# Accelerating the Evolution of Cognitive Architectures

**Organizer**
*Kevin Gluck*
Air Force Research Laboratory
711 HPW/RHAC – Cognitive Models and Agents Branch
2620 Q Street, Bldg 852
Wright-Patterson AFB, OH 45433-7955
937-938-3552
kevin.gluck@wpafb.af.mil

The idea that the mind should be rigorously studied in modeling and simulation traces its intellectual roots to the landmark Newell, Shaw, and Simon (1958) paper, in which they proposed information processing models implemented in computer code as explanations of human problem solving capabilities. Fifteen years later Newell (1973) adopted the stronger position that these information processing models must be developed as unified theories of cognition in order to achieve the desired goal of understanding the human mind. In the nearly four decades since, this idea has motivated dozens of new research programs intending to develop integrating, unifying theories, sometimes called *cognitive architectures* – see Byrne (2003), Gray (2008), Langley, Laird, and Rogers (2009), Taatgen and Anderson (2009), and Gluck (2010) for introductions and overviews on this topic. Cognitive architectures are broad, domain-general theories of the mechanisms and structures that enable mind and intelligent behavior. Often overlooked by those not working within cognitive architectural theory is the fact that they are also *evolving*, as functional capabilities expand and as various explanatory mechanisms are evaluated and subsequently incorporated, adapted, or discarded (Cooper, 2007). The breadth and depth of the accomplishments in cognitive architectural theory to date comprise an impressive collection of scientific contributions. While acknowledging and celebrating these achievements, we must also admit to some frustration and concern regarding the slow pace of progress in these integrative systems. This symposium is motivated by the idea that the time is right for us to reconsider the formalisms, methods, and technologies we use to develop and evaluate cognitive architectures, with the goal of accelerating their evolution. Each presenter will explicitly address some of the factors that hinder progress in this area and propose changes in thinking or approach that will overcome those hindrances.

This symposium should be of interest to a broad cross-section of the BRIMS community for at least two reasons. First, there is the fact that the requirement for formal computational implementation and the use of modeling and simulation aligns cognitive architectural theory and progress with the core mission of the BRIMS conference, as described in the call for papers. Second, the development of cognitive architectures is often motivated by an interest in *application*. That is, an objective of cognitive architects is often that the architectures have some applied utility. This may not be universally true, but the evidence for this is clear in some cases. For example, Anderson (1976) actually ends his book introducing the ACT theory with a statement of the importance of application for his research program by saying, "I would like to conclude this chapter with a remark about one of the ultimate goals I have set for my research efforts . . . that is, that it produce a theory capable of practical applications." (p. 535). Newell's (1990) position was that, "Applications are an important part of the frontier of any theory. . . . A unified theory of cognition is the key to successful applied cognitive science." (p. 498). A third example is found in the EPIC architecture, for which some of the earliest publications ( Kieras & Meyer, 1997 ; Kieras, Wood, & Meyer, 1997 ) make it clear that applications in system design served an important motivational role in its creation. The fact that this motivation persists is clear on the EPIC website, which states that EPIC is, ". . . for constructing models of human-system interaction that are accurate and detailed enough to be useful for practical design purposes." The emphasis among cognitive architects on application opportunities aligns this symposium with the BRIMS community's more applied researchers and technologists.

### How Can We Accelerate the Evolution of Cognitive Architectures?

Richard L. Lewis
University of Michigan
rickl@umich.edu

If we take the start of contemporary cognitive science to be the mid 1950s, it is sobering to note that cognitive architectures have been pursued for about half of the field's existence: it has been nearly 30 years since the publication of Anderson's *The Architecture of Cognition*, and over 20 years since Newell's *Unified*

*Theories of Cognition*. It seems potentially useful then to reflect about the state and rate of progress, and ask whether our current approaches to advancing cognitive architecture theory and practice are serving the field well, or in need of substantial modification. Newell considered a very similar question (what is required to move unified theories of cognition forward?) at the end of his 1990 book, and I have found it useful to revisit his injunctions, because they remain relevant:

N1. There should be many unified theories, at least for a while.
N2. We should develop consortia (it takes relatively large communities to work on a cognitive architecture).
N3. Be synthetic—incorporate local theories.
N4. Be prepared to modify existing theories—even strongly and in radical ways.
N5. Make UTCs easy to use.
N6. Acquire domains of application.

