An Implementation of Global Workspace Theory: Consciousness Interfered Episodic Memory for Cognitive Robots

Wenjie Huang, *PhD Student, the University of Manchester, UK*
Antonio Chella, *Professor of Robotics, University of Palermo, Italy*
Angelo Cangelosi, *Professor of Machine Learning and Robotics, the University of Manchester, UK*

***Abstract*—Artificial general intelligence revived in recent years after people achieved significant advances in machine learning and deep learning. This leads to the thinking of how real intelligence could be created. Consciousness theories believe that general intelligence is essentially conscious, yet no universal definition is agreed on the definition. In this work, a systematic theory of consciousness, Global Workspace Theory, is implemented and integrated with crucial cognitive components. The focus of this paper is episodic memory. With inspiration from the natures of episodic memory in psychology and neuroscience, the episodic memory component is implemented in the Global Workspace framework. The robotic agent operates in a real-world interactive context, forming episodic memory and demonstrating static, temporal and context memory capabilities during interactions. Consciousness in this work engages in all formation, maintenance, and retrieval processes of episodic memory. The novelties and contributions of this work are 1) this work is implementing episodic memory within the consciousness framework, suggesting the sustainable potential of such an integrated approach to cognitive agents with AGI; 2) Regarding the limited examples in consciousness-based cognitive architectures, this work attempts to contribute to the diversity of perspectives and approaches; 3) Extant episodic memory implementations are suffering from various limitations, while This work summarises some key features for modelling episodic memory within a cognitive architecture; 4) Authors discuss the relationship between episodic memory, consciousness and general intelligence, proposing the compatibility and relationship between machine consciousness and other AGI research. It is believed that a better alignment between them would further boost the fusion of diverse research for achieving the desired cognitive machines.**

***Index Terms*—General intelligence, Cognitive Robots, Consciousness, Global Workspace, Episodic memory**

I. Introduction

ARTIFICIAL general intelligence(AGI) research focuses less on the specific functions or algorithms like its variation in the current mainstream Artificial Intelligence community. Instead, the general cognitive capabilities of human beings are under the spotlight in this area. Many people have made attempts in cognitive theories and models for achieving AGI like [1] [2] [3] [4]. However, the outcomes of extant attempts are not yet applicable in the real world or not convincing to have reached AGI. Among various subtopics under this big title, this work is standing on the Global Workspace Theory [5], which was inspired by psychological and neuron-scientific research and proposed for consciousness. The initial purpose of exploiting this theory in intelligence systems is to create conscious machines by understanding and reproducing the consciousness correlates such as functions and mechanisms. This crazy idea for ordinary people should not shock the scientific community as it is not a research minority to build computing systems inspired by humans and animals [6] [7].

This work holds the belief that consciousness is the essence of general intelligence. However, this work does not take consciousness as a specific mechanism or function like attention, reasoning or others which are also crucial for intelligence. Instead, we attempt to emphasise consciousness as a systematic principle with which the whole system should observe during functioning. In other words, consciousness does not give rise to general intelligence directly, but all advanced cognitive modules of the human brain or the future cognitive robots operating and collaborating with each other within a consciousness framework would give birth to AGI. In this blueprint, the GWT plays as the cornerstone.

Another involved topic in this work is cognitive robotics, which nowadays is a popular way to study natural intelligence and build artificial systems [8] [9]. With the increasing interest in cognitive agents with consciousness, research on GWT has expanded to form a community. The work on this theory so far can be divided into three categories. The first one is theoretical studies based on subjects including psychology and neuroscience. This theory was firstly proposed by [5] [10], which was basically psychological work for conscious cognition. Later in recent years, this theory is further extended and validated by findings in neuroscience due to the advances in studies in brain dynamics.The main body of this kind is expressed by researchers for example in [11] [12] [13] [14]. In these studies, the role of Global Workspace is enhanced by the cortico-thalamic (C-T) evidence [15]. From a conclusive phrase in [16], the standing of this work is also partially consistent with that of GWT, viewing consciousness as the product of highly integrated and widespread cortico-thalamic (C-T) activity. Because of the consistency of the GWT in psychology and brain science, it becomes a guiding theory for cognitive computing, which is the second group of work on GWT. In this group, researchers do the implementation of the Global Workspace framework first but pay little attention to the complexities and capabilities of specialists. Some example attempts of this were made by [15] [17] [18]. In these works, GWT was exploited to simulate certain cognitive effects related to human consciousness. Works of this group have at least one profound implication. It is a validation of the potential of GWT in computing systems. For instance, [17] combines GWT and internal model theory, enhancing the agent's cognitive responses to the changing situation. The work of [18] integrates the attention mechanism, reproducing two consciousness-related effects of human cognition. The mentioned work for this group could be basically viewed as the study of GWT in the computing area. Through this group work, the viability of the theory of computing reproductions is strengthened, providing later researchers with prototypes. The third category of work is also from the perspective of computing systems, but, instead of taking GWT as the study subject, they are proposing cognitive architectures beyond the scope of GWT. Two significant examples in this group are LIDA [19] and IDyOT [20], in which the modules are organised by in the formula of GWT to perform in a cognitive way. By these, they put forward the work of cognitive architectures, providing concrete models for others. Based on the LIDA architecture, some implemented agents are summarised in [21]. These implementations to LIDA architecture repeat the common goal for the discussed second group work to GWT. Apart from the three groups discussed above, recent research also tries to combine GWT and deep learning. In [22], the authors proposed a roadmap to implement GWT with various deep learning-based modules. This paper does not categorise this kind of work due to the little attempt based on it. However, it does not mean deep learning techniques are excluded from the cognitive architectures discussed here. In fact, most extant cognitive architectures employ various tools including rule and statistics-based machine learning and deep learning ones. Regarding the advances in deep learning, The proposal of [22] is worth further exploration.

With the explanation of the extant work related to GWT, the authors of this work would like to claim it to be an attempt of consciousness-inspired cognitive systems with a different approach from the discussed. This work is beyond the validation study of GWT but does not propose a comprehensive architecture yet. Though it shares similarities with the third category that is working on cognitive agents, this work does it from a different perspective. Instead of implementing an agent based on a proposed architecture like LIDA [19] and IDyOT [20], this work investigates, incorporates and develops new specialists within the Global Workspace framework in an incremental manner. In this way, apart from taking GWT as a guide of the general framework, our work also examines the mutual influence of specialists and Global Workspace. We explicitly claim that the specialists of the system are functioning under the coordination of Global Workspace, and in return specialists also collectively determine some details of Global Workspace for it to support the information communications across the whole system. On the other hand, people may identify similarities between our work and those two examples. Firstly, the research scope of this work and its future extensions would definitely overlap with most of the example architectures though some may not. At a certain point, secondly, the extensions of this work would arrive at another cognitive architecture based on consciousness theories. Thus, this work would like to claim contributions to the community of cognitive agents with the innovations in a different approach we exploit towards consciously cognitive architecture and diversities of the pursued outcomes based on the similarities of this work to others, both of which would provide more community comparisons and discussions.

As one step in the incremental development of the cognitive agent with consciousness, this work further incorporates episodic memory based on our previously implemented model [18]. In the previous work, the Global Workspace framework was implemented and validated with experiments reproducing two cognitive phenomena, attentional blink and lag-1 sparing effect. However, the limitations with respect to width and depth of consciousness were recognised. Based on the limitations, we are seeking to integrate more components, which are essential for conscious cognition and further AGI, into the implemented framework. Episodic memory as a way of perceiving the past experience [23] [24] fits well with the limitation with respect to the width of consciousness. This is a simple but strong motivation for us to investigate episodic memory within our consciousness-based cognitive model. The purpose is to investigate how the episodic memory module could be implemented within the framework of GWT as well as what requirements episodic memory has on the coordination role of Global Workspace. By implementing episodic memory, the authors believe the agent would demonstrate richer conscious cognition.

Based on the existing work, the novelties and contributions of this work can be recognised as follows. Firstly, by implementing episodic memory and conducting the human-robot experiment based on the GWT consciousness framework, the authors aim at convincing other researchers of the potential of such an integrated approach with consciousness towards AGI. Secondly, the main extant projects of cognitive architectures based on consciousness are reviewed in this work. Regarding the limited examples in consciousness-based cognitive architectures, this work employs a different developmental approach, attempting to contribute to the research diversity of the community. Thirdly, extant episodic memory implementations are suffering from various limitations, while This work summarises some key features for modelling episodic memory within a cognitive architecture by reviewing past work. This provides others with a comprehensive reference on this issue. Last but not least, the authors discuss the relationship between episodic memory, consciousness and general intelligence, proposing the compatibility and relationship between machine consciousness and other AGI research. Through this discussion, this work appeals to a more realistic goal of machine consciousness research, which would better absorb machine consciousness research into the big picture of AGI or strong cognitive agents.

