of languages and cognition is limited, the neural mechanisms of these interactions are not known, and no computational models have existed (Perlovsky, 2006a, 2006b, 2006c, 2009).

This paper derives neurally motivated computational models of how conceptual and emotional contents of language affect cognition. This derivation is motivated by the knowledge about brain modules, rather than individual neurons. The next section reviews conceptual and emotional mechanisms of language and their interaction with cognition. Whereas direct experimental data are inadequate, I briefly review existing theoretical ideas and experimental evidence on language evolution, conceptualizing possible mechanisms, and emphasizing directions for future research. Section 3 summarizes previously developed neuro-mathematical theories of interaction between language and cognition (Perlovsky, 2004, 2006a, 2006b, 2006c, 2009), which correspond to recent experimental data; these models are extended toward heterarchy of the mind. Section 4 derives neurally motivated cultural evolutionary models and demonstrates that different cultural evolutionary paths are favored by differences in interaction between cognition and language. In conclusion I discuss future theoretical and experimental directions.

2. Language and cognition

Language is widely considered as a mechanism for communicating conceptual information. Emotional contents of language are less appreciated and their role in the mind and evolutionary significance are less known. Still their roles in ontology, evolution, and cultural differences are significant. Whereas much research concentrates on language-computation, sensory-motor, and concept-intention interfaces (Hauser, Chomsky, & Fitch, 2002), the current paper emphasizes that the primordial origin of language was a unified neural mechanism of fused voicing-behavior, emotion-motivation, and concept-understanding (Deacon, 1989; Lieberman, 2000; Mithen, 2007). It is likely that differentiation of mechanisms involved in language, voicing, cognition, motivation, and behavior occurred at different prehistoric times, in different lineages of our ancestors. This may be relevant to discussions of evolution of language and cognition (Botha, 2003; Botha & Knight, 2009).

I address the current differentiated state of these abilities in the human mind, as well as unifying mechanisms of interfaces-links, which make possible integrated human functioning. The paper concentrates on mechanisms of existing interfaces and their cultural evolution. Before describing in the next section mechanisms of language, concepts, and emotions mathematically I will summarize these mechanisms conceptually in correspondence with general knowledge documented in a large number of publications emphasizing certain aspects that have escaped close scientific attention in the previous research.

2.1. Primordial undifferentiated synthesis

Animals' vocal tract muscles are controlled mostly from the ancient emotional center (Lieberman, 2000). Vocalizations are more affective than conceptual. Mithen (Mithen, 2007) summarized the state of knowledge about vocalization by apes and monkeys. Calls could be deliberate, however their emotional-behavioral meanings are probably not differentiated; primates cannot use vocalization separately from emotional-behavioral situations; this is one reason they cannot have language.

Emotionality of voice in primates and other animals is governed from a single ancient emotional center in the limbic system (Deacon, 1989; Lieberman, 2000; Mithen, 2007). Cognition is less

differentiated than in humans. Sounds of animal cries engage the entire psyche, rather than concepts and emotions separately. An ape or bird seeing danger does not think about what to say to its fellows. A cry of danger is *inseparably* fused with recognition of a dangerous situation, and with a command to oneself and to the entire flock: "fly!". An evaluation (emotion of fear), understanding (concept of danger), and behavior (cry and wing sweep)—are not differentiated. Conscious and unconscious are not separated. Recognizing danger, crying, and flying away is a fused concept-emotion-behavioral *synthetic* form of cognition-action. Birds and apes can not control their larynx muscles *voluntarily*.

2.2. Language and differentiation of emotion, voicing, cognition, and behavior

Origin of language required freeing vocalization from uncontrolled emotional influences. Initial undifferentiated unity of emotional, conceptual, and behavioral (including voicing) mechanisms had to separate-differentiate into partially independent systems. Voicing separated from emotional control due to a separate emotional center in cortex which controls larynx muscles, and which is partially under volitional control (Deacon, 1989; Mithen, 2007). Evolution of this volitional emotional mechanism possibly paralleled evolution of language computational mechanisms. In contemporary languages the conceptual and emotional mechanisms are significantly differentiated, compared to animal vocalizations. The languages evolved toward conceptual contents, while their emotional contents were reduced. Cognition, or understanding of the world, is due to mechanisms of concepts, also referred to as internal representations or models. Barsalou calls this mechanism situated simulation (Barsalou, 2009). Perception or cognition consists of matching internal concept-models (simulations) with patterns in sensor data. Concept-models generate top-down neural signals that are matched to bottom-up signals coming from lower levels (Grossberg, 1988; Perlovsky, 2000). In this simulation process the vague internal models are modified to match concrete objects or situations (Bar et al., 2006; Perlovsky, 2006a).

