

The SAL Integrated Cognitive Architecture

Christian Lebiere¹, Randall O'Reilly², David J. Jilk³, Niels Taatgen¹ and John R. Anderson¹

¹Carnegie Mellon University
Psychology Department
Pittsburgh, PA 15213

²University of Colorado at Boulder
Psychology Department
Boulder, CO 80309

³eCortex Inc.
3118 Galileo Ln
Boulder, CO 80301

cl@cmu.edu; randy.oreilly@colorado.edu; david.jilk@e-cortex.com; taatgen@cmu.edu; ja@cmu.edu

Abstract

Over the last two decades, the complementary properties of symbolic and connectionist systems have led to a number of attempts at hybridizing the two approaches to leverage their strengths and alleviate their shortcomings. The fact that those attempts have generally fallen short of their goals largely reflects the difficulties in integrating computational paradigms of a very different nature without sacrificing their key properties in the process. In this paper, we propose that biological plausibility can serve as a powerful constraint to guide the integration of hybrid intelligent systems. We introduce a hybrid cognitive architecture called SAL, for "Synthesis of ACT-R and Leabra". ACT-R and Leabra are cognitive architectures in the symbolic and connectionist tradition, respectively. Despite widely different origins and levels of abstraction, they have evolved considerable commonalities in response to a joint set of constraints including behavioral, physiological, and brain imaging data. We introduce the ACT-R and Leabra cognitive architectures and their similarities in structures and concepts then describe one possible instantiation of the SAL architecture based on a modular composition of its constituent architectures. We illustrate the benefits of the integration by describing an application of the architecture to autonomous navigation in a virtual environment and discuss future research directions.

Introduction

A critical feature of robust systems is that they have multiple different methods of solving problems. Those different methods have complementary strengths and weaknesses that prevent catastrophic failures (because when one method fails, another can take over) and boost the overall performance of the system (by synthesizing the results of the different methods). For example, the human visual system has several mechanisms for obtaining depth cues. The system that results from the integration of these mechanisms is thus robust to a large variety of challenging visual conditions. At the broadest level, human intelligence gains considerable robustness by having both subsymbolic and symbolic capabilities. We have recently characterized these two capabilities in terms of analog and digital computational properties that emerge from computational models based on cognitive neuroscience data (O'Reilly, 2006). In the brain, subsymbolic, or analog, computation is supported by the posterior cortex, where graded, overlapping, distributed representations enable powerful

sensory-motor processing abilities that are robust to novelty, ambiguity and degradation. Furthermore, these analog properties are critical for supporting powerful learning mechanisms that can develop new representational structures through incremental, distributed changes in a malleable substrate. In contrast, symbolic, or digital, computation is supported by neural specializations in the prefrontal cortex and basal ganglia, that produce bistable (binary and discrete) neural activation states, and dynamically-gated processing that can perform rule-like operations on information with relative independence of its content (like a transistor gate). These symbolic processing abilities enable us to go beyond the limitations of purely analog processing, to reason abstractly about problems, apply declarative rules and generally engage in multi-step problem-solving processes. Through a comparison with other mammals, which have much of our subsymbolic capabilities but lesser amounts of the symbolic; or with traditional von Neumann architecture computer systems, which far exceed us in their symbolic capabilities but have little or none of the subsymbolic – we can see that it is the unique combination of these complementary abilities that makes human intelligence so powerful, and this is what have begun capturing in this research.

The symbolic/subsymbolic division corresponds largely to the distinction between control and representation subsystems, respectively. Those subsystems have complementary strengths and weaknesses, and a combination of the two is essential in implementing a broadly effective system. However, as one would expect and as some past efforts have shown, the specific manner in which they are combined is a critical determinant as to whether the potential functional benefits of the combination are realized, and in that respect, empirical guidance from neuroscience is essential. Advantages of subsymbolic systems for processing and representing large amounts of data include:

- Speed: distributed connectionist systems can process information in a few parallel steps much more rapidly than sequential symbolic algorithms.
- Capacity: massive parallelism at the cortical level enables the processing of much larger amounts of data in perceptual and memory areas than sequential processes.
- Robustness: unlike symbolic systems, distributed representations generalize naturally to new situations and their performance degrades gracefully in the face of

erroneous, imprecise or unexpected information, or damage to the system.

