2007

LIDA: A computational model of global workspace theory and developmental learning

S. Franklin

U. Ramamurthy

S. Mello

L. McCauley

A. Negatu

See next page for additional authors

Authors
S. Franklin, U. Ramamurthy, S. Mello, L. McCauley, A. Negatu, R. Silva, and V. Datla

LIDA: A Computational Model of Global Workspace Theory and Developmental Learning

Stan Franklin, Uma Ramamurthy, Sidney K. D'Mello, Lee McCauley, Aregahegn Negatu, Rodrigo Silva L., Vivek Datla

Institute for Intelligent Systems and the Department of Computer Science, The University of Memphis, Memphis, TN 38152, USA
franklin@memphis.edu, urmmrthy@memphis.edu, sdmello@memphis.edu, tmccauly@memphis.edu, asnegatu@memphis.edu, rsilva@memphis.edu, vdatla@memphis.edu

Abstract

In this paper, we present LIDA, a working model of, and theoretical foundation for, machine consciousness. LIDA's architecture and mechanisms were inspired by a variety of computational paradigms and LIDA implements the Global Workspace Theory of consciousness. The LIDA architecture's cognitive modules include perceptual associative memory, episodic memory, functional consciousness, procedural memory and action-selection. Cognitive robots and software agents controlled by the LIDA architecture will be capable of multiple learning mechanisms. With artificial feelings and emotions as primary motivators and learning facilitators, such systems will 'live' through a developmental period during which they will learn in multiple, human-like ways to act effectively in their environments. We also provide a comparison of the LIDA model with other models of consciousness.

Introduction

Here we describe the LIDA (Learning IDA) model of consciousness and cognition. LIDA implements Global Workspace Theory (GWT) (Baars 1988; 1997), which has become the most widely accepted psychological and neurobiological theory of consciousness (Baars 2002; Dehaene & Naccache 2001; Kanwisher 2001). In the process of implementing GWT, the LIDA model also implements imagination in the form of deliberation as a means of action selection (Franklin 2000a; Sloman 1999), as well as volition à la ideomotor theory (Franklin 2000a; James 1890). Comprising a complete control structure for software agents (Franklin & Graesser 1997) and, potentially for autonomous robots (Franklin & McCauley 2003), the computational LIDA can be thought of as a virtual machine (Sloman & Chrisley 2003), built on top of a series of other virtual machines: a Java development environment, an operating system, microcode, etc. The LIDA model employs feelings and emotions throughout, both as motivators (Sloman 1987) and as modulators of learning. With its emphasis on several forms of learning, a LIDA controlled software agent or autonomous robot would be expected to go through a developmental period (Franklin 2000b) as would a human infant. During this period the agent/robot would develop its own ontology. Falling clearly within the purview of embodied or enactive approaches to cognition (Varela, Thompson, & Rosch 1991), the LIDA model is quite consistent with a number of other psychological theories (Baddeley 1993; Barsalou 1999; Conway 2002; Ericsson & Kintsch 1995; Glenberg 1997).

In what follows, we'll review GWT, describe the LIDA architecture and the cognitive cycle by which it operates, and discuss how various multi-cyclic cognitive processes are implemented in the model. We'll then compare the LIDA model with several other computational and conceptual models of consciousness, and conclude with some final thoughts and intentions for future work.

LIDA as a model of GWT

Global Workspace Theory (GWT) attempts to integrate a large body of evidence into a single conceptual framework focused on the role of consciousness in human cognition (Baars 1988, 1997, 2002; Baars & Franklin 2003). Like other theories, GWT postulates that human cognition is implemented by a multitude of relatively small, special purpose processors, almost always unconscious (Edelman 1987; Jackson 1987; Minsky 1985; Ornstein 1986). Processors are comparatively simple, and communication between them is relatively rare, occurring over a narrow signal bandwidth. A coalition of such processors is a collection that works together to perform a specific task. Coalitions normally perform routine actions, in pursuit of sensory, motor, or other problem-solving tasks. GWT suggests that the brain supports a global workspace capacity which allows for the integration and distribution of separate processors (for neuroscience evidence, see Baars 2002; Schneider & Chein 2003). A coalition of processors that gains access to the global workspace can broadcast a message to all the unconscious processors, in order to recruit new components to join in interpreting a novel situation, or in solving the current problem.

