Chapter 7

The LIDA Model as a Foundational Architecture for AGI

Usef Faghihi and Stan Franklin

Computer Science Department & Institute for Intelligent Systems
The University of Memphis, Memphis, TN, 38152 USA

E-mail: {ufaghihi, franklin}@memphis.edu

Artificial intelligence (AI) initially aimed at creating "thinking machines," that is, computer systems having human level general intelligence. However, AI research has until recently focused on creating intelligent, but highly domain-specific, systems. Currently, researchers are again undertaking the original challenge of creating AI systems (agents) capable of human-level intelligence, or "artificial general intelligence" (AGI). In this chapter, we will argue that Learning Intelligent Distribution Agent (LIDA), which implements Baars' Global Workspace Theory (GWT), may be suitable as an underlying cognitive architecture on which others might build an AGI. Our arguments rely mostly on an analysis of how LIDA satisfies Sun's "desiderata for cognitive architectures" as well as Newell's "test for a theory of cognition." Finally, we measure LIDA against the architectural features listed in the BICA Table of Implemented Cognitive Architectures, as well as to the anticipated needs of AGI developers.

7.1 Introduction

The field of artificial intelligence (AI) initially aimed at creating "thinking machines," that is, creating computer systems having human level general intelligence. However, AI research has until recently mostly focused on creating intelligent, but highly domain-specific, computer systems. At present however, researchers are again undertaking the original challenge of creating AI systems (agents) capable of human-level intelligence, or "artificial general intelligence" (AGI).

To do so it may well help to be guided by following question, *how do minds work*? Among different theories of cognition, we choose to work from the Global Workspace

Theory (GWT) of Baars [1, 2] the most widely accepted psychological and neurobiological theory of the role of consciousness in cognition [3–6].

GWT is a neuropsychological theory of the role of consciousness in cognition. It views the nervous system as a distributed parallel system incorporating many different specialized processes. Various coalitions of these specialized processes facilitate making sense of the sensory data currently coming in from the environment. Other coalitions sort through the results of this initial processing and pick out items requiring further attention. In the competition for attention a winner emerges, and occupies what Baars calls the global workspace, the winning contents of which are presumed to be at least functionally conscious [7]. The presence of a predator, enemy, or imminent danger should be expected, for example, to win the competition for attention. However, an unexpected loud noise might well usurp consciousness momentarily even in one of these situations. The contents of the global workspace are broadcast to processes throughout the nervous system in order to recruit an action or response to this salient aspect of the current situation. The contents of this global broadcast enable each of several modes of learning. We will argue that Learning Intelligent Distribution Agent (LIDA) [8], which implements Baars' GWT, may be suitable as an underlying cognitive architecture on which to build an AGI.

The LIDA architecture, a work in progress, is based on the earlier IDA, an intelligent, autonomous, "conscious" software agent that does personnel work for the US Navy [9]. IDA uses locally developed artificial intelligence technology designed to model human cognition. IDA's task is to find jobs for sailors whose current assignments are about to end. She selects jobs to offer a sailor, taking into account the Navy's policies, the job's needs, the sailor's preferences, and her own deliberation about feasible dates. Then she negotiates with the sailor, in English via a succession of emails, about job selection. We use the word "conscious" in the functional consciousness sense of Baars' Global Workspace Theory [1, 2], upon which our architecture is based (see also [7]) .

7.2 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. The LIDA model has two faces, its science side and its engineering side.

LIDA's science side fleshes out a number of psychological and neuropsychological theories of human cognition including GWT [2], situated cognition 10, perceptual sym-

bol systems [11], working memory [12], memory by affordances [13], long-term working memory [14], and the H-CogAff architecture [15].

The LIDA architecture engineering side explores architectural designs for software agents that promise more flexible, more human-like intelligence within their domains. It employs several modules that are designed using computational mechanisms drawn from the "new AI." These include variants of the Copycat Architecture [16, 17], Sparse Distributed Memory [18, 19], the Schema Mechanism [20, 21], the Behavior Net [22], and the Subsumption Architecture [23]. However, an AGI developer using LIDA need make no commitment to any of these mechanisms. The computational framework of the LIDA architecture [24] allows free substitution of such mechanisms (see below). The required commitment is relatively modest, consisting primarily of a weak adherence to the LIDA cognitive cycle (see below). In addition, the LIDA architecture can accommodate the myriad features[1] that will undoubtedly be required of an AGI (see below). Thus the LIDA architecture, empirically based on psychological and biological principles, offers the flexibility to relatively easily experiment with different paths to an AGI. This makes us think of LIDA as eminently suitable for an underlying foundational architecture for AGI.

7.3 LIDA architecture

Any AGI will have to deal with tremendous amounts of sensory inputs. It will therefore need attention to filter the incoming sensory data to recruit resources in order to respond, and to learn. Note that this greatly resembles the Global Workspace Theory broadcast. By definition, every AGI must be able to operate in a wide variety of domains. It must therefore be capable of very flexible decision making. Flexible motivation, resulting from and modulated by feelings and emotions are in turn crucial to this end. The LIDA framework is setup accordingly. LIDA can be thought of as a proof of concept model for GWT. Many of the tasks in this model are accomplished by codelets [16] implementing the processors in GWT. Codelets are small pieces of code, each running independently. A class of codelets called attention codelets serves, with the global workspace, to implement attention. An attention codelet attempts to bring the contents of its coalition to the 'consciousness' spotlight. A broadcast then occurs, directed to all the processors in the system, to recruit resources with which to handle the current situation, and to learn.

