

DOI: 10.20098/j.cnki.1001-5132.2024.1112

Global workspace theory-based cognitive architecture for robots

HUANG Wenjie^{1,2}, CANGELOSI Angelo^{1,2,4*}, CHELLA Antonio^{3,4}

(1.The University of Manchester, Manchester, UK; 2.Manchester Centre for Robotics and AI, Manchester, UK;
3.University of Palermo, Italy; 4.ICAR-CNR, Palermo, Italy)

Abstract: This paper adopts the Global Workspace Theory as a neuro-scientifically plausible theory for developing conscious cognitive architecture. The Global Workspace Theory's compatibility with the working mechanisms underneath human brains is enhanced by the implementation of different cognitive features based on this framework. Amongst the topics in the literature for intelligent systems, we start with attention, memory and learning mechanisms, and corresponding experiments are summarized here. We also discuss how other topics of cognitive robotics could be developed based on these three basic components, and their correlations. This provides a foundation for future long-term development of cognitive architectures of cognitive robots. The research in this paper follows the incremental research pathway for the architecture implementation, which is consistent with the Biologically Inspired Cognitive Architecture roadmap.

Key words: Global Workspace Theory; cognitive architecture; incremental; biologically inspired cognitive architecture

CLC number: TP18

Document code: A

Article ID: 1001-5132 (2025) 02-0049-10

1 Introduction

Artificial intelligence is the ambition of all researchers to achieve human-level intelligence in machines. Taking us humans as the best intelligence (as we understand) on this planet, building machines according to theories inspired by human brains becomes a prevailing and powerful approach^[1-3], which is called the Biologically Inspired Cognitive Architecture (BICA) challenge. In the work of Samsonovich^[4], a roadmap for this challenge is explained. Sharing the same concern and approach, machine consciousness aims to reproduce consciousness-related cognitive activities from human brains. This is motivated by the fact that consciousness

is a key feature of human intelligence.

In this paper, the authors tend to introduce their previous work on GWT-based architecture implementations. Through this, readers are expected to understand the basic ideas behind authors' incremental research. In Section 2, works incorporating attention, episodic memory and continual learning based on the adopted theory are summarized, demonstrating their achievements with GWT-based cognitive architecture. By this, we define how these three mechanisms are implemented in the architecture. After this, we discuss other key features of human cognition and their relations to the three implemented mechanisms in Section 3. The discussion shows that

Received date: 2024-11-09.

JOURNAL OF NINGBO UNIVERSITY (NSEE): <http://journaljg.nbu.edu.cn/>

Foundation item: Supported by the European Union's Horizon Europe research and innovation program (101120727 - PRIMI).

Biography: HUANG Wenjie, research associate, specializing in: machine learning & robotics. Email: wenjie.huang@manchester.ac.uk

***Corresponding author:** CANGELOSI Angelo, professor, specializing in: machine learning & robotics. Email: angelo.cangelosi@manchester.ac.uk

the authors' research focuses are also consistent with the roadmap for BICA in [4]. Finally, the significance of building cognitive architectures based on GWT is re-stressed, and the possible future work to implement more cognitive features based on the developed architecture is discussed.

2 Global workspace theory

The review paper conducted by Reggia^[5] and the chapter on Robots and Machine Consciousness edited by Chella^[6] provide a general picture of the major research questions in machine consciousness. However, no conscious machine has been identified so far in any attempted implementations. Huang et al^[7] summarized three reasons for this, including individual-dependent definitions of consciousness, isolated research focuses and vague validation criteria. As suggested by them, work coping with any of these three problems is appreciated for its contribution to the research community. From the authors' perspective of engineering, breaking the isolation between different research focuses would shed light on achieving human-level intelligence with the whole cognitive system. For this purpose, Global Workspace Theory (GWT)^[8] is adopted as the basis for a series of works. This theory was proposed by [9-10] and a prototype of GWT-based architecture is illustrated in Fig. 1. From this structure, the whole brain consists of different specialist modules running in parallel. These specialists are domain-specific processors such as vision, hearing, motor, language, desire and so on. To coordinate the functioning of the whole brain, there is a global workspace node. To gain access to this node, all specialists forward domain signals and compete with each other based on corresponding salience values. At any time, only one domain-specific signal can enter the workspace node. After that, this signal is broadcast to all other specialist modules, making the whole brain work on this stimulus and demonstrate consciousness. Thus, instead of focusing on certain

mechanisms or functionalities that are correlated to consciousness in human brains, GWT emphasizes that consciousness towards a certain stimulus arises only when the stimulus gains exclusive access to the global workspace node and is broadcast to all specialist modules.

