sciendo

STUDIES IN LOGIC, GRAMMAR AND RHETORIC 67 (80) 2022

DOI: 10.2478/slgr-2022-0009

This work is licensed under a Creative Commons Attribution BY 4.0 License (http://creativecommons.org/licenses/by/4.0)

Włodzisław Duch
Department of Informatics, Institute of Technical Sciences
Faculty of Astronomy, Physics, and Informatics
and
Neurocognitive Laboratory, Center for Modern Interdisciplinary Technologies,
Nicolaus Copernicus University, Toruń, Poland
e-mail: wduch@umk.pl
ORCID: 0000-0001-7882-4729

CONCEPT REPRESENTATION AND THE GEOMETRIC MODEL OF MIND

Abstract. Current cognitive architectures are either working at the abstract, symbolic level, or the low, emergent level related to neural modeling. The best way to understand phenomena is to see, or imagine them, hence the need for a geometric model of mental processes. Geometric models should be based on an intermediate level of modeling that describe mental states in terms of features relevant from the first-person perspective but also linked to neural events. Concepts should be represented as geometrical objects that have sufficiently rich structures to show their properties and their relations to other concepts. The best way to create such geometrical representations of concepts is through the approximate description of the physical states of neural networks. The evolution of brain states is then represented as a trajectory linking successful concepts, and topological constraints on the shape of such trajectory define grammar and logic.

Keywords: Concepts, geometric model, neural modeling, physical states, neural networks.

1. Understanding mental processes

What is the best way of describing mental processes that can be linked to our inner perspective? Electrons and protons form atoms, atoms form molecules, molecules form cells and cells form tissues, neural layers, specialized regions, organs, and whole organisms with brains that control their homeostasis. Interacting organisms with brains form societies. The physical realm (in Karl Popper's terminology World 1) gives rise to the mental realm (World 2) and this leads to the cultural realm (World 3). This is enabled through the interaction of constituent elements. If the interactions

are sufficiently strong, higher-level entities may emerge. How do the mental concepts emerge from interactions, and what basic constituents do we need to create them?

Mathematics has two large branches. Algebra deals with symbolic, discrete structures. Geometry represents in some spaces shapes and continuous values of observed quantities of dynamic processes. Traditionally, mental processes have been described at the symbolic level, creating a sharp, seemingly insurmountable division between neuroscience and psychology. Newell and Simon (1976) advocated a symbolic approach, formalized in their Physical Systems Symbol Hypothesis as "Natural cognitive systems are intelligent in virtue of being physical symbol systems of the right kind." Symbolic Artificial Intelligence has failed in many ways, especially in a subject related to natural perception, such as vision, but also in higher mental functions, especially the use of language. The "Standard Model of Mind" (Laird, Lebiere and Rosenbloom, 2017) is perhaps the last attempt to formulate a model at the symbolic level, based on the interactions between different types of memory. Symbols have a great advantage because they can be combined into relations and complex networks of relations. Standard Model of Mind assumes that declarative and procedural long-term memory contain symbolic structures and communicate (interact), constituting associative networks.

Symbolic AI has been challenged by the connectionist and the dynamical systems' theory approaches to cognition. Treating the "physical" part of symbol representation seriously helps to elucidate connections between symbolic, connectionist, and dynamical approaches. Physical symbols may be understood not as tokens, but as patterns, in the spirit of Goertzel's "patternist philosophy" (Goertzel 2006). Patterns represent activation of the brain areas and may be defined with different spatial and temporal resolutions. For example, functional magnetic resonance (fMRI) scans provide patterns of the whole brain activation every second, with a high spatial resolution of about one cubic millimeter. Electroencephalography (EEG) can provide patterns every millisecond, but only of cortical surface activity, with a spatial resolution of a centimeter. Neuroscientists are convinced that mental states supervene on patterns of brain activity. When we recognize an object or a word, a pattern may persist for a small fraction of a second (be quasi-stable). When we think, talk or play a game, patterns may change rapidly. Quick transitions between different synchronized states of neuronal activity happen in less than 100 milliseconds.

Synchronized states are good candidates to create patterns that may be interpreted in a symbolic way. A mathematical technique called "symbolic dynamics" is using discretized sequences of similar patterns, converting

them into a series of symbols (Dale and Spivey 2005). A surprising result of recent research confirms "causal emergence": a macroscale representation of the system state (a symbol assigned to the low-resolution pattern) may carry more information than a detailed microscale representation (high-resolution pattern). Hoel's (2017) article "When the map is better than the territory", based on the application of information theory to causal analysis, shows that the "causal structure of some systems cannot be fully captured by even the most detailed microscale description". At the macroscale information may be converted from one type to another. The same information may be expressed in different ways. At the microscales' there is no way to define equivalent patterns that contain the same information. Even in Boolean networks, the synergistic information component of mutual information can increase at the macroscale (Varley and Hoel 2021). A similar phenomenon is known in quantum mechanics: the best possible knowledge of the states of the whole system is not sufficient to have the best possible knowledge of all parts of the system, and vice versa, perfect knowledge of all parts does not imply perfect knowledge of the whole system (Duch 1989; Varley and Hoel 2021). Causal emergence seems to be common in quantum as well as classical complex systems.