[1]Here we distinguish between features of an architecture such as one of its main components (e.g., Sensory Processing) and features of, say, an object such as its colors.

The LIDA architecture is partly symbolic and partly connectionist with all symbols being grounded in the physical world in the sense of Brooks [25]. Thus the LIDA architecture is embodied. (For further information on situated or embodied cognition, please see [26–29]. LIDA performs through its cognitive cycles (Figure 7.1), which occur five to ten times a second [30, 31], and depend upon saliency determination by the agent. A cognitive cycle starts with a sensation and usually ends with an action. The cognitive cycle is conceived of as an active process that allows interactions between the different components of the architecture. Thus, cognitive cycles are always on-going.

We now describe LIDA's primary mechanisms.

1) Perceptual Associative Memory (PAM): This corresponds neurallyto the parts of different sensory cortices in humans (visual, auditory and somatosensory), plus parts of other areas (e.g. enterhinal cortex). PAM allows the agent to distinguish, classify and identify external and internal information. PAM is implemented in the LIDA architecture with a version of the slipnet [16]. There are connections between slipnet nodes. Segments of the slipnet are copied into the agent's Workspace as parts of the percept [32]. In LIDA, perceptual learning is learning to recognize new objects, new categorizations, and new relationships. With the conscious broadcast (Figure 7.1), new objects, categories, and the relationships among them and between them and other elements of the agent's ontology are learned by adding nodes (objects and categories) and links (relationships) to PAM. Existing nodes and links can have their base-level activations reinforced. The conscious broadcast begins and updates the process of learning to recognize and to categorize, both employing perceptual memory [8].

2) Workspace: This roughly corresponds to the human preconscious buffers of working memory [33]. This is the "place" that holds perceptual structures, which come from perception. It also includes previous percepts not yet decayed away, and local associations from episodic memories. These local associations are combined with the percepts to generate a Current Situational Model, the agent's understanding of what is going on right now. Information written in the workspace may reappear in different cognitive cycles until it decays away.

3) Episodic memories: These are memories for events (what, where and when). When the consciousness mechanism broadcasts information, it is saved into transient episodic memory (TEM) and is later consolidated into LIDA's declarative memory (DM) [34]. In LIDA, episodic learningrefers to the memory of events – the what, the where and the when [12, 35]. In the LIDA model such learned events are stored in transient episodic

memory [34, 36] and in the longer-term declarative memory [34]. Both are implemented using sparse distributed memory [18], which is both associative and content addressable, and has other characteristics that correspond to psychological properties of memory. In particular it knows when it doesn't know, and exhibits the tip of the tongue phenomenon. Episodic learning in the LIDA model is also a matter of generate and test, with such learning occurring at the conscious broadcast of each cognitive cycle. Episodic learning is initially directed only to transient episodic memory. At a later time and offline, the undecayed contents of transient episodic memory are consolidated [37] into declarative memory, where they still may decay away or may last a lifetime.

4) Attentional Memory (ATM): ATM is implemented as a collection of a particular kind of codelet called an attention codelet. All attention codelets are tasked with finding their own specific content in the Current Situational Model (CSM) of the Workspace. For example, one codelet may look for a node representing fear. When an attention codelet finds its content it creates a coalition containing this content and related content. The coalition is added to the Global Workspace to compete for consciousness. Each attention codelet has the following attributes: 1) *concern:* that content, whose presence in the CSM, can trigger the codelet to act; 2) a base-level activation, a measure of the codelet's usefulness in bringing information to consciousness, as well as its general importance; and 3) a current activation which measures the degree of intersection between its concern and the content of the current situational model. The total activation measures the current saliency of its concern. We use a sigmoid function to both reinforce and decay the base-level and the current activations. The ATM includes several kinds of attention codelets. The *default* attention codelet reacts to whatever content it finds in the Current Situational Model in the Workspace, trying to bring its most energetic content to the Global Workspace. *Specific attention* codelets are codelets that may have been learned. They bring particular Workspace content to the Global Workspace. *Expectation codelets*, created during action selection, attempt to bring the result (or non-result) of a recently-executed action to consciousness. *Intention codelets* are attention codelets that bring any content that can help the agent reach a current goal to consciousness. That is, when the agent makes a volitional decision, an intention codelet is generated. There are attention codelets that react to the various dimensions of saliency, including novelty informativeness, importance, insistance urgency and unexpectedness.Attentional learning is the learning of what to attend to [38, 39]. In the LIDA model attentional learning involves attention codelets, small processes whose job it is to focus the agent's attention on some particular portion of its internal model of the

current situation. Again, learning occurs with the conscious broadcast with new attention codelets being created and existing attention codelets being reinforced.

