

University of Memphis

University of Memphis Digital Commons

CCRG Papers

Cognitive Computing Research Group

2023

Enabling an autonomous agent sharing its minds, describing its conscious contents

D. Dong

Follow this and additional works at: https://digitalcommons.memphis.edu/ccrg_papers

Recommended Citation

Dong, D. (2023). Enabling an autonomous agent sharing its minds, describing its conscious contents. [10.1016/j.cogsys.2023.03.001](https://digitalcommons.memphis.edu/ccrg_papers/10.1016/j.cogsys.2023.03.001)

This Document is brought to you for free and open access by the Cognitive Computing Research Group at University of Memphis Digital Commons. It has been accepted for inclusion in CCRG Papers by an authorized administrator of University of Memphis Digital Commons. For more information, please contact khggerty@memphis.edu.

Enabling an autonomous agent sharing its minds, describing its conscious contents

Daqi Dong
The University of Memphis
Memphis, TN, USA
daqi.dong@gmail.com

Abstract

We enable an autonomous agent (robot) to share its artificial mind, particularly its conscious content, with its audiences like humans or other agents. This effort is relying on the studies from cognitive modeling and supports the autonomous human robot interaction. LIDA is a biologically inspired systems-level cognitive model that explains and predicts how minds work. This model has been used in designing the controllers of intelligent autonomous agents. Here we propose to add a new description (sub) model into LIDA, allowing its agents to share their control structures, the artificial minds. We argue that the cognitive representations and processes introduced in LIDA may naturally serve as the source of the mind contents an agent can express out. We also proposed a type of description behavior based on LIDA's action modules, which support the interactions between the agent and its real-world environments including human audiences.

Through this description model, an agent's mind will be more visible and accessible, so the agent's intelligence can be assessed from not only its outside appearance and behaviors, but also its mind. In addition, the agent's mind activities may associate with, so help explain certain behaviors it acted/acts. This helps the agent earn more understanding from its audiences so to being accepted by and engage to its living society better. Finally, as an agent's controller, LIDA is designed heavily relying on the hypotheses of human minds, so its agent's artificial mind contents are conceptually easy to understand to its human audiences.

KEYWORDS

Artificial Minds, Cognitive Modeling, Autonomous Agents, BICA, LIDA model, Cognitive HRI

1. Introduction

We humans are recognized by our appearance, behaviors, and maybe most importantly, minds. The minds decide what we do next and form what we look like eventually. However, since our minds are observable only to ourselves, proactively sharing about it to others is an effective way helping other people knowing who we are and understanding us.

We propose to give an autonomous agent (Franklin & Graesser, 1997) the similar ability, sharing its own control-structure, a kind of artificial mind (Franklin, 1995), to others, like humans or other agents. As Franklin and Graesser defined (1997), an autonomous agent is a system situated within an environment where it interacts with the world and communicates with other agents, in pursuit of its own agenda overtime, and affecting what it senses in future. Here the capability of sharing the mind helps the agent pursue its agenda through the communications.

In addition, that enabling the agent to share its mind provides a straight channel to access the agent's control-structure, so to assess from its inside that how well this agent is intelligent and its similarity to human.

As the first step toward this sharing mind modeling work, we present a description cognitive model that supports the agent to describe its conscious contents, grounded upon a systems-level cognitive model LIDA¹ (Franklin et al., 2016).

The LIDA model hypothesizes how minds work. It integrates multiple cognitive modules, and each of which has different cognitive representations and processes. We argue that these cognitive components may naturally serve as the source of the mind content an agent may share out.

We may apply a LIDA model into an agent to control it, as its artificial mind. Borrowed from Global Workspace Theory (GWT) (Baars, 1988, 2002), a LIDA-based agent's conscious content is (1) formed from its understanding

¹ LIDA stands for Learning Intelligent Decision Agent.

of the situation, of both internal and external, (2) chosen as the most salient attention content, and (3) further used in action and learning parts² (Franklin et al., 2016).

Since this conscious content relates to most cognitive modules of the agent's mind, we can infer both what the agent's mind is and how it works from its conscious. Knowing the conscious content helps determine what (who) the agent is from its inside.

From a social perspective, giving an agent the ability to describe its conscious allows it to illustrate what its attentions were/are, from where these attentions may come, and potentially why the agent acted/acts certain associated behaviors. This will help the agent engage its audiences and earn more understanding from them (Chin-Parker & Bradner, 2010; Lombrozo, 2006; Matarese, Rea, & Sciutti, 2021), so being accepted by and adapt to its living society better (Umbrico et al., 2022).

In the next section, we review some related works of systems-level cognitive models in the direction of cognitive human robot interaction (HRI), followed by an overview of LIDA and its modules in Section 3. Then, we introduce the design of the new description model with examples in Section 4, and the initial experiments and some technical evaluations Section 5. Finally, we discuss and conclude about this sharing mind modeling work in Section 6 and list the next steps in Section 7.

2 Systems-level Cognitive Models in Cognitive HRI

From a recent survey of the cognitive HRI, “[it] is a research area that seeks to improve interactions between robots and their users by developing cognitive models for robots and understanding human mental models of robots.” (Mutlu, Roy, & Šabanović, 2016).