It is worthwhile to connect the formation of patterns as physical symbols with approximations to the neurodynamics of real brain processes. This may be done at a different level, from the symbolic dynamics logical approximations to graded, fuzzy types of description that allow for a more accurate approximation of neurodynamics (Duch 1989, 1996, 1997). Brain processes described by parameters derived from neuroimaging show the activity of many neural cell assemblies or individual neurons that are not directly related to mental events or inner experiences. The verbal, psychological description of mental events is usually a confabulation that ignores the real neurodynamical forces responsible for the creation of mental states. Transitions and relations between neurodynamic patterns may be represented in a geometrical model. However, instead of dimensions based on physical measurements, spaces in which mental events take place should be defined using dimensions that can be related to our inner experience. Such models, mapping brain activities to mental processes, have greater explanatory power than symbolic models (Duch 2010, 2012).

Attempts to describe the inner experience, or phenomenal consciousness, started in the 19th century by the introspective psychologist. Controversies between the first experimental psychology laboratories run by Wundt and Titchener on such issues as the existence of "imageless thoughts" are documented in Brock (2013). The development of phenomenology in phi-

losophy by Husserl, Ingarden, and others also did not help to resolve it. Hurlburt and Schwitzgebel (2007) conducted experiments with the random descriptive experience sampling technique. At random times Melanie, their subject, heard a sound signal, reminding her to note what was exactly going on in her mind. Transcripts of these notes cast doubt on the possibility to create a science of everyday inner experience. The reason for this difficulty is fairly clear: the mind, metaphorically speaking, is a shadow of neurodynamics. A lot of processes inaccessible to conscious introspection are hidden behind the scenes, controlling homeostasis and general behavior. In the famous allegory presented in Plato's *Republic* prisoners in a cave can see only shadows of real things projected on the wall, while the task of the philosopher is to perceive the true form of things. Mental events resemble such shadows of physical reality, which is represented more closely by neurodynamics), reflecting, or rather discovering actively those features of the environment that are important for survival. John Locke (1690) defined consciousness as "the perception of what passes in a man's own mind". What we can perceive are just peaks of neural activity that are sufficiently persistent to be categorized through association with quasi-stable brain activity patterns, expressed either by a verbal symbol (manifested as speech or silent thought) or by motor actions.

2. Geometrical models of cognition in the past

The great success of physics in understanding the physical world has inspired early psychologists to construct analogical theories for the inner world. Kurt Lewin in his book *The Conceptual Representation and the Measurement of Psychological Forces* (1938) tried to understand the behavior of an autonomous agent as a trajectory controlled by forces in the psychological space. His force field analysis is still used in social psychology to find factors that drive the movement towards the goal, block or divert it. The 3-stage description of the process of mental change: unfreezing or escaping the inertia, transition without a clear idea where it leads, and freezing or crystallizing new behaviors, can be directly linked to the process of skill learning and psychotherapy at the neural level. It can also be interpreted at the neural network level: desynchronization of the current state, chaotic exploration to find new associated states, and reaching a new synchronized state (attractor state in dynamical systems).

George Kelly in his book *The Psychology of Personal Constructs* (1955) proposed an explicit geometrical representation of personality and visual-

ization of psychological processes. The main assumption of his approach is that individual psychological processes are channeled by anticipations. In neuroscience anticipations are top-down influences on attractor neural states formed in the brain, necessary for resolving ambiguities in signal interpretation. Brains internalize some environmental regularities and can use them, starting from sensory recognition to predictions of rewards for planned actions. According to Kelly internal models may be described by "constructs" or dimensions that are used by people for discrimination, for example, "good-bad" or "happy-sad". Such dimensions may be inferred using principal component analysis on sets of synonyms and antonyms directly from texts, as shown by Ascoli and Samsonovich (2012) in their cognitive map. Kelly was convinced that mental states may be characterized by a small number of constructs (Neimeyer, R.A. and Neimeyer, G.J. 2002). This view is corroborated by the successes of the Latent Semantic Analysis which is used to reduce the dimensionality of the space in which words are embedded. The Natural Semantic Metalanguage theory of semantic universals (Wierzbicka 1996) is an attempt to find, by trial and error, a "semantic core" common to all languages, a set of primitives (about 60 have been identified) that is sufficient for the description of all other terms.

Kelly's idea was to use personal constructs as dimensions of psychological space that may be used to characterize people and mental events. Subjective reality is expressed in terms of constructs, and thus understanding people is possible if the constructs that they use are correctly recognized. Results are presented in a matrix form, called Repertory Grid, with rows representing constructs, while columns represent various types of elements, for example, people or mental states. Techniques based on personal construct psychology (PCP) have wide applications in social psychology, psychotherapy, personality assessment, and human resources in a business context.

