

# Reflections on Abstractions for General Artificial Intelligence

John E. Laird

Department of Electrical Engineering and Computer Science  
University of Michigan  
Ann Arbor, MI 48109-2121  
laird@umich.edu, shiwali@umich.edu

## Abstract

This paper proposes that the “right” abstraction for representing general intelligence depends on the timescale of behavior under study (Newell 1990) and overall goals of the research – is it to faithfully model the brain, the mind, or to achieve the same functionality? I briefly describe my approach, which focuses on functionality and time scales above .1 seconds. My strategy is to draw inspiration from neuroscience and cognitive psychology to achieve general intelligence through the study and development of the Soar symbolic cognitive architecture.

How should intelligence be abstracted in AI research? To answer this question, one must first look at the goals of the research. Is the research focused on solving a specific set of problems, or is it focused on understanding the class of problems for which an abstraction is most appropriate? Or is the research pursuing general intelligence, where the representations and associated processes cannot be optimized for a specific problem? If general intelligence is the goal, is the approach based on faithfully modeling human intelligence, or instead is the approach to develop computational systems that achieve the same functionality without replicating the underlying mechanisms?

The vast majority of researchers in AI are not engaged in the struggle to develop and evaluate the suitability of different representations and associated processes for achieving general intelligence. Instead, many of them attempt to develop and use representations for solving specific problems or limited classes of problems. Alternatively, many researchers study the performance (theoretically and empirically) of a given representation across classes of problems. These problems do not require general intelligence, but have characteristics that we associate with some aspect(s) of intelligent behavior. As a result, we see a rich diversity of approaches, tailored to the properties of problems. This divide-and-conquer approach

has led to great progress in the field and many important applications of AI research. I credit the willingness of the field to not be hampered by the goal of general intelligence as one of the reasons for AI’s spectacular success in the last decade (although I personally am unwilling to give up that constraint).

The original question of this paper has more “bite” when we consider general intelligence. Here the answer also depends on the level of intelligent activity being modeled. Newell (1990) observed that human activity can be classified by different levels of processing, grouped by timescales at twelve different orders of magnitude, starting with 1 ms. and extending through years. He grouped these into four bands: biological, cognitive, rational, and social. The lowest level corresponds to the timescale of processing for an individual neuron. What this hierarchy suggests, and is borne out in the diversity of research in neuroscience, psychology, AI, economics, sociology, political science, is that there are regularities at multiple levels that are productive for studying intelligence.

While neuron models, and the general abstraction of graphical models, have made great strides in the last ten years, especially in detecting statistical regularities in large data sets and discovering useful intermediate features for classifying complex items, there is still a large gap between that level of processing and the processing required in the cognitive and rational bands. Those bands are more generally associated with general intelligence – complete end-to-end autonomous taskable and goal-directed behavior that is informed by large bodies of diverse forms of knowledge. There are promising efforts, such as Leabra (O’Reilly et al., in press) and Spaun (Eliasmith, 2013); however, they struggle scale to large bodies of knowledge, achieve real-time reasoning across the breadth of symbolic reasoning, and learn new tasks.

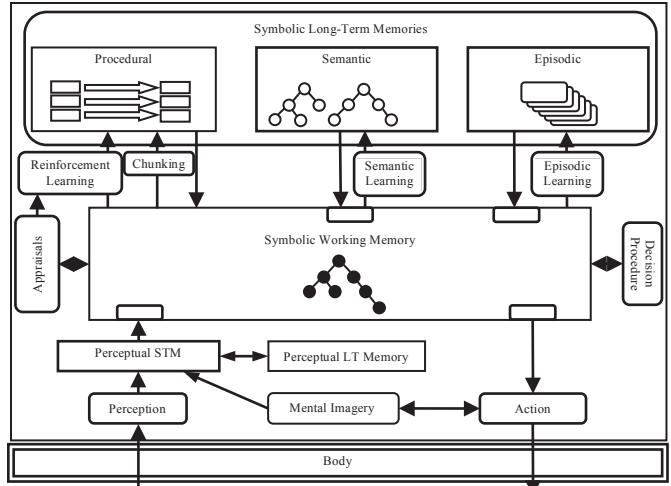
My own goal is not to faithfully model human intelligence, but draw inspiration for psychology and neuroscience in order to achieve the same functionality as achieved by humans. My levels of interest are the cognitive

and rational bands, spanning activity from approximately .1 sec. to hours. These are the levels that have been studied by much of the traditional research in AI, covering reactive behavior, goal-directed decision making, natural language processing, planning, and so on. Fundamental to these levels is the assumption that the abstraction afforded by symbol systems captures the essential capabilities required for human-level intelligence (Newell 1990), while abstracting away from the lower levels and the details of the underlying technology. Within research focused on these levels, the subfield of symbolic cognitive architectures has arisen (Langley, Laird, and Rogers 2009), whose goal is to create generally intelligent real-time agents, with end-to-end (usually embodied) behavior. By focusing on architecture, we study both individual components and their integration.

There are currently many different symbolic cognitive architectures (e.g., ACT-R, CogPrime, ICARUS, LIDA, Polyscheme, and Soar). Across those architectures, there are two shared representational commitments. The first is that relational symbolic structures are used for an agent's internal representation of the current situation. The second is that non-symbolic representations influence knowledge retrieval and decision making, as well as being the basis for perceptual processing and motor control.