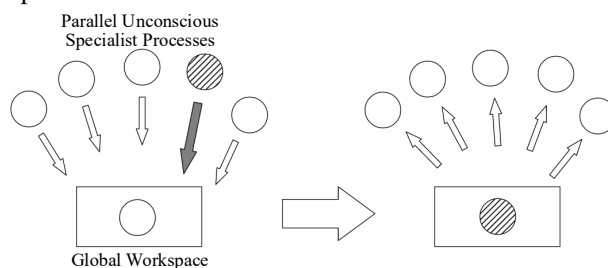


Fig. 1 The prototype of GW model illustrated in [11]

This theory has been studied by many researchers in neuroscience, psychology and computer science. With different motivations, they either validated the plausibility of this theory to human brains^[12-15] or testified to the feasibility of exploiting this theory in computing systems for achieving human-level cognitive agents^[11,16-18]. These studies strongly support the authors' development of cognitive architecture based on the theory. Though the general architecture is defined by GWT, researchers are confronted with narrower challenges in fleshing out the specialist modules. To do this, it is always questioned what modules and corresponding mechanisms should be implemented, as it is not feasible to make a complete list of them in human brains.

3 Cognitive architecture development

In this paper, we present an incremental cognitive architecture for consciousness based on the GW model. This aims to build and integrate a series of modules corresponding to basic functionalities contributing to machine consciousness^[19]. In particular, we first focus on modelling the mechanisms of selective attention, episodic memories, and continual learning within the same GWT framework. In the sub-sections below, three modelling and experimental works on the corresponding topics are summarized and discussed, which are used as building blocks of the incremental cognitive architecture.

3.1 Attention and Consciousness

In [7], the cognitive architecture is adapted from the prototype in Fig. 1 as illustrated in Fig. 2. Though this implementation does not strictly observe the definition of GWT, it focuses on the competing selection of consciousness access. This work illustrates two specialist modules, *Vision* and *Hearing*, competing with each other to gain access to the global workspace. The global workspace node here plays the role of signal cache, storing and broadcasting the attended information to other modules.

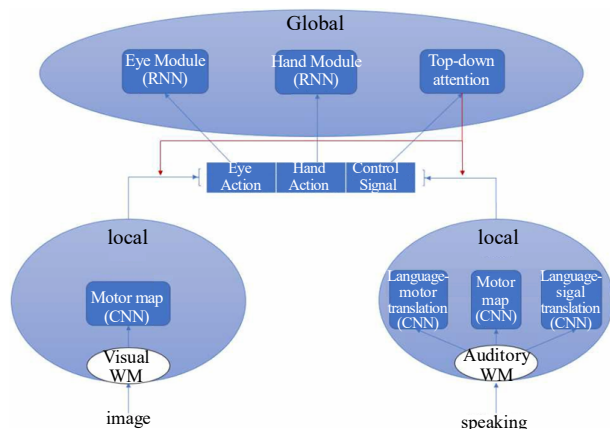


Fig. 2 The prototype of GW model implemented in [7]

The basic idea underneath this implementation is that attention is closely related to consciousness. The relation between these two concepts has been studied by many scientists^[20-22]. In addition, from GWT, the attention mechanism fits the competing selection for conscious activities. For validating the general idea of GWT for conscious cognition, two attention-related phenomena, attentional blink^[23-24] and lag-1 sparing effect^[25], are replicated in the experiments. The attentional phenomena are replicated with both top-down and bottom-up attention streams. This work shares the same concern with the BICA roadmap in [4] on attention, awareness, and intentionality.

3.2 Episodic Memory and Consciousness

With the implemented attention mechanism within the GWT-based cognitive architecture mentioned above, updates of the structure are inevitable due to certain limitations. Based on the previous implementation, two updates are introduced in later work in [26].

Firstly, the whole architecture is reformed to be more consistent with the definition of GWT as illustrated in Fig. 3. The new model retains all the key features of attentional phenomena studied in previous work. The minor adaptation is that specialist modules are categorized into three groups according to their communication ways with the global workspace node as illustrated by the arrow directions. Another update is the incorporation of the episodic memory^[27] module.