In GWT, consciousness allows the brain to deal with novel or problematic situations that can't be dealt with efficiently, or at all, by habituated unconscious processes. GWT also suggests an answer to the paradox of cognitive limited capacity associated with conscious experience, immediate memory, and immediate goals. GWT suggests that the compensating advantage is the ability to mobilize many unconscious resources in a non-routine way to address novel challenges.

GWT offers an explanation for consciousness being serial in nature rather than parallel as is common in the rest of the nervous system. Messages broadcast in parallel would tend to overwrite one another making understanding difficult. It similarly explains the limited capacity of consciousness as opposed to the huge capacity typical of long-term memory and other parts of the nervous system.

LIDA is a proof of concept model for GWT. Almost all the tasks in this model are accomplished by codelets (Hofstadter & Mitchell 1994) representing the processors in GWT. Codelets are small pieces of code, each running independently. A class of codelets called *attention codelets* - each looks out for situations of interest to them, and attempt to bring them to 'consciousness' spotlight. A broadcast then occurs to all the processors in the system to recruit resources to handle the current situation. The LIDA model also implements the seriality of consciousness in GWT with the *cognitive cycle* described later in this paper. Further the LIDA model addresses a vast expanse of cognitive processes including perception, various memory systems, action selection, developmental learning mechanisms, feelings and emotions, deliberation, voluntary action, non-routine problem solving and automatization.

The LIDA Architecture

The LIDA architecture is partly symbolic and partly connectionist with all symbols being grounded in the physical world in the sense of Brooks (1986). The mechanisms used in implementing the several modules have been inspired by a number of different 'new AI' techniques (Brooks 1986; Drescher 1991; Hofstadter & Mitchell 1994; Jackson 1987; Kanerva 1988; Maes 1989). We now describe LIDA's primary mechanisms.

Perceptual Associative Memory

LIDA perceives both exogenously and endogenously with Barsalou's perceptual symbol systems serving as a guide (1999). The perceptual knowledge-base of this agent, called perceptual associative memory, takes the form of a semantic net with activation called the slipnet, a la Hofstadter and Mitchell's Copycat architecture (1994). Nodes of the slipnet constitute the agent's perceptual symbols, representing individuals, categories, relations, etc. It should be noted that any node can be traced back to its primitive feature detectors that are grounded in reality and change according to the sensors of the agent in question. Pieces of the slipnet containing nodes and links, together with perceptual codelets with the task of copying what is being currently perceived to working memory, constitute Barsalou's perceptual symbol simulators (1999). Together they constitute an integrated perceptual system for LIDA, allowing the system to recognize, categorize and understand.

Workspace

LIDA's workspace is analogous to the preconscious buffers of human working memory. Perceptual codelets write to the workspace as do other, more internal codelets. Attention codelets watch what is written in the workspace in order to react to it. Items in the workspace decay over time, and may be overwritten.

Another pivotal role of the workspace is the building of temporary structures over multiple cognitive cycles (see below). Perceptual symbols from the slipnet are assimilated into existing relational and situational templates while preserving spatial and temporal relations between the symbols. The structures in the workspace also decay rapidly.

Episodic Memory

Episodic memory in the LIDA architecture is composed of a declarative memory for the long term storage of autobiographical and semantic information as well as a short term transient episodic memory similar to Conway's (2001) sensory-perceptual episodic memory with a retention rate measured in hours. LIDA employs variants of sparse distributed memory to computationally model declarative and transient episodic memory (Kanerva 1988; Ramamurthy, D'Mello, & Franklin 2004). Sparse distributed memory is a content addressable, associative memory that shares several functional similarities to human long term memory (Kanerva 1988).

Functional Consciousness

LIDA's 'consciousness' module implements Global Workspace theory's (Baars 1988) processes by codelets. These are specialized for some simple task and often play the role of a daemon watching for an appropriate condition under which to act. The apparatus for functional 'consciousness' consists of a coalition manager, a spotlight controller, a broadcast manager, and attention codelets that recognize novel or problematic situations. Please see the discussion above, and the description of the cognitive cycle below for more details.

Procedural Memory

Procedural memory in LIDA is a modified and simplified form of Drescher's schema mechanism (1991), the scheme net. Like the slipnet of perceptual associative memory, the scheme net is a directed graph whose nodes are (action) schemes and whose links represent the 'derived from' relation. Built-in primitive (empty) schemes directly controlling effectors are analogous to motor cell assemblies controlling muscle groups in humans. A scheme consists of an action, together with its context and its result. At the periphery of the scheme net lie empty schemes (schemes with a simple action, but no context or results), while more complex schemes consisting of actions and action sequences are discovered as one moves inwards. In order for a scheme to act, it first needs to be instantiated and then selected for execution in accordance with the action selection mechanism described next.