The whole script is structured as follows. The second part reviews relevant work on episodic memory and relevant attempts for intelligent agents. Based on the critical thinking towards those existing research, the extensive implementation based on our previous work is introduced in the third part. With this implemented model, the agent is expected to demonstrate cognitive capabilities related to memory, under the assistance of consciousness. To show them, the experiment is designed to be an interactive activity between the agent and the human commander. After the model is properly configured for the experiment, the results are organised and discussed in the fifth part. This is followed by the discussion of the implications of this work, providing the authors' opinions on the relationships between memory, consciousness, and general cognitive agents (or AGI). Finally, this paper summarises the achievements of this work and gives the limitations and future possibilities of the current implementation.

II. Computing Attempts on Episodic Memory

A very general definition of this episodic memory was made in [25] that episodic memory is recalling specific past events with what, where and when they happened. It is the only way in the cognitive system allows the agents to experience past experiences. Due to its uniqueness compared to other memory systems, its implications for cognition for humans and even across species are under effortful investigations including [26] [27]. These studies inspired many attempts in computer science to reproduce episodic memory. To have a clear understanding of them, in this part, some of the extant attempts during 2000-2022 are reviewed.

[28] reviews some early mathematical models for episodic memory involving context information retrieval. The focus of their work is on the recognition and recall functions of episodic memory, appealing for a unitary model to support both tasks. Though this is not the current interest of our work, the context information stored in episodic memory is also a key aspect of our work. On the other hand, [28] reflects that temporal information is another key aspect of episodic memory as studied by [29], which has not been implemented in any model in the review. This criticism could also be imposed on some other work. In [30] and [31], the hippocampal model implemented with the neural network technique is claimed to support the recalling of episodic memory. However, the experiments are constrained to context information, testing the agent's ability to recall the word pair patterns. [32] and [33] introduce a quantum mathematical model but also suffer from the same limitation. Though all of these mentioned lack the comprehensive design of episodic memory, they do provide various approaches to implementing non-temporal episodic memory. To account for multiple aspects of episodic memory, [34] proposes a neural network architecture called EM-ART with multiple layers. In their model, not only the context and temporal dimensions of episodic memory are implemented, but also the dynamics of forgetting are employed. With such implementation, the model was examined against other models in tasks of word recognition and shooting game, showing generally superior performance. The work presented in [35], which implements the episodic memory module with long-term memory cells [36], also demonstrates the temporal dimension of episodic memory. These two implementations are both appreciable from the perspective of functionalities of episodic memory, which could be regarded as the main implementation requirements for a cognitive architecture.

Apart from the functions of episodic memory discussed above, the policies of encoding, maintenance and retrieval are also studied. Most extant implementations implicitly encode fixed elements into memory. However, this is not the effective way in which episodic memory works. With the naive motivation to maintain the performance and control the storage consumption, [37] stores a subset of experienced samples for each task. A more intuitive idea in [38] is that episodic memory is updated with important information that is related to the current state and is crucial for the retrieval process. These designs are partial consistent with a common result from studies of psychology that people can remember information which is consciously perceived better and longer [39]. From another perspective, the work of [40] discusses when to encode and retrieve episodic memory. This is motivated by that most extant studies are conducted within naive and fixed tasks, missing the serious consideration of how episodic memory functions in real-time. This point is key to the performance of cognition in a real-world environment as episodic memory has to interact with other modules within a cognitive system. This means the maintenance of memory could not disrupt the continuous information flow of others or the impact has to be minimised. This leads us to think about the implementations of episodic memory within the scope of cognitive architecture with interacting modules.

One example work of episodic memory in cognitive architecture is the Soar [41]. In this project, the temporal dimension of episodic memory is implemented though with simplifications. Also, it covers other features of episodic memory missing or discussed above. Encoding memory is designed to happen when an action is taken by the agent. During the encoding, there is an activation threshold to filter out the information in working memory to be stored in the episodic memory system, and the dynamics of memory strength are considered. However, episodic memory is not well defined, basically driven by the task requirements of the agent. In their implementation, an episode includes the agent's input (sensing), internal data structures and output (actions in the world) [42]. This is quite different from the common sense of episodic memory. As discussed by [26], episodic memory is usually represented in the form of visual images. Though the implementation in Soar architecture might benefit their experiment performances, these over-inclusive episodic memory contents are not appreciated by our work. Another cognitive incorporating episodic memory is the project of LIDA [43] [21]. Compared to other reviewed work here focusing on the encoding, maintenance and retrieval mechanisms of episodic memory, the LIDA architecture emphasises the differences between long-term and short-term episodic memory. They proposed the transient episodic memory, via which the encoding of long-term memory could happen. Though this architecture does not provide much detail about the implementation of episodic memory, it stresses the role of consciousness in memory. This is attributed to the GWT employed in the architecture.

While the papers reviewed here have different focuses and limitations on the implementations, one can summarise the features attracting the majority of researchers' attention to episodic memory. Firstly, episodic memory should encode both context and temporal information. The encoded memory also needs to be filtered via certain mechanisms so that only useful memory is stored. This is an intuitive constraint for humans as we human beings do not encode all elements in all experiences into memory. With what should be encoded into episodic memory specified, when to enter the encoding process also needs consideration for the whole system functioning in a robust and smooth way [40]. Apart from these, the dynamics of memory strength are stressed by some researchers. This is an obvious requirement for implementation regarding the memory curves in our common senses. All of these key factors would guide the implementation of episodic memory in this work. To explicitly express the novelties of our model with respect to the implementation of episodic memory, the table below gives the comparison of different dimensions of the implementations between extant models and our model.

TABLE I
COMPARISON AGAINST EXTANT EPISODIC MEMORY MODELS

Models	Context Memory	Temporal Memory	Selective Encoding	Episodic Information	Maintenance	Retrieval
Foster, 2002	Y	N	N	Word Pairs	N	Y
Schapiro, 2017	Y	N	N	Letter Pairs	N	Y
Brainerd, 2015	Y	N	N	State Vectors	N	Y
Trueblood, 2017	Y	N	N	State Vectors	N	Y
Wang, 2012	Y	Y	N	State Vectors	Dynamics of decaying	Y
Kim, 2019	Y	Y	N	Sentences	Dynamics of decaying	Y
David, 2018	N	N	Subset of learnt Samples	Learning Samples	N	Y
Kasap, 2010	N	N	Important States	State Vectors	N	Y
Lu, 2022	Y	N	N	State Vectors	N	Y
Laird, 2019	Y	Y	N	Not Specified	Dynamics of decaying	Y
Franklin, 2016	Y	Y	Conscious subset of context	Not Specified	Dynamics of decaying	Y
Our Model	Y	Y	Conscious subset of context	First-Order Visual Experiences	Dynamics of decaying and strengthening	Y

To summarise, compared to extant implementations, our design realises more features of episodic memory consistent with the psychological evidence. Even though the work of [21] shares more common points with our model as we are both based on the GWT, the implementation in this work is more specified with respects to the define of episodic information and richer dynamics implemented.

III. Our Model

Based on the discussion of AGI, consciousness, episodic memory and its relevant implementations, it is fair to say that memory is essential to intelligence and consciousness, and consciousness in return interferes the memory formation and retrieval. This work implements a model based on the Global Workspace framework, incorporating sensory modules, attention mechanism, episodic memory and semantic knowledge. A simplified goal component is designed for increasing conscious activities. The overall structure of the agent is illustrated in Fig. 1. Those modules are running in parallel in an asynchronous way, while the lifetime of the whole agent is composed of infinite cognitive cycles. A cognitive cycle in this work is initialised by a signal that enters consciousness. After the signal gets broadcast to other specialists and processed the cognitive cycle ends with the beginning of the next cognitive cycle. The connections in the illustration are not all those implemented but the main channels for communications between those modules. Brief explanations of modules are given below. The details of each would be provided in the appendix for reproductive requirements. The code of the implementation of this whole agent would be also available online after this work is published.

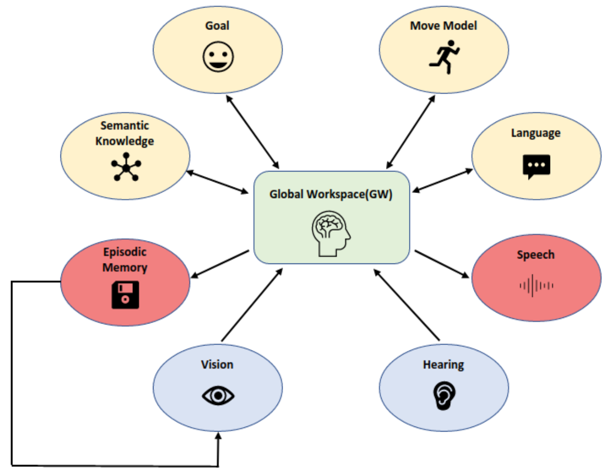

Fig. 1. the structure of the agent implemented in this work

Vision receives signals from either the external world or recalled scenes from episodic memory(frame by frame). It is expected to have multiple objects detected in a frame. Thus, these objects would compete with each with their salience values determined by both pixel values and consciousness interference. After the winner in the visual area is opted out, the sub-image of the winner would be encoded into a semantic representation with its position information together. The generated representation would join the competition into consciousness(GW).

