Proceedings of the Twelfth Annual Conference on Advances in Cognitive Systems ACS-2025 (83-104)

The STAR Cognitive Cycle in Context: A Comparative Analysis of 44 Cognitive Architectures

Iuliia Kotseruba KOTSERUBA@UOGUELPH.CA
School of Computer Science, University of Guelph

John K. Tsotsos TSOTSOS@YORKU.CA
Department of Electrical Engineering and Computer Science, York University

Abstract

This paper discusses the cognitive cycle of the Selective Tuning Attentive Reference (STAR) architecture and compares it to cognitive cycles of existing cognitive architectures at different levels of abstraction. First, we briefly discuss the purpose of the cognitive cycle and how it can be specified. Second, we propose a Core Model that includes modules and connections found in typical cognitive architectures. Third, we analyze descriptions of cognitive cycles of 43 other cognitive architectures and express them in the Core Model format to highlight existing gaps in research, such as perception, attention, learning from experience, and metacognition. A report on the current status of STAR is then provided. Lastly, we apply this framework to STAR cognitive cycle and discuss how it addresses some of the identified gaps.

1. Introduction

The cognitive architecture STAR (Tsotsos, 2013; Tsotsos & Kruijne, 2014) is a computational probe into the roles of attention and active perception in an embodied 3D real-world visual agent. We aim to build and test a cognitive architecture that models human visual abilities to a level of detail that may yield experimentally falsifiable human predictions (mostly at a behavioural level) and practical machine implementations. In this paper, we will focus on the cognitive cycle of STAR and provide a current description that highlights its differences from other architectures.

A cognitive cycle is a high-level specification of how cognitive architectures operate, describing the computational processes transforming system inputs into outputs. Despite its obvious importance, this concept has not received the attention it deserves. In the literature, most descriptions of cognitive architectures focus on their components rather than connections among them or the overall processing pipeline. Moreover, there is no standardized representation for cognitive cycles in the literature. As a result, both textual and diagrammatic depictions of cognitive cycles across different cognitive architectures vary significantly in the amount of detail and level of abstraction, making direct comparisons difficult.

To address this problem, we propose a Core Model intended as a unified representation of a cognitive cycle at the computational theory level (in Marr's terms). The model represents the cognitive cycle as a computational framework based on the core cognitive abilities, such as perception, memory, learning, reasoning, and motor control. We then summarize the characteristics of cognitive cycles of 43 cognitive architectures using the Core Model representation to reveal understudied abilities in existing cognitive architectures to guide their future development. Next, we provide an overview of the latest version of our STAR architecture. Lastly, we apply the Core Model to STAR and compare its cognitive cycle, representation, and tasks with those of other architectures. In addition, we discuss various sources of novelty in STAR, such as a 'first principles' foundation, an active embodiment, a sophisticated realization of visual attention, and a novel learning paradigm.

2. What Is a Cognitive Cycle?

There are many ways of describing how different living and artificial organisms operate throughout their lifespan. One widely accepted proposal by Newell (1990) is based on the following 4 time scales (bands) of human action: 1) *social band* spanning activities that take days to months, 2) *rational band* for actions lasting from minutes to hours, 3) *cognitive band* from 100 ms to 10 sec, and, 4) *biological band* for neural operations that take 10 ms or less. At each scale, various aspects of cognition are revealed, from continuous development, learning, and decision-making to perception, reflexive actions, and fine-grained motor control.

For intelligent agents the cognitive band is arguably the most important because it encompasses observable behaviors that occur within seconds, i.e. typical interactions with the environment, such as perception, motor actions, utterances, decision-making, etc. At this temporal scale, the basic operation cycle of any living or artificial organism consists of the following steps: perceiving the environment using available sensors, processing the information, and performing some action if necessary. This process received different names in the literature—in neuroscience it is known as the perception-action cycle and in robotics as the sense-plan-act cycle. In cognitive architectures, it is referred to as the cognitive cycle. Simply put, a cognitive cycle defines how inputs into the system result in observable behavior. We will use this term for the remainder of this paper.

