# The Role of Cognitive Architectures in General Artificial Intelligence



*Published version:*

DOI:10.1016/j.cogsys.2017.08.003



(Article begins on next page)

# The Role of Cognitive Architectures in General Artificial Intelligence

Antonio Lieto[a,*], Mehul Bhatt[b], Alessandro Oltramari[c], David Vernon[d]

[a] *University of Turin, Dept. of Computer Science and ICAR-CNR, Italy*
[b] *Örebro University, Sweden., and University of Bremen, Germany*
[c] *Bosch Research and Technology Center, 2555 Smallman Street, Pittsburgh, USA*
[d] *Carnegie Mellon University Africa, Rwanda*

**Abstract**

The term "Cognitive Architectures" indicates both abstract models of cognition, in natural and artificial agents, and the software instantiations of such models which are then employed in the field of Artificial Intelligence (AI). The main role of Cognitive Architectures in AI is that one of enabling the realization of artificial systems able to exhibit intelligent behavior in a general setting through a detailed analogy with the constitutive and developmental functioning and mechanisms underlying human cognition. We provide a brief overview of the status quo and the potential role that Cognitive Architectures may serve in the fields of Computational Cognitive Science and Artificial Intelligence (AI) research.

*Keywords:* Cognitive Architectures, Artificial Intelligence, Autonomous Systems, General Artificial Intelligence

## 1. Cognitive Architectures: Design Perspectives and Open Challenges

The design and development of Cognitive Architectures (CAs) is a wide and active area of research in Cognitive Science, Artificial Intelligence and, more recently, in the areas of Computational Neuroscience, Cognitive Robotics, and

*Corresponding author
*Email addresses:* lieto@di.unito.it (Antonio Lieto), bhatt@uni-bremen.de (Mehul Bhatt), oltramale@gmail.com (Alessandro Oltramari), vernon@cmu.edu (David Vernon)

Computational Cognitive Systems. Cognitive architectures[1] have been historically introduced for three main reasons: i) to capture, at the computational level, the invariant mechanisms of human cognition, including those underlying the functions of reasoning, control, learning, memory, adaptivity, perception and action [2] (this goal is crucial in the *cognitivist* perspective [3]), ii) to form the basis for the development of cognitive capabilities through ontogeny over extended periods of time (this goal is one of the main target of the so called *emergent* perspective), and iii) to reach human level intelligence, also called General Artificial Intelligence, by means of the realization of artificial artifacts built upon them (on the role of CAs for general intelligent systems see also [4]).

During the last few decades many different cognitive architectures — such as SOAR [5], ACT-R [6], CLARION [7, 8], the iCub [9] etc. — have been realized and agents based on such infrastructures have been widely tested in several cognitive tasks involving reasoning, learning, perception, action execution, selective attention, recognition etc. (for comprehensive reviews on the theme we refer to [10], [11], and [12]).

The design of these different CAs has obviously followed diverse approaches based on the specificity of the scientific objective pursued through these artifacts. In particular: the cognitive architectures aimed at building and implementing model of cognition and that, as such, focus on aspects concerning generality, completeness and on the attempt to identify a standard model of mind [13], are designed according to the so called "cognition in the loop" approach. In such a perspective, inspired by the cybernetics tradition and by the synthetic method [14], the computational simulation of biological and cognitive processes is assumed to play a central epistemological role in the development and refinement of theories about the elements characterizing the nature of intelligent behaviour. In particular, such an approach has a twofold goal: i) it aims at detecting novel and hidden aspects of the cognitive theories by building

[1] The term cognitive architecture was introduced by Allen Newell and his colleagues in their work on unified theories of cognition [1].

properly designed computational models of cognition and ii) it aims at providing technological advancement in the area of Artificial Intelligence (AI) of cognitive inspiration[2]. Within this framework, the debate between purely *functionalist* models [15], based on a weak equivalence (i.e. the equivalence in terms of functional organization) between cognitive processes and AI procedures, and "structural" models of our cognition (based on a more constrained equivalence between AI procedures and their corresponding cognitive processes) has seen the latter prevailing for both theoretical and practical reasons[16]. Despite some intrinsic differences, both the *cognitivist* and the *emergent* perspectives, follow a design perspective that is compliant with the structural approach and is, usually, driven by general desiderata [17, 18, 19].

A different design stance, on the other hand, is taken by agent architectures addressing the needs of an application without being concerned whether or not it is a faithful model of cognition. Such systems are effectively conventional system architecture, rather than a cognitive architecture *per se*, but they exhibit the required attributes and functionality that we usually recognize as "cognitive abilities". For example: typically the ability to autonomously perceive, to anticipate the need for actions and the outcome of those actions, and to act, learn, and adapt. In this case, the design principles of the system architecture is driven by user requirements, drawing on the available repertoire of AI and cognitive systems algorithms and data-structures.

