

LIDA and a Theory of Mind

By David Friedlander
Cognitive Computing Research Group
Houston, TX

and Stan Franklin
Computer Science Department
& Institute for Intelligent Systems
The University of Memphis

0. Abstract

Every agent aspiring to human level intelligence, every AGI agent, must be capable of a theory of mind. That is, it must be able to attribute mental states, including intentions, to other agents, and must use such attributions in its action selection process. The LIDA conceptual and computational model of cognition offers an explanation of how theory of mind is accomplished in humans and some other animals, and suggests how this explanation could be implemented computationally. Here we describe how the LIDA version of theory of mind is accomplished, and illustrate it with an example taken from an experiment with monkeys, chosen for simplicity.

1. Introduction

The theory of mind states that we ascribe minds to other individuals, and particularly attribute mental states to them even though, as individuals, we only have direct evidence of our own mental states (Premack, D. G. and G. Woodruff. 1978). We perceive the minds of others in the same way we perceive other aspects of our environment, by applying cognitive processes to sensory inputs. We use the results of such perception to select actions, and to learn from our perceptions of their effects. One can argue that a theory of mind process would be necessary for an artificial general intelligence (AGI) agent.

The central hypothesis of this paper is that the mind has the ability to build models of other cognitive agents in hypothetical environments and to reason about potential outcomes of possible actions as an aid to decision making. One possible objection is that, not only are our perceptions about other people's minds indirect, they are also sparse. However, the human mind is quite good at making sense of sparse data. There is a good analogy with visual perception. In vision, the fovea scans only a small part of the scene (Osterberg 1935). No detail is available for most of it and, for some of it, no data at all. Yet we perceive a continuous, fully filled in, detailed scene at all times. The brain supplies the missing perceptual qualities from a number of sources (Tse 2003): previous input from the recent past; world knowledge such as common sense or convention; abstract default information; similar information from the past. We propose that similar sources are used to make sense of sparse data while modeling other minds.

In this paper, we will use a cognitive model called LIDA (Franklin 2007), derived from Global Workspace Theory (Baars 1988, 1997) and based on recent theories from psychology, neuroscience and cognitive science, to develop our hypothesis of the theory of mind and to show how it can explain various psychological phenomena. Section 2 contains a brief description of the LIDA model, Section 3 shows how it can incorporate a theory of mind, Section 4 provides a specific example of how the model would work, and Section 5 contains conclusions and suggestions for future research.

2. LIDA

The LIDA model and its ensuing architecture are grounded in the LIDA cognitive cycle. As a matter of principle, every autonomous agent (Franklin and Graesser 1997), be it human, animal, or artificial, must frequently sample (sense) its environment, process (make sense of) this input, and select an appropriate response (action). The agent’s “life” can be viewed as consisting of a continual sequence of iterations of these cognitive cycles. Such cycles constitute the indivisible elements of attention, the least sensing and acting to which we can attend. A cognitive cycle can be thought of as a moment of cognition, a cognitive “moment.” Higher-level cognitive processes are composed of these cognitive cycles as cognitive “atoms.”

Just as atoms are composed of protons, neutrons and electrons, and some of these are composed of quarks, gluons, etc., these cognitive “atoms” have a rich inner structure. We’ll next concisely describe what the LIDA model hypothesizes as the rich inner structure of the LIDA cognitive cycle. More detailed descriptions are available elsewhere (Baars and Franklin 2003, Franklin et al 2005, Franklin and Patterson 2006, Ramamurthy et al 2006). Figure 1 should help the reader follow the description. It starts in the upper left corner and proceeds clockwise.

