Frontiers in Artificial Intelligence and Applications, Vol. 171
(Ed. by Pei Wang, Ben Goertzel, and Stan Franklin), IOS Press, pp. 122-136.

Cognitive Architectures: Where do we go from here?

Włodzisław DUCH[a,1], Richard J. OENTARYO[b], and Michel PASQUIER[b]

[a]*Nicolaus Copernicus University, Torun, Poland*
[b]*Nanyang Technological University, Singapore*

Abstract. Cognitive architectures play a vital role in providing blueprints for building future intelligent systems supporting a broad range of capabilities similar to those of humans. How useful are existing architectures for creating artificial general intelligence? A critical survey of the state of the art in cognitive architectures is presented providing a useful insight into the possible frameworks for general intelligence. Grand challenges and an outline of the most promising future directions are described.

Keywords: cognitive architectures, artificial general intelligence, neurocognitive models, intelligent agents.

1. Introduction

A long-term goal for artificial general intelligence (AGI) is to create systems that will exceed human level of competence in a large number of areas. There is a steady progress toward this goal in several domains, including recognition of specific patterns, memorization and retrieval of vast amount of information, interpreting signals and other types of numerical information, autonomous control, board games and reasoning in restricted domains. Yet even in lower level cognitive functions, such as object recognition or scene analysis artificial systems are still far behind the natural ones. Higher-level cognitive functions, such as language, reasoning, problem solving or planning, involve complex knowledge structures and are much more difficult to realize. Various types of memory stored by the brain facilitate recognition, association, semantic interpretation and rapid retrieval of large amounts of complex information patterns. At quite basic level organization of storage and retrieval of information in computers is completely different than in brains. Computing architectures are universal only in principle, in practice they always constrain information processing in specific ways. Computers are better in many tasks than brains, and vice versa, brains are better in many important tasks then computers. It is not clear at all whether cognitive architectures (CA) running on conventional computers, will reach the flexibility of the brain in lower or higher-level cognitive functions.

Traditionally higher cognitive functions, such as thinking, reasoning, planning, problem solving or linguistic competencies have been the focus of artificial intelligence (AI), relaying on symbolic problem solving to build complex knowledge structure. These functions involve sequential search processes [1], while lower cognitive functions, such as perception, motor control, sensorimotor actions, associative memory recall or categorization, are accomplished on a faster time scale in a parallel way, without stepwise deliberation. Embodiment is a powerful trend in robotics and there is now a general agreement that the meaning of many concepts should be grounded in embodied, sensorimotor representations. However, the actual limitations of the symbolic and the embodied approaches to cognition are not known. Perhaps the dream of creating a General Problem Solver [2] may be realized with relatively minor extensions to symbolic cognitive architectures, while understanding animal behavior and robotic applications may require embodied cognition. Analysis of existing cognitive architectures should facilitate understanding of limitations of different approaches. Many general ideas seem to explain everything but do not scale up well to real applications, therefore a clear notion what exactly AGI should do is necessary.

---

[1] Corresponding author, Department of Informatics, Nicolaus Copernicus University, Grudziądzka 5, 87100 Toruń, Poland, Google: "W. Duch" for more info.

2. Grand challenges for AGI

What should be required from an AI system to be worthy of the "Artificial General Intelligence" name? Artificial Intelligence has focused on many specific approaches to problem solving, useful for development of expert systems, neglecting its initial ambitious goals. One requirement for AGI, storing and manipulation of vast amount of knowledge, has been addressed by the Cyc project [3]. Started in 1984 a huge frame-based knowledge base has been constructed, but its "potential applications" list has not been replaced by actual applications for decades. The biggest failure of AI community is evident in the language-related domains, for example in general purpose conversational systems, developed mostly in the form of various chatterbots by commercial companies and enthusiastic individuals. Restricted form of the Turing test [4] (the full test being too difficult to try), called Loebner Prize competition [5], has been won for almost two decades by chatterbots based on old template matching techniques, or more recently contextual pattern matching techniques. Such programs have no chance to develop real understanding of language and use it in meaningful dialogs or texts analysis, but may be used for stereotyped question/answer systems or "impersonation". Carpenter and Freeman have proposed a "personal Turing test" [6], where a person tries to guess if the conversation is done with a program or a real personally known individual. Human behavior includes the ability to impersonate other people, and the personal Turing test may be an interesting landmark step on the road to general intelligence.

Another area that poses remarkable challenge to AI is word games, and in particular the 20-questions game. Word games require extensive knowledge about objects and their properties, but not about complex relations between objects. Different methods of knowledge representation may be used in different applications, from quite simple, facilitating efficient use of knowledge, to quite involved, needed only in deep reasoning. In fact simple vector-space techniques for knowledge representation are sufficient to play the 20-question game [7]. Success in learning language depends on automatic creation and maintenance of large-scale knowledge bases, bootstraping on the resources from the Internet. Question/answer systems pose even more demanding challenge, and in this area a series of competitions organized at Text Retrieval Conference (TREC) series may be used to measure progress. Intelligent tutoring systems are the next great challenge, but there seem to be no clearly defined milestones in this field.

Feigenbaum [8] proposed as a grand challenge building a super-expert system in a narrow domain. This seems to go in a direction of specialized, rather than general intelligence, but one may argue that a super-expert without general intelligence needed for communication with humans is not feasible. Sophisticated reasoning by human experts and artificial systems in such fields as mathematics, bioscience or law may be compared by a panel of experts who will pose problems, rise questions, and ask for further explanations to probe the understanding of the subject. A good example of such challenge is provided by the Automated Theorem Proving (ATM) System Competitions (CASC) in many sub-categories. An interesting step toward general AI in mathematics would be to create general theorem provers, perhaps using meta-learning techniques that rely on specialized modules. Automatic curation of genomic/pathways databases and creation of models of genetic and metabolic processes for various organisms poses great challenges for super-experts, as the amount of information in such databases exceeds by far human capacity to handle it.

Defining similar challenges and milestones towards AGI in other fields is certainly worthwhile. The ultimate goal would be to develop programs that will advice human experts in their work, evaluating their reasoning, perhaps even adding some creative ideas. DARPA in the "Personal Assistants that Learn" (PAL) program sponsors a large-scale effort in similar direction. Nilsson has argued [9] for development of general purpose educable systems that can be taught skills needed to perform human jobs, and to measure which fraction of these jobs can be done by AI systems. Building one such system replaces the need for building many specialized systems, as already Allan Turing has noted proposing a "child machine" in his 1950 paper [4]. Some human jobs are knowledge-based and can be done by information processing systems, where progress may be measured by passing a series of examinations, as is done in such fields as accounting. However, most human jobs involve manual labor, requiring senso-motoric coordination that should be mastered by household robots or autonomous vehicles. The DARPA Urban Challenge competition (2007) requires integration of computer vision, signal processing, control and some reasoning. It is still simpler than control of a humanoid robot, where direct interaction of robots with people will require an understanding of perception, controlling of attention, learning casual models from observations, and hierarchical learning with different temporal scales. Creation of partners or personal assistants, rather than complete replacements for human workers, may be treated as a partial success. Unfortunately specific milestones for this type of applications have yet to be precisely defined. Some ordering of different jobs from the point of view of difficulty to learn them could be worthwhile. In fact many jobs have already been completely automatized, reducing the number of people in manufacturing, financial services, printing houses etc. In most cases alternative organization of work is to be credited for reduction in the number of jobs (plant and factory automation, ATM machines, vending machines), not because of deployment of AI systems.

A detailed roadmap to AGI should thus be based on detailed analysis of the challenges, relationships between various functions that should be implemented to address them, system requirements to achieve these functions and classes of problems that should be solved at a given stage.