5) Procedural Memory, Action Selection and Sensory-motor Memory: LIDA's procedural memory deals with deciding what to do next. It is similar to Drescher's schema mechanism but with fewer parameters [20, 40]. The scheme net is a directed graph in which each of the nodes has a context, an action, and results. As a result of the conscious broadcast, schemes from Procedural Memory are instantiated and put into the Action Selection mechanism. The Action Selection mechanism then chooses an action and Sensory-Motor Memory executes the action (Figure 7.1). LIDA uses Maes' Behavior Network with some modifications [41] as its Action Selection mechanism [22]. Thus, in LIDA's architecture, while Procedural Memory and the Action Selection mechanism are responsible for deciding what will be done next, Sensory-Motor memory is responsible for deciding how tasks will be performed. Thus, each of these memory systems requires distinct mechanisms. In LIDA, procedural learning encodes procedures for possibly relevant behaviors into Procedural Memory (Figure 7.1). It is the learning of new actions and action sequences with which to accomplish new tasks. It is the learning of under what circumstances to perform new behaviors, or to improve knowledge of when to use existing behaviors. These procedural skills are shaped by reinforcement learning, operating by way of conscious processes over more than one cognitive cycle [8].

It must be noted that the LIDA model for the four aforementioned modes of learning, supports the instructionalist learning of new memory entities as well as the selectionist reinforcement of existing entities.

7.4 Cognitive architectures, features and the LIDA model

An AGI has to be built on some cognitive architecture, and by its nature, should share many features with other cognitive architectures. The most widely known cognitive architectures include Newell's Soar architecture [42–44], Anderson's ACT-R architecture [45–47], Sun's CLARION architecture [48], and Franklin's LIDA architecture [8]. The aforementioned cognitive architectures each have their strengths and weaknesses when it comes to defining a theory of mind [49, 50]. Some researchers have also tried to identify the most important features needed to construct biologically-inspired cognitive architectures (BICA). In particular, an AGI may well need the features listed by Sun (2004), and by Newell [51], as well as features from the BICA table [52]. The BICA table is not shown

in this paper because its size is beyond the size of this word document (readers can refer to the online version for full details). The LIDA model seems to have room for all these features. In the next sections, we describe Sun's desiderata and Newell's functional criteria for cognitive architectures and, in italics, the LIDA model's features that correspond to each of these cognitive architecture criteria. An assessment of LIDA against the features from the BICA table follows.

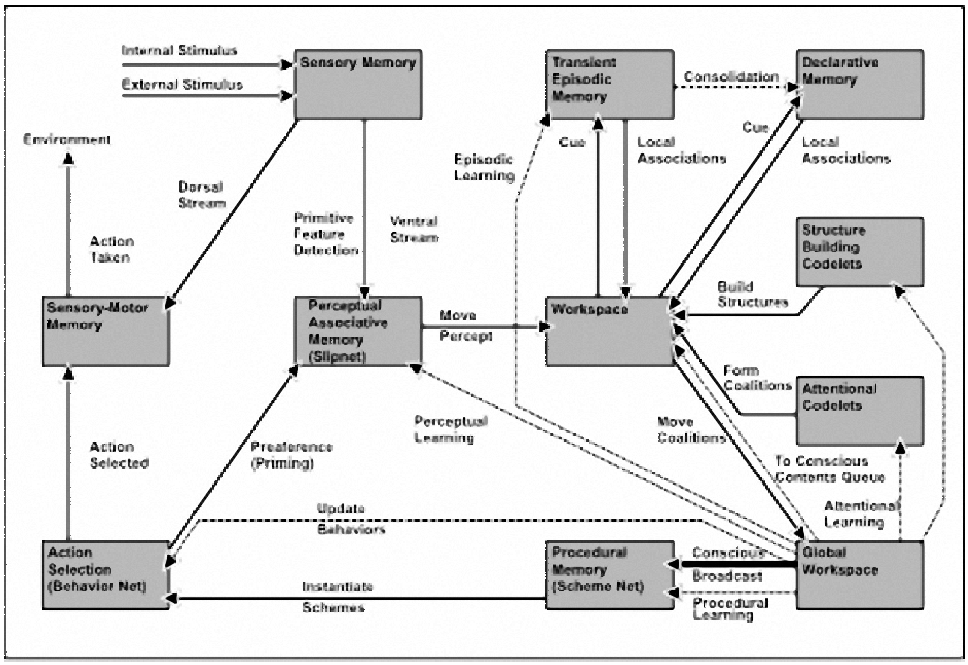

Fig. 7.1 LIDA's Architecture

7.4.1 *Ron Sun's Desiderata* [53]

In his article, Sun "proposes a set of essential desiderata for developing cognitive architectures," and argues "for the importance of taking into full consideration these desiderata in developing future architectures that are more cognitively and ecologically realistic." Though, AGI is not explicitly mentioned – the article predates the use of the term "AGI" – one could infer that Sun would consider them desiderata for an AGI as well. Here are some of his desiderata interspersed with our assessment of how well the LIDA model does, or does not, achieve them.

Ecological realism: "Taking into account everyday activities of cognitive agents in their natural ecological environments."

LIDA does allow ecological realism. It was designed for the everyday life of animals and/or artificial agents. Animals typically respond to incoming stimuli, often in search of food, mates, safety from predators, etc. Doing so requires frequent sampling of the environment, interpreting stimuli, attending to the most salient, and responding appropriately. The cascading LIDA cognitive cycles provides exactly this processing.

Bio-evolutionaryrealism supplements ecological realism. This feature refers to the idea that intelligences in species are on a continuum, and that cognitive models of human intelligence should not be limited to strictly human cognitive processes, but rather should be reducible to models of animal intelligence as well.

