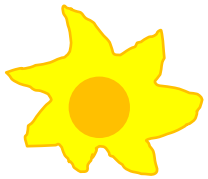
An autonomous agent sharing its mind, its conscious contents

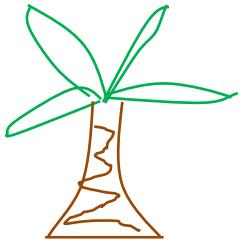
AAAI 2023, AI-HRI Symposium

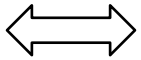
Daqi Dong

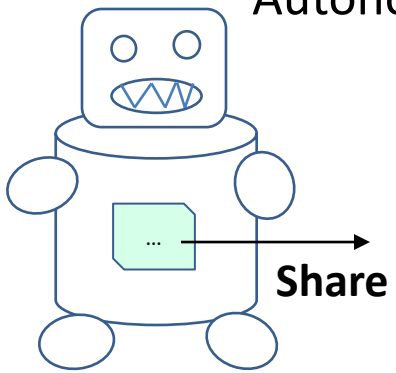
Oct. 25th, 2023

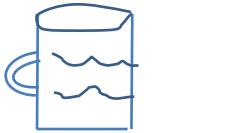
Knowing the agent

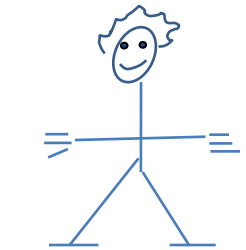








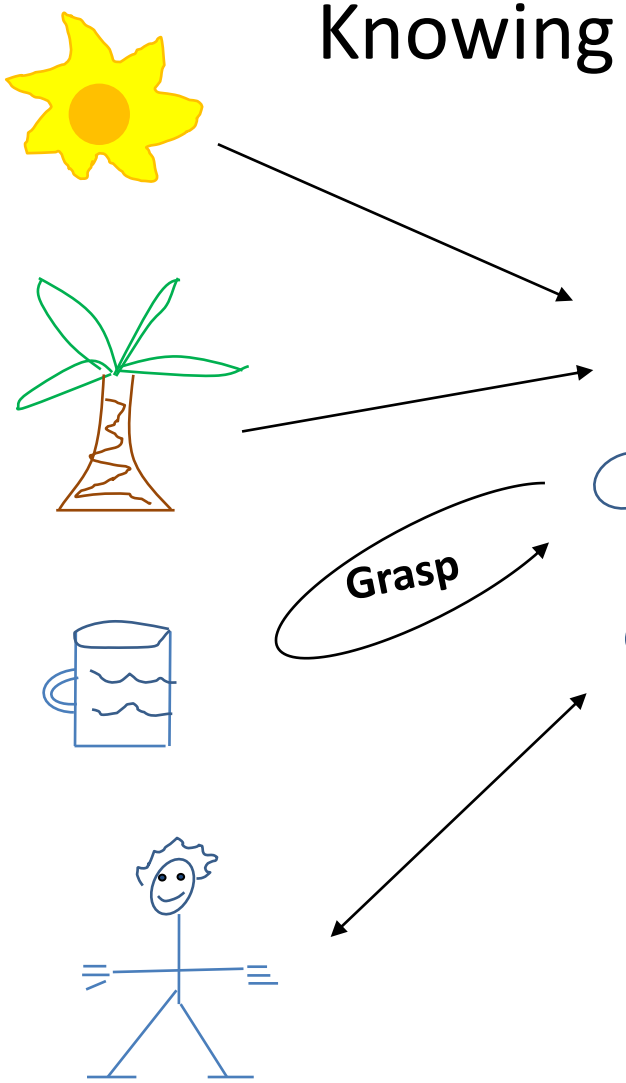




* Mind is a control structure for the autonomous agent [2]

[1] 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.
[2] Franklin, S. (1995). *Artificial Minds.* Cambridge, MA: MIT Press.