3. Cognitive architectures

Cognitive architectures are frequently created to model human performance in multimodal multiple task situations [1][10] rather than to create AGI. A short critical review of selected cognitive architectures that can contribute to development of AGI is provided below. Allen Newell in his 1990 book *Unified Theories of Cognition* [1] provided 12 criteria for evaluation of cognitive systems: adaptive behavior, dynamic behavior, flexible behavior, development, evolution, learning, knowledge integration, vast knowledge base, natural language, real-time performance, and brain realization. These criteria have been analyzed and applied to ACT-R, Soar and classical connectionist architectures [11] but such fine-grained categorization makes comparison of different systems rather difficult. Without going into such details we shall propose below a simpler taxonomy, give some examples of different types of cognitive systems that are currently under development, and provide a critique and some recommendations for better systems. Surveys on the system organization and working mechanisms of a few cognitive architectures that have already been published [12] were not written from the AGI point of view.

Two key design properties that underlie the development of any cognitive architecture are *memory* and *learning.* The importance of memory has been stressed from different perspectives in a few recent books [13]-[15]. Various types of memory serve as a repository for background knowledge about the world and oneself, about the current episode of activity, while learning is the main process that shapes this knowledge. Together learning and memory form the rudimentary aspects of cognition on which higher-order functions and intelligent capabilities, such as deliberative reasoning, planning, and self-regulation, are built. Organization of memory depends on the knowledge representation schemes. A simple taxonomy of cognitive architectures based on these two main features leads to a division of different approaches into three main groups (Fig. 1): *symbolic, emergent,* and *hybrid* models.

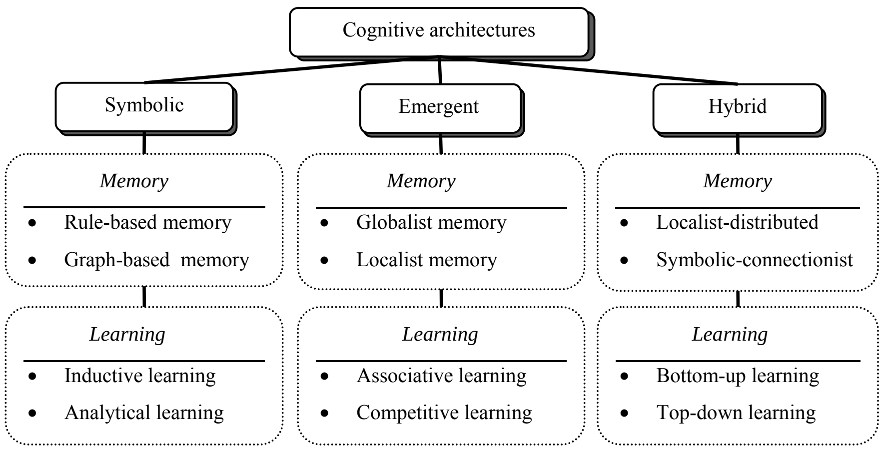

Fig. 1. Simplified taxonomy of cognitive architectures

Roughly speaking symbolic architectures focus on information processing using high-level symbols or declarative knowledge, in a classical AI top-down, analytic approach. Emergent architectures use low-level activation signals flowing through a network consisting of numerous processing units, a bottom-up process relaying on the emergent self-organizing and associative properties. Hybrid architectures result from combining the symbolic and emergent paradigms in one way or another. The memory and learning aspects of these three broad classes of approaches are investigated in details below.

3.1. Symbolic architectures

There is a strong link between the type of architecture and problems it is supposed to solve. The use of symbols as the key means to support information processing originates from the physical symbol system hypothesis [1], which has been motivated by research on memory and problem solving. A physical symbol system has the ability to input, output, store and alter symbolic entities, and to carry out appropriate actions in order to reach its goals. The majority of symbolic architectures utilize a centralized control over the information flow from sensory inputs through memory to motor outputs. This approach stresses the working memory executive functions, with an access to semantic memory for knowledge retrieval. *Rule-based* representations of perception-action memory in knowledge-based production systems embody the logical reasoning skills of human experts. *Graph-based* representations are typically encoded as directed graph structures comprising nodes for symbolic entities and their attributes, and edges for relationships among them. Main examples of this sort of knowledge representation are semantic networks and conceptual graphs [16], frames/schemata [17], and reactive action packages (RAPs) [18]. The underlying paradigms of these approaches remain very similar, and are sometimes even equivalent.

Substantial efforts have been made over the years to introduce *analytical and inductive* learning techniques to symbolic systems. The former aims at exploiting existing general/specific facts to infer other facts that they entail logically. Prominent examples include explanation-based learning (EBL) [19] and analogical learning [20]. Inductive machine learning, on the other hand, seeks to derive from specific facts or examples general rules which capture the underlying domain structure. Well-known examples of this kind include knowledge-based inductive learning (KBIL) [21] and delayed reinforcement learning [22]. Many ambitious cognitive architectures have been proposed and abandoned after a period of vigorous activity. Only those potential candidates for AGI that are still actively developed are reviewed below.

SOAR (*State, Operator And Result*) is a classic example of expert rule-based cognitive architecture designed to model general intelligence [1],[23]. Based on theoretical framework of knowledge-based systems seen as an approximation to physical symbol systems, SOAR stores its knowledge in form of production rules, arranged in terms of operators that act in the problem space, that is the set of states that represent the task at hand. The primary learning mechanism in SOAR is termed chunking, a type of analytical EBL technique for formulating rules and macro-operations from problem solving traces [23]. In recent years many extensions of the SOAR architecture have been proposed: reinforcement learning to adjust the preference values for operators, episodic learning to retain history of system evolution, semantic learning to describe more abstract, declarative knowledge, visual imagery, emotions, moods and feelings used to speed up reinforcement learning and direct reasoning [24]. SOAR architecture has demonstrated a variety of high-level cognitive functions, processing large and complex rule sets in planning, problem solving and natural language comprehension (NL-SOAR) in real-time distributed environments (see [25] for more references). At present SOAR architecture has not yet integrated all these extensions. A few additional ones, like memory decay/forgetting, attention and information selection, learning hierarchical representations, or handling uncertainty and imprecision, will also be useful. The design of the perceptual-motor systems within SOAR is fairly unrealistic, requiring users to define their own input and output functions for a given domain. It remains to be seen how well numerous problems that face such an extension can be handled using the existing architecture as a base.

EPIC (*Executive Process Interactive Control*) is a cognitive architecture for building computational models that subsume many aspects of human performance [10]. It aims at capturing human perceptual, cognitive and motor activities through several interconnected processors working in parallel, and to build models of human-computer interaction for practical purposes. 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. Although EPIC is focused on multiple simple tasks in one experiment it has been connected to SOAR for problem solving, planning and learning, and the EPIC-SOAR combination has been applied to air traffic control simulation [26].

ICARUS project [27] defines 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. Concepts are matched to percepts in a bottom-up way and goals are matched to skills in a top-down way. Conceptual memory contains knowledge about general classes of objects and their relationships, while skill memory stores knowledge about the ways of doing things. Each comprises a long-term memory (LTM) and a short-term memory (STM). The LTM is organized in a hierarchical fashion, with the conceptual memory directing bottom-up, percept-driven inference and skill memory controlling top-down, goal-driven selection of actions. The acquisition of knowledge in ICARUS is achieved through hierarchical, incremental reinforcement learning, propagating reward values backward through time. Since skills comprise sub-skills, the system learns to calculate the value estimates over a stack of state-action pairs, instead of just a single pair as in the traditional reinforcement learning. This hierarchical processing strategy yields in turn a much faster learning rate than that of the standard reinforcement learning. The hierarchical memory organization and learning procedure have equipped ICARUS with the ability to focus its attention on a specific object or event in its sensor range and to reduce the reaction time and the search space requirements. Through its planning and memory modules, ICARUS is also able to incrementally learn new concepts in an efficient manner by constructing a feature tree that the system can comprehend. Relational reinforcement learning that gives priority to high-utility beliefs and rapidly finds most useful inferences handles well large memory hierarchies [28]. Interesting applications of this architecture to in-city driving, blocks world or free-cell solitaire have been demonstrated. One vital problem is the lack of concurrent processing to cope with asynchronous inputs from multiple sensors while coordinating resources and actions across different modalities. Issues related to uncertainty have also been largely ignored.