Cognitive architectures are complex systems, thus it is useful to consider their operation at different levels of abstraction. Marr's three-level model (Marr, 1982) provides a suitable framework that isolates theoretical, computational, and implementation aspects of information-processing systems. When applied to cognitive architectures, the following needs to be specified at each level:

1. Computational theory: problem description, verbal and/or diagrammatic depiction of the system, its components, connections and information flow among them.
2. Representation and algorithm: concrete definition of task(s) that are required to solve the problem, expected inputs and outputs, data structures suitable for representing them, and algorithms for computing outputs from inputs.
3. Physical implementation: an instance of the system that can perform these tasks to solve a problem. The system can be implemented in a programming language on a hardware device (e.g., computer or robotic platform). At this level, additional language- or hardware-specific details that are not relevant for the earlier steps may be included.

Additional constraints may be imposed at every level to ensure a match with known or hypothesized properties of human cognition and biology. For example, verbal theories of cognition derived from psychophysical studies can guide high-level design of the system (e.g., basic modules and

Table 1: A list of cognitive architectures used for the analysis presented in this paper.

1. **3T** (Bonasso et al., 1997)	16. **COGNET** (Zachary et al., 1998)	31. **IMA** (Kawamura, 2023)
2. **ACT-R** (Anderson et al., 2004)	17. **CogPrime** (Goertzel et al., 2013)	32. **Kismet** (Breazeal, 2003)
3. **ADAPT** (Benjamin et al., 2004)	18. **CoJACK** (Ritter et al., 2012)	33. **LIDA** (Franklin et al., 2016)
4. **AIS** (Hayes-Roth, 1995)	19. **Copycat** (Hofstadter & Mitchell, 1994)	34. **MDB** (Bellas et al., 2010)
5. **ARCADIA** (Bridewell & Bello, 2016)	20. **CORTEX** (Bustos et al., 2019)	35. **MicroPsi** (Bach, 2009)
6. **ART** (Carpenter & Grossberg, 1987)	21. **DAC** (Verschure, 2012)	36. **NARS** (Wang, 2022)
7. **ATLANTIS** (Gat, 1998)	22. **DIARC** (Scheutz et al., 2013)	37. **PRS** (Georgeff & Lansky, 1987)
8. **BBD** (Edelman, 2007)	23. **Disciple** (Tecuci, 1991)	38. **SASE** (Weng, 2002)
9. **BECCA** (Rohrer, 2012)	24. **DUAL** (Kokinov, 1994)	39. **Soar** (Laird, 2022)
10. **CARACaS** (Huntsberger et al., 2011)	25. **EPIC** (Meyer & Kieras, 1997)	40. **SPA** (Eliasmith et al., 2012)
11. **CERA-CRANIUM** (Arrabales et al., 2009)	26. **ERE** (Bresina & Drummond, 1990)	41. **Subsumption** (Brooks, 1986)
12. **CHREST** (Gobet & Lane, 2012)	27. **FORR** (Epstein et al., 2002)	42. **TCA** (Simmons, 1994)
13. **CIRCA** (Musliner et al., 1993)	28. **GLAIR** (Shapiro & Ismail, 2003)	43. **Ymir** (Thórisson, 1998)
14. **Clarion** (Sun, 2007)	29. **HCA** (Haikonen, 2007)	44. **STAR** (Tsotsos & Kruijne, 2014)
15. **Cog** (Brooks et al., 1999)	30. **ICARUS** (Choi & Langley, 2018)	

information flow). At the lower levels, timings, concrete cognitive functions implemented by the algorithm, relation to areas of the brain involved in the computation, etc. can be introduced. These constraints determine computational complexity of the cognitive cycle in terms of time and memory resources required Tsotsos (2017), and help narrow down choices of representations and algorithms. Concrete specifications are also essential for designing tasks, scenarios, and metrics for qualitative and quantitative evaluation.

3. Cognitive Cycles of Existing Architectures

The remainder of the paper analyzes a diverse set of 44 architectures listed in Table 1. These architectures were selected out of the set of 84 surveyed in our past work (Kotseruba & Tsotsos, 2020, 2025) because they offered detailed diagrams and descriptions that allowed for our analysis.

We start by summarizing cognitive cycles of the 43 cognitive architectures at the first two levels of abstraction. The 44th architecture, STAR, will be discussed later. Although all of these architectures were implemented in some form, the exact technical specifications are not available for many, thus we omit the discussion of the physical implementation level.

3.1 Representations and Algorithms Level

Recalling the previous section, three aspects of cognitive architectures are defined at this level: representations, algorithms, and tasks. Nearly all cognitive architectures are an eclectic mix of representations and algorithms that are difficult to categorize due to the following reasons. First, specification at this level is often incomplete. Second, specifications are not expressed at the same level of detail and using the same terminology. Third, there are no established taxonomies that could cover the large variety of representations and algorithms found in different architectures.

