A Standard Model of the Mind: Toward a Common Computational Framework Across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics

John E. Laird, Christian Lebiere, Paul S. Rosenbloom

■ *A standard model captures a community consensus over a coherent region of science, serving as a cumulative reference point for the field that can provide guidance for both research and applications, while also focusing efforts to extend or revise it. Here we propose developing such a model for humanlike minds, computational entities whose structures and processes are substantially similar to those found in human cognition. Our hypothesis is that cognitive architectures provide the appropriate computational abstraction for defining a standard model, although the standard model is not itself such an architecture. The proposed standard model began as an initial consensus at the 2013 AAAI Fall Symposium on Integrated Cognition, but is extended here through a synthesis across three existing cognitive architectures: ACT-R, Sigma, and Soar. The resulting standard model spans key aspects of structure and processing, memory and content, learning, and perception and motor, and highlights loci of architectural agreement as well as disagreement with the consensus while identifying potential areas of remaining incompleteness. The hope is that this work will provide an important step toward engaging the broader community in further development of the standard model of the mind.*

A mind is a functional entity that can think, and thus support intelligent behavior. Humans possess minds, as do many other animals. In natural systems such as these, minds are implemented through brains, one particular class of physical device. However, a key foundational hypothesis in artificial intelligence is that minds are computational entities of a special sort — that is, cognitive systems — that can be implemented through a diversity of physical devices (a concept lately reframed as substrate independence [Bostrom 2003]), whether natural brains, traditional general-purpose computers, or other sufficiently functional forms of hardware or wetware.

ISSN 0738-4602

Artificial intelligence, cognitive science, neuroscience, and robotics all contribute to our understanding of minds, although each draws from a different perspective in directing its research. Artificial intelligence concerns building artificial minds, and thus cares most for how systems can be built that exhibit intelligent behavior. Cognitive science concerns modeling natural minds, and thus cares most for understanding cognitive processes that generate human thought. Neuroscience concerns the structure and function of brains, and thus cares most for how minds arise from brains. Robotics concerns building and controlling artificial bodies, and thus cares most for how minds control such bodies.

Will research across these disciplines ultimately converge on a single understanding of mind, or will the result be a large but structured space of possibilities, or even a cacophony of approaches? 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 are both investigating the same mind, or narrow class of minds, albeit at different levels of abstraction. Biologically, or cognitively, or psychologically inspired research in artificial intelligence and robotics also may fit within this particular class of minds, particularly if the class is slightly abstracted; but so may other work that has no aspiration to such inspiration yet still finds itself in the same neighborhood for functional reasons. This broader class comprises what can be called *humanlike minds,* with an overall focus more on the bounded rationality hypothesized to be central to human cognition (Simon 1957; Anderson 1990) than on the optimality that is the focus in much of artificial intelligence and robotics. The class is broader than the more familiar one of naturally inspired minds, as it also includes both natural minds and some artificial minds that are not necessarily naturally inspired yet functionally related. However, it is narrower in scope than human-level intelligence, as it excludes minds that are sufficiently inhuman in how they achieve this level of intelligence.

The purpose of this article is to begin the process of engaging the international research community in developing what can be called a *standard model of the mind,* where the mind we have in mind here is humanlike. The notion of a standard model has its roots in physics, where for over more than a half-century the international community has developed and tested a standard model that combines much of what is known about particles. This model is assumed to be internally consistent, yet still have major gaps. Its function is to serve as a cumulative reference point for the field while also driving efforts to both extend and break it.

As with physics, developing a standard model of the mind could accelerate work across the relevant disciplines by providing a coherent baseline that facilitates shared cumulative progress. For integrative researchers concerned with modeling entire minds, a standard model can help focus work on differences between particular approaches and the standard model, and on how to both extend and break the model. Also, instead of each such researcher needing to describe all the assumptions and constraints of their particular approach from scratch, given the standard model they can simply state how their own approach differs from it. Tables 1 and 2 in the summary of this article, for example, specify the standard model developed in this article and the standing of three distinct approaches with respect to it. In this process, the standard model itself could serve as something of an interlingua or shared ontology, providing a vehicle for mapping the common aspects, and possibly uncommon terminology, of disparate architectures onto a common base.

For theoretical and systems researchers who model/build specific components of mind — whether learning, memory, reasoning, language — a standard model can provide guidance when they seek to expand to include aspects of other components. For experimental researchers who tease out the details of how natural minds and brains work, a standard model can provide top-down guidance in interpreting the results, as well as suggesting new experiments that may be worth trying. For all researchers, a standard model can serve as a framework around which data that is used in evaluating single components or combinations of components may be organized and made available for use by the community; potentially growing to yield standard tests and testbeds. A standard model can also provide a sound basis for guiding practitioners in constructing a broad range of intelligent applications.

The intent, at least for the foreseeable future, is not to develop a single implementation or model of mind by which everyone concerned with humanlike minds would abide, or even a theory in which all of the details are agreed to as correct. What is sought though 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. Much of the existing work on integrative models of mind focuses on implementations rather than theory, with too little interchange or synthesis possible across these implementations. The development of a standard model provides an opportunity for the community to work together at a more abstract level, where such interchange and synthesis should be more practicable.

For this to transpire though will depend on researchers within the community being interested in relating their own approaches to the standard model and participating in its further evolution. In the process, it is fully expected that they will disagree with some aspects of the standard model presented here, leading ideally to efforts to either disprove or improve parts of it. It is also expected that the standard model will be incomplete in significant ways, not because those parts that are left out are unimportant, but because an adequate consensus on them has not yet been achieved. Omission from the standard model is thus often a statement of where a consensus is needed, rather than a consensus on a lack of either existence or importance.

Although the boundary around the class of humanlike minds is ill defined, at least at present, we do anticipate an evolving dialogue around this, driven by a sequence of challenges from ideas and data that conflict in substantive ways with the standard model. For each such challenge, it will be critical to determine whether the consensus is ultimately that the standard model should be altered — either changed to eliminate the conflict or abstracted to cover both old and new approaches — or that the new ideas or data should be deemed insufficiently humanlike, and thus outside of the class of interest. These will not necessarily be easy decisions, nor will the process as a whole be smooth, but the potential rewards for succeeding are real.

This article grew out of the 2013 AAAI Fall Symposium on Integrated Cognition that was initiated by two of us to bring together researchers across a set of disparate perspectives and communities concerned with an integrated view of human-level cognition (Burns et al. 2014). The full organizing committee included representatives from cognitive science, cognitively and biologically inspired artificial intelligence, artificial general intelligence, and robotics. The final activity during the symposium was a panel on Consensus and Outstanding Issues, at which two of us presented and the third participated. One of these presentations led to the startling finding that the wide range of researchers in the room at the time agreed that the content of the presentation was an appropriate consensus about the current state of the field. Given the field's history of stark differences between competing approaches, neither of the initiators of the symposium had anticipated this as a realistic outcome, and when it occurred, it startled those in attendance. It implied that a consensus had implicitly begun to emerge — perhaps signaling the dawning maturity of the field — and that an attempt to make it explicit could provide significant value.

This attempt is what fills the remainder of this article. The next section covers important background that largely predates the 2013 symposium and this effort, including several notable precursors to the concept of a standard model of the mind plus the critical notion of a *cognitive architecture* — a hypothesis about the fixed structure of the mind — which is at the heart of this attempt. We then introduce three cognitive architectures on which the effort here focused. That section is followed by a presentation of the proposed standard model that has been developed. In the summary section, we review what has been accomplished, including a précis of the proposed standard model, an analysis of where the same three cognitive architectures sit with respect to it, and a discussion of where we hope it will lead.

Background

This attempt at a standard model of the mind, although originating at the 2013 symposium, did not spring there from nothingness; and Allen Newell was at the root of much of what came before. One notable precursor from three decades earlier is the model human processor (Card, Moran, and Newell 1983), which defines an abstract model of structural and timing regularities in human perceptual, mental, and motor processes. It supports predicting approximate timings of human behavior, but does not include any details of the underlying computational processes.