The substantive points I wish to make are related to these injunctions and concern diagnoses of problems and conjectures for remedies. I will advance the following specific claims about problems:

P1. When it comes to exploring architectural theory, practice in the community has been too conservative and slow (N4), in part because it has not sufficiently exploited and kept pace with advances in local theory (N3).
P2. When it comes to making UTCs easy to use (N5), the community has tended to focus on improving software artifacts and tools (e.g. to make it easier to program architectures), but this focus on software has been misplaced. It does not address the key barriers to the use and programming of cognitive architectures, which are fundamentally theoretical, and not issues of software engineering. A major theoretical barrier is the perceived increase in degrees of freedom in accounting for data that is introduced by the architectural separation of strategy from fixed structure.
P3. When it comes to acquiring domains of application (N6), standard practice has been to build relatively large models of relatively complex task domains, but this practice is slow and has not led to significant cumulative scientific or applied benefit.

I will make the following conjectures about possible remedies for these problems:

R1. Exploit specific advances in other areas of computational and mathematical cognitive science and machine learning—especially control theory and decision theory perspectives—which will go some way toward addressing (P1) and (P2).
R2. Adopt a "lighter-weight" approach to the construction of cognitive architecture software artifacts that places greater emphasis on theoretical transparency (in part enabled by advances exploited in R1), a clear (even formal) specification of the adaptive/behavioral problem of interest, and rapid architecture exploration.

I will illustrate these remedies briefly with modeling results in one or two "basic science" domains and one "applied" domain. (Candidate domains include eye-movement control in reading, interference in short-term memory, "fast-and-frugal" decision making heuristics, and control of attention in piloted aircraft.)

## Guidelines for the Design of New Cognitive Architectures

Sashank Varma
University of Minnesota
sashank@umn.edu

Cognitive architectures are critical for the design of veridical models of behavior (Anderson, 1983; Newell, 1990). Therefore, the relatively small size and limited diversity of the population of architectures in cognitive science have been limiting factors on the evolution of fitter architectures. The question, then, is how to increase the number and diversity of available architectures (Varma, 2011)? This talk presents guidelines for the construction of new architectures. These guidelines are drawn from the design community, and supported by events in the history of cognitive architecture and computational modeling.

1. First, the architect invents a new style of cognitive information processing. This is a creative or imaginative act.
2. Next, the architect embodies the new style in a new architecture – expresses the desired function in a new form.
3. Finally, the architecture is made available to the cognitive science community. Those members who are transformed by its new view on cognitive information processing – who "actively receive it" – have their research programs reshaped and redirected.

These guidelines have been part of the "black art" of architectural design for the past 30 years, known to only a handful of researchers. This is perhaps one reason why the population of architectures has remained small in size and lacking in the diversity required for evolutionary progress. Public dissemination and critical discussion of these guidelines will enable researchers located outside "architectural hotspots" such as

Pittsburgh and Ann Arbor to contribute to the population of architectures. Increasing its size and diversity promises to accelerate progress towards better computational models of behavior.

### Towards Functionally Elegant, Grand Unified Architectures

Paul S. Rosenbloom
University of Southern California
rosenbloom@usc.edu

When developing cognitive architectures, the ultimate goal is typically a unified theory of intelligent behavior, with the working focus then being on integrating across the capabilities implicated within central cognition, and the result being a unified architecture for cognition. What can be called a *grand unified architecture* sets the bar higher, striving to also include the key non-cognitive aspects of intelligent behavior, such as perception, motor control, personality, motivation and affect. Such architectures can further be considered *functionally elegant* if they provide the requisite breadth of functionality in a simple and theoretically elegant manner, yielding a form of cognitive Newton's laws that provides broad coverage from interactions among a small set of general principles/mechanisms. The pursuit of functionally elegant, grand unified architectures provides a challenging research path, yet one that points the way towards rapid progress beyond today's state of the art, even within the more traditional cognitive focus; and which should also support both deep science and useful systems.