Action Selection

The LIDA architecture employs an enhancement of Maes' behavior net (1989) for high-level action selection in the service of feelings and emotions. Several distinct feelings and emotions operate in parallel, perhaps varying in urgency as time passes and the environment changes. The behavior net is a digraph (directed graph) composed of behaviors (instantiated action schemes) and their various links. As in connectionist models, this digraph spreads activation. The activation comes from four sources: from pre-existing activation stored in the behaviors, from the environment, from feelings and emotions, and from internal states. To be acted upon, a behavior must be executable, must have activation over threshold, and must have the highest such activation.

The LIDA Cognitive Cycle

Be it human, animal, software agent or robot, every autonomous agent within a complex, dynamic environment must frequently and cyclically sample (sense) its environment and act on it, iteratively, in what we call a cognitive cycle (Franklin et al., 2005). Cognitive cycles are flexible, serial but overlapping cycles of activity usually beginning in perception and ending in an action. We suspect that cognitive cycles occur five to ten times a second in humans, cascading so that some of the steps in adjacent cycles occur in parallel (Baars & Franklin 2003). Seriality is preserved in the conscious broadcasts. We now describe the cognitive cycle dividing it into nine steps.

1) Perception. Sensory stimuli, external or internal, are received and interpreted by perception producing the beginnings of meaning.

2) Percept to preconscious buffer. The percept, including some of the data plus the meaning, as well as possible relational structures, is stored in the preconscious buffers of LIDA's working memory (workspace). Temporary structures are built.

3) Local associations. Using the incoming percept and the residual contents of working memory, including emotional content, as cues, local associations are automatically retrieved from transient episodic memory and from declarative memory, and stored in long-term working memory.

4) Competition for consciousness. Attention codelets view long-term working memory, and bring novel, relevant, urgent, or insistent events to consciousness.

5) Conscious broadcast. A coalition of codelets, typically an attention codelet and its covey of related information codelets carrying content, gains access to the global workspace and has its contents broadcast. In humans, this broadcast is hypothesized to correspond to phenomenal consciousness.

6) Recruitment of resources. Relevant schemes respond to the conscious broadcast. These are typically schemes whose context is relevant to information in the conscious broadcast. Thus consciousness solves the relevancy problem in recruiting resources.

7) Setting goal context hierarchy. The recruited schemes use the contents of consciousness, including feelings/emotions, to instantiate new goal context hierarchies (copies of themselves) into the behavior net, bind their variables, and increase their activation. Other, environmental, conditions determine which of the earlier goal contexts also receive variable binding and/or additional activation.

8) Action chosen. The behavior net chooses a single behavior (scheme, goal context), from a just instantiated behavior stream or possibly from a previously active stream. Each selection of a behavior includes the generation of an expectation codelet (see the next step).

9) Action taken. The execution of a behavior (goal context) results in the behavior codelets performing their specialized tasks, having external or internal consequences, or both. LIDA is taking an action. The acting codelets also include at least one expectation codelet whose task it is to monitor the action, bringing to consciousness any failure in the expected results.

Multi-cyclic Processes in LIDA

Higher order cognitive processes such as reasoning, problem solving, imagination, etc., in LIDA, occur over multiple cognitive cycles. We now describe deliberation, voluntary action, non-routine problem solving, and automatization as some of the multi-cyclic processes that are accommodated by the LIDA system. The mechanisms that realize these processes are typically implemented as behavior streams in procedural memory.

Deliberation

When we humans are faced with a problem to solve, we often create in our mind different strategies or possible solutions. We imagine the effects of executing each strategy or trial solution without actually doing so. This is similar to a kind of internal virtual reality. Eventually, we decide upon one strategy or trial solution, and try solving the problem using it. This process is called deliberation (Sloman 1999). During the deliberation process several, possibly conflicting ideas compete to be selected as the strategy or solution of the problem. One such is chosen voluntarily. Deliberation in LIDA is implemented by utilizing conscious information to create scenarios and evaluate their utilities (Franklin 2000b).

