Neurocomputing ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

World survey of artificial brains, Part II: Biologically inspired cognitive architectures

Ben Goertzel [a,b,*], Ruiting Lian [b], Itamar Arel [c], Hugo de Garis [b], Shuo Chen [b]

[a] Novamente LLC, 1405 Bernerd Place, Rockville, MD 20851, USA
[b] Fujian Key Lab of the Brain-like Intelligent Systems, Xiamen University, Xiamen, China
[c] Machine Intelligence Lab, Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, USA

ARTICLE INFO

Keywords:
Artificial brains
Cognitive architectures

ABSTRACT

A number of leading cognitive architectures that are inspired by the human brain, at various levels of granularity, are reviewed and compared, with special attention paid to the way their internal structures and dynamics map onto neural processes. Four categories of Biologically Inspired Cognitive Architectures (BICAs) are considered, with multiple examples of each category briefly reviewed, and selected examples discussed in more depth: primarily symbolic architectures (e.g. ACT-R), emergentist architectures (e.g. DeSTIN), developmental robotics architectures (e.g. IM-CLEVER), and our central focus, hybrid architectures (e.g. LIDA, CLARION, 4D/RCS, DUAL, MicroPsi, and OpenCog). Given the state of the art in BICA, it is not yet possible to tell whether emulating the brain on the architectural level is going to be enough to allow rough emulation of brain function; and given the state of the art in neuroscience, it is not yet possible to connect BICAs with large-scale brain simulations in a thoroughgoing way. However, it is nonetheless possible to draw reasonably close function connections between various components of various BICAs and various brain regions and dynamics, and as both BICAs and brain simulations mature, these connections should become richer and may extend further into the domain of internal dynamics as well as overall behavior.

1. Introduction

In Part I of this paper we reviewed the leading large-scale brain simulations—systems that attempt to simulate, in software, the more or less detailed structure and dynamics of particular subsystems of the brain. We now turn to the other kind of "artificial brain" that is also prominent in the research community: "Biologically Inspired Cognitive Architectures," also known as BICAs, which attempt to achieve brainlike functionalities via emulating the brain's high-level architecture without necessarily simulating its low-level specifics.

The term BICA became commonplace via a 2005 DARPA funding program, but the concept is as old as the AI field. While no rigid demarcation separates Biologically Inspired Cognitive Architectures from Cognitive Architectures in general, the term is intended to distinguish cognitive architectures drawing significant direct inspiration from the brain, from those based exclusively (or nearly so) on models of mind. The line has become blurry in interesting new ways of late because some of the traditional mind-inspired cognitive architectures have begun to explore biological analogies to their functioning (ACT-R being one dramatic example, to be discussed below).

The line between BICAs and large-scale brain simulations is a little clearer. A brain simulation is intended to display not only closely similar functions to a brain or part thereof, but also closely similar internal structures and dynamics. A BICA is intended to display loosely similar functions to a brain, based on internal structures that are conceptually inspired by the brain (and not just the mind) but not necessarily extremely similar to the brain. One would not expect to be able to compare data drawn from the *internals* of a BICA and compare it, point for point, with data drawn from neurological instrumentation.

There are many BICAs out there and the reader may wish to peruse the proceedings of the 2008 and 2009 AAAI BICA symposia (http://members.cox.net/bica2009/). Here we review only a small but representative sampling. Also, the reader who desires a more thorough review of cognitive architectures, including BICAs and others, is referred to Duch's review paper from the AGI-08 conference [1]. Finally, another valuable resource is the spreadsheet organized and hosted by Alexei Samsonovich at http://members.cox.net/bica2009/cogarch/, which compares a number of cognitive architectures in terms of a feature checklist, and was created collaboratively with the creators of the architectures.

Duch, in his survey of cognitive architectures [1], divides existing approaches into three paradigms – symbolic, emergentist

* Corresponding author at: Novamente LLC, 1405 Bernerd Place, Rockville, MD 20851, USA.
E-mail addresses: ben@goertzel.org, bengoertzel@gmail.com (B. Goertzel).

0925-2312/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2010.08.012

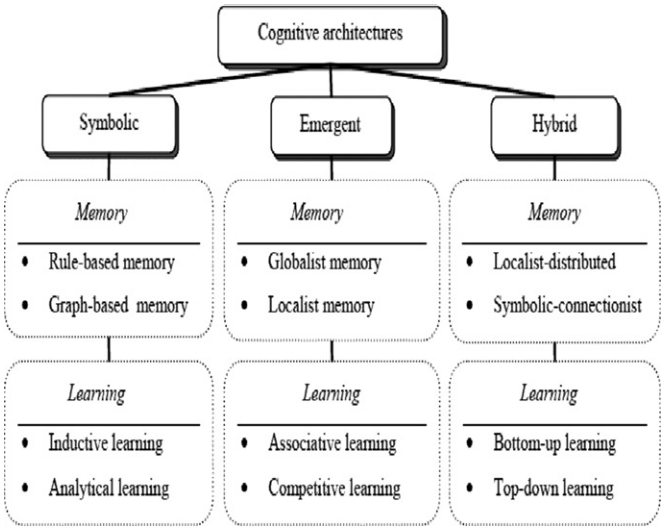


Fig. 1. Duch's simplified taxonomy of cognitive architectures.

and hybrid – as broadly indicated in Fig. 1 [1]. Drawing on his survey and updating slightly, here we give some key examples of each. We place little focus on the symbolic paradigm here because by and large these are not BICAs, but rather psychology-inspired cognitive architectures. However, we still need to pay symbolic cognitive architectures some attention because they are related to the hybrid architectures, which often do fall into the BICA category.

2. Symbolic architectures versus BICAs

A venerable tradition in AI focuses on the physical symbol system hypothesis [2], which states that minds exist mainly to manipulate symbols that represent aspects of the world or themselves. A physical symbol system has the ability to input, output, store and alter symbolic entities, and to execute appropriate actions in order to reach its goals. Generally, symbolic cognitive architectures focus on "working memory" that draws on long-term memory as needed, and utilize a centralized control over perception, cognition and action. Although in principle such architectures could be arbitrarily capable (since symbolic systems have universal representational and computational power, in theory), in practice symbolic architectures tend to be weak in learning, creativity, procedure learning, and episodic and associative memory. Decades of work in this tradition has not resolved these issues, which has led many researchers to explore other options—such as BICAs.

A few of the more important symbolic cognitive architectures are as follows:

- **SOAR** [3], a classic example of expert rule-based cognitive architecture designed to model general intelligence. It has recently been extended to handle sensorimotor functions, though in a somewhat cognitively unnatural way; and is not yet strong in areas such as episodic memory, creativity, handling uncertain knowledge, and reinforcement learning.
- **ACT-R** [4] is fundamentally a symbolic system, but in his review Duch classifies it as a hybrid system because it incorporates connectionist-style activation spreading in a significant role; and there is an experimental thoroughly connectionist implementation to complement the primary mainly symbolic implementation. Its combination of SOAR-style "production rules" with large-scale connectionist dynamics allows it to simulate a variety of human psychological phenomena, but abstract reasoning, creativity, and transfer learning are still missing. An article summarizing the strengths and shortcomings of ACT-R appeared recently in BBS [5].
- **EPIC** [6], a cognitive architecture aimed at capturing human perceptual, cognitive and motor activities through several interconnected processors working in parallel. The system is controlled by production rules for cognitive processor and a set of perceptual (visual, auditory, tactile) and motor processors operating on symbolically coded features rather than raw sensory data. It has been connected to SOAR for problem solving, planning and learning.
- **ICARUS** [7], an integrated cognitive architecture for physical agents, with knowledge specified in the form of reactive skills, each denoting goal-relevant reactions to a class of problems. The architecture includes a number of modules: a perceptual system, a planning system, an execution system, and several memory systems. Concurrent processing is absent, attention allocation is fairly crude, and uncertain knowledge is not thoroughly handled.
- **SNePS** (Semantic Network Processing System) [8] is a logic, frame and network-based knowledge representation, reasoning, and acting system that has undergone over three decades of development. While it has been used for some interesting prototype experiments in language processing and virtual agent control, it has not yet been used for any large-scale or real-world application.

While these architectures contain many valuable ideas and have yielded some interesting results, there is no clear consensus in the field regarding whether such systems will ever be capable *on their own* of giving rise to the emergent structures and dynamics required to yield humanlike general intelligence using feasible computational resources. Currently, there seems to be more of a trend toward incorporating components from symbolic architectures in hybrid architectures, rather than taking a purely symbolic approach.

As our focus here is on BICAs rather than cognitive architectures in general, we choose ACT-R as our example to review here in detail, because this is the symbolic architecture that has been most closely tied in with human brain structure and function.

2.1. ACT-R

ACT-R is defined in terms of declarative and procedural knowledge, where procedural knowledge takes the form of production rules (IF-THEN rules, somewhat similar to in classical expert systems), and declarative knowledge takes the form of "chunks" formed by combining previous rules into new ones. It contains a variety of mechanisms for learning new rules and chunks from old; and also contains sophisticated probabilistic equations for updating the activation levels associated with items of knowledge [9]. ACT-R in its original form did not say much about perceptual and motor operations, but recent versions have incorporated EPIC, an independent cognitive architecture focused on modeling these aspects of human behavior.

Fig. 2 [4] displays the current architecture of ACT-R. The flow of cognition in the system is in response to the current goal, currently active information from declarative memory, information attended to in perceptual modules (vision and audition are implemented), and the current state of motor modules (hand and speech are implemented).

The early work with ACT-R was based on comparing system performance to human behavior, using only behavioral measures, such as the timing of keystrokes or patterns of eye movements.

Please cite this article as: B. Goertzel, et al., World survey of artificial brains, Part II: Biologically inspired cognitive architectures, Neurocomputing (2010), doi:10.1016/j.neucom.2010.08.012

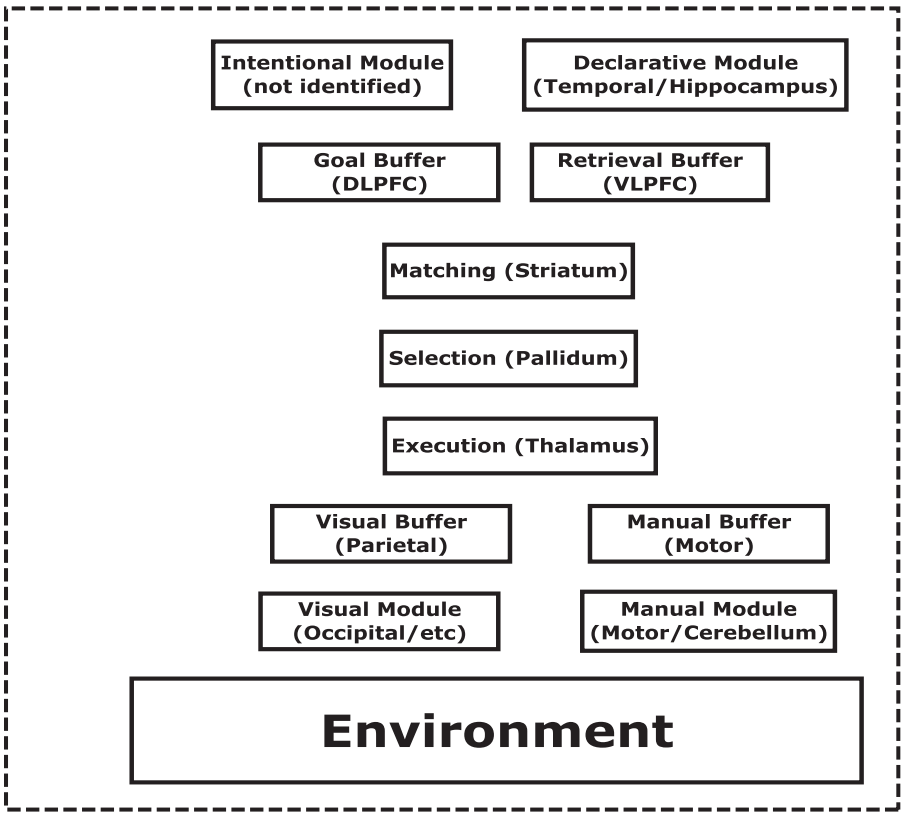


Fig. 2. ACT-R architecture.

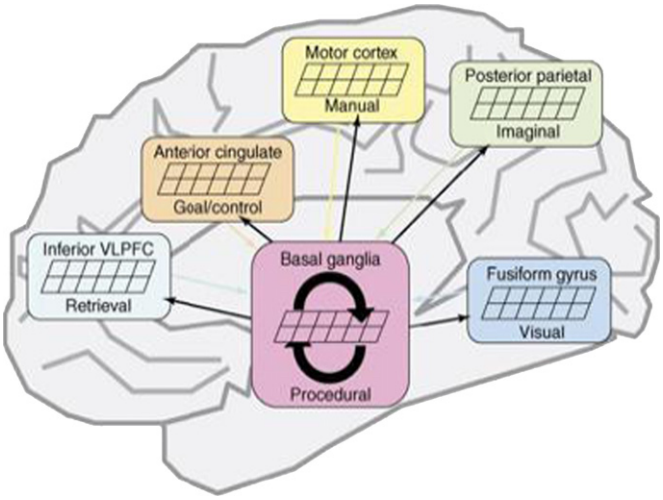


Fig. 3. An illustration of how the various cortical modules of ACT-R are coordinated through the procedural module that is associated with the basal ganglia. VLPFC, ventrolateral prefrontal cortex.

Using such measures, it was not possible to test detailed assumptions about which modules were active in the performance of a task. More recently, the ACT-R community has been engaged in a process of using imaging data to provide converging data on module activity. Fig. 3 [10] illustrates the associations they have made between the modules in Fig. 2 [4] and brain regions. Coordination among all of these components occurs through actions of the procedural module, which is mapped to the basal ganglia.