The LIDA model integrates diverse research on animal cognition. It has been used to provide both an ontology of concepts and their relations, and a working model of an animal's cognitive processes [54, 55]. *In addition to helping to account for a broad range of cognitive processes, the LIDA model can help to comparatively assess the cognitive capabilities of different animal species.*

Cognitive realism. Cognition is highly variable within species. Cognitive realism refers to the idea that cognitive architectures should seek to replicate only essential characteristics of human cognition.

The reader is referred to the "Why the LIDA model may be suitable for AGI" section above to see that this condition is satisfied.

Eclecticism of methodologies and techniques. Adhering too promptly to any methodology or paradigm will only prevent the creation of new or improved cognitive architectures. It is best to take a more broad-based approach.

Again the reader is referred to the "Why the LIDA model may be suitable for AGI" section.

Reactivity. Reactivity refers to fixed responses to given stimuli (as opposed to full-fledged analyses of stimuli to select a response) that characterize many human behaviors [15].

The LIDA model's reactive part is setup as a relatively direct connection between incoming sensory data and the outgoing actions of effectors. At the end of every cognitive cycle LIDA selects an appropriate behavior in response to the current situation. This consciously mediated, but unconscious, selection is always reactive.

Sequentiality. Sequentiallity refers to the chronological nature of human everyday activities.

In LIDA, cognitive cycles allow the agent to perform its activities in a sequential manner. The cycles can cascade. But, these cascading cognitive cycles must preserve the sequentiality of the LIDA agent's stream of functional consciousness, as well as its selection of an action in each cycle [56].

Routineness. Routineness refers to the fact that humans' every day behaviors are made of routines, which are constantly and smoothly adapting to the changing environment.

The LIDA model is equipped with both reactive and deliberative high-level cognitive processes. Such processes become routine when they are incorporated (learned) into LIDA's procedural memory as schemes representing behavior streams.

Trial-and-error adaptation. Trial-and-error adaptation refers to the trial-and-error process through which humans learn and develop reactive routines.

LIDA learns from experience, which may yield several lessons over several cognitive cycles. Such lessons include newly perceived objects and their relationship to already known objects and categories, relationships among objects and between objects and actions, effects of actions on sensation, and improved perception of sensory data. All of LIDA's learning be it, perceptual, episodic, or procedural, is very much trial and error (generate and test as AI researchers would say). LIDA is profligate in its learning, with new entities and reinforcement of existing entities learned with every broadcast. Those that are sufficiently reinforced (tested) remain. The others decay away as "errors."

7.4.2 *Newell's functional criteria (adapted from Lebiere and Anderson 2003)*

Newell proposed multiple criteria that a human cognitive architecture should satisfy in order tobe functional [57, 58]. Lebiere and Anderson [51] combined his two overlapping lists into the twelve criteria, phrased as questions, listed below. Each criterion described will be followed by an analysis of how LIDA does, or does not, satisfy it.

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

This criterion demands flexibility of action selection. In LIDA, motivation for actions, learning and perceiving, come from feelings and emotions. These provide a much more flexible kind of motivation for action selection than do drives, causations or rules, produc-

ing more flexible action selection. In LIDA, various types of learning, including learning to recognize or perform procedures, also contribute to flexible behavior. LIDA's sophisticated action selection itself allows such flexibility as switching back and forth between various tasks. LIDA is flexible in what to attend to, at any given time, increasing the flexibility of action selection as well. We suspect that LIDA cannot do anything that can be done with a Turing machine; it is not computationally universal. We also suspect that this is not necessary for AGI.

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

LIDA's cognitive cycles individually take approximately 300 ms, and theysample the environment cascading at roughly five to ten times per-second [59]. *There is considerable empirical evidence from neuroscience suggestive of, and consistent with, such cognitive cycling in humans* [60–66]. *An earlier software agent, IDA (see above), based on the LIDA architecture, found new billets for sailors in about the same time as it took a human "detailer"* [59].

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

In the LIDA model, feelings and emotions play important role in decision making. The LIDA model can feature both the affective and rational human-inspired models of decision making [67]. *LIDA's predecessor IDA, controlled by much the same architecture, was quite functional* [68], *promising the same for LIDA controlled agents.*

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?

In the LIDA model, selective attention filters potentially large amounts of incoming sensory data. Selective attention also provides access to appropriate internal resources that allow the agent to select appropriate actions and to learn from vast amounts of data produced during interactions in a complex environment. The model has perceptual, episodic, attentional and procedural memory for the long term storage of various kinds of information.

Knowledgeable in terms of variety: Does the agent integrate diverse knowledge? Is it capable of common examples of intellectual combination?

A major function of LIDA's preconscious Workspace is precisely to integrate diverse knowledge in the process of updating the Current Situational Model from which the con-

tents of consciousness is selected. Long-term sources of this knowledge include Perceptual Associative Memory and Declarative Memory. There are several sources of more immediate knowledge that come into play. LIDA is, in principle, "capable of common examples of intellectual combination," though work has only begun on the first implemented LIDA based agent promising such capability.

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?