Voluntary Action

Voluntary actions involve a conscious deliberation on the decision to take an action. William James proposed the *ideomotor theory* of voluntary action (1890). James suggests that any idea (internal proposal) for an action that comes to mind (to consciousness) is acted upon unless it provokes some opposing idea or some counter proposal. GWT adopts James' ideomotor theory "as is" (Baars 1988) and provides a functional architecture for it. The LIDA model furnishes an underlying mechanism that implements the ideomotor theory of volition (Franklin 2000b).

The players in this decision making process include *proposing* and *objecting* attention codelets and a *timekeeper* codelet. A proposing attention codelet's task is to propose a certain action on the basis of its particular pattern of preferences. The proposing attention codelet brings information about itself and the proposed action to "consciousness" so that if no other objecting attention codelet objects (by bringing itself to "consciousness" with an objecting message), and if no other proposing attention codelet makes a different proposal within a given span of time, the timekeeper codelet will decide on the proposed action. If an objection or a new proposal is made in a timely fashion, the timekeeper codelet stops timing or resets the timing for the new proposal.

Non-Routine Problem Solving

With the help of its consciousness mechanism, LIDA has the ability to deal with novel instances of routine situations. However, in order to efficiently handle novel, problematic, and unexpected situations, the model needs some form of non-routine problem solving. In general non-routine problem solving refers to the ability to devise solutions to novel problematic situations. This type of solution is generally referred to as *meshing*, where humans utilize chunks of prior knowledge towards obtaining solutions to novel problems (Glenberg 1997). Non-routine problem solving is quite similar to planning in classical AI. However, while planning in classical AI assumes that all the individual operators are continually available for consideration, we make no such assumption due to its cognitive implausibility. Instead our approach relies on consciousness to recruit unconscious pieces of knowledge that are potentially relevant to the solution. Non-routine problem solving in the LIDA architecture is best viewed as a unique behavior stream operating over multiple cycles, with the shaping of partial plans of action at each cycle.

Automatization

Automatization refers to the human (and animal) ability to learn a procedural task to an extent that the task can be accomplished without conscious intervention. Since consciousness is a limited resource, automatized tasks free up this resource for more pressing cognitive activities such as deliberation, problem solving, reasoning, etc.

In the LIDA architecture partial plans of actions are represented by behavior streams (goal context hierarchies consisting of behaviors operating roughly in a sequence). For a non-automatized task, consciousness is required to recruit for execution the next behavior in an instantiated stream. Automatization is implemented in LIDA by means of behaviors in a stream automatically building associations with one another, thereby eliminating the need for conscious intervention. Once a task is automatized, the execution of individual behaviors is monitored by expectation codelets. When a failed execution is noted and this information is brought to consciousness, the de-automatization process is recruited to temporarily suspend the automatization thereby restoring conscious intervention (Negatu, McCauley, & Franklin, in review).

Developmental Learning in LIDA

The LIDA model realizes three fundamental learning mechanisms that underlie much of human learning: 1) *perceptual learning*, the learning of new objects, categories, relations, etc., 2) *episodic learning* of events, the what, where, and when, 3) *procedural learning*, the learning of new actions and action sequences with which to accomplish new tasks. Although, the type of knowledge retained due to these three learning mechanisms differ, the mechanisms are founded on two basic premises. The first premise states that conscious awareness is sufficient for learning. Although subliminal acquisition of information appears to occur, the effect sizes are quite small compared to conscious learning. In a classic study, Standing (1973) showed that 10,000 distinct pictures could be learned with 96% recognition accuracy, after only 5 seconds of conscious exposure to each picture. No intention to learn was needed. Consciously learned educational material has been recalled after 50 years (Bahrick 1984). Conscious access greatly facilitates most types of learning. The second premise that is shared among the various learning mechanisms is that the learning is modulated by feelings and emotions, i.e. the learning rate varies with arousal (Yerkes & Dodson 1908).

Developmental learning in LIDA occurs during the conscious broadcast (step 5 of the cognitive cycle). The conscious broadcast contains the entire content of consciousness including the affective portions. The contents of perceptual associative memory are updated in light of the current contents of consciousness, including *feelings/emotions*, as well as objects, categories and relations (perceptual learning). *Up to a point, the stronger the affect, the stronger the encoding in memory*. Transient episodic memory is also updated with the current contents of consciousness, including *feelings/emotions*, as events (episodic learning). *Up to a point, the stronger the affect is, the stronger the encoding in memory*. Procedural memory (recent actions) is updated (reinforced) with *the strength of the reinforcement influenced by the strength of the affect* (procedural learning).