In practice, ACT-R seems to be used more as a programming framework for cognitive modeling than as an AI system. One can fairly easily use ACT-R to program models of specific human mental behaviors, which may then be matched against psychological and neurobiological data.

3. Emergentist cognitive architectures

Another species of cognitive architecture expects abstract symbolic processing to emerge from lower-level "subsymbolic" dynamics, which usually (but not always) are heavily biologically inspired, and designed to simulate neural networks or other aspects of human brain function. These architectures are typically strong at recognizing patterns in high-dimensional data, reinforcement learning and associative memory; but no one has yet shown how to achieve high-level functions such as abstract reasoning or complex language processing using a purely subsymbolic approach. A few of the more important subsymbolic, emergentist cognitive architectures are as follows:

- **Hierarchical Temporal Memory (HTM)** [10], is a hierarchical temporal pattern recognition architecture, presented as both an AI approach and a model of the cortex. So far it has been used exclusively for vision processing.
- **DeSTIN** [11,12] which contains a hierarchical perception network similar to (but more functional than) HTM, and also contains coordinated hierarchical networks dealing with action and reinforcement.
- **IBCA** (Integrated Biologically based Cognitive Architecture) [13], is a large-scale emergent architecture that seeks to model distributed information processing in the brain, especially the posterior and frontal cortex and the hippocampus. It has been used to simulate various human psychological and psycholinguistic behaviors, but has not been shown to give rise to higher-level behaviors like reasoning or subgoaling.

Please cite this article as: B. Goertzel, et al., World survey of artificial brains, Part II: Biologically inspired cognitive architectures, Neurocomputing (2010), doi:10.1016/j.neucom.2010.08.012

- **NOMAD** (Neurally Organized Mobile Adaptive Device) automata [14], are based on Edelman's "Neural Darwinism" model of the brain, and feature large numbers of simulated neurons evolving by natural selection into configurations that carry out sensorimotor and categorization tasks. The emergence of higher-level cognition from this approach seems particularly unlikely.
- Ben Kuipers has pursued an extremely innovative research program which combines qualitative reasoning [15] and reinforcement learning [16] to enable an intelligent agent to learn how to act, perceive and model the world. Kuipers' notion of "bootstrap learning" [17] involves allowing the robot to learn almost *everything* about its world, including for instance the structure of 3D space and other things that humans and other animals obtain via their genetic endowments.

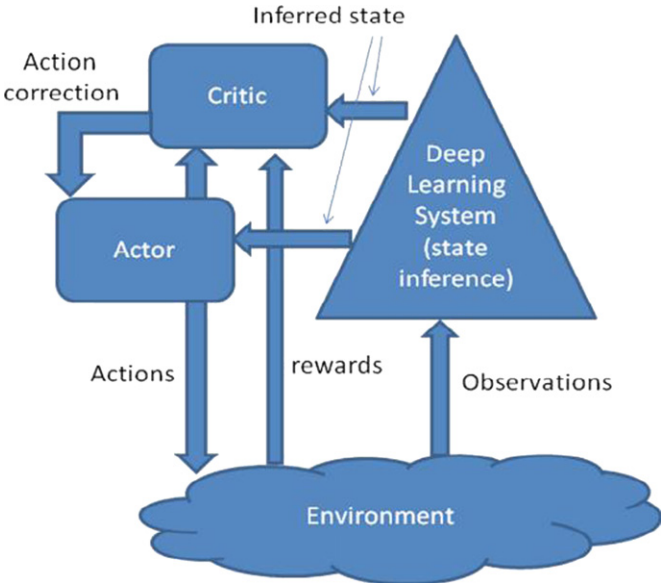


Fig. 4. High-level architecture of DeSTIN.

There is also a set of emergentist architectures focused specifically on developmental robotics, which we will review below in a separate section, as all of these share certain common characteristics.

As an example of this category, we now review the DeSTIN emergentist architecture in more detail, and then turn to the developmental robotics architectures.

3.1. *DeSTIN: a deep reinforcement learning based BICA*

The DeSTIN architecture, created by Itamar Arel and his colleagues, addresses the problem of general intelligence using hierarchical spatiotemporal networks designed to enable scalable perception, state inference and reinforcement-learning-guided action in real-world environments. DeSTIN has been developed with the plan of gradually extending it into a complete system for humanoid robot control, founded on the same qualitative information-processing principles as the human brain (though, distinguishing it from the large-scale brain simulations reviewed in Part I of this paper, without striving for detailed biological realism). The practical work with DeSTIN to date has focused on visual and auditory processing; and here we will discuss DeSTIN primarily in the perception context, only briefly mentioning the application to actuation which is conceptually similar.

In DeSTIN (see Figs. 4, 5), perception is carried out by a deep spatiotemporal inference network, which is connected to a similarly architected critic network that provides feedback on the inference network's performance, and an action network that controls actuators based on the activity in inference network. The nodes in these networks perform probabilistic pattern recognition according to algorithms to be described below; and the nodes in each of the networks may receive states of nodes in the other networks as inputs, providing rich interconnectivity and synergetic dynamics.

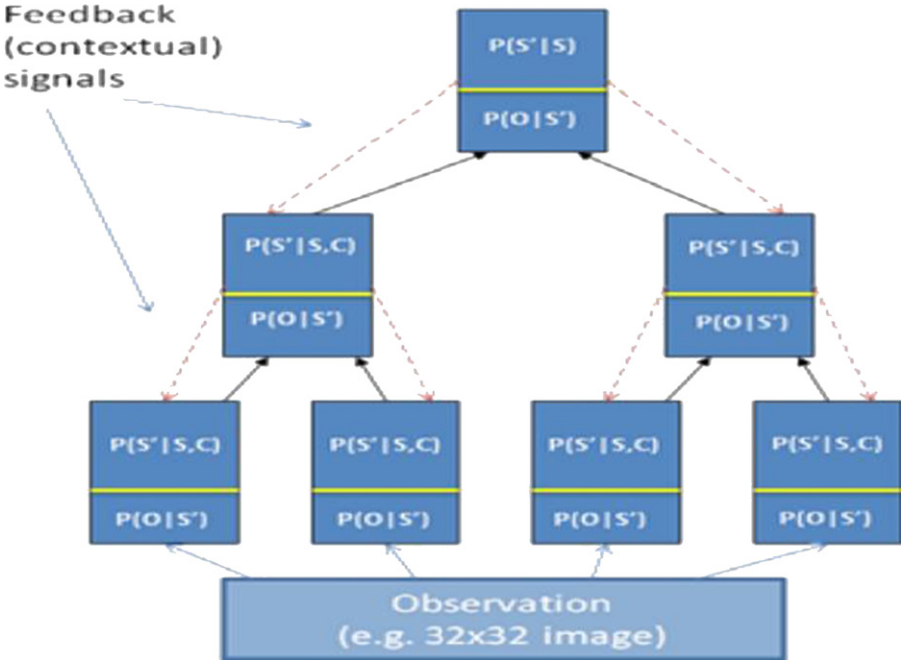


Fig. 5. Small-scale instantiation of the DeSTIN perceptual hierarchy. Each box represents a node, which corresponds to a spatiotemporal region (nodes higher in the hierarchy corresponding to larger regions). *O* denotes the current observation in the region, *C* is the state of the higher-layer node, and *S* and *S'* denote state variables pertaining to two subsequent time steps. In each node, a statistical learning algorithm is used to predict subsequent states based on prior states, current observations, and the state of the higher-layer node.

Please cite this article as: B. Goertzel, et al., World survey of artificial brains, Part II: Biologically inspired cognitive architectures, Neurocomputing (2010), doi:10.1016/j.neucom.2010.08.012

3.1.1. Deep versus shallow learning for perceptual data processing

The most critical feature of DeSTIN is its robust approach to modeling the world based on perceptual data. Mimicking the efficiency and robustness by which the human brain analyzes and represents information has been a core challenge in AI research for decades—and is one of the main reasons why AI researchers have turned to biologically inspired architectures. Humans are exposed to massive amounts of sensory data every second of every day, and are somehow able to capture critical aspects of it in a way that allows for appropriate future recollection and action selection. The brain apparently achieves this via neural computation on its massively parallel fabric, in which computation processes and memory storage are highly distributed. DeSTIN is one of a number of recent architectures that seeks to emulate the human brain's capability for massively parallel sensation-based world-modeling using a "deep learning" approach, in which detailed biological simulation is not attempted, but use is made of a hierarchical pattern recognition network architected similarly to relevant parts of the brain.

Humanlike intelligence is heavily adapted to the physical environments in which humans evolved; and one key aspect of sensory data coming from our physical environments is its **hierarchical** structure. However, most machine learning and pattern recognition systems are "shallow" in structure, not explicitly incorporating the hierarchical structure of the world in their architecture. In the context of perceptual data processing, the practical result of this is the need to couple each shallow learner with a pre-processing stage, wherein high-dimensional sensory signals are reduced to a lower-dimension feature space that can be understood by the shallow learner. The hierarchical structure of the world is thus crudely captured in the hierarchy of "preprocessor plus shallow learner". In this sort of approach, much of the intelligence of the system shifts to the feature extraction process, which is often imperfect and always application-domain specific.

Deep machine learning has emerged as a more promising framework for dealing with complex, high-dimensional real-world data. Deep learning systems possess a hierarchical structure that intrinsically biases them to recognize the hierarchical patterns present in real-world data. Thus, they hierarchically form a feature space that is driven by regularities in the observations, rather than by hand-crafted techniques. They also offer robustness to many of the distortions and transformations that characterize real-world signals, such as noise, displacement, scaling, etc.

Deep belief networks [18] and Convolutional Neural Networks [19] have been demonstrated to successfully address pattern inference in high dimensional data (e.g. images). They owe their success to their underlying paradigm of partitioning large data structures into smaller, more manageable units, and discovering the dependencies that may or may not exist between such units. However, the deep learning paradigm as manifested in these approaches has significant limitations; for instance, these approaches do not represent temporal information with the same ease as spatial structure. Moreover, some key constraints are imposed on the learning schemes driving these architectures, namely the need for layer-by-layer training, and oftentimes pre-training. DeSTIN seeks to overcome the limitations of prior deep learning approaches to perception processing, and also extends beyond perception to action and reinforcement learning.

3.1.2. DeSTIN for perception processing

The hierarchical architecture of DeSTIN's spatiotemporal inference network comprises an arrangement into multiple layers of "nodes" comprising multiple instantiations of an identical cortical circuit. Each node corresponds to a particular spatiotemporal region, and uses a statistical learning algorithm to characterize the sequences of patterns that are presented to it by nodes in the layer beneath it. More specifically,

- at the very lowest layer of the hierarchy nodes receive as input raw data (e.g. pixels of an image) and continuously construct a belief state that attempts to characterize the sequences of patterns viewed,
- the second layer, and all those above it, receive as input the belief states of nodes at their corresponding lower layers, and attempt to construct belief states that capture regularities in their inputs, and
- each node also receives as input the belief state of the node above it in the hierarchy (which constitutes "contextual" information).

DeSTIN's basic belief update rule, which governs the learning process and is identical for every node in the architecture, is as follows. The belief state is a probability mass function over the sequences of stimuli that the nodes learn to represent. Consequently, each node is allocated a predefined number of state variables each denoting a dynamic pattern, or sequence, that is autonomously learned. We seek to derive an update rule that maps the current observation (o), belief state (b), and the belief state of a higher-layer node (c), to a new (updated) belief state (b'), such that

$$b'(s') = \Pr(s'|o,b,c) = \frac{Pr(s' \cap o \cap b \cap c)}{Pr(o \cap b \cap c)} \tag{1}$$

alternatively expressed as

$$b'(s') = \frac{\Pr(o|s',b,c)Pr(s'|b,c)Pr(b,c)}{Pr(o|b,c)Pr(b,c)} \tag{2}$$

Under the assumption that observations depend only on true state, or $\Pr(o|s',b,c) = \Pr(o|s')$, we can further simplify the expression such that

$$b'(s') = \frac{\Pr(o|s')\Pr(s'|b,c)}{\Pr(o|b,c)} \tag{3}$$

where yielding the belief update rule

$$\Pr(s'|b,c) = \sum_{s \in S} \Pr(s'|s,c)b(s)$$

$$b'(s') = \frac{\Pr(o|s')\sum_{s \in S}\Pr(s'|s,c)b(s)}{\sum_{s'' \in S}\Pr(o|s'')\sum_{s \in S}\Pr(s''|s,c)b(s)} \tag{4}$$

where S denotes the sequence set (i.e. belief dimension) such that the denominator term is a normalization factor. One interpretation of (4) would be that the static pattern similarity metric $\Pr(o|s')$ is modulated by a construct that reflects the system dynamics $\Pr(s'|s,c)$. As such, the belief state inherently captures both spatial and temporal information. In our implementation, the belief state of the parent node c is chosen using the selection rule

$$c = argmax_s b_p(s) \tag{5}$$

where b_p is the belief distribution of the parent node. A closer look at Eq. (4) reveals that there are two core constructs to be learned, $\Pr(o|s')$ and $\Pr(s'|s,c)$.We show that the former can be learned via online clustering while the latter is learned based on experience by adjusting of the parameters with each transition from s to s'given c. The result is a robust framework that autonomously (i.e. with no human engineered pre-processing of any type) learns to

Please cite this article as: B. Goertzel, et al., World survey of artificial brains, Part II: Biologically inspired cognitive architectures, Neurocomputing (2010), doi:10.1016/j.neucom.2010.08.012

represent complex data patterns, such as those found in real-life robotics applications. Based on these equations, the DeSTIN perceptual network serves the critical role of building and maintaining a model of the state of the world. In a vision processing context, for example, it allows for powerful unsupervised classification. If shown a variety of real-world scenes, it will automatically form internal structures corresponding to the various natural categories of objects shown in the scenes, such as trees, chairs, people, etc., and also the various natural categories of events it sees, such as reaching, pointing, falling. And, as will be discussed below, it can use feedback from DeSTIN's action and critic networks to further shape its internal world-representation based on reinforcement signals.