A second, albeit rather different, precursor is Newell's (1990) analysis of how scale counts in cognition. Newell observed that human activity can be classified according to different levels of processing, and grouped by time scales at 12 different orders of magnitude, starting with 100 µs and extending up to months. While the many disciplines that have studied the nature of the mind have focused on different collections of levels, this analysis provides a coherent framework for integrating research into phenomena and mechanisms at different time scales. As with the notion of a standard model, this echoes the situation in physics, and in fact, all of the physical sciences and beyond, where the core phenomena of interest stratify according to time (and length) scales that when combined can yield models of more complex multiscale phenomena.

Newell grouped these levels into four bands: biological, cognitive, rational, and social. The lowest, biological, band corresponds to the time scale of processing for individual neurons and synapses, the functional building blocks of the human brain that have been the focus of neuroscience research. The next two bands up, the cognitive and rational bands span activity from approximately 100 ms to hours, covering the levels that have been studied by cognitive science as well as traditional AI research in reactive behavior, goal-directed decision making, natural language processing, planning, and so on. The highest, social band includes such higher-order capabilities as Theory of Mind, organizational behavior, and moral and ethical reasoning (as, for example, discussed from different perspectives in two articles in this special issue — Scheutz [2017] and Bello and Bridewell [2017]). What this hierarchy suggests, and what is borne out in the diversity of research in disciplines such as neuroscience, psychology, AI, economics, sociology, and political science, is that there are regularities at multiple time scales that are productive for understanding the mind.

For humans, the *deliberate act level,* at 100 ms, is roughly at the time scale of a simple reaction, although the roughness here obscures the fact that even simple reactions involve multiple internal processes, including perception, cognition, and action. More broadly, the deliberate act level is where elementary operations are selected and applied. Fundamental to this level and all levels above, is the assumption that computational capabilities similar to a physical symbol system are available. The physical symbol systems hypothesis states, "A physical symbol system has the necessary and sufficient means for general intelligent action" (Newell and Simon 1976). However, in a break with tradition, the standard model does not assume that computation at the deliberate act level is purely or perfectly symbolic. We know from the computational universality of symbol systems that they are logically sufficient; however, considerable evidence suggests that many types of reasoning that must be directly available at the deliberate act level, such as statistical and spatial, are best realized there through nonsymbolic processing.

In the standard model, the critical feature of symbols is that they are the primitive elements over which relations can be defined, and where their use across multiple relations enables the creation of complex symbol structures, including (but not limited to) structures such as semantic networks, ontologies, and taxonomies. This use mirrors the binding problem in cognitive neuroscience, which is concerned with how multiple elements can be associated in a structured manner (Treisman 1996). However, the model is agnostic as to whether symbols are uninterpreted labels, such as in Lisp, Soar (Laird 2012), and ACT-R (Anderson 2007), or whether they are patterns over vectors of distributed elements, such as semantic pointers in Spaun (Eliasmith et al. 2012) and holographic vectors in HDM (Kelly, Kwock, and West 2015), or whether both are available, such as in Clarion (Sun 2016) and Sigma (Rosenbloom, Demski, and Ustun 2016a). What is important is that they provide the necessary functionality to represent and manipulate relational structures.

In the standard model, nonsymbolic (that is, numeric) information has two roles. One is to represent explicitly quantitative task information, such as distances in spatial reasoning or times in temporal reasoning. The second is to annotate the representations of task information (symbolic and nonsymbolic) in service of modulating how it is processed. This second type of numeric information takes the form of (quantitative) metadata; that is, (numerical) data about data.

The mind then clearly comprises at least everything from the deliberate act level up; that is, the top three bands in Newell's hierarchy. Many conceptions of the mind, however, also include some portion of the biological band as well, whether in terms of an abstract neural model, or a close cousin such as a graphical model (Koller and Friedman 2009). Whether or not a portion of the biological band is included in the conceptualization, a model of the fixed structure at the deliberate act level, that defines a symbol system and more, is called a *cognitive architecture*. While models of the mind can be defined at different levels, we have situated ours at the deliberate act level because we believe that it represents a critical juncture between the neural processes that underlie it and the (boundedly) rational computations that it gives rise to. The standard model we are striving for here amounts to a consensus on what must be in a cognitive architecture in order to provide a humanlike mind.

In a significant break from much of the early work on cognitive architectures, this standard model involves a hybrid combination of symbolic and statistical processing to match the need introduced earlier for statistical processing in the architecture, rather than retaining a purely symbolic model of processing. In consequence, it also embodies forms of statistical learning, including Bayesian and reinforcement learning. It furthermore embraces significant amounts of parallelism both within modules and across them, while still retaining a serial bottleneck, rather than being strictly serial. Further explanations of these shifts, along with the remaining assumptions that define the standard model, can be found later in this article.

Typical research efforts on cognitive architectures (Langley, Laird, and Rogers 2009) are concerned with much more than just the architectural level — and thus may be more appropriately thought of as developing more comprehensive cognitive systems — although none has yet spanned the entire hierarchy. Often they start with one level, or a few, but over time expand, becoming multiyear — or even multidecade — research programmes (Lakatos 1970) that span larger and larger sequences of levels. However, the standard model will not come anywhere near to providing a direct model of the entire mind. If we again look to the situation in physics, the standard model there is also not a direct model of the entire physical world, focusing as it does only on the relatively low level of particles. Still it provides a critical foundation for the levels above it, up to and including the full universe (or multiverse), while being firmly grounded in, and constrained by, the levels below it. The standard model of the mind likewise directly concerns only one level, but in so doing provides a critical foundation for the higher levels of the mind, while being firmly grounded in, and constrained by, the levels below.

With respect to the higher levels of the mind, there is an ancillary hypothesis to the standard model that they are defined purely by the knowledge and skills that are acquired and processed by the architecture. In simple terms, the hypothesis is that intelligent behavior arises from a combination of an implementation of a cognitive architecture plus knowledge and skills. Processing at the higher levels then amounts to sequences of these interactions over time. Even complex cognitive capabilities — such as natural language processing (as, for example, discussed in another article in this special issue — McShane [2017]) and planning — are hypothesized to be constructed in such a fashion, rather than existing as distinct modules at higher levels. Specific mechanisms can sometimes be decomposed at multiple levels: for example, Forbus and Hinrichs' (2017, this issue) analogy process can be decomposed into a SME mechanism located at least partly at the deliberate act level, together with attendant search processes such as MAC/FAC and SAGE that operate at higher levels and could be decomposed into primitive acts.

The lower levels of the mind — in the biological band or its artificial equivalent — both implement and constrain the cognitive architecture. As the hierarchy shows, the concept of a cognitive architecture, and thus a standard model, need not be incompatible with neural modeling. Moreover, there is potential not only for compatibility, but also for useful complementarity. Aspects of neural processing, such as generalization from distributed representations, have been captured in cognitive architectures in the form of subsymbolic statistical mechanisms. Conversely, the standard model can define an architectural structure that can be beneficial in organizing and supplementing mechanisms such as deep learning when, for example, the need is recognized to move beyond the simple memory capabilities provided by feedforward or recurrent neural networks (for example, Vinokurov et al. [2012]). Furthermore, the traditional notion of a fixed cognitive architecture has always been tempered by the idea that it is fixed only relative to the time scale of normal reasoning processes, leaving open the possibility that a symbol system could emerge or change during development rather than necessarily being in place at birth.

The concept of cognitive architecture originated in Newell's even earlier criticism of task-specific models that induce a fragmented approach to cognitive science and the consequent difficulty of making cumulative progress (Newell 1973). As a solution, he advanced the concept of an integrated model of human cognition on top of which models of specific tasks could be developed in terms of a common set of mechanisms and representations, with the ultimate goal of achieving Unified Theories of Cognition (Newell 1990). Like a computer architecture, a cognitive architecture defines a general purpose computational device capable of running programs on data. The key differences are that: (1) the kinds of programs and data to be supported in cognitive architectures are limited to those appropriate for humanlike intelligent behavior; and (2) the programs and data are ultimately intended to be acquired automatically from experience — that is, learned — rather than programmed, aside from possibly a limited set of innate programs. Cognitive architectures thus induce languages, just as do computer architectures, but they are languages geared toward yielding learnable intelligent behavior, in the form of knowledge and skills. This is what distinguishes a cognitive architecture from an arbitrary — yet potentially quite useful — programming language.