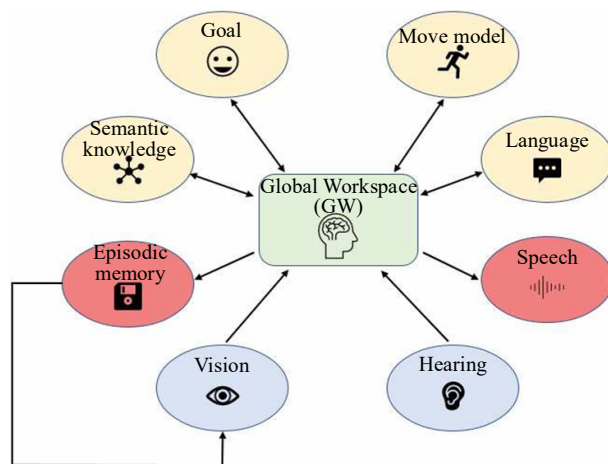


Fig. 3 The GWT-based cognitive architecture updated in [26]

The motivation behind this is that the implementation in [7] can only process the presented information, which is the working style of many artificial models and systems. On the other hand, memory entitles the agent to respond and use past experience. For this purpose, episodic memory is implemented in the updated architecture. By incorporating episodic memory, the agent is capable of referring to past experiences in memory and fulfilling inference tasks on those recalled memory traces with pre-trained semantic knowledge^[26]. Same as the previous work, this implementation also falls into the BICA roadmap in [4] on episodic memory.

The implemented agent shows ability to form memory into both long-term and short-term based on the strength of consciousness towards certain information. During the interactions between the agent and the environment, the goal motivation and the recalling effect of memory play the role of conscious

enhancement to certain pieces of memory.

3.3 Continual Learning and Consciousness

This part focuses on the learning ability, which is a basic component of conscious agents to develop intelligence. In this implementation, which is again based on the GWT framework, episodic memory and semantic knowledge are not implemented as individual modules anymore. The idea is that both of them could be regarded as certain transformations of information between different modules. For instance, the memory of the visual experience is the transformation from a speech cue to the corresponding experience episode. Also, knowledge of the name of a visual object transforms visual perception into language representation. Though they are both integrated into the transformative processes between different modules, the distinctions between these two cognitive mechanisms are retained. Memory is only about the recall of experience, while knowledge is a generalized usage of experience. This transformative definition is inspired by the work of Semantic Pointer Architecture (SPA)^[28-29]. The key idea of SPA is the transformation between different modalities and levels of cognition, which is illustrated in Fig. 4. The information transformation in the architecture hinges on the conceptual semantic pointer (SP) which is a low-dimension representation. This SP is connected to other SP in various modalities. Through the SP in perceptual modules, the conceptual SP is associated with high-dimension representations from the perceptual modules as well. For instance, In Fig. 4, the concept DOG could be transformed into other concepts in semantics and sensorimotor representations. On the contrary, sensorimotor representations can be transformed into semantic representations. Thus, three kinds of transformation could be summarized in this architecture, namely compression, decompression and within-semantics. Amongst them, for our interest, the compression and decompression transformations are inspiring as they are quite compatible with our definitions of memory and knowledge, though with differences in implementation from SPA.

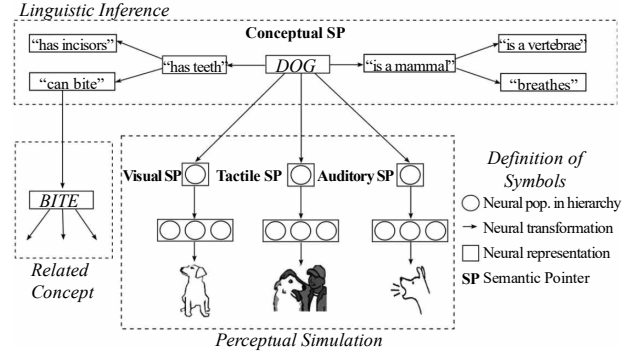


Fig. 4 Transformation diagram of concept DOG in Semantic Point Architecture^[28]

Inspired by SPA, modules in this model communicate with others via transformations via the Global Workspace. In this case, memory and knowledge are implemented as transformations between different modules, rather than individual modules. In Fig. 5, we focus on the interactions between modules vision, hearing, motor and speech with the coordination of Global Workspace. In this model, all modules are in bi-direction connected with the global workspace node. By this, it emphasizes that the information transformation between modules is bi-direction. The implementations of vision and hearing have been enriched compared to our previous model in the subsection. A sophisticated visual attention mechanism is incorporated.