There are three core research activities included: (1) human models of interaction where we build understandings of human expectations and cognitive responses to robot actions, (2) robot models of interaction where we develop cognitive representations and actions that allow robots to interact with people, and (3) models of HRI where we build models and platforms that guild and support the interactions between robots and humans (Mutlu et al., 2016).

“[A] systems-level [cognitive] model (cognitive architecture) attempts the full range of activities from incoming stimuli to outgoing actions, together with the full range of cognitive processes in between.” (Franklin et al., 2016). It models not only some separate functions of cognition, but also the relationships between them. The necessity of systems-level cognitive modeling has been argued from different disciplines such as artificial intelligence (Newell, 1973), cognitive modeling (Langley, Laird, & Rogers, 2009), and neuroscience (Bullock, 1993).

ACT-R³ is a cognitive architecture, providing a theory for simulating and understanding human cognition. Relating to HRI, it has been used in modeling human mental processes, for example, building models to understand human's cognition and fallibilities in order to help the robot to be the better teammates (Trafton et al., 2013), and generating promising instructions to robots by modeling a human's decision process and her expectations regarding the robot partner's actions (Lebiere, Jentsch, & Ososky, 2013). In the robot models of interaction, ACT-R was applied in building an autonomous agent (robot) which behaves in the HRI as a more human-like collaborator, providing a more efficient interface to the HRI tasks (Sofge et al., 2004). Also, a story-telling social robot was built to represent the story characters, through the definition of appropriate cognitive models replying on the ACT-R (Bono et al., 2020).

Soar is a cognitive architecture focusing on developing functional capabilities and applying them to tasks such as natural language processing, control of intelligent agents in simulations, virtual humans, and embodied robots (Laird, Lebiere, & Rosenbloom, 2017). It had been regularly used in robot interactive task learning (Laird, Gluck, et al., 2017). In the human models of interaction, it was applied to control a robot helping the human build better mental model of the partner robot (Ramaraj, 2021). In the robot models of interaction, Soar was used to build a robot that leverages a human's natural teaching skills by understanding her teaching intentions in HRI (Ramaraj, Klenk, & Mohan, 2020). Also, an interactive system was built within the Soar which provides the grounding language the agent (robot) may performs during the interactive tasks (Lindes, Mininger, Kirk, & Laird, 2017), where a common model of cognition and humanlike language processing had been introduced as well (Lindes, 2018).

² The LIDA model makes no claims regarding phenomenal consciousness.

³ ACT-R stands for Adaptive Control of Thought—Rational.

A dual process-inspired cognitive architecture has also been proposed for Adaptive HRI (Umbrico et al., 2022), and its authors plan to build an advanced mind-inspired system in next steps as a long-term goal.

LIDA is a systems-level cognitive model. It attempts to model minds, be they human, animal, or artificial (Franklin, 1995), which is taken to be the control structures for autonomous agent (Franklin & Graesser, 1997). In the robot models of interaction, Khayi and Franklin (Khayi & Franklin, 2018) have proposed and implemented a perceptual learning mechanism within LIDA, controlling an autonomous agent (robot) to simulate how an infant vervet monkey learns the meanings of vervet monkey alarm calls.

Here we propose to build a description cognitive model relying on LIDA, to extend an autonomous agent's intelligence by allowing it sharing its artificial minds. This work contributes to build a more humanlike collaborator autonomously expressing its artificial minds with its audiences like humans.

3. LIDA and its Modules

We give an overview about the LIDA model first and then its modules supporting to the new description model.

3.1 Overview

LIDA is a systems-level cognitive model that helps explain and predict the mental phenomena (Franklin et al., 2016). It implements and fleshes out a number of psychological and neuropsychological theories, and is primarily based on Global Workspace Theory (Baars, 1988, 2002).

The model has integrated three phases: perception and understanding, attention, and action and learning. These phases are functioning continually in a cognitive cycle (~10 Hz) and may (partially) overlap among multiple cycles (Madl, Baars, & Franklin, 2011).

In each cycle, the LIDA agent first senses the environment, recognizes objects, and builds its understanding of the current situation. Then by a competitive process, it decides what portion of the represented situation should be mostly attended to as the conscious contents, and be broadcasted to the rest of the system. Finally, these broadcasted conscious supplies information allowing the agent to choose an appropriate action to execute, and modulates learning (Franklin et al., 2016) (Fig. 1).