From this common origin, the concept of cognitive architecture took form in multiple subfields, each focused on different goals. In cognitive psychology, architectures such as ACT-R, Clarion, and LIDA (Franklin and Patterson 2006) attempt to account for detailed behavioral data from controlled experiments involving memory, problem-solving, and perceptual-motor interaction. In artificial intelligence, architectures such as Soar and Sigma focus on developing functional capabilities and applying them to tasks such as natural language processing, control of intelligent agents in simulations, virtual humans, and embodied robots. In neuroscience, architectures such as Leabra (O'Reilly, Hazy, and Herd 2016) and Spaun (Eliasmith 2013) adopt mechanisms and organizations compatible with the human brain, but primarily apply them to simple memory and decision-making tasks. In robotics, architectures such as 4D/RCS (Albus 2002) and DIARC (Schermerhorn et al. 2006) concern themselves with real-time control of physical robots.

However, there has historically been little agreement either across or within specialties as to the overall nature and shape of this architecture. The lack of such a consensus has hindered comparison and collaboration across architectures, prevented the integration of constraints across disciplines, and limited the guidance that could aid research on individual aspects of the mind. There is not even an agreed upon term for what is being built. In addition to cognitive architectures — a term that stems from cognitive science — relevant work also proceeds on architectures for intelligent agents, intelligent/cognitive robots, virtual humans, and artificial general intelligence. All these terms carry significantly different goals and requirements that span interaction with and control of, respectively, online resources, artificial physical bodies, and artificial virtual bodies, plus generality across domains. To the extent that the humanlike components of these divergent threads can (re)converge under combined behavioral, functional, and neural constraints, it yields a strong indication that a standard model is possible.

One recent attempt to bring several of these threads back together was work on a "generic architecture for humanlike cognition" (Goertzel, Pennachin, and Geisweiller 2014a), which conceptually amalgamated key ideas from the CogPrime (Goertzel, Pennachin, and Geisweiller 2014b), CogAff (Sloman 2001), LIDA, MicroPsi (Bach 2009), and 4D/RCS architectures, plus a form of deep learning (Arel, Rose, and Coop 2009). A number of the goals of that effort were similar to those identified for the standard model; however, the result was more of a pastiche than a consensus — assembling disparate pieces from across these architectures rather than identifying what is common among them — with a bias thus also more toward completeness than concord.

The standard model developed in this article is grounded in three other architectures and their associated research programs: ACT-R, Soar, and Sigma. The first two are the most complete, long-standing, and widely applied architectures in existence. ACT-R originated within cognitive science, although it has reached out to artificial intelligence as well (for example, Sanner et al. [2000]), been mapped onto regions of the human brain (Anderson 2007) — enabling it to be integrated with the Leabra neural architecture (Jilk et al. 2008) — and been used to control robots (for example, Kennedy et al. [2007]). Soar originated within artificial intelligence, although it has reached out to cognitive science (Newell 1990), and been used to control robots (Laird and Rosenbloom 1990; Laird et al. 2012). Sigma is a more recent development, based partly on lessons learned from the two others. It also originated within artificial intelligence, but has begun to reach out to cognitive science (for example, Rosenbloom [2014]), is based on a generalized notion of graphical models that has recently been extended to include neural networks (Rosenbloom, Demski, and Ustun 2016b), and been used to control virtual humans (Ustun and Rosenbloom 2016).

We selected these three architectures because we know them well. The ultimate goal is to ground the standard model in many more architectures and research programs, but in our experience, unless an expert on the architecture/program is directly involved in such a process, the results can be more problematic than useful, so our decision was to hold off on analyzing additional architectures until we can involve others, possibly through a focused symposium or workshop, and hopefully then follow up with a longer and more comprehensive article. Nevertheless, between these three architectures there is significant presence across artificial intelligence and cognitive science, plus extensions into neuroscience and robotics (and virtual humans), although it should be clear that none of the three architectures actually originated within either of the latter two disciplines.

Three Cognitive Architectures

The previous section introduced the general notion of a cognitive architecture. Here we introduce the three particular architectures we have focused on in extending the standard model beyond the initial synthesis at the Symposium. Each architecture is described in its own terms, along with a figure that provides a standard characterization of its structure. No attempt has been made to alter these figures to draw out their commonalities — for example, the Soar figure explicitly shows learning mechanisms while the other two don't — other than to use a common color scheme for the components: brown for working memory, red for declarative memory, blue for procedural memory, yellow for perception, and green for motor. The core work of identifying commonalities is left to the standard model, as described in the next section.

ACT-R is constructed as a set of modules that run asynchronously and in parallel around a central rule-based procedural module that provides global control (figure 1). Processing is often highly parallel within modules, but each yields only a single result per operation, which is placed in a module-specific working memory buffer, where it can be tested as a condition by the procedural module and transferred to other buffers to trigger further activity in the corresponding modules.

Soar is also comprised of a set of asynchronous internally parallel modules, including a rule-based procedural memory. Soar is organized around a broader-based global working memory (figure 2). It includes separate episodic and semantic declarative memories, in addition to visuospatial modules and a motor module that controls robotic or virtual effectors.

Sigma is a newer architecture that blends lessons from existing architectures such as ACT-R and Soar with what has been learned separately about graphical models (Koller and Friedman 2009). It is less modular architecturally, providing just a single long-term memory, which along with the working memory and perceptual and motor components is grounded in graphical models. It instead seeks to yield the distinct functionalities provided by the other two's modules by specialization and aggregation above the architecture (figure 3). Sigma's long-term memory, for example, subsumes a variety of both procedural and declarative functionalities, while also extending to core perceptual aspects and visuospatial imagery.

All three architectures structure behavior around a cognitive cycle that is driven by procedural memory, with complex behavior arising as sequences of such cycles. In each cycle, procedural memory tests the contents of working memory and selects an action that modifies working memory. These modifications can lead to further actions retrieved from procedural memory, or they can initiate operations in other modules, such as motor action, memory retrieval, or perceptual acquisition, whose results will in turn be deposited back in working memory.

Standard Model

In this section, we present the standard model, decomposed into structure and processing; memory and content; learning; and perception and motor (or, to use a robotics term, action). This model represents