DeSTIN's perceptual network offers multiple key attributes that render it a powerful approach to sensory data processing:

1. The belief space that is formed across the layers of the perceptual network inherently captures both *spatial and temporal regularities* in the data. Given that many applications require that temporal information be discovered for robust inference, this is a key advantage over existing schemes.
2. Spatiotemporal regularities in the observations are captured in a coherent manner (rather than being represented via two separate mechanisms).
3. All processing is both top-down and bottom-up, and both hierarchical and heterarchical, based on nonlinear feedback connections directing activity and modulating learning in multiple directions through DeSTIN's cortical circuits.
4. Support for multi-modal fusing is intrinsic within the framework, yielding a powerful state inference system for real-world, partially observable settings.
5. Each node is identical, which makes it easy to map the design to massively parallel platforms, such as graphics processing units.

3.1.3. DeSTIN for action and control

DeSTIN's perceptual network performs unsupervised world-modeling, which is a critical aspect of intelligence but of course is not the whole story. DeSTIN's action network, coupled with the perceptual network, orchestrates actuator commands into complex movements, but also carries out other functions that are more cognitive in nature.

For instance, people learn to distinguish between cups and bowls in part via hearing other people describe some objects as cups and others as bowls. To emulate this kind of learning, DeSTIN's critic network provides positive or negative reinforcement signals based on whether the action network has correctly identified a given object as a cup or a bowl, and this signal then impacts the nodes in the action network. The critic network takes a simple external "degree or success or failure" signal and turns it into multiple reinforcement signals to be fed into the multiple layers of the action network. The result is that the action network self-organizes so as to include an implicit "cup versus bowl" classifier, whose inputs are the outputs of some of the nodes in the higher levels of the perceptual network. This classifier belongs in the action network because it is part of the procedure by which the DeSTIN system carries out the action of identifying an object as a cup or a bowl.

This example illustrates how the learning of complex concepts and procedures is divided fluidly between the perceptual network, which builds a model of the world in an unsupervised way, and the action network, which learns how to respond to the world in a manner that will receive positive reinforcement from the critic network.

3.2. Developmental robotics architectures

Now we turn to another category of emergentist cognitive architectures: *developmental robotics* architectures, focused on controlling robots without significant "hard-wiring" of knowledge or capabilities, allowing robots to learn (and learn how to learn, etc.) via their engagement with the world. A significant focus is often placed here on "intrinsic motivation", wherein the robot explores the world guided by internal goals like novelty or curiosity, forming a model of the world as it goes along, based on the modeling requirements implied by its goals. Many of the foundations of this research area were laid by Schmidhuber's work in the 1990s [20–23], but now with more powerful computers and robots the area is leading to more impressive practical demonstrations.

We mention here a handful of the important initiatives in this area:

- Weng's **Dav** [24] and **SAIL** [25] projects involve mobile robots that explore their environments autonomously, and learn to carry out simple tasks by building up their own world-representations through both unsupervised and teacher-driven processing of high-dimensional sensorimotor data. The underlying philosophy is based on human child development [26], the knowledge representations involved are neural network based, and a number of novel learning algorithms are involved, especially in the area of vision processing.
- **FLOWERS** [27], an initiative at the French research institute INRIA, led by Pierre-Yves Oudeyer, is also based on a principle of trying to reconstruct the processes of development of the human child's mind, spontaneously driven by intrinsic motivations. Kaplan [28] has taken this project in a directly closely related to the present one via the creation of a "robot playroom." Experiential language learning has also been a focus of the project [29], driven by innovations in speech understanding.
- **IM-CLEVER**,[1] a new European project coordinated by Gianluca Baldassarre and conducted by a large team of researchers at different institutions, which is focused on creating software enabling an iCub [30] humanoid robot to explore the environment and learn to carry out human childlike behaviors based on its own intrinsic motivations. As this project is the closest to our own we will discuss it in more depth below.

IM-CLEVER is a humanoid robot intelligence architecture guided by intrinsic motivations, and using a hierarchical architecture for reinforcement learning and sensory abstraction. IM-CLEVER's motivational structure is based in part on Schmidhuber's information-theoretic model of curiosity [31]. On the other hand, IM-CLEVER's use of reinforcement learning follows Schmidhuber's earlier work RL for cognitive robotics [32,33], Barto's work on intrinsically motivated reinforcement learning [34,35], and Lee's [36,37] work on developmental reinforcement learning.

A skeptic of this research area might argue that, while the philosophy underlying developmental robotics is solid, the learning and representational mechanisms underlying the current systems in this area are probably not powerful enough to lead to human child level intelligence. Thus, it seems possible that these systems will develop interesting behaviors but fall short of robust human brain level competency, especially in areas like language and reasoning where symbolic systems have typically proved more effective. On the other hand, once the mechanisms underlying brains are better understood and robotic bodies are richer in

[1] http://im-clever.noze.it/project/project-description.

Please cite this article as: B. Goertzel, et al., World survey of artificial brains, Part II: Biologically inspired cognitive architectures, Neurocomputing (2010), doi:10.1016/j.neucom.2010.08.012

sensation and more adept in actuation, the developmental approach might grow into something more powerful.

4. Hybrid cognitive architectures

Finally, in response to the complementary strengths and weaknesses of the symbolic and emergentist approaches, in recent years a number of researchers have turned to integrative, hybrid architectures, which combine subsystems operating according to the two different paradigms. The combination may be done in many different ways, e.g. connection of a large symbolic subsystem with a large subsymbolic system, or the creation of a population of small agents each of which is both symbolic and subsymbolic in nature.

Nilsson expressed the motivation for hybrid cognitive architectures very clearly in his article at the AI-50 conference (which celebrated the 50th anniversary of the AI field) [38]. While affirming the value of the Physical Symbol System Hypothesis that underlies symbolic AI, he argues that "the PSSH explicitly assumes that, whenever necessary, symbols will be grounded in objects in the environment through the perceptual and effector capabilities of a physical symbol system." Thus, he continues,

> I grant the need for non-symbolic processes in some intelligent systems, but I think they supplement rather than replace symbol systems. I know of no examples of reasoning, understanding language, or generating complex plans that are best understood as being performed by systems using exclusively non-symbolic processes…
>
> AI systems that achieve human-level intelligence will involve a combination of symbolic and non-symbolic processing.

A few of the more important hybrid cognitive architectures are as follows:

- **CLARION** [39] is a hybrid architecture that combines a symbolic component for reasoning on "explicit knowledge" with a connectionist component for managing "implicit knowledge." Learning of implicit knowledge may be done via neural net, reinforcement learning, or other methods. The integration of symbolic and subsymbolic methods is powerful, but a great deal is still missing such as episodic knowledge and learning and creativity. Learning in the symbolic and subsymbolic portions is carried out separately rather than dynamically coupled, minimizing "cognitive synergy" effects.
- **DUAL** [40] is the most impressive system to come out of Marvin Minsky's "Society of Mind" paradigm. It features a population of agents, each of which combines symbolic and connectionist representation, self-organizing to collectively carry out tasks such as perception, analogy and associative memory. The approach seems innovative and promising, but it is unclear how the approach will scale to high-dimensional data or complex reasoning problems due to the lack of a more structured high-level cognitive architecture.
- **LIDA** [41] is a comprehensive cognitive architecture heavily based on Bernard Baars' "Global Workspace Theory". It articulates a "cognitive cycle" integrating various forms of memory and intelligent processing in a single processing loop. The architecture ties in well with both neuroscience and cognitive psychology, but it deals most thoroughly with "lower level'" aspects of intelligence, handling more advanced aspects like language and reasoning only somewhat sketchily. There is a clear mapping between LIDA structures and processes and corresponding structures and processing in OCP; so that it is only a mild stretch to view CogPrime as an instantiation of the general LIDA approach that extends further both in the lower level (to enable robot action and sensation via DeSTIN) and the higher level (to enable advanced language and reasoning via OCP mechanisms that have no direct LIDA analogues).
- **MicroPsi** [42] is an integrative architecture based on Dietrich Dorner's Psi model of motivation, emotion and intelligence. It has been tested on some practical control applications, and also on simulating artificial agents in a simple virtual world. MicroPsi's comprehensiveness and basis in neuroscience and psychology are impressive, but in the current version of MicroPsi, learning and reasoning are carried out by algorithms that seem unlikely to scale. OCP incorporates the Psi model for motivation and emotion, so that MicroPsi and CogPrime may be considered very closely related systems. But similar to LIDA, MicroPsi currently focuses on the "lower level" aspects of intelligence, not yet directly handling advanced processes like language and abstract reasoning.
- **PolyScheme** [43] integrates multiple methods of representation, reasoning and inference schemes for general problem solving. Each Polyscheme specialist models a different aspect of the world using specific representation and inference techniques, interacting with other specialists and learning from them. Polyscheme has been used to model infant reasoning including object identity, events, causality, and spatial relations. The integration of reasoning methods is powerful, but the overall cognitive architecture is simplistic compared to other systems and seems focused more on problem solving than on the broader problem of intelligent agent control.
- **Shruti** [44] is a fascinating biologically inspired model of human reflexive inference, represents in connectionist architecture relations, types, entities and causal rules using focal-clusters. However, much like Hofstadter's earlier Copycat architecture [45], Shruti seems more interesting as a prototype exploration of ideas than as a practical artificial brain system; at least, after a significant time of development it has not proved significant effective in any applications
- James Albus's **4D/RCS** robotics architecture shares a great deal with some of the emergentist architectures discussed above, e.g. it has the same hierarchical pattern recognition structure as DeSTIN and HTM, and the same three cross-connected hierarchies as DeSTIN, and shares with the developmental robotics architectures a focus on real-time adaptation to the structure of the world. However, 4D/RCS is not foundationally learning based but relies on hard-wired architecture and algorithms, intended to mimic the qualitative structure of relevant parts of the brain (and intended to be *augmented* by learning, which differentiates it from emergentist approaches).
- **OpenCogPrime** is a comprehensive architecture for cognition, language, and virtual agent control, created by the BenGoertzel and Cassio Pennachin and their collaborators during the period since 2001 (and building on their work from the 1990s). Conceptually founded on the systems theory of intelligence outlined in [46] and alluded to above, it is currently under development within the open-source OpenCog AI framework (http://opencog.org and [47]).

Many of the hybrid architectures are in essence "multiple, disparate algorithms carrying out separate functions, encapsulated in black boxes and communicating results with each other." For instance, PolyScheme, ACT-R and CLARION all display this "modularity" property to a significant extent. These architectures lack the rich, real-time interaction between the *internal dynamics* of various memory and learning processes that seems characteristics of the human brain, and that we conjecture is critical to achieving human brainlike general intelligence using realistic computational resources. On the other hand, there are also hybrid architectures that feature richer integration – such as DUAL,

Please cite this article as: B. Goertzel, et al., World survey of artificial brains, Part II: Biologically inspired cognitive architectures, Neurocomputing (2010), doi:10.1016/j.neucom.2010.08.012

Shruti, LIDA, OpenCog and MicroPsi – though many of these have the flaw of relying (at least in their current versions) on overly simplistic learning algorithms, which drastically limit their scalability.

It does seem plausible to us that some of these hybrid architectures could be dramatically extended or modified so as to produce humanlike general intelligence. For instance, one could replace LIDA's learning algorithms with others that interrelate with each other in a nuanced synergetic way, or one could replace MicroPsi's simple learning and reasoning methods with much more powerful and scalable ones acting on the same data structures. However, making these changes would dramatically alter the cognitive architectures in question on multiple levels.

We now review several hybrid architectures in more detail, focusing most deeply on LIDA, MicroPsi, and OpenCog.

4.1. CLARION

Ron Sun's CLARION architecture is interesting in its combination of symbolic and neural aspects—a combination that is used in a sophisticated way to embody the distinction and interaction between implicit and explicit mental processes. From a CLARION perspective, architectures like Soar and ACT-R are severely limited in that they deal only with explicit knowledge and associated learning processes.

As shown in Fig. 6 [39], CLARION consists of a number of distinct subsystems, each of which contains a dual representational structure, including a "rules and chunks" symbolic knowledge store somewhat similar to ACT-R, and a neural net knowledge store embodying implicit knowledge. The main subsystems are as follows:

- An action-centered subsystem to control actions.
- A non-action-centered subsystem to maintain general knowledge.
- A motivational subsystem to provide underlying motivations for perception, action, and cognition.
- A meta-cognitive subsystem to monitor, direct, and modify the operations of all the other subsystems.

4.2. DUAL

In his influential but controversial book *The Society of Mind* [48], Marvin Minsky describes a model of human intelligence as something that is built up from the interactions of numerous simple agents. He spells out in great detail how various particular cognitive functions may be achieved via agents and their interactions. He leaves no room for any central algorithms or structures of thought, famously arguing: "What magical trick makes us intelligent? The trick is that there is no trick. The power of intelligence stems from our vast diversity, not from any single, perfect principle."

This perspective was extended in the more recent work *The Emotion Machine* [49], where Minsky argued that emotions are "ways to think" evolved to handle different "problem types" that exist in the world. The brain is posited to have rule-based mechanisms (selectors) that turn on emotions to deal with various problems.

Overall, both of these works serve better as works of speculative cognitive science than as works of AI or cognitive architecture per se. As neurologist Richard Restak said in his review of *Emotion Machine*, "Minsky does a marvelous job parsing other complicated mental activities into simpler elements. ... But he is less effective in relating these emotional functions to what's going on in the brain." As Restak did not add, he is also not so effective at relating these emotional functions to straightforwardly implementable algorithms or data structures.