In any case, the majority of architectures do not take a strong stance on representation, that is, they do not assert that specific formalisms are necessary or sufficient. Even when computational parsimony is declared as a goal, additional representations are often introduced for practical reasons. For instance, ACT-R and Soar are architectures whose aim is finding a minimal set of predominantly symbolic computational approaches that enable human cognition, however, both have been interfaced with a wide variety of subsymbolic representations, such as neural networks. On the other end of the spectrum are cognitive architectures that allow any representation as long as it performs the required function, e.g. Polyscheme (Cassimatis et al., 2009).

Another aspect of architectures defined at this level of abstraction is the set of tasks they perform. Here, descriptions are also often imprecise, meaning that the acceptable inputs and desired inputs are not fully specified. An even bigger issue is that the majority of implementations are designed for a single task, i.e. a different instance of cognitive architecture is built each time the task changes. Only few architectures are implemented to support switching between tasks within a single instance (e.g. SPA, Eliasmith et al. (2012)).

In terms of task variety, cognitive architectures taken as a group possess many abilities ranging from solving abstract reasoning problems, categorization and pattern matching to real-world tasks that involve reasoning and motor skills. In our past works (Kotseruba & Tsotsos, 2020, 2025), we identified hundreds of unique tasks but found little overlap among them, even in relatively niche areas (e.g. analogical reasoning). This adds to the difficulties with comparing different representations and assessing their benefits and limitations.

3.2 Computational Theory Level

To analyze cognitive cycles at the computational theory level, we consider both theories of cognition and structure of the cognitive architectures.

3.2.1 High-Level Theories

Most of the cognitive architectures in our list were inspired by one or more high-level theories of cognition. This inspiration may take a direct form, i.e., the architecture is built to test a specific theory of cognition. Some examples of this are ACT-R based on the ACT theory (Ritter et al., 2019), CHREST and Chunking Theory (Gobet et al., 2001), BBD and Neural Darwinism, ART and Adaptive Resonance Theory (Grossberg, 2021), and STAR implementing the Selective Tuning model of attention (Tsotsos et al., 1995). Sometimes, multiple theories are combined. For instance, Clarion (Sun, 2020) merges elements of Soar (division between procedural and declarative memory) with distributed representations. Ymir's design is based on blackboard systems (Engelmore, 1988), schema theory (Arbib, 1981), and behavior networks (Maes, 1991). Other architectures contain a mix of indirect influences, such as folk-psychological theories or concepts, as well as representations and algorithm ideas from other architectures, without explicit commitment to any specific theories. Lastly, some architectures incorporate existing implementations for specific functions, e.g., ADAPT (Benjamin et al., 2004) delegates decision-making to an instance of Soar.

Theories themselves differ greatly in the level of detail and amount of empirical support they have received. Some contain only general guidelines but do not describe the computation or components explicitly. For example, Newell's Unified Theories of Cognition (Newell, 1990) that inspired many architectures is built on findings from cognitive science but focuses mainly on the high-level desiderata (e.g. symbolic computation, parsimony, rationality). Minsky's Society of Mind (Minsky, 1986) postulates that cognition is enabled by a constellation of interacting heterogeneous agents