Other influential approaches that tried to use geometric ideas to describe concepts and understand cognition can be found in the *Conceptual Spaces: The Geometry of Thought* (2004) and *The Geometry of Meaning: Semantics Based on Conceptual Spaces* (2014) books of Gärdenfors, *Mental Spaces: Aspects of Meaning Construction in Natural Language* book of Fauconnier (1994), and the book by Fauconnier and Turner (2003) on metaphors and conceptual blending based on simple bodily experiences (2003). Earlier, Johnson-Laird developed the idea of mental models (1983, 1995) that show how associations between concepts help in reasoning.

We are now reaching an understanding of neural processes that can justify such diverse ideas using computational models and verifying them with brain neuroimaging.

3. How brains create and use concepts

Linking neural activity to various categories of mental events should lead to a new level of modeling of mental processes. Fingerprints of brain states may be characterized in many ways, using computer simulations of neural networks, or analyzing real human brain neuroimaging and neurophysiological data. In recent years "brain reading" experiments have proved that the meaning of concepts can be decomposed in an interesting way. The semantic atlas, and the interactive tool to explore it, showed a distribution of functional magnetic resonance activations in the brain for over 1700 semantic categories (Huth, de Heer, Griffiths, Theunissen and Gallant 2016). Each small part of the brain (voxel, a few cubic millimeters) is activated by many concepts. It is a small part of complex patterns composed of the activity of tens of thousands of such voxels. In the case of some concepts contribution of brain regions can be identified in terms of qualia associated with them. For example, Binder et al. (2016) have defined 65 aspects of mental experience related to sensory perceptions and affective responses to entities and situations, including motor actions, causality, perception of spatial and temporal phenomena, social experiences, drive states and other internal cognitive phenomena. Estimation of how salient is each of these aspects in relation to the semantics of a concept may be done either through co-occurrence analysis in a large corpus, answers to queries, or analysis of brain activations. In the space of brain-relevant dimensions (attributes) derived from such analysis, concepts may be represented as clouds of points, with the value assigned to each attribute derived from experimental data. We can combine the activities of many voxels correlated with such qualia as recognition of sensory sensations, or emotions, and use them as dimensions in mental spaces.

Descriptions of concepts may be modeled using neurofuzzy systems that estimate the probability density of a set of attributes assigning symbolic value to the peaks of such combinations (Duch 1997). Brain-based semantics may be useful in the vector-based representation of concepts in natural language processing, adding human-like qualities to formal semantics. Previous attempts to build a semantic cognitive map (including the patent of Acoli and Samonovich 2012) were based on statistical techniques, such as the latent semantic analysis, applied to dictionaries of synonyms and antonyms. This leads to different abstract attributes, the most important being valence, arousal, dominance, and dimensions based on human value systems such as good vs. bad, or calming vs. exciting. It is hard to assign the value of these more abstract attributes to specific brain activity.

Cognitive attributes are related to specific brain activations. Fernandino et al. (2022) have investigated how activation of different representational systems in the brain help to implement semantic cognition. For both object and event-related concepts, the activity of the heteromodal cortex could be linked to the information about affective, sensory-motor, and other aspects of phenomenal experience. What is needed now is to link the description of brain activations, with attributes based on the activity of regions of interest, with models based on brain-based semantics derived from linguistic analysis. Mapping between these two spaces should not only help to understand the mind-brain-body relations, but also bridge the gap between results of neuroimaging and models of neural activity on the physical side, and description of mental events from the first person's inner perspective. The flexibility of knowledge representation by patterns of brain activity has not yet been matched by any other knowledge representation framework. Systematic approximation to neurodynamics should help to understand essential features of brain activations that are aroused by concepts.

In a large population, there will be many clusters of people with similar mental maps. Induction of concepts by education, media, or social interactions, combined with hereditary predispositions that influence perception and motor activity, leads to the creation of a specific structure of mental maps. Depending on the resolution that is used to analyze them, maps may show broad differences among cultures, subcultures, subjects, detailed levels of knowledge, political and religious orientations, or general interests. This leads to sociology. The growing world population and availability of communication technologies, especially social networks, led to a great increase in the number of different clusters, and their internal coherence.

