Overview

Unified theories of cognition

Michael D. Byrne*

Unified theories of cognition (UTCs) offer an alternative to the modal 'divide and conquer' methodology within cognitive science and attempt to address the full range of cognitive activity within a single theoretical framework. These theories, also termed 'cognitive architectures' are generally computational in nature and are intended to model, at some degree of fidelity, human cognition in a broad range of tasks. This style of research has numerous advantages, not the least of which being that the actual human cognitive system is itself an integrated system and many important tasks require bringing integrated capabilities to bear. There are also drawbacks, particularly dealing with the incompleteness of the knowledge base in cognitive science and the difficulty of evaluating such theories. Three architectures are profiled, each one representing a different 'home' discipline: from AI, Soar; from cognitive psychology, Adaptive Control of Thought-Rational; and from neuroscience, Leabra. Future directions for UTCs include expansion into branches of cognition not already well represented, such as spatial cognition, and increasing attention to cognitive moderators such as emotion and fatigue. Overall, this is a powerful research strategy that is likely to remain an important part of cognitive science for the foreseeable future.

How to cite this article:
WIREs Cogn Sci 2012, 3:431–438. doi: 10.1002/wcs.1180

> The theory that can absorb the greatest number of facts, and persist in doing so, generation after generation, through all changes of opinion and detail, is the one that must rule all observation.
>
> –Adam Smith

INTRODUCTION

The term 'unified theories of cognition' (UTCs) comes from Newell's (1990) book of the same title.[1] He defines UTC as a theory that gains its 'power by positing a single system of mechanisms that operate together to produce the full range of human cognition'. (p. 1). This definition is almost painfully short, given the scope of what is actually proposed. The modal style of research in cognitive science over the last 40 years has been and is still in research within a single, fairly narrow domain such as attentional blink, implicit memory, logic, visual search, and so on. Most theories in cognitive science are rooted in a divide-and-conquer strategy that tends to generate highly specific theories of a very limited range of phenomena. The notion of a UTC works directly at odds with this style of research, being explicitly concerned with breadth of coverage—the 'full range'—rather than a specific domain. The domain, instead, is *all* of cognition.

UTCs are also termed 'cognitive architectures', in part because of Anderson's (1983) *The Architecture of Cognition*,[2] and the two terms will be used interchangeably here. Young and colleagues[3,4] define a cognitive architecture as an embodiment of 'a scientific hypothesis about those aspects of human cognition that are relatively constant over time and relatively independent of task'. That is, it is a theory of cognition that is both integrative and task-independent. Instead of asking 'how can we describe this isolated phenomenon?' people working with cognitive architectures can ask 'how does this phenomenon fit in with what we already know about other aspects of cognition?'

WHAT ARE UTCs?

Anderson (2007)[5] uses this definition: 'A cognitive architecture is a specification of the structure of the

*Correspondence to: byrne@rice.edu

Departments of Psychology and Computer Science, Rice University, Houston, USA

brain at a level of abstraction that explains how it achieves the function of the mind'. Note that this definition makes explicit reference to the physical implementation of the mind, the human brain. Generally speaking, cognitive architectures have not taken their inspiration from neural or neuroscientific approaches but have rather approached the problem as more of an analog to understanding computer architecture: what are the functional components and how do they fit together? However, as neuroscience advances, its influence is being felt in cognitive architectures as well.

Another important feature of UTCs is that they specify only the human 'virtual machine', the fixed architecture. A UTC alone typically cannot do anything. Generally, the architecture has to be supplied with the knowledge needed to perform a particular task. The combination of a UTC and a particular set of knowledge is generally referred to as a computational cognitive model (or just a 'model' to shorten the current presentation). It is often possible to construct more than one model for any particular task using a UTC, representing different strategies available to the person or people being modeled. The specific knowledge incorporated into a particular model is determined by the modeler. Because the relevant knowledge must be supplied to the architecture, the knowledge engineering task facing modelers attempting to model performance on complex tasks can be formidable.