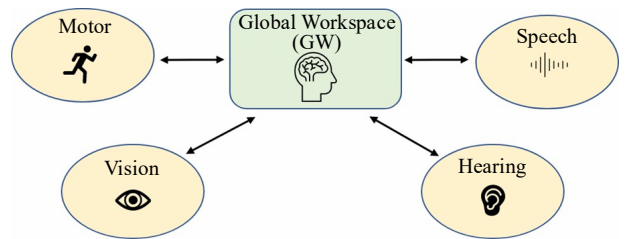


Fig. 5 The structure of the model

In addition, in the hearing module, the continual learning scheme is integrated. By implementing the continual learning scheme, the agent is expected to react to not only the present stimuli but also the stimuli in past. Moreover, continual learning approaches aim to reduce the memory burden but retain the learning performance of the AI system^[30-32]. During learning, the agent associates perceived visual information with the language inputs, while

dynamically maintaining its memory based on the importance of knowledge.

In summary, motor and speech modules enable the robotic agent to engage in physical interactions with human peers. The vision part is implemented as a receptor and pre-processor of visual stimuli, without a learning process. The focus of our new model is on learning associations from visual stimuli to auditory ones. This is not only achieved by the continual learning scheme of memory replay but also by the attention mechanisms implemented before. The attention mechanism within this Global Workspace framework enables the agent also to automatically filter the information to be learnt, demonstrating the capability of autonomous learning^[33] in interaction at a beginning stage of intelligence. This work suggests the reciprocal relations between memory and learning. Also, the association learning in this GWT framework takes the conscious perception as a pre-requisite.

4 Discussion

In the first experimental study revisited in the last section, the focus is on the GWT architecture at the minimal level. The attention mechanism is employed to support the competing selection process for certain information to enter consciousness. Thus, it contributes basically to grounding the research basis of GWT for the later development of richer cognitive architectures. The second study on episodic memory is initially motivated by the limitation of the previous model concerning the width of cognitive behaviors. The study by [34] also indicates the relation between episodic memory and consciousness from a psychological perspective. Following this, the learning capability in a continuous manner is another basic component of cognitive agents to expand experience and knowledge. To summarize these works together, they indicate that the cognitive architecture based on consciousness should span along both time and modality dimensions. In other words, a conscious cognitive agent is expected to process

information of different modalities from past, current and even future time.

However, the updated implementation of the GWT-based architecture is still far from human-like intelligence. All those mechanisms implemented so far have significant space to improve. For the attention mechanism, sophisticated top-down and bottom-up attention streams are required for rich cognition. This depends on different modules to modulate the unconscious as well as conscious behaviors of the agent. As for the memory part, the debate of semantic and episodic memory is always there. Though it is open to implementations, experiments with more complicated data and the interactive environment for agents should be studied to enhance the feasibility of the transformative implementation presented in this paper. In addition, the authors suggest a wider scope for continual learning in this paper at the agent-system level. It is proposed here two continual learning schemes should be both emphasized. One is the continual learning impacting the functionalities of only specialist modules, which is called the accumulation of domain knowledge. This is partially discussed in Section 2.3. The other is continual learning of consciousness interactions between different modules. Babies are learning what reactions are expected from given stimuli from the caregivers, thus adjusting their responding strategies in different cognitive scenarios. For instance, when a conversation is initialized by the human partner with the robotic agent, an attention enhancement is supposed to be imposed on the hearing module of the agent. Also, this conscious enhancement could be assigned to different modules in various contexts. A good example is the eye gaze understanding emerging in babies' cognition during 8–12 months^[35-36]. To interpret this with the architecture (e.g., Fig. 3), a top-down attention mechanism is important to coordinate the conscious activities between different specialist modules, which has been explained in [7]. However, it is impossible to manually design a fixed strategy to decide which module should be enhanced and/or suppressed in

certain cognitive scenarios when the cognitive architecture becomes increasingly complicated. Thus, it requires the agent to learn and adapt its top-down-enhancement and suppression strategies through continuous experiences. From the view of engineering, this aspect of continual learning could also be implemented within certain specialist modules to be domain-specific knowledge accumulation. However, the authors suggest its distinction in its impact on the coordination of the whole architecture. On the other hand, these two learning schemes should jointly contribute to the general knowledge and learner critical mass proposed in the BICA roadmap in [4].

This discussion of the implementations indicates two principles for a pathway the cognitive architecture would evolve for achieving the ultimate goal. Firstly, the authors emphasize the substrates for high-level intelligence to emerge more than the comprehensive and complicated cognitive functionalities. It is believed that implementing the primitive mechanisms for high-level cognition in the architecture would eventually benefit the growth of domain-specific functionalities. Secondly, the implementation is going through a revision-update cycle because of the difficulty in defining all necessary modules and mechanisms for a human-like cognitive agent. The later step in the incremental implementation process is motivated by the limitations of the previous model. This makes the work practical when there is always a clear sub-goal to achieve and also feasible for the implementation and validation of sub-tasks.