How these cognitive processes are determined and affected by language? Primates' cognitive abilities are independent from language. Language is fundamental to human cognitive abilities (Perlovsky, 2006a). A possible mathematical mechanism of language guiding and enhancing cognition have been discussed in Perlovsky (2004, 2006a, 2006c, 2007a, 2007b, 2009), Fontanari and Perlovsky (2005, 2007, 2008a, 2008b) and Fontanari, Tikhonoff, Cangelosi, Perlovsky, and Ilin (2009). This is a mechanism of the dual model whereby every concept-model has two parts: cognitive and language. The language models (words, phrases) are acquired from surrounding language by age of five or seven. They contain cultural wisdom accumulated through millennia. During the rest of life the language models guide the acquisition of cognitive models.

2.3. Emotions, instincts, and the knowledge instinct

The word emotion refers to several neural mechanisms in the brain (Juslin & Västfjäll, 2008); in this paper I always refer to instinctual-emotional mechanism described in Grossberg and Levine (1987), which is consistent with Cabanac (2002). The word instinct in this paper is used in correspondence with this reference to denote a simple inborn, non-adaptive mechanism of internal "sensor", which measures vital body parameters, such as blood pressure, and indicates 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 Grossberg and Seidman (2006) and Gnadt and Grossberg (2008) and references therein. We

have dozens of such sensors, measuring sugar level in blood, body temperature, pressure at various parts, etc.

Mechanisms of concepts evolved for instinct satisfaction. According to instinctual–emotional theory (Grossberg & Levine, 1987), communicating satisfaction or dissatisfaction of instinctual needs from instinctual “sensors” to decision making parts of the brain is performed by emotional neural signals. Perception and understanding of concept-models corresponding to objects or situations that potentially can satisfy an instinctual need receive preferential attention and processing resources in the mind. In this paper emotions refer to neural signals connecting conceptual and instinctual brain regions.

Perception and cognition requires matching top-down signals from concept-models to bottom-up signals coming from sensory organs. Perception is required for survival. Therefore humans and high animals have an inborn drive to fit top-down and bottom-up signals, the knowledge instinct (Perlovsky, 1997, 2000, 2006a, 2007a; Perlovsky & McManus, 1991). These references discuss specific emotions related to satisfaction or dissatisfaction of the knowledge instinct. These emotions are related purely to knowledge, not to bodily needs; since Kant (1790/1914) this type of emotions are called aesthetic emotions. According to our theory of the knowledge instinct they are inseparable from every act of perception and cognition.

Biologists and psychologists have discussed various aspects of this mechanism: a need for positive stimulations, curiosity, a motive to reduce cognitive dissonance, a need for cognition (Berlyne, 1960; Cacioppo, Petty, Feinstein, & Jarvis, 1996; Festinger, 1957; Harlow & Mears, 1979; Levine & Perlovsky, 2008). Until recently, however, this drive was not mentioned among ‘basic instincts’ on a par with instincts for food and procreation. The fundamental nature of this mechanism became clear during mathematical modeling of workings of the mind. Our knowledge always has to be modified to fit the current situations. We don’t usually see exactly the same objects as in the past: angles, illumination, and surrounding contexts are different. Therefore, our internal representations have to be modified; adaptation-learning is required (Grossberg, 1988; Kosslyn, Ganis, & Thompson, 2001).

All learning and adaptive algorithms maximize correspondence between the algorithm internal structure (knowledge in a wide sense) and objects of recognition; the psychological interpretation of this mechanism is the knowledge instinct. The mind–brain mechanisms of the knowledge instinct are discussed in Levine and Perlovsky (2008). As we discuss below, it is a foundation of our higher cognitive abilities, and it defines the evolution of consciousness and cultures (Perlovsky, 2006a, 2006b, 2006c).