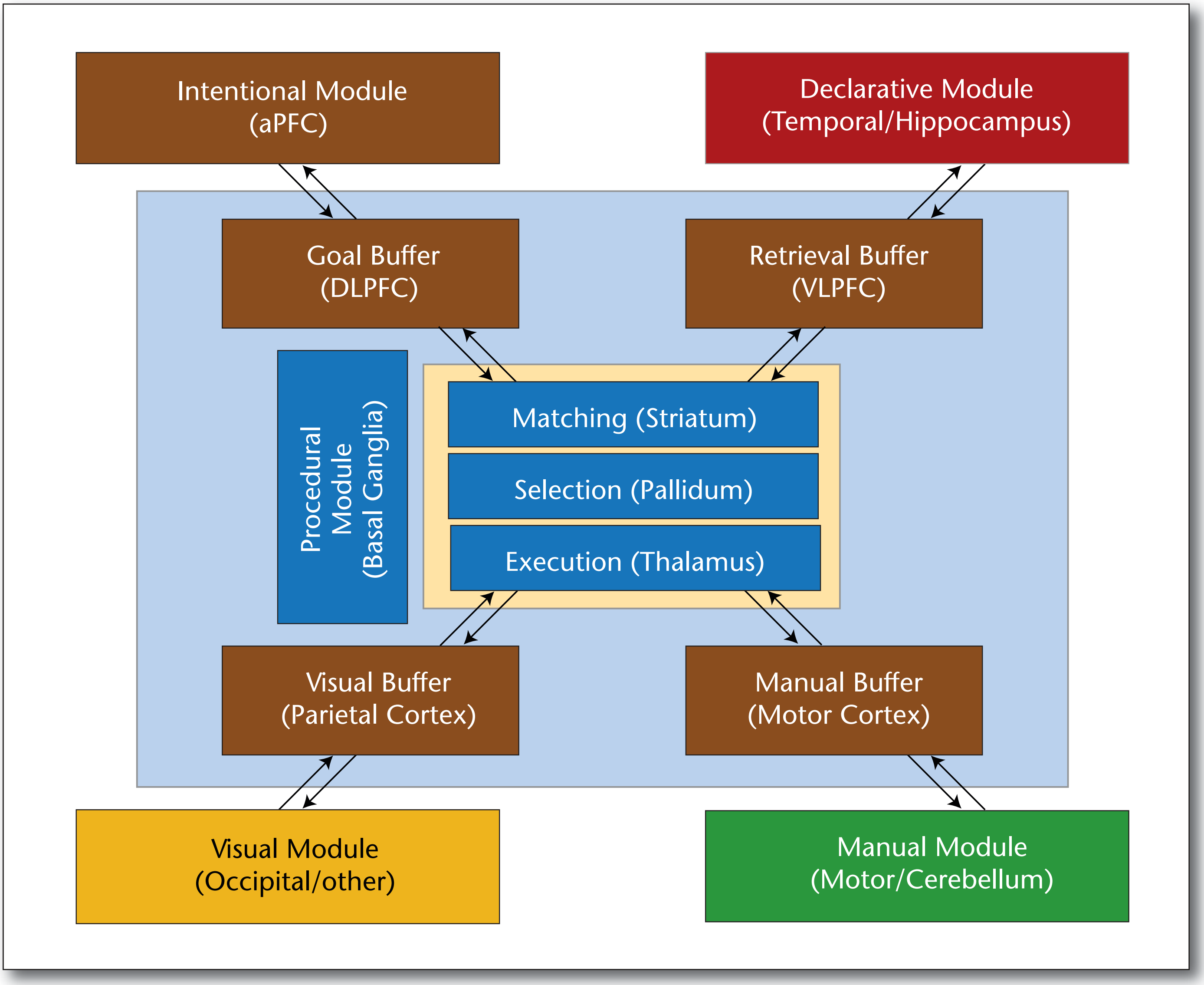

Figure 1. ACT-R Cognitive Architecture.

our understanding of the consensus that was introduced skeletally at the AAAI symposium, as fleshed out based on our understanding of the three architectures of concern in this article. While individuals, including the three of us, might disagree with specific aspects of what is presented here — consensus after all does not require unanimity — it is our attempt at providing a coherent summary along with a broadly shared set of assumptions held in the field. Specific areas of disagreement plus open issues are discussed in the final section.

Structure and Processing

The structure of a cognitive architecture defines how information and processing are organized into components, and how information flows between components. The standard model posits that the mind is not an undifferentiated pool of information and processing, but is built of independent modules that have distinct functionalities. Figure 4 shows the core components of the standard model, which include perception and motor, working memory, declarative long-term memory, and procedural long-term memory. At this granularity, not a great deal of progress can be seen compared to what might have appeared in a Standard Model several decades ago, aside from the distinction here between procedural and declarative long-term memory. However, as will be seen in the rest of this section and summarized in table 1, there is substantial further progress when one looks deeper.

Each of the modules in figure 4 can be seen as unitary or further decomposed into multiple modules or submodules, such as multiple perceptual and motor modalities, multiple working memory buffers, semantic versus episodic declarative memory, and

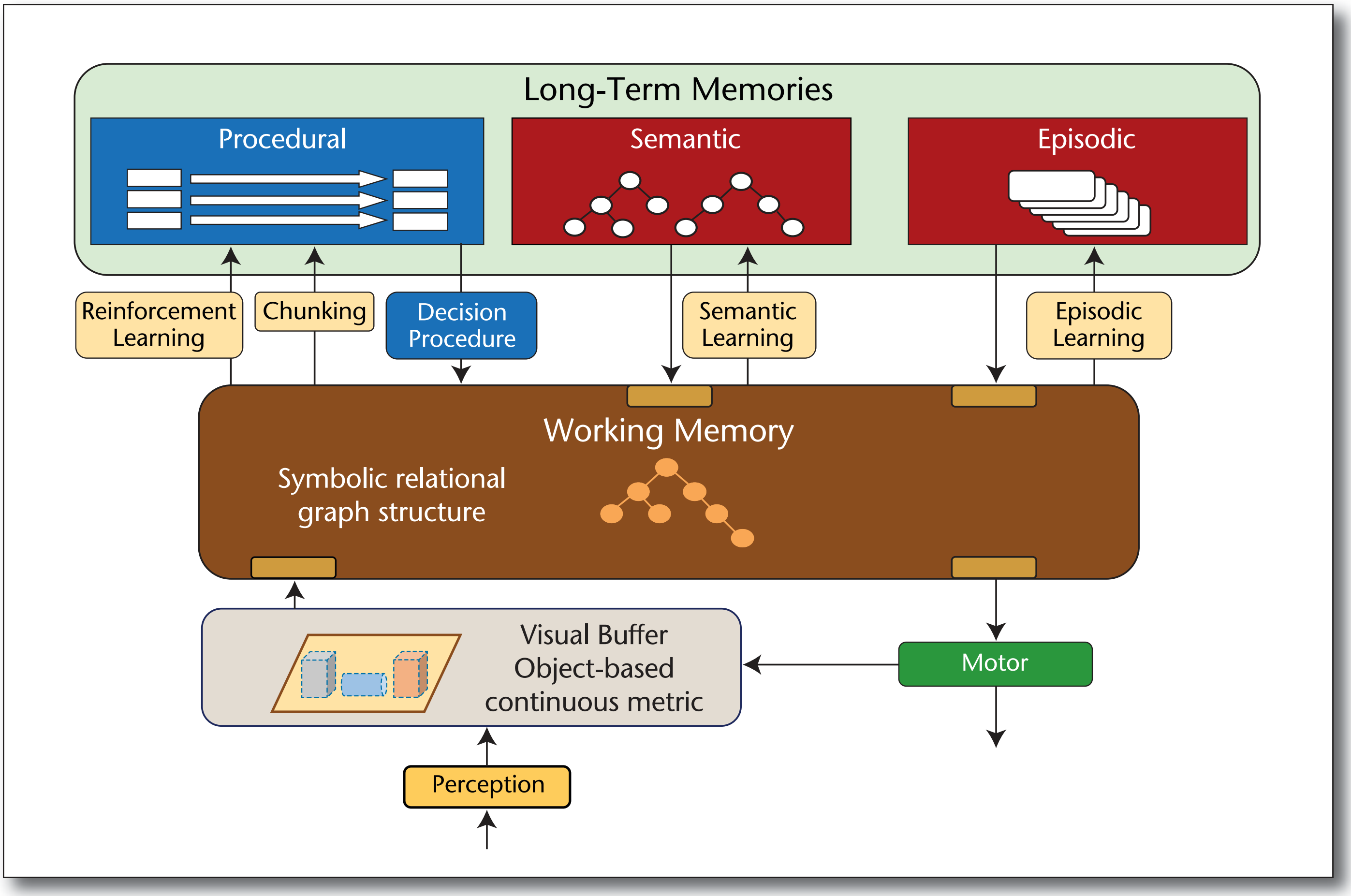

Figure 2. Soar Cognitive Architecture.

various stages of procedural matching, selection and execution. Outside of direct connections between the perception and motor modules, working memory acts as the intercomponent communication buffer for components. It can be considered as unitary, or consist of separate modality-specific memories (for example, verbal, visual) that together constitute an aggregate working memory. Long-term declarative memory, perception, and motor modules are all restricted to accessing and modifying their associated working memory buffers, whereas procedural memory has access to all of working memory (but no direct access to the contents of long-term declarative memory or itself). All long-term memories have one or more associated learning mechanisms that automatically store, modify, or tune information based on the architecture's processing.