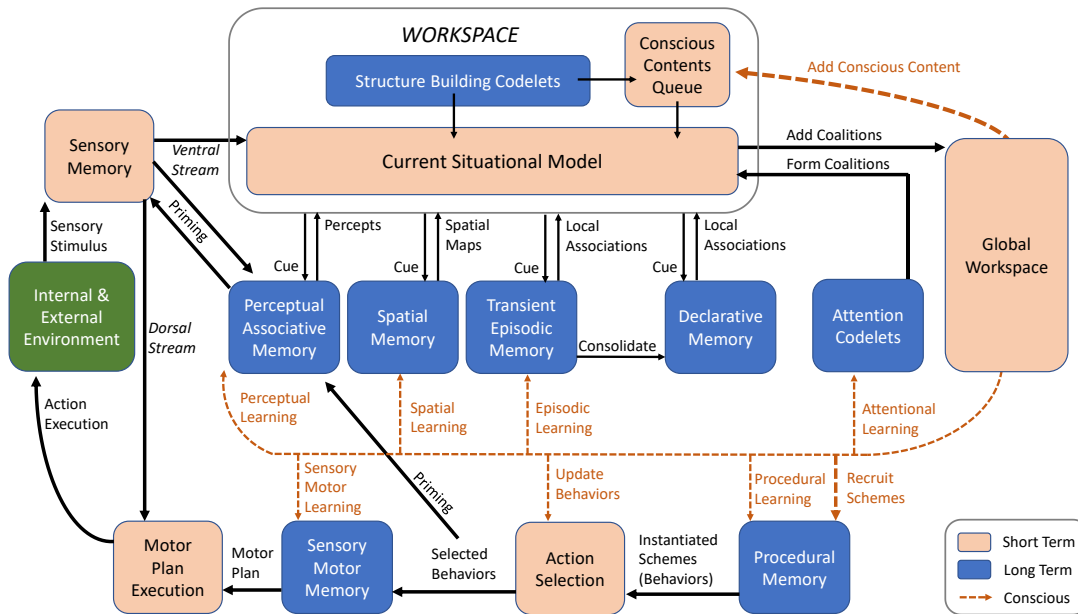


Figure 1: LIDA Cognitive Cycle Diagram (Franklin et al., 2016)

The fundamental data type in each of LIDA modules is “the digraph, consisting of nodes and links. More complex structures are built from these” (Franklin et al., 2016). For example, we may represent objects using nodes and the relationships between them the links. Each of these represented entities has the activation variables attached to it. These activations are used to measure how important, useful, and salient an entity is, from different perspectives, like base level for the past, current level the present, etc.

We may apply this LIDA model as a controller to drive a software robot. The model provides the hypothesized mind activities supplying the intelligence to the robot, driving it making decisions about what to do next and execute it, so make it an intelligential agent.

3.2 Understanding & Attention Modules (Fig. 1)

In LIDA’s understanding phase, its Current Situational Model (CSM) repeatedly taking in both stimuli sensations from Sensory Memory, and more abstract entity perceptions from Perception Associative Memory (PAM). Together with the local associations retrieved from Spatial Memory, Transient Episodic Memory, and Declarative Memory, CSM continually updates itself so keep tracks and represents the LIDA agent’s current situation (Franklin et al., 2016).

Additionally, Structure Building Codelets⁴ (SBC) can create new structures in the CSM. The SBC continually monitor the CSM to build new structures using the contents of interests, so enable the agent to update its recognitions about the current situation, such as new relationships between concepts and objects, new content of category, etc. (Franklin et al., 2016)

In the attention phase, Attention Codelets choose the salient portion of structures from the CSM, form them into new data structures of *coalitions* respectively, and bring them to the Global Workspace (GW) to compete to become the conscious contents. Each attention codelet continually monitor the CSM, looking for the structures based on its own concerns for saliency. In the GW, the most salient attention structure wins from a competition to be the conscious content and be globally broadcasted to almost entire system for the following actions and modulating learnings.

Conscious Contents Queue (CCQ) is a very short-term memory, last about three seconds in humans, which stores the past few tens of conscious contents in an order of first comes first in (Franklin et al., 2016). SBC may take the contents from CCQ to build new structures. For example, a particular causation building codelet may find the Event 1 newly in the CSM and another relevant one, Event 2 recently in CCQ, then it may create a causal link from Event 2 to Event 1 (Snaider, McCall, & Franklin, 2010). Most importantly, CCQ provides the functions grounding the time related concepts.

We proposed the new *description attention codelets* to form and bring part of the represented current situation to the consciousness to be the describing contents. (Details in Section 4.1)

3.3 Action Modules (Fig. 1)

LIDA models what to do and how to do it in its action phase. The Procedural Memory stores some templates, so called the schemes, to respond what to do under a certain scenario to achieve a certain result. A scheme consists of a context, an action, a result component, and the likelihood the action may help achieve the result under the context.

Driven by the arrival conscious content, some schemes are matched, recruited and then instantiated to behaviors. The behaviors are sent to Action Selection module to compete for execution based on the conscious contents and the behaviors’ activations. When a behavior is selected, an expectation codelet is sent to the Attention Codelets module to monitor the CSM for the results of the behavior (Franklin et al., 2016).

The Sensory Motor Memory stores the Motor Plan Templates (MPTs). One MPT is selected among others based on the selected behavior, and then will be specified to a particular motor plan (MP) using both the context of the behavior and the environmental data directly coming from Sensory Memory.

In Motor Plan Execution module, driven by the sensory data coming from the Sensory Memory, the MP generates an order of motor commands applying on the agent’s actuators, as an online running process (Dong & Franklin, 2015). The MP is designed based on the subsumption architecture (Brooks, 1991), a type of reactive motor control mechanism directly linking the sensory data to the selection of motor commands.