Conversely, advantages of symbolic systems for controlling the operations of the system include:

- **Tractability:** Enforcing sequentiality sacrifices speed to enable tractable control over the flow of execution and reduce combinatorial complexity.
- **Inspectability:** Sequential control steps enable explicit memory of past processing and the metacognitive introspection essential to avoiding local solution minima.
- **Efficiency:** learning sequential control is considerably more efficient for symbolic systems than it is for subsymbolic systems.

Past attempts to develop symbolic/subsymbolic hybrid systems (see e.g. Wermter and Sun, 2000 for a review) have typically not reflected this division of labor. Instead of applying symbolic techniques to control functions and subsymbolic techniques to representation functions, they often use both techniques for both functions. This dilutes rather than exploits the benefits of each type of technique, resulting in systems that struggle in both control and representation of information required for robust cognition. We have started to develop a new cognitive architecture, called SAL, that integrates widely used and validated architectures at the symbolic and subsymbolic levels, ACT-R and Leabra. Reflecting the division outlined above, ACT-R has focused on the control aspects of synchronizing individual perception, cognitive and motor actions to accomplish complex tasks in dynamic environments, while Leabra has focused on learning and processing robust representations of challenging stimuli. As we detail later, for each of these architectures there is an associated large-scale mapping of specialized subsystems onto various parts of the brain, and the mappings of the two architectures are highly compatible with each other. Furthermore, ACT-R already includes some important graded subsymbolic elements, while Leabra has some initial models of symbol-like processing supported by the prefrontal cortex and basal ganglia. Thus, relative to past approaches to developing symbolic/subsymbolic hybrid models, SAL leverages not only the extensive independent successes of each of these existing architectures, but also their considerable and principled compatibility for integration. Because our approach incorporates strong constraints and inspiration from neuroscience and cognitive data, it remains close to the only known working example of the functionality we seek to replicate, the human brain. Indeed, the strong role that cognitive neuroscience data has played in constraining these architectures is likely why they have independently arrived at a similar view of the cognitive system.

We have applied an initial, relatively shallow integration of the SAL architecture to the task of navigation and object recognition in a virtual environment. Even this relatively straightforward and low-risk first-pass integration has produced important tangible benefits. However, ultimately

we need to go beyond this simple modular hybridization scheme toward a deeper synthesis where elements of each architecture interact in a more tightly coupled and biologically inspired manner.

SAL Architecture

Our approach to integrated intelligence is based on the integration of two existing cognitive architectures, ACT-R (Anderson et al., 2004; Anderson, 2007) and Leabra (O'Reilly & Munakata, 2000) into a hybrid symbolic-connectionist architecture (SAL, for Synthesis of ACT-R and Leabra) distinguished by its biological plausibility. It will inherit both efficiency and robustness to noise, uncertainty, and change from the combination of diverse methods and representations. ACT-R is a high-functionality architecture based on the combination of a symbolic production system with an subsymbolic activation calculus, and already performs well on the goal of arbitrary taskability, including interpreting explicit instructions from humans. Leabra is a high neural-fidelity architecture, and already performs well in adapting to new environments through a combination of powerful learning mechanisms. The ACT-R and Leabra architectures are both characterized by their attempt to account for a wide range of cognitive and neural phenomena using a small and therefore strongly constrained set of computational primitives, as contrasted with the predominant “one-off” and “grab-bag” cognitive models in the field. These architectures have been focused on largely complementary domains: Leabra on the neural mechanisms underlying the processing of individual stimuli and short sequences thereof, and ACT-R on more abstract, controlled cognition unfolding over longer time scales. These architectures are each arguably the most successful in their domain at rigorously accounting for a wide range of cognitive and neural phenomena, with each model providing detailed accounts of hundreds of distinct types of data. Although they are targeted at different levels of abstraction, ACT-R and Leabra have deep compatibilities because of their common conception of the brain mapping of their components that enables them to be synthesized into the SAL architecture. Given their independent success at describing the human cognitive system, it is reassuring, and quite remarkable, that they have arrived at very convergent views of the overall cognitive architecture. This convergence is particularly significant given that the Leabra architecture is derived from more bottom-up neuro-computational constraints about the kinds of processing different parameterizations of a common neural substrate can support, while ACT-R is derived more top-down based on regularities and constraints present in human cognitive performance. Their combination into a modular system that leverages their properties in computationally tractable form while preserving their mathematical foundations will yield a system that is able to achieve robust decision making in concert with human decision makers.