Hearing receives signals from external world. It was designed and implemented to detect speech sentences over time. Due to the generalisation problem of this audio processing, it is replaced by feeding pre-recorded audio to the agent as the speech inputs. The Hearing module encodes audio signals into semantic representations and, with the language inputs, generates a task signal which would be accepted by the Language module. Both semantic representation and the task signal would be wrapped together, competing with other modules for consciousness.

Episodic Memory functions with the interference of consciousness(GW). Each time the interactive environment changes, the contents in GW would be moved to memory along with their strength values. As the contents to be stored have already been filtered by the competitive mechanism into consciousness, all contents inside GW are consciously perceived. The episodic memory in this work is simply implemented by a trace table, which intrinsically retains the temporal feature. Over the life cycle of the agent, the contents inside Episodic Memory would decay with a threshold function, while the revisited memory would be further strengthened. This is a simulation of the forgetting curves of memory. For retrieval, a cue would be provided for memory searching and a retrieval mode would be set for retrieving static context or temporal episodes. The retrieved episodes would become the possible inputs to Vision.

Speech is a simple module independent from Hearing and Language modules. It receives signals from Global Workspace and generates speech outputs if the received signals are compatible with the speech generation.

Language module is only for the analysis of speech inputs. When GW broadcasts speech signals which are characterised by a task signal generated by the Hearing module, the Language module analyses the subject and task indicated by the speech. For instance, the speech signal of 'what is the colour of apple' would be extracted to 'apple' as the subject and 'colour inference' as the task. The task analysed here is usually accompanied by a task clock signal, which indicates how many enquiry cycles are requested by the speech enquiry. Apart from these two analysis results, when the speech signal requests the engagement of Episodic Memory, a cue and a retrieval mode signals would also be generated. The resulting signals from the analysis would be in effect only if the Language module wins the consciousness competition in the next cognitive cycle.

Semantic Knowledge in this implementation contains two inference networks for colour and object recognition. When this module receives the broadcast signal from Global Workspace, it takes the task signal for choosing the responsible network for inference and feeds the subject signal into the chosen network. The output would re-join the competition into consciousness with a fixed salience value.

Goal is functioning as an affective stimuli detector. With a pre-defined global goal of the agent, the goal module would analyse the coming signal from Global Workspace by direct recruitment of the Semantic Knowledge module. However, the goal module does not respond to Global Workspace with generated semantic representations like those generated by the Vision, Hearing or other responsive modules. Instead, if a signal meets the pre-defined goal, a feedback signal would be sent back to Global Workspace to strengthen the conscious activation of the held information in working memory.

Move Model is equipped with a temporal cache. When the agent is required to analyse the temporal pattern of objects, this module would cache the broadcast signals and do the pattern recognition when the inference trigger condition is met. It is fair to consider it as a kind of semantic knowledge for temporal information, though it is implemented independently in current work.

Global Workspace is the core node of the implemented model. At the beginning of each cognitive cycle, Global Workspace reads the signals from all specialist modules except Episodic Memory and Speech. Based on the salience values of signals and the voluntary attention signal generated by Global Workspace based on the signal pattern in the last cognitive cycle, the winner of the competition for consciousness would be broadcast to all modules. Apart from this hub function, Global Workspace also holds a working memory, which caches perceptual signals from Vision, which is the raw images captured of the environment differentiated from the semantic representations. An accompanying conscious score is attached to each held working memory element. The conscious score reflects the transient memory strength inside working memory, and will also decay over time faster than those in Episodic Memory. This is consistent with many studies on working memory, short-term memory and long-term memory. This score will be enhanced if any module is responsive to its encoded signal and will decay over time by a weakening function. Each time the interactive environment changes, the contents held in working memory would be moved to Episodic Memory, and the conscious scores would be stored together as the memory strength values.

The key feature of this implementation is that Global Workspace(or consciousness) is assisting the filtering, formation and retrieval of episodic memory. For filtering, the competitive selection of conscious perceptions intrinsically makes only important information enter the transient memory (working memory). The decaying mechanism within working memory further drops out some memory candidates. For formation, it simulates the enhancement of conscious processing upon memory strength. Each module responsive to the broadcast signal from Global Workspace would enhance the conscious score of the corresponding memory element. Thus, it implicitly takes the effective spread (resulting in responses) of the signal across the whole system as the extent to which the agent is conscious of it. The conscious score would determine the initial memory strength of the episodic memory, which would to some extent determine the memory trace to be short-term or long-term. For retrieval, in this work, only if the agent is consciously cued by certain signals, the retrieval process would be initialised. More detailed descriptions of modules are provided in Appendix.

IV. Experiment Design and Configuration

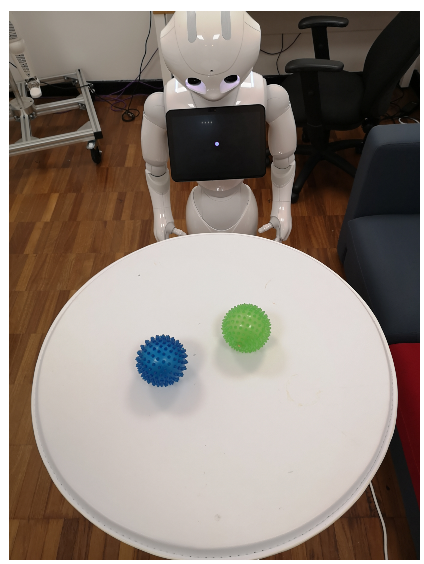

Fig. 2. Pepper robot within the experiment environment

To validate the implementation of the cognitive agent with consciousness-assisted episodic memory, tasks requiring recalling of three aspects of episodic memory are designed. The tasks are human-robot interactions, asking the robotic agent to answer questions about past experiences. Firstly, the robot agent is exposed to the changing experiment environment.

During this period, the agent perceives various visual stimuli and accumulates memory. The environment is illustrated above.

After this, the agent is asked questions requiring her to recall episodic memory about cued objects, the static contexts and the movement histories of objects. Testing the episodic memory functions within this consciousness architecture is consistent with the natures of episodic memory discussed by [26]. For the three task requirements, the interactive questions are set to be 'what is the colour of X', 'what else are around X' and 'how does X move'.

In the exposure phase, the agent is facing a table with 4 distinct objects. The human commander changed the positions of the 4 objects. The changes were made in 2 times, and each time 2 of the 4 are moved. In this configuration, the memory transformation from working memory in Global Workspace to Episodic Memory is expected to happen at least 3 times corresponding to the three different visual environments. More episodes might be formed when the agent consciously perceives the moving commander, which also incurs changes in the visual environment. After the exposure phase, the agent would be exposed to a visual environment irrelevant to the interactive tasks to avoid unexpected disruptions. To configure the agent, networks for processing are exploited from pre-trained repositories online or trained with created dataset for the experiments. They are summarised as follows.

Vision Detector has two candidates during configuration phase, YoLOv5 and Detectron2 [44]. This network is responsible for detecting areas in visual field which possibly contain an object but is not required to recognise them. Though the available detection networks possess functions much beyond this, only the position predictions are retained. The performances of position prediction of these two models are tested using the same series of frames. It turns out that YoLov5 missed more objects while Detectron2 rarely missed. On the other hand, it was obvious that Detectron2 runs much slower compared to the swift predictions of YoLov5, which is attributed to the differences in network structure and model size. For our experiment, the speed is a inferior requirement compared to the accuracy, as the latter will significantly impact the cognitive capability of the agent. Thus, Detectron2 is recruited here.

Vision Encoder takes the inputs from Vision Detector, encoding the sub-image winning the local competition within Vision into a semantic representation. The semantic representations designed in this work is to be the representations in a hidden layer of a classifier network, which is not constrained by accuracy requirement. Instead of the prediction result of the classifier, the hidden layer states are much more sparse. As expected, the semantic representations for different objects, and the same object in different positions and poses would be different from each other. This requirement could be met by many open-source algorithms. The employed one here is Resnet50, which could be directly instantiated from the PyTorch repository.

Hearing Encoder is similar to the Vision Encoder, though it has less input channels. The audio signals are firstly transformed into image format based on Mel-frequency cepstral coefficients (MFCC) features. The transformation result is encoded by Hearing Encoder into semantic the representation.

Language Analyser is trained on the experiment data. The inputs and outputs are both semantic representations generated by the Hearing Encoder. While the inputs and outputs contains different information. The inputs are representations of sentence speeches. The outputs consist of a representation of words, concatenated with task signal, retrieval mode signal and task clock signal which are explained in the previous part.

Colour Classifier and ***Object Classifier*** are trained with experiment data. The inputs are semantic representations of sub-images generated by the Vision Encoder. The outputs are semantic representations of words generated by the Hearing Encoder. This is a simulation of associations between visual and auditory signals. For the model implemented in this work, the space of semantic representations is playing the role of the intermediate space for associations and communications between different modalities.