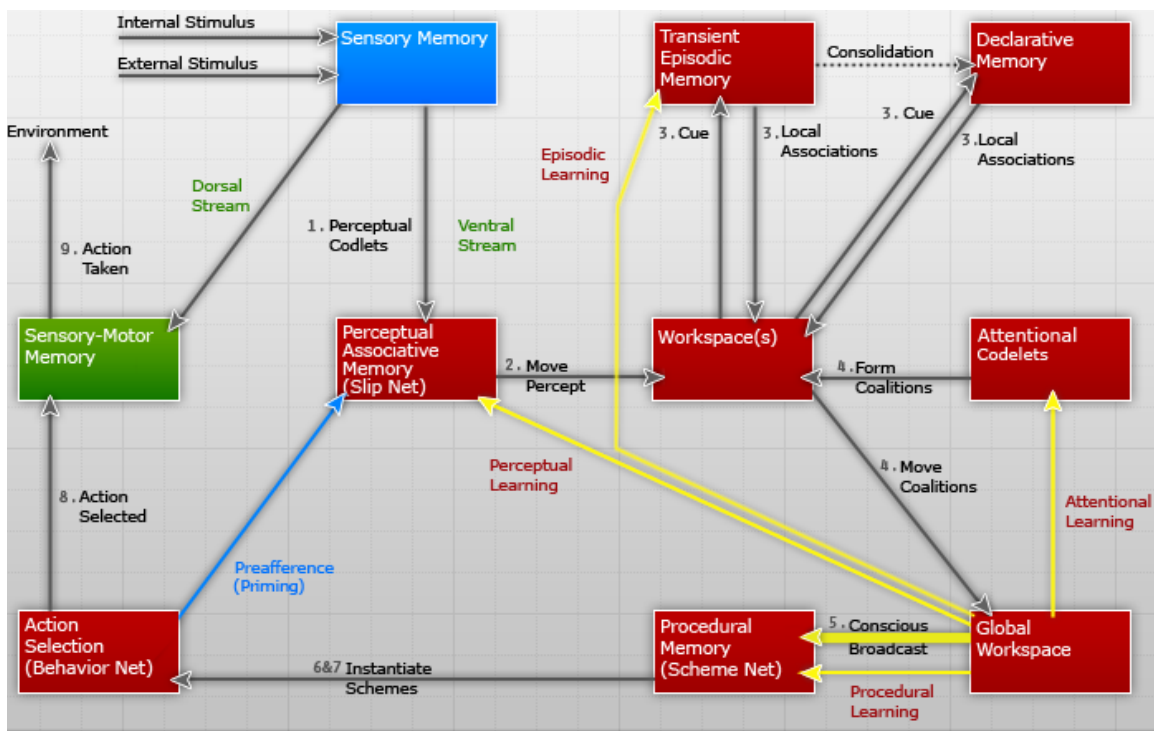


Figure 1. The LIDA Cognitive Cycle

During each cognitive cycle the LIDA agent first makes sense of its current situation as best as it can. It then decides what portion of this situation is most in need of attention. Broadcasting this portion, the current contents of consciousness, enables the agent to finally chose an appropriate action and execute it. Let’s look at these three processes in a little more detail.

The cycle begins with sensory stimuli from the agent's environment, both an external and an internal environment. Low-level feature detectors in sensory memory begin the process of making sense of the incoming stimuli. These low-level features are passed to perceptual memory where higher-level features, objects, categories, relations, actions, situations, etc. are recognized. These recognized entities, comprising the percept, are passed to the workspace, where a model of the agent's current situation is assembled. Workspace structures serve as cues to the two forms of episodic memory, yielding both short and long term remembered local associations. In addition to the current percept, the workspace contains recent percepts that haven't yet decayed away, and the agent's model of the then current situation previously assembled from them. The model of the agent's current situation is updated from the previous model using the remaining percepts and associations. This updating process will typically require looking back to perceptual memory and even to sensory memory, to enable the understanding of relations and situations. This assembled new model constitutes the agent's understanding of its current situation within its world. It has made sense of the incoming stimuli.

For an agent "living" in a complex, dynamically changing environment, this updated model may well be much too much for the agent to deal with all at once. It needs to decide what portion of the model should be attended to. Which are the most relevant, important, urgent or insistent structures within the model? Portions of the model compete for attention. These competing portions take the form of coalitions of structures from the model. One such coalition wins the competition. The agent has decided on what to attend.

But, the purpose of all this processing is to help the agent to decide what to do next (Franklin 1995, chapter 16). To this end, the winning coalition passes to the global workspace, the namesake of Global Workspace Theory (Baars 1988, 1997), from which it is broadcast globally. Though the contents of this conscious broadcast are available globally, the primary recipient is procedural memory, which stores templates of possible actions including their contexts and possible results (D'Mello et al 2006). It also stores an activation value for each such template that attempts to measure the likelihood of an action taken within its context producing the expected result. Templates whose contexts intersect sufficiently with the contents of the conscious broadcast instantiate copies of themselves with their variables specified to the current situation. These instantiations are passed to the action selection mechanism, which chooses a single action from these instantiations and those remaining from previous cycles. The chosen action then goes to sensory-motor memory, where it picks up the appropriate algorithm by which it is then executed. The action so taken affects the environment, and the cycle is complete.

The LIDA model hypothesizes that all human cognitive processing is via a continuing iteration of such cognitive cycles. These cycles occur asynchronously, with each cognitive cycle taking roughly 200 ms. The cycles cascade, that is, several cycles may have different processes running simultaneously in parallel. Such cascading must, however, respect the serial order of consciousness in order to maintain the stable, coherent image of the world with which consciousness endows us (Merker 2005, Franklin 2005). This cascading, together with the asynchrony, allows a rate of cycling in humans of five to ten cycles per second. A cognitive "moment" is quite short! There is considerable empirical evidence from neuroscience studies suggestive of such cognitive cycling in humans (Lehmann et al 1998, Massimini et al 2005, Sigman and Dehaene

2006, Uchida, Kepecs and Mainen 2006, Willis and Todorov 2006, Melloni et al 2007). None of this evidence is conclusive.