5 Future Work

5.1 Cognitive Modules

Future work is expected to extend the system with more modules accounting for different functionalities of human cognition. Apart from cognitive modules implemented in the presented works, other sensorimotor and high-level modules are also key to high-level intelligence.

One sensorimotor module that could be further

implemented is tactile perception^[37]. Apart from visual and hearing perceptions, tactile perception is another modality important for information understanding, which includes the detection of temperature, texture, shape, force, friction, pain, and other related physical properties^[38-39]. From another perspective, one application of tactile sensors in the robotic system is used for preventing dangerous injuries during physical human-robot interaction^[40]. Beyond the understanding of stimuli, this use of tactile perception further contributes to interactive intentions in that the robotic agent can use this information to adjust actions towards different stimuli. In the physiology of mankind, tactile is the sensory effect regarded as one kind of homeostasis, which is defined as the state of steady internal, physical, chemical, and social conditions maintained by living systems^[41]. Thus, extension to other aspects of homeostasis, such as internal physical and chemical conditions is also within this direction. In human physiology, emotions like anger and happiness, feelings of thirst and hunger are all closely related to internal conditions of the body^[42]. Though robotic agents do not have the corresponding physiological system, those features and mechanisms can still be implemented as cognitive modules in robotic agents to facilitate behaviors.

Also, the implementation of these would contribute to other high-level modules, for instance, the goal module illustrated in the architecture. For instance, the avoidance of injuries^[40], the positive feedback from human peers^[43] and so on could all be implemented based on the simulation of the human physiological system. Stimuli associated with positive emotions, feelings or external responses would be considered as strong candidates for the goal of the agent. In addition, more high-level modules for cognition are expected for a robotic agent. Specific modules to be implemented in future work require broader and deeper research.

5.2 Attention Mechanisms

Though attention mechanisms are well defined by reviewing relevant literature^[44], the presented

implementations in this work could be improved. This could be done from two aspects. Firstly, the bottom-up stream of attention is modality-specific. The implemented works show much effort in the design of visual attention. However, the presented design based on expertise rules is way not sufficient according to studies on visual attention. As studies suggest, visual attention is subject to many factors such as color, illumination, size, novelty, curvature, optic flow, intersection, angular coverage, contiguity, etc^[45-46]. Though implementing more factors into the architecture can improve the performance of the agent, it is highly unlikely to include a sufficient set of them in this expertise-based way. An alternative could be the connectionist approach. Transformer, an artificial neural network model, learns to prioritize important information in the given input^[47]. Though it lacks explainability compared to the expertise approach, neural network implementation of the bottom-up attention mechanism could be taken as an integrated effect of countless factors. Another advantage of the neural network model compared to the expertise design that requires investigation into modality-specific factors is that the developer does not need to research bottom-up attention factors for different modalities. Because of these, the neural network implementation of the bottom-up attention is worth research. Secondly, top-down attention, different from the bottom-up stream that is purely externally driven factors, is based on the combined effect of prior knowledge, plans, goals, emotions and so on^[48-51]. The incremental implementation of more modules within the architecture discussed in the future work of cognitive modules serves as a pre-requisite for this. They would function as the modulating factors of the top-down attention stream. However, different from the bottom-up factors that could be represented by a single neural network model, the modules such as emotion, goal, knowledge and so on are separate from each other in the modular design of the architecture. Because of this, the weights of different top-down factors and how they are integrated are the problems to be resolved. For this, reinforcement

learning^[52] might be a solution to model the evolution process of the joint attention. Thus, this is also one of the future research points.

5.3 Episodic and Semantic Memory

To investigate the neural network implementation of contextual and temporal memory is key to validating the definition of episodic memory as information transformation. Norman et al^[53] and Schapiro et al^[54] studied the feasibility of a hippocampal model for episodic memory. One feature of the implemented hippocampal model is the retrieval of memory with partial cues by recurrent connections. Wang^[55] also proposes a hierarchical neural network architecture for episodic memory. This model applies also to episodic memory retrieval with partial cues and accounts for the temporal features of the memory. However, the distinction between these works and the defined episodic memory is that they take the episodes of experience as a fusion of various information. In contrast, the episodic memory defined in Section 3.3 emphasizes that episodic memory is the re-activation of information in specific modalities, based on the information transformation process between modalities.