LIDA's predecessor, IDA, was developed for the US Navy to fulfill tasks performed by human resource personnel called detailers. At the end of each sailor's tour of duty, he or she is assigned to a new billet. This assignment process is called distribution. The Navy employs almost 300 full time detailers to effect these new assignments. IDA's task is to facilitate this process, by automating the role of detailer. IDA was tested by former detailers and accepted by the Navy [68].

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

IDA communicates with sailors by email in unstructured English. However, we think of this capability as pseudo-natural-language, since it is accomplished only due to the relatively narrow domain of discourse. Language comprehension and language production are high-level cognitive processes in humans. In the LIDA model, such higher-level processes are distinguished by requiring multiple cognitive cycles for their accomplishment. In LIDA, higher-level cognitive processes can be implemented by one or more behavior streams; that is, streams of instantiated schemes and links from procedural memory. Thus LIDA should, in principle, be capable of natural language understanding and production. In practice, work on natural language has just begun.

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

Researchers in, philosophy, neuroscience and psychology postulate various forms of a "self" in humans and animals. All of these selves seem to have a basis in some form of consciousness. GWT suggests that a self-system can be thought of "... as the dominant context of experience and action." [2] *. Following others (see below) the various selves in an autonomous agent may be categorized into three major components, namely:* 1) *the Proto-Self;* 2) *the Minimal (Core) Self; and* 3) *the Extended Self. The LIDA model provides*

for the basic building blocks from which to implement the various parts of a multi-layered self-system as hypothesized by philosophers, psychologists and neuroscientists [69]. *In the following, we discuss each component, and their sub-selves, very briefly* (*for more detail see* [69]).

1) The Proto-Self: *Antonio Damasio conceived the Proto-self as a short-term collection of neural patterns of activity representing the current state of the organism* [70]. *In LIDA, the Proto-self is implemented as the set of global and relevant parameters in the various modules including the Action Selection and the memory systems, and the underlying computer system's memory and operating system;* ***2) the Minimal (Core) Self:*** *The minimal or core self* [71] *is continually regenerated in a series of pulses, which blend together to give rise to a continuous stream of consciousness. The Minimal Self can be implemented as sets of entities in the LIDA ontology, that is, as computational collections of nodes in the slipnet of LIDA's perceptual associative memory; and* ***3) the Extended Self:*** *The extended self consists of* (a) *the autobiographical self,* (b) *the self-concept,* (c) *the volitional or executive self, and* (d) *the narrative self. In human beings, the autobiographical self develops directly from episodic memory. In the LIDA model, the autobiographical self can be described as the local associations from transient episodic memory and declarative memory which come to the workspace in every cognitive cycle. The self-concept consists of enduring self-beliefs and intentions, particularly those relating to personal identity and properties. In the LIDA model, the agent's beliefs are in the semantic memory and each volitional goal has an intention codelet. The volitional self provides executive function. In the LIDA model, deliberate actions are implemented by behavior streams. Thus, LIDA has a volitional self. The narrative self is able to report actions, intentions, etc., sometimes equivocally, contradictorily or self-deceptively. In the LIDA model, feeling, motivation, and attention nodes play a very important role in the Narrative Self. That is, after understanding a self-report request, the LIDA model could, in principle, generate a report based on its understanding of such a request.*

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

LIDA is equipped with perceptual, episodic, procedural, attentional learning, all modulated by feelings and emotions. As humans do, LIDA learns continually and implicitly with each conscious broadcast in each cognitive cycle.

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

Since LIDA learns as humans do, we would expect a LIDA controlled agent to go through a developmental period of rapid learning as would a child. The work on replicating data from developmental experiments has just begun.

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

Since LIDA is attempted to model humans and other animals, presumably the model should be evolvable and comparative, at least in principle.

Be realizable within the brain: Do the components of the theory exhaustively map onto brain processes?

Shanahan [6] *devotes two chapters to a compelling argument that the brain is organized so as to support a conscious broadcast. LIDA is beginning to build on this insight, using the work of Freeman and colleagues* [72]*, to create a non-linear dynamical systems bridge between LIDA and the underlying brain. Whether this bridge will lead to an exhaustive mapping is not at all clear as yet.*

7.4.3 *BICA table*

Any AGI is likely to be produced by a very diverse collection of cognitive modules and their processes. There is a computational framework for LIDA [24] that requires only a modest commitment to the underlying assumptions of the LIDA architecture. One can introduce into LIDA's framework a large variety of differently implemented modules and processes so that, many possible AGI architectures could be implemented from the LIDA framework. One advantage of doing it in this way is that all of these AGI architectures implemented on the top of the LIDA's framework, would use a common ontology based on the LIDA model as presented to AGI 2011 [24].

In the following, we will give an assessment of the LIDA model against the features of the BICA Table of Implemented Cognitive Architectures [52]. Column 1 of the BICA Table contains a list of features proposed by developers of cognitive architectures to be at least potentially useful, if not essential, for the support of an AGI. Subsequent columns are devoted to individual cognitive architectures with a cell describing how its column architecture addresses its row feature. The rest of this section is an expansion of the column devoted to LIDA in the BICA table.

