Précis of 'A Standard Model of the Mind'

Paul S. Rosenbloom ROSENBLOOM@USC.EDU
Institute for Creative Technologies & Department of Computer Science, University of Southern California, 12015 Waterfront Dr., Playa Vista, CA 90094 USA

John E. Laird LAIRD@UMICH.EDU
Department of Electrical Engineering and Computer Science, University of Michigan, 2260 Hayward St., Ann Arbor, MI 48109 USA

Christian Lebiere CL@CMU.EDU
Psychology Department, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213 USA

Abstract

A *standard model* captures a community consensus over a coherent region of science, such as particle physics. Here we summarize the key points from a longer article (Laird, Lebiere & Rosenbloom, 2017) that proposes developing such a model for *human-like minds*.

1. Introduction

A mind is a functional entity that can think, and thus support intelligent behavior. Artificial intelligence, cognitive science, neuroscience, and robotics all contribute to our understanding of minds, although each draws from a different perspective. Artificial intelligence concerns building artificial minds, and thus cares most about how systems can be built that exhibit intelligent behavior. Cognitive science concerns modeling natural minds, and thus cares most about understanding cognitive processes that yield human thought. Neuroscience concerns the structure and function of brains, and thus cares most about how brains induce minds. Robotics concerns building and directing artificial bodies, and thus cares most about how minds control such bodies.

Will research across these disciplines ultimately converge on a single understanding of mind? This is a deep scientific question to which there is as yet no answer. However, there must at least be a single answer for cognitive science and neuroscience, as they both investigate the same mind, or narrow class of minds, albeit at different levels of abstraction. Research that is inspired by natural systems also may fit within this class of minds, particularly if the class is slightly abstracted; but so may research that has no such aspiration yet still finds itself in the same neighborhood for functional reasons. This broader class effectively comprises *human-like minds*.

The purpose of the effort summarized here is to engage the international community in developing a *standard model of the mind*, focused on human-like minds. The notion of a standard model has its roots in physics, where such a model has been developed for particles. This standard model is assumed to be internally consistent, yet still have major gaps. It serves as a cumulative reference point for the field while also driving efforts to both extend and break it.

As with physics, a standard model of the mind could provide a coherent baseline that facilitates shared cumulative progress, including a framework for organizing and sharing data for use in evaluating it and its alternatives. For integrative researchers, it could also help focus work on differences between particular approaches and the standard model, and on how to both extend and break the model. For theoretical and systems researchers, it could also provide guidance when they seek to expand to include aspects of other components. For experimental researchers, it could also provide top-down guidance in interpreting results, as well as suggesting new experiments. For practitioners, it could also provide a sound basis for guiding development.

Our hypothesis is that *cognitive architectures* provide the appropriate computational abstraction for defining a standard model, where cognitive architectures are models of the fixed structure of the mind. However, the standard model itself is not such an architecture. What is sought instead is a statement of the best consensus given the community's current understanding of the mind, plus a sound basis for further refinement as more is learned. Achieving this depends on researchers within the community relating their own approaches to the standard model and participating in its evolution. In the process, disagreements with aspects of the standard model presented here are expected, leading ideally to efforts to either disprove or improve parts of it. The standard model is also incomplete in significant ways, particularly in aspects where an adequate consensus has not yet been achieved. While individuals may disagree with aspects of the proposed standard model – consensus after all does not require unanimity – it is our attempt to provide a coherent summary along with a broadly shared set of assumptions held in the field.

2. The Standard Model

The standard model summarized here first began to emerge at the 2013 AAAI Fall Symposium on Integrated Cognition, but has since been further grounded in three particular architectures: ACT-R (Anderson, 2007), Soar (Laird, 2012) and Sigma (Rosenbloom, Demski, & Ustun, 2016). The first two are the most complete, long-standing, and widely applied architectures, while Sigma is a more recent development. Although ACT-R originated within cognitive science, and Soar and Sigma within artificial intelligence, all three have since reached out to more of the relevant disciplines. The ultimate goal is to ground the standard model much more broadly; however, these three were chosen for now because we know them well. In our experience, unless an expert on an architecture is involved in such a process, the interpretations of it can be problematic, so inclusion of other architectures should ideally wait until others are involved.

The structure of a cognitive architecture defines how information and processing are organized into components, and how information flows between them. The standard model posits that the mind is not an undifferentiated pool of information and processing, but is built out of a relatively small number of independent modules with distinct functionalities. Figure 1 shows the core modules of the standard model, with each potentially either unitary or further decomposed into sub-modules, e.g.: multiple perceptual and motor modalities, multiple working memory buffers, and semantic vs. episodic declarative memory. Outside of direct connections between the perception and motor modules, working memory acts as the inter-component communication buffer for modules. Procedural memory has access to all of working memory, while the other modules are restricted to accessing and modifying their associated working memory buffers. All long-term memories have one or more associated learning mechanisms. One of the major points of consensus that was not possible in earlier days is the distinction between procedural and declarative long-term memories seen in Figure 1.