Push Singh, in his Ph.D. thesis [50], did the best job so far of creating a concrete AI design based oclosely n Minsky's ideas. Singh's system was certainly interesting, it was also noteworthy for its lack of any learning mechanisms, and its exclusive focus on explicit rather than implicit knowledge. Singh's work was never completed due to his tragic death; and it seems fair to say that there has not yet been a serious cognitive architecture posed based *closely* on Minsky's ideas.

The nearest thing to a Minsky-style cognitive architecture is probably DUAL, which takes the Society of Mind concept and adds to it a number of other interesting ideas. DUAL integrates symbolic and connectionist approaches at a deeper level than CLARION, and has been used to model various cognitive functions such as perception, analogy and judgment. Computations in DUAL emerge from the self-organized interaction of many micro-agents, each of which is a hybrid symbolic/connectionist device. Each DUAL agent plays the role of a neural network node, with an activation level and activation spreading dynamics; but also plays the role of a symbol, manipulated using formal rules. The agents exchange messages and activation via links that can be learned and modified, and they form coalitions which collectively represent concepts, episodes, and facts.

The structure of the model is sketchily shown in Fig. 7 [40], which covers the application of DUAL to a toy environment

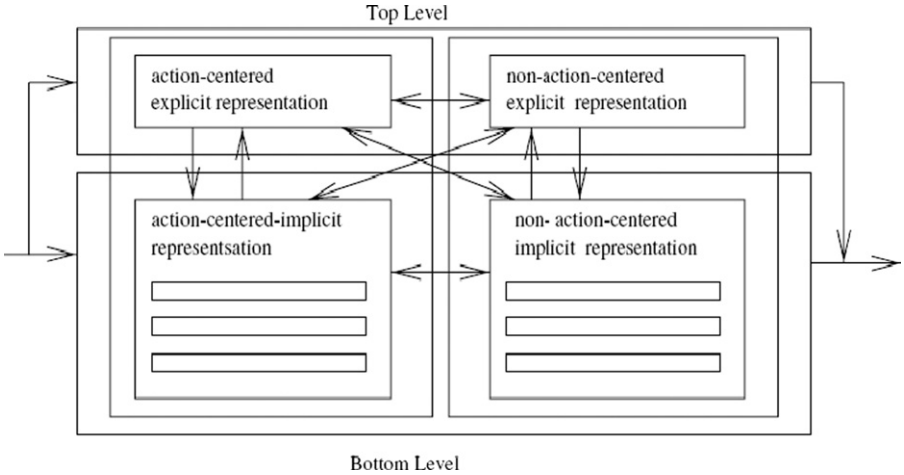


Fig. 6. The CLARION architecture.

Please cite this article as: B. Goertzel, et al., World survey of artificial brains, Part II: Biologically inspired cognitive architectures, Neurocomputing (2010), doi:10.1016/j.neucom.2010.08.012

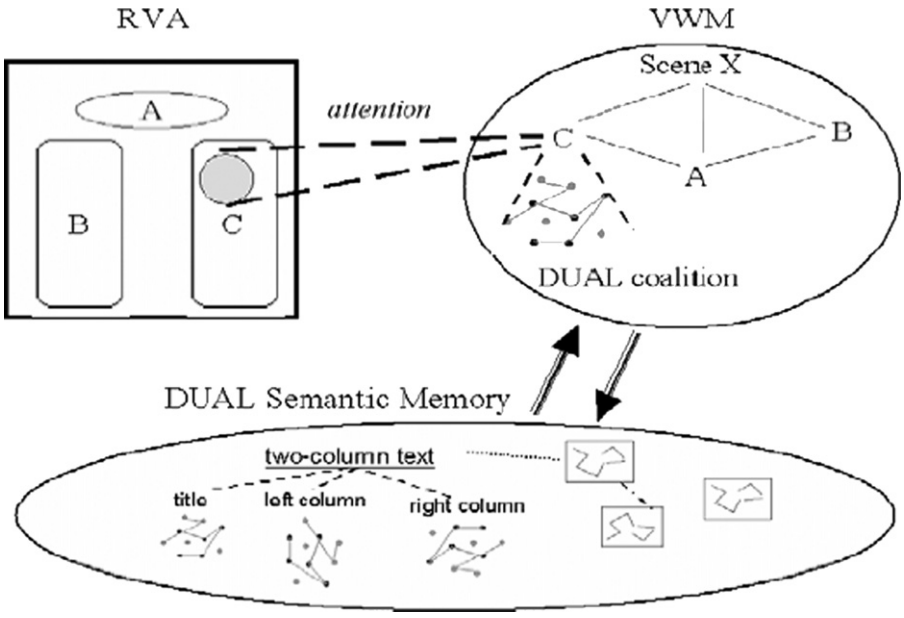


Fig. 7. The three main components of the model: the retinotopic visual array (RVA), the visual working memory (VWM), and DUAL's semantic memory. Attention is allocated to an area of the visual array by the object in VWM controlling attention, while scene and object categories corresponding to the contents of VWM are retrieved from the semantic memory.

called TextWorld. The visual input corresponding to a stimulus is presented on a two-dimensional visual array representing the front end of the system. Perceptual primitives like blobs and terminations are immediately generated by cheap parallel computations. Attention is controlled at each time by an object which allocates it selectively to some area of the stimulus. A detailed symbolic representation is constructed for this area which tends to fade away as attention is withdrawn from it and allocated to another one. Categorization of visual memory contents takes place by retrieving object and scene categories from DUAL's semantic memory and mapping them onto current visual memory representations.

4.3. 4D/RCS

In a rather different direction, James Albus, while at the National Bureau of Standards, developed a very thorough and impressive architecture for intelligent robotics called 4D/RCS, which was implemented in a number of machines including unmanned automated vehicles. This architecture lacks critical aspects of intelligence such as learning and creativity, but combines perception, action, planning, and world-modeling in a highly effective and tightly integrated fashion (Figs. 8 and 9).

In a striking parallel to DeSTIN as reviewed above, the architecture has three hierarchies of memory/processing units: one for perception, one for action and one for modeling and guidance. Each unit has a certain spatiotemporal scope, and (except for the lowest level) supervenes over children whose spatiotemporal scope is a subset of its own. The action hierarchy takes care of decomposing tasks into subtasks, whereas the sensation hierarchy takes care of grouping signals into entities and events. The modeling/guidance hierarchy mediates interactions between perception and action based on its understanding of the world and the system's goals.

In his book [51] Albus describes methods for extending 4D/RCS into a complete cognitive architecture, but these extensions have not been elaborated in full detail nor implemented.

The traditional implementations of 4D/RCS are not closely brain-like except in overall conception, but in recent work Albus has tied 4D/RCS in more tightly with detailed theories of brain function. Fig. 12 [65] shows a neural implementation of a portion of Albus's 4D/RCS perceptual hierarchy, where the nodes in the hierarchy are implemented by Cortical Computational Units (CCUs) mimicking brain structure. Figs. 10 and 12 [65] shows the software implementation of a CCU, and Fig. 11 [65] shows how Albus conjectures the same functionality to be implemented in the human brain.

4.4. LIDA

The LIDA architecture developed by Stan Franklin and his colleagues [52] is based on the concept of the cognitive cycle—a notion that is important to nearly every BICA and also to the brain, but that plays a particularly central role in LIDA. As Franklin says, as a matter of principle, every autonomous agent, 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.

The simplest cognitive cycle is that of an animal, which senses the world, compares sensation to memory, and chooses an action, all in one fluid subjective moment. But the same cognitive cycle structure/process applies to higher-level cognitive processes as well. The LIDA architecture is based on the LIDA model of the cognitive cycle, which posits a particular structure underlying the cognitive cycle that possess the generality to encompass both simple and complex cognitive moments.

Fig. 13[2] shows the cycle pictorially, starting in the upper left corner and proceeding clockwise. At the start of a cycle, the LIDA agent perceives its current situation and allocates attention differentially to various parts of it. It then broadcasts information about the most important parts (which constitute the agent's consciousness), and this information gets features extracted from it, when then get passed along to episodic and semantic memory,

[2] http://ccrg.cs.memphis.edu/tutorial/synopsis.html.

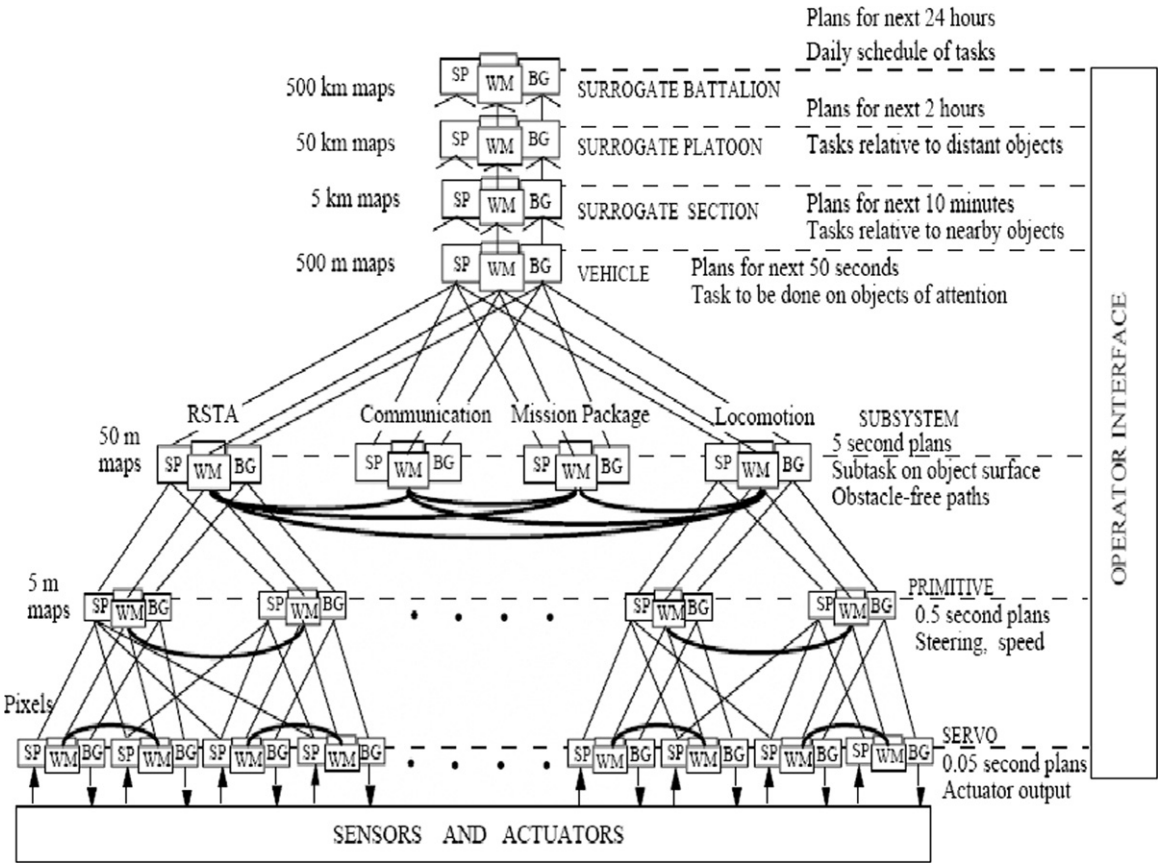


Fig. 8. [64]: Albus's 4D-RCS architecture for a single vehicle.

that interact in the global workspace to create a model of the agent's current situation. This model then, in interaction with procedural memory, enables the agent to choose an appropriate action and execute it the critical section-selection phase!

4.5. The LIDA cognitive cycle in more depth[3]

We now run through the LIDA cognitive cycle in more detail. It begins with sensory stimuli from the agent's external 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, called percepts, 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 have not 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. Via constructing the model, the agent has made sense of the incoming stimuli.

Now attention allocation comes into play, because a real agent lacks the computational resources to work with all parts of its world-model with maximal mental focus. Portions of the model compete for attention. These competing portions take the form of (potentially overlapping) coalitions of structures comprising parts the model. Once one such coalition wins the competition, the agent has decided what to focus its attention on.

And now comes the purpose of all this processing: to help the agent to decide what to do next. The winning coalition passes to the global workspace, the namesake of Global Workspace Theory, 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.

Procedural memory also stores an activation value for each such template—a value that attempts to measure the likelihood of an action taken within its context producing the expected result. It is worth noting that LIDA makes a rather specific assumption here.

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 sensorimotor 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. It acknowledges that other cognitive processes may also occur, refining and building on the knowledge used in the cognitive cycle (for instance, the cognitive cycle itself does not mention

[3] This section paraphrases in some places from [53].