As pointed out by [17], the two research agendas pursued by such different stances are not necessarily complementary since they are beneficial for different important purposes: advancements in science and in engineering. In our opinion, it is important to keep alive both these different souls of research in the area of cognitive architectures to see how/if/to what extent the elements characterizing the success in one of these approach can be plausibly adapted or reused in the other one. A possible common ground for evaluating the provided

[2]This implies building systems able to solve/deal with a particular problem in a better way with respect to other artificial systems due to the adopted cognitively-inspired design.

advancements of the different frameworks (or the encountered problems that prevent to obtain advancements) is that one of focusing on classes of problems that are easily solvable for humans but very hard to solve for machines. For instance, these could involve aspects concerning common sense reasoning about space, action, change and language categorization [20, 21, 22]; selective attention; integration of multi-modal perception; learning from few examples; robust integration of mechanisms involving planning, acting, monitoring and goal reasoning. Such complementarities could also be explored in specialised cognitive problem-solving contexts [23], e.g., involving computational visuo-spatial cognition in particular domains such as design cognition [24].

These aspects, and in particular those arising from the general integration of all such distinct cognitive functions, raise research challenges that go beyond the study of each single component (in fact they require an architectural level of abstraction). Solving such challenges is crucial for building systems that can take the form of general virtual humans for companionship, versatile service or personal robots, software agents for interactive tutoring or personal assistant. Such systems need to be robust and resilient, and they have to meet the quality of service constraints. The AI and the Cognitive Systems research communities are nowadays posing an increasing level of attention on these problem. It is worth noting, for example, the recent AAAI 2017 special track on Integrated Systems (http://www.aaai.org/Conferences/AAAI/2017/aaai17integrated.php and the EUCognition 2016 Conference on Robot Architectures [25]. Other relevant venues explicitly addressing such issues are the Advances inf Cognitive Systems Conference series (http://www.cogsys.org/) as well as the AIC workshop series on Artificial Intelligence and Cognition, that played, in this perspective, a recognized role of promotion and development of such themes at the cross border of the AI, Cognitive Modelling and Cognitive Robotics[3] communities [26]. We believe that the road traced is the

[3]In the area of Cognitive Robotics a particularly active role of promotion of such research themes is played by the IEEE Technical Committee on Cognitive Robotics http:

way to follow in order to make progresses towards the realization of human-level intelligent systems in general setting. Despite the continuous warnings coming from the popular press, in fact, this goal is still far from being achieved.

In the following we provide a quick tour of the work appearing in the Special Issue: the article "Evolution of the Icarus Cognitive Architecture" by Choi and Langley [27] presents an overview of the development of one of the most known cognitive architectures of the cognitivist tradition: ICARUS. The authors present the main elements of the architecture and focus their attention of the evolution of such system over the last three decades by discussing the representational and processing assumptions made by different versions of the architecture, their relation to alternative theories, and some promising directions for future research.

The article "The Knowledge Level in Cognitive Architectures: Current Limitations and Possible Developments" by Lieto, Lebiere and Oltramari [28] proposes a critical overview of the current state of the art of the knowledge level of cognitive architectures pointing out two constitutive problems: the limited size and the homogeneous typology of knowledge that is encoded and processed by such systems. In order to address the current limitations the authors propose three research directions that can be explored.

The work "Modeling valuation and core affect in a cognitive architecture: The impact of valence and arousal on memory and decision-making" by Juvina, Larue and Hough [29] presents a novel approach to adding primitive evaluative capabilities to a cognitive architecture that impacts on the affective valence and arousal on memory and decision-making.

Finally, the article "An architecture for ethical robots inspired by the simulation theory of cognition" by Vanderelst and Winfield [30] presents an original attempt to embed, in a physical robotic architecture, an ethic layer inspired by the simulation theory of cognition able to work without recurring to standard logic-based approaches that are usually employed to perform meta-cognitive

`//www.ieee-coro.org/`

computation.

## References


[1] A. Newell, Unified theories of cognition, Harvard University Press, 1994.

[2] A. Oltramari, C. Lebiere, Pursuing artificial general intelligence by leveraging the knowledge capabilities of act-r, in: Artificial General Intelligence, Springer, 2012, pp. 199–208.

[3] D. Vernon, Artificial cognitive systems: A primer, MIT Press, 2014.

