Artificial Intelligence Reading Club

Chapter 07

Ruoding Wang

411624184@qq.com

2020.11

Chapter 07

The LIDA Model as a Foundational Architecture for AGI

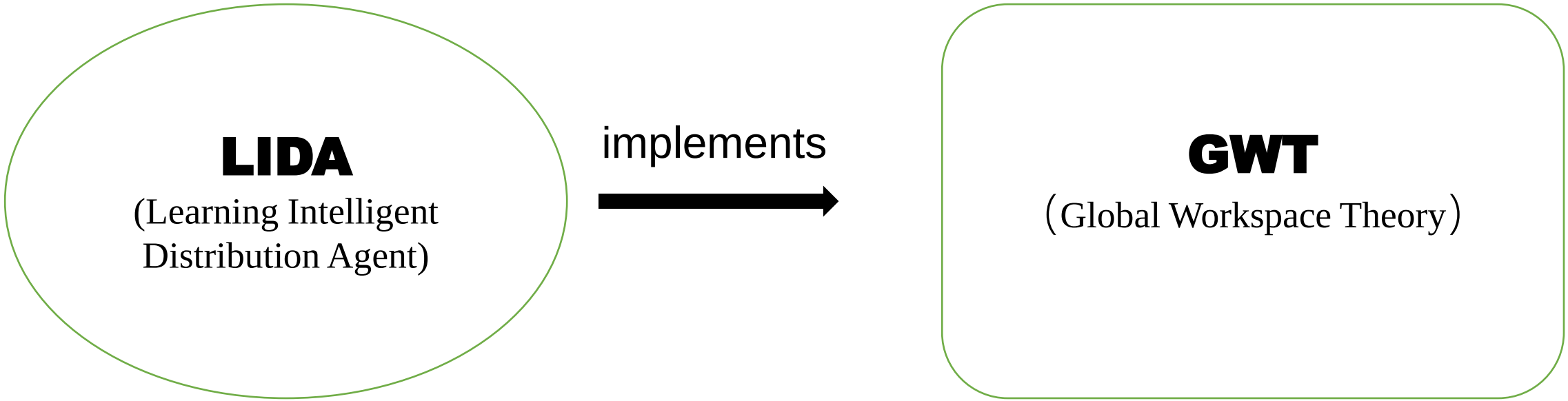
1. LIDA architecture

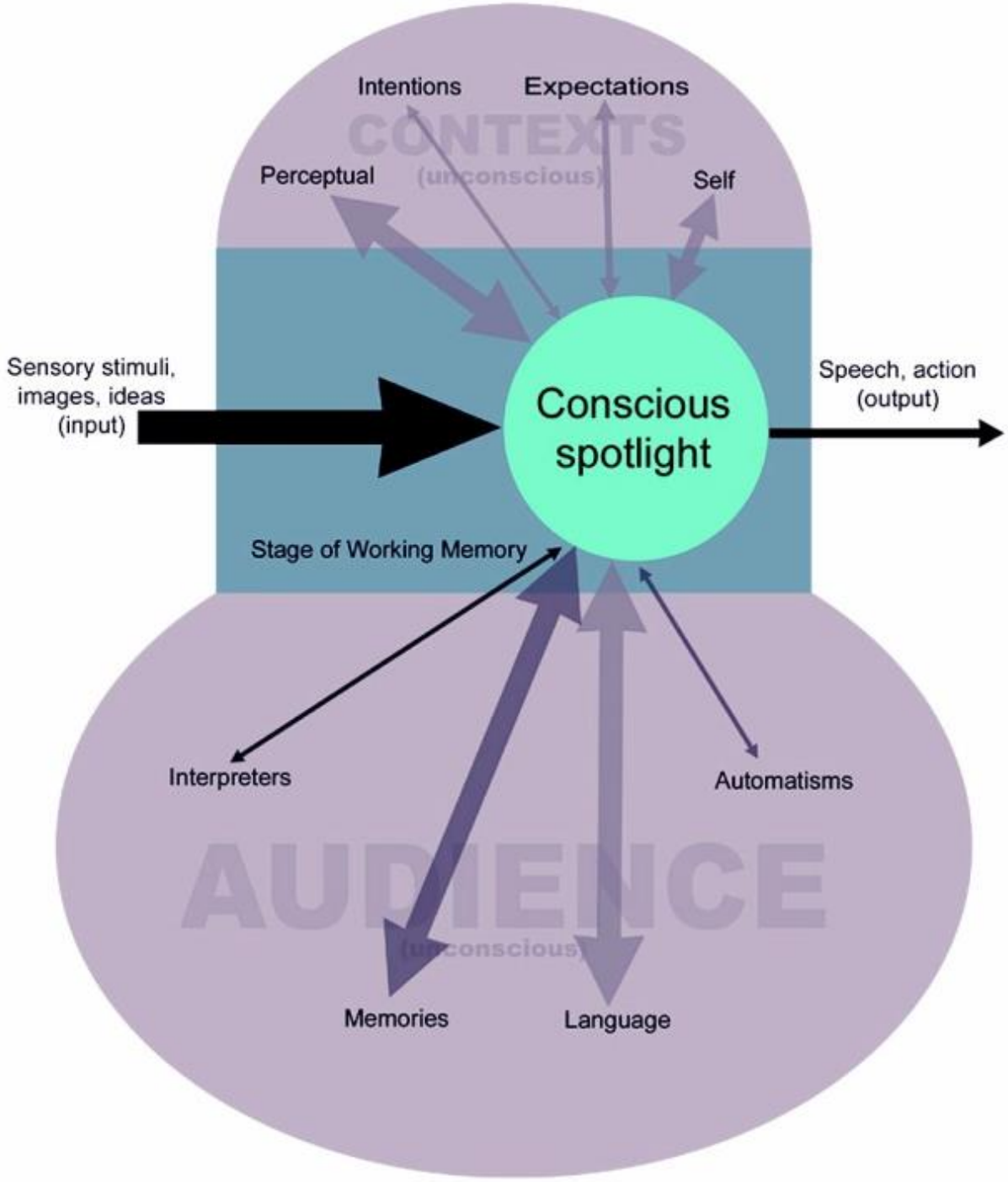
2. Cognitive architectures and the LIDA model