NARS (Non-Axiomatic Reasoning System) [29] project has been developed for over two decades. It is a reasoning system based on a language for knowledge representation, an experience-grounded semantics of the language, a set of inference rules, a memory structure, and a control mechanism, carrying out various high-level cognitive tasks as different aspects of the same underlying process. The non-axiomatic logic is used for adaptation with insufficient knowledge and resources, operating on patterns that have the "truth-value" evaluated according to the system's "experience" with using these patterns. This approach allows for emergence of experience-grounded semantics, and inferences defined on judgments. While several working NARS prototypes of increasing complexity have been built they are solving only rather simple problems.

SNePS (Semantic Network Processing System) [30] is a logic, frame and network-based knowledge representation, reasoning, and acting system that went through over three decades of development. It stores knowledge and beliefs of some agent in form of assertions (propositions) about various entities. Each knowledge representation has its own inference scheme, logic formula, frame slots and network path, integrated in SNIP, the SNePS Inference Package. When a belief revision system discovers contradiction some hypotheses that led to the contradiction have to be unasserted by the user and the system removes all those assertions that depend on it. The SNePS Rational Engine controls plans and sequences of actions using frames and believing/disbelieving propositions. The natural language processing system works with English morphological analyzer/synthesizer and generalized augmented transition network grammar interpreter. SNePS has been used for commonsense reasoning, natural language understanding and generation, contextual vocabulary acquisition, control of simulated cognitive agent that is able to chat with the users, question/answer system and other applications. Interesting inferences have been demonstrated, but the program has not yet been used in a large-scale real application, for example a chatterbot.

3.2. Emergent Paradigm Architectures

Emergent cognitive architectures are inspired by connectionist ideas [31]. Their relation to the processes that take place in the brain may be rather distant. Processing elements (PEs) form network nodes that interact with each other in a specific way changing their internal states and revealing interesting emergent properties. There are two complementary approaches to memory organization, *globalist* and *localist*. The Multi-Layer Perceptron (MLP) and other neural networks based on delocalized transfer functions process information in a distributed, global way. All parameters of such networks influence their outputs. Generalization of learned responses to novel stimuli is usually good, but learning new items may lead to catastrophic interference with old knowledge [32]. The basis set expansion networks that use localized functions (such as Gaussians) are examples of localist networks; the output signals for a given input depend only on a small subset of units that are activated. However, it should be remembered that a modular organization of globalist network will easily create a subgroups of processing elements that react in a local way [33].

The learning methodologies for emergent architectures are quite diverse [31],[32]. A*ssociative learning* creates a mapping of specific input representation to specific output representation and in this sense remembers the reactions, heteroassociations or enables pattern completion (autoassociations). In this case, learning may be either guided directly by a set of correct "target" signals or indirectly by certain critic inputs, which correspond to the supervised and reinforcement learning paradigms in AI, respectively. In *competitive learning* all PEs compete to become active and learn in an unsupervised fashion. The simplest form of this procedure is the winner-takes-all (WTA) rule, which permits only one winning PE (or one per group) to learn at a time, while inhibiting others that lose the competition. Correlation-based learning using Hebb learning rule captures statistical properties of incoming signals creating an internal model of the environment [32]. A few emergent architectures that are still under active development are presented below. In applications to complex reasoning they have not yet reached the same level of maturity as symbolic architectures, but are much closer to natural perception and reasoning based on perceptions, rather than symbolic knowledge.

IBCA (*Integrated Biologically-based Cognitive Architecture*) is a large-scale emergent architecture that epitomizes the automatic and distributed notions of information processing in the brain [34]. The role of three regions in the brain is emphasized: posterior cortex (PC), frontal cortex (FC), and hippocampus (HC). The PC module assumes an overlapping, distributed localist organization that focuses on sensory-motor as well as multi-modal, hierarchical processing. The FC module employs a non-overlapping, recurrent localist organization in which working memory units are isolated from one another and contribute combinatorially (with separate active units representing different features). The HC module utilizes a sparse, conjunctive globalist organization where all units contribute interactively (not combinatorially) to a given representation. It permits rapid binding of all activation patterns across PC and FC (i.e., episodic memory), while reducing interference. The LEABRA learning algorithm includes error-driven learning of skills and Hebbian learning with inhibitory competition dynamics. In this framework, the PC and FC modules employ a slow integrative learning that blends many individual experiences to capture the underlying regularities of the world and to support sensory-motor activities. The HC module adds fast learning that retains and discriminates (instead of integrating) the individual experiences. This cooperation between HC and FC/PC reflects in turn the complementary learning paradigms in the brain.

A salient trait of IBCA is the knowledge-dependency merits (generalization, parallelism, flexibility) of content-specific distributed representation in the brain, missing in symbolic models. Complementary learning capacity resolves problems with knowledge consolidation, or transfer of short-term into long-term memory. Higher-level cognition (variable binding, chaining of operation sequences etc.) is a result of emergent power of activation-based processing (invocation, maintenance, and updating of active representations for self-regulation) in the FC module. These capacities have been validated in basic psychological tasks, such as Stroop test, dynamic sorting task, or visual feature binding [32]. However, the fine-grained structure of the architectural modules requires a large number of neurons to simulate cognitive functions, and raises the issues of system scalability. The current architecture is limited to learning the ANN weight parameters, but not the network local structure. Representation of emotions for motivation and setting goals, as well as motor coordination and timing, is still missing. While this type of architecture may be used to explain human behavior probed by psychological or psycholinguistic experiments no-one has yet demonstrated how to use it for tasks that require reasoning, where many symbolic architectures reach the human competence level.

Cortronics is a new emergent architecture that models the biological functions of the cerebral cortex and thalamus systems (jointly termed thalamocortex) in the human brain [15]. Its memory organization consists of modular feature attractor circuits called lexicons. Each lexicon comprises further a localist cortical patch, a localist thalamic patch, and the reciprocal connections linking them. Essentially, each lexicon implements a large stable set of attractive states called symbols, each represented by a specific group of neurons. The number of neurons overlapping between each pair of symbols is relatively small, and each neuron representing one symbol may be used to represent many symbols. Accordingly, knowledge in Cortronics takes the form of parallel, indirect unidirectional links between the neurons representing one symbol in a lexicon and those describing a symbol in another lexicon. Each such knowledge link is termed an item of knowledge, and the collection of all these links is called a knowledge base. The union of cortical patches of all lexicons constitutes in turn the entire cortex, while that of the thalamic patches of all lexicons forms only a portion of thalamus. A competitive activation of symbols of lexicons, called confabulation, is used for learning and information retrieval. Confabulation is carried out by every lexicon when appropriate knowledge link inputs, and operation command inputs, arrive at the lexicon at once. The states of involved neurons evolve dynamically via the parallel, mutual interactions of these neurons, and the minority that ends up in the excited/active state denote conclusions, the symbol(s) that won the competition, or a null symbol, implying a "don't know" answer. The model predicts or anticipates the next state, move or word that should follow. This is quite new architecture and it is not yet clear how it can be extended to create AGI, as confabulation is not sufficient for reasoning with complex knowledge. However, it is an interesting process involved in anticipation, imagination and creativity [35][36], a process at a shorter time scale than reasoning processes.