[4] P. Langley, Cognitive architectures and general intelligent systems, AI magazine 27 (2) (2006) 33.

[5] J. Laird, The Soar cognitive architecture, MIT Press, 2012.

[6] J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, Y. Qin, An integrated theory of the mind., Psychological review 111 (4) (2004) 1036.

[7] R. Sun, The CLARION cognitive architecture: Extending cognitive modeling to social simulation, Cognition and multi-agent interaction (2006) 79–99.

[8] R. Sun, Anatomy of the Mind: Exploring Psychological Mechanisms and Processes with The: Exploring Psychological Mechanisms and Processes with the Clarion Cognitive Architecture, Oxford University Press USA, 2016.

[9] D. Vernon, G. Metta, G. Sandini, The icub cognitive architecture: Interactive development in a humanoid robot, in: Development and Learning, 2007. ICDL 2007. IEEE 6th International Conference on, Ieee, 2007, pp. 122–127.

[10] 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.

[11] P. Langley, J. E. Laird, S. Rogers, Cognitive architectures: Research issues and challenges, Cognitive Systems Research 10 (2) (2009) 141–160.

[12] K. Thórisson, H. Helgasson, Cognitive architectures and autonomy: A comparative review, Journal of Artificial General Intelligence 3 (2) (2012) 1–30.

[13] J. E. Laird, C. Lebiere, P. S. Rosenbloom, A standard model of the mind: Toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics, AI Magazine (in press) 1–19.

[14] R. Cordeschi, The discovery of the artificial: Behavior, mind and machines before and beyond cybernetics, Vol. 28, Springer Science & Business Media, 2002.

[15] H. Putnam, Minds and machines, MacMillan, 1960.

[16] A. Lieto, D. P. Radicioni, From human to artificial cognition and back: New perspectives on cognitively inspired ai systems, Cognitive Systems Research 39 (2016) 1–3.

[17] D. Vernon, Two ways (not) to design a cognitive architecture, Cognitive Robot Architectures (2017) 42.

[18] D. Vernon, C. von Hofsten, L. Fadiga, Desiderata for developmental cognitive architectures, Biologically Inspired Cognitive Architectures 18 (2016) 116–127.

[19] R. Sun, Desiderata for cognitive architectures, Philosophical Psychology 17 (3) (2004) 341–373.

[20] M. Bhatt, Reasoning about space, actions and change: A paradigm for applications of spatial reasoning, in: Qualitative Spatial Representation and Reasoning: Trends and Future Directions, IGI Global, USA, 2012.

[21] M. Bhatt, C. Schultz, C. Freksa, The 'Space' in Spatial Assistance Systems: Conception, Formalisation, and Computation. Book: Representing Space in Cognition: Behaviour, Language, and Formal Models., Series: Explorations in Language and Space, Oxford University Press, 2013, pp. 171–214.

[22] A. Lieto, D. P. Radicioni, V. Rho, Dual peccs: a cognitive system for conceptual representation and categorization, Journal of Experimental & Theoretical Artificial Intelligence 29 (2) (2017) 433–452.

[23] M. Bhatt, Between sense and sensibility: Declarative narrativisation of mental models as a basis and benchmark for visuo-spatial cognition and computation focussed collaborative cognitive systems, CoRR abs/1307.3040.

[24] M. Bhatt, C. Schultz, People-Centered Visuospatial Cognition: Next-generation Architectural Design Systems and their Role in Conception, Computing, and Communication., In (edited volume): The Active Image: Architecture and Engineering in the Age of Modeling, Series: Philosophy of Engineering and Technology, Springer International Publishing, 2017.

[25] R. Chrisley, V. Muller, Y. Sandamirskaya, M. Vincze, Proceedings of the eucognition meeting (european society for cognitive systems), in: EUCognition 2016 Cognitive Robot Architectures, Vol. 1855, 2016.

[26] A. Lieto, M. Bhatt, A. Oltramari, D. Vernon, Proceedings of aic 2016, 4th international workshop on artificial intelligence and cognition.

[27] D. Choi, P. Langley, Evolution of the icarus cognitive architecture, Cognitive Systems Research.

[28] A. Lieto, C. Lebiere, A. Oltramari, The knowledge level in cognitive architectures: Current limitations and possible developments, Cognitive Systems Research.

[29] I. Juvina, O. Larue, A. Hough, Modeling valuation and core affect in a cognitive architecture: The impact of valence and arousal on memory and decision-making, Cognitive Systems Research.

[30] D. Vanderelst, A. Winfield, An architecture for ethical robots inspired by the simulation theory of cognition, Cognitive Systems Research.