1. LIDA architecture

1. GWT（Global Workspace Theory）
2. Why the LIDA model may be suitable for AGI
3. LIDA’s primary mechanisms

1. LIDA architecture




Intentions
Expectations
CONTEXTS
(unconscious)
Perceptual
Self
Sensory stimuli,
images, ideas
(input)
Conscious
spotlight
Speech, action
(output)
Stage of Working Memory
Interpreters
Automatisms
AUDIENCE
(unconscious)
Memories
Language

1. LIDA architecture

Why the LIDA model may be suitable for AGI

The LIDA model of cognition is a fully integrated artificial cognitive system capable of reaching across a broad spectrum of cognition, from low-level perception/action to high-level reasoning.

(1) science side

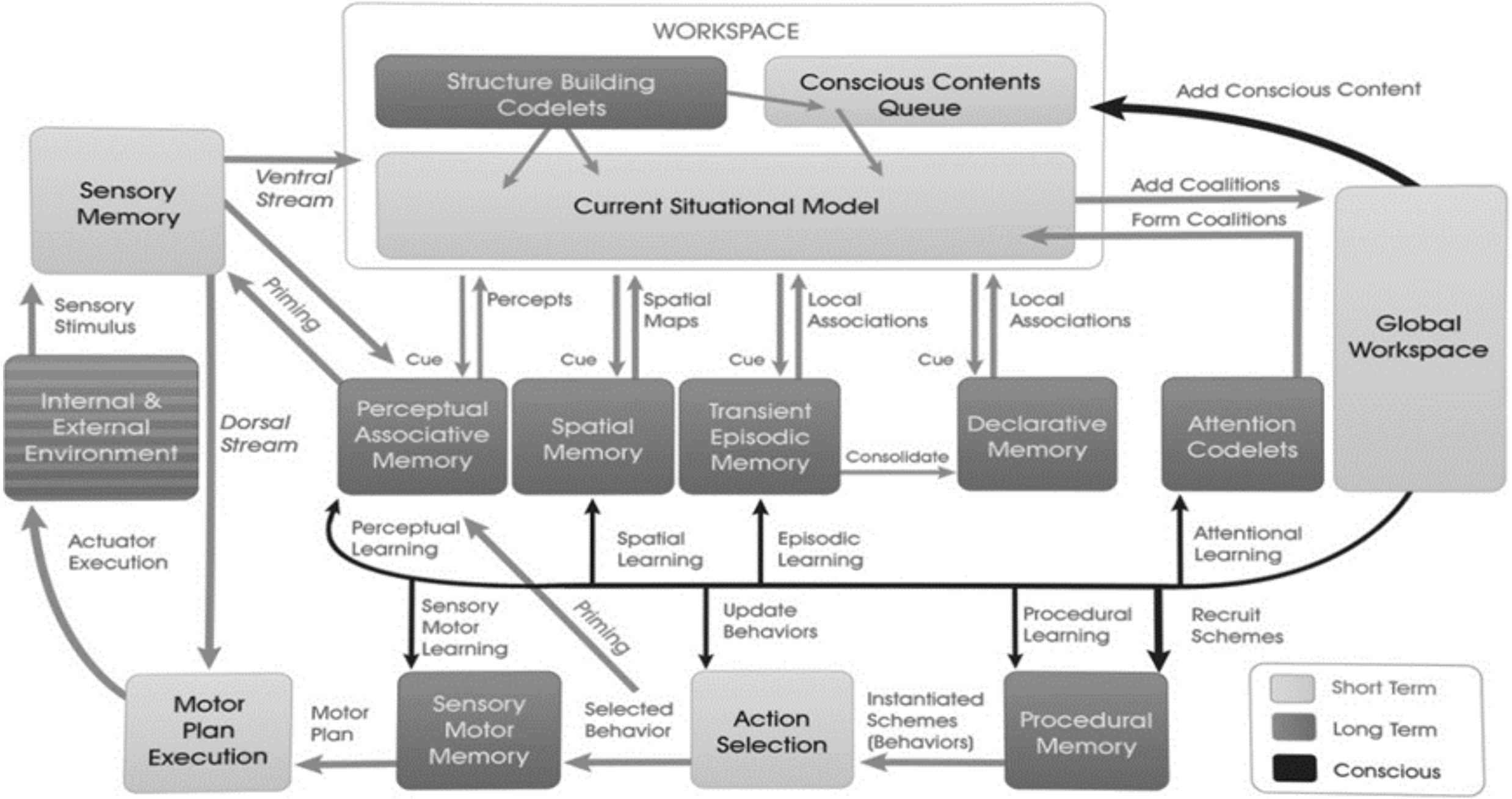
Many psychological and neuropsychological theories： GWT, situated cognition, perceptual symbol systems, working memory, memory by affordances, long-term working memory, and the H-CogAff architecture.

(2) engineering side

Modules: variants of the Copycat Architecture, Sparse Distributed Memory, the Schema Mechanism, the Behavior Net, and the Subsumption Architecture.

(3) Integration and flexibility

LIDA architecture can accommodate the myriad features that will undoubtedly be required of an AGI.
LIDA architecture offers the flexibility to relatively easily experiment with different paths to an AGI.


WORKSPACE
