Contents lists available at ScienceDirect

Biologically Inspired Cognitive Architectures

journal homepage: www.elsevier.com/locate/bica

Research article

Approaches to cognitive architecture of autonomous intelligent agent

Yuriy Dyachenko[a,*], Nayden Nenkov[b], Mariana Petrova[c], Inna Skarga-Bandurova[a], Oleg Soloviov[a]

[a] *Volodymyr Dahl East Ukrainian National University, Severodonetsk, Ukraine*
[b] *"Konstantin Preslavsky" University of Shumen, Shumen, Bulgaria*
[c] *St. Cyril and St. Methodius University of Veliko Tarnovo, Veliko Tarnovo, Bulgaria*

ARTICLE INFO

Keywords:
Cognitive architecture
Artificial intelligence
Autonomous intelligent agent
Systems theory

ABSTRACT

Taking into account that the human intelligence is the only available intelligence we will find the functional relationship between neuronal processes and psychic phenomena to reproduce intelligence in artificial system. The autonomous behavior of an agent may be the consequence of a gap between physical processes and self-referential meaningful processing of information which is related but not determined by physical processes. This indeterminism can be reproduced in a cognitive architecture through the self-referential processing of information with consideration of itself as a meaningful model. We propose embodiment of cognitive architecture of autonomous intelligent agent as an artificial neural network with a feedback loop in meaningful processing of information.

Introduction

To solve a problem of reproducing of natural intelligence we distinct the problem of creating of autonomous cognitive agent (Freksa, 2015). Here we used the term agent for an entity that pursues its goal(s) by interacting with its environment both external (other agents, objects and so on) and internal (emotions, reasoning, knowledge, internal state and so on) (Chella, Cossentino, Gaglio, & Seidita, 2012). Such autonomous human-like agents are going to be used in human-computer interaction, as intelligent assistants and for the execution of cognitive tasks. Human-like or general AI must demonstrate autonomous intelligent activity based on analysis of environment and its place in it. An autonomous agent is a system able to choose its actions independently and efficiently through its sensors and effectors in the environment (Davidsson, 1995).

Cognitive architecture is defined as a computational model, based on the structure of the human mind. In this research we intend to design a cognitive architecture of the highest, fifth type of the (Samsonovich, De Jong, & Kitsantas, 2009) classification (Table 1). More generally, we agree with Miłkowski that "successful cognitive modelling is a question of satisfying multiple constraints from multiple fields of inquiry, levels of organization, and theories" (Milkowski, 2017). On this base, we put a target to describe the fundamental features of cognitive architecture of autonomous intelligent agent. The autonomous intelligent agent is defined by us as the entity that could perform purposeful non-predetermined actions based on the previous experience. For this purpose, following to the position that creating of general AI "will emerge from the intersection of two major pursuits: the reverse engineering of the brain and the burgeoning field of artificial intelligence" (Hawkins, 2017), in this paper we plan to develop cognitive architecture based on systems theory, informatics and cognitive science.

Fundamentals of the autonomous intelligent agent

One of the possible architecture of autonomous intelligent agent's structure with a body that mediating the mind and the environment, and with an emotional system interacting with the cognitive system described in Perotto et al. paper (Perotto, Vicari, & Alvares, 2004). In this architecture agent's mind receives sensorial inputs and action outputs. The cognitive system has a set of schemas as cognitive constructions which select and activate by the agent for each situation the agent experiences and depending on the desire of changing this situation (Perotto et al., 2004). The cognitive system also has a mechanism to build schemas which may be initially empty because the proposed mechanism is able to produce all its knowledge by interacting with the environment, while it carries out its activity does not require any pre-programming (Perotto et al., 2004). But up to date, we have an unclear gap between physically caused determination in brain processes and human's free will. The human brain can be considered as an extremely complicated network of synaptically connected neurons providing

[*] Corresponding author at: Volodymyr Dahl East Ukrainian National University, 59a, pr. Tsentral'nyi, Severodonetsk 93406, Ukraine.
E-mail address: yuriy.y.dyachenko@gmail.com (Y. Dyachenko).

https://doi.org/10.1016/j.bica.2018.10.004
Received 18 September 2018; Accepted 16 October 2018
2212-683X/ © 2018 Elsevier B.V. All rights reserved.

Table 1
Hierarchy of intelligent agent architectures due to (Samsonovich et al., 2009).