Comparison with other models

As described above, the LIDA model includes several cognitive mechanisms resulting in a working model of consciousness. In this section, we will compare the LIDA model with other models of consciousness.

CLARION vs. LIDA

First we consider the *Connectionist Learning with Adaptive Rule Induction ON-line*, CLARION (Sun 2003) architecture. This architecture takes the two-systems view of consciousness, i.e., conscious processes (top-level knowledge) are directly accessible while unconscious processes (bottom-level knowledge) are not. In contrast, the LIDA model takes the unitary-system view of the conscious and the unconscious. LIDA implements GWT's view that the primary function of consciousness is to solve the *relevance* problem, namely, finding the resources needed to handle the current situation.

While the CLARION architecture supports several memory systems including working memory, semantic memory and episodic memory, the LIDA model has a distinct transient episodic memory (TEM) based on the hypothesis that only conscious contents are stored in TEM to be consolidated at a later time into declarative memory. Both the models support various types of learning mechanisms, while the capability of developmental learning exists in the LIDA model. The LIDA model also supports multi-cyclic processes like deliberation, voluntary action, non-routine problem solving and automatization providing this model with a depth and vastness in modeling several aspects of cognition.

Schacter's model vs. LIDA

Schacter's model (1990) has strong neuropsychological motivations with respect to dissociation of various types of knowledge in the system. The different knowledge modules in this system perform specialized and unconscious tasks and send the output to the "conscious awareness system". The LIDA model's approach to conscious awareness is clearly different from this. In LIDA, the action to be taken is chosen after the conscious broadcast (refer to LIDA's cognitive cycle above). Further, there is no clear computational distinction between conscious and unconscious processes in Schacter's model.

Damasio's model vs. LIDA

Damasio's model (1990) is neuro-anatomically motivated, with several "sensory convergence zones" integrating information from sensory modalities through forward and backward synaptic connections. The activation passes through the entire system, with the resulting "broadcast" making the information stored about an entity available. This is described as "accessibility of consciousness". The model has no central information store. In contrast, the LIDA model has a 'consciousness' module to implement functional consciousness. Compared to the LIDA model, Damasio's model is much more narrow in scope with respect to the cognitive mechanisms addressed. However, an advantage to Damasio's model is that it addresses multi-modal sensory convergence which has not yet been explicitly addressed in the LIDA model's perception module.

Cotterill's model vs. LIDA

The "master-module" model of consciousness by Cotterill (1997) postulates that consciousness arises from the planning of movements. The master-module in this system is the brain's motor planner. Movement is the central aspect of this model. To that extent, this model is comparable to the LIDA model, where the cognitive cycle is continuously acting and sensing its environment. The LIDA model has a much broader scope in modeling cognition compared to the master-module model.

ICARUS vs. LIDA

Similar to the LIDA model, the ICARUS architecture (Langley, P., in press) has comparative breadth in modeling cognition. ICARUS is based on Newell's view (1990) that agent architectures should incorporate strong theoretical assumptions about the nature of the mind. This is also true of the LIDA model as it attempts to integrate what we know about cognition from neuroscience, cognitive science and AI.

ICARUS has several memory systems, both long-term and short-term, similar to LIDA. It has a separate memory module for "skills or procedures" similar to LIDA's procedural memory. Similar to the CLARION model, there is no transient episodic memory system (which plays a unique role in the LIDA model) in ICARUS, though there are perceptual and motor buffers in the short-term memories of this model. Similar to LIDA, the ICARUS architecture has a strong correspondence amongst the contents of its short-term and long-term memories.

ICARUS has separate performance modules for conceptual inference, skill execution and problem solving. These modules are very inter-related with respect to building on each others' output and operating on the same subsystems/structures of the architecture, including the long-term memories. This is directly compatible with LIDA's cognitive cycle.

The ICARUS architecture has a learning module which generates a new skill whenever a goal is achieved with problem-solving and execution. In contrast to the LIDA architecture, there are no multiple learning mechanisms. While the learning in ICARUS is incremental, there is no developmental learning. The ICARUS architecture does not mention a 'consciousness' or awareness module.