The heart of the standard model is the *cognitive cycle*. Procedural memory induces the processing required to select a single deliberate act per cycle. Each action can perform multiple modifications to working memory. Changes to working memory can correspond to a step in abstract reasoning or the internal simulation of an external action, but they can also initiate the retrieval of knowledge from long-term declarative memory, initiate motor actions in an external environment, or provide top-down influence to perception. Complex behavior, both external and internal, arises from sequences of such cycles. In mapping to human behavior, cognitive cycles operate at roughly 50 ms, corresponding to the deliberate-act level in Newell's hierarchy, although the activities that they trigger can take significantly longer to execute.

The restriction to selecting a single deliberate act per cycle yields a serial bottleneck in performance, although significant parallelism can occur during procedural memory's internal processing. Significant parallelism can also occur across components, each of which has its own time course and runs independently once initiated. The details of the internal processing of these components are not specified as part of the standard model, although they usually involve significant parallelism. The cognitive cycle that arises from procedural memory's interaction with working memory provides the seriality necessary for coherent thought in the face of the rampant parallelism within and across components.

Although the expectation is that for a given system there can be additional perceptual and motor mod-

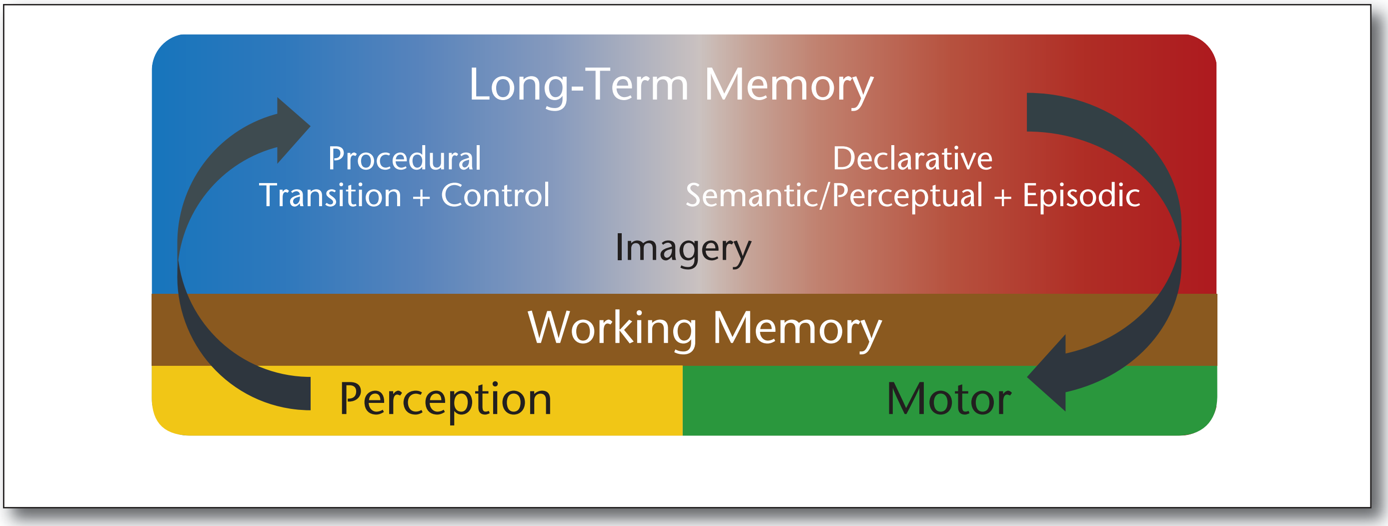

Figure 3. Sigma Cognitive Architecture.

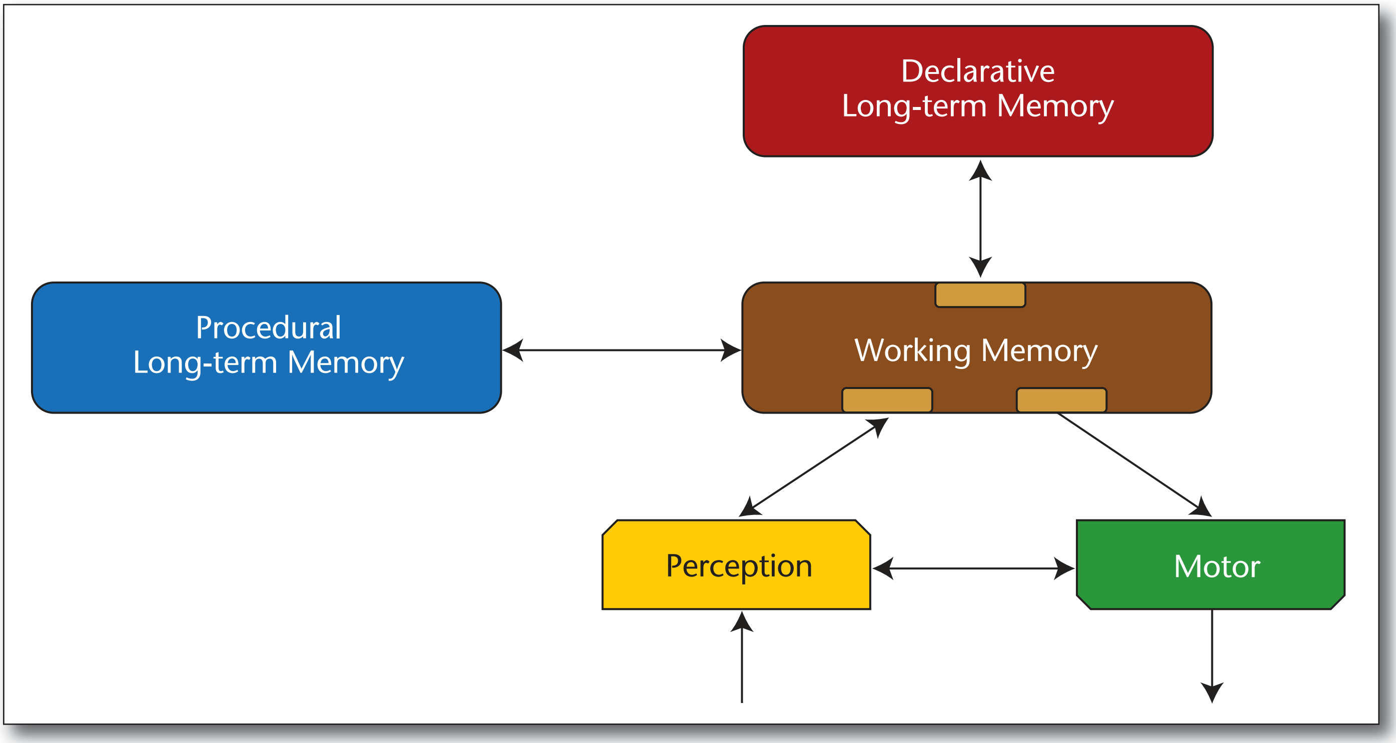

Figure 4. The Structure of the Standard Model.

ules as part of an agent's embodiment, and additional memory modules, such as an episodic memory, there is a strong commitment that no additional specialized architectural modules are necessary for performing complex cognitive activities such as planning, language processing and Theory of Mind, although architectural primitives specific to those activities (for example, visuospatial imagery for planning, or the phonological loop for language processing) can be included. All such activities arise from the composition of primitive acts; that is, through sequences of cognitive cycles. The existence of a cognitive cycle, along with an appropriate procedural memory to drive it, has become definitional for a cognitive architecture.

Memory and Content

The memory components store, maintain, and retrieve content to support their specific functionalities. The core of this content is represented as relations over symbols. However, supplementing these relational structures is quantitative metadata that annotates instances of symbols and relations for the purpose of modulating decision making as well as the storage, retrieval, and learning of symbols and relations. Frequency information is a pervasive form of metadata, yielding a statistical aspect to the knowledge representation (for example, Anderson and Schooler [1991]). Other examples of metadata include recency, co-occurrence, similarity, utility, and more general notions of activation. The inclusion of quantitative metadata, resulting in tightly integrated hybrid symbolic-subsymbolic representations and processing, is perhaps the most dramatic evolution from the early days of (purely) symbolic cognitive architectures (Newell, Rosenbloom, and Laird 1989). There is a strict distinction between domain data — symbols and relations — and such metadata. The metadata only exists in support of the symbolic representations, and relations cannot be defined over quantitative metadata. The set of available metadata for symbols and relations and the associated mechanisms are fixed within the architecture. In a reflective architecture, there may be symbolic relations at a metalevel that can be used to reason about the domain relations, but that is quite different from the architecturally maintained metadata described here, and is not part of the current standard model. A brief summary of each of the three memory components follows.