Cognitive architecture type	The agent is capable of	Level
5. Meta-cognitive and self-aware	Modelling mental states of agents, including own mental states, based on the self concept	Highest
4. Reflective	Modelling internally the environment and behavior of entities in it	High
3. Proactive, or deliberative	Reasoning, planning, exploration and decision making	Middle
2. Reactive, or adaptive	Sub-cognitive forms of learning and adaptation	Low
1. Reflexive	Pre-programmed behavioral responses	Lowest

direction of transmission of the controlling electrical signals mostly from the sensory structures toward the motor ones. In this case, however, the existence of numerous recurrent neuronal networks should also be taken into consideration with taking into account the mental phenomena (Soloviov, 2015). Under these conditions, the flow of impulses that controls motor phenomena, which has been formed in the brain, should, somehow, take into account the subjective "bias" to certain stimuli which implicated in the formation of motor responses (Soloviov, 2015). Also, this flow of impulses should be controlled according to the experience gained during the lifetime (Kandel, 2006). So, following (Soloviov, 2015), we are faced with the necessity to explain how the subjective factor of significant selectivity with respect to the objects and phenomena of the environment is involved in the regulation of the flows of information in the brain. According to the factor of subjectivity, the living system is able to accumulate in its brain information estimated as biologically meaningful and to use it effectively in the future (Soloviov & Soloviov, 2009).

Meaning

We suppose that meaning is the instrument for the formalization of subjectivity. In this approach meaning is the reflection of subjective estimation of the environment. Mathematician Stan Ulam assumed that "AI-problem" lies at word "as" in seeing of "object as a key," "a man in a car as a passenger," "some sheets of paper as a book." And the solution is a formalization of the word "as": "until you do that, you will not get very far with your Al-problem" (Rota, 1986). We consider that "as" could be described in a relationship between internal and external representations of the cognitive agent. So we suppose (based on the Ogden and Richards' meaning triangle (Ogden & Richards, 1923)), that meaning is a relationship between concept(s) (internal (with respect to an agent), subjective model) and symbol(s) (in external symbolical system). Also Putnam contrary psychological theory of meaning position "meanings just ain't in the head" noted that meaning largely determined by the external environment (Putnam, 1973). Also we are taking into account Frege's proposition of the context principle, which is characterized by Dummett as "the thesis that it is only in the context of a sentence that a word has a meaning: the investigation therefore takes the form of asking how we can fix the senses of sentences containing words for numbers" (Dummett, 1993) and Shanon's position about context dependence of meaning (Shanon, 1987; Shanon, 1988). In the field of ontological engineering and semantic technology, "context is defined to be the circumstances that form the setting for an event, statement, or idea, and in terms of which it can be better understood and assessed" (Ontology Summit 2018 Summary, 2018). Due to Prokopchuk, the meaning is the relation between the subject and the object or the phenomenon of reality, which is determined by the place of the object (phenomenon) in the life of the subject, identifies this object (phenomenon) in the image of the world and is embodied in personal structures that regulate the subject's behavior in relation to the given object (phenomenon) (Prokopchuk, 2012). It is close to semantic spaces aims to create representations that are capable of capturing meaning. So we define meaning as a relationship between internal and external semantic spaces. In vector space models, meanings are represented as points in vector spaces, also referred to as semantic spaces. Osgood, Suci, and Tannenbaum (1957) in order to measure meaning, propose creating of semantic spaces by a method of the semantic differential as a type of a rating scale designed to measure the meaning of objects, events, and concepts. Also, we could take words meanings through vector representations of words (Rumelhart, Hinton, & Williams, 1986) produced by Artificial Neural Networks (ANN) from collections of texts (Bengio, Ducharme, Vincent, & Janvin, 2003). Following Luhmann, "if the meaning selection is attributed to the system itself, then what occurs is characterized as action" (Luhmann, 1995), i.e., actions originated and formed by meaning. Luhmann considered meaning as a shaping factor of intelligence system and described meaning as the processing of itself by itself, according to differences, i.e., transforming itself (self-modification) due to meaning. For example, human brain can learn by rewiring through changing of brain structure due to acquiring new information. Based on Luhmann's thesis "meaning equips an actual experience or action with redundant possibilities" (Luhmann, 1995), we conclude that the meaning is the result of acting of subjective evaluation of information. Therefore, intelligence is an emergent property of a system in the world which manifested in effecting that formed on the base of meaning. Thus, the first condition of autonomous activity of intelligent agent is a meaning.