The **NuPIC** (*Numenta Platform for Intelligent Computing*) is an emergent architecture based on the Hierarchical Temporal Memory (HTM) technology, which is modeled on the putative algorithm used by neocortex [13]. Network nodes are organized in a hierarchical way, with each node implementing learning and memory functions. Hierarchical organization is motivated by the growing size of cortical receptive fields in the information streams that connect primary sensory cortices with secondary and higher-level association areas. This feature is also present in the IBCA architecture, where specific connectivity between different layers leads to growing and invariant object representation. HTM is unique in stressing the temporal aspect of perception, implementing memory for sequences of patterns that facilitate anticipation. Each level in the hierarchical network is trained separately to memorize spatio-temporal objects, and is able to recognize new, similar objects in a bottom-up/top-down process. The architecture has not yet been tested in larger applications, therefore it is hard to evaluate its potential and its limitations.

NOMAD (Neurally Organized Mobile Adaptive Device) automata are based on "neural Darwinism" theory [37]. Nomads, also known as Darwin automata, demonstrate the principles of emergent architectures for pattern recognition task in real time. They use many sensors for vision, range finders to provide a sense of proximity, prioproceptive sense of head direction and self-movement, artificial whiskers for texture sensing, artificial taste (conductivity) sensors. NOMAD is controlled by a large ($\sim 10^5$ neurons with $\sim 10^7$ synapses) simulated nervous system running on a cluster of powerful computers. The "Brain-Based Robotic Devices" that develop through behavioral tasks has elucidated a role of value systems based on reward mechanisms in adaptation and learning, importance of self-generated movement in development of perception, the role of hippocampus in spatial navigation and episodic memory (Darwin X-XI models), invariant visual object recognition (Darwin VI-VII), binding visual features of complex objects into a unified scene by neural synchrony due to the recurrent connections in visual pathways, implementation of concurrent, real-time processes. However, the emergence of higher-level cognition does not seem likely in this architecture.

A number of emergent architectures based on the global workspace theory of Baars [38] have been formulated in the last two decades, but very few reached implementation level. Shanahan has described a very simple implementation based on weightless neural network built from generalizing random access memory processing units and used to control simulated robot [39]. Other examples of architectures inspired by the global workspace theory are discussed in the hybrid systems subsection below.

Many other interesting emergent architectures have been discussed in recent years, but there is little experience with them due to the lack of software implementation to experiment with. Haikonen has written a book outlining an approach to conscious machines and discussing cognitive architecture for robot brains [40]. Anderson and his colleagues proposed the Erzatz brain project [41]. The autonomous mental development movement, motivated by the human mental development from infancy to adulthood, has been active for about a decade now [42], going in similar direction as the Edelman's Darwin projects, and Brooks' Cog project [43][44], that is creating a robotic agent for real-time interaction. Korner and Matsumoto argue [45] that cognitive architecture should control constraints that are used to select a proper algorithm from existing repertoire to solve a specific problem, or to create a new one if the stereotyped behaviors are not sufficient. This is very much in agreement with the meta-learning ideas in computational intelligence [46], where solving hard learning problems is done by learning which transformations should be composed to achieve the goal. The DARPA Biologically-Inspired Cognitive Architectures (BICA) program has already resulted in a several interesting proposals, such as the "TOSCA Comprehensive brain-based model of human mind" [47] written by a number of experts from leading USA institutions, which essentially came to the same conclusion.

3.3. Hybrid Paradigm Architectures

Given the relative strengths of the symbolic and emergent paradigms, it becomes clear that combining the two would offer a promising venue for developing a more complete framework for cognition [48]. Symbolic architectures are able to process information and realize high-level cognitive functions, such as planning and deliberative reasoning, in a way that resembles human expertise. However, the major issues in this approach are the formulation of symbolic entities from low-level information, as well as the handling of large amount of information and uncertainty. Emergent architectures are better suited for capturing the context-specificity of human performance and handling many pieces of low-level information simultaneously. Yet their main shortcoming is the difficulty in realizing higher-order cognitive functions. The potential benefit of a combined approach is therefore to have each method address the limitations of the other, allowing creation of a complete brain architecture that covers all levels of processing, from stimuli to higher-level cognition.

Research in this area has led to many proposals of hybrid cognitive architectures, which can be roughly divided in two classes based upon the memory type of the constituent modules: *localist-distributed* and *symbolic-connectionist* [48]. The first class of hybrid architectures comprises a combination of localist modules (with each concept specified by one PE node) and distributed modules (with each concept represented by a set of overlapping nodes). In comparison, the second class involves a mixture of symbolic modules (i.e., rule- or graph-based memory) and connectionist modules (either of localist or distributed type). Correspondingly, hybrid architectures can be categorized into two main classes according to their direction of learning: *top-down* and *bottom-up learning* [49]. The former involves a transition of knowledge from explicit (accessible) conceptual level to implicit (inaccessible) sub-conceptual level, while the latter goes from sub-conceptual level to conceptual level. The top-down learning can be achieved by pre-coding a set of expert rules at the top level (localist/symbolic module) and allowing the bottom-level (distributed ANN) to learn by observing actions guided by the top-level [49]. Conversely, bottom-up learning may be accomplished by extracting or translating implicit knowledge coded by a bottom-level module into a set of conceptual rules [33][50]. A few examples of hybrid cognitive architectures follow, focused on the memory organizations, learning methodologies, and key strengths and issues.

ACT-R (*Adaptive Components of Thought-Rational*) is a hybrid cognitive architecture and theoretical framework for emulating and understanding human cognition [11]. It aims at building a system that can performs the full range of human cognitive tasks and describe in detail the mechanisms underlying perception, thinking, and action. The central components of ACT-R comprise a set of perceptual-motor modules, memory modules, buffers, and a pattern matcher. The perceptual-motor modules basically serve as an interface between the system and the world. There are two types of memory modules in ACT-R: declarative memory (DM) and procedural memory (PM), which encode factual knowledge about the world and that about how to do things respectively. Both are realized as a symbolic-connectionist structures, where the symbolic level consists of productions (for PM) or chunks (for DM), and the sub-symbolic level of a massively parallel connectionist structure. Each symbolic construct (i.e., production or chunk) has a set of sub-symbolic parameters that reflect its past usage and control its operations, thus enabling an analytic characterization of connectionist computations using numeric parameters (associative activation) that measure the general usefulness of a chunk or production in the past and current context. Finally, the ACT-R buffers serve as a temporary storage for inter-module communications (excluding PM), while the pattern matcher is used to find a production in PM that matches the present state of the buffers.

ACT-R utilizes a top-down learning approach to adapt to the structure of the environment. In particular, symbolic constructs (i.e., chunks or productions) are first created to describe the results of a complex operation, so that the solution may be available without recomputing the next time a similar task occurs. When a goal, declarative memory activation or perceptual information appears it becomes a chunk in the memory buffer, and the production system guided by subsymbolic processes finds a single rule that responds to the current pattern. Sub-symbolic parameters are then tuned using Bayesian formulae to make the existing symbolic constructs that are useful more prominent. In this way chunks that are often used become more active and can thus be retrieved faster and more reliably. Similarly, productions that more likely led to a solution at a lower cost will have higher expected utility, and thus be more likely chosen during conflict resolution (i.e., selecting one production among many that qualify to fire). This architecture may be partially mapped on the brain structures. It has been applied in a large number of psychological studies, and in intelligent tutoring systems, but ambitious applications to problem solving and reasoning are still missing.