Movement Analyser is trained with samples generated from the collected experiment data. The input for this analyser is a concatenation of two semantic representations of two sub-images and the output is a label able to represent 16 different movement directions. Though this network is designed to only analyse the movement of the same object over time, the fed training data includes pairs of different objects. Theoretically, this network is generally able to classify the positional relations between two objects.

The descriptions in this part are for brief explanations only. The configurations of model details including the training of networks, data, and hyper-parameters of the agent would be explained in Appendix.

V. RESULTS AND ANALYSIS

In this part, the results of 3 recall-inference tasks corresponding to different aspects of episodic memory are illustrated. Five trials are analysed with the episode traces, memory content, interactive performance and the memory strength curves. For the memory content analysis, though the agent is exposed to a real interactive environment, perceiving continuous frames, the analysis of the episodes only displays the episodes incurring memory formation. This is because only frames changing from the previous perception will result in the memory store process. However, the full experiences of the agent would be provided via open source for any people interested in analysing by themselves.

A. Trial 1

In the memory trace analysis, there are four episodes triggering the store of episodic memory. Surprisingly, the phone in each episode is not consciously perceived by the agent as shown in the memory contents analysis. This is the only exception throughout the 5 trials, which might be attributed to the occasional performance flaw of the detector. Interestingly, there are still objects recognised as phones in this trial, generating unexpected mistakes in the interactive tasks. That the answer for context enquiry is correct when asked with the phone is understandable as the agent successfully accumulates the memory and knowledge of the context in the exposure phase.

Fig. 3. Memory trace analysis of episodes incurring memory formation in trial 1. The visual perceptions happened to the agent from the left to the right.

Fig. 4. Memory trace analysis of memory contents in trial 1. The traces are memorised by the agent from top to bottom corresponding to the sequence in Fig. 3.

TABLE II
INTERACTIVE PERFORMANCE OF TRIAL 1

	Mouse	Phone	Car	Toy
Movement	Wrong ans	Still	Still	Right
Colour	Black	Red	Yellow	Red
Context	Phone, Toy, Car, Unknown	Toy, Mouse, Car, Unknown	Phone, Toy, Mouse, Unknown	Phone, Mouse, Car, Unknown

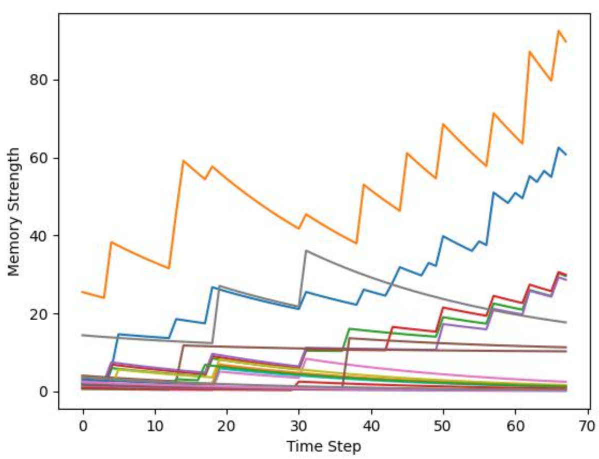

Fig. 5. Memory trace analysis of episodes incurring memory formation in trial 1. The visual perceptions happened to the agent from the left to the right.

B. Trial 2

In this trial, only three episodes incur the memory storage process. That the frames disrupted by the human commander do not result in episodic memory formation could be attributed to the attentional blink effect studied in our previous work [18]. While the consciousness is dedicated to previous frames, the frames with disruptions are not consciously perceived by the agent. In this trial, all 4 wanted objects are successfully perceived by the agent in consciousness as shown in the figure of the memory contents analysis. However, according to the inference results of the phone, it seems that a wrong object is recognised as a phone and results in the wrong answers for movement and colour classifications. Apart from this, all of the other inferences on recalled memory are correct.

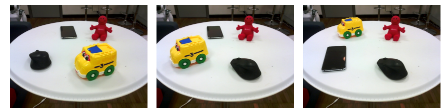

Fig. 6. Memory trace analysis of episodes incurring memory formation in trial 2. The visual perceptions happened to the agent from the left to the right.

Fig. 7. Memory trace analysis of memory contents in trial 2. The traces are memorised by the agent from top to bottom corresponding to the sequence in Fig. 6.

TABLE III
INTERACTIVE PERFORMANCE OF TRIAL 2

	Mouse	Phone	Car	Toy
Movement	Right	Still	Farther and Left	Still
Colour	Black	Yellow	Yellow	Red
Context	Phone, Car, Toy, Unknown, Unknown	Car, Mouse, Toy, Unknown, Unknown	Phone, Mouse, Toy, Unknown, Unknown	Phone, Car, Mouse, Unknown, Unknown

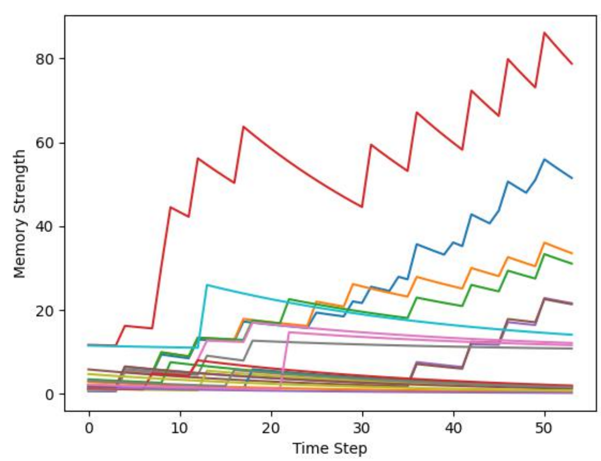

Fig. 8. Memory trace analysis of episodes incurring memory formation in trial 2. The visual perceptions happened to the agent from the left to the right.

C. Trial 3

Similarly to Trial 1, there are four episodes resulting in episodic memory formation, one of which is caused by the disruption of the human commander. In this trial, all objects are successfully perceived in consciousness and all the inferences on the recalled memory are correct except the movement classification of the toy. This is attributed to the prediction error of the Movement Analyser.

Fig. 9. Memory trace analysis of episodes incurring memory formation in trial 3. The visual perceptions happened to the agent from the left to the right.

Fig. 10. Memory trace analysis of memory contents in trial 3. The traces are memorised by the agent from top to bottom corresponding to the sequence in Fig. 9.

TABLE IV
INTERACTIVE PERFORMANCE OF TRIAL 3

	Mouse	**Phone**	**Car**	**Toy**
Movement	Left	Closer and Right	Still	Still
Colour	Black	Black	Yellow	Red
Context	Phone, Toy, Car, Unknown, Unknown	Mouse, Toy, Car, Unknown, Unknown	Phone, Mouse, Toy, Unknown, Unknown	Phone, Mouse, Car, Unknown, Unknown

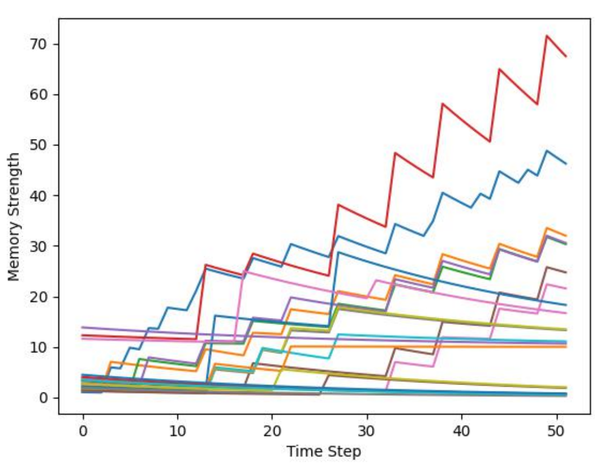

Fig. 11. Memory trace analysis of episodes incurring memory formation in trial 3. The visual perceptions happened to the agent from the left to the right.

D. Trial 4

Three episodes are perceived as changing the interactive environment, resulting in memory formation. The inference mistakes happen in the movement classification, with an invalid prediction made for the movement of the mouse. The invalid prediction (which also happens in Trial 1) indicates the object is classified as moving in conflicting directions, such as moving to the right also the left or move to farther also closer. This is attributed to the imperfect performance of the Move Analyser network. On the other hand, the overlapped detections in this trial are obvious. The impact of this is not well investigated. This is because of that the results are generated by the collaboration of multiple modules, which makes the analysis hard. However, this is not the focus of this work.

Fig. 12. Memory trace analysis of episodes incurring memory formation in trial 4. The visual perceptions happened to the agent from the left to the right.

Fig. 13. Memory trace analysis of memory contents in trial 4. The traces are memorised by the agent from top to bottom corresponding to the sequence in Fig. 12.