UTCs are generally distinct from engineering approaches to artificial intelligence, which strive to construct intelligent computer systems by whatever technologies that best serve the purpose. Instead, UTCs were originally intended to simulate human intelligence in a human-like way.[1] However, this distinction is now somewhat less clear. Some cognitive architectures are tied closely to psychological data whereas others are more AI-oriented and exploratory in terms of the mappings between their mechanisms and human cognition.

A critical concern for any UTC is obviously breadth. Different theories have had somewhat different ideas about what it means to approach the 'full range' of human cognition. For instance, early versions of the Adaptive Control of Thought-Rational (ACT-R)[6] and Soar[1] architectures were concerned exclusively with 'in-the-head' cognition, and not really the full range of human cognitive behavior; they were primarily aimed at domains like memory and problem solving. More recently, architectures like executive process-interactive control (EPIC)[7] and later versions of ACT-R[8] have attempted to broaden even further and have taken seriously the idea of end-to-end modeling and included both perceptual and motor components as well, in part on the argument that cognition is so intertwined with these things that it is impossible to meaningfully compartmentalize cognition from perception and action.

Finally, UTCs are computational. That is, they are running, implemented pieces of software. Generally, a model of a task constructed in a cognitive architecture is runnable and produces a sequence of behaviors. These behavior sequences can be compared with the sequences produced by human subjects to help assess the quality of a particular model. Many UTCs produce not just behavior streams, but timestamped streams of behavior, which provides another index for comparison with times generated by humans performing the same task. Being running programs, cognitive architectures guarantee internal consistency within the theory; many verbally specified theories in cognitive science have no such assurance. Being implemented as simulation models is particularly important to UTCs because as theories grow in size and number of mechanisms, the interactions of those mechanisms becomes increasingly difficult to predict analytically. Computer simulations permit relatively rapid evaluations of complex mechanisms and their interactions. (For an excellent discussion of this topic, see Ref 9.)

So, a UTC is a computationally implemented broad and inclusive theory of cognition intended to be applicable across multiple tasks, generally aimed at modeling human cognition.

ADVANTAGES AND DISADVANTAGES OF UTCs

UTCs do not represent the modal research strategy for cognitive scientists. What is it that draws those who do so to the endeavor? Similarly, why are they not more popular? The most straightforward way to address these questions is to look at the strengths and weaknesses of pursuing UTCs as a research approach.

First, the most obvious reason that people develop UTCs is that human cognition is generated by a unified, integrated system. People use the same sets of mechanisms for memory, attention, reasoning, language, problem solving, skill acquisition, and so on. Because the mechanisms of human cognition do not exist in a vacuum—instead, they exist as part of an integrated system—there is a great deal of obvious face validity in studying them and trying to replicate or simulate them as part of a unified system. Even if, as neuroscience tells us, the brain has many relatively specialized areas, all of these areas function as part of a coherent whole. People do not have stand-alone memory systems. If we know the human cognitive

systems works as a complete system, why not study it in the way that it actually works?

Single-task theories are inherently limited. They do not cumulate knowledge across the field, they do not address concerns of how high-level capabilities are achieved, and many of the benchmark tasks upon which single-task theories are based may not reflect the modal operation even if the cognitive components are intended to cover. Cognitive scientists could run thousands of simple experiments or build thousands of single-domain systems without making meaningful progress in understanding how people actually perform what they are expected to do. Newell (1973)[10] recognized this problem when he prophesied that if something was not done, cognitive science would remain a fragmented and immature field. He argued in Newell (1990)[1] that enough data had been collected and enough smaller models built that the field must move on to larger and more encompassing theories if it was ever going to become a mature science at the level of physics or chemistry.