This picture of concepts' semantics is based on the statistical distributions of observed brain activation patterns and features derived from these patterns. Concepts represented in high-dimensional spaces become points, or clouds of points, because most concepts cannot be described by a unique combination of fixed features (only abstract, mathematical concepts may have such unique character). Features may have many discrete values (like the number of legs in a chair) or distribution of values (like the colors, shapes, or sizes). The similarity is measured by the distance between points. Synonyms make larger clouds, with several smaller clouds representing words that have a very similar meaning. Two interactive models of geometrical lexical representation have been presented (Hyungsuk, Ploux and Wehrli 2003, Ploux, Boussidan and Ji 2010). A Semantic Atlas was generated using various dictionaries and thesauri by applications

of correspondence factor analysis to discover 'cliques', "a fine-grained infra linguistic subunit of meaning" (Ploux, Boussidan and Ji 2010). Associations are based on similarity of meanings, but also context, or co-occurrence of words. The second tool is called Automatic Contexonym Organizing Model. It helps to find associations that were never compiled in the traditional lexical approaches. Using cliques contextually related words (called contexonyms) are searched in large corpora. Contexonyms reflect contextual usage of concepts (Hyungsuk, Ploux and Wehrli 2003) and can also be embedded in high-dimensional spaces. Such models have many applications. For example, they provide maps of similar words in different languages, showing the network of concepts that span similar ideas expressed in ways that depend on culture (Ploux and Ji 2003).

Representations derived from large text corpora average the experiences of many people. Understanding semantics depends also on personal experience, especially in the case of more abstract concepts that require complex descriptions, like democracy or wisdom. Understanding emotions is also very subjective. The patterns of brain activity that are clustered and labeled as one emotional state are blurry, and we cannot manifest them in many ways. Some people, suffering from alexithymia, are not able to identify the emotions they experience. Associations between concepts result from learning, general knowledge contained in a conceptual network, our semantic memory. Semantic memory is largely acquired through episodic learning and these two types of memory are not completely separate. In the model of individual conceptual network the average saliency of attributes and their variance have to be adjusted to the individual "mental map" of each person. Moreover, priming effects dynamically change these values, accounting for idiosyncratic dependence on the context. Observations may become memes if they fit in the individual conceptual network. They become easily and deeply encoded. Confirmation bias is one of many effects illustrating this process. Such processes can be modeled in the spreading-activation neural networks that exhibit attractor states (quasi-stable activity patterns). The concept is represented by a word and its synonyms, Wordnet synsets, and contexonyms. It creates a cluster of similar patterns, a "Markov blanket" covering this concept, called in our neurocognitive approach to language "a coset" (Duch, Matykiewicz and Pestian 2008).

Such understanding of concepts connects psychological and neuroscience perspectives. Experiences become memories and contribute to the semantics of concepts (Heusser, Fitzpatrick and Manning 2021). Combining dynamic and static aspects helps to understand the influence of media, viral events, the spreading of memes, and the formation of conspiracy theories

(Duch 2021). New concepts may increase or decrease the order of conceptual networks, either individual or common to social groups, a phenomenon that requires more attention and simulations in neural networks (Hutchins 2012).

4. Dynamics: brain states and in the mind spaces

In the *Continuity of Mind* book Michael Spivey (2007) described mental events as continuous trajectories in the state spaces based on the activity of neurons or neural assemblies. Understanding high-dimensional neurodynamical systems is hopelessly difficult unless some way of dimensionality reduction is applied. He recommends symbolic dynamics applied to the trajectories in the state space of brain activations, but this is a crude approximation that suffers from combinatorial explosion of the number of symbols in high-dimensional spaces. Moreover, trajectories in the brain-based state space are not related to qualities of experience that are used for the description of inner experience. Neurofuzzy models (Duch 1997) combined with the transformation of neuroimaging data to spaces spanned by features related to qualia (such as those used in the brain-based semantics, Binder et al. 2016) should allow for a much better representation of conceptual spaces and networks of semantic maps. Analysis of brain dynamics requires visualization and analysis of trajectories of brain activation patterns in low-dimensional spaces (Duch 1997, 2012). This idea is finally gaining popularity in neuroscience (Varley and Sporns 2022).

A simplified, but quite useful, representation of the brain state is derived from measures of the average activity of groups of neurons, following the dynamics of their activity. Such neurodynamical description ignores a lot of details at the molecular and biochemical levels. With over 16 billion neurons in the neocortex it will be anyway too complex to handle. In some parts of the brain (for example in the motor cortex) coordinated activity of a large number of neurons is needed to create a signal correctly recognized by other parts of the brain. In other brain areas small groups of neurons (neural ensembles) may be sufficient to form distinct states. The average activity of neuronal ensembles is collected in a matrix with elements that are used to characterize the global bioelectrical state of the brain. Some components of this matrix describe the state of sensory cortices, and thus partially depend on the external stimuli. Other components represent associative, executive, and motor processes. Event segmentation theory postulates that what we perceive as continuous experience is composed of discrete events, rapidly changing by context, new scenes, discussion topics, etc. This theory has