CLARION (The *Connectionist Learning Adaptive Rule Induction ON-line*) is a hybrid architecture that incorporates a distinction between explicit (symbolic) and implicit (sub-symbolic) processes and captures the interactions between the two [48]-[50]. The design objective is two-fold: to develop artificial agents for certain cognitive task domains, and to understand human learning and reasoning processes in similar domains. The CLARION architecture contains four memory modules, each comprising a dual explicit-implicit representation: action-centered subsystem (ACS), non-action-centered subsystem (NCS), motivational subsystem (MS), and metacognitive subsystem (MCS). Essentially, the ACS module serves to regulate the agent's actions, while NCS maintain the general system knowledge (either explicit or implicit). On the other hand, MS functions to provide a motivation/impetus for perception, action and cognition, while MCS monitor, direct and alter the operations of the other three modules. Each of these modules adopts a localist-distributed representation, where the localist section encodes the explicit knowledge and the distributed section (e.g. an MLP network) the implicit knowledge. CLARION also employs different learning methods for each level of knowledge. Learning of implicit knowledge is achieved using reinforcement learning methods such as Q-learning or supervised methods such as the standard back-propagation, both of which can be implemented using an MLP network [50]. The implicit knowledge already acquired at the bottom level is then utilized to craft the explicit knowledge at the top level via a bottom-up learning. This can in turn be viewed as a rational reconstruction of implicit knowledge at the explicit level. Top-down learning may also be achieved by precoding/fixing some rules at the top level and allowing the bottom-level to accumulate knowledge by observing actions guided by these rules [49]. As such, the system's decision making that relies initially on the top level gradually becomes more dependent on the bottom level. Software is available for experimentation with CLARION. A lot of psychological data has been simulated with this architecture, but also a complex sequential decision-making for a minefield navigation task.

LIDA (The *Learning Intelligent Distribution Agent*) is a conceptual and computational framework for intelligent, autonomous, "conscious" software agent that implements some ideas of the global workspace (GW) theory [51]. The architecture is built upon a bit older IDA framework, which was initially designed to automate the whole set of tasks of a human personnel agent who assigns sailors to new tours of duty. LIDA employs a partly symbolic and partly connectionist memory organization, with all symbols being grounded in the physical world in the sense of Brooks [44]. LIDA has distinct modules for perception, working memory, emotions, semantic memory, episodic memory, action selection, expectation and automatization (learning procedural tasks from experience), constraint satisfaction, deliberation, negotiation, problem solving, metacognition, and conscious-like behavior. Most operations are done by codelets implementing the unconscious processors (specialized networks) of the global workspace theory. A codelet is a small piece of code or program that performs one specialized, simple task. The LIDA framework incorporates three new modes of learning into the older IDA model: perceptual, episodic, and procedural learning, which are all of bottom-up type. Perceptual learning concerns learning of new objects, categories, relations, etc, and takes two forms: strengthening or weakening of the base-level activation of nodes, as well as creation of new nodes and links in the perceptual memory. Episodic learning, on the other hand, involves learning to memorize specific events (i.e., the what, where, and when). It results from events taken from the content of "consciousness" being encoded in the (transient) episodic memory. Finally, procedural learning concerns learning of new actions and action sequences with which to accomplish new tasks. It combines selectionist learning (i.e., selecting from an obsolete repertoire) and instructionalist learning (i.e., constructing new representations), with functional consciousness providing reinforcements to actions. There is no doubt that this architecture may explain many features of mind, however, it remains to be seen high competence ti will achieve in understanding language, vision, and common sense reasoning based on perceptions.

DUAL architecture [52] has been inspired by Minsky's "Society of Mind" theory of cognition [53]. It is a hybrid, multi-agent general-purpose architecture supporting dynamic emergent computation, with a unified description of mental representation, memory structures, and processing mechanisms carried out by small interacting micro-agents. As a result of lack of central control the system is constantly changing, depending on the environment. Agents interact forming larger complexes, coalitions and formations, some of which may be reified. Such models may be evaluated at different levels of granularity, the microlevel of micro-agents, the mesolevel of emergent and dynamic coalitions, and the macrolevel of the whole system and models, where psychological interpretations may be used to describe model properties. Micro-frames are used for symbolic representation of facts, while relevance or activation level of these facts in a particular context is represented by network connections with spreading activation that changes node accessibility. Links between microagents are based on their frame slots and weights control the influence of agents on each other's activity. DUAL architecture has been used in a number of projects: AMBR, a model of human reasoning that unifies analogy, deduction, and generalization, including a model of episodic memory; a model of human judgment; a model of perception, analysis of interactions between analogy, memory, and perception; understanding the role of context and priming effects for the dynamics of cognitive processes. This is certainly a very interesting architecture that is capable of explaining many cognitive phenomena. It is not clear how well it will scale up to real problems requiring complex reasoning, as nothing in this area has yet been demonstrated.

Polyscheme [54] integrates multiple methods of representation, reasoning and inference schemes in 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. Scripts, frames, logical propositions, neural networks and constraint graphs can be used to represent knowledge. A reflective specialist guides the attention of the whole system, providing various focus schemes that implement inferences via script matching, backtracking search, reason maintenance, stochastic simulation and counterfactual reasoning. High-order reasoning is guided by higher-level policies for focusing attention. Many problem solving algorithms use forward inference, subgoaling, grounding, representing alternate worlds and identity matching as their basic operations. Such operations are handled by specialists who are equipped with different representations but focus on the same aspect of the world, and may integrate also lower-level perceptual and motor processes. Thus Polyscheme may be used both in abstract reasoning and also in common sense physical reasoning in robots. It has been used to model infant reasoning including object identity, events, causality, spatial relations. This meta-learning approach combining different approaches to problem solving is certainly an important step towards AGI and common sense reasoning.

4CAPS architecture [55] has plausible neural implementation and is designed for complex tasks, such as language comprehension, problem solving or spatial reasoning. A unique feature is the ability to compare the activity of different 4CAPS modules with functional neuroimaging measures of brain's activity. It has been used to model human behavioral data (response times and error rates) for analogical problem solving, human–computer interaction, problem solving, discourse comprehension and other complex tasks solved by normal and mentally impaired people. Its first operating principle, "Thinking is the product of the concurrent activity of multiple centers that collaborate in a large scale cortical network", leads to the architecture based on a number of centers (corresponding to particular brain areas) that have different processing styles, for example Wernicke's area is specialized for the associative retrieval/design, constructing and selectively accessing structured sequential and hierarchical representations. Each center can perform and be a part of multiple cognitive functions, but has a limited computational capacity constraining its activity. Functions are assigned to centers depending on the resource availability, therefore the topology of the whole large-scale network is not fixed. Although 4CAPS contains many interesting ideas it is not aimed at achieving intelligent behavior, but rather tries to model human performance; software written in Lisp is available for experimentation. See the discussion in [55] of other models that are aimed at explanation of behavioral data.

Shruti [56], biologically-inspired model of human reflexive inference, represents in connectionist architecture relations, types, entities and causal rules using focal-clusters. These clusters encode universal/existential quantification, degree of belief, and the query status. The synchronous firing of nodes represents dynamic binding, allowing for representations of quite complex knowledge and inferences. This architecture may have great potential, but after rather long time of development it has not yet found any serious applications to problem solving or language understanding.

The Novamente AI Engine is based on system-theoretic ideas regarding complex mental dynamics and associated emergent patterns, inspired by the *psynet* model [57] and more general "patternist philosophy of mind" [58]. Similarly as in the "society of minds" and the global workspace, self-organizing and goal-oriented interactions between patterns are responsible for mental states. Emergent properties of network activations should lead to hierarchical and relational (heterarchical) pattern organization. Probabilistic term logic (PTL), and the Bayesian Optimization Algorithm (BOA) algorithms are used for flexible inference. Actions, perceptions, and internal states are represented by tree-like structures. This is still an experimental architecture that is being developed, seems to be in a fluid state, and its scaling properties are not yet known.

