Two Ways (Not) To Design a Cognitive Architecture

David Vernon[†]
Carnegie Mellon University Africa
Rwanda
Email: vernon@cmu.edu

***Abstract*—In this short paper, we argue that there are two conflicting agendas at play in the design of cognitive architectures. One is principled: to create a model of cognition and gain an understanding of cognitive processes. The other is practical: to build useful systems that have a cognitive ability and thereby provide robust adaptive behaviour that can anticipate events and the need for action. The first is concerned with advancing science, the second is concerned with effective engineering. The main point we wish to make is that these two agendas are not necessarily complementary in the sense that success with one agenda may not necessarily lead, in the short term at least, to useful insights that lead to success with the other agenda.**

I. Introduction

There are two aspects to the goal of building a cognitive robot [1]. One is to gain a better understanding of cognition in general — the so-called synthetic methodology — and the other is to build systems that have capabilities that are rarely found in technical artifacts (i.e. artificial systems) but are commonly found in humans and some animals. The motivation for the first is a principled one, the motivation for the second is a practical one. Which of these two aspects you choose to focus on has far-reaching effects on the approach you will end up taking in designing a cognitive architecture. One is about advancing science and the other is more about effective engineering. These two views are obviously different but they are not necessarily complementary. There is no guarantee that success in designing a practical cognitive architecture for an application-oriented cognitive robot will shed any light on the more general issues of cognitive science and it is not evident that efforts to date to design general cognitive architectures have been tremendously successful for practical applications.

The origins of cognitive architectures reflects the former principled synthetic methodology. In fact, the term cognitive architecture can be traced to pioneering research in cognitivist cognitive science by Allen Newell and his colleagues in their work on unified theories of cognition [2]. As such, a cognitive architecture represents any attempt to create a theory that addresses a broad range of cognitive issues, such as attention, memory, problem solving, decision making, and learning, covering these issues from several aspects including psychology, neuroscience, and computer science, among others. A cognitive architecture is, therefore, from this perspective at least, an over-arching theory (or model) of human cognition. It continues today under the banner of artificial general intelligence, emphasizing human-level intelligence. The term cognitive architecture is employed in a slightly different way in the emergent paradigm of cognitive science where it is used to denote the framework that facilitates the development of a cognitive agent from a primitive state to a fully cognitive state. It is a way of dealing with the intrinsic complexity of a cognitive system by providing a structure within which to embed the mechanisms for perception, action, adaptation, anticipation, and motivation that enable development over the systems life-time. Nevertheless, even this slightly different usage reflects an endeavour to construct a viable model that sheds light on the natural phenomenon of cognition.

From these perspectives - cognitivist and emergent - a cognitive architecture is an abstract meta-theory of cognition and, as such, focusses on generality and completeness (e.g. see [3]). It reflects Krichmar's first aspect of the goal of building a cognitive robot: to gain a better understanding of cognition in general [1]. We draw from many sources in shaping these architectures. They are often encapsulated in lists of desirable features (sometimes referred to as desiderata) or design principles [4], [5], [6], [7]. A cognitive architecture schema is not a cognitive architecture: it is a blueprint for the design of a cognitive architecture, setting out the component functionality and mechanisms for specifying behaviour. It describes a cognitive architecture at a level of abstraction that is independent of the specific application niche that the architecture targets. It defines the necessary and sufficient software components and their organization for a complete cognitive system. The schema is then instantiated as a cognitive architecture in a particular environmental niche. This, then, is the first approach to designing a cognitive architecture (or a cognitive architecture schema). We refer to it as *design by desiderata.*

The second approach is more prosaic, focussing on the practical necessities of the cognitive architecture and designing on the basis of user requirements. We refer to this as *design by use case*. Here, the goal is to create an architecture that addresses the needs of an application without being concerned whether or not it is a faithful model of cognition. In this sense, it is effectively a conventional system architecture, rather than a cognitive architecture per se, but one where the system exhibits the required attributes and functionality, typically the ability to autonomously perceive, to anticipate the need for actions and the outcome of those actions, and

[†]Much of the work described in this paper was conducted while the author was at the University of Skövde, Sweden. This research was funded by the European Commission under grant agreement No: 688441, RockEU2.

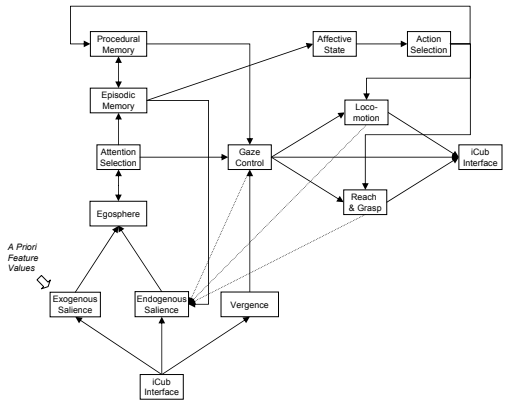

Fig. 1. The iCub cognitive architecture (from [9]).