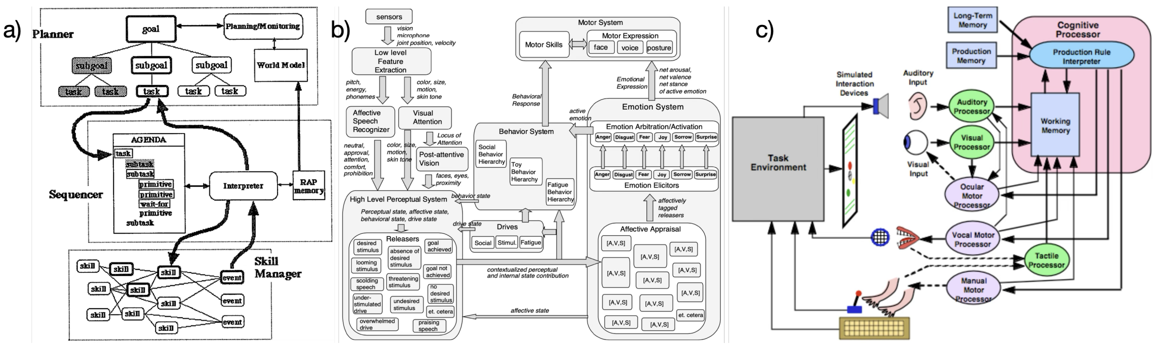

Figure 1: Diagrams of several cognitive architectures: a) 3T (Bonasso et al., 1997), b) Kismet (Breazeal, 2003), c) EPIC (Kieras et al., 2016). Note the differences in graphical representations and naming of the modules, as well as varying granularity at which components are depicted.

but neither refers to human data nor provides a concrete computational framework to build upon. Other theories, such as Global Workspace Theory (Baars, 1988), Theory of Cerebellar Function (Albus, 1971), and Selective Tuning (Tsotsos et al., 1995) are grounded in cognitive science and provide more concrete descriptions suitable for implementation and testing. Blackboard architectures (Engelmore, 1988), BDI (Rao & Georgeff, 1995), and behavior-based robotics (Brooks, 1986) are examples of formalized computational frameworks that were motivated by some aspects of human or animal cognition but even more so by the need to solve real-world engineering problems.

There are no theories of cognition that provide a complete description of all known cognitive processes. Thus, in any given architecture only some components have a theoretical foundation, while the rest are either omitted or hypothesized. The latter may lead to filling the gaps in theory, which is one of the main goals of building the cognitive architectures in the first place. Another consequence of the incomplete theories is that most cognitive architectures focus on investigating specific areas of cognition, e.g., memory or decision-making, rather than all of them.

3.2.2 Core Model of Computation

Since cognitive architectures are often based on vague, incomplete, or overlapping theories, comparisons at this level are difficult. Thus, here we will consider organization and information flow. To do so, we gathered textual and diagrammatic representations of 43 architectures in Table 1 (excluding STAR). However, direct comparisons could not be made due to the following issues:

1. There is no common terminology for naming modules or components of the architectures. For example, modules of the architectures originating from neuroscience (e.g. BBD, Leabra) are often labelled as corresponding brain areas, whereas agent and robotic architectures use a wide variety of terms, e.g. advisors, blackboards, world model, schemas, etc., that do not have direct cognitive or neural correlates. More recent projects that borrow concepts and implementations from multiple past works operate with an even more eclectic mix of terms.
2. The number and granularity of components varies across architectures. For example, memory in some systems is a unified short- and long-term storage, while in others it is divided into

numerous specialized components. Similarly, decision-making and learning may be represented as module(s) or processes, or their functions may be distributed among other components (e.g. reward assignment, motor control, memory).

3. Despite some commonalities in the visual language used in the diagrams (e.g. boxes for modules and arrows for information flow) there is no common level of abstraction across the diagrams. As a result, some architectures are visualized in very broad strokes while others are highly granular and even include implementation details (see Fig. 1). Even within the same diagram, different components may be shown with different levels of detail.
4. Diagrams do not always match the text. Sometimes components or connections are not explained and sometimes modules or processes described in papers do not appear in the diagrams.
5. For many architectures, there is no complete description of the cognitive cycle from inputs to outputs. This is somewhat unexpected given that architectures in our selection are implemented. Therefore specification should exist in some form for all of them.

To compare the architectures, it is necessary to express them in the same terms and visualization format. We propose a Core Model—a representation based on the set of core components we identified in our previous work, which includes modules found in a broad and diverse set of cognitive architectures, such as perception, short-term memory (STM), long-term memory (LTM), decision-making, and motor module, as well as flows of information among the modules (see Figure 2). We applied the Core Model to each system in our set of cognitive architectures by following these steps:

1. Identified structures in the published diagrams that correspond to the modules in our template;
2. Traced inputs and outputs into each module and highlighted them in the template;
3. In our template, removed elements without correspondences in the architecture diagram;
4. Traced flow from inputs to outputs (cognitive cycle) based on diagrams and descriptions.