ACT-R

ACT-R (Anderson et al, 2004) is a cognitive architecture whose initial development was driven by modeling phenomena from the psychology laboratory. As can be seen by visiting the ACT-R web site (<http://act-r.psy.cmu.edu/>), successful models have been developed for a wide range of tasks involving attention, learning, memory, problem solving, decision making, and language processing. Under the pressure of accommodating this range of tasks the architecture has developed fairly detailed modules that represent perceptual attention, motor programming, long-term declarative memory, goal processing, mental imagery and procedural competence. Recent years have seen a major effort to apply detailed modeling approach in ACT-R to the performance of significant real-world tasks. These applications have included driving (Salvucci, 2006), aircraft maneuvering (Byrne & Kirlik, 2005), simulated agents for computer-generated forces (Best & Lebiere, 2006), and tutoring systems of academic skills, particularly high school mathematics (Anderson & Gluck, 2001).

In response to these efforts, a new variant of the ACT-R architecture called ACT-R 6.0 has been created (Anderson, 2007). While it continues the basic tenets of the earlier models, it cleanly breaks out components into separate modules, including declarative, procedural, goal, imaginal, visual and manual modules (see Figure 1). The information processing in each of these modules is largely encapsulated from the information processing in others and proceeds asynchronously in parallel across all modules. These modules communicate with one another by putting information into limited-capacity buffers. A system of production rules coordinates the action of these modules via these buffer contents – for example, production rules can recognize patterns in the buffers and make requests of modules.

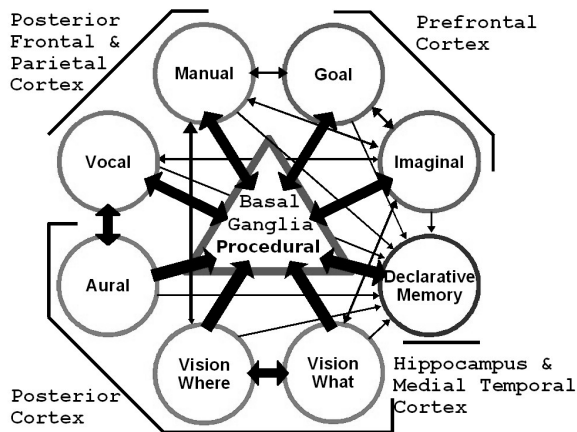


Figure 1. Overview of ACT-R architecture, showing different modules classified according to tripartite organization of Leabra (see Figure 2). Limited-capacity buffers (not shown) mediate communication between procedural system and other modules.

Three major gains have resulted from this explicit modular design. First and obviously, it has been easier to develop components of the architecture separately and implement them efficiently. Second and less obviously, it has enabled us to try out different modules, often borrowed from other architectures, and to insert specific ACT-R modules into other architectures. Third and perhaps least obvious, the modules appear to have mappings to brain regions which has enabled us to use cognitive neuroscience data, particularly brain imaging, to guide further development of models and the architecture.

The perceptual and motor modules are critically important for modeling “embodied” cognition (the visual and manual modules are illustrated in Figure 1) by enabling cognition to interact with its environment in a principled way. Our development of these has been heavily influenced by the EPIC architecture (Meyer & Kieras, 1997). A critical aspect of learning involves optimizing the perceptual-motor operations of the model. A number of applications have now appeared where one or more ACT-R models interact with other agents and complex external systems. These applications depend critically on the ability for external events to interrupt and direct processing.