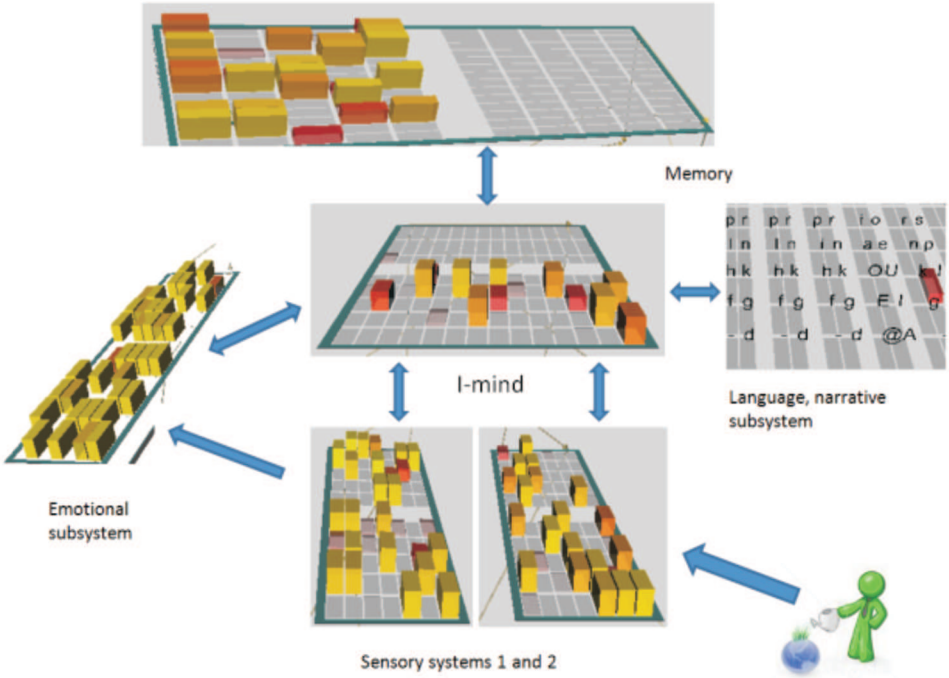


Figure 1. Neural layers representing two sensory inputs, emotional and language subsystems and memory, with I-mind layer that represents and binds percepts. Each box represents neural cell assembly, activity is represented by the hights of the box. Strength of connections between these assemblies is learned during network training.

been verified by fMRI and EEG experiments that discover active brain subnetworks and transitions between them when people listen to stories or watch videos (Speer et al. 2009; Zacks et al. 2010; Tian et al. 2021).

The description of brain activity should be sufficiently detailed to be useful in predicting the changes in the global state of the brain over time. It requires the knowledge of detailed connectivity between the brain areas that give rise to the averaged neural activity. Shun-ichi Amari has developed continuous models of neural tissue, analyzed them using topological information geometry ideas, and applied them to model concept formation in neural networks (Amari 1977). Simulations of brain functions help to understand this process (Fig. 1). The network, created using the Emergent software (O'Reilly et al. 2020), has several layers of mutually connected neurons, analyzing sensory signals (possibly each composed of several hierarchically oriented layers), an emotional subsystem that evaluates the saliency of different signals, inputs from memory, and coupling to the

language layer that includes separate semantic and phonological layers. Semantic representation is based on the activity of the whole brain (Huth et al. 2016; Binder et al. 2016). Regions that respond to experiential features of concepts are activated, and a subnetwork of synchronized brain regions emerges for a relatively short time. At each moment only a small percentage of neurons in the brain is highly active. They are shown as yellow boxes in Fig. 1. Synchronized semantic brain state has an influence on the sensory cortices, recreating images similar to those that arise during perception. It also affects auditory and motor cortices, activating phonological labels that provide symbolic reference to the semantic state. To save energy neurons desynchronize and recruit other neurons, creating a different quasi-stable state that activates another phonological label. This mechanism leads to a creation of a stream of symbols and images, narrative comments on the neurodynamics of brain states.

Percepts result from binding all this information in a distributed network, but may also involve links to more localized representations, involving a hippocampus that activates episodic memories through links to many cortical areas, or phonological representations that help to categorize diverse internal states. When a percept is established, in each layer a quasi-stable synchronized local attractor state is formed. Understanding brain activity is thus reduced to the analysis of the dynamics of its global state, finding characteristic responses, identifying attractor states of these dynamics, and charting basins of attractions in low-dimensional spaces constructed from brain signals that are compared to prototype attractor states. We are working on tools that will help us to analyze real brain signals and visualize such processes (Rykaczewski, Nikadon, Duch and Piotrowski 2021; Komorowski et al. 2021).

Memory plays an important role in the formation of networks of concepts. It was known that the hippocampus plays a major role in spatial memory, but recently it was found that it is also responsible for more abstract non-spatial memories. Memory spaces seem to link various representations of events, and attempts to understand the topological nature of these connections have started. Memory spaces are more general than cognitive maps, and models that link them directly to neural processes have already been created (Babichev and Dabaghian 2018). Conscious perception requires neural space that serves as an arena of mental events, a "theater of consciousness", with the neurodynamics that coordinates the activity of the whole brain. Let's call this internal mental stage the I-mind, a reflection of all brain processes that may be "internally perceived", not by a homunculus, but by creating associations that allow for internal reportability. Such

processes lead to actions, either in the form of direct motor activity or in the form of association with phonological representations, producing verbal, narrative comments. I am aware of the state of my higher sensory cortices that have already filtered relevant information from the raw signals that have been received by my senses.