2.4. Grammar, language emotionality, and meanings

Language and voice started separating from ancient emotional centers possibly millions of years ago. Nevertheless, emotions are present in language. Most of these emotions originate in cortex and are controllable aesthetic emotions. Their role in satisfying the knowledge instinct is considered in the next section. Emotional centers in cortex are neurally connected to old emotional limbic centers, so both influences are present. Emotionality of languages is carried in language sounds, what linguists call prosody or melody of speech. This ability of human voice to affect us emotionally is most pronounced in songs. Songs and music, however, is a separate topic (Perlovsky, 2006d, 2008) not addressed in this paper.

Emotionality of everyday speech is low, unless affectivity is specifically intended. We may not notice emotionality of everyday “non-affective” speech. Nevertheless, “the right level” of emotionality is crucial for developing cognitive parts of models. If language parts of models were highly emotional, any discourse

would immediately resort to blows and there would be no room for language development (as among primates). If language parts of models were non-emotional at all, there would be no motivational force to engage into conversations, to develop language models. The motivation for developing higher cognitive models would possibly be reduced. Lower cognitive models, say for object perception, would be developed because they are imperative for survival and because they can be developed independently from the language, based on direct sensory perceptions, like in animals. But models of situations and higher cognition are developed based on language models (Perlovsky, 2004, 2006a, 2006c, 2007b, 2009). As discussed later, this requires emotional connections between cognitive and language models.

Primordial fused language–cognition–emotional models, as discussed, have differentiated long ago. The involuntary connections between voice–emotion–cognition have dissolved with emergence of language. They have been replaced with habitual connections. Sounds of all languages have changed and, it seems, sound–emotion–meaning connections in languages should have severed. Nevertheless, if the sounds of a language change slowly, connections between sounds and meanings persist and consequently the emotion–meaning connections persist. This persistence is a foundation of meanings because meanings imply motivations. If the sounds of a language change too fast, the cognitive models are severed from motivations, and meanings disappear. If the sounds change too slowly the meanings are nailed emotionally to the old ways, and culture stagnates.

This statement is controversial, and indeed, it may sound puzzling. Doesn’t culture direct language changes or is the language the driving force of cultural evolution? Direct experimental evidence is limited; it will have to be addressed by future research. Theoretical considerations suggest no neural or mathematical mechanism for culture directing evolution of language through generations; just the opposite, most of cultural contents are transmitted through language. Cognitive models contain cultural meanings separate from language (Perlovsky, 2009), but transmission of cognitive models from generation to generation is mostly facilitated by language. Cultural habits and visual arts can preserve and transfer meanings, but they contain a minor part of cultural wisdom and meanings comparative to those transmitted through the language. Language models are major containers of cultural knowledge shared among individual minds and collective culture.

The arguments in the previous two paragraphs suggest that an important step toward understanding cultural evolution is to identify mechanisms determining changes of the language sounds. As discussed below, changes in the language sounds are controlled by grammar. In inflectional languages, affixes, endings, and other inflectional devices are fused with sounds of word roots. Pronunciation-sounds of affixes are controlled by few rules, which persist over thousands of words. These few rules are manifest in every phrase. Therefore every child learns to pronounce them correctly. Positions of vocal tract and mouth muscles for pronunciation of affixes (etc.) are fixed throughout population and are conserved throughout generations. Correspondingly, pronunciation of whole words cannot vary too much, and language sound changes slowly. Inflections therefore play a role of “tail that wags the dog” as they anchor language sounds and preserve meanings. This, I think is what Humboldt (1836/1967) meant by “firmness” of inflectional languages. When inflections disappear, this anchor is no more and nothing prevents the sounds of language to become fluid and change with every generation.

This has happened with English language after transition from Middle English to Modern English (Lerer, 2007), most of inflections have disappeared and sound of the language started changing within each generation with this process continuing today. English evolved into a powerful tool of cognition unencumbered

by excessive emotionality. English language spreads democracy, science, and technology around the world. This has been made possible by conceptual differentiation empowered by language, which overtook emotional synthesis. But the loss of synthesis has also led to ambiguity of meanings and values. Current English language cultures face internal crises, uncertainty about meanings and purposes. Many people cannot cope with diversity of life. Future research in psycholinguistics, anthropology, history, historical and comparative linguistics, and cultural studies will examine interactions between languages and cultures. Initial experimental evidence suggests emotional differences among languages consistent with our hypothesis (Gutfreund, 1990; Harris et al., 2003).