There are other motivations for UTCs as well. One of them is that some people in cognitive science—though again, a minority—are concerned about applying cognitive science to so-called "real world" problems. Consider an example of the modern jetliner pilot. The task faced by the pilot is complex, safety-critical, and takes many years of intensive training to master. The environment in which this task is performed is staggeringly complex, visually rich, places extreme time demands on the flight crew, and is partially managed by a fairly opaque piece of automation. This is not only a challenging domain for the pilots but also for human factors engineers who have the task of trying to make the pilot's job easier and safer. Those human factors engineers would certainly be well-served by a theory that could scale-up to tasks as complex and integrated as the ones performed by actual pilots. There are other applications besides human factors, of course, such as tutoring and instruction, that would be well-served by a good unified theory. No isolated theory of memory or attention or logic can possibly hope to handle such a requirement.

On the other hand, not everyone is interested in these kinds of situations. For those who are not, despite the fact that the human system is a unified whole, the problem often seems too daunting to be taken all at once. The human cognitive/perceptual/motor system is obviously incredibly complex; perhaps, it is the most complex system humans have ever tried to understand. The problem of understanding it all is so vast, it seems intractable taken all at once. Specialization happens in all sciences by necessity, and the science of cognition is no exception. It is legitimately impossible to be an expert in every aspect of human performance.

Furthermore, while the empirical and theoretical literature in cognitive science is vast, it still does not cover everything. Any theorist attempting to construct a UTC will encounter areas of the theory where there is little principled guidance for how to proceed, that is, how certain parts of the theory should be constructed. Because working with a unified theory causes researchers to ask questions that have not been asked previously, until there are more people working on unified theories, these questions will continue to go unexplored.

Another issue with working on UTCs is that it is not always clear as to what the appropriate method for moving forward is. Given the scope of the problem, how should one go about generating a cognitive architecture? Attempt to invent the entire theory alone? Borrow other smaller theories, adapt them, and try to integrate them? A similar problem comes when dealing with the issue of expanding or altering a UTC. By necessity, all UTCs are incomplete; we do not yet actually have any full-range theories yet. When a UTC researcher runs into a domain or a problem that cannot be modeled with the current theory, what is the correct course of action? Modify the current architecture, or extend it? Or perhaps start over entirely? There is no simple answer to this question.

Related to this is the problem of evaluation; how can one evaluate a UTC? Because these systems are so large, it is extremely difficult to directly falsify an entire UTC; it can even be difficult to falsify the components. Cooper (2007)[11] proposed that we should not try to evaluate cognitive architectures on the basis of falsification, but rather on Lakatosian criteria. In this view, theories are not monolithic and straightforwardly falsifiable, but instead consist of a mix of central assumptions and peripheral hypotheses. A failure of prediction does not necessarily invalidate a theory, but can cause reevaluation of the peripheral hypotheses. Under this view, the focus should not be on falsification, but rather cumulation of knowledge.

Other approaches have been proposed, such as looking at the complete theory and evaluating it on multiple criteria. There are older versions of this approach[12] as well as more recent ones, such as the 'Newell test'.[13] Despite these efforts, there is as yet, no standard method for evaluating a UTC.

Furthermore, there is also a concrete cost to working within a cognitive architecture. These systems are large and complex and there is a significant learning curve to become an expert with any one of them. This is hardly surprising; the human mind is

no simple system itself, so anything which purports to model the whole system is bound to be complex in its own right. However, this difficulty—and it is a difficulty—is probably more overstated than necessary. It is typically possible to learn the fundamentals of a UTC in a one-day tutorial and basic proficiency is possible as a part of a single-semester undergraduate course. Constructing a detailed model of a highly sophisticated task is still a considerable effort requiring definite expertise, but it is not unattainable.

Finally, there are in-principle criticisms of the UTC endeavor in its entirety. For example, some critics take issue with the idea that any form of computational system can adequately capture human cognition, particularly subjective aspects like consciousness.[14,15] Other critics have argued on other axes, for example, that the non-falsifiability of architectures makes them fundamentally unscientific[16] or simply unsuitable as predictive entities.[17] Still others, particularly Roberts and Pashler (2000),[18] have argued that model-fitting is an inappropriate methodology. This impugns much of the UTC effort, as a substantial proportion of the research on cognitive architectures involves model-fitting.