Working memory provides a temporary global space within which symbol structures can be dynamically composed from the outputs of perception and long-term memories. It includes buffers for initiating retrievals from declarative memory and motor actions, as well as buffers for maintaining the results of perception and declarative memory retrieval. It also includes temporary information necessary for behavior production and problem solving, such as information about goals, intermediate results of a problem, and models of a task. All of working memory is available for inspection and modification by procedural memory.

Procedural memory contains knowledge about actions, whether internal or external. This includes both how to select actions and how to cue (for external actions) or execute (for internal actions) them, yielding what can be characterized as skills and procedures. Arbitrary programs can be thought of generically as a form of procedural memory, but they provide a rigid control structure for determining what to do next that is difficult to interrupt, acquire, and modify. In the standard model, procedural memory is instead based on pattern-directed invocation of actions, typically cast in the form of rules with conditions and actions. Rule conditions specify symbolic patterns over the contents of working memory and rule actions modify working memory, including the buffers used for cuing declarative memory and motor actions. There is variation in how the knowledge from multiple matching rules is integrated together, but agreement that a single deliberate act is the result, with metadata influencing the selection.

Declarative memory is a long-term store for facts and concepts. It is structured as a persistent graph of symbolic relations, with metadata reflecting attributes such as recency and frequency of (co-)occurrence that are used in learning and retrieval. Retrieval is initiated by the creation of a cue in the designated buffer in working memory, with the result being deposited in that buffer. In addition to facts, declarative memory can also be a repository of the system's direct 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 distinction between those terms roughly maps to semantically abstract facts versus contextualized experiential knowledge, respectively, but its precise meaning is the subject of current debate.

Learning

Learning involves the automatic creation of new symbol structures, plus the tuning of metadata, in long-term — procedural and declarative — memories. It also involves adaptation of nonsymbolic content in the perception and motor systems. The standard model assumes that all types of long-term knowledge are learnable, including both symbol structures and associated metadata. All learning is incremental, and takes place online over the experiences that arise during system behavior. What is learned is typically based on some form of a backward flow of information through internal representations of these experiences. Learning over longer time scales is assumed to arise from the accumulation of learning over short-term experiences. These longer time scales can include explicit deliberation over past experiences. Learning mechanisms exist for long-term memory, and although they are not yet fully implemented in current architectures, they are also assumed to exist for the perception and motor modules.

There are at least two independent learning mechanisms for procedural memory: one that creates new rules from the composition of rule firings in some form, and one that tunes the selection between competing deliberative acts through reinforcement learning. Declarative memory also involves at least two learning mechanisms: one to create new relations and one to tune the associated metadata.

Perception and Motor

Perception converts external signals into symbols and relations, with associated metadata, and places the results in specific buffers within working memory. There can be many different perception modules, each with input from a different modality — vision, audition — and each with its own perceptual buffer. The standard model assumes an attentional bottleneck that constrains the amount of information that becomes available in working memory, but does not embody any commitments as to the internal representation (or processing) of information within perceptual modules, although it is assumed to be predominantly nonsymbolic in nature, and to include learning. Information flow from working memory to perception is possible, providing expectations or possible hypotheses that can be used to influence perceptual classification and learning.

Motor converts the symbol structures and their metadata that have been stored in their buffers into external action through control of whatever effectors are a part of the body of the system. As with perception, there can be multiple motor modules (arms, legs). Much is known about motor control from the robotics and neuroscience literature, but there is at present no consensus as to the form this should take in the standard model, largely due to a relative lack of focus on it in humanlike architectures.

Summary

Table 1 summarizes the key assumptions that underlie the standard model of humanlike minds proposed in this article. It is derived from the 2013 symposium session plus an extensive post hoc discussion among the authors of this article centered around ACT-R, Soar, and Sigma. In the table, the standard model has been decomposed into (A) structure and processing, (B) memory and content, (C) learning, and (D) perception and motor systems.

Table 2 provides an analysis, tabulated by the assumptions in table 1, of the extent ACT-R, Soar, and Sigma agree in theory with the standard model and implement the corresponding capabilities. Versions of ACT-R and Soar from the early 1990s have been included to show the evolution of those architectures in relation to the standard model. The convergence is striking. Although there was significant disagreement (or lack of theory, especially in the case of perception and motor) in the early 1990s for both ACT-R and Soar, their current versions are in total agreement in terms of theory and only substantially differ in the extent to which they implement perception and motor systems. Sigma is also in agreement on most of these assumptions as well. However, because it defines some of the standard model's capabilities not through specialized architectural modules but through combinations of more primitive architectural mechanisms plus specialized forms of knowledge and skills, three cells are colored blue to indicate a partial disagreement in particular with the strong architectural distinction between procedural and declarative memories, and the complete architectural nature of reinforcement learning.

This standard model reflects a very real consensus over the assumptions it includes, but it 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 mechanisms necessary for social cognition. However, even with these gaps, the standard model captures much more than did precursors such as the model human processor, and much more than could have been agreed upon even ten years ago. It thus reflects a significant point of convergence, consensus, and progress.

The hope is that the presented model will yield a sound beginning upon which the field can build by folding into the mix additional lessons from a broader set of architectures. Such an effort ideally should focus on architectures that: (1) are under active (or recent) development and use; (2) have strong architectural commitments that yield a coherence of assumptions rather than being just a toolkit for construction of intelligent systems; (3) are concerned with humanlike intelligence; and (4) have been applied across diverse domains of human endeavor. Architectures worth considering for this include, but are not limited to, CHREST (Gobet and Lane 2010), Clarion (Sun 2016), Companions (Forbus and Hinrichs 2006), EPIC (Kieras and Meyer 1997), ICARUS (Langley and Choi 2006), Leabra (O'Reilly, Hazy, and Herd 2016), LIDA (Franklin and Patterson 2006), MicroPsi (Bach 2009), MIDCA (Cox et al. 2013), and Spaun (Eliasmith 2013).

Newell's (1973) warning about trying to approach full intelligence through a pastiche of task-specific models applies not only to cognitive science — and, in particular, psychology and AI — but also to any other discipline that ultimately seeks or depends on such comprehensive models of intelligent behavior, including notably neuroscience and robotics. A comprehensive standard model of the human mind could provide a blueprint for the development of robotic architectures that could act as true human companions and teammates as well as a high-level structure for efforts to build a biologically detailed computational reconstruction of the workings of the brain, such as the Blue Brain project. The standard model could play an integrative role to guide research in related disciplines — for example, ACT-R is already being applied to modeling collections of brain regions and being integrated with neural models, and both ACT-R and Soar have been used in robotics (and Soar and Sigma in the sister discipline of virtual humans) — but the existence of a standard model can enable more generalizable results and guidance. Conversely, those disciplines can provide additional insights and constraints on the standard model, leading to further progress and convergence. In addition, the standard model potentially provides a platform for the integration of theoretical ideas without requiring realization in complete cognitive architectures.

It is hoped that this attempt at a standard model, based as it is on extending the initial sketch from the Symposium through a focus on three humanlike architectures, will grow over time to cover more data, applications, architectures, and researchers. This is

A. Structure and Processing

1. The purpose of architectural processing is to support bounded rationality, not optimality
2. Processing is based on a small number of task-independent modules
3. There is significant parallelism in architectural processing
   a. Processing is parallel across modules
      i. ACT-R & Soar: asynchronous; Sigma: synchronous
   b. Processing is parallel within modules
      i. ACT-R: rule match, Sigma: graph solution, Soar: rule firings
4. Behavior is driven by sequential action selection via a cognitive cycle that runs at ~50 ms per cycle in human cognition
5. Complex behavior arises from a sequence of independent cognitive cycles that operate in their local context, without a separate architectural module for global optimization (or planning).