Neural mechanisms of grammar, language sound, related emotions–motivations, and meanings hold a key to connecting neural mechanisms in the individual brains to evolution of cultures. Studying them experimentally is a challenge for future research. It is not even so much a challenge, because experimental methodologies are at hand; they just should be applied to these issues. The following sections develop mathematical models based on existing evidence that can guide this future research.

3. Heterarchy of the mind and cultural dynamics

This section summarizes mathematical models of the mind mechanisms corresponding to the discussion in the previous section. These models are based on the available experimental evidence and theoretical development by many authors summarized in Perlovsky (1987, 1994a, 1997, 1998, 2000, 2006a, 2006b, 2006c, 2007b, 2009) and Perlovsky et al. (1997) and it corresponds to recent neuro-imaging data (Bar et al., 2006; Franklin et al., 2008).

3.1. Mathematical model of cognition

Mechanisms of concepts, instincts, and emotions were described in above references and summarized in Section 2.3. To briefly summarize, concepts operate like internal models of objects and situations; 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 retina (bottom-up signal). Perception occurs when top-down and bottom-up signals match. Concepts evolved for instinct satisfaction. The word instinct denotes here a simple inborn, non-adaptive mechanism of internal “sensor”, which measures vital body parameters, such as blood pressure. Satisfaction or dissatisfaction of instinctual needs is communicated from instinctual parts of the brain to decision making parts of the brain by emotional neural signals. Perception and understanding of concept-models corresponding to objects or situations that potentially can satisfy an instinctual need receive preferential attention and processing resources. Here I summarize a mathematical description of these mechanisms according to the cited references and Perlovsky (2006a, 2007a, 2009).

Matching top-down and bottom-up signals is essential for perception. Therefore humans and high animals have an inborn drive to fit top-down and bottom-up signals. This is a mechanism of the instinct for knowledge (Perlovsky, 1994b, 2000, 2006a; Perlovsky & McManus, 1991). Brain areas participating in the knowledge instinct were discussed in Levine and Perlovsky (2008). The mathematical description of the knowledge instinct maximizes similarity between top-down and bottom-up signals, L :

$$L = \prod_{n \in N} \sum_{m \in M} r(m) l(n|m) p_e(N, M) o(N, M) v. \quad (1)$$

Here $l(n|m)$ is a partial similarity of a bottom-up signal in pixel n given that it originated from the object or concept described

by the top-down concept-model m (we refer to this below as object m). This partial similarity is normalized on object-concept m being definitely present, which is not necessarily true; therefore coefficient $r(m)$ models a probability of object-concept m actually be present; they are called rates. Function $p_e(N, M)$, penalizes for the number of parameters in models, $o(N, M)$ penalizes for the number of computations, and v is Vapnik’s penalty function (Vapnik, 1998) discussed below.

Similarity (1) mathematically models the mind’s “sensor” mechanism, which measures knowledge as a correspondence between the minds concepts–representations and perceptual signals (top-down and bottom-up signals). According to instinctual–emotional theory of Grossberg and Levine (1987), this value is an emotional measure of satisfaction or dissatisfaction of the knowledge instinct.

Modeling perception and cognition by maximizing this expression was described in Perlovsky (2006a); Perlovsky, Webb, Bradley, and Hansen (1998). This maximization procedure underlies the following development; its principal aspect, called dynamic logic, proceeds from unconscious and vague concept-models to conscious and concrete. As mentioned, this process was confirmed in brain imaging experiments (Bar et al., 2006). Dynamic logic is a general mechanism of the brain–mind; it corresponds to phase transitions from highly chaotic neurodynamic states to lower chaotic states (Freeman, 2007; Perlovsky & Kozma, 2007); and it eliminates reductionism of consciousness (Freeman, 2001; Levine & Perlovsky, 2008), which many authors take as an unavoidable consequence of the scientific analysis of the mind (Chalmers, 1997).