3. Theory of Mind in LIDA Agent

Procedural memory in LIDA is implemented as a scheme net (D'Mello et al 2006) of templates called schemes. Information about how another cognitive agent does something is stored in the scheme net. Representations of other agents in perceptual memory and of their actions allow them to be recognized (activated) when that other agent is perceived. They remain in the workspace for a few 10's of seconds or a few hundred cognitive cycles. Structures may form as part of the updated internal model and win the competition for consciousness, which will cause them to be stored in episodic memory. This allows LIDA to reconstruct them after they decay from the workspace. This process creates a kind of virtual reality in the workspace, and allows LIDA to model other agents and predict what they are likely to do in a given situation. This ability to build internal structures predicting the actions of other agents is a high-level adaptation to the environment, which provides feedback on these predictions. In LIDA, schemes and perceptual categories used to make successful predictions have their activations increased. The opposite is true for components involved in unsuccessful predictions.

The ability to represent and model other cognitive agents is known as a theory of mind. It helps explain some aspects of human cognition such as one-shot learning by observing a teacher and the ability of people to read subtle social queues. There is a theory that the mirror neurons in the cortex are involved in this process (Frey and Gerry 2006, Buckner and Carroll 2007). Other animals are also known to exhibit theory of mind (e.g. Santos, Nissen, and Ferrugia 2006).

There are two mechanisms for implementing the theory of mind in LIDA. Perceptual memory has concept nodes (representations) for both the self and for other agents. Thus, if an action were observed in a given context, similar percepts would be produced whether the cognitive agent itself or another agent took the action. The most significant difference would be that one percept would contain an instantiation of the "self" node, whereas the other would contain a representation for the concept of the other agent.

The second mechanism is a set of codelets, small independently running programs, and schemes from procedural memory that contain slots that can bind to any agent allowing LIDA to perceive and reason about actions by other agents in the same way as it does about its own perceptions and actions.

These two mechanisms, combined with the learning procedures in LIDA, allow humans and, perhaps to some extent, other primates, to form complex representations of other agents. Simply put, a procedural learning mechanism converts a series of percepts containing another agent's concept node into a stream of schemes where the other agent's node is converted to a slot that can bind to any similar agent, including the self. This highlights one difference between learning in humans and in some lower animals, which must be trained by repeatedly rewarding the desired behavior until it is learned. This latter type of learning is similar to traditional statistical machine learning techniques, in contrast to human-like learning, represented in cognitive models that take into account research in psychology, neuroscience and cognitive science (D'Mello et al 2006).

4. Example

Thinking about other peoples' cognitive processes involves using partial knowledge about what is actually true, and ideas about how the mind works, to create a narrative about what happened (or would happen), and what a given individual would do in a particular situation. Some aspects of other minds are, or appear to be, perceived directly. Others require

deliberation. An example of the former would be a "gut" feeling about whether a person is telling the truth based on their demeanor. An example of the latter would be reasoning about a person's beliefs and motives.

A person's demeanor is related to their facial expressions, body language, tone of voice,

etc. Information about a person's state of mind may be detected through the mirror neuron system, which reacts to subtle social signals whether expressed by the person themselves or observed by the person in others (Arbib and Rizzolatti 1997). These signals accompany primal emotions such as anger, fear, or guilt and are difficult to mask (Ekman 1993). Recognition of such signals has survival value at the individual level, though warning of threats, and at the group level, by increasing trust through the ability to detect false allegiances. There is no evidence that the mirror system is used in "higher-level" tasks such as reasoning about another person's beliefs. There is evidence that people possess a relatively common set of beliefs about how minds work (Saxe 2005).

For example, there is an over-emphasis on logic that people use to explain both their own actions and those of others. This can be referred to as "naïve psychology" in analogy to the "naïve physics" that people use in reasoning about the physical world.

Both sets of beliefs contain heuristics that work well in

most cases but are shown to be false in others. A famous example from naïve physics is that "heavier objects fall faster than lighter ones." A similar example from naïve psychology is that "people act in their own self interest."

In order to avoid the complexities of the theory of mind in humans, we will use as an example a primate experiment that showed evidence of it in monkeys.