Evolutionary feasibility determines a physical embodiment of human mind as an example of intelligence based on natural neural nets in the brain. For this purpose, intelligence needs to build a model of the world to predict world's changing and behave on the base of subjective intentions. Human sensing is acquiring of external information from environment that form observations. In human mind on the basis of observations, experience and concepts as models are formed. The nature of produced ideas describes as representations that are product of cognitive work (Shanon, 1987; Shanon, 1988). Following Barsalou (1999), meaning occurs by the re-enactment of observations aroused during the acquisition of models. On the basis of meanings, predictions and expectations as intentions concerning the future are formed. Intentions are manifested in actions. And finally, emotions provide feedback by relationships between intentions and actions.

Feedback

In addition to sequential processing of feedforward signals, information in mind processed involving feedbacks to influence sources of signals to adjust processing (Varbanov et al., 2018). It is noted in (Petro & Muckli, 2017), "the integration of feedforward and feedback signals is important for healthy cognition and consciousness." Particularly, feedback processing is central to the enticing narrative that the brain predicts its environment (Park & Friston, 2013). In Adaptive Resonance Theory, feedback provides hypotheses for object representations in the sensory signal (Grossberg, 2013). We assume that in artificial creating of observations, models, and even intentions considerable progress has been made especially last years. But up to now, we couldn't grasp forming of meaning alongside the consideration of itself as a meaningful model. Further, we intend to describe a core of cognitive architecture to reach autonomous behavior of the intelligent agent. Thus, the second condition of autonomous activity of intelligent agent is self-reference of intelligence.

Realization of autonomous nature of intelligent agent can be achieved through causing of activity and attainment of freedom. Due to "the possibility of free action depends on the possibility of free will"

(Free Will, 2018) we need for reproduction of latter. Free will as freedom from causal determinism in human is the result of a split between neuronal processes and psychic phenomena which carried out but not determined by neuronal processes (Soloviov, 2015). Chrisley and Sloman (2016) noted that cognitive architecture with reactive, deliberative and meta-management components required "at least two layers of meta-cognition: (i) detection and use of various states of internal virtual machine components; and (ii) holding beliefs/theories about those components." These components we could interpret as a self-awareness. Also, Graziano in attention schema theory proposed that "the brain constructs not only a model of the physical body but also a model of its own internal, information handling processes" (Graziano, 2017). And "crucial difference between a personal computer and a human brain lies in the subjective experience" that associated with subjective awareness "to refer to this phenomenological property that people claim is associated with some select events and information in the brain" (Graziano, 2017). Due to Graziano "we sometimes call it intention, choice, or free will" (Graziano, 2017). Consequently, "the machine is responding on the basis of an internal model, the response can be flexible, self-consistent, and meaningful" (Graziano, 2017).

Conditions for development of autonomous intelligent agent

We state that central condition for the indeterminate activity of intelligence is self-reference of intelligence as consideration of itself as a meaningful model (Varbanov et al., 2018). Due to classical "The Cambridge Handbook of Consciousness", "consciousness is the property a system has by virtue of modelling itself as having sensations and making free decisions" (The Cambridge Handbook, 2007). Thus we suppose that general artificial intelligence could be achieved as a conscious processing of information. Dehaene et al. distinguish two essential dimensions of conscious computation: global information sharing and self-monitoring, if jointly and correctly implemented, may provide machines with conscious subjectivity (Dehaene, Lau, & Kouider, 2017). In our opinion, global availability is an understanding of something as a whole through meaningful processing of information. A working hypothesis of Dehaene et al. is that subjective experience comes down to nothing but a combination of specific forms of processing (including reality monitoring as a crucial component) (Dehaene, Lau, & Kouider, 2018). Based on information processing nature of subjective experience we suppose that self-referential meaningful processing of information is a key feature of intelligence.