Despite these difficulties, cognitive architectures have been highly successful in cognitive science in terms of both generation of publishable research and theoretical impact. To make the discussion more concrete, the next section will briefly profile three leading cognitive architectures.

PROFILES OF COGNITIVE ARCHITECTURES

The purpose of this section is to provide the reader with an overview of some of the cognitive architectures currently available. Because of space limitations, only three architectures will be discussed in any length. The three to be covered in slightly more detail represent the three different approaches to UTCs: a computational approach, more from the AI side of cognitive science; a psychological approach, obviously with its roots in psychological theory; and a newer approach, more biological or neuroscientific in nature.

The AI Approach: Soar

Soar has the distinction of being the first system explicitly dubbed as UTC, as it is the candidate architecture put forth by Newell[1] when he coined the term. Soar is now most closely associated with John Laird, one of its original developers.[19] Soar has been used to model a wide variety of human cognitive activity from syllogistic reasoning[20] to flying combat aircraft in simulated wargames against actual human opponents[21] and many tasks in between.

Soar is a production system, meaning it is driven by two kinds of symbolic memory: production rule memory and working memory. Production rules are IF-THEN condition-action pairings; when a certain condition is matched in working memory, which represents the current state of both the world and the system, then an action is taken. This was the basis for many of the 'expert systems' of the 1980s. While Soar does indeed contain these types of structures, it uses them in a more complex way than was typical of prior expert systems.

To understand this, it is useful to think of Soar at a more abstract level. The Problem Space Principle is the guiding principle behind the design of Soar. Thus, Soar casts all cognitive activity as occurring in a problem space, which consists of a number of states. States are transformed through the application of operators. Consider Soar playing a simple game like tic-tac-toe as player 'X'. The problem space is the set of all the states of the tic-tac-toe board—not a very large space. The operators available at any given state of that space are placing an X at any of the available open spaces on the board. Obviously, this is a simplified example; the problem space and the available operators for flying an F-16 in a simulated wargame are radically more complex.

Other than the ubiquitous application of the problem space principle, Soar's most defining characteristics come from two mechanisms developed specifically in Soar, universal subgoaling and a general-purpose learning mechanism. When progressing through a problem space, Soar can reach a state where the next action is not clear, termed an *impasse*. Perhaps the system does not know any acceptable operators for the current state, or perhaps the system lacks the knowledge of how to apply the best operator. Rather than halting or entering some kind of failure state, Soar sets up a new state in a new problem space with the goal of resolving the impasse. For example, if multiple operators were proposed, the goal of the new problem space is to choose between the proposed operators. In the course of resolving one impasse, Soar may encounter another impasse and create another new problem space, and so on. As long as the system is provided with some fairly generic knowledge about resolving degenerate cases (e.g., if all else fails, choose randomly between the two good operators), this universal subgoaling allows Soar to continue even in cases where there is little knowledge.

Learning in Soar is a by-product of universal subgoaling. Whenever an impasse is resolved, Soar

creates a new production rule. This rule summarizes the processing that went on in the substate. The resolution of an impasse makes a change to the superstate (the state in which the impasse originally occurred), and this change is called a result. This result becomes the action, or THEN, side of the new production. The condition, or IF, side of the production is generated through a dependency analysis by looking at any declarative memory item matched in the course of determining this result. When Soar learns, it learns only new production rules, and it only learns as the result of resolving an impasse. It is important to realize that an impasse is not equated with failure or an inability to proceed in the problem-solving, but may arise simply because, for example, there are multiple good actions to take and Soar has to choose one of them. Soar impasses regularly when problem-solving and thus learning is pervasive in Soar.

For most of its history, Soar consisted of only two entirely symbolic memory systems, working memory and production memory, but more recent proposals[22] have included mechanisms for imagery, reinforcement learning, and more. While Soar has not always been closely tied to human data, Soar has always been inspired by and attempted to achieve human-like performance.