Perception is ill-defined, sensory signals are not sufficient to recognize words, objects, or people. This is why it was so difficult to create artificial computer vision systems – they had no prior knowledge that is needed to fit observations to recognition. Words in text-to-speech systems are recognized with low accuracy, but the context in which they appear helps to increase the recognition. To do this artificial systems need prior knowledge, they are trained on large databases. Rich prior knowledge acquired from experience is stored in our brains. It enables anticipation, rapid learning, understanding, and creating. The development of very large "multimodal foundation models", huge neural networks with billions and even trillions of parameters, trained on text, images, and videos is the next challenge for artificial intelligence. Embedding human experience in such models requires similar data that our brains receive. Within a few months of Summer 2022 tens of text-to-image systems emerged (Dall-E2 by OpenAI, Imagen by Google, and Midjourney are the most popular), capable of creating novel images based on detailed text prompts describing the image[1]. It has already revolutionized the creation of all forms of visual arts. The imagery of artificial systems is based on billion images used to train such systems. Our visual cortex is composed of at least 32 cortical regions, processing information in ways we still do not understand. We have conscious access to the final result of complex transformations of visual inputs, but our imagery may change activations on the deeper, subconscious level. This leads to hallucinations, images that are not directly related to signals from the eyes. Text-to-image systems are not blending pieces of existing images, but use text descriptions to change activations in the deep layers of the networks, resulting in novel images that no human has ever imagined.

Brains require a good prior hypothesis, stored in the memory, providing top-down expectations. I-mind gets inputs from both directions: real-time sensory cortex, and the activation of past episodes stored in the memory. Elements in the I-mind layer receive information about the similarity of signals to specific prototypes and thus linguistic comments and the states of this layer have to be expressed in terms of comparisons: as sweet as honey, as red as blood, etc. Whenever we try to express a new experience, it is by trying to compare it to other experiences that may be recalled. Recently this idea has been proved using neuroimaging experiments (Zhang, Han,

Worth and Liu 2020). Semantic categories and relations between them are represented in the brain by a distributed network of spatially overlapping cortical patterns.

5. Conclusions

Concepts provide basic categories that facilitate cognitive states. We have come a long way from linguistic construction, and symbolic description that tries to define concepts by referring to other concepts. Concept maps reveal important relations between concepts, but correlations derived from analysis of texts show only the surface level. Surprisingly, causal emergence shows that the map may be more than the territory. Symbolic level, natural language, can be quite a powerful means of passing information. Kurt Lewin (1936, 1938) tried to create mental models based on inspirations from field theory, as used in physics and topology. His formal models were not connected with physical neural processes in the brain, offering only descriptive accounts of human experience (Duch 2018). This is not sufficient to understand the mechanisms responsible for mental phenomena. Harnad (1990) has formulated the "symbol grounding problem", asking: How can the semantic interpretation of a formal symbol system be made intrinsic to the system, rather than just parasitic on the meanings in our heads?" A deeper understanding of semantics requires linking concepts to experience. This can be achieved using a sensorimotor approach to perception, as described by O'Regan (2011). Sensory experiences are a result of interaction with the world. The sensorimotor approach can be fully implemented only in robots capable of active exploration of the environment and forming concept representations based on signals from their sensors.

From the engineering perspective understanding requires building a functional model. Linking concepts with experience may be done to some degree using brain-based semantics vector representations of concepts based on salient attributes that our brains care about (Binder et al. 2012). Brain-mind transformations are based on mapping neural activity into a space of mental concepts, where each dimension in the "mental space" is related to some quality of our inner experience, such as intentional feelings, tactile, auditory, visual, and other sensory features, goes only halfway towards real symbol, or category, grounding in experience. However, it should endow symbols with qualities that will make their use similar to human-like semantics. Suppose that a robot has learned the meaning of various concepts categorizing its neural network states as a result of interaction with

its environment. This structure can then be copied to the brain of another robot, providing a semantic interpretation of concepts in his artificial brain.

Understanding how mental processes are related to neurodynamics is already providing a new language for describing mental experiences, linking brain activity to the results of psychological research. The physical symbols in the I-mind are labels of discretized patterns of activity (Zacks et al. 2010). Systems with this architecture may associate symbols with their internal activation states, claiming awareness of physical I-mind states. Because brain states are continuous they cannot be precisely verbalized, segmentation creates broad categories. Such systems will have to claim qualia (Duch 2005). With a high degree of sophistication, they should be able to label their internal states with symbols that form a series analogous to the stream of inner thoughts. This is already happening in the text-to-speech large language models combined with images, that demonstrate unprecedented imagery, and can verbally describe what they create. The trajectory of the I-mind state defined in the space of phenomenologically meaningful variables reflects particular associations with the past episodes, giving each inner state specific qualities or qualia.