Self-referencing in processing of information

Locke defines consciousness as "the perception of what passes in a man's own mind." Lamme noted central importance of recurrent interaction as a functional basis for consciousness and even define consciousness as recurrent processing (Lamme, 2006). We suppose that human mode of conscious consideration of itself is self-awareness. From a psychological point of view, self-awareness is defined by Alain Morin as "the capacity to become the object of one's own attention" (Morin, 2006). By Luhmann's view, distinguishing the system and the environment (Luhmann, 1992) and contradictions are central to the existence of autopoietic systems because they survive as the unities of contradiction that they have mastered (Bausch, 2001). In their reproduction, systems find their boundaries in basal self-reference; they proceed to reflexivity as communication about communication; and then to reflection when they recognize their separations from their environments (Bausch, 2001). Cox (2005a) noted that "some insight into the content of one's mind resulting in an internal feedback for the cognition being performed and a judgment of progress (or lack thereof)". On the Hofstadter's point of view, a model of reality of an active brain must include a model of itself – of "bodies" and internal "algorithms," that creates a (by (Hofstadter, 2007) term) "strange loop" in terms of model representations (Ruffini, 2007) as a cyclic with self-reference. Hofstadter stated that in brain-mind "there is a kind of perceiver of the symbols' activity" and "this "perceiver" is itself just further symbolic activity" (Hofstadter, 2007). Such positing of self as one of the meaningful models corresponds to self-awareness (Hofstadter, 2007). Due to Hofstadter's position, "we are self-perceiving, self-inventing, locked-in mirages that are little miracles of self-reference" (Hofstadter, 2007). On this basis, Ruffini proposed a simple architecture of agent (Ruffini, 2017) driven on the base of internal models with feedback. In Allen and Friston neuronal architecture (Allen & Friston, 2016) individual sub-components are distinguished by a feed-forward specialization without functionalist 'goals,' 'desired outputs' or 'motor commands' (Allen & Friston, 2016). Lateral connections and global precision-carrying signals link the network into a 'centrifugal' hierarchy with 'outer' layers, and 'inner' layers and last may be described as a global or self-model (Allen & Friston, 2016). This architecture does not need external goal setting to establish autonomous behavior. By Tegmark, conscious, controlled global thinking of intelligence "has high integration because of feedback loops, where all the information you're aware of right now can affect your future brain states" (Tegmark, 2017). Such position refers to Tononi's Integrated Information Theory (Oizumi et al., 2014), whereby consciousness requires a grouping of elements within a system that has physical cause-effect power upon one another. This, in turn, implies that only reentrant architecture consisting of feedback loops, whether neural or computational, will realize consciousness (Integrated Information Theory of Consciousness, 2018). Also by Lewis (2017), Cox stated that meta-cognitive feedback loop is required for self-awareness; he argues that being aware of oneself is not merely about possessing information, but also about using that information to modify goals, including information to gather in future (Cox, 2005b).

Meaningfulness in processing of information

Above mentioned definitions focus on feedback loop as a core of self-awareness. This feedback related to the meaningful processing of information. By Tegmark, "conscious beings giving meaning" (Tegmark, 2017). As Luhmann stated, "meaning can insert itself into a sequence that is bound to bodily feelings; then it appears as consciousness. But meaning can also insert itself into a sequence that involves others' understanding; then it appears as communication. Whether meaning is actualized as consciousness or as communication does not reveal itself "only afterwards", but determines any respective actualization of meaning, because meaning is always constructed self-referentially and therefore always includes reference to others as the way to self-reference" (Luhmann, 1995). Consequently, the development of autonomous intelligent agent is impossible without the acquisition of self-awareness as self-referencing of intelligence due to consideration of itself as a meaningful model. According to the Luhmann's systems theory, self-awareness is a product of uncertainty and unpredictability in the form of double contingency as consequences of the interaction of contingent systems. "Neither necessary nor impossible; it is just what it is (or was or will be), though it could also be otherwise" (Luhmann, 1995). In double contingency "two black boxes, by whatever accident, come to have dealings with one another. Each determines its own behavior by complex self-referential operations within its own boundaries" (Luhmann, 1995).

Reproduction of autonomous intelligent agent in artificial system

The main difficulty during the realization of general AI is the contradiction between predictability of classical physical systems and autonomous behavior of an agent as freedom from causal determinism. We suppose that it can result from a split between physically determined processes and self-referential meaningful processing of information which related but not determined by physical processes. Earlier we stated that autonomous intelligent activity is ensured by an acquisition of self-awareness. There is a common approach (Cox, 2005a) to describe self-aware computer system as exhibited self-awareness as generating of explicit goals. Self-aware computing systems are computing systems that learn models through capturing knowledge and reason using the models in accordance with higher-level purposes, which also be subject to change (Kounev et al., 2017). We emphasize the internal feedback as an essential feature of self-awareness. The advantage of this approach is that it gives more possibilities than, for example, self-learning neural networks (Pukala, 2016).