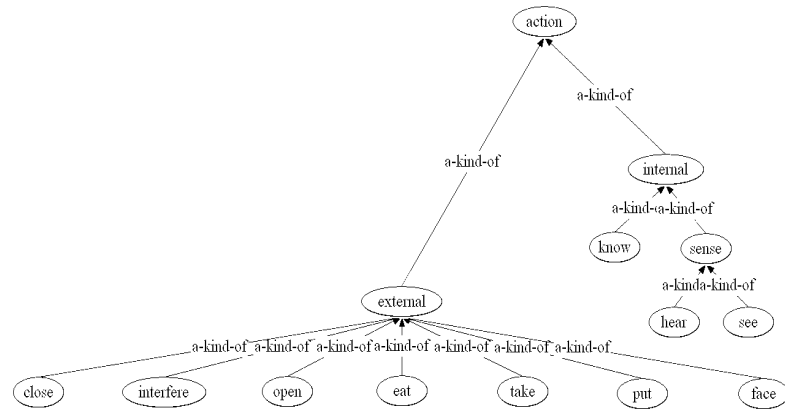


Figure 2. Perceptual Memory Actions Needed for the Experiment

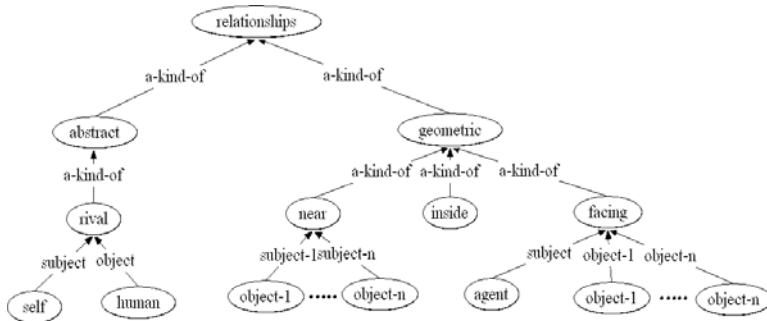


Figure 3. Perceptual Memory Relationships Needed for the Experiment

4.1. Theory of Mind in Macaca Mulatta (Rhesus Monkeys)

A recent animal experiment (Santos et al 2006) has shown evidence for the theory of mind in monkeys. In this section we describe the experiments and show how the LIDA model represents this process in a natural and realistic manner. By using experimental data from higher animals instead of humans, we can simplify the description while keeping the essential elements.

The experimenters approached a monkey with two containers, a “noisy” container with bells attached to the lid and body, and a “silent” container without the bells. A grape was placed in each container. The action of placing the grape caused noise

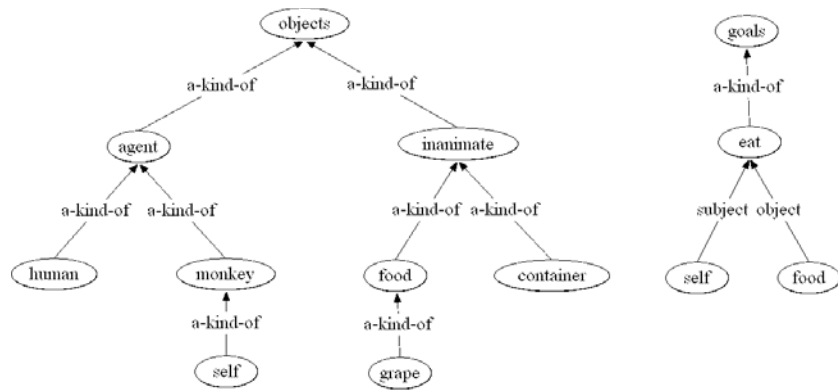


Figure 4. Perceptual Memory Objects and Goals Needed for the Experiment

for the first container but not the second. The experimenters then backed away from the containers but remained in view. In one set of experiments, the experimenters faced the containers and in another set, they faced the ground. Data were collected when the monkey took the grape from one of the containers. When the experimenters faced the ground, the monkey preferentially chose the silent container. When the experimenters faced the containers, the monkey chose randomly.

The researchers concluded that these data are evidence that the monkeys had a mental model of the human experimenters, who were considered potential rivals, i.e. competitors for the food. If the humans were not looking the containers, the monkey could get the food without alerting them by choosing the silent container. If the humans were looking at the containers, they would know of the monkey’s actions in either case, so it wouldn’t matter which container was chosen.

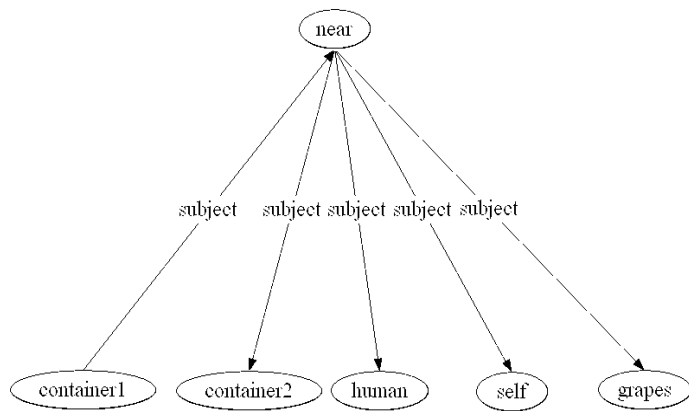


Figure 5. Initial Percept in Experiment

The cognitive states and actions of the monkey will now be illustrated using the LIDA model. Figures 2 to 4 show a simplified version of the hypothetical perceptual memory components needed to recognize the events in the experiments. As shown in the figures, there are four root concepts, actions, relationships, goals, and objects. Perceptual memory, and the other conceptual structures described in this paper, will be represented as semantic nets, that is, directed graphs with labels on both the nodes and links (Franklin 2005b). The graphs are simplified to trees by repeating certain nodes. For example, the “eat” node under “external action” is the same as the “eat” node under “goals.”