Conclusion

Here we have presented the LIDA model as a case study of models of consciousness (MoC), and compared it with other such MoC. Though the various models often have, by necessity, many areas of similarity, the comparisons above suggest a number of possible lacunae in many of the models, many or most of which could perhaps be filled, to the improvement of the models. Here we will enumerate some of these suggested lacunae.

In recent years, artificial intelligence and cognitive science have both been moving rapidly toward an embodied/enactive/situated view of intelligence and cognition (Barsalou 1999; Franklin 1997; Glenberg 1997; Varela Thompson, & Rosch 1991). (The neuroscientists have long adhered to that view.) The case study of the LIDA model presented here and its comparison with other MoC, suggests that designers of other models of consciousness might well adopt this same view. While the LIDA model begins with sensation and ends with action, many of the other models assume that sensation/perception is provided and go on to model higher-level processes only, thereby running the risk of avoiding some of the really significant problems associated with perception.

Global Workspace Theory (GWT) has become the currently dominant bio-psychological theory of consciousness. Our case study model, LIDA, implements a major portion of GWT. The theories of Dehaene (Dehaene & Naccache 2001) and of Shanahan (2006) are the only other MoC to avail themselves of the insights of GWT, particularly of its solution to the relevance problem, a major function of consciousness.

Learning and development is becoming an ever more significant part of cognitive modeling, both computational and conceptual, and the several varieties of learning form an integral part of the LIDA (Learning IDA) model. Most of the other MoC discussed here provide some form of procedural learning, usually selectionist procedural learning via reinforcement, but some also include instructionist procedural learning. Episodic learning is included in a few of the other MoC, while perceptual learning is almost universally omitted, to the detriment of these models, at least in our view.

Finally, the LIDA model implementation of the multicyclic processes of conscious volition and deliberation, a form of imagination, seem unique among the other MoC to which we compared our model. Our conclusion is that many of the MoC might well benefit by taking perception, including perceptual learning, more seriously, by availing themselves of the insights of GWT, including volition via ideomotor theory, by the use of feelings and emotions to facilitate learning and development.

A final comparison of the LIDA model to humans, whom we model, reveals even more lacunae than those so belabored in the preceding paragraphs. For a start, LIDA is missing metacognition, perhaps a half-dozen senses of self, attentional learning, and a host of multi-cyclic cognitive processes, as well a clear account of how new skills are learned first with the active participation of consciousness, and later progress to becoming sensory-motor automatisms that operate completely unconsciously (Goodale and Milner 2004). There's so much work to be done to take the beam from our own eye.

References

Baars, B. J. 1988. *A Cognitive Theory of Consciousness*. Cambridge.: Cambridge University Press.

Baars, B. J. 1997. In the Theatre of Consciousness. Oxford: Oxford University Press.

Baars, B. J. 2002. The conscious access hypothesis: origins and recent evidence. Trends in Cognitive Science 6:47–52.

Baars, B., and Franklin, S. 2003. How conscious experience and working memory interact. *Trends in Cognitive Science* 7:166-172.

Baddeley, A. D. (1993) Working memory and conscious awareness. In Theories of memory, ed. A. Collins, S. Gathercole, M. Conway, and P. Morris. Howe: Erlbaum.

Barsalou, L. W. 1999. Perceptual symbol systems. *Behavioral and Brain Sciences* 22:577–609.

Bahrick, H. P. 1984. Semantic memory content in permastore: fifty years of memory for Spanish learned in school. J. Exp. Psychol. Gen. 113:1-29.

Brooks, R. A. 1986. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*. RA-2:14-23.

Conway, M. A. (2002) Sensory-perceptual episodic memory and its context: autobiographical memory. In Episodic Memory, ed. A. Baddeley, M. Conway, and J. Aggleton. Oxford: Oxford University Press.

Cotterill, R. (1997) On the mechanism of consciousness. *Journal of Consciousness Studies*, 4:231-247.

Damasio, A. et. al. (1990) Neural regionalization of knowledge access. In: *Cold Spring Harbor Symposium on Quantitative Biology*, Vol. LV. CSHL Press.

Dehaene, S., and L. Naccache. 2001. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. Cognition 79:1-37.

Drescher, G. 1991. *Made Up Minds: A Constructivist Approach to Artificial Intelligence*, Cambridge, MA: MIT Press.

Edelman, G. M. (1987) Neural Darwinism. New York: Basic Books.