Leabra

Leabra can be described at two distinct levels: the large-scale architecture of the cognitive system (at the level of brain areas), and the basic neural mechanisms upon which this architecture operates. The two interact in important ways, in that the neural mechanisms constrain and inform the nature of the large-scale architecture. Each brain area uses the same set of core neural mechanisms, yet they differ in their parameterization to maximize different mutually incompatible cognitive functions.

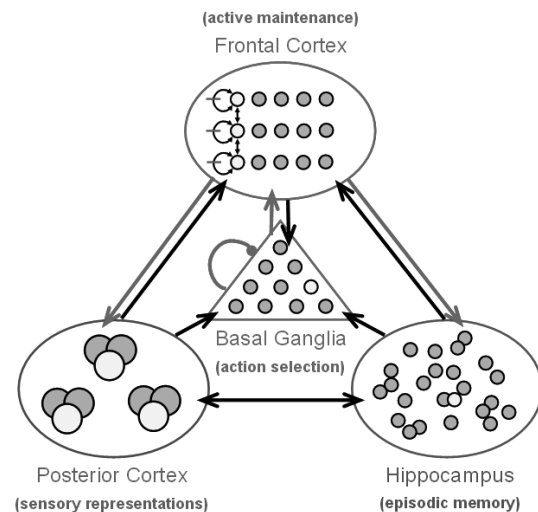


Figure 2. Overview of the Leabra architecture

The core of the large-scale architecture includes three major brain systems: the posterior cortex, specialized for perceptual and semantic processing using slow, integrative learning; the hippocampus, specialized for rapid encoding

of novel information using fast, arbitrary learning; and the frontal cortex/basal ganglia, specialized for active (and flexible) maintenance of goals and other context information, which serves to control (bias) processing throughout the system (Figure 2). This latter system also incorporates various neuromodulatory systems (dopamine, norepinephrine, acetylcholine) that are driven by cortical and subcortical areas (e.g., the amygdala, ventral tegmental area (VTA), substantia nigra pars compacta (SNc), locus ceruleus (LC)) involved in emotional and motivational processing. These neuromodulators are important for regulating overall learning and decision-making characteristics of the entire system. Properties of basic neural mechanisms suggest this large-scale specialization of the cognitive architecture. For example, a single neural network cannot both learn general statistical regularities about the environment and quickly learn arbitrary new information (e.g., new facts, people's names, etc.; McClelland, McNaughton, & O'Reilly, 1995; O'Reilly & Rudy, 2001; O'Reilly & Norman, 2002). Specifically, rapid learning of arbitrary new information requires sparse, pattern-separated representations and a fast learning rate, whereas statistical learning requires a slow learning rate and overlapping distributed representations. These properties correspond nicely to known biological properties of the hippocampus and neocortex, respectively. Many empirical studies by ourselves and other researchers, specifically motivated by our computational modeling work, have tested and confirmed these and other more detailed properties (for details, see the Leabra web site at <http://grey.colorado.edu/emergent/index.php/Leabra>). A similar kind of reasoning has been applied to understanding the specialized properties of the frontal cortex (particularly focused on the prefrontal cortex) relative to the posterior neocortex and hippocampal systems. The tradeoff in this case involves specializations required for maintaining information in an active state (i.e., maintained neural firing, supported by the frontal cortex) relative to those required for performing semantic associations and other forms of inferential reasoning (supported by the posterior cortex). The prefrontal cortex system also requires an adaptive gating mechanism (O'Reilly & Frank, 2006), to be able to rapidly update some (new) information, such as a new subgoal, while simultaneously maintaining other information that remains relevant, such as the super-ordinate goal. The basal ganglia have appropriate neural properties to provide this function (Frank, Loughry, & O'Reilly, 2001).