This autonomous intelligent agent can be implemented through ANN with a feedback loop in the meaningful processing of information. Whereas traditional computing is based on pre-determined logical reasoning, ANN represents evolution-based reasoning after learning. ANNs are related to cognitive modelling because in human brain cognition emerges from the activity of neural networks that carries information from one cell assembly or brain region to another. Thus we could distinguish two features of a proposed core of cognitive architecture (Varbanov et al., 2018): 1) self-referential loop in 2) processing of meaning. As it is marked above, implementation of self-referential loop related to meanings could acquire general abilities of intelligence. Based on this we propose a realization of cognitive architecture of autonomous intelligent agent (Fig. 1). In this scheme, artificial cognitive functions are realized through observing and modelling (Varbanov et al., 2018). Observing ANN acquires external information from the environment as input, produces perceptions as interpretations of sensory information and recognizes symbols from external symbolical system (Varbanov et al., 2018). In Modelling ANN experience and concepts are formed and memorized as the models from Observing ANN output with feedback that change Observing ANN through learning (Varbanov et al., 2018). Meanings produces in Meaning ANN as a relationship between internal concepts that acquiring from Modelling ANN and symbols that observing from external environment through Observing ANN (Varbanov et al., 2018). On this basis meaningful actions as output are formed. In this scheme we suppose arrows as 1) informational influence on ANN; 2) structural and parametric influence on ANN such as neural plasticity in natural neural nets. Idea of this core of cognitive architecture is not far from Bengio's proposition about using of ANN with feedback loops: recurrent neural network "as a tool for exploring interpretations or plans or to sample predictions about the future" or "tool to isolate a particular high-level abstraction and extract the information about it" (Bengio, 2017). Also mentioned, "task to involve meaningful abstractions which have a high predictive power" (Bengio, 2017). Operation with meaning could be proposed within knowledge representations in semantic spaces. One of the approaches dealing with semantic spaces is the representation of knowledge in geometrical structure. For example, Gärdenfors proposed Conceptual Space as a metric space in which entities are characterized by quality dimensions (Gärdenfors, 2000) which can be directly related to perceptual information or can be more abstract (Lieto, Chella, & Frixione, 2017). Concepts are represented as regions in Conceptual Spaces. Conceptual Spaces enable a more transparent interpretation of underlying neural network representations, by limiting the opacity problems of this class of formalism, and it may constitute a sort of blueprint for the design of such networks (Lieto et al., 2017).

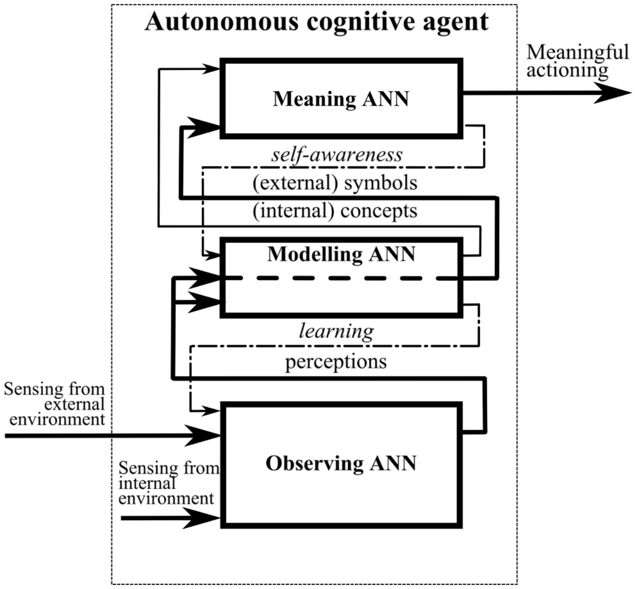


Fig. 1. Functional structure of core of cognitive architecture.

Discussion

Integrated Information Theory of consciousness claims that, at the fundamental level, consciousness is linked to integrated information, which can be represented by a precise mathematical quantity called Φ. Integrated information Φ quantifies how irreducible a system's cause-effect structure is to those of its parts and serves as a general measure of complexity. Φ calculated as an effective information ei of the minimum information partition P (MIP) in a system X with state x:Φ

$$\Phi(X,\mathrm{x}) = \mathrm{ei}(\mathrm{X},\mathrm{x},\mathrm{MIP}(\mathrm{x})), \tag{1}$$