B. Memory and Content

1. Declarative and procedural long-term memories contain symbol structures and associated quantitative metadata
   a. ACT-R: chunks with activations and rules with utilities; Sigma: predicates and conditionals with functions; Soar: triples with activations and rules with utilities
2. Global communication is provided by a short-term working memory across all cognitive, perceptual,and motor modules
3. Global control is provided by procedural long-term memory
   a. Composed of rule-like conditions and actions
   b. Exerts control by altering contents of working memory
4. Factual knowledge is provided by declarative long-term memory
   a. ACT-R: single declarative memory; Sigma: unifies with procedural memory; Soar: semantic and episodic memories

C. Learning

1. All forms of long-term memory content, whether symbol structures or quantitative metadata, are learnable
2. Learning occurs online and incrementally, as a side effect of performance and is often based on an inversion of the flow of information from performance
3. Procedural learning involves at least reinforcement learning and procedural composition
   a. Reinforcement learning yields weights over action selection
   b. Procedural composition yields behavioral automatization
      i. ACT-R: rule composition; Sigma: under development; Soar: chunking
4. Declarative learning involves the acquisition of facts and tuning of their metadata
5. More complex forms of learning involve combinations of the fixed set of simpler forms of learning

D. Perception and Motor

1. Perception yields symbol structures with associated metadata in specific working memory buffers
   a. There can be many different such perception modules, each with input from a different modality and its own buffer
   b. Perceptual learning acquires new patterns and tunes existing ones
   c. An attentional bottleneck constrains the amount of information that becomes available in working memory
   d. Perception can be influenced by top-down information provided from working memory
2. Motor control converts symbolic relational structures in its buffers into external actions
   a. As with perception, there can be multiple such motor modules
   b. Motor learning acquires new action patterns and tunes existing ones

Table 1. Standard Model Architectural Assumptions.

partially a scientific process and partially a social process. The scientific side is driven by what is learned about humanlike minds from studying both human minds and humanlike artificial minds. The social side needs to be driven by spanning more and more of the community concerned with humanlike cognitive architectures, and possibly even beyond this to other communities with related interests. This could happen incrementally, by expanding to a single new architecture and proponent at a time, or in bursts, through symposia or workshops at which multiple such come together to see what new consensus can be found. Communitywide surveys are also possible, but it is our sense that by sidestepping the hard part of working out differences interactively, this would likely not yield what is desired. Rather, it is our hope that the shared benefits of a standard model of the mind will lead to a virtuous cycle of community contributions and incremental refinements.

	A1	A2	A3a	A3b	A4	A5	B1	B2	B3a	B3b	B4	C1	C2	C3a	C3b	C4	C5	D1a	D1b	D1c	D1d	D2a	D2b
ACT-R 1993	🟩	🟩	🟥	🟩	🟥	🟩	🟩	🟦	🟩	🟦	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟥	🟥
SOAR 1993	🟩	🟩	▒	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟥	▒	🟩	🟥	🟩	🟥	🟩	▒	░	░	░	▒	░
SIGMA 2016	🟩	🟩	🟩	🟩	▒	🟩	🟦	🟩	🟩	🟩	🟦	▒	🟩	🟦	░	🟩	🟩	🟩	▒	▒	▒	▒	░
ACT-R 2016	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	░	🟩	▒	🟩	░
SOAR 2016	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	▒	░	▒	🟩	░

🟥 Disagree (or unspecified by theory)
░ Agree but not implemented
▒ Agree but partially implemented
🟩 Agree and implemented
🟦 Agree partially (some key aspects are above architecture), implemented

Table 2. Analysis of Soar, ACT-R, and Sigma with Respect to the Standard Model.

Acknowledgments

The work described in this article was sponsored by the Air Force Office of Scientific Research (under award FA9550-15-1-0157), the US 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

Albus, J. 2002. 4D/RCS: A Reference Model Architecture for Intelligent Unmanned Ground Vehicles. Paper presented at the 16th Annual SPIE International Symposium on Aerospace/Defense Sensing, Simulation, and Controls, Orlando, FL, 1–5 April. doi.org/10.1117/12.474462

Anderson, J. R. 1990. *The Adaptive Character of Thought*. Hillsdale, NJ: Lawrence Erlbaum and Associates.

Anderson, J. R. 2007. *How Can the Human Mind Exist in the Physical Universe?* Oxford, UK: Oxford University Press. doi.org/10.1093/acprof:oso/9780195324259.001.0001

Anderson, J. R., and Schooler, L. J. 1991. Reflections of the Environment in Memory. *Psychological Science* 2(6): 396–408. doi.org/10.1111/j.1467-9280.1991.tb00174.x

Arel, I.; Rose, D.; and Coop, R. 2009. Destin: A Scalable Deep Learning Architecture with Application to High-Dimensional Robust Pattern Recognition. In *Biologically Inspired Cognitive Architectures II: Papers from the 2009 AAAI Fall Symposium,* ed. Alexei V. Samsonovich. Technical Report FS-09-01. Menlo Park, CA: AAAI Press.

Bach, J. 2009. *Principles of Synthetic Intelligence*. Oxford, UK: Oxford University Press.

Bello, P., and Bridewell, W. 2017. There Is No Agency Without Attention. *AI Magazine* 38(4). doi.org/10.1609/aimag.v38i4.2742

Bostrom, N. 2003. Are You Living in a Computer Simulation? *Philosophical Quarterly* 57(211): 243–255. doi.org/10.1111/1467-9213.00309

Burns, G. A. P. C.; Gil, Y.; Villanueva-Rosales, N.; Liu, Y.; Risi, S.; Lehman, J.; Clune, J.; Lebiere, C.; Rosenbloom, P.; Van Harmelen, F.; Hendler, J.; Hitzler, P.; Janowicz, K.; and Swarup, S. 2014. Reports on the 2013 AAAI Fall Symposium Series. *AI Magazine* 35(2): 69–74.

Card, S.; Moran, T.; and Newell, A. (1983). *The Psychology of Human Computer Interaction.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Cox, M. T.; Maynord, M.; Paisner, M.; Perlis, D.; and Oates, T. 2013. The Integration of Cognitive and Metacognitive Processes with Data-driven and Knowledge-rich Structures. Paper presented at the Annual Meeting of the International Association for Computing and Philosophy, College Park, MD, July 15–17.

Eliasmith, C. 2013. *How to Build a Brain*. Oxford: Oxford University Press. doi.org/10.1126/science.1225266

Eliasmith, C.; Stewart, T. C.; Choo, X.; Bekolay, T.; DeWolf, T.; Tang, Y.; and Rasmussen, D. 2012. A Large-Scale Model of the Functioning Brain. *Science* 338(6111): 1202–1205.

Forbus, K. D., and Hinrichs, T. R. 2006. Companion Cognitive Systems: A Step Toward Human-Level AI. *AI Magazine* 27(2): 83–95. doi.org/10.1609/aimag.v27i2.1882

Forbus, K. D., and Hinrichs, T. R. 2017. Analogy and Relational Representations in the Companion Cognitive Architecture. *AI Magazine* 38(4). doi.org/10.1609/aimag.v38i4.2743

Franklin, S., and Patterson, F. G., Jr. 2006. The LIDA Architecture: Adding New Modes of Learning to an Intelligent, Autonomous, Software Agent. *Integrated Design and Process Technology*. San Diego, CA: Society for Design and Process Science.

Gobet, F., and Lane, P. C. 2010. The CHREST Architecture of Cognition: The Role of Perception in General Intelligence. In *Proceedings of the Third Conference on Artificial General Intelligence,* Lugano, Italy. Amsterdam, The Netherlands: Atlantis Press. doi.org/10.2991/agi.2010.20

Goertzel, B.; Pennachin, C.; and Geisweiller, N. 2014a. *Engineering General Intelligence, Part 1: A Path to Advanced AGI via Embodied Learning and Cognitive Synergy,* Paris. Amsterdam, The Netherlands: Atlantis Press.