Structure Building Codelets
Conscious Contents Queue
Current Situational Model
Add Conscious Content
Add Coalitions
Form Coalitions
Sensory Memory
Ventral Stream
Priming
Sensory Stimulus
Internal & External Environment
Dorsal Stream
Percepts
Spatial Maps
Local Associations
Local Associations
Cue
Cue
Cue
Cue
Perceptual Associative Memory
Spatial Memory
Transient Episodic Memory
Consolidate
Declarative Memory
Attention Codelets
Global Workspace
Perceptual Learning
Spatial Learning
Episodic Learning
Attentional Learning
Actuator Execution
Sensory Motor Learning
Priming
Update Behaviors
Procedural Learning
Recruit Schemes
Motor Plan Execution
Motor Plan
Sensory Motor Memory
Selected Behavior
Action Selection
Instantiated Schemes (Behaviors)
Procedural Memory
Short Term
Long Term
Conscious

2. Cognitive architectures and the LIDA model

1. Ron Sun’s Desiderata
2. Newell’s functional criteria
3. BICA table

（1）Ron Sun’s Desiderata

Ecological realism

everyday activities of cognitive agents in natural ecological environments.

Bio-evolutionary realism supplements ecological realism

intelligences in species are on a continuum, cognitive models should be reducible to models of animal intelligence.

Cognitive realism

cognitive architectures should seek to replicate only essential characteristics of human cognition.