When the experimenter approaches the monkey with the containers and grapes, certain structures will be excited in perceptual memory and instantiated in the workspace as a percept. This is shown in figure 5, which simply contains a representation of the objects perceived in the scene. When the experimenter puts the grape in the noisy container, the perceptual structures in figure 6 are added.

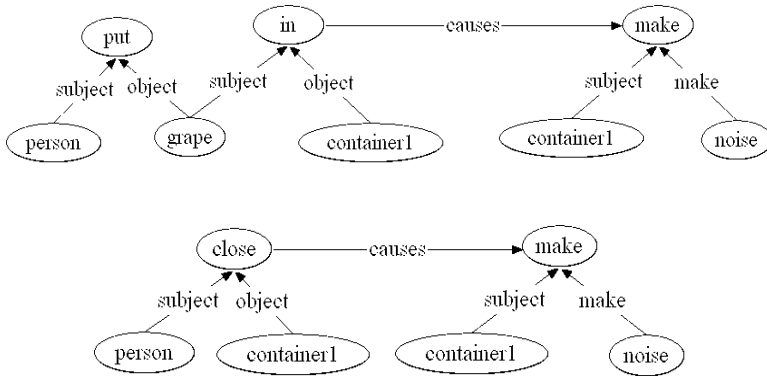


Figure 6. Percepts for Placing a Grape in the Noisy Container

One advantage of the representation in the figures is that it can be translated into natural language. For example, the perceptual structure on the upper left in Figure 6 could be translated to: The person put a grape in container1, causing it to make noise. In order to avoid a proliferation of diagrams, only the natural language description will be used in some cases.

When the experimenter puts a grape in container 2, similar perceptual structures appear in the workspace except that there is no noise because container 2 doesn't have bells attached to it. These percepts translate to: The person put a grape in container2 and the person closed container2.

Structure Building Codelets act upon the elements in the workspace. These codelets are simple computational processes that scan the workspace for elements that match class variables. In this paper we will use class names beginning with a question mark. The codelets also contain probable results, also represented as graph structures. The results can combine the triggering elements and/or include a new structure. The activation of the codelet depends on the likelihood of the result, given the trigger, which includes elements of the context implicit in the matched workspace structures. This likelihood based activation is constantly adjusted by feedback from the environment. The activation of the new structure depends on both the activation of the codelet and that of its

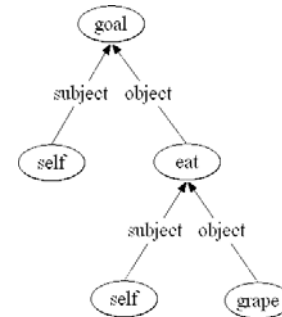


Figure 7. The monkey's goal is to eat a grape

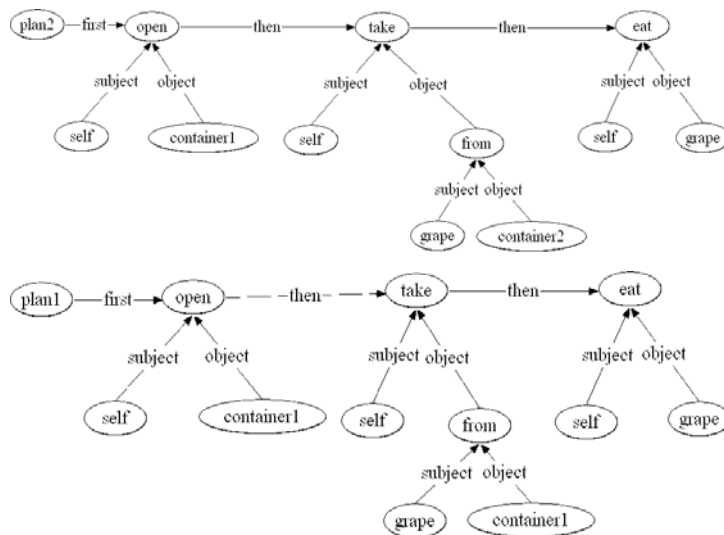


Figure 8. Alternative plans

triggering elements. Other structure building codelets can combine the new structures to build larger structures. This allows a reasonable number of codelets to build complex models. Attention Codelets are triggered by structures in the workspace. These codelets and the structures that triggered them form coalitions in the model. The coalitions compete for consciousness based their activation. The winner goes to consciousness and may eventually trigger actions by the agent.

The monkey sees the grapes, which are a kind of food. If the monkey is hungry or likes grapes, this perception may activate a goal structure in the workspace to eat the grape, as shown in Figure 7. Goal structures from perceptual memory are only potential goals. If this structure makes it to “consciousness,” it will be broadcast to all components of the model, including Procedural Memory, where it could activate schemes with results that may achieve the goal.

Instances of the activated schemes will be incorporated in the behavior net (Maes 1989, Negatu and Franklin 2002). This will excite other components of the net, where the goals can be converted into behaviors to achieve them through the activation of streams of behavior codelets. In this case, the higher-level behavior is likely to result in volitional decision making to construct a plan to obtain and eat the grape. Its actions are internal and include activation of codelets to search, write to, and build structures in the workspace.