It is hard to avoid the conclusion that deep representation of concepts based on experience will lead us to conscious machines.

Acknowledgment
This work was supported by the National Science Center grant UMO-2016/20/W/NZ4/00354.

NOTE

[1] At the time of writing only the Internet sources can be cited, for example collected at this site: https://flipboard.com/@wlodzislaw/art-music-ai-brain-ih9t1m8py

REFERENCES

Amari, S.-I. (1977). "Neural theory of association and concept-formation". *Biological Cybernetics,* 26(3), 175–185. https://doi.org/10.1007/BF00365229

Ascoli, G.A. and Samsonovich, A. (2012). Semantic cognitive map. US Patent 8,190,422.

Babichev, A. and Dabaghian, Y.A. (2018). "Topological schemas of memory spaces". *Frontiers in Computational Neuroscience*, 12, 27.

Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M. and Desai, R. H. (2016). "Toward a brain-based componential semantic representation". *Cognitive Neuropsychology*, 33(3–4), 130–174.

Brock, A. C. (2013). "The history of introspection revisited". In: Clegg, J.W. (ed.), *Self-Observation in the Social Sciences*. London: Routledge, 25–43.

Dabaghian, Y. (2019). "Through synapses to spatial memory maps via a topological model". *Scientific Reports*, 9(1), 572.

Dale R. and Spivey M.J. (2005). "From apples and oranges to symbolic dynamics: a framework for conciliating notions of cognitive representation". *Journal of Experimental & Theoretical Artificial Intelligence*, 17(4), 317–342.

Duch, W. (1989). "Schrödinger's thoughts on perfect knowledge". In: Bitsakis, E.I. and Nicolaides, C.A. (eds.), *The Concept of Probability*. Amsterdam: Kluwer Academic Publishers, 5–14.

Duch, W. (1996). "Categorization, prototype theory and neural dynamics". In: Yamakawa, T.Y. and. Matsumoto, G. (eds), *Methodologies for the Conception, Design, and Application of Intelligent Systems*, *Proceedings of the 4th International Conference on Soft Computing*, Iizuka: Singapore/River Edge, N.J., 482–485.

Duch, W. (1997). "Platonic model of mind as an approximation to neurodynamics". In: Amari, S.I. and Kasabov, N. (eds), *Brain-Like Computing and Intelligent Information Systems*. Singapour: Springer, 491–512.

Duch, W. (2005). "Brain-inspired conscious computing architecture". *Journal of Mind and Behavior*, 26, 1–22.

Duch, W. (2012). "Mind-brain relations. Geometric perspective and neurophenomenology", *American Philosophical Association Newsletter*, 12(1), 1–7.

Duch, W. (2018). "Kurt Lewin, psychological constructs and sources of brain cognitive activity". *Polish Psychological Forum*, 23(1), 5–19.

Duch, W. (2021). "Memetics and neural models of conspiracy theories". *Patterns*, 2(11), 100353.

Duch, W., Matykiewicz, P. and Pestian, J. (2008). "Neurolinguistic approach to natural language processing with applications to medical text analysis". *Neural Networks*, 21(10), 1500–1510.

Fauconnier, G. (1994). *Mental Spaces: Aspects of Meaning Construction in Natural Language*. Cambridge: Cambridge University Press.

Fauconnier, G. and Turner, M. (2003). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities* (Reprint edition). New York: Basic Books.

Fernandino, L., Tong, J.-Q., Conant, L. L., Humphries, C. J. and Binder, J. R. (2022). "Decoding the information structure underlying the neural representation of concepts". *Proceedings of the National Academy of Sciences*, 119(6). https://doi.org.10.1073/pnas.2108091119

Gärdenfors, P. (2004). *Conceptual Spaces: The Geometry of Thought*. Cambridge Mass: The MIT Press.

Gärdenfors, P. (2014). *The Geometry of Meaning: Semantics Based on Conceptual Spaces*. Cambridge Mass.: The MIT Press.

Goertzel. B. (2006). *The Hidden Pattern*. Boca Raton, Fl.: BrownWalker.

Harnad, S. (1990). "The symbol grounding problem". *Physica D: Nonlinear Phenomena*, 42(1), 335–346.

Heusser, A. C., Fitzpatrick, P. C. and Manning, J. R. (2021). "Geometric models reveal behavioural and neural signatures of transforming experiences into memories". *Nature Human Behaviour*, 5(7), 905–919.

Hoel, E. P. (2017). "When the map is better than the territory". *Entropy*, 19(5), 188.

Hurlburt, R.T. and Schwitzgebel, E (2007). *Describing Inner Experience? Proponent Meets Skeptic*. Cambridge Mass.: The MIT Press.