The semantic knowledge mechanism defined in the proposed architecture is generalized from the episodic memory. For the implemented agent to accumulate knowledge through interactive experience, it depends on the implementation of the continual learning scheme. Because of this, memory limitation is a common obstacle in AI systems. For this, the selective policy for memory replay is one of the key points^[56-57]. Although the presented work adopted this generic idea, the criterion based on whether the sample is learnt knowledge is naive. With more modules implemented as discussed for cognitive modules, the selective policy should be refined. Another possible solution is the generative model used for pseudo-reply^[58]. Future work should investigate the combined solution of selective replay and generative reply of memory for learning. In addition, the continual learning ability of human cognition relies not only on memory replay but also on

other neuronal characteristics like plasticity and modularity^[59-60]. Thus, future work should take into account different mechanisms for continual learning.

6 Conclusions

In this paper, the authors revisit GWT with their implementations based on this theory. From this, the plausibility of GWT-based cognitive architecture is further enhanced. After this, the limitations of the latest implementation are discussed. It is stressed that the implemented cognitive architecture should be able to continuously update both domain-specific knowledge and consciousness interaction strategy between different modules. It is worth emphasizing that these implementations are all consistent with the BICA roadmap^[4]. With the discussion of implementation, the authors also summarize two principles of their work in developing large-scale cognitive architectures. They are substrate priority and revision-update cycle. These two principles, to some extent, guarantee the feasibility of implementations. Several directions for future work are also elaborated, including richer cognitive modules, advanced attention mechanisms, and improving memory systems for higher-level intelligence.

Acknowledgement

This work was funded by the European Union's Horizon Europe research and innovation program under Grant Agreement No. 101120727 - PRIMI.

References:

- [1] SAMSONOVICH A V. Toward a unified catalog of implemented cognitive architectures[M]//SAMSONOVICH A V, JÓHANNSDÓTTIR K R, CHELLA A, et al. *Biologically Inspired Cognitive Architectures* 2010. Amsterdam: IOS Press, 2010, 221:195-244.
- [2] HASSABIS D, KUMARAN D, SUMMERFIELD C, et al. Neuroscience-inspired artificial intelligence[J]. *Neuron*, 2017, 95(2):245-258.
- [3] CANGELOSI A, SCHLESINGER M. From babies to robots: the contribution of developmental robotics to developmental psychology[J]. *Child Development Perspectives*, 2018, 12(3):183-188.
- [4] SAMSONOVICH A V. On a roadmap for the BICA challenge[J]. *Biologically Inspired Cognitive Architectures*, 2012, 1:100-107.
- [5] REGGIA J A. The rise of machine consciousness: studying consciousness with computational models[J]. *Neural Networks*, 2013, 44:112-131.
- [6] CHELLA A. Robots and machine consciousness[M]//CANGELOSI A, ASADA M. *Cognitive robotics*. Cambridge MA: MIT Press, 2022.
- [7] HUANG W J, CHELLA A, CANGELOSI A. A design of global workspace model with attention: simulations of attentional blink and lag-1 sparing[J]. *Journal of Artificial Intelligence and Consciousness*, 2022, 9(1):29-57.
- [8] BAARS B J, GELD N, KOZMA R. Global workspace theory (GWT) and prefrontal cortex: recent developments [J]. *Frontiers in Psychology*, 2021, 12:749868.
- [9] BAARS B J. *A cognitive theory of consciousness*[M]. Cambridge, UK: Cambridge University Press, 1993.
- [10] BAARS B J. *In the theater of consciousness: the workspace of the mind*[M]. New York: Oxford University Press, 1997.
- [11] SHANAHAN M. A cognitive architecture that combines internal simulation with a global workspace[J]. *Consciousness and Cognition*, 2006, 15(2):433-449.
- [12] MASHOUR G A, ROELFSEMA P, CHANGEUX J P, et al. Conscious processing and the global neuronal workspace hypothesis[J]. *Neuron*, 2020, 105(5):776-798.
- [13] BAARS B J, FRANKLIN S, RAMSOY T Z. Global workspace dynamics: cortical "binding and propagation" enables conscious contents[J]. *Frontiers in Psychology*, 2013, 4:200.
- [14] BAARS B J. The global workspace theory of consciousness: predictions and results[M]//SCHNEIDER S, VELMANS M. *The blackwell companion to consciousness*, 2ed. Hoboken, NJ: Wiley, 2017:227-242.
- [15] DECO G, VIDAURRE D, KRINGELBACH M L. Revisiting the global workspace orchestrating the hierarchical organization of the human brain[J]. *Nature Human Behaviour*, 2021, 5:497-511.
- [16] FRANKLIN S, MADL T, D'MELLO S, et al. LIDA: a systems-level architecture for cognition, emotion, and learning[J]. *IEEE Transactions on Autonomous Mental Development*, 2014, 6(1):19-41.
- [17] WIGGINS G A, FORTH J. IDyOT: a computational theory of creativity as everyday reasoning from learned information[M]//Besold T R, Schorlemmer M, Smaill A.