The monkey’s mind goes through a planning process to construct a causally linked structure from perceptual structures in the workspace to the goal structure (Baars 1997, Hofstadter 1995, Franklin 2000). There are choices on how to achieve the goal, whether to take the grape from container 1 or container 2. The resulting plans are shown in Figure 8.

According to ideomotor theory (James 1890, Baars 1988), a cognitive agent will consider one plan of action at a time. In the LIDA implementation of this theory (Franklin. 2000), once the consideration begins, a timer is set. If no objection to the plan comes to mind before the timer ends, the plan will be performed. If there is an objection, the timer is reset and an alternative plan is considered.

In this experiment, any objections to the monkey’s plans would likely concern possible interference by the human, which the monkey considers a rival. The monkey

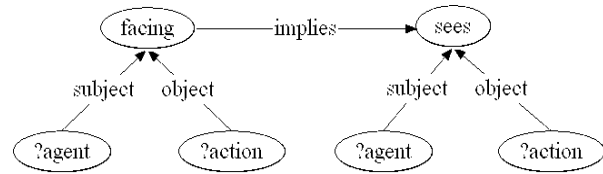


Figure 9. If an agent is facing an action, the agent is likely to see the action

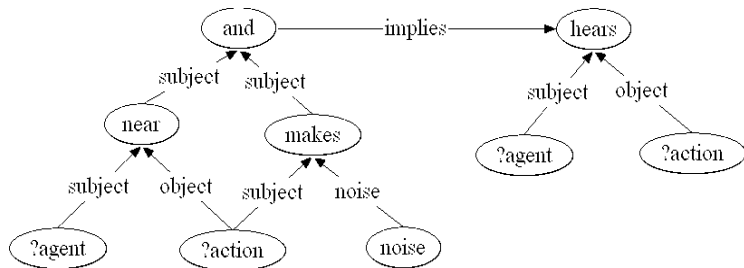


Figure 10. If an agent is near an action and the action makes noise, the agent will hear the action.

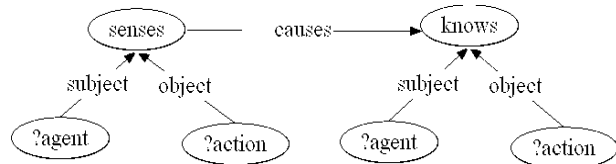


Figure 11. If an agent senses an action the agent knows the action (has occurred).

reasons about what the human would know about potential actions the monkey could take. This is the essence of the Theory of Mind. In the experiment, it can

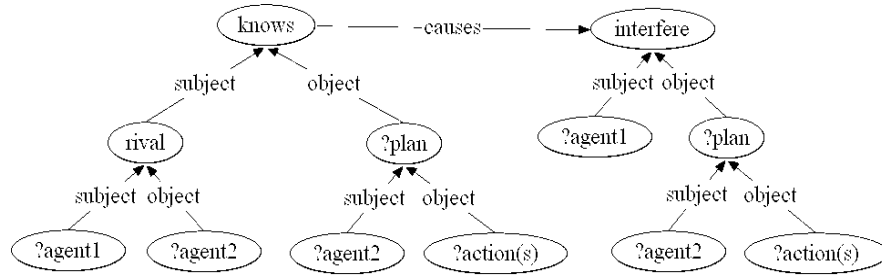


Figure 12. If agent 1 is a rival of agent 2 and agent 1 knows of planned actions by agent 2, agent 1 will interfere with agent 2's actions.

be accomplished with the following structures (Figures 9 to 12).

If the monkey considers plan two first, no objection is created and it will go to consciousness (unless some coalition with higher activation wins the competition). It will then be broadcast to procedural memory and activate schemes that are sent to the behavior net. The resulting behavior stream will instantiate a series of external behavior codelets enabling the monkey to physically open the silent container and take the grape.

If the monkey considers plan one first, however, an objection will be raised. This is due to a structure related to the concept of a rival (Figure 12): “?agent1 is a rival of ?agent2 causes ?agent1 to interfere with the plans of ?agent2 if they are known to ?agent1.” At this point, the monkey abandons plan one and considers plan two. Since there are no objections, the plan will be carried out. This explains the experimental results for the case where the human faces the ground.

In the case when the experimenter is facing the scene, the monkey's plan will be known whether or not it chooses the noisy box. This is a cognitive dilemma in the sense that the monkey could consider the first plan, find it blocked, go the second plan, find that it is also blocked, then go back to the first plan, resulting in an oscillation. There are three mechanisms for breaking the deadlock (Franklin 2000). First, the plans for eating the grape compete with other potential activities represented in the workspace. The winning coalition makes it to consciousness and is acted upon. In the experiment, for example, there were cases when another monkey drove the subject from the scene and both plans were abandoned.