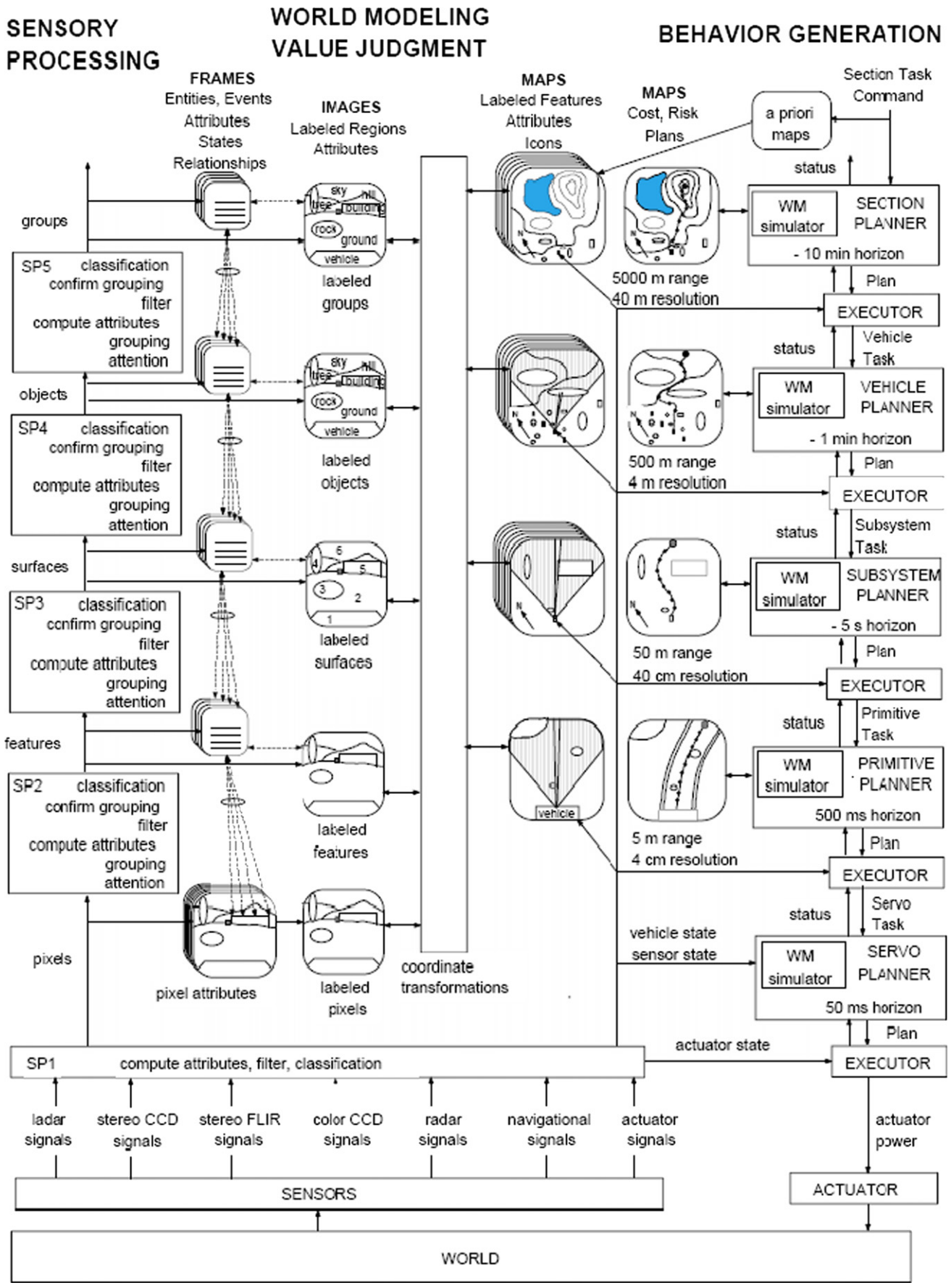


Fig. 9. [64]: Albus's perceptual, motor, and modeling hierarchies.

Please cite this article as: B. Goertzel, et al., World survey of artificial brains, Part II: Biologically inspired cognitive architectures, Neurocomputing (2010), doi:10.1016/j.neucom.2010.08.012

abstract reasoning or creativity). But the idea is that these other processes occur in the context of the cognitive cycle, which is the main loop driving the internal and external activities of the organism.

4.5.1. Neural correlates of LIDA cognitive processes

Stan Franklin and the other developers of LIDA, together with Bernard Baars and other neuroscientists have conducted an ongoing effort towards creating a concordance between the modules and processes of the LIDA model and their possible neural correlates. Table 1a and b presents some of the conclusions of this investigation, but the reader is encouraged to peruse the full list online at [61] to get a sense of the full impressiveness of this initiative. In addition to the connection with LIDA specifically, this is one of the most carefully constructed systematic mappings between brain regions and cognitive functions of which we are aware.

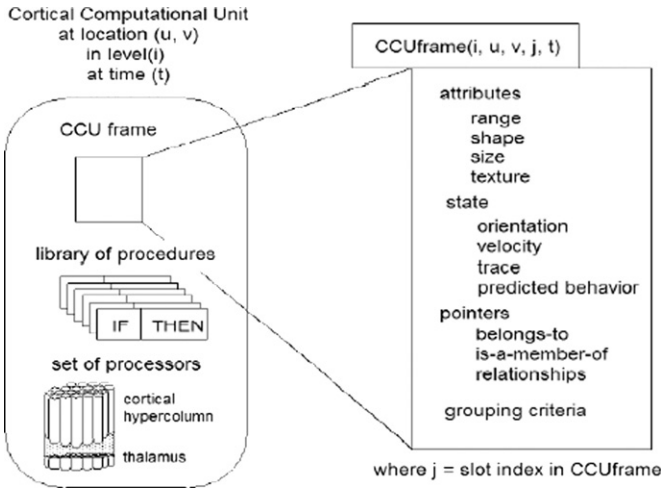


Fig. 10. Internal structure of a Cortical Computational Unit (CCU) consisting of a CCUframe, a library of procedures that maintain the frame, and a set of processors that implement the procedures.

4.5.2. Avoiding combinatorial explosion via adaptive attention allocation

As the worst problem plaguing traditional symbolic cognitive architectures is "combinatorial explosion" – the overabundance of possible combinations of memory items and new concepts and processes to evaluate in the context of choosing actions to achieve goals – it is interesting to ask how various BICAs use brainlike methods to overcome this issue. LIDA has a fairly compelling solution to the problem, at least conceptually (as with other BICAs, the experimental work with LIDA to date has not been thorough enough to fully validate the solution). In essence, LIDA avoids combinatorial explosions in its inference processes via two methods:

- combining reasoning via association with reasoning via deduction and
- foundational use of uncertainty in reasoning.

One can create an analogy between LIDA's workspace structures and codelets and a symbolic architecture's assertions and functions or production rules. 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.

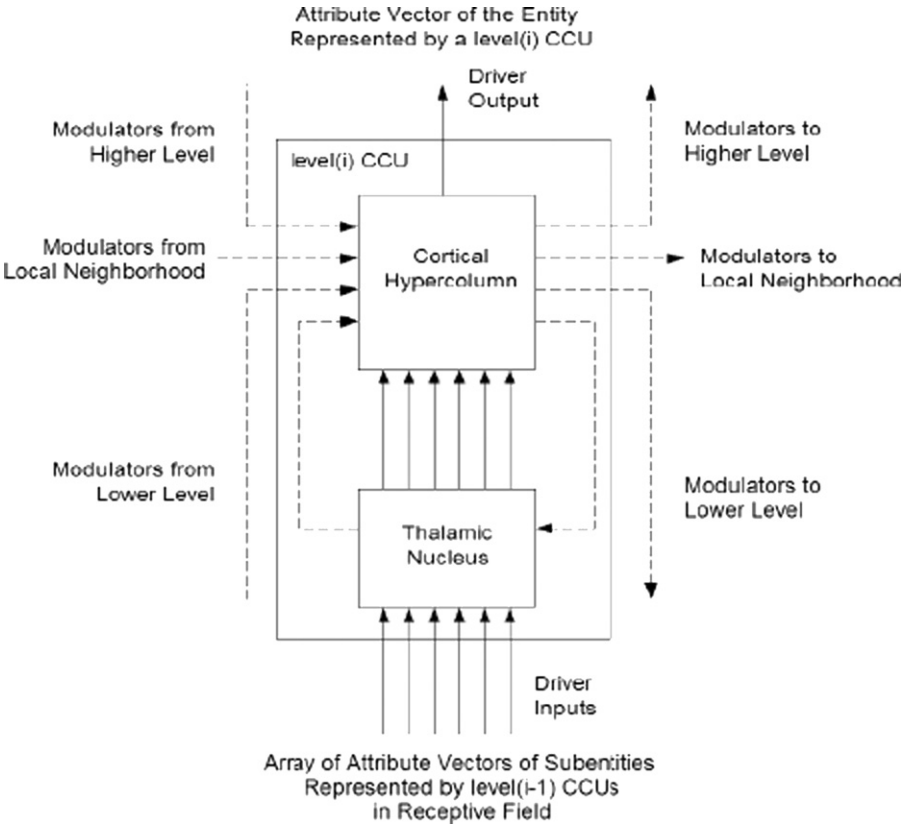


Fig. 11. Inputs and outputs of a Cortical Computational Unit (CCU) in posterior cortex.

Please cite this article as: B. Goertzel, et al., World survey of artificial brains, Part II: Biologically inspired cognitive architectures, Neurocomputing (2010), doi:10.1016/j.neucom.2010.08.012

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

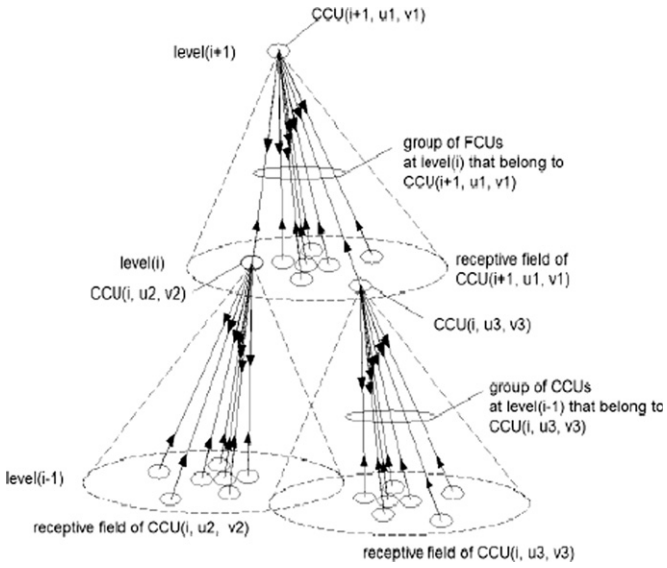


Fig. 12. Two levels of segmentation and grouping of entity frames.

4.6. Psi and microPsi

Next we consider Joscha Bach's MicroPsi architecture, which is closely based on Dietrich Dorner's Psi theory, roughly shown in Fig. 14 [60]. One could argue that MicroPsi is more *psychologically* inspired than biologically inspired, but the boundary becomes extremely blurry, e.g. because Dorner's psychological models involve networks of processing units with many broadly neuron-like properties.

Psi's motivational system begins with **Demands**, which are the basic factors that motivate the agent. For an animal these would include things like food, water, sex, novelty, socialization, protection of one's children, etc. For an intelligent robot they might include things like electrical power, novelty, certainty, socialization, well-being of others and mental growth.

Psi also specifies two fairly abstract demands and posits them as psychologically fundamental (see Fig. 15 [60]):

- **competence**, the effectiveness of the agent at fulfilling its Urges, and
- **certainty**, the confidence of the agent's knowledge.

Each demand is assumed to come with a certain "target level" or "target range" (and these may fluctuate over time, or may change as a system matures and develops). An **Urge** is said to develop when a demand deviates from its target range: the urge then seeks to return the demand to its target range. For instance, in an animal-like agent the demand related to food is more clearly described as "fullness," and there is a target range indicating that the agent is neither too hungry nor too full of food. If the agent's

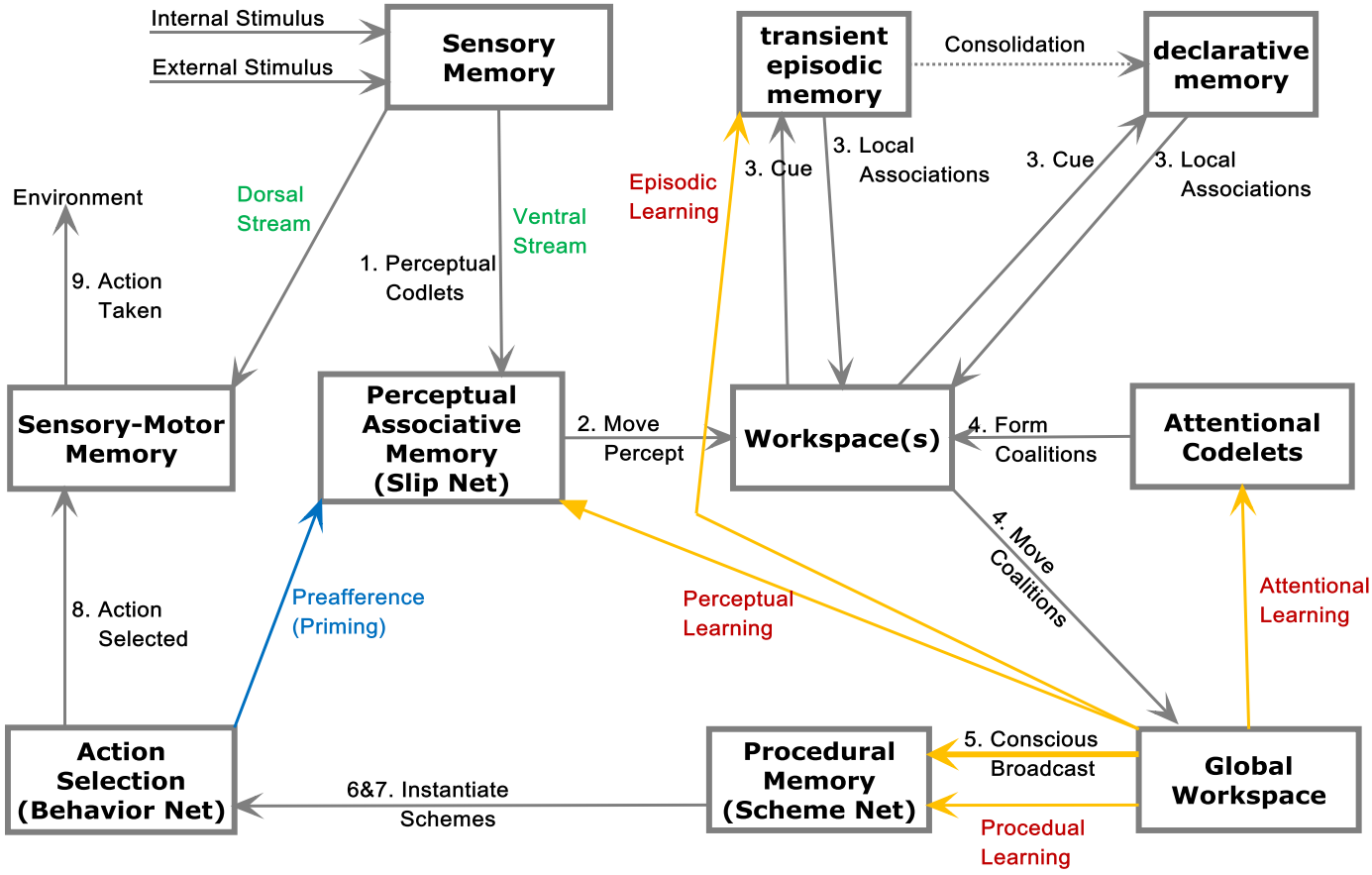


Fig. 13. The LIDA cognitive cycle.