Hutchins, E. (2012). "Concepts in practice as sources of order". *Mind, Culture, and Activity*, 19(3), 314–323. https://doi.org/10.1080/10749039.2012.694006

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E. and Gallant, J. L. (2016). "Natural speech reveals the semantic maps that tile human cerebral cortex". *Nature*, 532(7600), 453–458.

Hyungsuk, J., Ploux, S. and Wehrli, E. (2003). Lexical Knowledge Representation with Contexonyms. 9th MT Summit Machine Translation, New Orleans, September 2003, 194–201. https://hal.archives-ouvertes.fr/hal-00933207

Johnson-Laird, P.N. (1983). *Mental models: Towards a Cognitive Science of Language, Inference and Consciousness*. Harvard: Harvard University Press.

Johnson-Laird, P.N. (1995). *Mental Models, Deductive Reasoning, and the Brain*. Cambridge Mass: The MIT Press: 999–1008.

Kelly, G. (1955). *The Psychology of Personal Constructs.* New York: Norton.

Komorowski, M. K., Rykaczewski, K., Piotrowski, T., Jurewicz, K., Wojciechowski, J., Keitel, A., Dreszer, J. and Duch, W. (2021). "ToFFi-Toolbox for frequency-based fingerprinting of brain signals". *Neurocomputing* (in revision), and ArXiv:2110.09919

Landauer, T. and Dumais, S. A. (1997). "Solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge". *Psychological Review*, 104(2), 211–240.

Laird, J. E., Lebiere, C. and Rosenbloom, P. S. (2017). "A standard model of the mind: Toward a common computational framework across artificial intelligence. Cognitive science, neuroscience, and robotics". *AI Magazine*, 38(4), 13–26.

Lewin, K. (1936). *Principles of Topological Psychology*. New York: McGraw-Hill. https://pl.scribd.com/book/262688082/Principles-of-Topological-Psychology

Lewin, K. (1938). *The conceptual representation and the measurement of psychological forces.* Durham N.C.: Duke University Press.

Neimeyer, R.A. and Neimeyer, G.J. (eds.), (2002). *Advances in Personal Construct Psychology.* New York: Praeger.

Newell, A. and Simon, H.A. (1976). "Computer science as empirical inquiry: symbols and search". *Communications of the ACM*, 19(6), 113–126.

O'Regan, J. K. (2011). *Why Red Doesn't Sound Like a Bell: Understanding the Feel of Consciousness.* Oxford: Oxford University Press.

O'Reilly, R. C., Munakata, Y., Frank, M. J., Hazy, T. E., and Contributors (2020). *Computational Cognitive Neuroscience.* Wiki Book, 4th edition. https://CompCogNeuro.org

Ploux, S. and Ji, H. (2003). "A model for matching semantic maps between languages (French/English, English/French)". *Computational Linguistics*, 29(2), 155–178.

Ploux, S., Boussidan, A. and Ji, H. (2010). The Semantic Atlas: An Interactive Model of Lexical Representation. *Proceedings of the Seventh Conference of International Language Resources Ans Evaluation*: 1–5. https://hal.archives-ouvertes.fr/hal-00933294

Rykaczewski, K., Nikadon, J., Duch, W. and Piotrowski, T. (2021). "supFunSim: Spatial filtering toolbox for EEG". *Neuroinformatics*, 19(1), 107–125.

Speer, N. K., Reynolds, J. R., Swallow, K. M. and Zacks, J. M. (2009). "Reading stories activates neural representations of visual and motor experiences". *Psychological Science*, 20(8), 989–999. https://doi.org/10.1111/j.1467-9280.2009.02397.x

Spivey, M.J. (2007). *The Continuity of Mind.* New York: Oxford University Press.

Tian, F., Wang, H., Cheng, W., Zhang, W. and Li, Y. (2021). "A high-density EEG study investigating VR Film editing and cognitive event segmentation theory". *Sensors*, 21(21), 7176. https://doi.org/10.3390/s21217176

Varley, T., and Hoel, E. (2021). "Emergence as the conversion of information: A unifying theory". ArXiv:2104.13368 [Cs, Math]. http://arxiv.org/abs/2104.13368

Varley, T. F., and Sporns, O. (2022). "Network nalysis of time series: novel approaches to network neuroscience". *Frontiers in Neuroscience*, 15. https://www.frontiersin.org/article/10.3389/fnins.2021.787068

Wierzbicka, A. (1996). *Semantics: Primes and Universals.* Oxford: Oxford University Press.

Zacks, J. M., Speer, N., Swallow, K. and Maley, C. (2010). "The brain's cutting-room floor: segmentation of narrative cinema". *Frontiers in Human Neuroscience*, 4. https://www.frontiersin.org/article/10.3389/fnhum.2010.00168

Zhang, Y., Han, K., Worth, R. and Liu, Z. (2020). "Connecting concepts in the brain by mapping cortical representations of semantic relations". *Nature Communications*, 11(1), 1877.