Second, every time the same plan is reconsidered, the length of time on the timer is shortened so that it could end before the objection has time to form. Finally, if the other mechanisms do not force a decision, a metacognitive process will be triggered to accept whichever plan is under consideration. This multiplicity of tie-breaking mechanisms illustrates the importance of decisiveness for survival in higher animals and results in a random choice. It explains the case where the experimenter is facing the scene.

5. Discussion, Conclusions and Future Work

The LIDA model can explain the results of the primate experiment in a psychologically realistic way. It offers an explanation of the theory of mind in monkeys in the sense that the actions depending on where the experimenters were facing were derived from a representation of the experimenter's mental state and the monkey's belief in their potential to interfere with the monkey's plans.

LIDA also accounts for the results that were not analyzed in the experiment, the times when the monkey abandoned the scene without trying to retrieve either grape. In LIDA, this is a result of losing the competition for consciousness.

LIDA's cognitive cycle is based on human cognition and is consistent with the older *Sense-Plan-Act* cycle, but contains many more cognitive functions. LIDA's perceptions of the ongoing activities of itself and other agents create percepts in the workspace that trigger codelets for building additional structures, taking actions, and other cognitive functions that result in the model's interpretation of the current situation.

The basic representations in LIDA for implementing the theory of mind are found in both perceptual and procedural memories. A percept involving a given action and context can contain either the "self" node or the node for another agent, depending on who performed it. Schemes in procedural memory can contain slots that can be bound to any agent, the self or another. This is more general than what is known about the mirror neuron system, which responds strongly only to basic actions such as eating or grasping objects with the hands (Gallese, et al 1996),

While a detailed explanation of all the types of learning and adaptation in LIDA is beyond the scope of this paper, LIDA's ability for self-organization results from: a large number of simple behaviors and primitive features that can be combined in arbitrary ways; feedback from the environment; decay of learned concepts and procedures, including the ability to forget; and mechanisms that include both competitive and cooperative learning, i.e., competition between coalitions of cognitive structures.

Uncertainty plays a role in LIDA's reasoning through the base activation of its *behavior codelets*, which depend on the model's estimated probability of the codelet's success if triggered. LIDA observes the results of its behaviors and updates the base activation of the responsible codelets dynamically.

It avoids combinatorial explosions by combining reasoning via association with reasoning via deduction. One can create an analogy between LIDA's workspace structures and codelets and a logic-based architecture's assertions and functions. However, LIDA's codelets only operate on the structures that are active in the workspace during any given cycle. This includes recent perceptions, their closest matches in other types of memory, and structures recently created by other codelets. The results with the highest estimate of success, i.e. *activation*, will then be selected. No attempt is made to find all possible solutions or reason until either a solution is found it is shown that none exists. If reasoning takes too long, time-keeping mechanisms such as the ones described above will force termination. The disadvantage is that incorrect conclusions are possible and potential solutions to problems can be overlooked, similar to the way in which human cognition works.

One would expect higher level cognition to be more sophisticated in humans than in other primates and perhaps lacking in some lower mammals and other animals. The primate experiment was selected for this paper because of the simplicity of the situation involving evidence for the theory of mind. Its implementation in LIDA provides a mechanism to explain "one-shot" learning by observing a teacher.

In this paper, the model is being tested only qualitatively, showing that it explains the behaviors exhibited by the monkeys. Further research will involve enhancing the existing LIDA implementation so that the experiment can be simulated and quantitatively confirm the predicted results, or not.

6. References

- Arbib, M.A. and G. Rizzolatti. 1997. Neural expectations: A possible evolutionary path from manual skills to language. *Communication and Cognition* 29: 393-423.
- Baars, Bernard J. 1988. *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.
- Baars, Bernard. 1997. In the theatre of consciousness. Global workspace theory, a rigorous scientific theory of consciousness. *Journal of Consciousness Studies* 4: 292–309.
- Baars, B. J., and S. Franklin. 2003. How conscious experience and working memory interact. *Trends in Cognitive Science* 7:166-172.
- Buckner, Randy L. and Daniel C. Carroll. 2007. Self-projection and the brain. *Trends in Cognitive Sciences* 11, no. 2: 49-57.
- D'Mello, Sidney K, S Franklin, U Ramamurthy, and B J Baars. 2006. A cognitive science based machine learning architecture. In *AAAI 2006 Spring Symposium Series Sponsor: American Association for Artificial Intelligence*. Stanford University, Palo Alto, California, USA.
- D'Mello, Sidney K, U Ramamurthy, A Negatu, and S Franklin. 2006. A procedural learning mechanism for novel skill acquisition. In *Proceeding of adaptation in artificial and biological systems, aisb'06*, ed. Tim Kovacs and James A R Marshall, 1:184–185. Bristol, England: Society for the Study of Artificial Intelligence and the Simulation of Behaviour.
- Ekman, P. 1993. Facial expression of emotion. *American Psychologist* 48: 384-392.
- Franklin, Stan. 1995. *Artificial minds*. Cambridge MA: MIT Press.
- Franklin, Stan. 2000. Deliberation and voluntary action in 'conscious' software agents. *Neural Network World* 10: 505–521.
- Franklin, Stan. 2005. Evolutionary pressures and a stable world for animals and robots: A commentary on merker. *Consciousness and Cognition* 14: 115–118.
- Franklin, S. 2005a. Cognitive robots: Perceptual associative memory and learning. In *Proceedings of the 14th annual international workshop on robot and human interactive communication (Ro-Man 2005)*:427-433.
- Franklin, Stan. 2005b. Perceptual memory and learning: Recognizing, categorizing, and relating. In *Symposium on Developmental Robotics: American Association for Artificial Intelligence (AAAI)*. Stanford University, Palo Alto CA, USA.
- Franklin, Stan. 2007. A foundational architecture for artificial general intelligence. In *Advances in artificial general intelligence: Concepts, architectures and algorithms, proceedings of the AGI workshop 2006*, ed. Ben Goertzel and Pei Wang:36-54. Amsterdam: IOS Press.
- Franklin, S., B. J. Baars, U. Ramamurthy, and M. Ventura. 2005. The Role of Consciousness in Memory. *Brains, Minds and Media* 1:1-38, pdf.
- Franklin, S., and A. C. Graesser. 1997. Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents. In *Intelligent Agents III*. Berlin: Springer Verlag.
- Franklin, S. and F. G. Patterson, Jr. (2006). The Lida Architecture: Adding New Modes of Learning to an Intelligent, Autonomous, Software Agent. *Integrated Design and Process Technology, IDPT-2006*, San Diego, CA, Society for Design and Process Science.