Note that all the **Basic overview** features listed in the BICA's first column are detailedearlier in this chapter. We will discuss the rest of the features in the following:

Support for Common Components: *The LIDA model supports all features mentioned in this part such as episodic and semantic memories. However, the auditory mechanism is not implemented in a LIDA-based agent as yet.*

Support for Common Learning Algorithms: *The LIDA model supports different types of learning such as episodic, perceptual, procedural, and attentional learning. However, the Bayesian Update and Gradient Descent Methods* (*e.g., Backpropagation*) *are not implemented in a LIDA-based agent.*

Common General Paradigms Modeled: *The LIDA model supports features listed in this part such as decision making and problem solving. However, perceptual illusions, meta-cognitive tasks, social psychology tasks, personality psychology tasks, motivational dynamics are not implemented in a LIDA-based agent.*

Common Specific Paradigms Modeled columns: 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.

Although the Common Specific Paradigms Modeled features listed above are not implemented in LIDA, in principle LIDA is capable of implementing each of them. For instance, a LIDA-based agent is replicating some attentional tasks à la Van Bockstaele's and his colleagues [73].

Meta-Theoretical Questions:

1) Uses only local computations? *Yes, throughout the architecture with the one exception of the conscious broadcast which is global;*
2) Unsupervised learning? *Yes. The LIDA model supports four different modes of learning, perceptual, episodic, attentional and procedural;*
3) Supervised learning? *While in principle possible for a LIDA agent, supervised learning per se is not part of the architecture;*
4) Can it learn in real time? *Yes* (*see above*)*;*
5) Can it do fast stable learning; i.e., adaptive weights converge on each trial without forcing catastrophic forgetting? *Yes. One shot learning in several modes occurs with the conscious broadcast during each cognitive cycle. With sufficient affective support and/or sufficient repeated attention, such learning can be quite stable;*

6) Can it function autonomously? *Yes. A LIDA-based agent can, in principle,operate machines and drive vehicles autonomously;*
7) Is it general-purpose in its modality; i.e., is it brittle? *A LIDA-based agent can, in principle, be developed to be general purpose and robust in real world environments;*
8) Can it learn from arbitrarily large databases; i.e., not toy problems? *Yes, this question is already answered in the previous sections;*
9) Can it learn about non-stationary databases; i.e., environmental rules change unpredictably? *Yes, a LIDA-based agent is, in principle, capable of working properly in an unpredictable environment;*
10) Can it pay attention to valued goals? *Yes, already explained earlier in this chapter;*
11) Can it flexibly switch attention between unexpected challenges and valued goals? *Yes. A LIDA-based agent attends to what is most salient based on its situational awareness;*
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? *In principle, yes, but it has yet to be implemented in particular domains with multiple senses.*

7.5 Discussion, Conclusions

In this chapter, we argue that LIDA may be suitable as an underlying cognitive architecture on which others might build an AGI. Our arguments rely mostly on an analysis of how LIDA satisfies Sun's "desiderata for cognitive architectures" as well as Newell's "test for a theory of cognition." We also measured LIDA against the architectural features listed in the BICA Table of Implemented Cognitive Architectures, as well as to the anticipated needs of AGI developers.

As can be seen in Section 7.4 above, the LIDA model seems to meet all of Sun's "... essential desiderata for developing cognitivearchitectures," and Newell's criteria that a human cognitive architecture should satisfy in order tobe functional. In addition, the LIDA architecture seems to be able, at least in principle, of incorporating each of the features listed in the BICA Table of Implemented Cognitive Architectures. Thus the LIDA architecture would seem to offer the requisite breadth of features.

The LIDA computational framework offers software support for the development of LIDA based software agents, as well as LIDA based control systems for autonomous

robots [24]. As described in Section 7.2 above, developing an AGI based loosely on the LIDA architecture requires only a modest commitment. Higher-level cognitive processes such as reasoning, planning, deliberation, etc., must be implemented by behavior streams, that is, using cognitive cycles. But this is not a strong commitment, since, using the LIDA computational framework, any individual module in LIDA's cognitive cycle can be modified at will, or even replaced by another designed by the AGI developer. Thus we are left with the contention that various AGI systems can effectively be developed based loosely on the LIDA architecture and its computational framework. Such systems would lend themselves to relatively easy incremental improvements by groups of developers and, due to their common foundation and ontology, would also allow relatively straightforward testing and comparison. Thus the LIDA architecture would seem to be an ideal starting point for the development of AGI systems.

Bibliography

[1] B.J. Baars. *In the Theater of Consciousness: The Workspace of the Mind.* Oxford: Oxford University Press (1997).
[2] B.J. Baars. *A cognitive theory of consciousness.* Cambridge: Cambridge University Press (1988).
[3] B.J. Baars. The conscious access hypothesis: origins and recent evidence. *Trends in Cognitive Science* 47–52 (2002).
[4] S. Dehaene & L. Naccache. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* **79**, 1–37 (2001).
[5] N. Kanwisher. Neural events and perceptual awareness. *Cognition* **79:89**, 89–113 (2001).
[6] M. Shanahan. *Embodiment and the Inner Life.* Oxford: Oxford University Press (2010).
[7] S. Franklin. IDA: A Conscious Artifact? *Journal of Consciousness Studies* **10**, 47–66 (2003).
[8] S. Franklin & F.G.J. Patterson. 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* (2006).
[9] S. Franklin, A. Kelemen & L. McCauley. IDA: A Cognitive Agent Architecture. *IEEE Conf. on Systems, Man and Cybernetics*, 2646–2651 (1998).
[10] F.J. Varela, E. Thompson & E. Rosch. *The embodied mind: Cognitive Science and Human Experience.* MIT Press, Cambridge, MA, USA (1991).
[11] L.W. Barsalou. *Perceptual Symbol Systems.* Vol. **22** (MA: The MIT Press, 1999).
[12] A.D. Baddeley. The episodic buffer: a new component of working memory? *Trends in Cognitive Science* **4**, 417–423 (2000).
[13] A.M. Glenberg. What memory is for. *Behavioral and Brain Sciences*, 1–19 (1997).
[14] K.A. Ericsson & W. Kintsch. Long-term working memory. *Psychological Review* **102**, 21–245 (1995).
[15] A. Sloman. 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* (1999).