Please cite this article as: B. Goertzel, et al., World survey of artificial brains, Part II: Biologically inspired cognitive architectures, Neurocomputing (2010), doi:10.1016/j.neucom.2010.08.012

Table 1a
Possible neural correlates of cognitive processes and modules from the LIDA model of cognition. The first column of the table lists the LIDA model module or process (function) that's involved with the cognitive process(es) listed in the second column; the third column lists possible neural correlates of the cognitive process in question, meaning that there is evidence of the stated neural assemblies being involved with that process (clearly, neural assemblies other than those listed by are also involved).

LIDA module or function	Cognitive processes	Neural correlates include
Sensory motor automatism	Sensory motor automatism	Cerebellum
	Sensory memory	Temporal-prefrontal (echoic)
Slipnet	Perceptual associative memory (PAM)	Perirhinal cortex
Slipnet object nodes	PAM-visual object recognition	Inferotemporal cortex (IT), perirhinal cortex
Slipnet	PAM-phenomenal visual experience	Medial temporal lobe
Slipnet	PAM-categorization-sensory	Sensory neocortex
Slipnet	PAM-categorization-abstract	Lateral and anterior prefrontal structures
Slipnet	PAM-categorization-motion	lateral intraparietal
Slipnet face nodes	PAM-visual face recognition	Inferotemporal cortex (IT)
Slipnet face and other object nodes	PAM-recognition of faces and other objects	Fusiform face area
Slipnet emotion nodes	PAM-emotions	Amygdala and orbito-frontal cortex
Slipnet	PAM-emotion affecting learning	Basolateral amygdala, perirhinal cortex, entorhinal cortex
Slipnet fear nodes	PAM-memory of fear	Lateral nucleus of the amygdala
Slipnet novelty node	PAM-novelty	Substantia nigra/ventral tegmental area
Slipnet action situation nodes	PAM-recognize action situation—e.g. eating a peanut	Mirror neurons in the perisylvian cortical region
Slipnet reward nodes	PAM-reward for action or its avoidance	Medial orbitofrontal cortex
Slipnet reward nodes	PAM-stimulus-reward associations	Orbitofrontal cortex/ventral striatum
Slipnet emotion nodes	PAM-romantic love	Brainstem right ventral tegmental area and right postero-dorsal body of the caudate nucleus
Slipnet self movement nodes & place nodes	PAM-self movement input to location cognitive map	Entorhinal cortex, hippocampus
Slipnet category nodes	Perceptual learning of concepts	Hippocampus
Slipnet feeling nodes	Feelings, bodily states, social emotions	Insula
Attention codelets	Attention to objects	Posterior parietal cortex
Attention codelets	Higher visual scene processing	Frontal eye fields

Table 1b
Possible neural correlates of cognitive processes and modules from the LIDA model of cognition. The first column of the table lists the LIDA model module or process (function) that is involved with the cognitive process(es) listed in the second column; The third column lists possible neural correlates of the cognitive process in question, meaning that there is evidence of the stated neural assemblies being involved with that process (clearly, neural assemblies other than those listed by are also involved).

Global workspace	Global broadcast	Long-distance synchronization of gamma oscillations
Schemenet	Procedural learning	Striatum
Schemenet	Procedural memory	Basal ganglia, cortical motor regions, cerebellar structures
PL attention codelets	Procedural learning	Corticostriatal synapses
Scheme net	Procedural learning	Ventral striatum
Scheme net	Procedural memory	Anterior cingulate cortex, Striatum
Behavior net	Action selection	Prefrontal cortex
Behavior net	Action selection	Striatum, basal ganglia
Behavior net	Emotions modulating action selection	Amygdala, Orbital and medial prefrontal cortex
	Sensory-motor memory	Cerebellum
	Sensory-motor control of balance and navigation	Semicircular canals
	Sensory-motor innate behaviors	command chemical-central peptidergic ensembles
Sensory-motor memory	Action execution	Dorsal striatum
	Motor representations of sequential procedures	Right intraparietal sulcus
Cognitive cycle	Cognitive cycle	Few hundred ms
Cognitive cycle	Cognitive cycle	200–300 ms
Cognitive cycle	Cognitive cycle	Coupled theta–gamma rhythms
Cognitive cycle	EEG microstate	
	Reflexive decision making	amygdala, basal ganglia, ventromedial prefrontal cortex, dorsal anterior cingulate cortex, lateral temporal cortex
	Reflective decision making	Lateral prefrontal cortex, posterior parietal cortex, medial prefrontalal cortex, rostral anterior cingulate cortex, and hippocampus and surrounding medial temporal lobe region
	Emotions in moral decision making	Ventromedial prefrontal cortex

fullness deviates from this range, an Urge to return the demand to its target range arises. Similarly, if an agent's novelty deviates from its target range, this means the agent's life has gotten either too boring or too disconcertingly weird, and the agent gets an Urge for either more interesting activities (in the case of below-range novelty) or more familiar ones (in the case of above-range novelty).

There is also a primitive notion of **Pleasure** (and it is opposite, displeasure), which is considered as different from the complex emotion of "happiness." Pleasure is understood as associated with Urges: pleasure occurs when an Urge is (at least partially) satisfied, whereas displeasure occurs when an urge gets increasingly severe. The degree to which an Urge is satisfied is not necessarily defined instantaneously; it may be defined, for instance, as a time-decaying weighted average of the proximity of the demand to its target range over the recent past.

So, for instance if an agent is bored and gets a lot of novel stimulation, then it experiences some pleasure. If it is bored and then the monotony of its stimulation gets even more extreme, then it experiences some displeasure.

Please cite this article as: B. Goertzel, et al., World survey of artificial brains, Part II: Biologically inspired cognitive architectures, Neurocomputing (2010), doi:10.1016/j.neucom.2010.08.012

Note that, according to this relatively simplistic approach, any decrease in the amount of dissatisfaction causes some pleasure; whereas if everything always continues within its acceptable range, there is not any pleasure. This may seem a little counterintuitive, but it's important to understand that these simple definitions of "pleasure" and "displeasure" are not intended to fully capture the natural language concepts associated with those words. The natural language terms are used here simply as heuristics to convey the general character of the processes involved. These are very low level processes whose analogues in human experience are largely below the conscious level.

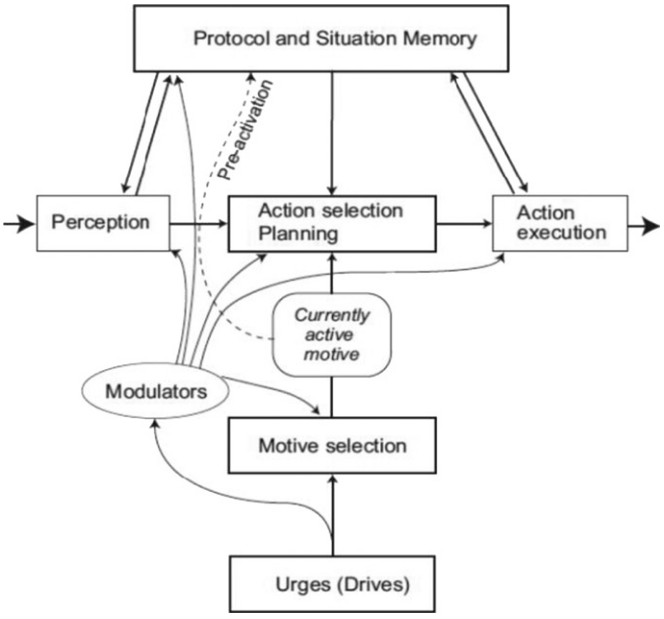


Fig. 14. High-level architecture of the Psi model.

A **Goal** is considered as a statement that the system may strive to make true at some future time. A **Motive** is an (*urge*, *goal*) pair, consisting of a goal whose satisfaction is predicted to imply the satisfaction of some urge. In fact one may consider Urges as top-level goals, and the agent's other goals as their subgoals.

In Psi an agent has one "ruling motive" at any point in time, but this seems an oversimplification more applicable to simple animals than to human-like or other advanced AI systems. In general one may think of different motives having different weights indicating the amount of resources that will be spent on pursuing them.

Emotions in Psi are considered as complex systemic response-patterns rather than explicitly constructed entities. An emotion is the set of mental entities activated in response to a certain set of urges. Dorner conceived theories about how various common emotions emerge from the dynamics of urges and motives as described in the Psi model. "Intentions" are also considered as composite entities: an intention at a given point in time consists of the active motives, together with their related goals, behavior programs, etc.

The basic logic of action in Psi is carried out by "triplets" that are very similar to production rules in classical symbolic systems like SOAR and ACT-R, in a manner to be described below. However, an important role is also played by four **modulators** that control how the processes of perception, cognition, and action selection are regulated at a given time:

- *activation*, which determines the degree to which the agent is focused on rapid, intensive activity versus reflective, cognitive activity,
- *resolution level*, which determines how accurately the system tries to perceive the world,
- *certainty*, which determines how hard the system tries to achieve definite, certain knowledge, and

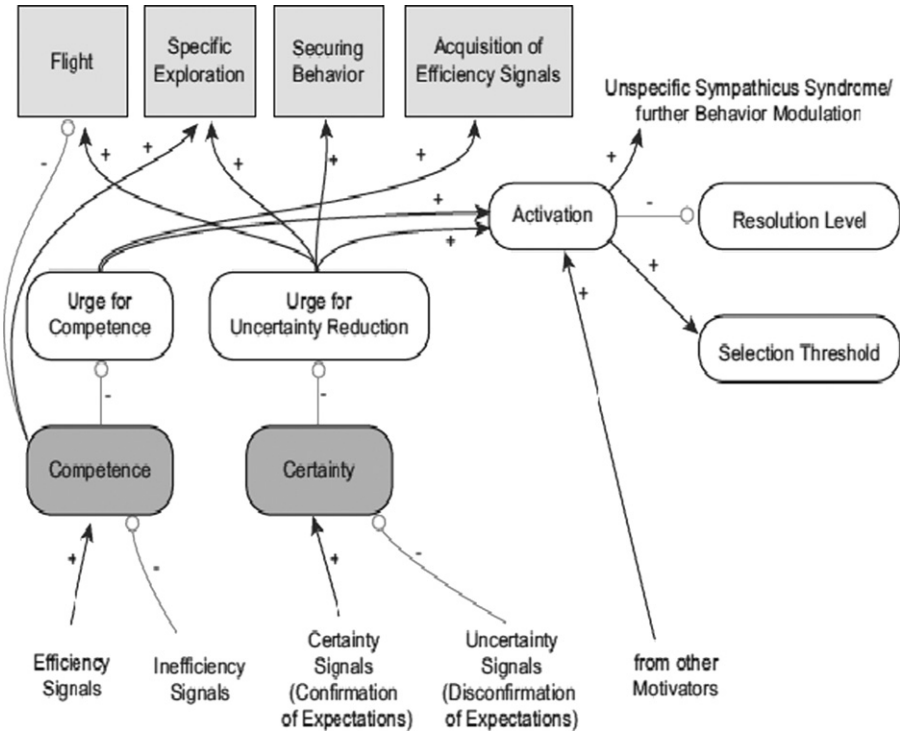


Fig. 15. Primary interrelationships between Psi modulators.

Please cite this article as: B. Goertzel, et al., World survey of artificial brains, Part II: Biologically inspired cognitive architectures, Neurocomputing (2010), doi:10.1016/j.neucom.2010.08.012

- *selection threshold*, which determines how willing the system is to change its choice of which goals to focus on.

These modulators characterize the system's emotional and cognitive state at a very abstract level; they are not emotions per se, but they have a large effect on the agent's emotions. Their intended interaction is shown in Fig. 15.

4.6.1. Knowledge representation, action selection, and planning in Psi

On the micro level, Psi represents knowledge using closely brain-inspired structures called "quads." Each quad is a cluster of 5 neurons containing a core neuron, and four other neurons representing before/after and part-of/has-part relationships in regard to that core neuron. Quads are naturally assembled into spatiotemporal hierarchies, though they are not required to form part of such a structure.

Psi stores knowledge using quads arranged in three networks, which are conceptually similar to the networks in Albus's 4D/RCS and Arel's DeSTIN architectures:

- A sensory network, which stores declarative knowledge: schemas representing images, objects, events and situations as hierarchical structures.
- A motor network, which contains procedural knowledge by way of hierarchical behavior programs.
- A motivational network handling demands.

Perception in Psi, which is centered in the sensory network, follows principles similar to DeSTIN (which are shared also by other systems), for instance the principle of *perception as prediction*. Psi's "HyPercept" mechanism performs hypothesis-based perception: it attempts to predict what is there to be perceived and then attempts to verify these predictions using sensation and memory. Furthermore HyPercept is intimately coupled with actions in the external world, according to the concept of "Neisser's perceptual cycle," the cycle between exploration and representation of reality. Perceptually acquired information is translated into schemas capable of guiding behaviors, and these are enacted (sometimes affecting the world in significant ways) and in the process used to guide further perception. Imaginary perceptions are handled via a "mental stage" analogous to CogPrime's internal simulation world.