At the lower level of fundamental neural mechanisms, Leabra integrates into one coherent framework a set of basic neural learning and processing mechanisms (see O'Reilly, 1996; O'Reilly, 1998; O'Reilly & Munakata, 2000; O'Reilly, 2001) that have been otherwise separately investigated in the neural modeling community. Making all these elements work together in a biologically plausible manner is non-trivial, and requires some novel mechanisms, including: a point-neuron activation function that uses simulated ion channels to update a membrane

potential with a nonlinear thresholded output to other neurons; bidirectional (i.e., interactive, recurrent) excitatory projections (which are ubiquitous in neocortex) that propagate information throughout the network, integrating information processing on a cycle-by-cycle basis across the different specialized brain areas; inhibitory competition that greatly constrains and speeds up the constraint satisfaction process that generates a good representation of the current perceptual inputs; and a synthesis of error-driven, Hebbian, and reinforcement learning, which together produce better overall learning than any of them alone (O'Reilly, 2001; O'Reilly & Munakata, 2000; O'Reilly & Frank, 2006; O'Reilly, Frank, Hazy, & Watz, 2007).

Cognitive Synergies

Essential to our hypothesis that SAL, in contrast with prior attempts at hybrid models, can successfully integrate symbolic/digital and subsymbolic/analog processing, is the principled and biologically mapped agreement between ACT-R and Leabra. At the level of large-scale neural systems, this agreement is evident in a comparison between Figures 1 and 2. Both posit a central role for the basal ganglia in receiving converging inputs from a wide range of cortical processing areas, which then drives the performance of specific motor or cognitive actions. Anatomically, the basal ganglia output primarily to the frontal cortex, which is associated with active maintenance of task relevant information in Leabra, and with the homologous buffers of ACT-R. Similarly, both architectures highlight the importance of the declarative/episodic memory system supported by the hippocampus and related anatomical structures. Finally, with little possible controversy, both adopt specialized sensory and motor processing pathways that are well characterized in posterior cortex.

At the cognitive level, the division of the cognitive architecture into procedural and declarative components is shared by both frameworks. This distinction and dissociation has clear cognitive validity: people can possess abstract declarative knowledge of how to do something, yet be procedurally incapable of doing so (e.g., driving a car or playing golf), and vice-versa (e.g., touch typists often cannot recall where the keys are located). In the Leabra architecture, different types of processing are supported by the neural specializations present in the basal ganglia, compared with those present in the hippocampus and cortex. The basal ganglia system is strongly modulated by dopamine, which signals reward and punishment information. Positive reward reinforces associated procedural actions, while negative feedback reduces the likelihood of producing associated actions. A similar, more abstract form of reinforcement learning is present in the ACT-R procedural system, where the integrated history of past success and time-cost are the major determinants in selecting which production will fire. On the other hand, as

mentioned above, the neural properties of the hippocampus have been shown in the Leabra framework to be critical for the rapid learning of new arbitrary information without interfering with existing knowledge. The declarative system in ACT-R integrates both of these properties: new chunks of knowledge, encoded as combinations or modifications of existing chunks, can be rapidly formed and accessed unambiguously; chunks that are used more frequently over time gain higher levels of activation and correspond to more expert knowledge; numerical similarities can be defined between symbolic chunks to drive semantic generalization to related situations.

Although dissociable, the procedural and declarative systems interact intimately in any complete cognitive process. In ACT-R, the firing of productions is driven by the active contents of the declarative and other information buffers, and the result of production firing is the updating of these buffers and the creation of new declarative memory chunks. In Leabra, the basal ganglia procedural system is tightly linked with the prefrontal cortex, which maintains task-relevant information in an active state over time. One of the primary functions of the basal ganglia in the brain is to drive the updating of these prefrontal active memory states. These prefrontal areas then influence activation states throughout the rest of the cortex via strong top-down excitatory projections. Each area of posterior cortex has an associated prefrontal area, with which it has strong bidirectional excitatory connectivity. Thus, we associate the buffers of ACT-R with these prefrontal representations of corresponding posterior cortical areas.