[16] D. Hofstadter, R & M. Mitchell. 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, N.J. Norwood: Ablex.* (1994).
[17] J. Marshall. Metacat: A self-watching cognitive architecture for analogy-making. *Proceedings of the 24th Annual Conference of the Cognitive Science Society* 631–636 (2002).
[18] P. Kanerva. *Sparse Distributed Memory.* Cambridge MA: The MIT Press (1988).
[19] R.P.N. Rao & O. Fuentes. Hierarchical Learning of Navigational Behaviors in an Autonomous Robot using a Predictive Sparse Distributed Memory. *Machine Learning* **31**, 87–113 (1998).
[20] G.L. Drescher. *Made-Up Minds: A Constructivist Approach to Artificial Intelligence.* Cambridge, MA: MIT Press (1991).
[21] H.H. Chaput, B. Kuipers & R. Miikkulainen. Constructivist Learning: A Neural Implementation of the Schema Mechanism. *Workshop for Self-Organizing Maps, Kitakyushu, Japan* (2003).
[22] P. Maes. How to do the right thing. *Connection Science* **1**, 291–323 (1989).
[23] R.A. Brooks. Intelligence without Representation. *Artificial intelligence. Elsevier* (1991).
[24] J. Snaider, R. McCall & S. Franklin. The LIDA Framework as a General Tool for AGI. *Paper presented at the The Fourth Conference on Artificial General Intelligence, Mountain View, California, USA* (2011).
[25] R.A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation* **2**, pp. 14–23 (1986).
[26] M.L. Anderson. Embodied Cognition: A Field Guide. *Artificial Intelligence* **149**, 91–130 (2003).
[27] T. Ziemke, J. Zlatev & R. M. Frank. Body, *Language and Mind: Volume 1: Embodiment.* (Mouton de Gruyter, 2007).
[28] S. Harnad. The Symbol Grounding Problem. *Physica D* **42**, 335–346 (1990).
[29] M. de Vega, A. Glenberg & A. Graesser. *Symbols and Embodiment:Debates on meaning and cognition.* Oxford: Oxford University Press (2008).
[30] B.J. Baars & S. Franklin. How conscious experience and working memory interact. *Trends in Cognitive Sciences* **7** (2003).
[31] S. Franklin, B.J. Baars, U. Ramamurthy & M. Ventura. The Role of Consciousness in Memory. *Brains, Minds and Media* **1**, *bmm150* (*urn:nbn:de:0009-3-1505*) (2005).
[32] R. McCall, S. Franklin & D. Friedlander. Grounded Event-Based and Modal Representations for Objects, Relations, Beliefs, Etc. *Paper presented at the FLAIRS-23, Daytona Beach, FL* (2010).
[33] A.D. Baddeley. Working memory and conscious awareness. *In Theories of memory*, (eds. Alan Collins, S. Gathercole, M. A Conway, & P. Morris) 11–28 (Erlbaum, 1993).
[34] S. Franklin. Cognitive Robots: Perceptual associative memory and learning. *In Proceedings of the 14th Annual International Workshop on Robot and Human Interactive Communication* (2005).
[35] E. Tulving. *Elements of Episodic Memory.* New York: Oxford University Press (1983).
[36] M.A. Conway. Sensory-perceptual episodic memory and its context: autobiographical memory. *In Episodic Memory, ed. A. Baddeley, M. Conway, and J. Aggleton. Oxford: Oxford University Press* (2002).
[37] R. Stickgold & M.P. Walker. Memory consolidation and reconsolidation: what is the role of sleep? *Trends Neurosci* **28**, 408–415 (2005).
[38] W.K. Estes. *Classification and Cognition.* Oxford: Oxford University Press (1993).
[39] Z. Vidnyánszky & W. Sohn. Attentional learning: learning to bias sensory competition. *Journal of Vision* **3** (2003).
[40] G.L. Drescher. Learning from Experience Without Prior Knowledge in a Complicated World. *Proceedings of the AAAI Symposium on Parallel Models. AAAI Press* (1988).