4. Where do we go from here?

The previous sections has presented a number of very interesting models of cognition that have the potential to develop general intelligence. Many excellent projects have already been formulated, some have been developed over many decades, while others are just starting. So far cognitive architectures are used in very few real-world applications. Grand challenges, as discussed in section two, and smaller steps that lead to human and super-human levels of competence should be formulated to focus the research. Extending small demonstrations in which a cognitive system reasons in a trivial domain to larger-scale applications, for example generating results that may be of interest to experts, or acting as an assistant to human expert, is one important direction. Without a set of demanding test problems it is very hard to evaluate new projects, compare their capabilities and understand their limitations. Integrative models of human performance are of great interest in the defense and aerospace industries. A recent project on the Agent-Based Modeling and Behavior Representation (AMBR) Model Comparison resulted in quantitative data comparing the performance of humans and cognitive architectures in a simplified air traffic controller environment [59]. Some efforts have been expended on the evaluation of software agents and several proposals in this direction has been put forth during the 2007 AAAI Workshop "Evaluating Architectures for Intelligence" [60]. Ideas ranged from using in-city driving environment as a testbed for evaluating cognitive architectures, to measuring incrementality and adaptivity components of general intelligent behavior.

Perhaps a measure of "cognitive age" could be established, with a set of problems that children at a given age are able to solve. Problems should be divided into several groups: e.g. vision and auditory perception, understanding language, common-sense reasoning, abstract reasoning, probing general knowledge about the world, learning, problem solving, imagination, creativity. Solving all problems from a given group that children at some age are able to solve will qualify cognitive system to pass to the next grade in this group of problems. It should be expected that some systems will show advanced age in selected areas, and not in the others. For example, solving problems requiring vision may require addition of specialized computer vision modules, while mathematical reasoning in many reasoning systems may be fairly advanced comparing to children. Experts in human intelligence largely agree to the original Gardner's proposal [61] that seven kinds of intelligence should be distinguished: logical-mathematical, linguistic, spatial, musical, bodily-kinesthetic, interpersonal and intrapersonal intelligence, perhaps extended by emotional intelligence and a few others.

General world knowledge is fairly difficult to collect and could be probed using a question/answer system. If a 5-year old child could get all the answer to general questions from an avatar controlled by some cognitive architecture one should assume that the mental age of the control system in this respect is at least 5. Knowledge bases in cognitive systems are usually quite limited and require very different kind of organization and knowledge representation methods. Huge CyC knowledge base is an exception [3], and using it to construct large knowledge bases suitable for other cognitive systems is certainly worth the effort.

Such analysis should certainly help to understand what type of intelligence may be expected from embodied cognitive robotic projects and what the limitations of symbolic approaches are. Brooks has made a good point that elephants do not play chess [43], and expressed hope [44] that a robot with integrated vision, hearing and dextrous manipulation controlled by large scale parallel MIMD computer "will learn to 'think' by building on its bodily experiences to accomplish progressively more abstract tasks". His Cog project based on grounding the meaning of concepts in deep embodiment has many followers although after 15 years it has stayed at the level of reactive agent and there are no good ideas how to extend it to higher cognitive levels. While behavioral intelligence in robotic devices may be difficult to achieve without embodiment experiences with this approach in the last two decades are not very encouraging for AGI. Elephants are intelligent, but cannot learn language or be personal assistants. It is also possible that ideas on which cognitive architectures are based are not sufficient to solve the problems in computer vision or language and more specific models of some brain functions are needed.

The survey presented above showed several trends that will probably dominate in the research on cognitive architectures. First, the number of hybrid architectures is already quite large, biological inspirations are becoming increasingly important and this will lead to domination of BICA architectures. Even hard core symbolic architecture proponents base now further extension of their architectures on inspirations from the brain [24]. They focus on the role of cortex and limbic system, but completely neglect the regulatory role of the brain stem which may provide overall meta-control selecting different types of behavior. Second, there may be many BICA architectures, but several key features need to be preserved. Different types of memory are certainly important, as has been already stressed by several symbolic, emergent and hybrid architectures. Processing of speech or texts requires recognition of tokens, or mapping from sounds or strings of letters to unique terms; resolving ambiguities and mapping terms to concepts in some ontology; and a full semantic representation of the text, that facilitates understanding and answering questions about its content. These three steps are roughly based on several kinds of human memory.

First, recognition memory that helps to focus quickly attention when something is wrong, for example a strangely spelled word that could be a misspelling, a foreign word, personal name, or an attempt to avoid spam filters. This may be implemented by simple neural networks without hidden layer or by correlation matrix memories [35]. The role of recognition memory has also been largely forgotten.

Second, there is a need for semantic memory that serves not only as hierarchical ontology, but approximates spreading activation processes in real brains, and thus activates various types of associations providing background knowledge that humans use for token to concept mapping and disambiguation. Unfortunately large-scale semantic memories that contain both structural properties of concepts (chairs have legs, seat, etc) and their relations and associations (chair – table, sit, etc) and could be used in computationally efficient way do not exist. While significant progress has been made in drawing inspirations from neuroscience in analysis of auditory, visual and olfactory signals much less has been done at the higher cognitive function level. Although neurocognitive approach to linguistics has been formulated as "an attempt to understand the linguistic system of the human brain, the system that makes it possible for us to speak and write, to understand speech and writing, to think using language …" [62], in practice it has been used only to analyze specific linguistic phenomena. A practical algorithm to discover these "pathways of the brain" has been introduced recently [63], opening the way for construction of large-scale semantic networks that will approximate symbolic knowledge stored in human brain, although creating such large-scale memories will require a large effort. Efforts to build concept descriptions from electronic dictionaries, ontologies, encyclopedias, results of collaborative projects and active searches in unstructured sources have been described in [7].

Third, episodic memory is required to store experiences from interactions with individual users, to understand the context of current interactions and interpret all events in view of this context. Various elements of semantic and episodic memories are kept in the working memory. All types of memory are intimately connected. Recognition of tokens is facilitated by the active part of semantic memory and made easier by the expectations resulting from episodic memory. Reading text leads to priming effects: expectation and anticipation of a few selected words, and inhibition of many others that do not come to the mind of the reader. Episodic memory is based on semantic relations of the concepts found in the text. Although several proposals for memory-based cognitive architectures have been formulated the role of different types of memory has not been stressed and no effort to create appropriate large-scale knowledge bases has been made. AGI requires such memories, and as a step towards such memory-based architecture an avatar that uses large semantic memory to play word games has been demonstrated [7].

What is the best practical way to implement these ideas? Template matching proved to be amazingly effective in simulation of dialogues and is still dominating in chatterbots [64], but it obviously does not lead to real understanding of the concepts that appear in the discussion. Neural template matching, or templates approximating the distribution of neuronal group activities in the brain during concept comprehension, is the simplest technique that goes beyond symbolic template matching, leading to sets of consistent concepts. Words are ambiguous and form concepts that have meanings modified by their contexts. In the brain a word $w = (w_f, w_s)$ has phonological component w_f (the memorized form of the word, string of phonemes or characters), and an extended semantic representation w_s (extended activations related to the use and category of the word, including immediate associations). The extended activation is not unique, only when the current context *Cont* is specified (specific activations of other concepts are determined) the meaning of the word is established, resulting from spreading activation in the brain to form a global state $\Psi(w,Cont)$. This state changes with each new word received in sequence, with quasi-stationary states formed after each sentence is processed and understood. It is quite difficult to decompose the $\Psi(w,Cont)$ state into components, because the semantic representation w_s is strongly modified by the context. The state $\Psi(w,Cont)$ may be regarded as a quasi-stationary wave, with its core component centered on the phonological/visual brain activations w_f and with quite variable extended representation w_s. As a result the same word in a different sentence creates quite different states of activation, and the lexicographical meaning of the word may be only an approximation of an almost continuous process. To relate states $\Psi(w,Cont)$ to lexicographical meanings, one can cluster all such states for a given word in different contexts and define prototypes $\Psi(w_k,Cont)$ for different meanings w_k. These prototypes are neural templates that should replace symbolic templates. The form of the word w_f identifies several candidate templates with different meanings, and the one that fits to other templates, maximizing overall consistency of interpretations, is selected.