Similarity (1) describes the knowledge instinct operating at a single level of the mind heterarchy. The word heterarchy (Grossberg, 1988) refers to the fact that the mind is not a strict hierarchy; it involves cross-interaction among multiple layers. When concentrating on higher and lower level structure of the mind–brain, for simplicity I will use the word hierarchy. To describe the hierarchy, I denote a single-layer similarity (1) and all characteristics of this layer by index $h = 1, \dots, H$. The total similarity, specifying the instinct for knowledge for the entire hierarchy,

$$L = \prod_h L_h. \quad (2)$$

Expressions (1), (2), as discussed in Perlovsky (2000, 2006a), model neural functioning of the mind in correspondence with a large number of publications and neuro-imaging data discussed in these references and in Section 2. It leads to Zipf’s law of word rank distribution (Fontanari & Perlovsky, 2004). Mathematical models connecting this neural brain modeling to cultural evolution can proceed by simulating societies of interacting agents, each one satisfying its instinct for knowledge Eq. (2), similar to Fontanari and Perlovsky (2005, 2007, 2008a, 2008b) and Fontanari et al. (2009) and communicating through language (Perlovsky, 2009). In this paper a different approach is taken, deriving simplified expressions for similarity averaged over a population, so that maximization of similarity (2) could be studied analytically, including processes of cultural evolution. Similarity (2) determines the dynamics of multi-agent societies not unlike Lagrangian in physics determines the behavior of complex systems. Correspondingly, I use a technique inspired by mean field theories in physics, which has been developed for studying complex systems by substituting certain parameters in Lagrangian by their average values.

3.2. The mean field hierarchical dynamics

Considering (1) as a layer in (2), bottom-up signals are substituted by activated models at a lower layer, $N_h = M_{h-1}$. Parameter

penalty function is taken according to Akaike (1974),

$$pe(h) = \exp\{-p * M_h/2\}. \quad (3)$$

Here p is an average number of parameters per model (the layer index h is sometimes omitted for shortness). A penalty for the number of computations, $o(h) = 1/(\text{number of operations})$; the number of operations is proportional to the product of bottom-up and top-down signals,

$$o(h) = c2(h)/(M_{h-1} * M_h * p), \quad (4)$$

where $c2(h)$ is a constant. At every layer h , only a tiny part of all possible combinations of bottom-up signals, M_{h-1} , are organized into meaningful concepts M_h ; a majority of combinations do not have any meaning; they are assigned to a “clutter” model. The clutter model is homogeneous (does not depend on input data, and is only characterized by its proportion of signals, or rate, r_c . Concept-model rates at layer h , $r(m, h)$, are proportions of M_{h-1} signals associated with model $m(h)$; they are replaced by their average values, r_h . According to the rate normalization (Perlovsky, 2006a),

$$\sum_{m \in M(h)} r(m, h) + r_c = 1, \quad \text{or} \quad M_h * r_h + r_c = 1. \quad (5)$$

Psychologically, at level h , $M_h r_h$ is proportional to the total amount of knowledge, therefore we introduce a notation, $K_h = M_h r_h$; correspondingly, clutter is proportional to the “unknown”. Eq. (5) is equivalent to

$$r_c = 1 - K_h; \quad K_h = M_h r_h. \quad (6)$$

Vapnik’s penalty penalizes “too flexible” models, which can explain everything. In a simplified way, it penalizes for $K_h \rightarrow 1$. Accordingly, as an approximation, we define it as

$$v(h) = \exp\{-v/(1 - K_h)\}. \quad (7)$$

The average value of $l(m|n)$ can be computed as follows. Following Perlovsky (2006a), $l(m|n)$ can be modeled by a Gaussian function of $\Delta\mathbf{X}$, deviations of data, $\mathbf{X}(n)$, from the model m , \mathbf{M}_m , with covariance matrix \mathbf{C} ,

$$l(m|n) = (1/2\pi)^{p/2} \det(\mathbf{C})^{-p/2} \exp\{-\Delta\mathbf{X}\mathbf{C}^{-1}\Delta\mathbf{X}/2\}. \quad (8)$$

Here dimensionality is taken equal to the number of model parameters, p . For evaluating of an average value of $l(m|n)$ we assume that concept recognition is nearly perfect, so $l(m|n) \sim \delta_{mn}$. The average value of $\det(\mathbf{C})$ is substituted with σ^{2p} , σ being an average standard deviation. In the exponent, $\langle \Delta\mathbf{X}\Delta\mathbf{X} \rangle = \mathbf{C}$, and

$$\langle -\Delta\mathbf{X}\mathbf{C}^{-1}\Delta\mathbf{X}/2 \rangle = -1/2\text{Tr}(\mathbf{1}) = -p/2. \quad (9)$$

So the average value of partial similarity,

$$\langle l(m|n) \rangle = (1/2\pi)^{p/2} (1/\sigma) \exp\{-p/2\} \delta_{mn}. \quad (10)$$

Psychologically, this partial similarity is an emotional certainty that data n originates from object m . We denote it

$$E = \langle l(m|n) \rangle. \quad (11)$$

Emotionality of knowledge, as discussed depends on emotionality of language: language drives details vs. generality of cognitive models and determines ranges of σ and E . Detailed mathematical models of this interaction suitable for modeling of the hierarchical dynamics is beyond the limits of this paper.