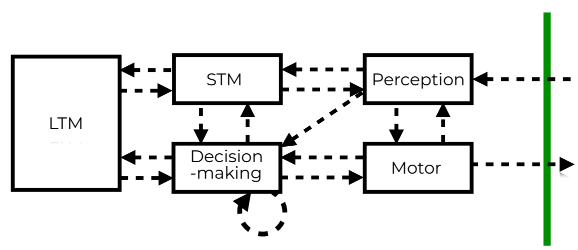

Figure 2: Core Model with modules (black rectangles) and connections (dashed arrows) found in 43 cognitive architectures. STM and LTM stand for short- and long-term memory, respectively. A looped arrow denotes metacognition. The green line separates the agent from the environment.

After completing this process, we combined all diagrams by counting the frequencies of connections (arrows) and identified typical cognitive cycles, i.e., as defined earlier, cycles that specify how inputs into the system result in observable behavior. Figure 3 shows a graphical summary highlighting the most common connections among modules and two most common cycles. A shorter cycle for reactive actions involves only perception and motor modules to ensure the quickest possible responses to stimuli. A longer cycle consists of multiple steps: 1) processing perceptual input, 2) depositing the results into temporary storage (STM), 3) decision-making using information in STM and LTM and self-reflection, and 4) planning and executing motor commands.

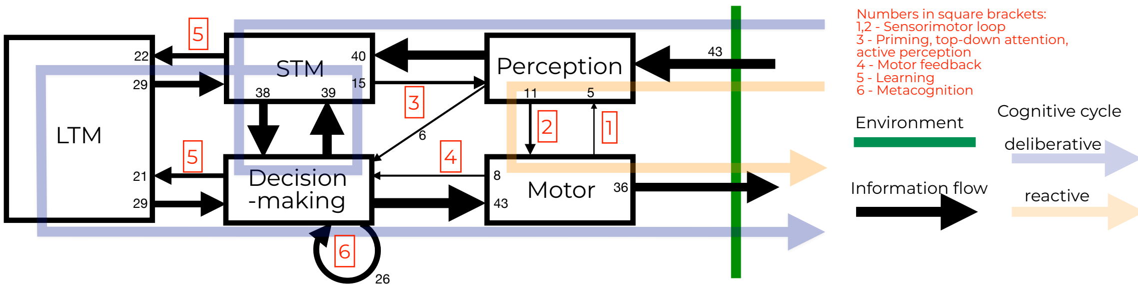

Figure 3: Modules and information flow of 43 cognitive architectures (excluding STAR) expressed in the Core Model format. STM and LTM stand for short- and long-term memory, respectively. Small numbers next to each arrow are counts of architectures that contain said arrow; thicker arrows represent more common connections and vice versa. Arrows labelled with red boxed numbers show under-represented connections and the text will refer to them by that number. Blue and and orange arrows show two typical cognitive cycles: a long one for deliberative actions and a short one for reactive actions, respectively.

The main benefit of expressing all architectures in a Core Model representation is that it provides a visual summary of existing gaps in theory and implementations. A closer look at Fig. 3 shows several areas where improvements are needed:

- Sensorimotor integration (arrows 1 and 2) is a well-known property of biological systems but is conspicuously missing in many architectures. This includes sensory information for motor commands and feedback from the motor system (efference copy).
- Priming, top-down attention, and active perception (arrow 3) are an important part of biological perception, which is recurrent and incorporates internal biases, task, and knowledge. Most implementations of perception are feedforward and bottom-up, meaning that it is single pass and driven mainly by the properties of the scene.
- Motor feedback (arrow 4), similarly to sensorimotor integration, is necessary for decision-making. Current systems rely mostly on perception to assess changes in the environment and struggle to determine which changes were caused by the agent itself.
- Learning (arrow 5) in cognitive architectures is treated mainly as accumulation of declarative and procedural knowledge. Broadly speaking, this is represented by two pathways in the diagram: 1) transfer of perceived observations from short-term to long-term memory (e.g. memorizing instructions or locations of objects in the room) and 2) learning from experience (e.g. prioritizing successful decisions/actions made in the past in similar situations).
- Metacognition (arrow 6) to some extent is found in about half of the cognitive architectures we analyzed, mainly for debugging purposes. Fewer use the information about the internal processes to aid in decision-making and learning.

A common feature of the missing components is that all of them are necessary for solving everyday problems that require perception and navigation in 3D, adaptation to new environments, and learning from experience.

3.3 Core Model and Common Model of Cognition

The Core Model has some similarities with the Common Model of Cognition (CMC)[1] (Laird et al., 2017). Both focus on the high-level organization of the system and flow of information within it. There is also overlap in terms of basic modules out of which the system is composed.