where

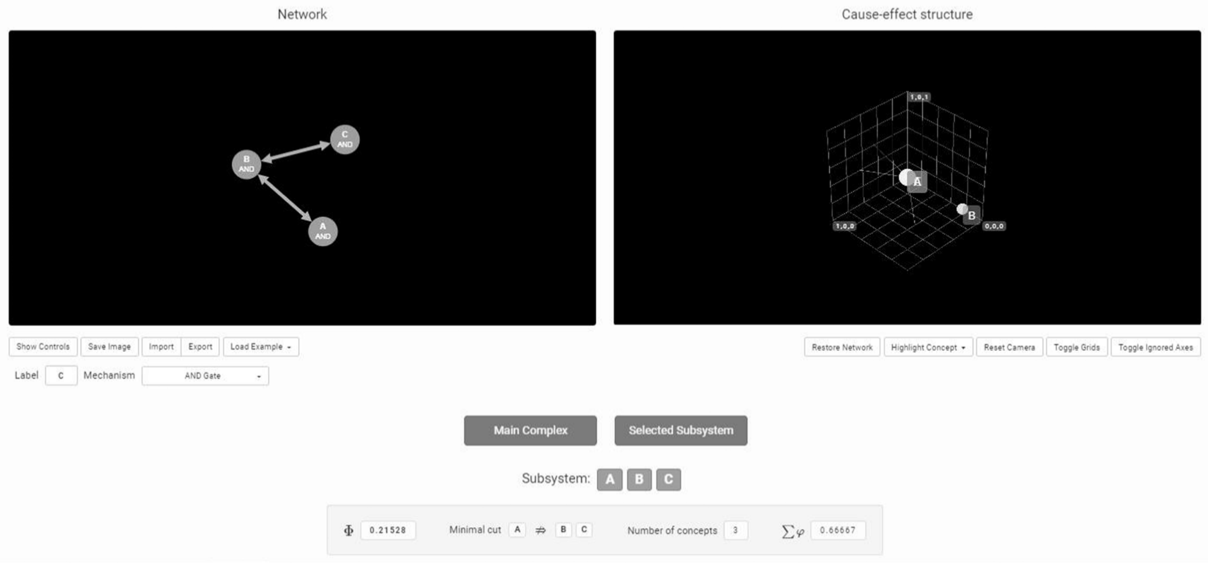
$$\mathrm{MIP}(\mathrm{x}) = \arg\ \min\{\mathrm{ei}(\mathrm{X},\ \mathrm{x},\mathrm{P})\}. \tag{2}$$

Φ captures how much the cause-effect repertoires of the system's mechanisms are altered and is a quantitative measure referred to consciousness ability (Oizumi et al., 2014).

We calculate value of integrated information Φ with using Python library PyPhi and visual interface to this code available on http://integratedinformationtheory.org/calculate.html. For proposed core of cognitive architecture we find Φ = 0,22 (Fig. 2) that confirms possibility of consciousness ability for system based on proposed core of cognitive architecture.

On the Lieto's position, in designing a general cognitive architecture, the Conceptual Spaces offer the common ground where all the representations find a theoretically and geometrically interpretation (Chella et al., 2012). Many approaches proposed in computational linguistics and distributional semantics (Mikolov, , 2013) aiming at learning vector structures, called word embeddings, from amounts of textual documents that represent the meaning of words as points in high-dimensional Euclidean space (Lieto et al., 2017). However, the dimensions of a word embedding space are essentially meaningless since they correspond, given an initial word, to the most statistically relevant words co-occurring with it, while quality dimensions in Conceptual Spaces directly reflect salient cognitive properties of the underlying domain (Lieto et al., 2017). One of most prominent approach dealing with meanings could be based on Eliasmith's semantic pointers as neural representations that carry partial semantic content and are composable into the representational structures necessary to support complex cognition (Eliasmith, 2013).

Conditions of stability of behavior of autonomous intelligent agent based on a system with feedback could be investigated in the frames of computational neuroscience approach that "…tries to describe the dynamics of networks of neurons and synapses with realistic models and plasticity which carried out by means of mental processes to reproduce emergent properties or predict observed neurophysiology… and associated behavior" (Deco, Jirsa, & Friston, 2012). Particularly, attractor theory (Brunel & Wang, 2001) introduced "to capture the neural computations inherent in cognitive functions like attention, memory, and decision making" (Deco et al., 2012). In (Cruse & Schilling, 2015) "speculated that subjective experience might occur in a recurrent neural network that is equipped with attractor properties. Following this hypothesis, subjective experience would occur if such a network approached its attractor state".



Label C Mechanism AND Gate

Subsystem: A B C

Φ 0.21528 Minimal cut A ⇏ B C Number of concepts 3 $\sum\varphi$ 0.66667

Fig. 2. Model of core of cognitive architecture and visual interface of its Φ calculation.

Both in the brain “endogenous fluctuations are an important example of dynamics and structure and may disclose fundamental principles of self-organization that underwrite the brain’s remarkable capacity to support itinerant and adaptive dynamics” (Deco et al., 2012) and in, on our opinion, it may be true for ANNs.