The symbolic approach to language is a poor substitute for neurolinguistic processes, and high-dimensional vector model of language, popular in statistical approach to natural language processing (NLP) [65], is a very crude approximation that does not reflect essential properties of the perception-action-naming activity of the brain [66][67]. The process of understanding words (spoken or read) starts from activation of word form representation (the phonological or grapheme representation) in the temporal lobe, quickly spreading neural activation to further brain areas, including the non-dominant (usually right) hemisphere, that does not contain representations of word forms, but learns to evaluate clusters of activations, forming constraints on the way words may be used, and forming general, higher-level concepts [63]. This continuous process may be approximated through a series of snapshots of patterns created by microcircuit activations $\phi_i(w,Cont)$ that can be treated as basis functions for the expansion of the state $\Psi(w,Cont) = \Sigma_i\, \alpha_i\, \phi_i(w,Cont)$, where the summation extends over all patterns that show significant activity resulting after presentation of the word w. The high-dimensional vector model used in NLP measures only the co-occurrence of words $\mathbf{V}_{ij} = \langle \mathbf{V}(w_i),\mathbf{V}(w_j)\rangle$ in small window neighborhood, averaged over all contexts, while human knowledge includes also structural properties of concepts that are important, but do not appear explicitly in texts. The use of wave-like representation in terms of basis functions to describe neural states makes this formalism similar to that used in quantum mechanics, although no real quantum effects are implied here. Objects of discourse and actual episodes are memorized by hippocampus that links to the cortex and is able to recreate the original activations at the moment the episode has been experienced.

An outline of the road from single neurons, to brain modules, to societies of brains has been presented in [69]. Single neurons have little internal knowledge and very simple interactions via weighted links; assemblies of neurons at different levels form coalitions and may be regarded as specialized processors, passing structured information and attaining rather complex internal states. This may be approximated by interacting agents, with internal knowledge and means of communication, with coalitions of simple agents creating dynamic, higher-order units that may be viewed as "soft agents", with new competencies arising at demand by variation on the main theme. Such meta-learning ideas for solving pattern recognition and reasoning based on partial observations has been recently described [46], and preliminary implementation of general system architecture to support such approach has been presented [70].

The role of imagination, creativity, learning from partial observations and using this knowledge in an intuitive way, and the role of the right hemisphere in the linguistic processes, has only recently been discussed [24][35][36]. The AIM1 (ArtIficial Mind1) architecture based on these ideas is under development and will be presented in near future. This architecture will draw inspirations from some of the projects presented in this paper, but will be primarily aimed at ambitious applications requiring natural language processing. It is quite likely that this approach will lead to creation of conscious artifacts [40][71].

Acknowledgment: W.D. thanks Polish Committee for Scientific Research for research grant 2005-2007.

References

[1] A. Newell, *Unified Theories of Cognition*: Harvard University Press, 1990.
[2] A. Newell, H.A. Simon, GPS: A program that simulates human thought. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and Thought*. New York: McGraw-Hill, 1963.
[3] K. Panton, C. Matuszek, D. Lenat, D. Schneider, M. Witbrock, N. Siegel, B. Shepard, Common Sense Reasoning – From Cyc to Intelligent Assistant. In: Y. Cai and J. Abascal (Eds.): Ambient Intelligence in Everyday Life, LNAI **3864** (2006) 1–31.
[4] A. Turing, Computing Machinery and Intelligence, *Mind* **49** (1950) 433-460.
[5] See http://www.loebner.net for information on the Loebner competition.
[6] R. Carpenter, J. Freeman, Computing Machinery and the Individual: the Personal Turing Test, 2005, paper available at http://www.jabberwacky.com/.
[7] J. Szymański, T. Sarnatowicz, W. Duch, Towards Avatars with Artificial Minds: Role of Semantic Memory. *Journal of Ubiquitous Computing and Intelligence* **1** (2007, in print),
[8] E.A. Feigenbaum, Some Challenges and Grand Challenges for Computational Intelligence. *J. of the ACM* **50**(1) (2003) 32–40.
[9] N. Nilsson, Human-Level Artificial Intelligence? Be Serious! *The AI Magazine* 26(4) (2005) 68-75.
[10] D.E. Meyer, D.E. Kieras, A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review,* **104**(1), (1997) 3-65.
[11] J.R. Anderson, C. Lebiere, The Newell test for a theory of cognition. *Behavioral and Brain Science* **26** (2003) 587-637.
[12] D. Vernon, G. Metta, G. Sandini, A survey of artificial cognitive systems: Implications for the autonomous development of mental capabilities in computational agents. *IEEE Transactions on Evolutionary Computation* **11**(2) (2007) 151-180.
[13] J. Hawkins, S. Blakeslee, On intelligence: How a New Understanding of the Brain will Lead to the Creation of Truly Intelligent Machines. Times Books 2004.
[14] T. Hoya, *Artificial Mind System. Kernel Memory Approach*. Springer, 2005.
[15] R. Hecht-Nielsen, *Confabulation Theory: The Mechanism of Thought*. Springer 2007.
[16] J.F. Sowa, *Conceptual Structures*. Reading, Mass, Addison-Wesley, 1984.
[17] M. Minsky, A Framework for Representing Knowledge. In: P.H. Winston, Ed. *The Psychology of Computer Vision*. New York: McGraw-Hill, 1975.
[18] R.J. Firby, *Adaptive Execution in Complex Dynamic Worlds.* Ph.D. Thesis, Yale University, 1989.
[19] T.M. Mitchell, R., Keller, S. Kedar-Cabelli, Explanation-based generalization: A unifying view. *Machine Learning* **1** (1986) 47-80.
[20] M.M.Veloso, J.G. Carbonell, Integrating analogy into a general problem-solving architecture. In M. Zemankova & Z. Ras (Eds.), *Intelligent Systems* (pp. 29-51). Chichester, England: Ellis Horwood, 1990.
[21] N., Larvac, S. Dzeroski, *Inductive Logic Programming: Techniques and Applications*. New York: Ellis Horwood, 1994.
[22] J. Shavlik, T. Dietterich, *Readings in machine learning*. San Mateo, CA: Morgan Kaufmann, 1990.
[23] J.E. Laird, P.S. Rosenbloom, A. Newell, Soar: An architecture for general intelligence. *Artificial Intelligence* **33** (1987) 1-64.
[24] John Laird, Current Soar Research, see http://ai.eecs.umich.edu/people/laird/current-research.html.
[25] Soar community: http://sitemaker.umich.edu/soar/home
[26] J. Rosbe, R.S. Chong, D.E. Kieras, Modeling with Perceptual and Memory Constraints: An EPIC-Soar Model of a Simplified Enroute Air Traffic Control Task, SOAR Technology Inc. Report, Ann Arbor, Michigan, 2001.
[27] P. Langley, An adaptive architecture for physical agents. In *Proc. of the 2005 IEEE/WIC/ACM Int. Conf. on Intelligent Agent Technology*. Compiegne, France: IEEE Computer Society Press, pp. 18-25, 2005.
[28] P. Langley, D. Choi, Learning recursive control programs from problem solving. *J. of Machine Learning Res.* **7** (2006) 493-518.
[29] P. Wang, Rigid flexibility. The Logic of Intelligence. Springer 2006
[30] S.C. Shapiro, W.J. Rapaport, M. Kandefer, F.L. Johnson, A. Goldfain, Metacognition in SNePS, *AI Magazine* **28** (2007) 17-31.
[31] J.L. McClelland, D.E. Rumelhart and the PDP Research Group. *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*. Cambridge, MA: MIT Press, 1986.
[32] R.C. O'Reilly, Y. Munakata, *Computational Explorations in Cognitive Neuroscience: Understanding of the Mind by Simulating the Brain*. Cambridge, MA: MIT Press, 2000.
[33] W. Duch, R. Adamczak, K. Grąbczewski, A new methodology of extraction, optimization and application of crisp and fuzzy logical rules. *IEEE Transactions on Neural Networks,* **12** (2001) 277-306.
[34] R.C. O'Reilly, T.S. Braver, J.D. Cohen A biologically-based computational model of working memory. In A. Miyake & P. Shah (Eds.), *Models of Working Memory*. Cambridge University Press, pp. 375-411, 1999.