- Frey, Scott H and Valerie E. Gerry. 2006. Modulation of neural activity during observational learning of actions and their sequential orders. *Journal of Neuroscience* 26: 13194-13201.
- Gallese, V., Fadiga, L., Fogassi, L., Rizzolatti, G.. Action recognition in the premotor cortex. *Brain* (1996), 119, 593-609.
- Hofstadter, D. R. & the Fluid Analogies Research Group (1995). *Fluid concepts and creative analogies*. New York: Basic Books.
- James, W. 1890. *The Principles of Psychology*. Cambridge, MA: Harvard University Press.
- Lehmann D, Strik WK, Henggeler B, Koenig T, and Koukkou M. 1998. Brain electric microstates and momentary conscious mind states as building blocks of spontaneous thinking: I. Visual imagery and abstract thoughts. *Int. J Psychophysiology* 29, no. 1: 1-11.
- Maes, P. 1989. How to do the right thing. *Connection Science* 1:291-323.
- Massimini, M., F. Ferrarelli, R. Huber, S. K. Esser, H. Singh, and G. Tononi. 2005. Breakdown of Cortical Effective Connectivity During Sleep. *Science* 309:2228-2232.
- Merker, Bjorn. 2005. The liabilities of mobility: A selection pressure for the transition to consciousness in animal evolution. *Consciousness and Cognition* 14: 89–114.
- Negatu, Aregahegn and Stan Franklin. 2002. An action selection mechanism for 'conscious' software agents. *Cognitive Science Quarterly* 2, no. special issue on "Desires, goals, intentions, and values: Computational architectures." Guest editors Maria Miceli and Cristiano Castelfranchi.: 363–386.
- Osterberg, G. 1935. Topography of the layer of rods and cones in the human retina. *Acta Ophthalmol. suppl.* 6, no. 1-103.
- Premack, D. G. and G. Woodruff. 1978. Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences* 1: 515-526.
- Ramamurthy, U., Baars, B., D'Mello, S. K., & Franklin, S. (2006). LIDA: A Working Model of Cognition. *Proceedings of the 7th International Conference on Cognitive Modeling*. Eds: Danilo Fum, Fabio Del Missier and Andrea Stocco; pp 244-249. Edizioni Goliardiche, Trieste, Italy.
- Saxe, Rebecca. 2005. Against simulation: the argument from error. *TRENDS in Cognitive Sciences* Vol. 9 No. 4, pp 174-179.
- Santos, Laurie R, Aaron G Nissen, and Jonathan A Ferrugia. 2006. Rhesus monkeys, macaca mulatta, know what others can and cannot hear. *Animal Behaviour* 71: 1175–1181.
- Sigman, M., and S. Dehaene. 2006. Dynamics of the Central Bottleneck: Dual-Task and Task Uncertainty. *PLoS Biol.* 4.
- Tse, P. U. (2003). If vision is 'veridical hallucination', what keeps it veridical? Commentary (p. 426-427) on Gestalt isomorphism and the primacy of subjective conscious experience: a Gestalt Bubble model, by Steven Lehar. *Behavioral and Brain Sciences*, 26(4):375-408.
- Uchida, N., A. Kepecs, and Z. F. Mainen. 2006. Seeing at a glance, smelling in a whiff: rapid forms of perceptual decision making. *Nature Reviews Neuroscience* 7:485-491.
- Willis, J., and A. Todorov. 2006. First Impressions: Making Up Your Mind After a 100-Ms Exposure to a Face. *Psychological Science* 17:592-599.