As mentioned above, both ACT-R and Leabra include reinforcement learning mechanisms to shape the procedural processing system. This form of learning uses success and failure information to shape the probability of selecting a given action in the future, and is dissociable from the form of learning that shapes cortical and hippocampal declarative representations. Both architectures also make use of associative learning mechanisms to modulate the strength of representations in declarative memory. Subsymbolic declarative quantities in ACT-R are learned according to Bayesian statistical algorithms, while new declarative representations in Leabra are learned using a combination of error-driven and Hebbian learning. Such learning mechanisms are based on the history of activation of the information stored in declarative memory, but, critically, not on the success, failure, or costs of a particular action taken using that memory. This fact clearly dissociates these mechanisms from procedural reinforcement learning, and numerous cognitive experiments have validated this property of declarative memory. In terms of processing information already stored in declarative memory, the concept of spreading activation is critical to both architectures. In ACT-R, activation spreads among declarative chunks in proportion to their associative strength, and similarities between chunks determine the degree of match. In Leabra,

a similar activation spreading dynamic emerges, in that coarse-coded distributed representations in posterior cortical areas cause associated representations to overlap and share activation states.

It is worth pointing out that in the combined architecture, most major machine learning techniques are represented, and in forms that are motivated and informed by human psychology and biology. However, in SAL these learning methods have the additional advantage that the knowledge representations upon which they operate can be learned and manipulated directly by the architecture; while in most computationally motivated machine learning, the representations are external to the algorithms and must be specified by the human developers. This difference has the potential to provide an unparalleled flexibility and adaptability in behavior.

In summary, the ACT-R and Leabra frameworks overlap to a considerable degree in their basic claims about the nature of the cognitive and neural architecture, despite having been developed from very different perspectives. This overlap is what makes their integration in the SAL architecture both possible and desirable; combined with their independent successes in modeling cognitive phenomena and in performing important cognitive functions, this suggests that development of SAL is a promising approach to integrated, embodied cognition.

Initial Integration and Demonstration

The SAL team has built a demonstration model representing a preliminary synthesis of the two architectures. This demonstration performs a simple version of a seemingly ubiquitous challenge task for embodied robotic systems: searching for a target object within an environment containing multiple rooms, with different objects distributed in the different rooms.

The integration is of the simplest form, whereby the visual module in an existing ACT-R model of navigation is replaced with a Leabra vision model, which is capable of processing raw bitmap images in a way that the ACT-R visual module was not capable of doing (see Figure 3). Similarly, extant Leabra models are not capable of organizing problem solving behavior over a period of several minutes, as the ACT-R model does in searching for the target object in a complex environment. Thus, this hybrid SAL model represents a new level of functionality that goes beyond the capabilities of its constituent architectures. Given the success of this simplest form of integration, we are optimistic that much more interesting and powerful forms of cognition can be captured as our integration efforts develop further.

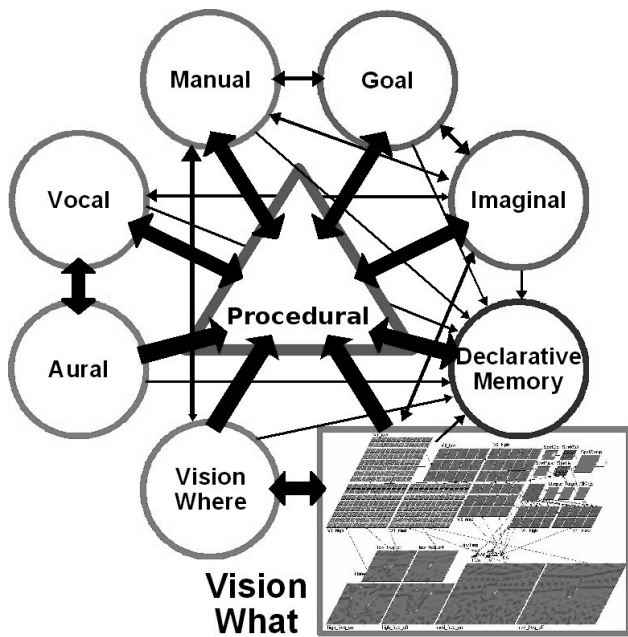


Figure 3: The SAL demonstration model architecture.