Knowing the agent's behaviors



** Conscious contents are the most salient attending contents the agent has
-- Global Workspace Theory [3, 4]

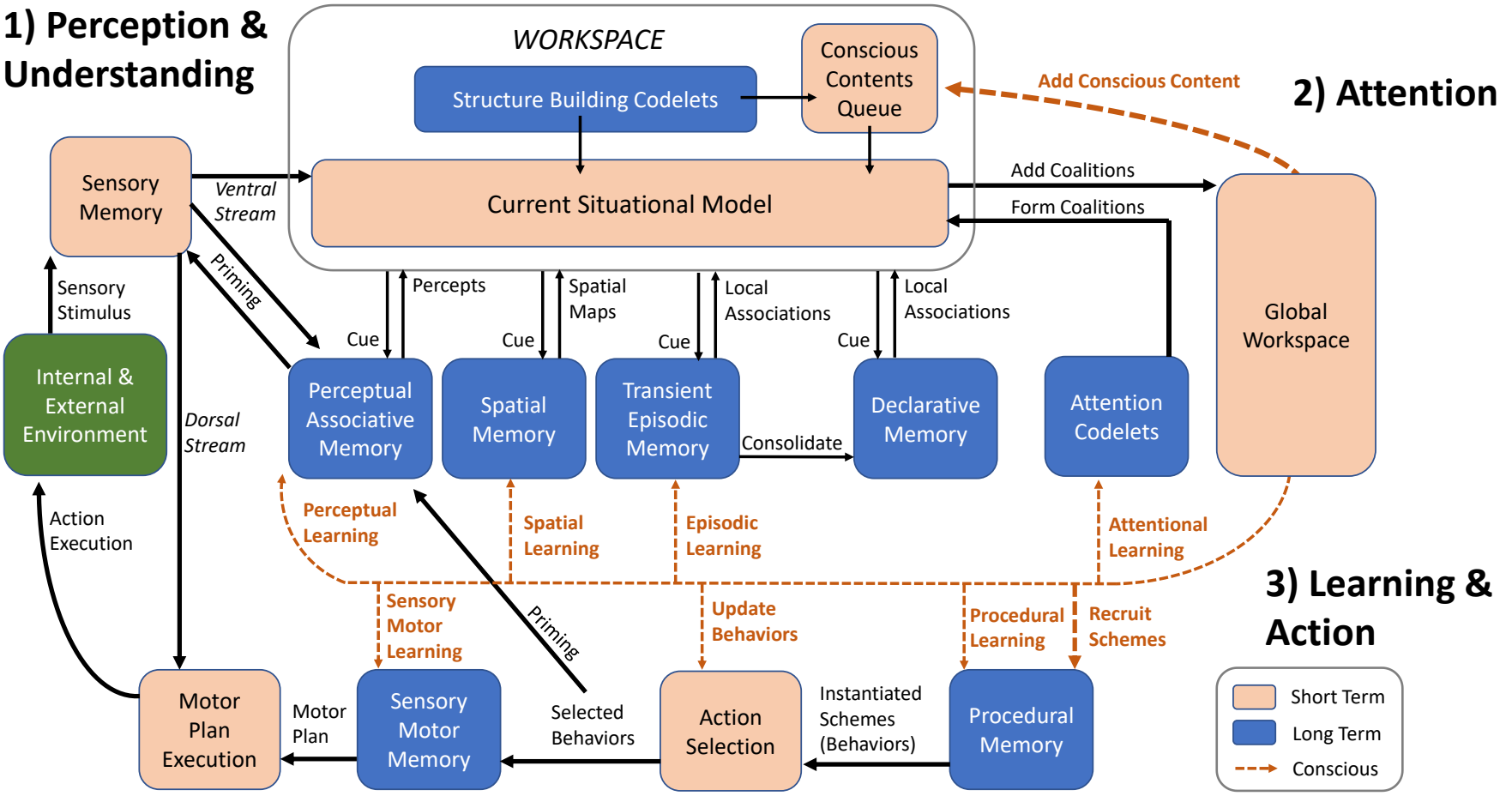
[3] Baars, B. J. (1988). *A cognitive theory of consciousness.* New York: Cambridge University Press.
[4] Baars, B. J. (2002). The conscious access hypothesis: origins and recent evidence. *Trends in cognitive sciences*, 6(1), 47-52.

Controlling the agent using a cognitive architecture, LIDA [5]

- A systems-level cognitive model, LIDA
  - How minds work and its output
- Added the sharing mind components
  - Motivation & Perception
  - Attention
  - Behaviors

[5] 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

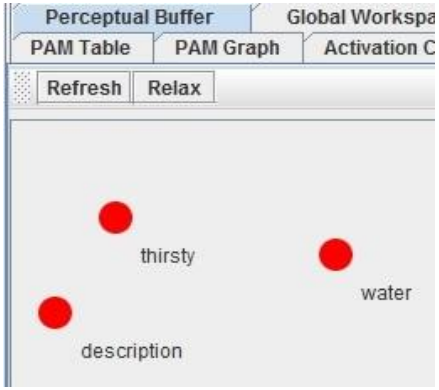
LIDA architecture diagram [5]