There are other AI-oriented UTCs in various stages of maturity, such as CLARION,[23] ICARUS,[24] and PolyScheme.[25] Each of these has a somewhat different focus, but they all attempt to be broad theories capable of a wide range of behaviors, applying a mixture of different AI techniques.

The Psychological Approach: ACT-R

While Soar was the first architecture to be put forth as a UTC, ACT-R is in some sense the oldest, as its roots go back to the 1970s when it was primarily a theory of memory called Human Associative Memory (HAM;[26]). A precursor to ACT-R that was highly influential in the 1980s, though not a full architecture in the sense that it was never completely implemented, was ACT*.[2] ACT-R was introduced in the early 1990s[6] and has gone through multiple revisions, the most recent version termed ACT-R 6.0.[5] ACT-R is primarily intended as a psychological model, and a hallmark of ACT-R research is fitting ACT-R models to data collected from human subjects. This both serves as a method of validating ACT-R models as well as providing directions for future research; data that cannot be modeled often inspire changes to the architecture.

ACT-R is arguably the most successful UTC in cognitive science. It has a sizable research community and literally hundreds of published research papers are based on ACT-R. Like Soar, ACT-R is at its heart a production system, operating with IF-THEN rules. However, ACT-R has many other components (termed 'modules') and the symbolic aspects of ACT-R are augmented with numerical values (termed 'subsymbolic') that control strength and accessibility of those symbolic structures. ACT-R also learns. Where Soar has historically had only one learning mechanism (production learning), ACT-R has four: learning of both new productions and declarative memory elements, and adjustment of the subsymbolic quantities for both types of memory.

ACT-R has been used to model an impressive variety of tasks, from the most straightforward list memory experiments[27] to complex dynamic tasks such as driving a car while dialing a cellular telephone[28] and many other domains in between; see http://act-r.psy.cmu.edu for a nearly complete listing of publications. One of the more recent features of ACT-R is that most ACT-R modules have now been associated with brain regions and ACT-R models have been used to predict the bloodflow response from human subjects in those brain regions with moderate success.[5]

Other psychologically oriented UTCs include EPIC,[7] CHREST,[29] and QN-MHP.[30] Like ACT-R, these architectures are generally focussed on matching human data, but differ on one or more critical axes, such as underlying assumptions, phenomena of core interest, contact with neuroscientific findings, intended breadth of coverage, and so on. While a comprehensive review of the psychological architectures is beyond the scope of the current presentation, it is useful to consider an illustration of some of the differences. For example, EPIC is more focussed on perceptual-motor performance and thus has a more elaborate visual system than ACT-R, but it does not include any learning mechanisms or theoretical commitments on the workings of long-term declarative memory. It also differs in certain core theoretical assumptions. The most critical of these is that ACT-R stipulates that only one production rule may fire in each production cycle, while in EPIC, there is no upper bound on the number of rules that can fire. This has important performance implications in some domains, in particular those that involve multi-tasking. So, while ACT-R currently occupies the dominant position in this space, there are viable alternatives.

The Biological Approach: Leabra

Connectionism is not an area typically associated with UTCs. While there have been a wide variety of connectionist models in many domains of cognitive

science, they have almost universally been models of fairly constrained phenomena. Generally speaking, connectionist models have not attempted the breadth typical of the UTC enterprise.

However, there is an exception to this, the Leabra system.[31,32] While still not as full-featured and broad as many of the other architectures, this one distinguishes itself in that its focus is neuroscientific. Leabra is an attempt to synthesize findings from various areas of neuroscience and implement them in a set of interconnected connectionist layers. This system makes use of multiple kinds of learning found in connectionist systems including Hebbian learning, error-driven learning (similar in some ways to backpropagation, but more biologically plausible), and reinforcement learning. The primary brain regions modeled are the frontal cortex, which performs active maintenance of representations, the hippocampus, responsible for episodic memory, the posterior cortex, which provides sensory representations, and the basal ganglia, responsible for action selection. Each of these regions is implemented as a separate set of recurrent networks, each one optimized for a different kind of learning and different forms of representation.