- Computational creativity research: towards creative machines. Paris: Atlantis Press, 2014:127-148.
- [18] FRANKLIN S, MADL T, STRAIN S, et al. A LIDA cognitive model tutorial[J]. *Biologically Inspired Cognitive Architectures*, 2016, 16:105-130.
- [19] CHELLA A, CANGELOSI A, METTA G, et al. Editorial: consciousness in humanoid robots[J]. *Frontiers in Robotics and AI*, 2019, 6:17.
- [20] HAIKONEN P. *The cognitive approach to conscious machines*[M]. Exeter, UK: Imprint Academic, 2003.
- [21] TINSLEY C J. Using topographic networks to build a representation of consciousness[J]. *BioSystems*, 2008, 92(1):29-41.
- [22] STARZYK J A, PRASAD D K. A computational model of machine consciousness[J]. *International Journal of Machine Consciousness*, 2011, 3(2):255-281.
- [23] RAYMOND J E, SHAPIRO K L, ARNELL K M. Temporary suppression of visual processing in an RSVP task: an attentional blink?[J]. *Journal of Experimental Psychology Human Perception and Performance*, 1992, 18(3):849-860.
- [24] DEHAENE S, SERGENT C, CHANGEUX J P. A neuronal network model linking subjective reports and objective physiological data during conscious perception [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2003, 100(14):8520-8525.
- [25] HOMMEL B, AKYÜREK E G. Lag-1 sparing in the attentional blink: benefits and costs of integrating two events into a single episode[J]. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 2005, 58(8):1415-1433.
- [26] HUANG W J, CHELLA A, CANGELOSI A. A cognitive robotics implementation of global workspace theory for episodic memory interaction with consciousness[J]. *IEEE Transactions on Cognitive and Developmental Systems*, 2024, 16(1):266-283.
- [27] TULVING E. Episodic memory: from mind to brain[J]. *Annual Review of Psychology*, 2002, 53:1-25.
- [28] BLOUW P, SOLODKIN E, THAGARD P, et al. Concepts as semantic pointers: a framework and computational model[J]. *Cognitive Science*, 2016, 40(5):1128-1162.
- [29] ELIASMITH C. *How to build a brain: a neural architecture for biological cognition*[M]. Oxford, UK: Oxford University Press, 2013.
- [30] VAN DE VEN G M, TOLIAS A S. Three scenarios for continual learning[EB/OL]. [2024-03-21]. <http://arxiv.org/abs/1904.07734>.
- [31] HADSELL R, RAO D, RUSU A A, et al. Embracing change: continual learning in deep neural networks[J]. *Trends in Cognitive Sciences*, 2020, 24(12):1028-1040.
- [32] De LANGE M, ALJUNDI R, MASANA M, et al. A continual learning survey: defying forgetting in classification tasks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(7):3366-3385.
- [33] ANGELOV P. *Autonomous learning systems: from data streams to knowledge in real-time*[M]. Hoboken, NJ: John Wiley & Sons, 2012.
- [34] TULVING E. What is episodic memory?[J]. *Current Directions in Psychological Science*, 1993, 2(3):67-70.
- [35] WOODWARD A L. Infants' developing understanding of the link between looker and object[J]. *Developmental Science*, 2003, 6(3):297-311.
- [36] CSIBRA G, VOLEIN Á. Infants can infer the presence of hidden objects from referential gaze information[J]. *British Journal of Developmental Psychology*, 2008, 26(1):1-11.
- [37] BREMNER A J, SPENCE C. The development of tactile perception[J]. *Advances in Child Development and Behavior*, 2017, 52:227-268.
- [38] BUTTAZZO G, BICCHI A, DARIO P. *Robot tactile perception*[M]//Lee C S G. *Sensor-based robots: algorithms and architectures*. Berlin: Springer, 1991:25-40.
- [39] DARGAHI J, PAYANDEH S. Surface texture measurement by combining signals from two sensing elements of a piezoelectric tactile sensor[C]//*Proceedings of SPIE 3376, Sensor fusion: architectures, algorithms, and applications II*. 1998:122-128.
- [40] FRITZSCHE M, ELKMANN N, SCHULENBURG E. Tactile sensing: a key technology for safe physical human robot interaction[C]//*Proceedings of the 6th International Conference on Human-robot Interaction*. 2011:139-140.
- [41] DELMONTE D W, KIM T. Anatomy and physiology of the cornea[J]. *Journal of Cataract and Refractive Surgery*, 2011, 37(3):588-598.
- [42] WIRTH M M, GAFFEY A E. Hormones and emotion: stress and beyond[M]//ROBINSON M D, WATKINS E R, HARMON-JONES E. *Handbook of cognition and emotion*. New York: Guilford Publications, 2013:69-94.
- [43] MOREIRA I, RIVAS J, CRUZ F, et al. Deep reinforcement learning with interactive feedback in a human-robot environment[J]. *Applied Sciences*, 2020, 10(16):5574.
- [44] KNUDSEN E I. Fundamental components of attention[J]. *Annual Review of Neuroscience*, 2007, 30:57-78.