⁴ The codelets are some special purpose processes.

The Sensory Memory collects the environmental stimuli, internally or externally, and feed them to the Motor Plan Execution through a dorsal stream⁵ directly (Franklin et al., 2016; Goodale & Milner, 1992; Milner & Goodale, 2008).

We proposed the new *description schemes* and *description motor plan templates* for choosing and executing the right behavior to convey the description. Note that, here an internal perception is necessary and has been proposed too using LIDA's sensation and understanding modules. It helps the agent sense the conscious contents it is describing, a type of internal environmental data, to support the online execution of the description (Dong & Franklin, 2014). (Details in Section 4.2)

4. The Designs of the Description Model with Examples

We apply the LIDA model into an autonomous agent as its controller. The agent senses its environment and acts on it in pursuit of its goal. For example, she may sense the thirsty feeling internally, and luckily a glass of water on the table externally too, then she will choose to reach out the water to consume it.

This LIDA-based agent keeps a *thirsty* node structure and a *glass of water* structure—it contains both the *glass cup* and the *water* nodes, in its Current Situational Model (CSM). One of its Attention Codelets chooses these two structures and forms them into a coalition sending to the Global Workspace (GW) to compete to be the conscious content.

When this coalition wins from the competition and became the conscious content, it recruits the relevant schemes from the Procedural Memory and instantiates them to the behaviors. Through the Action Selection module, one behavior such as the *grasp* is selected, which will be used to choose a Motor Plan Template (MPT) in the Sensory Motor Memory, to be specified to a concrete Motor Plan to generate a sequence of motor commands onto the agent actuators, such as its simulated hands or grippers, to execute the grasp.

When we add the description capability into this agent's controller, a LIDA model, we proposed new LIDA description cognitive representations and processes (1) to form and choose the conscious contents to describe, the describing contents, and (2) to carry the describing behaviors.

4.1 Describing Contents

Similar to having a thirsty node, we build in a *description-need* node into the agent's CSM, representing that the agent is being understanding that it has the description need internally.

We add a new description attention codelet specifically concerning the availability of the description-need node and its relationships with other structures like the thirsty node. This attention codelet may choose the description-need and the thirsty nodes, to form a new coalition to represent the concept of *needing to describe the thirsty (feeling)*.

Based on this description attention codelet, different description type of coalitions may be formed, such as that of needing to describe thirsty, a glass cup, or a glass of water. The coalition activation value depends on the activations of its sub structures like the description-need and thirsty nodes, the attention codelet base-level activation, and others. (Franklin et al., 2016). Within the Global Workspace, one of these description coalitions may win from the competition to be the conscious content.

In detail, we proposed three types of the description coalitions, each of which represents a different type of conscious content an agent may describe:

4.1.1 The regular perceptions

This has just been explained above, such as the coalition of needing to describe thirsty, where the thirsty node is a structure representing a perception of the internal environment.

4.1.2 The being monitored previous action results

In the above example of a thirsty agent grasping a glass of water, when the agent selected the behavior of *grasp* to reach the water, an expectation (attention) codelet will be created in the agent's CSM to monitor the grasping result and form a coalition to bring the result to its consciousness. Here our new description attention codelet will combine

⁵ In the LIDA Model, the concept of ventral and dorsal streams for the transmission of visual information has been extended to multimodal transmission.

forces with this expectation codelet to create a joint coalition, where the description-need node is attached to the action-result coalition, to represent the concept of *needing to describe how well its grasping achieved*.

4.1.3 The near past conscious content events

A Structure Building Codelet (SBC) may monitor the Conscious Contents Queue (CCQ) to build an event specifically relating to the time concept. For example, the thirsty agent might have attended on its *thirsty* feeling a while like couple of seconds, so multiple near past conscious contents listed in its CCQ may contain the thirsty node. A SBC will find these thirsty-node-involved conscious contents from the CCQ and count them to build a new structure in the CSM, to represent the duration of being thirsty in the near past. Again, our new description attention codelet may choose this new *thirsty duration* structure and the description-need node, to form a coalition to represent the concept of *needing to describe how long the thirsty has been available*.

4.2 Describing Behaviors

When a description coalition became the conscious content, that involves the description-need node, some schemes stored in Procedural Memory will be recruited. These schemes have their contexts highly overlapping with the arrival conscious contents such as holding a description-need attribute, and their actions are capable of accomplishing the need of describing, such as in the type of *draw*, *speak*, or *write*. These schemes are instantiated to the corresponding behaviors by plugging in details from the conscious contents, such as that a speak behavior is attached with the context of description-need and the thirsty nodes.

In the Action Selection module, one of these description-capable behaviors is selected depending on their contexts and activations. For example, using an analogy to humans, a young child may choose *draw* to describe the water, *speak* will be a common choice for many adults, and *write* for the specific offline letter communication channel. After one of these behaviors had been selected, an expectation codelet will be created in the CSM, to monitor the action results of this specific type of description.