Leabra has been used to model a variety of tasks in visual recognition, language processing, cognitive control, and memory, including learning in all of these domains. Considering that this approach is somewhat younger than the more traditionally symbolic efforts, this is a compelling accomplishment and is so far the exception rather than the rule in the connectionist literature. Whether or not such systems will be able to scale-up to more complex problem solving, reasoning, and decision making is not yet known, but there is not necessarily any in-principle reason to believe that they cannot.

Interestingly, there has been some work attempting to integrate ACT-R and Leabra in a number of different ways.[33] For example, while ACT-R has a visual system, that system must still be programmed with symbolic representations of the visual scene. Leabra, on the other hand, can 'see' raw bitmaps, so one version of the integration has been to use the visual parts of Leabra in place of ACT-R's usual visual system. This is not unique in the world of cognitive architectures, as other systems also integrate symbolic and connectionist mechanism (e.g., CLARION;[23]), and there have been other attempts to integrate two independently developed architectures (e.g., EPIC-Soar;[34]). The extent to which these and other hybrids will persist in the future is still an open research question and will depend on the extent to which researchers continue to pursue explicit hybridization.

WHAT'S NEXT? THE FUTURE OF UTCs IN COGNITIVE SCIENCE

There is no reason to think that UTCs will lose momentum in the near future. As should be clear even from this brief survey, there are a variety of candidate systems from multiple perspectives, and several of these systems are more than two decades old.

One obvious future direction for UTCs is further inspiration and integration with new findings and theories from neuroscience. Leabra may be joined by other broad connectionist theories, and existing symbolic or symbolic-connectionist hybrids are likely to follow ACT-R's lead and address more neuroscientific data.

Inclusion of perceptual and motor capabilities as components on an equal footing with cognitive capacities is an important step in the development of cognitive architectures, and further expansion of these capabilities is another reasonably likely axis for future development. There is an enormous literature in visual science that has for the most part still not been incorporated into modern UTCs. There are likely to be related developments in other senses as well. For example, there has been recent research into providing a module giving ACT-R the ability to explicitly monitor the passage of time.[35]

Another area of recent interest is the relationship between cognition and emotion. There are now multiple approaches to incorporating emotion in cognitive architectures. Hudlicka (2007)[36] presents a general emotional system that might work with multiple systems, and there has also been recent research on emotions in Soar.[37] This is in some sense going beyond the 'cognitive' label applied to such systems, but that label is increasingly more symbolic than limiting, particularly with the incorporation of perception and action. Emotion is, at the very least, a moderator of cognition. Of particular interest in applied contexts is the effect on cognition and performance of various behavioral moderators such as fatigue and caffeine. The effects of both of those have been investigated in at least ACT-R (for fatigue, see Ref 38; for caffeine, see Ref 39) and possibly other UTCs as well.

One area more traditionally within the bounds of cognition that is also receiving increasing attention is spatial cognition and imagery. This has not been a focus of cognitive architectures in the past, but recent work in Soar[40] and multiple projects based at least in part in ACT-R[41,42] have also taken substantial steps in this direction. This reflects a general interest in situating cognitive architectures in 'real world' contexts, such as using cognitive architectures to control mobile robots.[43]

Of course, there are many areas where UTCs have, for the most part, not yet ventured in any serious way. For example, there have not yet been cognitive architectures that have run and learned on their own for, say, months or years at a time; that is, there are as yet no prominent long-term cognitive models derived from UTCs. Despite their goals, cognitive architectures are still a considerable way away from covering the 'full range' of cognition. While there have certainly been UTC-derived models of natural language processing,[44] this has not traditionally been a strength. Generally speaking, cognitive architectures have had little to say about consciousness, creativity, culture, or numerous other areas of cognition which are studied under the cognitive science banner. This is not to say that they will never get to these areas, but so far progress has been extremely limited in some cognitive domains, and in many of them it is not clear the extent to which the near future holds any promise.