- [45] RASHID U, NACENTA M A, QUIGLEY A. Factors influencing visual attention switch in multi-display user interfaces: a survey[C]//Proceedings of the 2012 International Symposium on Pervasive Displays. 2012:1-6.
- [46] WOLFE J M, HOROWITZ T S. Five factors that guide attention in visual search[J]. *Nature Human Behaviour*, 2017, 1:58.
- [47] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st Conference on Neural Information Processing Systems, 2017:6000-6010.
- [48] BREAZEAL C, SCASSELLATI B. A context-dependent attention system for a social robot[J]. *International Joint Conference on Artificial Intelligence*, 1999, 255(3):1146-1151.
- [49] KATSUKI F, CONSTANTINIDIS C. Bottom-up and top-down attention: different processes and overlapping neural systems[J]. *The Neuroscientist*, 2014, 20(5):509-521.
- [50] VIJAYAKUMAR S, CONRADT J, SHIBATA T, et al. Overt visual attention for a humanoid robot[C]//Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. 2001:2332-2337.
- [51] TAYLOR J G. CODAM: a neural network model of consciousness[J]. *Neural Networks*, 2007, 20(9):983-992.
- [52] WIERING M A, VAN OTTERLO M. Reinforcement learning[J]. *Adaptation, Learning, and Optimization*, 2012, 12(3):729.
- [53] NORMAN K A, O'REILLY R C. Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach[J]. *Psychological Review*, 2003, 110(4):611-646.
- [54] SCHAPIRO A C, TURK-BROWNE N B, BOTVINICK M M, et al. Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning[J]. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 2017, 372(1711):20160049.
- [55] WANG X J. Neural dynamics and circuit mechanisms of decision-making[J]. *Current Opinion in Neurobiology*, 2012, 22(6):1039-1046.
- [56] YOON J, MADAAN D, YANG E, et al. Online coreset selection for rehearsal-based continual learning[EB/OL]. [2024-05-08]. <http://arxiv.org/abs/2106.01085>.
- [57] ISELE D, COSGUN A. Selective experience replay for lifelong learning[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018, 32(1):11595.
- [58] SHIN H, LEE J K, KIM J, et al. Continual learning with deep generative replay[C]//Proceedings of the 31st Conference on Neural Information Processing Systems, 2017:2994-3003.
- [59] LESORT T, LOMONACO V, STOIAN A, et al. Continual learning for robotics: definition, framework, learning strategies, opportunities and challenges[J]. *Information Fusion*, 2020, 58:52-68.
- [60] De LANGE M, ALJUNDI R, MASANA M, et al. A continual learning survey: defying forgetting in classification tasks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(7):3366-3385.

基于全局工作空间理论的机器人认知架构

HUANG Wenjie^{1,2}, CANGELOSI Angelo^{1,2,4}, CHELLA Antonio^{3,4}

(1.曼彻斯特大学, 英国 曼彻斯特; 2.曼彻斯特机器人与人工智能中心, 英国 曼彻斯特;

3.巴勒莫大学, 意大利 巴勒莫; 4.意大利国家研究委员会高性能计算与网络研究所, 意大利 巴勒莫)

摘要: 采用适用于神经科学的全局工作空间理论开发有意识的认知架构. 基于这一架构实现了不同的认知要素, 使得全局工作空间理论与人脑工作机制的兼容性得以增强. 从注意力、记忆和学习机制这些针对智能系统的研究主题开始进行探讨, 并对相关实验进行总结. 随后讨论了如何基于这三个基本要素及其相互关系研究认知机器人开发这一新主题. 这些工作为认知机器人认知架构的未来长期发展提供了基础. 此外, 采用渐进式研究路线用以上述架构的实现, 这与生物启发式认知架构的研究路线图相一致.

关键词: 全局工作空间理论; 认知架构; 渐进; 生物启发式认知架构

(责任编辑 韩 超)