In Action Execution side, a Motor Plan Template (MPT) represents a specific skill the agent owns to execute an action (behavior in LIDA) such as speaking the thirsty. A MPT is selected based on the selected behavior. Usually this is a one-to-one mapping since an agent may born with only one skill to execute an action, like using only the mother language to speak out the “thirsty”. However, for a bilingual agent, she may have two skills (MPTs) to execute that speaking the thirsty, where one MPT will be selected from others based on its activations and the context contents taken from the behaviors’ contexts.

The selected MPT will be specified to a Motor Plan to run, where the template’s variable values are initiated using the context content. For example, when the agent choses to speak about the thirsty, the context of the duration of being thirsty may help determine the values of a variable modifier, to render the describing object thirsty: longer duration may give the terms like “very much”, shorter the term “a little bit”. Here the variable of modifier is exemplated as human language terms, while broader types of modifiers may be applied, such as the voice tones, etc.

The Motor Plan (MP) constructs a set of agent’s actuators and assigns motor commands to each of them in an order. For example, the MP executes the action of speaking the thirsty by controlling the agent’s voice actuators, kind of Throat and Tongue muscles of humans, in different patterns over time. These patterns are dynamically formed based on (1) the MP’s structure that organizes a set of triggers that will activate in certain conditions to generate motor commands, and (2) the input online environmental data coming from the Sensory Memory that drive the plan’s triggers.

The Sensory Memory collects the environmental stimuli and feed it to the Motor Plan Execution module through the dorsal stream directly. The agent may sense its internal environmental data—for example, its describing contents being maintained in the Current Situational Model—and send it to the running description motor plan to support its online process. When the describing contents changes, like the duration of being thirsty grows, this change will cause the running description motor plan to update its corresponding variable values, strengthening the modifier’s values that indicates the degree of thirsty. This fact forms the dynamic patterns to control the actuators during the description execution online.

If more structures were involved into the description behavior’s context, like involving thirsty, water, and glass of cup structures, a more complex motor plan is necessary to structure so to execute the description. Currently, we propose to build in the motor plan (templates) while the agent may learn more types of it over time (Next steps for the learning in Section 7)

So far, we illustrated how the description model works conceptually with examples, showing that an agent describes out what is its conscious content so giving a kind of snapshot about who it is at a moment. In addition, this description may help give insights to explain the agent's outside behaviors. For example, the agent may describe about its being thirsty followed by an action of grasping a glass of water. It may execute these two actions in parallel (partially), if their actuators are not overlapping. These two joint actions tell a potentially reasonable causal relationship from the agent's inside mind activities to its outside behaviors.

5. Initial Implementations and Technical Evaluations

Here we first give an overview about the LIDA framework that helps build the LIDA-based software agent and support to probe the agent's minds in a developer view. Then we share an initial implementation of an agent autonomously sharing its conscious contents, followed by some technical feasibility evaluations regarding the further implementation and experiments.

5.1 The LIDA Framework and Probing the Agent's Mind

The LIDA Framework is an underlying computational software framework (Snaider et al., 2011), which provides the domain independent modules and processes of LIDA. It supports generic and configurable design principles, patterns, and good practices to help build the LIDA based agents, including its controller and a virtual environment.

The LIDA modules are implemented in the Framework as asynchronously running computational modules, using an *observer* software design pattern. Each module named as the subject, registers and maintains a list of its relevant dependents as its observers, such as other modules or processes. The subject notifies its observers automatically of any of its state changes online.

This framework also provides an experimental tool, where the developers can observe how the agent's controller works in time. It delivers a set of GUIs displaying the status of the agent's internal cognitive modules, such as Perceptual Associative Memory and the Global Workspace. Also, it logs some important cognitive processes such as attentions competition, behavior selection, and consciousness arrivals for learning (Fig. 2).

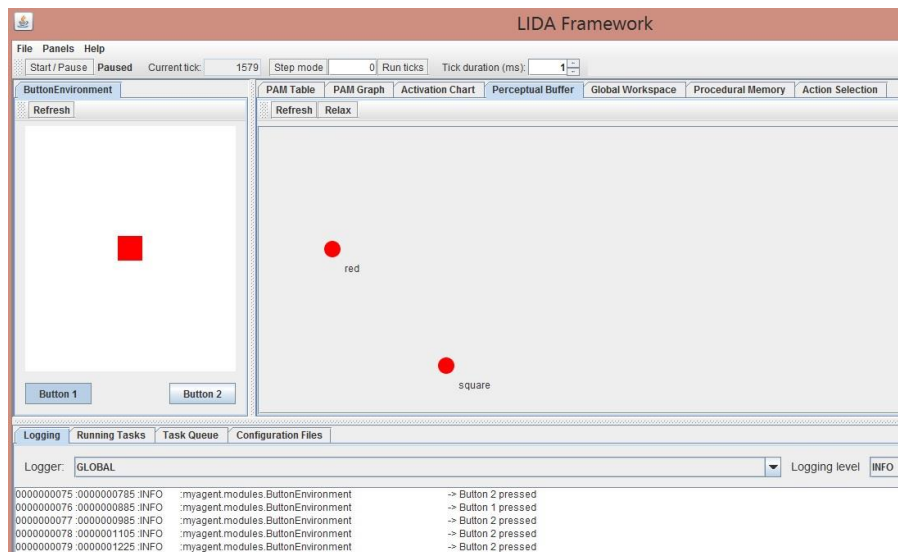


Figure 2: The LIDA Framework experimental tool (Snaider, McCall, & Franklin, 2011)