[35] W. Duch, M. Pilichowski, Experiments with computational creativity. *Neural Information Processing – Letters and Reviews* **11** (2007) 123-133.
[36] W. Duch, Intuition, Insight, Imagination and Creativity. *IEEE Computational Intelligence Magazine* **2**(3) (2007) 40-52.
[37] G.M. Edelman, Neural Darwinism: Selection and reentrant signaling in higher brain function. *Neuron* **10** (1993) 115-125.
[38] B.J. Baars, *A Cognitive Theory of Consciousness*. New York: Cambridge University Press, 1988.
[39] M.P. Shanahan, A cognitive architecture that combines internal simulation with a global workspace. Consciousness and Cognition 15 (2006) 157-176.
[40] P. Haikonen, *Robot brains; circuits and systems for conscious machines*. Wiley, 2007.
[41] J.A. Anderson, P. Allopenna, G.S. Guralnik, D. Sheinberg, J.A. Santini, Jr., D. Dimitriadis, B.B. Machta, and B.T. Merrit, Programming a Parallel Computer: The Ersatz Brain Project. In W Duch, J Mandzuik (Eds.), Challenges to Computational Intelligence. Springer: Berlin, pp. 61-88, 2007.
[42] J. Weng and W. S. Hwang, From Neural Networks to the Brain: Autonomous Mental Development. *IEEE Computational Intelligence Magazine* **1**(3) (2006) 15-31.
[43] R. Brooks, Elephants don't play chess. *Robotics and Autonomous Systems* **6** (1986) 3-15.
[44] R. Brooks, L.A. Stein, Building Brains for Bodies. *Autonomous Robotics* **1** (1994) 7-25.
[45] E. Korner, G. Matsumoto, Cortical architecture and self-referential control for brain-like computation. *IEEE Engineering in Medicine and Biology Magazine,* **21**(5) (2002) 121-133.
[46] W. Duch, Towards comprehensive foundations of computational intelligence. In: W. Duch and J. Mandziuk, *Challenges for Computational Intelligence*. Springer Studies in Computational Intelligence, **63** (2007) 261-316.
[47] TOSCA: A comprehensive brain-based cognitive architecture: *Biologically-Inspired Cognitive Architecture* (BICA) Phase 1 Architecture Report, DARPA-IPTO 2006.
[48] R. Sun, F. Alexandre, *Connectionist symbolic integration*. Hillsdale, NJ: Erlbaum, 1997.
[49] R. Sun, X. Zhang, Top-down versus bottom-up learning in cognitive skill acquisition. *Cognitive Systems Research* **5** (2004) 63-89.
[50] R. Sun, E. Merrill, T. Peterson, From implicit skills to explicit knowledge: A bottom-up model of skill learning. *Cognitive Science,* **25**(2) (2001) 203-244.
[51] S. Franklin, The LIDA architecture: Adding new modes of learning to an intelligent, autonomous, software agent. In *Proc. of the Int. Conf. on Integrated Design and Process Technology*. San Diego, CA: Society for Design and Process Science, 2006.
[52] A. Nestor, B. Kokinov, Towards Active Vision in the DUAL Cognitive Architecture. *International Journal on Information Theories and Applications* **11** (2004) 9-15.
[53] M. Minsky, *The Society of Mind*. Simon and Schuster, New York, 1986.
[54] N.L. Cassimatis, Adaptive Algorithmic Hybrids for Human-Level Artificial Intelligence. Advances in Artificial General Intelligence. IOS Press. Eds. B. Goertzel and P. Wang, 2007.
[55] M.A. Just, S. Varma, The organization of thinking: What functional brain imaging reveals about the neuroarchitecture of complex cognition. Cognitive, Affective, and Behavioral Neuroscience **7** (2007) 153-191.
[56] L. Shastri, V. Ajjanagadde, From simple associations to systematic reasoning: A connectionist encoding of rules, variables, and dynamic bindings using temporal synchrony. Behavioral & Brain Sciences, 16(3) (1993) 417-494.
[57] B. Goertzel, From Complexity to Creativity. New York, NY: Plenum Press, 1997.
[58] B. Goertzel, The Hidden Pattern, BrownWalker Press, 2006.
[59] K.A. Gluck, R.W. Pew (Eds.), *Modeling Human Behavior with Integrated Cognitive Architectures: Comparison, Evaluation*. Lawrence Erlbaum Associates 2005.
[60] G.A. Kaminka, C.R. Burghart (eds.), Evaluating Architectures for Intelligence. Technical Report WS-07-04, AAAI Press, Menlo Park, CA, 2007.
[61] H. Gardner, *Multiple intelligences: The theory in practice*. New York: Basic Books, 1993.
[62] S. Lamb, *Pathways of the Brain: The Neurocognitive Basis of Language*. Amsterdam: J. Benjamins Publishing Co. 1999.
[63] W. Duch, P. Matykiewicz, J. Pestian, Towards Understanding of Natural Language: Neurocognitive Inspirations. Lecture Notes in Computer Science **4669** (2007) 953–962.
[64] R. Wallace, *The Elements of AIML Style*, ALICE A.I. Foundation, 2003.
[65] C.D. Manning, H. Schütze, *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press, 1999.
[66] F. Pulvermuller, The Neuroscience of Language. On Brain Circuits of Words and Serial Order. Cambridge, UK: Cambridge University Press, 2003.
[67] S. Dehaene, L. Cohen, M. Sigman, F. Vinckier, The neural code for written words: a proposal. *Trends in Cognitive Science* 9, 335-341, 2005.
[68] M. Meeter, J.M.J. Murre, TraceLink: A model of consolidation and amnesia. *Cognitive Neuropsychology* **22** (5), 559-587, 2005.
[69] W. Duch, J. Mandziuk, Quo Vadis Computational Intelligence? In: *Machine Intelligence. Quo Vadis*? Eds: P. Sinčák, J. Vaščák, K. Hirota. Advances in Fuzzy Systems - Applications and Theory - Vol. 21, World Scientific, 3-28, 2004.
[70] K. Grąbczewski, N. Jankowski, Versatile and Efficient Meta-Learning Architecture: Knowledge Representation and Management in Computational Intelligence. *IEEE Symposium Series on Computational Intelligence* (SSCI 2007), Honolulu, HI, IEEE Press, pp. 51-58.
[71] W. Duch, Brain-inspired conscious computing architecture. *Journal of Mind and Behavior* 26(1-2) (2005) 1-22.