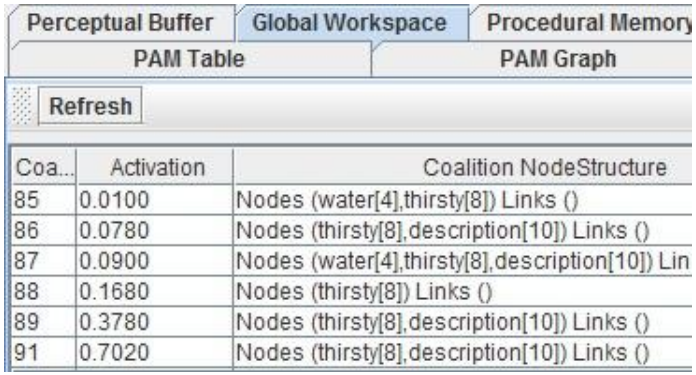


[5] Franklin, S., Madl, T., Strain, S., Faghihi, U., Dong, D., Kugele, S., Snaider, J., Agrawal, P., & Chen, S. (2016). A LIDA cognitive model tutorial. *Biologically Inspired Cognitive Architectures*, 105-130. doi: 10.1016/j.bica.2016.04.003

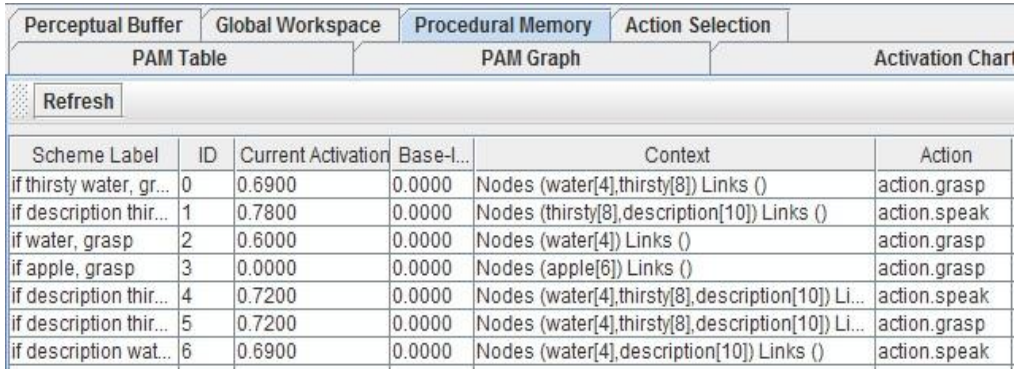
An initial sharing-mind agent

- LIDA Computational Framework [6]





Coa...	Activation	Coalition NodeStructure
85	0.0100	Nodes (water[4],thirsty[8]) Links ()
86	0.0780	Nodes (thirsty[8],description[10]) Links ()
87	0.0900	Nodes (water[4],thirsty[8],description[10]) Lin
88	0.1680	Nodes (thirsty[8]) Links ()
89	0.3780	Nodes (thirsty[8],description[10]) Links ()
91	0.7020	Nodes (thirsty[8],description[10]) Links ()



Scheme Label	ID	Current Activation	Base-l...	Context	Action
if thirsty water, gr...	0	0.6900	0.0000	Nodes (water[4],thirsty[8]) Links ()	action.grasp
if description thir...	1	0.7800	0.0000	Nodes (thirsty[8],description[10]) Links ()	action.speak
if water, grasp	2	0.6000	0.0000	Nodes (water[4]) Links ()	action.grasp
if apple, grasp	3	0.0000	0.0000	Nodes (apple[6]) Links ()	action.grasp
if description thir...	4	0.7200	0.0000	Nodes (water[4],thirsty[8],description[10]) Li...	action.speak
if description thir...	5	0.7200	0.0000	Nodes (water[4],thirsty[8],description[10]) Li...	action.grasp
if description wat...	6	0.6900	0.0000	Nodes (water[4],description[10]) Links ()	action.speak

[6] 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

Next steps and Q&A

- Communication context knowledge
- The motivation [7] of sharing the mind

Contacts:
daqi.dong@gmail.com

Cognitive Computing Research Group
https://ccrg.cs.memphis.edu/

[7] McCall, R., Franklin, S., Faghihi, U., Snaider, J., & Kugele, S. (2020). Artificial Motivation for Cognitive Software Agents. Journal of Artificial General Intelligence, 11(1), 38-69. doi: https://doi.org/10.2478/jagi-2020-0002