As the developers of a LIDA-based agent, we can easily probing the agent's artificial mind during its running time using this Framework. For example, we can directly observe the perceptual nodes being in Current Situational Model (CSM), including the description-need node we plan to add; we can track the behaves of each attention codelet in the logs, including the designed description attention codelet, about what their current activations are and

which attention wins from the competition so to form the conscious content; also we can observe and examine the behavior selection process, such as whether a description behavior is selected to execute, if so exactly which type (instance) of description behaviors it is.

5.2 An Agent Autonomously Sharing its Minds

Relying on the LIDA Framework, we borrowed and re-implemented a LIDA-based demo agent as introduced by (Snaider et al., 2011). This agent senses a virtual environment where one of the two objects, blue circle or red square, appears randomly overtime (Fig. 2). The object appearing refreshing interval is set to 1 second as default. Based on its conscious content, the agent will press one of the two labeled buttons to express its responses to the appearing of different objects.

We implemented the new description model, its description contents and behaviors (high-level only), into this agent. We list the implemented functions of each module below. Figure 1 gives an intuitive feel for the relationship of these modules and see Section 3 for more details.

The Environment Module

The environment constantly 1) retrieves the data detected by the robot's sensors, and 2) sends out the motor commands provided by the SMS to the robot's actuators. We added an internal environment module where the description need is built in.

Sensory Memory (SM)

SM gets sensory data from the Environment Module and provides the SMS with the current data. We added the sensing process taking in the agent's inner needs of the description.

Feature detectors (FDs) and Perceptual Associate Memory (PAM)

PAM stores a set of nodes, each of them representing a specific aspect of an environmental state of concern to the agent. These nodes are object nature nodes Red, Blue, Circle, and Square, and the additional inner Description-Need node. FDs constantly obtain the current state from the SM, activating relevant nodes in PAM.

The Current Situational Model (CSM)

The CSM receives currently activated nodes from PAM, and builds the agent's understanding of the current situation. Driven by the internal description needs introduced above, we had a Description-Need node activated in CSM so it is able to be combined into the agent's current understanding (Fig. 3 (a)).

Attention codelets and the Global Workspace (GW)

Besides the color and shape object attention codelets, we also added a description attention codelet concerned for the structures where the description-need node is involved in the CSM (Fig. 3 (b)). These attention codelet form their concerns to coalitions and bringing it into the GW. In the GW, the description coalitions may win the competition to produce the agent's conscious content to describe.

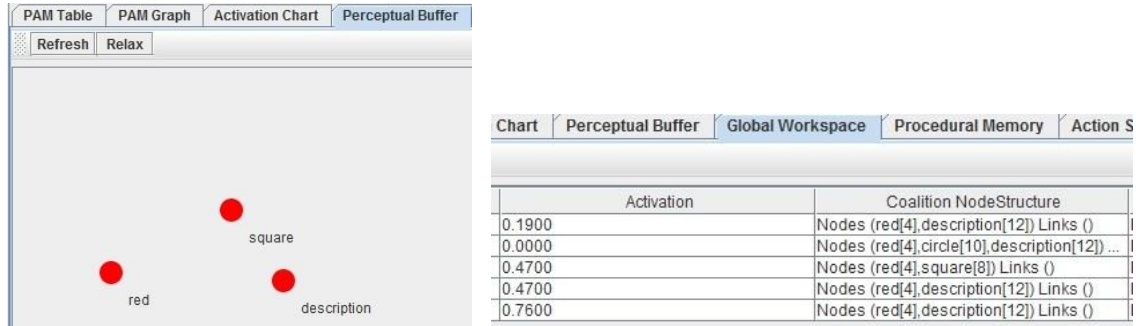
Procedural Memory (PM) and Sensor Motor Memory (in-process)

Following the broadcast of conscious content from the GW, a button pressing scheme is recruited in the PM, and then a relevant behavior (pressing button 1 or 2) is instantiated that is selected by Action Selection module and sent to SMS, which initiates a motor command generation mechanism for executing the button pressing in the SMS. We also added two description schemes: their contexts contain the description-need node, so the schemes are prepared to be recruited by the arrival conscious content holding the description-need attribute; and their actions are specified as the types of *draw* and *speak*, respectively (Fig. 3 (c)).

The detailed description behaviors and motor plan templates are not implemented yet. That selecting between different description behaviors such as draw and speak requires adapting the agent behavior to different contexts. First it is about the agent's "skill level" of acting a behavior, how much the agent is familiar of a behavior. This skill level is represented as the activation of a behavior, which relies on both its scheme that instantiates the behavior, and the arrival conscious contents that providing the detailed information supporting the instantiation. Second, the outside environmental (communication) context may also impact the description behavior selection: for example, whether it is an official talking channel needing a quick responds through speak, or a casual chat situation having a

room allowing an expression through draw. We need to build more extensive description behaviors and communication environment to implement the description behaviors selection next.