TABLE V
INTERACTIVE PERFORMANCE OF TRIAL 4

	Mouse	**Phone**	**Car**	**Toy**
Movement	Wrong ans	Still	Still	Farther and Left
Colour	Black	Black	Yellow	Red
Context	Phone, Toy, Car, Unknown	Toy, Mouse, Car, Unknown	Phone, Toy, Mouse, Unknown	Phone, Mouse, Car, Unknown

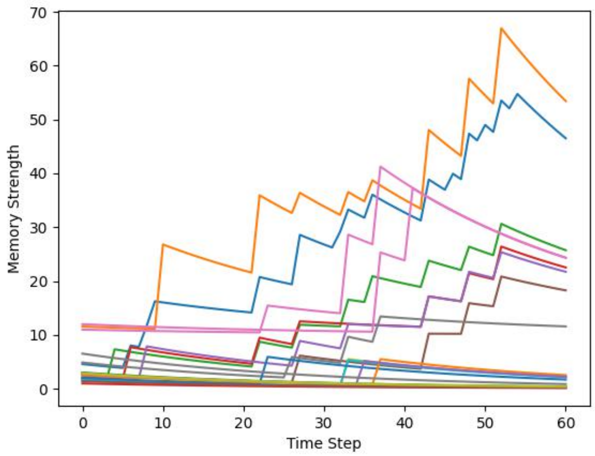

Fig. 14. Memory trace analysis of episodes incurring memory formation in trial 4. The visual perceptions happened to the agent from the left to the right.

E. Trial 5

Similar to other trials, based on the successful perceptions of the desired objects over episodes, mistakes only happen in the movement classification.

Fig. 15. Memory trace analysis of episodes incurring memory formation in trial 5. The visual perceptions happened to the agent from the left to the right.

Fig. 16. Memory trace analysis of memory contents in trial 5. The traces are memorised by the agent from top to bottom corresponding to the sequence in Fig. 15.

TABLE VI
INTERACTIVE PERFORMANCE OF TRIAL 5

	Mouse	**Phone**	**Car**	**Toy**
Movement	Still	Still	Closer and Right	Farther and Left
Colour	Black	Black	Yellow	Red
Context	Phone, Car, Toy, Unknown, Unknown	Mouse, Car, Toy, Unknown, Unknown	Mouse, Phone, Toy, Unknown, Unknown	Mouse, Phone, Car, Unknown, Unknown

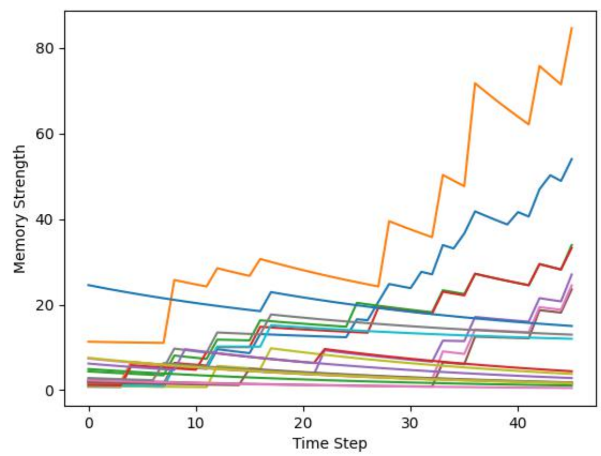

Fig. 17. Memory trace analysis of episodes incurring memory formation in trial 5. The visual perceptions happened to the agent from the left to the right.

From the experiments, some basic features of the Global Workspace framework [5] could be figured out. The interaction experiments involve the coordination between multiple modules implemented in this model, in which the exclusive permission among parallel specialist modules into consciousness is the basic rule of the model. Another feature of global availability is more implicit, while it is also functioning here considering the agent is simultaneously getting tasks done and goals fulfilled in multiple modules. Interestingly, the attentional blink effect [45], which was validated in other GWT research [17] [46] [47], is again generated as a side effect in this model. This is demonstrated during the phase in which the agent is exposed to a continuous visual experience in the real environment. By comparing the memory trace and real experience, some frames are found not consciously perceived by the agent. This attentional blink effect is mainly attributed to the suppression of consciousness permission from the global workspace, which has been well discussed in our previous work [18]. In this model, this suppression mechanism is formally implemented as a block signal in this model maintained by all the specialists.

As for the results, the episodic memory [26] part is the focus of this work. According to the interactions, the agent shows us the ability to recall visual experiences for different task situations. The interaction tasks require the agent to answer questions with her knowledge of the colour, movement history and experience context of certain cued information. They demand the ability to recall static, temporal and context information respectively. As shown in the results, though with minor mistakes, the agent recalls a memory about a static object to answer the colour, about the temporal events to answer the movement pattern of objects, and about the context to answer the objects in a certain episode. From these, episodic memory is successfully implemented in this model in terms of memory contents.

What should be emphasised is the role of consciousness in the formation of memory in this work. Firstly, only the experience attending consciousness would be encoded into working memory in the Global Workspace node of this model. After this, when the consciousness about the whole environment is changed, the contents of the working memory would be moved into the memory area. By this, consciousness basically plays the role of filter for efficient and effective memory formation. Moreover, the dynamics of episodic memory are also stressed in this work. The memory trace is autonomously forming short-term and long-term ones under the interference of consciousness. As explained in the model part, memory strengths are determined by conscious exposure as well as affection enhancement. As shown in the memory strength curves in each trial, some starting above 10 units of memory strength always remains above the threshold level. These are objects meeting the affection goal (redness) during the exposure phase. Besides the affection enhancement, memory could sometimes get enhanced to be stronger, being equal to or more than 10 units and transform into long-term memory if that memory is revisited in later interactive activities. However, memory contents starting below the threshold and never being lifted above the threshold would eventually fade away from the brain of the agent. This means the memory is forgotten to be inaccessible.

VI. Discussion

Instead of building a cognitive agent based on a given architecture such as LIDA [19] and IDyOT [20], this work appreciates implementing specialist modules related to conscious cognition in an incremental way. Though overlaps with those extant architectures are inevitable, authors expect to contribute to the diversity of cognitive architecture research.

As explained in the model part, differences between the implementation in this work and the extant models can be identified with respect to the design of the episodic memory module. Compared to the reviewed models in past work, this implementation successfully demonstrates different forms of episodic memory, including the retrieval of static, context and temporal information. Importantly, this work shows that consciousness is essential to all the formation, maintenance and retrieval processes of episodic memory, which is consistent with the psychological evidence (e.g., [39]). From this perspective, this work is complementing the work in [43]. With the assistance of consciousness, episodic memory also contributes to the consciousness of the agent in return. Implementing episodic memory into the cognitive system implies that the agent can demonstrate cognitive capabilities over time, which are validated by the interactive experiments in this work. Interestingly, this work does not explicitly implement short-term and long-term memory systems. However, episodic memory automatically transforms into long-term memory or fades away as short-term memory. This is partially realised by consciousness. Based on this implementation, the agent is much stronger than systems only responding to the present time. One example of the weaker agent is the precedent of this implementation [18]. Without episodic memory, the agent can only react to the present stimuli. From this point, one would definitely agree that the implemented agent in this work performs more consciously and intelligently.

Strengthened by this work, the authors would like to stress the potential of the consciousness framework for potent cognitive agents or AGI. Though it is still hard to define consciousness and intelligence, it is believed that the consciousness framework is the substrate of intelligence regarding its role of integrating all the specialists key to general intelligence. By comparing this implementation to its precedent version, it is reasonable to conclude that richer dimensions or contents of consciousness are important for achieving the ultimate goal. This leads us to think about the essence of a conscious machine. An attempted opinion is that the Global Workspace framework in our research makes the consciousness phenomenon possible to happen while the contents of consciousness define what we feel about consciousness. On the other hand, many researchers are struggling to find out the key correlates of consciousness. In other words, they attempt to explain what exactly makes consciousness arise [48]. However, they all yet generate no convincing outcomes. From another perspective, researchers in psychological, medical and neuroscientific studies never propose the loss of consciousness accompanying the loss of certain brain functions. For instance, [49] [50] [51] point out the disorders of consciousness due to certain brain impairments, while only a subset of consciousness is lost in each case. This is consistent with the opinion in this paper. Though many processes are believed to be conscious such as thinking, emotion, self-reflection and etc., none of them is the producer of consciousness. The consciousness rises from the structure within which all processors function in a way coordinated by systematic principles. These principles are competition and broadcast cycle in the definition of GWT, though others may propose other frameworks. However, the system implemented with only a processing structure without any or with an insufficient subset of consciousness correlates seems unreliable and hardly convinced to be conscious. The extant models reviewed in this paper are all examples of this group. This is simply because of the elusive definition of consciousness. Basically, people including the authors intuitively expect conscious machines to behave in a human-like way. Otherwise, we all will claim the implemented agent to be conscious. However, that intuitive expectation seems very similar to the one for a generally intelligent machine. This is also a big confusion for the authors in the research of artificial consciousness. Thus, we appeal for an explicit detachment of consciousness from intelligence. Here, the authors provide our opinions for discussion. A framework like GWT in this work is the substrate of consciousness, while the specialist processors provide various contents into consciousness which are the embodiment of consciousness. In other words, only if certain information from the external or internal world is entering consciousness, consciousness is reportable. This leads to that both the framework and specialists are essential to consciousness. However, the authors propose that not all dimensions of consciousness are necessary and researchers should not, at least for now, pursue the reproduction of all kinds of consciousness of humans. Within the big picture of AGI, researchers only need a minimal subset of consciousness correlates which are sufficient to meet the requirements of developing human-like cognitive agents. Also, no matter what specialist modules are implemented, the researchers should be confident to claim the created machine consciousness as long as the framework employed is consistent with human consciousness.