Combining the above, a mean value of a layer h similarity,

$$L_h = [1 - K_h + K_h E_h]^{M(h-1)} \times \exp[-p_h M_h/2 - v/(1 - K_h)] o(h). \quad (12)$$

Here K and M characterize the breadth and differentiation of knowledge, whereas E characterizes emotional certainty about

validity of knowledge. This mean-field expression for similarity, together with Eq. (2) can be used now to derive hierarchical dynamics of the knowledge instinct, which defines emotional and knowledge-oriented “spiritual” individual ontological development (on average) as well as social dynamics and cultural evolution. This dynamics according to the knowledge instinct is given by the standard procedure of defining temporal derivatives along the gradient of similarity. This dynamics leads to evolution that satisfies the knowledge instinct,

$$\begin{aligned} dE_h/dt &= \delta dL/dE_h = \delta L * d(\ln L_h)/dE_h \\ &= \delta L * M_{h-1} * K_h/[1 - K_h + K_h E_h]. \end{aligned} \quad (13)$$

$$\begin{aligned} dK_h/dt &= \delta dL/dK_h \\ &= \delta L * \{M_{h-1} * (E_h - 1)/[1 - K_h + K_h E_h] - v/(1 - K_h)^2\} \end{aligned} \quad (14)$$

$$\begin{aligned} dM_h/dt &= \delta dL/dM_h \\ &= \delta L * \{\ln[1 - K_{h+1} + K_{h+1} E_{h+1}] - p_h/2 - 1/M_h\} \end{aligned} \quad (15)$$

where δ is a coefficient that would have to be determined empirically.

In addition to this knowledge–instinct driven dynamics, the hierarchy grows or shrinks depending on expansion or contraction of the number of general concepts at each layer. More general concepts move to higher levels of the hierarchy, and vice versa. The generality of a concept is determined by its standard deviation, related to emotionality, Eqs. (11) and (12). Detailed description of this part of hierarchical dynamics would require accounting for standard deviations varying from a typical value for each layer. Modeling this process in the future will account for interaction between language and cognition, and for the distribution of standard deviations, σ_h , at every layer. Taking a simple assumption that the distribution of σ_h at every layer is similar, would lead to a number of models moving between layers proportional to the number of models at each layer

$$dM_h/dt \sim (M_{h+1} - 2M_h + M_{h-1}). \quad (16)$$

Since the number of concepts at lower layers is much larger than at higher ones, this equation might lead to a growing hierarchy; however, combining this dynamics with Eqs. (13)–(15) would require a detailed numerical study.

Maximizing Eq. (1) even for a single layer in case of few specific objects is a highly complex problem, rarely solved. Deriving relatively simple Eq. (13) through (16) for the evolution of the entire hierarchy is a major step. Still, this section misses important mechanisms of interaction between cognition and language (Perlovsky, 2009). In the future research we will derive the necessary more comprehensive equations, and explore their solutions. In the following sections we use the above equations as an intuitive, qualitative guide for deriving simpler equations, which can be explored within the limits of the present paper.

4. Differentiation and synthesis

Qualitative examination of Eqs. (13)–(15) indicates two mechanisms with opposing tendencies: differentiation and synthesis. Differentiation drives creation of a large number of detailed models, whereas synthesis unifies these detailed models at higher hierarchical levels. 3 regimes or solution types can be identified. The first, $E, K, M \sim 0$ and their time derivatives are also near 0. This could be characterized as primordial consciousness. The second, $K \sim < 1, E \gg 1$; time derivatives are near 0. This could be characterized as traditional consciousness, there is no strivings for unknown, everything seems understood and fixed, emotional certainty in this limited knowledge is high. The third, is a knowledge-acquiring consciousness, with $(1 - K) \sim KE$ and a non-trivial dynamics.