The differences between these two frameworks stem from the strong influence of production systems on CMC. First, CMC is based on three cognitive architectures—ACT-R, Soar, and Sigma, whereas the Core Model is more data-driven and is a result of analyzing dozens of architectures. Second, CMC emphasizes the distinction between procedural and declarative memory, which is a distinguishing feature of production systems. Our representation is more general and combines all types of memory (procedural, declarative, episodic, etc.) within the LTM module. Third, CMC does not include a decision-making component explicitly since in production systems most processing occurs in memory. However, in many architectures decision-making is a separate module, thus we include it in the Core Model.

4. STAR Cognitive Cycle

STAR is a visuospatial, purposefully behaving, mobile agent. The STAR cognitive architecture is not intended as a framework for building new applications that display intelligent behavior. Rather it should be regarded as a computational probe into the roles that active vision and attention play in an embodied intelligent agent in a real 3D world. Here we document the most recent version of STAR .

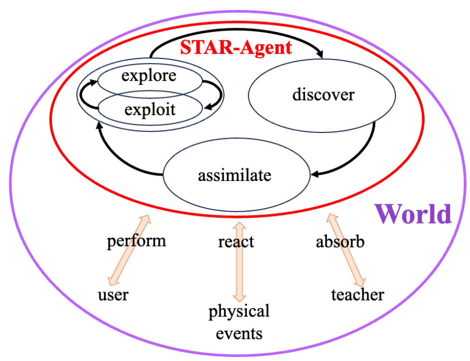

Figure 4: High-level diagram of the STAR cognitive cycle. By default, STAR Agent (SA) explores, discovers and assimilates information about its world. When a user/teacher provides a task/information or event happens, SA performs or reacts and returns to its default state.

4.1 A Brief Introduction to STAR

The history of cognitive architectures provides a strong foundation for STAR as is clear from the review of Kotseruba & Tsotsos (2025). At the computational level, STAR belongs to a small group of architectures that are grounded in a theory of cognition with ample empirical support. STAR is based on a number of strong foundational elements.

First, STAR was founded on the success of the Selective Tuning (ST) model of visual attention, which has predicted many novel aspects of human visual attention, now strongly supported experimentally (summarized in Carrasco (2011); Tsotsos (2011); for 2 recent results see Bartsch et al. (2023); Schulz et al. (2024)). Within ST is a very influential saliency model, AIM (Bruce & Tsotsos, 2005). STAR is envisioned as the body and brain within which ST would operate.

1. It was initially introduced as the Standard Model of the Mind (SMM) by analogy with the Standard Model of particle physics. However, shortly after, SMM has been renamed to CMM.

Secondly, Visual Routines (Ullman, 1983) inspires the use of programs to encode step-by step processes for solving visuospatial problems. They require modernizing and has led to our Cognitive Programs (Tsotsos & Kruijne, 2014).

The third foundational element is the classic and well-proved Means-Ends Analysis problem solving method (Newell et al., 1959). As with Visual Routines, some updating is needed here too because its apparent assumptions regarding its connection to perception are no longer in agreement with current understanding of vision (in ways similar to the same issue with Visual Routines).

Fourth, STAR is all about active perception, active observers that solve visual problems in 3D (Bajcsy et al., 2018). This provides the basic closed-loop perception-action cycle within STAR as is featured in the active perception literature since the mid-1980's.

Fifth, STAR is intended to be an embodied visual agent and thus our past experience with robotic binocular camera systems (Milios et al. 1983, Herpers et al. 1999) roots the design and implementation of a new convergent binocular camera system and robot head with human-like form and performance (submitted).

Finally, the developmental literature has motivated our Active Developmental Bootstrapping learning strategy (Aslin et al., 2023; Gopnik et al., 2017; Siu & Murphy, 2018). This blends active perception, development of visual capabilities, and learning methods in a novel experiential framework.

4.2 A 'First Principles' Description of STAR

The basic STAR cognitive cycle is shown in Figure 4. Without a user task, a STAR Agent (SA) explores, discovers and assimilates knowledge about its world. It reacts to events in the world (reactive behavior), it can accept and absorb knowledge (from a teacher), and it can perform a user task, exploiting knowledge that it already has learned supplemented by exploration as required (deliberative behavior).