Navigation through the known environment is performed by the ACT-R component, and visual object identification is handled by a Leabra neural network component. The neural network replaces ACT-R's "vision: what" module; the ACT-R production system replaces Leabra's prefrontal cortex / basal ganglia element.

In the demonstration (see Figure 4), the SAL agent is embodied within an Unreal Tournament simulation environment consisting of three rooms containing three categories of objects. It is familiar with the environment in that it has access to navigation points and object location points in symbolic form. Further, its Leabra-based vision module has been trained to perceptually identify the three object categories from bitmaps under a variety of viewing angles and distances. An operator instructs SAL to find the desired target via a typed command ("find armor"). SAL must then navigate the rooms; view and perceptually identify each object; and, when it recognizes the desired target, navigate to it and pick it up. ACT-R must decide which navigation point to visit next and which objects to view from that navigation point. Upon selection of the object to view, ACT-R provides the digital image that is a snapshot of the view of that object from the SAL agent's current location. Leabra attempts to identify the object in the center of the image, and responds with its conclusion in symbolic form, which is derived from a localist neural representation. If the object matches the specified target, ACT-R navigates to it, picks it up, and navigates back to the starting point; otherwise it selects a new object to view or a new location to which to navigate. The ACT-R component of the model deals with the challenges of performing the search in an efficient manner. It plans where to begin searching, and uses its episodic memory

and inhibition of return capabilities to remember where it has already searched. The Leabra component of the model was trained to recognize the object categories by repeated presentations of each object from a variety of perspectives and distances and in different room backgrounds. Its non-embodied performance on novel examples is 96%, when using a variety of backgrounds and object angles; simplifying the object angles and ensuring perfect foveation increases its performance to 100%.

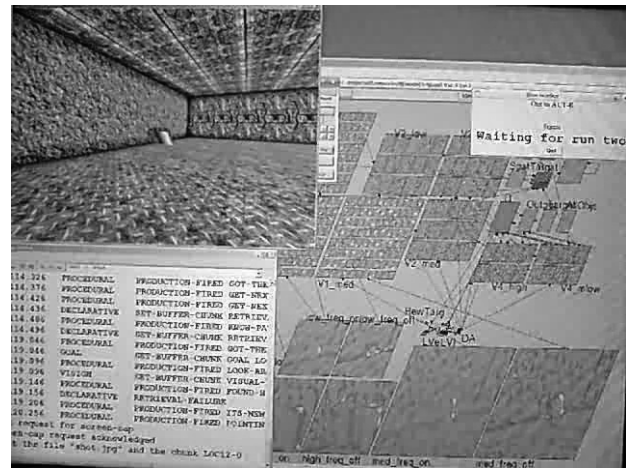


Figure 4: A view of the SAL demonstration system. At top left is a sample view of a room with a single object in the virtual environment. At bottom left is a trace of the ACT-R architecture controlling the Leabra visual module. On the right is the state of the Leabra network as it finishes correctly identifying the armor in the scene shown. The network mimics the properties of the visual system, such as: a hierarchy of visual areas, with shortcut connections between them; representations of the image at several scales; and learning from visual experience.

This simple initial effort provides a number of tangible benefits:

- It illustrates a simple connection of the two architectures operating together, and demonstrates that it is possible to bridge the gap between their different levels of description.
- It serves as a first "embodiment" of a combined model operating in a simulation environment.
- It demonstrates that the system can adapt to new environments and be tasked with new instructions without reprogramming, two essential cognitive capabilities.

There are also a number of clear directions for improvement in this tasks, including using the visual system for navigation as well as object identification, using higher-level movement operations, learning the characteristics of new perceptual inputs, integrating percepts into episodic memory, increasing robustness to perceptual and motor errors in the symbolic controller and more flexibly learning from experience.