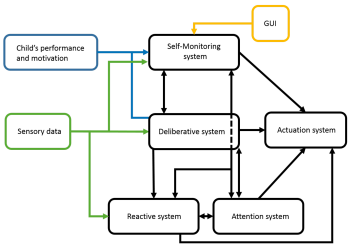

Fig. 2. Project DREAMs cognitive architecture (from [11]).

to act, learn, and adapt. In this case, the design principles, or desiderata, do not drive the cognitive architecture — the requirements do that — but it helps to be aware of them so that you know what capabilities are potentially available and might be deployed to good effect. Significantly, design by use case implies that it is not feasible to proceed by developing a cognitive architecture schema and then instantiating it as a specific cognitive architecture because routing the design through the meta-level schema tacitly abstracts away many of the particularities of the application that makes this approach useful.

We can recast the distinction between the two motivations for building cognitive robots and designing cognitive architectures by asking the following question. Should a cognitive architecture be a specific or a general framework? This is an important design question because a specific instance of a cognitive architecture derived from a general schema will inherit relevant elements but it may also inherit elements that are not strictly necessary for the specific application domain. Also, it is possible that it is not sufficient, i.e. that it does not have all the elements that are necessary for the specific application domain.

To illustrate this argument, consider two architectures that were designed in these two different manners: the iCub cognitive architecture (Figure 1) [8], [9] which was designed by desiderata [9], [7] for use in a general-purpose open cognitive robot research platform, and the DREAM system architecture with its cognitive controller (2) [10], [11] which was designed by use case [12] for use in Robot-Enhanced Therapy targetted at children with autism spectrum disorder. The former comprises components that reflect generic properties of a cognitive system; the latter comprises several functional components that directly target the needs of therapists who can control the cognitive architecture through a GUI.

II. Conclusion

There are two ways not to design a cognitive architecture. If your focus is on creating a practical cognitive architecture for a specific application, you should probably not try to do so by attempting to instantiate a design guided by desiderata; you are probably better off proceeding in a conventional manner by designing a system architecture that is driven by user requirements, drawing on the available repertoire of AI and cognitive systems algorithms and data-structures. Conversely, if your focus is a unified theory of cognition — cognitivist or emergent — then you should probably not try to do so by developing use-cases and designing a matching system architecture. You are likely to miss some of the key considerations that make natural cognitive systems so flexible and adaptable, and it is unlikely that you will shed much light on the bigger questions of cognitive science.

References

[1] J. L. Krichmar, "Design principles for biologically inspired cognitive architectures," *Biologically Inspired Cognitive Architectures*, vol. 1, pp. 73–81, 2012.

[2] A. Newell, *Unified Theories of Cognition*. Cambridge MA: Harvard University Press, 1990.

[3] J. E. Laird, C. Lebiere, and P. S. Rosenbloom, "A standard model of the mind: Toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics." *AI Magazine*, vol. In Press, 2017.

[4] R. Sun, "Desiderata for cognitive architectures," *Philosophical Psychology*, vol. 17, no. 3, pp. 341–373, 2004.

[5] J. L. Krichmar and G. M. Edelman, "Principles underlying the construction of brain-based devices," in *Proceedings of AISB '06 - Adaptation in Artificial and Biological Systems*, ser. Symposium on Grand Challenge 5: Architecture of Brain and Mind, T. Kovacs and J. A. R. Marshall, Eds., vol. 2. Bristol: University of Bristol, 2006, pp. 37–42.

[6] J. E. Laird, "Towards cognitive robotics," in *Proceedings of the SPIE — Unmanned Systems Technology XI*, G. R. Gerhart, D. W. Gage, and C. M. Shoemaker, Eds., vol. 7332, 2009, pp. 73 320Z–73 320Z–11.

[7] D. Vernon, C. von Hofsten, and L. Fadiga, "Desiderata for developmental cognitive architectures," *Biologically Inspired Cognitive Architectures*, vol. 18, pp. 116–127, 2016.

[8] D. Vernon, G. Sandini, and G. Metta, "The iCub cognitive architecture: Interactive development in a humanoid robot," in *Proceedings of IEEE International Conference on Development and Learning (ICDL)*, Imperial College, London, 2007.

[9] D. Vernon, C. von Hofsten, and L. Fadiga, *A Roadmap for Cognitive Development in Humanoid Robots*, ser. Cognitive Systems Monographs (COSMOS). Berlin: Springer, 2010, vol. 11.

[10] D. Vernon, E. Billing, P. Hemeren, S. Thill, and T. Ziemke, "An architecture-oriented approach to system integration in collaborative robotics research projects - an experience report," *Journal of Software Engineering for Robotics*, vol. 6, no. 1, pp. 15–32, 2015.

[11] P. Gomez Esteban, H. Cao, A. De Beir, G. Van De Perre, D. Lefeber, and B. Vanderborght, "A multilayer reactive system for robots interacting with children with autism," in *Proceedings of the Fifth International Symposium on New Frontiers in Human-Robot Interaction*, 2016.

[12] D. David, "Intervention definition," vol. DREAM Deliverable D1.1, 2014.