For future work, the implementation of the cognitive architecture could be enhanced in two directions. Firstly, though the episodic memory module implemented here is sufficient for the demonstration, the table-based structure is very constrained. People are usually not convinced with this storage form inside human brains. For this, the neural network models such as [34] [31] [35] might shed light on our future attempts. On the other hand, the current implementation is still way far from a reliable cognitive agent. More correlates are expected to be implemented. An example is learning ability which is usually emphasised by researchers [52]. As discussed above, though we can claim the implemented agent to be conscious, finding out a minimal subset of consciousness correlates is crucial for further developing strong cognitive machines.

Acknowledgments

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952324 (TRAINCREASE).

Appendix A. Model Descriptions

In this appendix, the model is explained in detail for people who are interested. The information processes are illustrated with diagrams when necessary. This appendix only gives the details at processing level. For people who are interested in reproduction or doing research based on this model, in case there are any missing details, the code would be available online as an open-source resource after this paper is published.

Vision receives signals from either the external world or recalled scenes from episodic memory(frame by frame). The competition between these two sources are determined by a control signal ***G*** according to the interactive task. This signal is one of the element in the analysis result of Language module which will be explained in detail later. As the information is in different format in the two sources. When the agent is perceiving an image from the external world, it goes through a detector network ***Detectron2***, resulting in the detected objects. Otherwise, the detected objects are directly retrieved from ***Episodic Memory***. Importantly, the array data of an object consists of not only the sub-image but also the position information. The position information is represented by two channels appended to the RGB channels of the image. Also, for computational convenience, the data arrays are reshaped into ***[64, 64, channelNum]***. Thus, each array data is of ***[64, 64, 5]***.

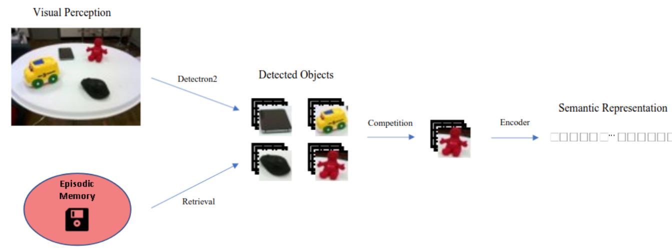

Fig. 18. Information process of Vision

When there are more than one object, these objects would compete with each other based on their salience values. For stimuli from the visual field, the salience values are partially determined by the difference of pixel values between the sub-image of certain object ***Sub*** and its local background ***Local***, which is defined as the square area surrounding the object. The size of the area is as three times as the sub-image in both terms of height and width as illustrated in figure below.

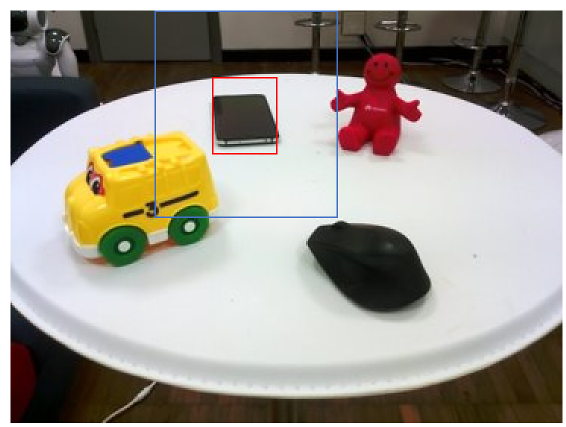

Fig. 19. the local background for a detected object

The calculation of the pixel salience is

$$PixelSalience = \frac{|Sum(Local) - 2 * Sum(Sub)|}{MaxPixelSalience} \tag{1}$$

Where ***MaxPixelSalience*** is held in ***Vision***. It is updated every time by the molecular if ***PixelSalience*** is greater than ***1.0***. For the objects from ***Episodic Memory***, the cued object has the highest initial salience ***1.0*** while the rest have the same initial salience ***0.8***. After the pixel salience values are calculated, the consciousness interference is another factor determining the salience values of those stimuli. This interference will decrease the salience value of the objects which have previously entered ***GW*** to be lower than the smallest salience value among all detected objects. This is assited by the workign memory ***wm*** held by ***GW***.

$$obj.Salience = \begin{cases} min(PixelSalienceValues), & obj.in(\boldsymbol{wm}) \\ obj.PixelSalienceValue, & else \end{cases} \tag{2}$$

Then, those stimuli compete by the salience values. The winner is encoded into a semantic representation. The generated representation would join the competition into consciousness(GW).

Hearing receives signals from external world. The received audio input is firstly transformed into the MFCC feature map. Then, the module encodes this feature map into a semantic representation and, with the language input, generates a task signal which would be accepted by the ***Language*** module. Both semantic representation and the task signal would be wrapped together, competing with other modules for consciousness. Also, ***Hearing*** holds a ***MaxAmplitudeSalience***. The salience of the audio input is calculated with the strongest amplitude of the audio signal and normalised by ***MaxAmplitudeSalience***.

$$Salience = \frac{Max(AudioSignal)}{MaxAmplitudeSalience} \tag{3}$$

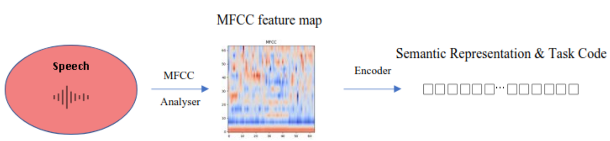

Fig. 20. Information process of Hearing

Episodic Memory functions with the interference of consciousness(GW). Each time the interactive environment changes, the contents in GW would be moved to memory along with their strength values. As the contents to be stored have already been filtered by the competitive mechanism into consciousness, all contents inside GW are consciously perceived. The episodic memory in this work is simply implemented by a trace table, which intrinsically retains the temporal feature. Over the life cycle of the agent, the contents inside Episodic Memory would decay with a threshold function, while the revisited memory would be further strengthened. This is a simulation of the forgetting curves of memory. During the maintenance, a part of the memory would automatically transform into long-term memory, while others fade away which can be regarded as short-term memory. The memory strength of each memory element at each time step would be updated by the decaying function below.

$$Strength- = \begin{cases} \frac{Strength-10.}{30.}, & Strength > 10. \\ \frac{Strength}{30.}, & else \end{cases} \tag{4}$$

Where ***10.*** is the threshold distinguishing short-term and long-term memory and ***30.*** is the decaying factor.

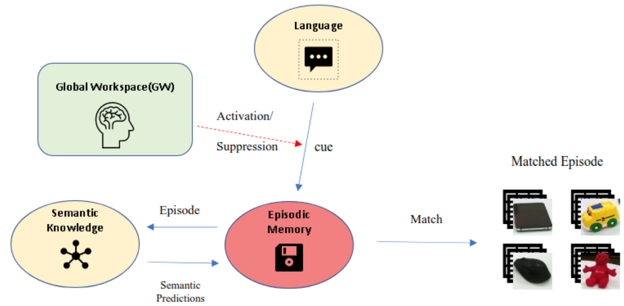

Fig. 21. Information process of Episodic Memory

For retrieval, a cue would be provided for memory searching and a retrieval mode would be set for retrieving static context or temporal episodes. For the former, the only one memory frame would be retrieved, while, with the latter mode, retrieval process would go through all memory traces one by one, returning the trace if cued information is contained. The retrieved episodes would become the inputs to Vision.

Speech is a simple module independent from ***Hearing*** and ***Language*** modules. It receives signals from Global Workspace and generates speech outputs. However, this will take effect only when the agent is expected to answer a question.

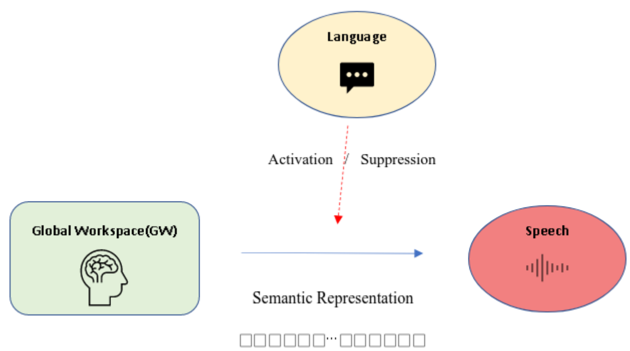

Fig. 22. Speech Generation

Language module is only for the analysis of speech inputs. When GW broadcasts speech signals which are characterised by a task signal generated by the ***Hearing*** module, ***Language*** module analyses the subject and task indicated by the speech. For instance, the speech signal of 'what is the colour of apple' would be extracted to 'apple' as the subject and 'colour inference' as the task. Apart from these, the result also contains a part of control signal, consisting of a source gate signal, a task clock signal, and a retrieval mode signal. The source gate signal indicates if the retrieved memory should be taken as the perceived stimuli, imposed on ***Vision***. The task clock signal indicates how many enquiry cycles are requested by the speech enquiry. Each enquiry cycle ends when a the inference task is completed, the clock value held in ***GW*** will decrease while an enquiry cycle is completed. When the task clock value is less than 1, the task is totally completed. For instance, when the agent is asked to answer all the objects in a certain frame, the 'object inference' task signal would be held for multiple enquiry cycles. Only when all objects are answered indicated by the task clock value less than 1, the task signal would be released. Another contained in the control signal is, when the speech signal requests the engagement of ***Episodic Memory***, a retrieval mode signal. Also, in this case, the semantic representation of the analysed subject would be taken as the cue for memory retrieval. Importantly, these resulting signals from the analysis would be in effect only if the ***Language*** module wins the consciousness competition in the next cognitive cycle.