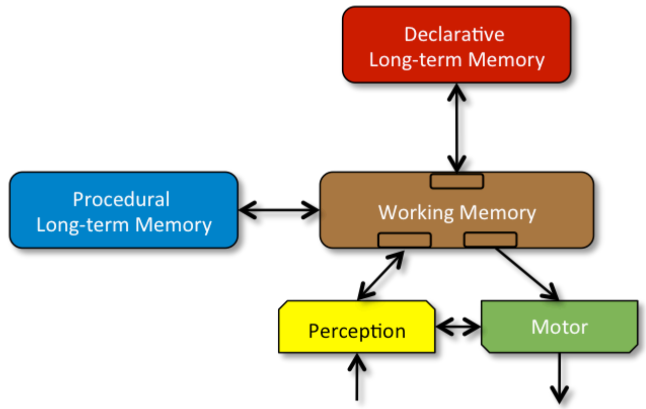

Figure 1. The structure of the standard model.

The heart of the standard model is the *cognitive cycle*, which operates at ~50 ms/cycle when mapped to human behavior. Procedural memory induces the selection of a single deliberate act per cycle, which can modify working memory, initiate the retrieval of knowledge from long-term declarative memory, initiate motor actions in an external environment, and provide top-down influence to perception. Complex behavior, such as planning, arises from sequences of cycles rather than from additional architectural mechanisms that operate over longer time scales. The restriction to a single deliberate act per cycle yields a serial bottleneck; however, significant parallelism can occur within any of the components, as well as across components.

The memory components in the model store, maintain, and retrieve content to support their specific functionalities. The core of this content is represented as relations over symbols. However, both symbols and relations can be annotated by *quantitative metadata* for the purpose of modulating their storage (i.e., for learning) and retrieval, and for guiding decision making. Frequency information is a pervasive form of metadata, yielding a statistical aspect to the knowledge representation. Other examples include recency, co-occurrence, similarity, utility, and more general notions of activation. The inclusion of quantitative metadata, which yields tightly integrated hybrid symbolic-subsymbolic representations and processing, is perhaps the most dramatic evolution from the early days of (purely) symbolic cognitive architectures.

Working memory provides a temporary global space within which symbol structures can be composed from the outputs of perception and long-term memory. Procedural memory contains knowledge about actions, including how to select them and how to cue (for external actions) or execute (for internal actions) them, to yield both skills and procedures. It is based on pattern-directed invocation of actions, typically in the form of rules with conditions and actions, although also with quantitative metadata that influences action selection. Declarative memory is a long-term store for facts and concepts. It comprises a persistent graph of symbolic relations, with metadata reflecting attributes such as recency and (co-)occurrence that are relevant to learning and retrieval. Declarative memory can also store the system's experiences in the form of episodic knowledge. There is not yet a consensus concerning whether there is a single uniform declarative memory or whether there are two memories, one semantic and the other episodic.

The standard model assumes that all types of long-term knowledge are learnable, including both symbol structures and their associated metadata. All learning is incremental, and takes place online over the experiences that arise during system behavior. There are at least two learning mechanisms for procedural memory: one that creates new rules from the composition of rule firings, and one that tunes action selection via reinforcement learning. Declarative memory also includes at least two learning mechanisms: for creating new relations and tuning metadata.

The standard model also includes more assumptions about the perception and motor modules, including the role of learning in them, but no more will be said here due to lack of space.

3. Conclusion

The standard model remains incomplete in a number of ways. It is silent, for example, concerning metacognition, emotion, mental imagery, direct communication and learning across modules, the distinction between semantic and episodic memory, and social cognition. However, even with these gaps, it captures much more than could have been agreed upon in earlier decades, reflecting a significant point of convergence, consensus and progress. There was, e.g., significant disagreement in the early-90s on the part of both ACT-R and Soar, yet their current versions are in total agreement with respect to the aspects covered in this summary. Sigma is also mostly in agreement, but differs in defining some of the standard model's modules in terms of more primitive architectural mechanisms plus particular bits of knowledge and skill.

It is hoped that this attempt at a standard model will grow over time to cover more data, applications, architectures, and researchers. A comprehensive standard model of the human mind could play an integrative role in guiding research across the relevant disciplines, enabling more generalizable results and guidance. Conversely, those disciplines could also provide additional insights and constraints on the standard model, leading to further progress and convergence.

Acknowledgements

The work described in this article was sponsored by the Air Force Office of Scientific Research (under award FA9550-15-1-0157), the U.S. Army (under contract W911NF-14-D-0005), the Army Research Laboratory (under contract W911NF-10-2- 0016), and the Office of Naval Research (under awards N00014-15-1-2468, N00014-15-12151, N00014-12-C-0638, and N00014-08-1-0099). Statements and opinions expressed may not reflect the position or policy of the United States Government, and no official endorsement should be inferred.

References

Anderson, J. R. (2007). *How Can the Human Mind Exist in the Physical Universe?* New York, NY: Oxford University Press.

Laird, J. E. (2012). *The Soar Cognitive Architecture*. Cambridge, MA: MIT Press.

Laird, J. E., Lebiere, C. & Rosenbloom, P. S. (2017). A standard model of the mind: Toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. *AI Magazine*. In press

Rosenbloom, P. S., Demski, A. and Ustun, V. (2016). The Sigma cognitive architecture and system: Towards functionally elegant grand unification. *Journal of Artificial General Intelligence*, 7, 1-103.