Eclecticism of methodologies and techniques

It is best to take a more broad-based approach----science and engineering

（1）Ron Sun’s Desiderata

Reactivity

fixed responses to given stimuli that characterize many human behaviors.

Sequentiality

the chronological nature of human everyday activities.

Routineness

humans’ every day behaviors are made of routines, which are constantly and smoothly adapting to the changing environment.

Trial-and-error adaptation

the trial-and-error process through which humans learn and develop reactive routines.

（2）Newell’s functional criteria(12 questions)

Flexible behavior

Does the architecture behave as an (almost) arbitrary function of the environment?
Is the architecture computationally universal with failure?

Real-time operation

Does the architecture operate in real time?
Given its timing assumptions, can it respond as fast as humans?

Rationality

Does the architecture exhibit rational, i.e., effective adaptive behavior?
Does the system yield functional behavior in the real world?

（2）Newell’s functional criteria

Knowledgeable in terms of size

Can it use vast amounts of knowledge about the environment?
How does the size of the knowledge base affect performance?

Knowledgeable in terms of variety

Does the agent integrate diverse knowledge?
Is it capable of common examples of intellectual combination?

Behaviorally robust

Does the agent behave robustly in the face of error, the unexpected, and the unknown?
Can it produce cognitive agents that successfully inhabit dynamic environments?

（2）Newell’s functional criteria

Linguistic

Does the agent use (natural) language? Is it ready to take a test of language proficiency?

Self-awareness

Does the agent exhibit self-awareness and a sense of self?
Can it produce functional accounts of phenomena that reflect consciousness?

1) ***The Proto-Self:*** *In LIDA, the Proto-self is implemented as the set of global and relevant parameters in the various modules.*

2) the Minimal (Core) Self: *The Minimal Self can be implemented as computational collections of nodes in the slipnet of LIDA’s perceptual associative memory.*

3) the Extended Self: (a) *the autobiographical self,* (b) *the self-concept,* (c) *the volitional or executive self, and* (d) *the narrative self.*

（2）Newell’s functional criteria

Adaptive through learning

Does the agent learn from its environment? Can it produce the variety of human learning?

Developmental

Does the agent acquire capabilities through development? Can it account for developmental phenomena?

Evolvable

Can the agent arise through evolution?
Does the theory relate to evolutionary and comparative considerations?

Be realizable within the brain

Do the components of the theory exhaustively map onto brain processes?

（3）BICA table

Support for Common Components

Support: *all features such as episodic and semantic memories…*
Not support: *the auditory mechanism.*

Support for Common Learning Algorithms

Support: *episodic, perceptual, procedural, and attentional learning…*
Not support: *the Bayesian Update and Gradient Descent Methods* (*e.g., Backpropagation*).

Common General Paradigms Modeled

Support: *decision making and problem solving …*
Not support: *perceptual illusions, meta-cognitive tasks, social psychology tasks, personality psychology tasks, motivational dynamics*

（3）BICA table

Common Specific Paradigms Modeled columns

Not support:

1） Stroop;
2) Task Switching;
3) Tower of Hanoi/London;
4) Dual Task;
5) *N*-Back;
6) Visual perception with comprehension;
7) Spatial exploration;
8) Learning and navigation;
9) Object/feature search in an environment;
10) Learning from instructions;
11) Pretend-play.

But

In principle the LIDA can implement *each of them*.

(3) BICA table

Meta-Theoretical Questions

1) Uses only local computations? *Yes;*
2) Unsupervised learning? *Yes;*
3) Supervised learning? *Yes in principle;*
4) Can it learn in real time? *Yes;*
5) Can it do fast stable learning; i.e., adaptive weights converge on each trial without forcing catastrophic forgetting? *Yes;*
6) Can it function autonomously? *Yes;*
7) Is it general-purpose in its modality; i.e., is it brittle? *Yes in principle;*
8) Can it learn from arbitrarily large databases; i.e., not toy problems? *Yes;*
9) Can it learn about non-stationary databases; i.e., environmental rules change unpredictably? *Yes in principle;*
10) Can it pay attention to valued goals? *Yes;*
11) Can it flexibly switch attention between unexpected challenges and valued goals? *Yes;*
12) Can reinforcement learning and motivation modulate perceptual and cognitive decision making? *Yes;*
13) Can it adaptively fuse information from multiple types of sensors and modalities? *Yes in principle.*

Thank you for your time!