This demonstration project has been useful in revealing some challenges that arise in constructing a truly embodied

and flexible cognitive architecture. These challenges highlight gaps in existing psychological theory, and suggest that building embodied architectures may be crucial to progress in the field. The project has also established the surprising power and flexibility available with even a simple combination of ACT-R and Leabra architectures. The two architectures are quite compatible despite two different levels of abstraction, suggesting that these two approaches have converged on a correct overall theory of cognition. The combined SAL architecture is capable of accomplishing tasks that are fundamentally beyond the reach of either ACT-R or Leabra in isolation. Deeper integration of other modules, including procedural, declarative, motor and visual modules will both raise additional issues but also deliver increasing benefits.

Acknowledgments

This research was funded by grants to John Anderson and Randall O'Reilly under the DARPA/IPTO BICA program.

References

- Anderson, J. R. 2007. How can the human mind occur in the physical universe? New York, NY: Oxford Univ. Press.
- Anderson, J. R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y. (2004) An integrated theory of Mind. *Psychological Review*, 111(4), 1036-1060.
- Anderson, J. R. & Gluck, K. 2001. What role do cognitive architectures play in intelligent tutoring systems? In D. Klahr & S. Carver (Eds.) *Cognition & Instruction: Twenty-five years of progress*, 227-262. Erlbaum.
- Anderson, J. R. & Lebiere, C. 1998. *The Atomic Components of Thought*. Mahwah, NJ: Erlbaum.
- Best, B. J. & Lebiere, C. 2006. Cognitive agents interacting in real and virtual worlds. In R. Sun (ed.), *Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation*. Cambridge University Press; New York, NY, 186-218.
- Byrne, M. D., & Kirlik, A. 2005. Using computational cognitive modeling to diagnose possible sources of aviation error. *International Journal of Aviation Psychology*, 15, 135-155.
- Frank, M.J., Loughry, B. & O'Reilly, R.C. 2001. Interactions between the frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, and Behavioral Neuroscience*, 1, 137-160.
- McClelland, J.L., McNaughton, B.L. & O'Reilly, R.C. 1995. Why There are Complementary Learning Systems in the Hippocampus and Neocortex: Insights from the Successes and Failures of Connectionist Models of Learning and Memory. *Psychological Review*, 102, 419-457.
- Meyer, D. E., & Kieras, D. E. 1997. A computational theory of executive control processes and human multiple-task performance: Part 1. Basic Mechanisms. *Psychological Review*, 104, 3-65.
- O'Reilly, R.C. 1996. Biologically Plausible Error-driven Learning using Local Activation Differences: The Generalized Recirculation Algorithm. *Neural Computation*, 8, 895-938.
- O'Reilly, R.C. 1998. Six Principles for Biologically-Based Computational Models of Cortical Cognition. *Trends in Cognitive Sciences*, 2, 455-462.
- O'Reilly, R.C. 2001. Generalization in Interactive Networks: The Benefits of Inhibitory Competition and Hebbian Learning. *Neural Computation*, 13, 1199-1242.
- O'Reilly, R.C. (2006). Biologically-Based Computational Models of High-Level Cognition. *Science*, 314, 91-94.
- O'Reilly, R.C. & Frank, M.J. 2006. Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, 18, 283-328.
- O'Reilly, R.C., Frank, M.J., Hazy, T.E. & Watz, B. 2007. PVLV: the primary value and learned value Pavlovian learning algorithm. *Behavioral Neuroscience*, 121, 31-49.
- O'Reilly, R.C. & Norman, K.A. 2002. Hippocampal and Neocortical Contributions to Memory: Advances in the Complementary Learning Systems Framework. *Trends in Cognitive Sciences*, 6, 505-510.
- O'Reilly, R.C. & Rudy, J.W. 2001. Conjunctive Representations in Learning and Memory: Principles of Cortical and Hippocampal Function. *Psychological Review*, 108, 311-345.
- O'Reilly, R.C. & Munakata, Y. 2000. *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. Cambridge, MA: MIT.
- Salvucci, D. D. 2006. Modeling driver behavior in a cognitive architecture. *Human Factors*, 48, 362-380.
- Wermter, S. and Sun, R. eds. 2000. *Hybrid Neural Systems*. Springer-Verlag, Heidelberg.