Conclusion and future work

Autonomous behavior is the result of meaningful interacting of intelligent agent with environment. The core of architecture of autonomous intelligent agent is the self-referential meaningful processing of information. This architecture could be embodied in the artificial system through using of ANN with a feedback loop in the processing of semantic vectors.

We saw a possibility to check the abilities of the proposed core of cognitive architecture in a software package for simulating large-scale neural systems Nengo that helps implement complex high-level cognitive algorithms (Eliasmith, 2013). For example, Semantic Pointer Architecture Unified Network (SPAUN) model use semantic pointers that are neural representations that carry partial semantic content and are composable into the representational structures necessary to support complex cognition (Eliasmith, 2013). SPAUN based on Nengo demonstrates how a wide variety of cognitive and non-cognitive tasks can be integrated with a single large-scale, spiking neuron model (Eliasmith, 2013). But up to now, we couldn’t estimate the scale of ANN that needs to operate with a meaning of self in the self-referential loop.

Acknowledgements

This work was supported by a research project RD-08-104/02.02.18, Applied models of artificial intelligence for data analysis and optimization of administrative systems of “Konstantin Preslavsky” University of Shumen (Bulgaria).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bica.2018.10.004.

References

Allen, M., & Friston, K. J. (2016). From cognitivism to autopoiesis: Towards a computational framework for the embodied mind. *Synthese,* 1–24.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*(4), 577–660.

Bausch, K. (2001). *The emerging consensus in social systems theory.* New York: Kluwer Academic.

Bengio, Y. (2017). The Consciousness Prior < https://arxiv.org/abs/1709.08568 > .

Bengio, Y., Ducharme, R., Vincent, P., & Janvin, C. (2003). A neural probabilistic language model. *The Journal of Machine Learning Research, 3*, 1137–1155.

Brunel, N., & Wang, X. J. (2001). Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *Journal of Computational Neuroscience, 11*, 63–85.

Chella, A., Cossentino, M., Gaglio, S., & Seidita, V. (2012). A general theoretical framework for designing cognitive architectures: Hybrid and meta-level architectures for BICA. *Biologically Inspired Cognitive Architectures, 2*, 100–108.

Chrisley, R., & Sloman, A. (2016). Architectural requirements for consciousness. *Proceedings of EU Cognition Meeting “Cognitive Robot Architectures”,* 31–36.

Cox, M. T. (2005a). Perpetual Self-Aware Cognitive Agents. In Metacognition in Computation: Papers from 2005 AAAI Spring Symposium (pp. 42–48).

Cox, M. T. (2005b). Metacognition in computation: A selected research review. *Artificial Intelligence, 169*(2), 104–141.

Cruse, H., & Schilling, M. (2015). Mental states as emergent properties. From walking to consciousness. *Open Mind, 9*(C), 1–38.

Davidsson, P. (1995). On the concept of concept in the context of autonomous agents. In WOCFAI 95: Proceedings on the Second World Conference on the Fundamentals of Artificial Intelligence, Angkor (pp. 95–145).

Deco, G., Jirsa, V., & Friston, K. J. (2012). *The dynamical and structural basis of brain activity. Principles of brain dynamics: Global state interactions.* Cambridge: MIT Press.

Dehaene, S., Lau, H., & Kouider, S. (2017). What is consciousness, and could machines have it? *Science, 358*, 486–492.

Dehaene, S., Lau, H., & Kouider, S. (2018). Response. *Science, 359*, 400–402.

Dummett, M. (1993). *The origins of analytical philosophy.* Cambridge: Harvard University Press.

Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition.* New York, NY: Oxford University Press.

Free Will. (2018). In The Internet Encyclopedia of Philosophy, < http://www.iep.utm.edu/freewill/ > .

Freksa, C. (2015). Strong spatial cognition. In 12th International Conference, COSIT 2015, Santa Fe, NM, USA, October 12-16, 2015, Proceedings (pp. 65–86).

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought.* MIT Press.

Graziano, M. S. A. (2017). The attention schema theory: A foundation for engineering artificial consciousness. *Frontiers in Robotics and AI, 4*, 60.

Grossberg, G. (2013). Adaptive resonance theory: How a brain learns to consciously attend, learn, and recognize a changing world. *Neural Networks, 37*, 1–47.

Hawkins, G. (2017). What intelligent machines need to learn from the neocortex. *IEEE Spectrum Magazine,* 35–40.

Hofstadter, D. R. (2007). *I Am a Strange Loop.* New York: Basic Books.