Similarly, to implement and examine the description motor plan templates, we need to further build an agent having richer actuators which are capable of executing different patterns of motor commands generated through a description motor plan.



(a)

(b)

Figure 3(c) shows a table with tabs for PAM Table, PAM Graph, Activation Chart, Perceptual Buffer, Global Workspace, Procedural Memory, and Action Selection. It contains a 'Refresh' button and a table with columns for Scheme Label, ID, Current Activation, Base-level Activation, Context, and Action.

Scheme Label	ID	Current Activation	Base-level Activation	Context	Action
if red description, speak	0	0.8000	0.0000	Nodes (red[4],descriptio...	action.speak
if red circle description, ...	1	0.4867	0.0000	Nodes (red[4],circle[10],...	action.draw
if red square, press 1	2	0.3300	0.0000	Nodes (red[4],square[8])...	action.pressOne
if blue circle, press 2	3	0.0000	0.0000	Nodes (blue[6],circle[10]...	action.pressTwo

(c)

Figure 3: (a) Percepts of the Current Situational Model, (b) Global Workspace, and (c) Procedural Memory

6. Conclusions and Discussions

As the first step of enabling an autonomous agent sharing its mind, we proposed a conscious contents description model based on LIDA, outlined the conceptual designs of the new description cognitive processes and representations with examples, followed by the initial implementations and some technical evaluations. If considering this agent a functional assistant tool, then we may use it more precisely by knowing its driver's logic, and if we see it an intelligent entity companying us, then we may cooperate with it easier and please being with it by understanding its mind.

This description capability originates from the subject agent itself without any supervising assistance. The describing contents are supplied from the agent's artificial mind, and the describing behaviors are carried by the agent autonomously. This autonomous feature is inherited from LIDA's modeling hypotheses of minds; the minds both serve as the source to share and control the sharing action.

We build the description model into a systems-level cognitive model so its driving agent may understand, attend to, and act on the descriptions across different necessary cognitive representations and processes with a consistent and mutual-compatible modeling standard. LIDA supports this systems-level necessity by accounting for different cognitive modules and their relationships in one unified architecture.

In addition, because the design of LIDA is biologically inspired, heavily based on the hypotheses of human minds, its agents' artificial minds are potentially easy-to-understand to the human users. This helps the agent engage to and being accepted by its living and working human societies better.

7. Next steps

First we will continue implement the software agent, so it may fully autonomously describe out its conscious content as part of its mind.

Second, we argued that for an autonomous agent, sharing its mind such as its conscious contents may help the agent express who it is and engage more with its audiences. But in the view of agent itself, that how it may take and link these benefits to its description capability so to build the describing motivations, is not yet clarified in detail. A feeling-based motivation system had been studied in LIDA (Franklin et al., 2016; McCall et al., 2020), which provided bridges between LIDA and other existing motivation related concepts. We plan to continue this motivation study on the mind sharing part.

Third, in LIDA, its broadcast conscious is used for learning. Every LIDA memory module updates itself to incorporate appropriate materials from the contents of this conscious (Franklin et al., 2016). Regarding the incorporation part, the description-involved conscious may reinforce (1) the understanding of the description-needs, (2) the attentions that chose the description conscious contents, and (3) the description behaviors and its execution plans. We will address these learning mechanisms more in next.

As the further plans, we plan to study this mind sharing capabilities upon among different cognitive architectures and intelligent systems such as ACT-R, Soar, etc., to build more general intelligent agents sharing their minds.

ACKNOWLEDGMENTS

The authors would like to truly thank Stan Franklin for his ideas, comments, and questions giving to this paper during the initial idea's discussions and the earlier draft buildings. Specific thanks also go to Steve Strain who gave helpful suggestions to the earlier draft too.

References

- Baars, B. J. (1988). *A cognitive theory of consciousness*. New York: Cambridge University Press.
- Baars, B. J. (2002). The conscious access hypothesis: origins and recent evidence. *Trends in cognitive sciences*, 6(1), 47-52.
- Bono, A., Augello, A., Pilato, G., Vella, F., & Gaglio, S. (2020). An ACT-R based humanoid social robot to manage storytelling activities. *Robotics*, 9(2), 25.
- Brooks, R. A. (1991). How to build complete creatures rather than isolated cognitive simulators. *Architectures for Intelligence: The Twenty-second Carnegie Mellon Symposium on Cognition*, 225-239.
- Bullock, T. H. (1993). Goals and Strategies in Brain Research: The Place of Comparative Neurology *How do Brains Work?* (pp. 1-8): Springer.
- Chin-Parker, S., & Bradner, A. (2010). Background shifts affect explanatory style: How a pragmatic theory of explanation accounts for background effects in the generation of explanations. *Cognitive processing*, 11(3), 227-249.
- Dong, D., & Franklin, S. (2014). *Sensory Motor System: Modeling the process of action execution*. Paper presented at the Proceedings of the 36th Annual Conference of the Cognitive Science Society (2145-2150). Quebec, Canada.
- Dong, D., & Franklin, S. (2015). A New Action Execution Module for the Learning Intelligent Distribution Agent (LIDA): The Sensory Motor System. *Cognitive Computation*, 1-17. doi: 10.1007/s12559-015-9322-3
- Franklin, S. (1995). *Artificial Minds*. Cambridge, MA: MIT Press.
- Franklin, S., & Graesser, A. (1997). Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents *Intelligent agents III agent theories, architectures, and languages* (pp. 21-35). London, UK: Springer-Verlag.