[41] A. Negatu & S. Franklin. An action selection mechanism for 'conscious' software agents. *Cognitive Science Quarterly* **2**, 363–386 (2002).
[42] P. Rosenbloom, J. Laird & A. Newell. *The Soar Papers: Research on Integrated Intelligence.* Cambridge, Massachusetts: MIT Press (1993).
[43] J.E. Laird, A. Newell & P.S. Rosenbloom. Soar: an architecture for general intelligence. *Artificial Intelligence* **33**, 1–64 (1987).
[44] J.F. Lehman, J.E. Laird & P.S. Rosenbloom. A gentle introduction to Soar, an architecture for human cognition. *In Invitation to Cognitive Science Methods, Models, and Conceptual Issues,* Vol. **4** (eds. S. Sternberg & D. Scarborough) (MA: MIT Press, 1998).
[45] J.R. Anderson. *Rules of the mind.* (Mahwah, NJ: Lawrence Erlbaum Associates, 1993).
[46] J.R. Anderson. *The Architecture of Cognition.* Cambridge, MA: Harvard University Press (1983).
[47] J.R. Anderson, D. Bothell, M.D. Byrne, S. Douglass, C. Lebiere & Y. Qin. An integrated theory of the mind. *Psychological Review* **111**, 1036–1060 (2004).
[48] R. Sun. *The CLARION cognitive architecture: Extending cognitive modeling to social simulation Cognition and Multi-Agent interaction.* Cambridge University Press, New York (2006).
[49] U. Faghihi. *The use of emotions in the implementation of various types of learning in a cognitive agent.* Ph.D thesis, University of Quebec at Montreal (UQAM), (2011).
[50] D. Vernon, G. Metta & G. Sandini. A Survey of Artificial Cognitive Systems: Implications for the Autonomous Development of Mental Capabilities in Computational Agents. *IEEE Transactions on Evolutionary Computation, Special Issue on Autonomous Mental Development* **11**, 151–180 (2007).
[51] J.R. Anderson & C. Lebiere. The Newell Test for a theory of cognition. *Behavioral And Brain Sciences* **26** (2003).
[52] A.V. Samsonovich. Toward a Unified Catalog of Implemented Cognitive Architectures. *Proceeding of the 2010 Conference on Biologically Inspired Cognitive Architectures*, 195–244 (2010).
[53] R. Sun. Desiderata for cognitive architectures. *Philosophical Psychology* **17**, 341–373 (2004).
[54] S. Franklin & M.H. Ferkin. An Ontology for Comparative Cognition: a Functional Approach. *Comparative Cognition & Behavior Reviews* **1**, 36–52 (2006).
[55] S. D'Mello & S. Franklin. A cognitive model's view of animal cognition. *Cognitive models and animal cognition. Current Zoology.* (in press) (2011).
[56] J. Snaider, R. McCall & S. Franklin. Time production and representation in a conceptual and computational cognitive model. *Cognitive Systems Research.* (in press).
[57] A. Newell. *Unified Theory of Cognition.* Cambridge, MA: Harvard University Press (1990).
[58] A. Newell. Precis of Unified theories of cognition. *Behavioral and Brain Sciences* (1992).
[59] T. Madl, B.J. Baars & S. Franklin. The Timing of the Cognitive Cycle. *PLoS ONE* (2011).
[60] S. Doesburg, J. Green, J. McDonald & L. Ward. Rhythms of consciousness: binocular rivalry reveals large-scale oscillatory network dynamics mediating visual perception. *PLoS One. 4: c6142* (2009).
[61] W. Freeman. The limbic action-perception cycle controlling goal-directed animal behavior. *Neural Networks* **3**, 2249–2254 (2002).
[62] J. Fuster. Upper processing stages of the perception-action cycle. *Trends in Cognitive Sciences* **8**, 143–145 (2004).
[63] M. Massimini, F. Ferrarelli, R. Huber, S.K. Esser & H. Singh. Breakdown of Cortical Effective Connectivity During Sleep. *Science* **309** (2005).
[64] M. Sigman & S. Dehaene. Dynamics of the Central Bottleneck: Dual-Task and Task Uncertainty. *PLoS Biol.* **4** (2006).
[65] N. Uchida, A. Kepecs & Z. F. Mainen. Seeing at a glance, smelling in a whiff: rapid forms of perceptual decision making. *Nature Reviews Neuroscience* **7**, 485–491 (2006).

[66] J. Willis & A. Todorov. First Impressions: Making Up Your Mind After a 100-Ms Exposure to a Face. *Psychological Science* **17**, 592–599 (2006).

[67] W. Wallach, S. Franklin & C. Allen. *In Topics in Cognitive Science, special issue on Cognitive Based Theories of Moral Decision Making* (eds. W. Wallach & S. Franklin) 454–485 (Cognitive Science Society, 2010).

[68] L. McCauley & S. Franklin. A Large-Scale Multi-Agent System for Navy Personnel Distribution. *Connection Science* **14**, 371–385 Comments: special issue on agent autonomy and groups. (2002).

[69] U. Ramamurthy & S. Franklin. Self System in a model of Cognition. *Proceedings of Machine Consciousness Symposium at the Artificial Intelligence and Simulation of Behavior Convention (AISB'11), University of York, UK*, 51–54 (2011).

[70] A.R. Damasio. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness.* New York: Harcourt Inc (1999).

[71] S. Gallagher. Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Science* **4**, 14–21 (2000).

[72] C. Skarda & W.J. Freeman. How Brains Make Chaos in Order to Make Sense of the World. *Behavioral and Brain Sciences* **10**, 161–195 (1987).

[73] B. Van Bockstaele, B. Verschuere, J.D. Houwer & G. Crombez. On the costs and benefits of directing attention towards or away from threat-related stimuli: A classical conditioning experiment. *Behaviour Research and Therapy* **48**, 692–697 (2010).