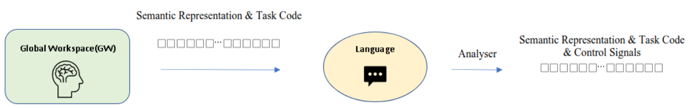

Fig. 23. Information process of Language module

Semantic Knowledge in this implementation contains two inference networks for colour and object recognition. When this module receives the broadcast signal from Global Workspace, it takes the task signal for choosing the responsible network for inference and feeds the semantic representation which is the subject of inference into the chosen network. The output would re-join the competition into consciousness with a fixed salience value. This result would be accepted by ***Speech*** if this module wins the competition into consciousness, making the inference answer be articulated.

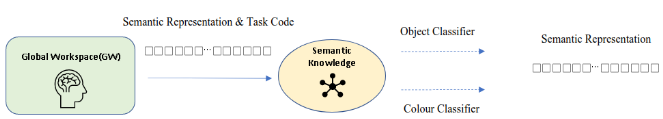

Fig. 24. Information process of Semantic Knowledge

Goal is functioning as an affective stimuli detector. With a pre-defined global goal of the agent, the goal module would analyse the coming signal from ***GW*** by direct recruitment of the ***Semantic Knowledge*** module. However, the goal module does not respond to ***GW*** with generated semantic representations like other responsive modules. Instead, if a signal meets the pre-defined goal, an enhancement signal would be sent back to ***GW*** to strengthen the conscious activation of the held information in working memory. This is also assited by the working memory of ***GW***, which holds the raw information of the analysed semantic representation.

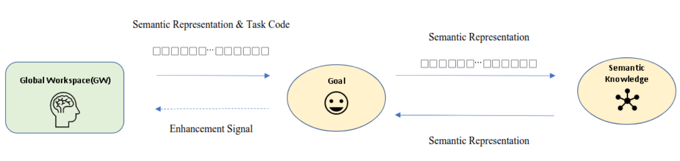

Fig. 25. Information process of Goal module

Move Model is equipped with a temporal cache. When the agent is required to analyse the temporal pattern of objects, this module would cache the broadcast signals and do the pattern recognition when the inference trigger condition is met. It is fair to consider it as a kind of semantic knowledge for temporal information, though it is implemented independently in current work. This module firstly receives broadcast signal from ***GW***, when the task code is recognised, ***Move Model*** firstly checks the received semantic representation with the cue signal generated in ***Language***. As explained, the cue signal is also the resulting subject signal in ***Language***. By employing the object inference network of ***Semantic Knowledge***, if the semantic representation is verified, it is cached in ***Move Model***. After retrieval of the temporal sequence of a cued information, only the oldest and latest two frames would be fed into the movement analysis network, resulting in a semantic representation of direction. This result would be accepted by ***Speech*** if this module wins the competition into consciousness, making the analysed direction be articulated.

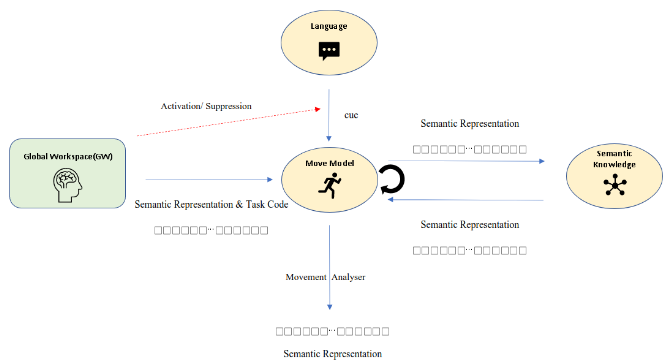

Fig. 26. Information process of Move Model

Global Workspace is the core node of the implemented model. At the beginning of each cognitive cycle, Global Workspace reads the signals from all specialist modules except Episodic Memory and Speech. Based on the salience values of signals and the voluntary attention signal generated by Global Workspace based on the signal pattern in the last cognitive cycle, the winner of the competition for consciousness would be broadcast to all modules. Apart from this hub function, Global Workspace also holds a working memory, which caches perceptual signals from Vision, which is the raw images captured of the environment differentiated from the semantic representations. An accompanying conscious score is attached to each held working memory element. The conscious score reflects the transient memory strength inside working memory, and will also decay over time faster than those in Episodic Memory. This is consistent with many studies on working memory, short-term memory and long-term memory. This score will be enhanced if any module is responsive to its encoded signal and will decay over time by a weakening function. Each time the interactive environment changes, the contents held in working memory would be moved to Episodic Memory, and the conscious scores would be stored together as the memory strength values. The information processes of

GW is complemented by above all explained modules.

Appendix B. Model Configurations

In this appendix, the configuration of 4 self-trained networks would be introduced.

Language Analyser is convolutional neural network trained on the experiment data, with input size 8192 and output size 8203. The inputs and outputs are both semantic representations generated by the Hearing Encoder. While the inputs and outputs contains different information. The inputs are representations of sentence speeches. The outputs consist of a semantic representation of a word, concatenated with task signal, retrieval mode signal and task clock signal which are explained in the previous part. For the input data, the predefined three questions are recorded for each subject. In this experiment, 10 subjects are prepared for training while only 4 are used in the interactive experiments. For each sentence, at least 10 variations are recorded, with minor differences in numbers for randomness. Furthermore, each speech sample, for enhancing the data set to achieve a good accuracy, is augmented by adding noise and shifting the pitches, resulting in 10 more variations for each sample. This results in the input data set of 6043 speeches. All speeches are then encoded by the Hearing Encoder into semantic representations, the size of each is 8192. For the output data, the first 8192 elements are the semantic representation of a word, which here is an object name. The task code, source gate, retrieval mode and task clock value consist of 3, 1, 2, 5 elements after the semantic representation respectively. After the hyper-parameter pruning, this network is trained by the Adam optimiser with learning rate 0.001 for 800 epochs.

Colour Classifier and ***Object Classifier*** are both convolutional neural networks trained with the same input data collected in the experiment environment by the Pepper robot. The input and output have the same size of 8192. Firstly, we use the Pepper robot to take pictures of the visual field containing the experiment objects. Then those images are processed by the Vision Detector, resulting in 532 sub-images containing certain objects, appended with the position information. After this, those sub-images are manually annotated. The inputs are semantic representations of sub-images generated by the Vision Encoder. The outputs are semantic representations of words generated by the Hearing Encoder. This is a simulation of associations between visual and auditory signals. After the hyper-parameter pruning, these two networks are trained by SGD optimiser, with learning rate 0.001 for 200 epochs.

Movement Analyser is trained with same data for colour and object classifier. It is also a convolutional neural network with input size of 2*8192 and output size of 4. The input for this analyser is a concatenation of two semantic representations of two sub-images and the output is a label representing 8 different movement directions. For the input data, the pairs are generated from those sub-images produced in the training of colour and object classifiers. According to the position information appended to those sub-images as explained above, those pairs are manually annotated. Finally, 93856 pairs are generated for training. For the output, though the representation space of 4 elements is 16, not all combinations are valid. The 4 elements represent respectively closer, farther, right, left. Thus, there are only 8 valid results. Notably, this network is designed to only analyse the movement of the same object over time, while the fed training data includes pairs of different objects. Theoretically, this network is generally able to classify the positional relations between two objects. After the pruning of hyper-parameters, this analyser is trained by Adam optimiser with learning rate 0.001 for 500 epochs.

References

[1] B. Goertzel, "Artificial general intelligence: concept, state of the art, and future prospects," *Journal of Artificial General Intelligence*, vol. 5, no. 1, p. 1, 2014.

[2] A. Chella and R. Manzotti, "Agi and machine consciousness," in *Theoretical foundations of artificial general intelligence*. Springer, 2012, pp. 263–282.

[3] U. Faghihi and S. Franklin, "The lida model as a foundational architecture for agi," in *Theoretical foundations of artificial general intelligence*. Springer, 2012, pp. 103–121.

[4] P. Wang, *Non-axiomatic logic: A model of intelligent reasoning*. World Scientific, 2013.

[5] B. J. Baars, *A cognitive theory of consciousness*. Cambridge University Press, 1993.