CONCLUSION

While still incomplete, UTC have been both productive and influential in cognitive science and provide a meaningful alternative to piecemeal theorizing that still largely characterizes the field. Despite the fact that a majority of cognitive science research is not conducted within the context of a UTC, this is a strong research area with many architectures staking a claim as candidate unified theory. These architectures bring with them different strengths and perspectives from many areas within cognitive science, and provide a path forward for cumulation of knowledge between traditionally separate subdisciplines.

REFERENCES

1. Newell A. *Unified theories of cognition*. Cambridge, MA: Harvard University Press; 1990.
2. Anderson JR. *The Architecture of Cognition*. Mahwah NJ: Lawrence Erlbaum Associates, Inc.; 1983.
3. Gray WD, Young RM, Kirschenbaum SS. Introduction to this special issue on cognitive architectures and human-computer interaction. *Hum-Comput Interact* 1998, 12:301–309.
4. Ritter FE, Young RM. Embodied models as simulated users: Introduction to this special issue on using cognitive models to improve interface design. *Int J Hum-Comput St* 2001, 55:1–14.
5. Anderson JR. *How can the human mind occur in the physical universe?* New York: Oxford; 2007.
6. Anderson JR. *Rules of the mind*. Hillsdale, NJ: Erlbaum; 1993.
7. Meyer DE, Kieras D E. A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychol Rev* 1997, 104:3–65.
8. Anderson JR, Bothell D, Byrne MD, Douglass S, Lebiere C, Qin Y. An integrated theory of the mind. *Psychol Rev* 2004, 111:1036–1060.
9. Simon HA. *The Sciences of the Artificial*. 3rd ed. Cambridge: Mass MIT Press; 1996.
10. Newell A. You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In: Chase WG, ed. *Visual Information Processing*. New York: Academic Press; 1973, 283–308.
11. Cooper RP. The role of falsification in the development of cognitive architectures: insights from a Lakatosian analysis. *Cognitive Sci* 2007, 31:509–533.
12. Norman DA. Twelve issues for cognitive science. In: Aitkenhead AM, Slack JM, eds. *Issues in Cognitive Modeling*. Hillsdale, NJ: Erlbaum; 1985, 309–336.
13. Anderson JR, Lebiere C. The Newell test for a theory of cognition. *Behav Brain Sci* 2003, 26:587–640.
14. Block N. On a confusion about a function of consciousness. *Behav Brain Sci* 1995, 18:227–287.
15. Bringsjord S. Animals, zombanimals, and the total Turing test: the essence of artificial intelligence. *JoLLI* 2000, 9:397–418.
16. Hunt E, Luce RD. Soar as a world view, not a theory. *Behav Brain Sc* 1992, 15:447–448.
17. Tadepalli P. Cognitive architectures have limited explanatory power. *Behav Brain Sc* 2003, 26:622–623.
18. Roberts S, Pashler H. How persuasive is a good fit? A comment on theory testing. *Psychol Rev* 2000, 107:358–367.
19. Laird JE, Newell A, Rosenbloom PS. Soar: an architecture for general intelligence. *Artif Intell* 1987, 33:1–64.
20. Polk TA, Newell A. Deduction as verbal reasoning. *Psychol Rev* 1995, 102:533–566.
21. Jones RM, Laird JE, Nielsen PE, Coulter K J, Kenny P, Koss FV. Automated intelligent pilots for combat flight simulation. *AI Mag* 1999, 20:27–41.
22. Laird, JE. *Extending the Soar Cognitive Architecture*. Paper presented at the First Conference on Artificial General Intelligence, Memphis, TN: IOS Press; 2008.
23. Sun R. The CLARION cognitive architecture: extending cognitive modeling to social simulation In: Sun R, ed. *Cognition and Multi-Agent Interaction*. New York: Cambridge University Press; 2006.