Integrated Information Theory of Consciousness. (2018). In The internet encyclopedia of philosophy, http://www.iep.utm.edu/int-info/.

Kandel, E. R. (2006). *The search of memory. The emergence of a new science of mind.* New York & London: Norton & Co.

Kounev, S., et al. (2017). The notion of self-aware computing. *Self-aware computing systems* (pp. 3–16). Berlin: Springer Verlag.

Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences, 10*, 494–501.

Lewis, P. R. (2017). Self-aware computing systems: from psychology to engineering. In Proceedings of the 2017 Design, Automation and Test in Europe (DATE 2017) (pp. 1044–1049).

Lieto, A., Chella, A., & Frixione, M. (2017). Conceptual spaces for cognitive architectures: A lingua franca for different levels of representation. *Biologically Inspired Cognitive Architectures, 19*, 1–9.

Luhmann, N. (1992). The concept of society. *Thesis Eleven, 31*, 67–80.

Luhmann, N. (1995). *Social systems.* Stanford: Stanford University Press.

Mikolov, T., et al. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems, 26*, 3111–3119.

Milkowski, M. (2017). The false dichotomy between causal realization and semantic computation. *Hybris,* 1–21.

Morin, A. (2006). Levels of consciousness and self-awareness: A comparison and integration of various neurocognitive views. *Consiousness and Cognition, 15*(2), 358–371.

Ogden, C. K., & Richards, I. A. (1923). *The meaning of meaning. A study of the influence of language upon thought and of the science of symbolism.* New York: Harcourt Brace and World.

Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0. *PLoS Computational Biology, 10*(5).

Ontology Summit 2018 Summary. (2018). http://ontologforum.org/index.php/OntologySummit2018_ResearchSummary.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning.* Urbana: University of Illinois Press.

Park, H. J., & Friston, K. (2013). Structural and functional brain networks: From connections to cognition. *Science, 342*, 1238411.

Perotto, F. P., Vicari, R., & Alvares, L. O. (2004). An autonomous intelligent agent architecture based on constructivist AI. *Artificial Intelligence Applications and Innovations, 154*, 103–115.

Petro, L. S., & Muckli, L. (2017). The laminar integration of sensory inputs with feedback signals in human cortex. *Brain and Cognition, 112*, 54–57.

Prokopchuk, Y. (2012). *Principle of limiting generalizations: Methodology, tasks, Appications.* Dnepropetrovsk: Institute of Technical Mechanics of NASU and NCAU.

Pukala, R. (2016). Use of neural networks in risk assessment and optimization of insurance cover in innovative enterprises. *Economics and Management, 8*(3), 43–56.

Putnam, H. (1973). Meaning and reference. *The Journal of Philosophy, 70*(19), 699–711.

Rota, G. C. (1986). In memoriam of Stan Ulam: The barrier of meaning. *Physica, 22D*, 1–3.

Ruffini, G. (2007). Information, complexity, brains and reality ("Kolmogorov Manifesto"). < https://arxiv.org/abs/0704.1147 > .

Ruffini, G. (2017). An algorithmic information theory of consciousness. *Neuroscience of Consciousness, 3*(1).

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by backpropagating errors. *Nature, 323*(6088), 533–536.

Samsonovich, A. V., De Jong, K. A., & Kitsantas, A. (2009). The mental state formalism of GMU-BICA". *International Journal of Machine Consciousness, 1*, 111–130.

Shanon, B. (1987). On the place of representations in cognition. In Thinking: The Second International Conference (pp. 33–49).

Shanon, B. (1988). Semantic representation of meaning: A critique. *Psychological Bulletin, 104*(1), 70–83.

Soloviov, O. V. (2015). Neuronal networks responsible for genetic and acquired (ontogenetic) memory: Probable fundamental differences. *Neurophysiology, 47*(5), 419–431.

Soloviov, O. V., & Soloviov, S. O. (2009). On crucial dissimilarities of determinism in informational networks of the human brain and computers. *Shtuchnyi Intelekt, 3*, 11–22.

Tegmark, M. (2017). *Life 3.0: Being human in the age of artificial intelligence.* London: Allen Lane.

The Cambridge Handbook of Consciousness. (2007). New York: Cambridge University Press.

Varbanov, S., Dyachenko, Y., Skarga-Bandurova, I. (2018). Functional structure of autonomous intelligent agent. In Information technologies, management and society: Theses of 16th international scientific conference "information technologies and management" April 26–27, Riga, Latvia (pp. 24–25).