[6] H. Basgol, I. Ayhan, and E. Ugur, "Time perception: A review on psychological, computational and robotic models," *IEEE Transactions on Cognitive and Developmental Systems*, 2021.

[7] I. Lourenço, R. Mattila, R. Ventura, and B. Wahlberg, "A biologically-inspired computational model of time perception," *IEEE Transactions on Cognitive and Developmental Systems*, 2021.

[8] J. S. Hall, "The robotics path to agi using servo stacks," in *2nd Conference on Artificiel General Intelligence (2009)*. Atlantis Press, 2009, pp. 20–25.

[9] M.-J. Escobar, N. Navarro-Guerrero, J. Ruiz-Del-Solar, and G. Sandini, "Special issue on emerging topics on development and learning," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, no. 2, pp. 255–257, 2022.

[10] B. J. Baars *et al.*, *In the theater of consciousness: The workspace of the mind*. Oxford University Press, USA, 1997.

[11] G. A. Mashour, P. Roelfsema, J.-P. Changeux, and S. Dehaene, "Conscious processing and the global neuronal workspace hypothesis," *Neuron*, vol. 105, no. 5, pp. 776–798, 2020.

[12] B. J. Baars, S. Franklin, and T. Z. Ramsoy, "Global workspace dynamics: cortical "binding and propagation" enables conscious contents," *Frontiers in psychology*, vol. 4, p. 200, 2013.

[13] B. J. Baars, "The global workspace theory of consciousness: Predictions and results," *The blackwell companion to consciousness*, pp. 227–242, 2017.

[14] G. Deco, D. Vidaurre, and M. L. Kringelbach, "Revisiting the global workspace orchestrating the hierarchical organization of the human brain," *Nature human behaviour*, vol. 5, no. 4, pp. 497–511, 2021.

[15] S. Dehaene, M. Kerszberg, and J.-P. Changeux, "A neuronal model of a global workspace in effortful cognitive tasks," *Proceedings of the national Academy of Sciences*, vol. 95, no. 24, pp. 14 529–14 534, 1998.

[16] B. J. Baars, N. Geld, and R. Kozma, "Global workspace theory (gwt) and prefrontal cortex: Recent developments," *Frontiers in Psychology*, p. 5163, 2021.

[17] M. Shanahan, "A cognitive architecture that combines internal simulation with a global workspace," *Consciousness and cognition*, vol. 15, no. 2, pp. 433–449, 2006.

[18] W. Huang, A. Chella, and A. Cangelosi, "A design of global workspace model with attention: Simulations of attentional blink and lag-1 sparing," *Journal of Artificial Intelligence and Consciousness*, vol. 9, no. 01, pp. 29–57, 2022.

[19] S. Franklin, T. Madl, S. D'mello, and J. Snaider, "Lida: A systems-level architecture for cognition, emotion, and learning," *IEEE Transactions on Autonomous Mental Development*, vol. 6, no. 1, pp. 19–41, 2013.

[20] G. A. Wiggins and J. Forth, "Idyot: a computational theory of creativity as everyday reasoning from learned information," in *Computational creativity research: Towards creative machines*. Springer, 2015, pp. 127–148.

[21] S. Franklin, T. Madl, S. Strain, U. Faghihi, D. Dong, S. Kugele, J. Snaider, P. Agrawal, and S. Chen, "A lida cognitive model tutorial," *Biologically Inspired Cognitive Architectures*, vol. 16, pp. 105–130, 2016.

[22] R. VanRullen and R. Kanai, "Deep learning and the global workspace theory," *Trends in Neurosciences*, vol. 44, no. 9, pp. 692–704, 2021.
[23] E. Tulving, "What is episodic memory?" *Current directions in psychological science*, vol. 2, no. 3, pp. 67–70, 1993.
[24] E. Tulving *et al.*, "Episodic memory: From mind to brain," *Annual review of psychology*, vol. 53, no. 1, pp. 1–25, 2002.
[25] E. Tulving, "Episodic and semantic memory." 1972.
[26] M. A. Conway, "Episodic memories," *Neuropsychologia*, vol. 47, no. 11, pp. 2305–2313, 2009.
[27] T. A. Allen and N. J. Fortin, "The evolution of episodic memory," *Proceedings of the National Academy of Sciences*, vol. 110, no. supplement_2, pp. 10 379–10 386, 2013.
[28] J. K. Foster, *the Oxford Handbook of Memory*, 2002, vol. 125, no. 2.
[29] M. W. Howard, K. H. Shankar, W. R. Aue, and A. H. Criss, "A distributed representation of internal time." *Psychological review*, vol. 122, no. 1, p. 24, 2015.
[30] K. A. Norman and R. C. O. Reilly, "modeling hippocampal and neocotical contributions to recognition memory," 2002.
[31] A. C. Schapiro, N. B. Turk-Browne, M. M. Botvinick, and K. A. Norman, "Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 372, no. 1711, 2017.
[32] C. J. Brainerd, Z. Wang, V. F. Reyna, and K. Nakamura, "Episodic memory does not add up: Verbatim-gist superposition predicts violations of the additive law of probability," *Journal of Memory and Language*, vol. 84, pp. 224–245, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.jml.2015.06.006
[33] J. S. Trueblood and P. Hemmer, "The Generalized Quantum Episodic Memory Model," *Cognitive Science*, vol. 41, no. 8, pp. 2089–2125, 2017.
[34] W. Wang, B. Subagdja, A. H. Tan, and J. A. Starzyk, "Neural modeling of episodic memory: Encoding, retrieval, and forgetting," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 10, pp. 1574–1586, 2012.
[35] U. H. Kim and J. H. Kim, "Integration of Semantic and Episodic Memories for Task Intelligence," *Communications in Computer and Information Science*, vol. 1015, no. 12, pp. 85–100, 2019.
[36] J. A. Starzyk and H. He, "Anticipation-based temporal sequences learning in hierarchical structure," *IEEE Transactions on Neural Networks*, vol. 18, no. 2, pp. 344–358, 2007.
[37] David and Marc, "Gradient Episodic Memory," *Mississippi Legislature*, no. February, p. 2, 2018. [Online]. Available: http://arxiv.org/abs/1712.01169
[38] Z. Kasap and N. Magnenat-Thalmann, "Towards episodic memory-based long-term affective interaction with a human-like robot," *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, pp. 452–457, 2010.
[39] H. P. Bahrick, "Semantic memory content in permastore: fifty years of memory for spanish learned in school." *Journal of experimental psychology: General*, vol. 113, no. 1, p. 1, 1984.
[40] Q. Lu, U. Hasson, and K. A. Norman, "A neural network model of when to retrieve and encode episodic memories," *eLife*, vol. 11, pp. 1–43, 2022.
[41] J. E. Laird, *The Soar cognitive architecture*. MIT press, 2019.
[42] A. M. Nuxoll and J. E. Laird, "Enhancing intelligent agents with episodic memory," *Cognitive Systems Research*, vol. 17-18, pp. 34–48, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.cogsys.2011.10.002
[43] S. Franklin, B. J. Baars, U. Ramamurthy, and M. Ventura, "The role of consciousness in memory," *Memory*, vol. 1, no. 1, pp. 1–38, 2005. [Online]. Available: http://cogprints.org/5806/
[44] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, "Detectron2," https://github.com/facebookresearch/detectron2, 2019.
[45] J. E. Raymond, K. L. Shapiro, and K. M. Arnell, "Temporary suppression of visual processing in an rsvp task: An attentional blink?" *Journal of experimental psychology: Human perception and performance*, vol. 18, no. 3, p. 849, 1992.
[46] M. Shanahan, "The brain's connective core and its role in animal cognition," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 367, no. 1603, pp. 2704–2714, 2012.
[47] S. Dehaene, C. Sergent, and J.-P. Changeux, "A neuronal network model linking subjective reports and objective physiological data during conscious perception," *Proceedings of the National Academy of Sciences*, vol. 100, no. 14, pp. 8520–8525, 2003.
[48] J. A. Reggia, "The rise of machine consciousness: Studying consciousness with computational models," *Neural Networks*, vol. 44, pp. 112–131, 2013.
[49] R. Gil, E. Arroyo-Anllo, P. Ingrand, M. Gil, J. Neau, C. Ornon, and V. Bonnaud, "Self-consciousness and alzheimer's disease," *Acta Neurologica Scandinavica*, vol. 104, no. 5, pp. 296–300, 2001.
[50] R. Llinas and U. Ribary, "Consciousness and the brain: The thalamocortical dialogue in health and disease," *Annals of the New York Academy of Sciences*, vol. 929, no. 1, pp. 166–175, 2001.
[51] J. T. Giacino, J. J. Fins, S. Laureys, and N. D. Schiff, "Disorders of consciousness after acquired brain injury: the state of the science," *Nature Reviews Neurology*, vol. 10, no. 2, pp. 99–114, 2014.
[52] P. Wang, "On defining artificial intelligence," *Journal of Artificial General Intelligence*, vol. 10, no. 2, pp. 1–37, 2019.