Action selection in Psi works based on what are called "triplets," each of which consists of

- a sensor schema (pre-conditions, "condition schema"),
- a subsequent motor schema (action, effector), and
- a final sensor schema (post-conditions, expectations).

What distinguishes these triplets from classic production rules as used in (say) Soar and ACT-R is that the triplets may be partial (some of the three elements may be missing) and may be uncertain. The difference lies in the underlying knowledge representation used for the schemata, and the probabilistic logic used to represent the implication.

The work of figuring out what schema to execute to achieve the chosen goal in the current context is done in Psi using a combination of processes called the "Rasmussen ladder" (named after Danish psychologist Jens Rasmussen). The Rasmussen ladder describes the organization of action as a movement between the stages of skill-based behavior, rule-based behavior and knowledge-based behavior, as follows:

- If a given task amounts to a trained routine, an automatism or skill is activated; it can usually be executed without conscious attention and deliberative control.
- If there is no automatism available, a course of action might be derived from rules; before a known set of strategies can be applied, the situation has to be analyzed and the strategies have to be adapted.
- In those cases where the known strategies are not applicable, a way of combining the available manipulations (operators) into reaching a given goal has to be explored at first. This stage usually requires a recomposition of behaviors, that is, a planning process.

The planning algorithm used in the Psi and MicroPsi implementations is a fairly simple hill-climbing planner. While it is hypothesized that a more complex planner may be needed for advanced intelligence, part of the Psi theory is the hypothesis that most real-life planning an organism needs to do is fairly simple, once the organism has the right perceptual representations and goals.

4.7. OpenCogPrime

Finally, OpenCogPrime (OCP), a cognitive architecture developed by several of the co-authors of this paper, combines multiple AI paradigms such as uncertain-logic, computational linguistics, evolutionary program learning and connectionist attention allocation in a unified cognitive-science-based architecture. Cognitive processes embodying these different paradigms interoperate together on a common neural-symbolic knowledge store called the Atomspace.

The high-level architecture of OCP, shown in Fig. 16, involves the use of multiple cognitive processes associated with multiple types of memory to enable an intelligent agent to execute the procedures that it believes have the best probability of working toward its goals in its current context. In a robot preschool context, for example (the context for which OCP is currently being primarily developed), the top-level goals will be simple things such as pleasing the teacher, learning new information and skills, and protecting the robot's body.

4.7.1. Memory and cognition in OpenCogPrime

OCP's memory types are the declarative, procedural, sensory, and episodic memory types that are widely discussed in cognitive neuroscience [54], plus attentional memory for allocating system resources generically, and intentional memory for allocating system resources in a goal-directed way. Table 2 overviews these memory types, giving key references and indicating the corresponding cognitive processes, and also indicating which of the generic patternist cognitive dynamics each cognitive process corresponds to (pattern creation, association, etc.).

The essence of the OCP design lies in the way the structures and processes associated with each type of memory are designed to work together in a closely coupled way, yielding cooperative intelligence going beyond what could be achieved by an architecture merely containing the same structures and processes in separate "black boxes".

Specifically, the inter-cognitive-process interactions in OpenCog are designed so that

- conversion between different types of memory is possible; though sometimes computationally costly (e.g. an item of declarative knowledge may with some effort be interpreted procedurally or episodically, etc.), and
- when a learning process concerned centrally with one type of memory encounters a situation where it learns very slowly, it can often resolve the issue by converting some of the relevant knowledge into a different type of memory: i.e. **cognitive synergy.**

Please cite this article as: B. Goertzel, et al., World survey of artificial brains, Part II: Biologically inspired cognitive architectures, Neurocomputing (2010), doi:10.1016/j.neucom.2010.08.012

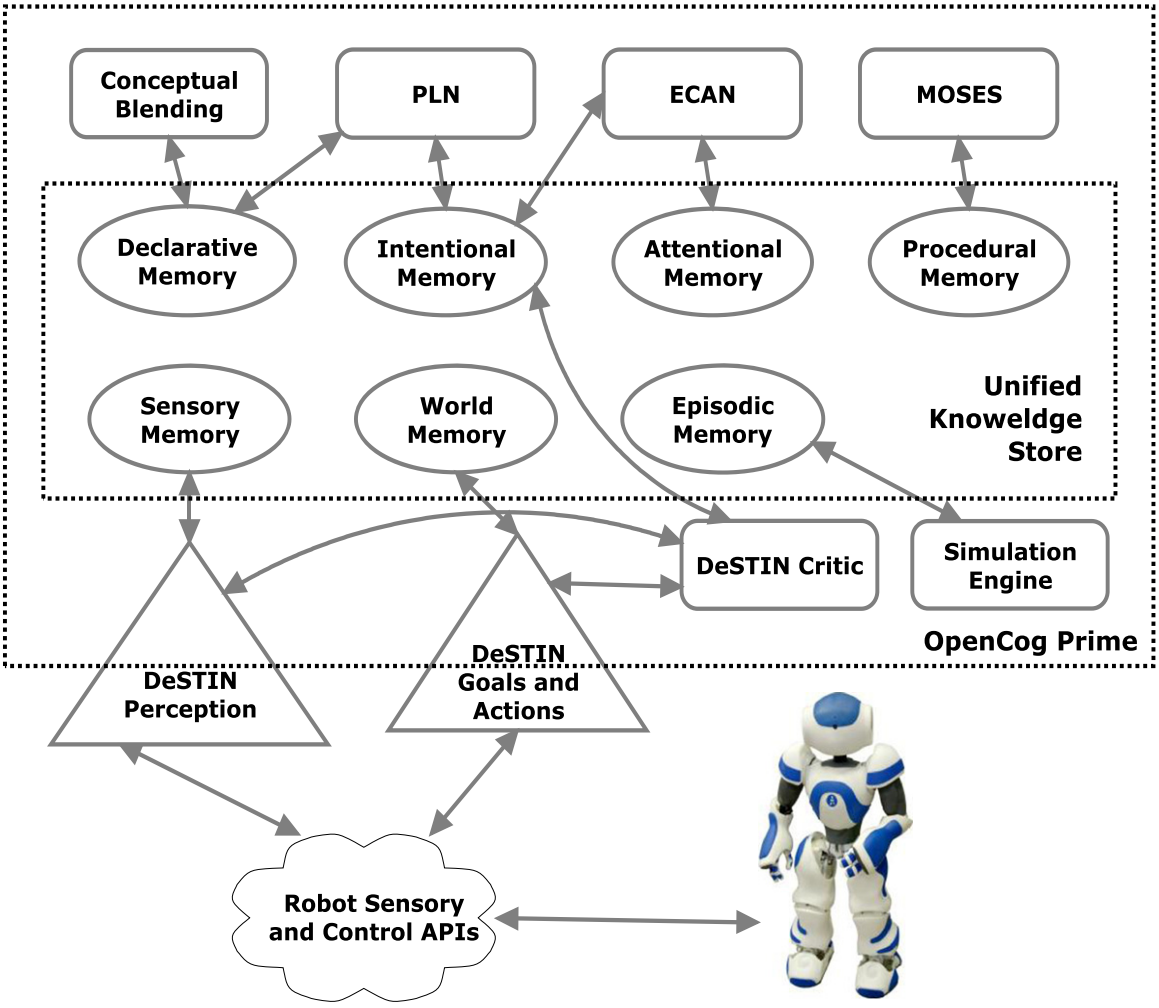


Fig. 16. High-level CogBot architecture diagram.

Table 2
Memory types and cognitive processes in OpenCogPrime. The third column indicates the general cognitive function that each specific cognitive process carries out, according to the patternist theory of cognition.

Memory type	Specific cognitive processes	General cognitive functions
Declarative	Probabilistic logic networks (PLN) [55]; conceptual blending [56]	Pattern creation
Procedural	MOSES (a novel probabilistic evolutionary program learning algorithm) [57]	Pattern creation
Episodic	Internal simulation engine [58]	Association, pattern creation
Attentional	Economic attention networks (ECAN) [59]	Association, credit assignment
Intentional	Probabilistic goal hierarchy refined by PLN and ECAN, structured according to MicroPsi [60]	Credit assignment, pattern creation
Sensory	In CogBot, this will be supplied by the DeSTIN component	Association, attention allocation, pattern creation, credit assignment

4.7.2. OpenCog versus MicroPsi

OCP's dynamics has both goal-oriented and "spontaneous" aspects; here for simplicity's sake we will focus here on the goal-oriented ones. The basic goal-oriented dynamic of the OCP system, within which the various types of memory are utilized, is driven by implications known as "cognitive schematics", which take the form

Context & Procedure $\Rightarrow$ Goal$\langle p \rangle$

summarized $C \wedge P \Rightarrow G$). Semi-formally, such implication may interpret to mean: "If the context C appears to hold currently, then if I enact the procedure P, I can expect to achieve the goal G with certainty p." There is a strong parallel with production rules, and an even stronger parallel with MicroPsi's triplets, which incorporate uncertainty in a similar way.

Apart from the incorporation of uncertainty and the ability to cope with missing information, the biggest difference between OCP/MicroPsi schematics/triplets and production rules or other similar constructs, is that in OCP this level of knowledge representation is not the only important one. CLARION seeks to work around the limitations of production rules by using them only for explicit knowledge representation, and then using a totally separate subsymbolic knowledge store for implicit knowledge.

Please cite this article as: B. Goertzel, et al., World survey of artificial brains, Part II: Biologically inspired cognitive architectures, Neurocomputing (2010), doi:10.1016/j.neucom.2010.08.012

In OCP, similarly to MicroPsi, both explicit and implicit knowledge are stored in the same graph of nodes and links, with

- explicit knowledge stored in probabilistic logic based nodes and links such as cognitive schematics, and
- implicit knowledge stored in patterns of activity among these same nodes and links, defined via the activity of the "importance" values associated with nodes and links and propagated by the ECAN attention allocation process.

The meaning of a cognitive schematic in OCP is hence not entirely encapsulated in its explicit logical form, but resides largely in the activity patterns that ECAN causes its activation or exploration to give rise to. And this fact is important because the synergetic interactions of system components are in large part modulated by ECAN activity. Without the real-time combination of explicit and implicit knowledge in the system's knowledge graph, the synergetic interaction of different cognitive processes would not work so smoothly, and the emergence of effective high-level structures in the system's knowledge base would be less likely.

4.7.3. Current and prior applications of OpenCog

OpenCogPrime has been used for commercial applications in the area of natural language processing and data mining; for instance, see [62] where OpenCog's PLN reasoning and RelEx language processing are combined to do automated biological hypothesis generation based on information gathered from PubMed abstracts. It has also been used to control virtual agents in virtual worlds [58], using an OpenCog variant called the OpenPetBrain (see Fig. 17) for a screenshot of an OpenPetBrain-controlled virtual dog, and see http://novamente.net/example for some videos of these virtual dogs in action. The CogBot project is a natural extension to humanoid robotics of this prior work in virtual worlds.

While the OpenCog virtual dogs do not display intelligence closely comparable to that of real dogs (or humans!), they do demonstrate a variety of relevant functionalities including

- learning new behaviors based on imitation and reinforcement,
- responding to natural language commands and questions, with appropriate actions and natural language replies, and
- spontaneous exploration of their world, remembering their experiences and using them to bias future learning and linguistic interaction.

A current research project involves hybridizing OCP with other systems that handle lower-level sensorimotor intelligence, and using the combination to control a humanoid robot. DeSTIN is being experimented within this role, along with evolvable neural networks as described in [63]. Fig. 17 shows the basic architecture of such an integration, in the DeSTIN context.

5. Concluding discussion

In this two-part paper, we have surveyed only a representative sample of the extant work in the field of artificial brains—there is a lot more out there. And furthermore, it seems likely that this current crop of artificial brain research projects will be dwarfed in the next few years by new entrants, due to the combination of Moore's Law and its allies with our equally rapidly increasing knowledge of neural architecture and micro-circuitry.

But even from the limited amount of work surveyed here, one may draw some interesting conclusions. One fairly obvious phenomenon worth drawing attention to is the gap between the BICA and brain-simulation approaches. Both approaches seek to

Fig. 17. Screenshot of OpenCog-controlled virtual dog.

Please cite this article as: B. Goertzel, et al., World survey of artificial brains, Part II: Biologically inspired cognitive architectures, Neurocomputing (2010), doi:10.1016/j.neucom.2010.08.012

leverage recent advances in hardware and neuroscience to create artificial brains, but so far the two approaches display very different strengths.

Brain simulations tell us about cortical columns and about the way collections of neurons "spontaneously" organize into collectives, but they do not yet tell us anything specific about how brains achieve goals, select actions or process information. On the other hand, BICAs tell us how brains may do things, but so far their intelligent behaviors are quite simplistic compared to real brains. This differential may be due to processing power, or as we conjecture it may be because the BICAs lack the chaotic, complex generativity that comes from neural nonlinear dynamics—i.e. they have the sensible and brainlike higher-level structures, but lack the lower-level complexity and emergence that one sees in large-scale brain simulations.

Based on these observations, our conjecture is that the future of artificial brain research lies in three directions:

1. Large-scale brain simulations that simulate multiple brain regions and their interconnections, thus verging on being BICAs.
2. BICAs that integrate more detailed neural dynamics into their processing, enabling greater creativity and flexibility of response.
3. Hybrid architectures that link BICA elements with brain simulation elements.

By bringing large-scale brain simulations and BICAs together, we suggest, we will most rapidly progress toward the twin goals of understanding the brain and emulating human-like intelligence in digital computers.