As an example, Figure 1 shows the structure of the Soar cognitive architecture (Laird 2012). Symbolic structures are extracted from perception, and added to the symbolic working memory. Symbolic representations are also used for procedural knowledge (production rules), and long-term declarative knowledge (semantic and episodic). Soar accesses its procedural knowledge by matching the conditions of rules to working memory, and firing all rules that match. Decision making is based on preferences retrieved from procedural memory, and those preferences can be symbolic ( $A > B$ ), but they can also be numeric quantities that represent the expected values of proposed operators. These expected values are automatically updated by reinforcement learning as an agent interacts with its environment. Data from long-term declarative memories is retrieved by the creation of symbolic cues which are matched to the contents of the memory, biased by numeric activations. Activation information is also maintained for working memory and is used to support forgetting.

Soar's perception system includes a mental imagery component (called SVS), which uses a scene graph to represent perceived and imagined objects, and associated spatial metric information. Soar agents can extract symbolic spatial relations from SVS, as well as project objects into SVS where the agent can perform non-symbolic spatial reasoning operations, such as imagining the path of an object and detecting if it would collide with other objects in the environment.



*Figure 1. The structure of the Soar.*

Using Soar, we have developed agents that embody many capabilities associated with intelligence.<sup>1</sup> The most advanced agent in terms of the combinations and integration of features is Rosie,<sup>2</sup> an agent learns from situated interactive instruction (Mohan et al. 2012; Mohan, Kirk, and Laird 2013). The agent is embodied with a Kinect sensor and a robotic arm, and manipulates foam blocks. The agent starts with procedural knowledge for general language processing, instruction interpretation and learning, dialog management, and performing primitive actions (pickup and putdown), and some limited declarative knowledge of locations in its world. Using interactive instruction, Rosie learns new nouns/adjectives (relating to color, size, and shape), prepositions (relating to spatial relations), and verbs (compositions of the primitive actions). After learning, Rosie can describe object properties and their spatial relations, and execute the learned verbs. Moreover, the learned words can be used in future instructions and interactions with the agent. In addition, Rosie can learn new tasks, which it then solves using its robotic arm. For example, in learning Towers of Hanoi, a human instructor describes the spatial relations, actions, and goal for Towers of Hanoi. Following instruction, the agent solves the problem internally using iterative deepening, and then executes the solution by moving blocks.<sup>3</sup> The agent learns semantic, procedural, and episodic knowledge, and uses mental imagery for internally projecting the actions. To date, Rosie has been taught ten different games and puzzles and is able to

<sup>1</sup> See Laird, (2012) for an analysis of the requirements I associate with intelligence and for an evaluation of how well Soar supports them.

<sup>2</sup> RObotic Soar Instructable Entity.

<sup>3</sup> The robot does not have the dexterity to stack disks on pegs, so we teach it an isomorphism of the puzzle that involves stacking blocks in three towers.

transfer knowledge learned in one to others that use the same concepts.

Our continued success in extending Soar and using it to create agents that embody many of the capabilities we associate with intelligence makes me optimistic that symbolic systems are an appropriate abstraction for achieving general intelligence. However, I see the value in the capabilities of graphical architectures, and there is always the chance that symbolic architectures will fail to capture some aspects of intelligence that require a lower level of abstraction. Thus, one interesting line of research is to consider how cognitive architectures can be realized in lower-level abstractions, such as graphical models and neural networks. This is the line of research being pursued with Sigma, which is a graphical-model architecture inspired by Soar (Rosenbloom 2013), and SAL (Jilk et al. 2008), which integrates ACT-R and Leabra.

My current hypothesis is that neural/graphical models are necessary for modeling perception and some aspects of motor control, but that symbolic representations, augmented with non-symbolic metadata (such as activation or appraisals) for biasing retrieval and decision making, are sufficient for achieving the breadth we see in general intelligence. They also are much easier to work with, more efficient, and can scale to very large bodies of knowledge. In terms of functionality, the symbol structures make it possible for Rosie to learn completely new tasks from natural language instruction – not just new policies as in reinforcement learning, but the set of available actions, goal conditions, and constraints that define a task. After learning a task, Rosie can solve it (through internal search) and transfer some of what it learned to new tasks. This ability of *cognitive taskability* appears to be a major strength of the cognitive architecture approach, but remains a challenge for other approaches.

As it is, I have not experienced sufficient shortcomings in our development and use of Soar to lead me to consider lower-level alternatives. There needs to be compelling functionality that arises from a lower-level of abstraction (without sacrifice of functionality or computational efficiency) before I will make a shift in my overall approach to modeling intelligence through symbolic cognitive architecture.

## References

- Eliasmith, C. (2013). *How to Build a Brain*. Oxford: Oxford University Press.
- Jilk, D. J., Lebiere, C., O'Reilly, R.C., and Anderson, J. R. (2008). SAL: an explicitly pluralistic cognitive architecture. *Journal of Experimental and Theoretical Artificial Intelligence* 20, 3.
- Laird, J. E. (2012). *The Soar Cognitive Architecture*. MIT Press.
- Langley, P., Laird, J. E., and Rogers, S. (2009). Cognitive architectures: Research issues and challenges. *Cognitive Systems Research*, 10 (2).
- Mohan, S., Mininger, A., Kirk, J., & Laird, J. (2012). Acquiring grounded representations of words with situated interactive instruction. *Advances in Cognitive Systems*, 2012.
- Mohan, S., Kirk, J., & Laird, J. (2013). A computational model for situated task learning with interactive instruction, *International Conference on Cognitive Modeling*, 2013.
- Newell, A. (1990). *Unified Theories of Cognition*. Harvard University Press.
- O'Reilly, R.C., Hazy, T.E. & Herd, S.A. (in press). The Leabra cognitive architecture: How to play 20 principles with nature and win! S. Chipman (Ed.) *Oxford Handbook of Cognitive Science*, Oxford: Oxford University Press.
- Rosenbloom, P. S. (2013). The Sigma cognitive architecture and system. AISB Quarterly, 136, 4-13.