I am currently approaching this goal by rethinking architectures from the ground up, leveraging the interactions between a pair of very general mechanisms – graphical models (factor graphs, in particular) and piecewise continuous multivariate functions – to yield a parameterized space of state-of-the-art capabilities over the processing of symbols, probabilities and signals (Rosenbloom, 2011a-b). The availability of this broad parameterized space promises to accelerate the evolution of cognitive architectures by facilitating the exploration of a wider range of the requisite capabilities and their variations; and without the need to explicitly implement a whole new module for each. Work to date – much of which will be summarized here – demonstrates that within the resulting space can be found: standard flavors of long-term memory, such as a procedural rule-based memory and declarative semantic and episodic memories, plus other variations and blends (Rosenbloom, 2010); forms of knowledge-based (Rosenbloom, 2011c), decision-theoretic (Chen *et al*., 2011) and social problem solving; perception (Chen *et al*., 2011) and mental imagery (Rosenbloom, 2011d); and key bits of language processing. Much more is still required on many of these topics, and additional capabilities must also be added, but the already proven applicability of graphical models to many of these problems shines a bright light on the path towards their rapid incorporation into such a grand synthesis. It also may help understand other topics – such as personality, motivation and affect – that have not previously been investigated via these kinds of techniques. For some capabilities – such as learning – more principles/mechanisms will likely be required, but functional elegance still looks to be within reach, with the inclusion of only a small number of additional general principles/mechanisms.

### Acknowledgments

The participants in this symposium thank the following organizations for the financial support that enabled participation in this symposium and the substantive research on cognitive architecture described therein: the Air Force Research Laboratory's Human Effectiveness Directorate and Office of Scientific Research, the United States Army and the Army Research Laboratory, the Defense Advanced Research Projects Agency, the Federal Aviation Administration, the National Aeronautics and Space Administration, the National Science Foundation, and the Office of Naval Research.

### References

Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1983). *The architecture of cognition*. Mahwah, N.J.: Lawrence Erlbaum.

Byrne, M. D. (2003). Cognitive architecture. In J. Jacko & A. Sears (Eds.) *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications* (pp. 97–117). Mahwah, NJ: Lawrence Erlbaum .

Chen, J., Demski, A., Han, T., Morency, L-P., Pynadath, D., Rafidi, N. & Rosenbloom, P. S. (2011). Fusing symbolic and decision-theoretic problem solving + perception in a graphical cognitive architecture. *Proceedings of the 2nd International Conference on Biologically Inspired Cognitive Architectures* (pp. 64-72). Arlington: VA: IOS Press.

Cooper, R. P. (2007). The role of falsification in the development of cognitive architectures: Insights from a Lakatosian analysis. *Cognitive Science*, *31*(3), 509-533.

Gluck, K. A. (2010). Cognitive architectures for human factors in aviation. In E. Salas & D. Maurino (Eds.) *Human Factors in Aviation, 2nd Edition* (pp. 375-400). New York, NY: Elsevier.

Gray, W. D. (2008). Cognitive architectures: Choreographing the dance of mental operations with the task environment. *Human Factors*, *50* (3), 497–505.

Kieras, D., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction*, *12*, 391–438.

Kieras, D. E., Wood, S. D., & Meyer, D. E. (1997). Predictive engineering models based on the EPIC architecture for a multimodal high-performance human-computer interaction task. *ACM Transactions on Computer-Human Interaction*, *4*, 230–275.

Langley, P., Laird, J. E., & Rogers, S. (2009). Cognitive architectures: Research issues and challenges. *Cognitive Systems Research*, *10*, 141-160.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual Information Processing* (pp. 283-308). New York: Academic Press.

Newell, A. (1990). *Unified Theories of Cognition*. Harvard University Press, Cambridge, MA.

Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review*, *65*, 151–166.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Rosenbloom, P. S. (2010). Combining procedural and declarative knowledge in a graphical architecture. *Proceedings of the 10$^{th}$ International Conference on Cognitive Modeling* (ICCM 2010).

Rosenbloom, P. S. (2011a). Rethinking cognitive architecture via graphical models. *Cognitive Systems Research*, 12(2), 198-209.

Rosenbloom, P. S. (2011b). Bridging dichotomies in cognitive architectures for virtual humans. *Proceedings of the AAAI Fall Symposium on Advances in Cognitive Systems*.

Rosenbloom, P. S. (2011c). From memory to problem solving: Mechanism reuse in a graphical cognitive architecture. *Proceedings of the 4$^{th}$ Conference on Artificial General Intelligence* (pp. 143-152). Mountain View. CA: Springer.

Rosenbloom, P. S. (2011d). Mental imagery in a graphical cognitive architecture. *Proceedings of the 2$^{nd}$ International Conference on Biologically Inspired Cognitive Architectures* (pp. 314-323). Arlington, VA: IOS Press.

Taatgen, N. A. & Anderson, J. R. (2009). The past, present, and future of cognitive architectures. *topiCS in Cognitive Science*, 1-12.

Varma, S. (2011). The design and evaluation of cognitive architectures. *Cognitive Science*, *35*, 1329-1351.