24. Langley P, Choi D. A unified cognitive architecture for physical agents. *Proceedings of the Twenty-First National Conference on Artificial Intelligence*. Boston: AAAI Press; 2006.
25. Cassimatis NL. A cognitive substrate for human-level intelligence. *AI Mag* 2006, 27:45–56.
26. Anderson JR, Bower G. *Human Associative Memory*. Washington, DC: Winston; 1973.
27. Anderson JR, Bothell D, Lebiere C, Matessa M. An integrated theory of list memory. *J Mem Lang* 1998, 38:341–380.
28. Salvucci DD. Modeling driver behavior in a cognitive architecture. *Hum Factors* 2006, 48:362–380.
29. Gobet F, Lane P. The CHREST architecture of cognition: listening to empirical data. In: Davis D, ed. *Visions of Mind: Architectures for Cognition and Affect*. Hershey, PA: Information Science Publishing; 2005, 204–224.
30. Liu Y, Feyen R, Tsimhoni O. Queueing network-model human processor (QN-MHP): a computational architecture for multitask performance in human-machine systems. *ACM T on Comput-Hum Int* 2006, 13:37–70.
31. O'Reilly RC, Munakata Y. *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. Cambridge, MA: MIT Press; 2000.
32. Atallah HE, Frank MJ, O'Reilly R. Hippocampus, cortex, and basal ganglia: Insights from computational models of complementary learning systems. *Neurobiol Learn Mem* 2004, 82:253–267.
33. Jilk DJ, Lebiere C, O'Reilly R, Anderson JR. SAL: an explicitly pluralistic cognitive architecture. *J Exp Theor Artif In* 2008, 20:197–218.
34. Chong RS, Laird JE. Identifying dual-task executive process knowledge using EPIC-Soar. In: Shafto M, Langley P, eds. *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum; 1997, 107–112.
35. Taatgen NA, Rijn HV, Anderson JR. An integrated theory of prospective time interval estimation: the role of cognition, attention and learning. *Psychol Rev* 2007, 114:577–598.
36. Hudlicka E. Reasons for emotions: modeling emotions in integrated cognitive systems. In: Gray WD, ed. *Integrated Models of Cognitive Systems*. New York: Oxford University Press; 2007, 263–278.
37. Marinier RP, Laird JE. Computational modeling of mood and feeling from emotion. *Proceedings of 29th Meeting of the Cognitive Science Society*. Nashville: Cognitive Science Society; 2007, 461–466.
38. Gunzelmann G, Gross JB, Gluck KA, Dinges DF. Sleep deprivation and sustained attention performance: Integrating mathematical and cognitive modeling. *Cognitive Sci* 2009, 33:880–910.
39. Ritter FE, Reifers A, Klein LC, Quigley K, Schoelles M. Using cognitive modeling to study behavior moderators: Pre-task appraisal and anxiety. In *Proceedings of the Human Factors and Ergonomics Society*. Santa Monica, CA: Human Factors and Ergonomics Society; 2004, 2121–2125.
40. Lathrop SD, Laird JE. Towards incorporating visual imagery into a cognitive architecture, *Eighth International Conference on Cognitive Modeling*, 2007.
41. Lyon DR, Gunzelmann G, Gluck KA. A computational model of spatial visualization capacity. *Cognitive Psychol* 2008, 57:122–152.
42. Harrison AM, Schunn CD. ACT-R/S:Look Ma, no "cognitive map"! In: Detje F, Doerner D, Schaub H, eds. *Proceedings of the Fifth International Conference on Cognitive Modeling*. Bamberg, Germany: Universitats-Verlag Bamberg; 2003, 129–134.
43. Kennedy WG, Bugajska MD, Harrison AM, Trafton JG. "Like-Me" simulation as an effective and cognitively plausible basis for social robotics. *Int J Soc Econ* 2009, 1:181–194.
44. Lewis RL, Vasishth S. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Sc* 2005, 29:375–419.