Goertzel, B.; Pennachin, C.; and Geisweiller, N. 2014b. *Engineering General Intelligence, Part 2: The CogPrime Architecture for Integrative, Embodied AGI,* Paris. Amsterdam, The Netherlands: Atlantis 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): 197–218. doi.org/10.1080/09528130802319128

Kelly, M. A.; Kwock, K.; and West, R. L. 2015. Holographic Declarative Memory and the Fan Effect: A Test Case for a New Memory Module for ACT-R. Paper presented at the 2015 International Conference on Cognitive Modeling, Groningen, The Netherlands, April 9–11.

Kennedy, W. G.; Bugajska, M. D.; Marge, M.; Fransen, B. R.; Adams, W.; Perzanowski, D.; Schultz, A. C.; and Trafton, J. G. 2007. Spatial Representation and Reasoning for Human-Robot Interaction. In *Proceedings of the Twenty-Second Conference on Artificial Intelligence,* Vancouver, Canada, 1554–1559. Palo Alto, CA: AAAI Press.

Kieras, D. E., and Meyer, D. E. 1997. An Overview of the EPIC Architecture for Cognition and Performance with Application to Human-Computer Interaction. *Human-Computer Interaction* 12(4): 391–438. doi.org/10.1207/s15327051hci1204_4

Koller, D., and Friedman, N. 2009. *Probabilistic Graphical Models: Principles and Techniques.* Cambridge, MA: The MIT Press.

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

Laird, J. E., and Rosenbloom, P. S. 1990. Integrating Execution, Planning, and Learning in Soar for External Environments. In *Proceedings of the Eighth National Conference of Artificial Intelligence,* Boston, MA, 1022–1029. Menlo Park, CA: AAAI Press.

Laird, J. E.; Kinkade, K. R.; Mohan, S.; and Xu, J. Z. 2012. Cognitive Robotics Using the Soar Cognitive Architecture. In *Cognitive Robotics: Papers from the 2012 AAAI Workshop,* ed. H. Palacios, P. Haslum, and J. Baier. Technical Report WS-12-12. Palo Alto, CA: AAAI Press.

Lakatos, I. 1970. Falsification and the Methodology of Scientific Research Programmes. In *Criticism and the Growth of Knowledge,* ed. I. Lakatos and A. Musgrave. Cambridge, UK: Cambridge University Press. doi.org/10.1017/CBO9781139171434 .009

Langley, P., and Choi, D. 2006. A Unified Cognitive Architecture for Physical Agents. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence,* Boston, MA. Menlo Park, CA: AAAI Press.

Langley, P.; Laird, J. E.; and Rogers, S. 2009. Cognitive Architectures: Research Issues and Challenges. *Cognitive Systems Research* 10(2): 141–160. doi.org/10.1016/j.cogsys.2006.07 .004

McShane, M. 2017. Natural Language Understanding (NLU, not NLP) in Cognitive Systems. *AI Magazine* 38(4). doi.org/10.1609 /aimag.v38i4.2745

Newell A. 1973. You Can't Play 20 Questions with Nature and Win: Projective Comments on the Papers of This Symposium. In *Visual Information Processing,* ed. W. G. Chase. New York: Academic Press. 283–310.

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

Newell, A., and Simon, H. A. 1976. Computer Science as Empirical Inquiry: Symbols and Search. *Communications of the ACM* 19(3): 113–126. doi.org/10.1145/ 360018. 360022

Newell, A.; Rosenbloom, P. S.; and Laird, J. E., 1989. Symbolic Architectures for Cognition. In *Foundations of Cognitive Science,* ed. M. Posner. Cambridge, MA: The MIT Press.

O'Reilly, R. C.; Hazy, T. E.; and Herd, S. A. 2016. The Leabra Cognitive Architecture: How to Play 20 Principles with Nature and Win! In *Oxford Handbook of Cognitive Science,* ed. S. Chipman. Oxford: Oxford University Press.

Rosenbloom, P. S. 2014. Deconstructing Episodic Learning and Memory in Sigma. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society,* Québec City, Québec, Canada, 1317–1322. Austin, TX: Cognitive Science Society Inc.

Rosenbloom, P. S.; Demski, A.; and Ustun, V. 2016a. The Sigma Cognitive Architecture and System: Towards Functionally Elegant Grand Unification. *Journal of Artificial General Intelligence* 7(1): 1–103. doi.org/10. 1515/jagi-2016-0001

Rosenbloom, P. S.; Demski, A.; and Ustun, V. 2016b. Rethinking Sigma's Graphical Architecture: An Extension to Neural Networks. In *Proceedings of the Ninth Conference on Artificial General Intelligence,* Lecture Notes in Computer Science, volume 9782, New York, NY, ed. B. Steunebrink, P. Wang, B. Goertzel, 84–94. Berlin: Springer. doi.org/10.1007/ 978-3-319-41649-6_9

Sanner, S.; Anderson, J. R.; Lebiere, C.; and Lovett, M. C. 2000. Achieving Efficient and Cognitively Plausible Learning in Backgammon. In *Proceedings of the Seventeenth International Conference on Machine Learning.* San Francisco, CA: Morgan Kaufmann.

Schermerhorn, P.; Kramer, J.; Brick, T.; Anderson, D.; Dingler, A.; and Scheutz, M. 2006. DIARC: A Testbed for Natural Human-Robot Interactions. In *AAAI Mobile Robot Competition and Exhibition: Papers from the 2006 AAAI Workshop,* ed. B. Avanzato, P. Rybski, and J. Forbes. Technical Report WS-06-15. Menlo Park, CA: AAAI Press.

Scheutz, M. 2017. The Case for Explicit Ethical Agents. *AI Magazine* 38(4). doi.org/10. 1609/aimag.v38i4.2746

Simon, H. 1957. A Behavioral Model of Rational Choice. In *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting.* New York: John Wiley and Sons.

Sloman, A. 2001. Varieties of Affect and the CogAff Architecture Schema. Paper presented at the Symposium on Emotion, Cognition, and Affective Computing, York, UK, 21–24 March.

Sun, R. 2016. *Anatomy of the Mind: Exploring Psychological Mechanisms and Processes with the Clarion Cognitive Architecture.* Oxford, UK: Oxford University Press. doi.org/10. 1093/acprof:oso/9780199794553.001.0001

Treisman, A. 1996. The Binding Problem. *Current Opinion in Neurobiology* 6(2): 171–178. doi.org/10.1016/S0959-4388(96) 80070-5

Ustun, V., and Rosenbloom, P. S. 2016. Towards Truly Autonomous Synthetic Characters with the Sigma Cognitive Architecture. In *Integrating Cognitive Architectures into Virtual Character Design,* ed. J. O. Turner, M. Nixon, U. Bernardet, and S. DiPaola, 213–237. Hershey, PA: IGI Global. doi.org/10. 4018/978-1-5225-0454-2.ch008

Vinokurov, Y.; Lebiere, C.; Wyatte, D.; Herd, S.; and O'Reilly, R. 2012. Unsupervised Learning in Hybrid Cognitive Architectures. In *Neural-Symbolic Learning and Reasoning: Papers from the 2012 AAAI Workshop,* ed. A. d'Avila Garcez, P. Hitzler, and L. C. Lamb. AAAI Technical Report WS-12-11. Palo Alto, CA: AAAI Press.

John E. Laird is the John L. Tishman Professor of Engineering at the University of Michigan. He is one of the original developers of the Soar architecture and leads its continued evolution. He is a founder and chairman of the board of Soar Technology, Inc., and a Fellow of AAAI, AAAS, ACM, and the Cognitive Science Society.

Christian Lebiere is in the research faculty in the Psychology Department at Carnegie Mellon University. He is one of the original developers of the ACT-R cognitive architecture and is coauthor with John R. Anderson of The Atomic Components of Thought. He is a founding member of the Biologically Inspired Cognitive Architectures Society.

Paul S. Rosenbloom is a professor of computer science at the University of Southern California and director for cognitive architecture research at USC's Institute for Creative Technologies. He is one of the original developers of the Soar architecture and the primary developer of the Sigma architecture. He is the author of *On Computing: The Fourth Great Scientific Domain* and a Fellow of both AAAI and the Cognitive Science Society.