References

[1] W. Duch, R. Oentaryo, M. Pasquier, Cognitive architectures: where do we go from here? Proceedings of the Second Conference on AGI, 2008.
[2] M. Pelikan, Hierarchical Bayesian Optimization Algorithm: Toward a New Generation of Evolutionary Algorithms, Springer, 2005.
[3] J. Laird, P. Rosenbloom, A. Newell, Soar: an architecture for general intelligence, Artificial Intelligence 33 (1987).
[4] J.R. Anderson, C. Lebiere, The newell test for a theory of cognition, Behavioral and Brain Sciences 26 (2003).
[5] W. Gray, M. Schoelles, C. Myers, Meeting newells other challenge: cognitive architectures as the basis for cognitive engineering, Behavioral and Brain Sciences (2009).
[6] D.E. Meyer, D.E. Kieras, A computational theory of executive cognitive processes and multiple-task performance: Part 1, Psychological Review 104 (1997).
[7] P. Langley, An adaptive architecture for physical agents, in: Proceedings of the 2005 IEEE/WIC/ACM International Conference on Intelligent Agent Technology, 2005.
[8] S. Shapiro, et al., Metacognition in sneps, AI Magazine 28 (2007).
[9] J.R. Anderson, J.M. Fincham, Y. Qin, A. Stocco, A central circuit of the mind, Trends in Cognitive Science 12 (4) (2008) 136–143.
[10] J. Hawkins, S. Blakeslee, On intelligence, Brown Walker (2006).
[11] I. Arel, D. Rose, R. Coop, Destin: a scalable deep learning architecture with application to high-dimensional robust pattern recognition, in: Proceedings of the AAAI Workshop on Biologically Inspired Cognitive Architectures, November 2009.
[12] I. Arel, D. Rose, T. Karnowski, A deep learning architecture comprising homogeneous cortical circuits for scalable spatiotemporal pattern inference, in: NIPS 2009 Workshop on Deep Learning for Speech Recognition and Related Applications, December 2009.
[13] R.C. O'Reilly, T.S. Braver, J.D. Cohen, in: A. Miyake, P. Shah (Eds.), A biologically-based computational model of working memory, in: Models of Working Memory, 1999, pp. 375–411.
[14] J. Fleischer, J. Gally, G. Edelman, and J. Krichmar, Retrospective and prospective responses arising in a modeled hippocampus during maze navigation by a brain-based device, in: Proceedings of the National Academy of Sciences, 104, 2007.
[15] J. Modayil, B. Kuipers, Autonomous development of a grounded object ontology by a learning robot, AAAI-07 (2007).
[16] J. Mugan, B. Kuipers, Towards the application of reinforcement learning to undirected develop-mental learning, in: International Conference on Epigenetic Robotics, 2008.
[17] J. Mugan, B. Kuipers, Autonomously learning an action hierarchy using a learned qualitative state representation, IJCAI-09 (2009).
[18] G.E. Hinton, S. Osindero, Y. Teh, A fast learning algorithm for deep belief nets, Neural Computation 18 (2006) 1527–1554.
[19] Y. LeCun, B. Boser, J.S. Denker, et al., Handwritten digit recognition with a back-propagation network, Advances in Neural Information Processing Systems 2 (1990).
[20] Juergen Schmidhuber, Curious model-building control systems, in: Proceedings of the International Joint Conference on Neural Networks, 1991.
[21] Juergen Schmidhuber, A possibility for implementing curiosity and boredom in model-building neural controllers, in: Proceedings of the International Conference on Simulation of Adaptive Behavior: From Animals to Animats, 1991.
[22] Juergen Schmidhuber, Reinforcement-driven information acquisition in non-deterministic environments, in: Proceedings of the ICANN'95, 1995.
[23] Juergen Schmidhuber, Exploring the Predictable, Advances in Evolutionary Computing, Springer, 2002.
[24] J. Han, S. Zeng, K. Tham, M. Badgero, J.Weng, Dav: a humanoid robot platform for autonomous mental development, in: Proceedings of the Second International Conference on Development and Learning, 2002.
[25] J. Weng, W.S. Hwang, Y. Zhang, C. Yang, R. Smith, Developmental humanoids: humanoids that develop skills automatically, in: Proceedings of the First IEEE-RAS International Conference on Humanoid Robots, 2000.
[26] J. Weng, W.S. Hwangi, From neural networks to the brain: autonomous mental development, IEEE Computational Intelligence Magazine (2006).
[27] A. Baranes, P.-Y. Oudeyer, R-iac: robust intrinsically motivated active learning, in: Proceedings of the IEEE International Conference on Learning and Development, vol. 33, Shangha, China, 2009.
[28] F. Kaplan, Neurorobotics: an experimental science of embodiment, Frontiers in Neuroscience (2008).
[29] P. Oudeyer, F. Kaplan, Discovering communication, Connection Science (2006).
[30] G. Metta, G. Sandini, D. Vernon, L. Natale, F. Nori, The icub humanoid robot: an open platform for research in embodied cognition, in: Performance Metrics for Intelligent Systems Workshop (PerMIS 2008), 2008.
[31] J. Schmidhuber, Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts, Connection Science (2006).
[32] B. Bakker, J. Schmidhuber, Hierarchical reinforcement learning based on subgoal discovery and subpolicy specialization, in: Proceedings of the Eighth Conference on Intelligent Autonomous Systems, 2004.
[33] B. Bakker, V. Zhumatiy, G. Gruener, J. Schmidhuber, Quasi-online reinforcement learning for robots, in: Proceedings of the International Conference on Robotics and Automation, 2006.
[34] J. Simsek, A. Barto, An intrinsic reward mechanism for efficient exploration, in: Proceedings of the Twenty-Third International Conference on Machine Learning, 2006.
[35] S. Singh, A. Barto, N. Chentanez, Intrinsically motivated reinforcement learning, in: Proceedings of the Neural Information Processing Systems, vol. 17, 2005.
[36] M.H. Lee, Q. Meng, F. Chao, Developmental learning for autonomous robots, Robotics and Autonomous Systems 55 (2007) 750–759.
[37] M.H. Lee, Q. Meng, F. Chao, Staged competence learning in developmental robotics, Adaptive Behavior 15(3) (2007) 241–255.
[38] N. Nilsson, The physical symbol system hypothesis: status and prospects, 50 Years of AI, Festschrift, Lecture Notes in Artificial Intelligence 33 (2009) 48–50.
[39] R. Sun, X. Zhang, Top-down versus bottom-up learning in cognitive skill acquisition, Cognitive Systems Research 5 (2004).
[40] A. Nestor, B. Kokinov, Towards active vision in the dual cognitive architecture, International Journal on Information Theories and Applications 11 (2004).
[41] S. Franklin, The lida architecture: adding new modes of learning to an intelligent, autonomous, software agent, in: International Conference on Integrated Design and Process Technology, 2006.
[42] J. Bach, in: Principles of Synthetic Intelligence, Oxford University Press, 2009.
[43] N. Cassimatis. Adaptive algorithmic hybrids for human-level artificial intelligence, in: B. Goertzel, P. Wang (Eds.), Advances in Artificial General Intelligence, IOS Press, 2007, pp. 151–163.
[44] L. Shastri, V. Ajjanagadde, From simple associations to systematic reasoning: a connectionist en-coding of rules, variables, and dynamic bindings using temporal synchrony, Behavioral and Brain Sciences 16(3) (1993).
[45] D. Hofstadter, Fluid Concepts and Creative Analogies, Basic Books, 1996.
[46] B. Goertzel, The hidden pattern. Brown Walker, 2006.
[47] B. Goertzel, Opencog prime: a cognitive synergy based architecture for embodied artificial general intelligence, in: Proceedings of the ICCI-09, Hong Kong, 2009.
[48] M. Marvin, The Society of Mind, Simon and Schuster, New York, March 15, 1988.
[49] M. Marvin, The Emotion Machine, Simon and Schuster, 2006.
[50] P. Singh, EM-ONE: an architecture for reflective commonsense thinking, Ph.D. Thesis, MIT, June 2005.
[51] J. Albus, A. Meystel, Engineering of Mind: An Introduction to the Science of Intelligent Systems, Wiley and Sons, 2001.
[52] D. Friedlander, S. Franklin, in: Ben Goertzel, Pei Wang (Eds.), LIDA and a theory of mind, in: Artificial General Intelligence (AGI-08), IOS Press, Memphis, TN, USA, 2008.

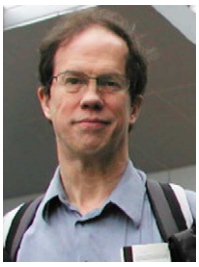
Please cite this article as: B. Goertzel, et al., World survey of artificial brains, Part II: Biologically inspired cognitive architectures, Neurocomputing (2010), doi:10.1016/j.neucom.2010.08.012

[53] S. Franklin, The lida architecture: adding new modes of learning to an intelligent, autonomous, software agent, in: International Conference on Integrated Design and Process Technology, 2006.
[54] E. Tulving, R. Craik, The Oxford Handbook of Memory, Oxford University Press, 2005.
[55] B. Goertzel, I.G.M. Ikl_e, A. Heljakka, Probabilistic Logic Networks, Springer, 2008.
[56] G. Fauconnier, M. Turner, The way we think: conceptual blending and the mind's hidden complexities, Basic (2002).
[57] M. Looks, Competent program evolution, Ph.D. Thesis, Computer Science Department, Washington University, 2006.
[58] B. Goertzel C.P., et al, An integrative methodology for teaching embodied non-linguistic agents, applied to virtual animals in second life, in: Proceedings of the AGI-08, 2008.
[59] B. Goertzel, J. Pitt, M. Ikle, C. Pennachin, R. Liu, Glocal memory: a design principle for artificial brains and minds, Neurocomputing, this issue, doi:1016/j.neucom.2009.10.033.
[60] J. Bach, Principles of Synthetic Intelligence, Oxford University Press, 2009.
[61] 〈http://ccrg.cs.memphis.edu/tutorial/correlates.html〉.
[62] B. Goertzel, H. Pinto, C. Pennachin, Using dependency parsing and probabilistic inference to extract relationships between genes, proteins and malignancies implicit among multiple biomedical research abstracts, in: Proceedings of the BioNLP Workshop/HLT-NAACL, 2006.
[63] B. Goertzel H. de Garis, in: XIA-MAN: an extensible, integrative architecture for intelligent humanoid robotics, in: Proceedings of the BICA-08, 2008, pp. 86–90.
[64] J.S. Albus, A.J. Barbera, RCS: a cognitive architecture for intelligent multi-agent systems, Annual Reviews in Control 29 (1) (2005) 87–99.
[65] J.S. Albus, Reverse engineering the brain, in: Proceedings of the AAAI Fall Symposium, Washington, DC, 2008.

Dr. Ben Goertzel is CEO of AI software company Novamente LLC and bioinformatics company Biomind LLC; Chief Technology Officer of biopharma firm Genescient Corporation; leader of the open-source OpenCog AI software project; Chairman of Humanity+; Advisor to the Singularity University and Singularity Institute; Research Professor in the Fujian Key Lab for Brain-Like Intelligent Systems at Xiamen University, China; and general Chair of the Artificial General Intelligence conference series. His research work encompasses artificial general intelligence, natural language processing, cognitive science, data mining, machine learning, computational finance, bioinformatics, virtual worlds and gaming and other areas, He has published a dozen scientific books, nearly 90 technical papers, and numerous journalistic articles. Before entering the software industry he served as a university faculty in several departments of mathematics, computer science and cognitive science, in the US, Australia and New Zealand. He has three children and too many pets, and in his spare time enjoys creating avant-garde fiction and music, and the outdoors.

Ruiting Lian is a Ph.D. student in the Fujian Key Lab for Brain-Like Intelligent Systems at Xiamen University, China. Her general research interests are Natural Language Processing and Artificial Intelligence.

Itamar Arel received the B.Sc., M.Sc., and Ph.D. degrees in electrical and computer engineering and the M.B.A. degree from the Ben-Gurion University, Israel, in 1995, 1998, 2003, and 2002, respectively. He is currently an Associate Professor with the Department of Electrical Engineering and Computer Science at The University of Tennessee. During 2000–2003, he was with TeraCross, Inc., a semiconductor company developing terabit per-second switch fabric integrated circuits, where he held several key positions, including chief scientist. His research interests include theoretical and practical aspects of machine learning, biologically inspired cognitive architectures, and high-performance computing.



Hugo de Garis (born on 1947, Sydney, Australia) is a researcher in the sub-field of artificial intelligence (AI) known as evolvable hardware. He became known in the 1990s for his research on the use of genetic algorithms to evolve neural networks using three dimensional cellular automata inside field programmable gate arrays. He claimed that this approach would enable the creation of what he terms "artificial brains" which would quickly surpass human levels of intelligence. He has also recently been noted for his exploration of the concept of technological singularity, and its potential social and political implications. de Garis originally studied theoretical physics, but he abandoned this field in favour of artificial intelligence. In 1992 he received his Ph.D. from Université Libre de Bruxelles, Belgium. He worked as a researcher at Advanced Telecommunications Research institute international (ATR), Japan from 1994 to 2000, a researcher at Starlab, Brussels from 2000 to 2001, and an associate professor of computer science at Utah State University From 2001 to 2006. He has recently retired from Xiamen University, where he taught theoretical physics and computer science, and ran the Artificial Brain lab.

Mr. Shuo Chen was born in Hubei, China. He got his MS degree in software engineering from Wuhan University, China in 2007. Presently he is a Ph.D. candidate at the Department of Cognitive Science of Xiamen University. His general interests are in the areas of computational vision and artificial neural networks.

Please cite this article as: B. Goertzel, et al., World survey of artificial brains, Part II: Biologically inspired cognitive architectures, Neurocomputing (2010), doi:10.1016/j.neucom.2010.08.012