- Franklin, S., Madl, T., Strain, S., Faghihi, U., Dong, D., Kugele, S., Snaider, J., Agrawal, P., & Chen, S. (2016). A LIDA cognitive model tutorial. *Biologically Inspired Cognitive Architectures*, 105-130. doi: 10.1016/j.bica.2016.04.003
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in neurosciences*, 15(1), 20-25.
- Khayl, N. A., & Franklin, S. (2018). Initiating language in LIDA: learning the meaning of vervet alarm calls. *Biologically Inspired Cognitive Architectures*, 23, 7-18.
- Laird, J. E., Gluck, K., Anderson, J., Forbus, K. D., Jenkins, O. C., Lebiere, C., Salvucci, D., Scheutz, M., Thomaz, A., & Trafton, G. (2017). Interactive task learning. *IEEE Intelligent Systems*, 32(4), 6-21.
- Laird, J. E., Lebiere, C., & Rosenbloom, P. S. (2017). A standard model of the mind: Toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. *Ai Magazine*, 38(4), 13-26.
- Langley, P., Laird, J. E., & Rogers, S. (2009). Cognitive architectures: Research issues and challenges. *Cognitive Systems Research*, 10(2), 141-160.
- Lebiere, C., Jentsch, F., & Ososky, S. (2013). *Cognitive models of decision making processes for human-robot interaction*. Paper presented at the International Conference on Virtual, Augmented and Mixed Reality (285-294).
- Lindes, P. (2018). The Common Model of Cognition and humanlike language comprehension. *Procedia Computer Science*, 145, 765-772.
- Lindes, P., Mininger, A., Kirk, J. R., & Laird, J. E. (2017). *Grounding language for interactive task learning*. Paper presented at the Proceedings of the First Workshop on Language Grounding for Robotics (1-9).
- Lombrozo, T. (2006). The structure and function of explanations. *Trends in cognitive sciences*, 10(10), 464-470.
- Madl, T., Baars, B. J., & Franklin, S. (2011). The timing of the cognitive cycle. *PloS one*, 6(4), e14803.
- Matarese, M., Rea, F., & Sciutti, A. (2021). A user-centred framework for explainable artificial intelligence in human-robot interaction. *arXiv preprint arXiv:2109.12912*.
- McCall, R. J., Franklin, S., Faghihi, U., Snaider, J., & Kugele, S. (2020). Artificial motivation for cognitive software agents. *Journal of Artificial General Intelligence*, 11(1), 38-69.
- Milner, D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, 46(3), 774-785.
- Mutlu, B., Roy, N., & Šabanović, S. (2016). Cognitive human–robot interaction. *Springer handbook of robotics*, 1907-1934.
- Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium.
- Ramaraj, P. (2021). *Robots that Help Humans Build Better Mental Models of Robots*. Paper presented at the Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (595-597).
- Ramaraj, P., Klenk, M., & Mohan, S. (2020). *Understanding intentions in human teaching to design interactive task learning robots*. Paper presented at the RSS 2020 Workshop: AI & Its Alternatives in Assistive & Collaborative Robotics: Decoding Intent.
- Snaider, J., McCall, R., & Franklin, S. (2010). *The immediate present train model time production and representation for cognitive agents*. Paper presented at the 2010 AAI Spring Symposium Series.
- Snaider, J., McCall, R., & Franklin, S. (2011). The LIDA framework as a general tool for AGI *Artificial General Intelligence* (pp. 133-142). Berlin Heidelberg: Springer
- Sofge, D., Trafton, J. G., Cassimatis, N., Perzanowski, D., Bugajska, M., Adams, W., & Schultz, A. (2004). *Human-robot collaboration and cognition with an autonomous mobile robot*. Paper presented at the In Proceedings of the 8th Conference on Intelligent Autonomous Systems (IAS-8) (80-87).
- Trafton, J. G., Hiatt, L. M., Harrison, A. M., Tamborello, F. P., Khemlani, S. S., & Schultz, A. C. (2013). Act-r/e: An embodied cognitive architecture for human-robot interaction. *Journal of Human-Robot Interaction*, 2(1), 30-55.
- Umbrico, A., De Benedictis, R., Fracasso, F., Cesta, A., Orlandini, A., & Cortellessa, G. (2022). A Mind-inspired Architecture for Adaptive HRI. *International Journal of Social Robotics*, 1-21.