Ericsson, K. A., and W. Kintsch. 1995. Long-term working memory. Psychological Review 102:21–245.

Franklin, S. 1997. Autonomous Agents as Embodied AI. *Cybernetics and Systems* 28:499-520.

Franklin, S. 2000a. Modeling Consciousness and Cognition in Software Agents. In Proceedings of the Third International Conference on Cognitive Modeling, Groeningen, NL, March 2000, ed. N. Taatgen. Veenendal, NL: Universal Press.

Franklin, S. 2000b. Deliberation and Voluntary Action in 'Conscious' Software Agents. Neural Network World 10:505-521.

Franklin, S. 2003. IDA: A Conscious Artifact? *Journal of Consciousness Studies* 10:47-66.

Franklin, S., Baars, B.J., Ramamurthy, U., & Ventura, M. 2005. The Role of Consciousness in Memory. Brains, Minds and Media, Vol.1, bmm150 (urn:nbn:de:0009-3-1505)

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. 21-35.

Franklin, S., and L. McCauley. 2003. Interacting with IDA. In Agent Autonomy, ed. H. Hexmoor, C. Castelfranchi, and R. Falcone. Dordrecht: Kluwer.

Glenberg, A. M. 1997. What memory is for. *Behavioral and Brain Sciences* 20:1–19.

Goodale, M. A., and D. Milner. 2004. *Sight Unseen*. Oxford: Oxford University Press.

Hofstadter, D. R., and M. Mitchell. 1994. The Copycat Project: A model of mental fluidity and analogy-making. In *Advances in connectionist and neural computation theory*, Vol. 2: logical connections, ed. K. J. Holyoak, and J. A. Barnden. Norwood N.J.: Ablex.

Jackson, J. V. 1987. *Idea for a Mind*. Siggart Newsletter, 181:23–26.

James, W. 1890. The Principles of Psychology. Cambridge, MA: Harvard University Press.

Kanerva, P. 1988. *Sparse Distributed Memory*. Cambridge MA: The MIT Press.

Kanwisher, N. 2001. Neural events and perceptual awareness. Cognition 79:89-113.

Langley, P. (in press). Artificial intelligence and cognitive systems. In P. Cohen (Ed.), AI: The First Hundred Years. Menlo Park, CA: AAAI Press.

Maes, P. 1989. How to do the right thing. *Connection Science* 1:291-323.

Minsky, M. (1985) The Society of Mind. New York: Simon and Schuster.

Newell, A. (1990) *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Negatu, A., T. L. McCauley, and S. Franklin. in review. Automatization for Software Agents.

Ornstein, R. (1986) Multimind. Boston: Houghton Mifflin.

Ramamurthy, U., D'Mello, S., and Franklin, S., (2004) Modified Sparse Distributed Memory as transient episodic memory for cognitive software agents. In *Proceedings of the International Conference on Systems, Man and Cybernetics*, Piscataway, NJ: IEEE.

Schacter, D. (1990) Toward a cognitive neuropsychology of awareness: implicit knowledge and anosagnosia. *Journal of Clinical and Experimental Neuropsychology*. 12 (1), 155-178.

Schneider, Walter and Chein, Jason M. (2003) Controlled & automatic processing: behavior, theory, and biological mechanisms. Cognitive Science, 27:525-559.

Shanahan, M P. 2006. A cognitive architecture that combines internal simulation with a global workspace. *Consciousness and Cognition* 15: 433-449.

Sloman, A. 1987. Motives Mechanisms Emotions. *Cognition* 1:217-234.

Sloman, A. 1999. What Sort of Architecture is Required for a Human-like Agent? In Foundations of Rational Agency, ed. M. Wooldridge, and A. Rao. Dordrecht, Netherlands: Kluwer Academic Publishers.

Sloman, A. and Chrisley, R. 2003. Virtual machines and consciousness. Journal of Consciousness Studies. 10:133-172.

Sun, R. (2003) A Tutorial on CLARION. http://www.cogsci.rpi.edu/~rsun/sun.tutorial.pdf

Standing, L. 1973. Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology* 25:207-222.

Varela, F. J., E. Thompson, and E. Rosch. 1991. *The Embodied Mind*. Cambridge, MA: MIT Press.

Yerkes, R. M., and J. D. Dodson. 1908. The Relationship of Strength of Stimulus to Rapidity of Habit Formation. *Journal of Comparative Neurology and Psychology* 18:459–482.