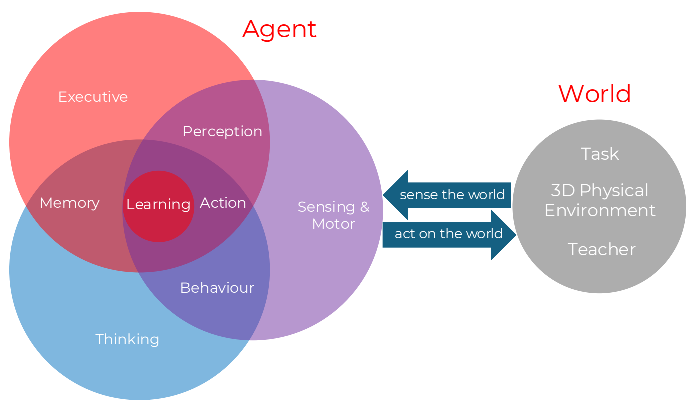

Figure 5: Components of STAR: Executive, Thinking, and Sensing & Motor (functions in white font). Overlapping areas indicate bidirectional flow of control and data.

At a less abstract level of description, Figure 5 depicts SA's four primary interacting components, drawn as colored disks in the figure and labelled Executive, Thinking, Sensing & Motility, and Learning. The intersecting disk regions highlight secondary components that require multidirectional coordination among intersecting components, Memory, Perception, Action and Behavior. The Core Model above has the main components of STM, LTM, Decision-Making, Perception and Motor. This overlaps significantly with ours which is to be expected; however the interactions seen in past architectures seem more limited. We approach the problem of detailing these components in a first principles manner, by rooting to several basic questions whose answers we believe would be relevant to any such agent; however, the replies presented are specific to STAR.

With respect to the Core Model, much of what is presented here conforms to commonalities with past architectures. However, much is also intended to address the weaker aspects denoted by the 6 red numbers of Figure 3 and these will be noted.

What comprises the physical world external to the agent and what roles can it play interacting with the agent's behaviour? The real 3D physical world is represented by the World disk in the figure. It includes not only the physical environment SA lives in, but also a Teacher that provides declarative and procedural knowledge to SA as needed (including advice and feedback while performing a task), and a User that can pose a task for SA to solve. SA can sense the world, can move within the world and can manipulate the world. SA can communicate with the User and Teacher using Imperative English (Kunić, 2017). Sensing & Motility, Perception and Executive handle the processing of these inputs. Currently, we are not considering any dynamic aspects of the physical world, i.e, it is static, and SA's embodiment does not include a manipulator nor other means for world change.

How does the agent interact with the world? SA has a physical embodiment. The Sensing & Motility components of the embodiment physically interact with and sense the world. SA actively interacts with its world by moving in the environment in any manner it wishes, and employing any sensing viewpoint needed in order to solve its task. There are two kinds of tasks: reactive and deliberative. The former is sensed by the Sensing component, recognized by Perception, but connected directly to Action and then Motility bypassing the Executive or other components. Action needs to have table-lookup-like capabilities to decide on reactions quickly, and these can be learned. It is important that the reactive action and any side-effects be recorded because they may change the agent's or world state. Deliberative tasks, on the other hand, require the full set of components in Figure 5.

In order for the STAR Agent to be an active observer, it is necessary to enable eye, head and body movements. To this end, we have designed and constructed a novel binocular robotic camera system (submitted). This has 9 DOF mechanically (including neck motion) and 4 DOF optically, with camera baseline and overall size similar to a human head. It has convergent stereo and we have developed the first algorithm for computing horizontal and vertical disparities from convergent stereo images (forthcoming). We have a new method for controlling this head for saccadic eye movements achieving accuracy similar to human (although a bit slower) (submitted). This head will be mounted on a commercial robot dog platform that would provide motion along the floor in our test environment as well elevation changes for the head. Thus, the STAR embodiment will be a fully active observer that can cover the space of all the observed human eye, head and body motions that we documented experimentally and briefly describe in the next section.

The primary sensor will be the visible light binocular camera system just mentioned. There will also be a form of proprioception via motor feedback from the robots. Although the current state-of-the-art in computer vision seems of high quality and we can build upon it, there are specific aspects missing, primarily, visual attention. Attention in modern systems seems re-imagined and has little to do with actual attention as humans embody it (Mehrani & Tsotsos, 2023). The Selective Tuning visual attention model will be used in STAR. Briefly, ST does more than selection which is classi-

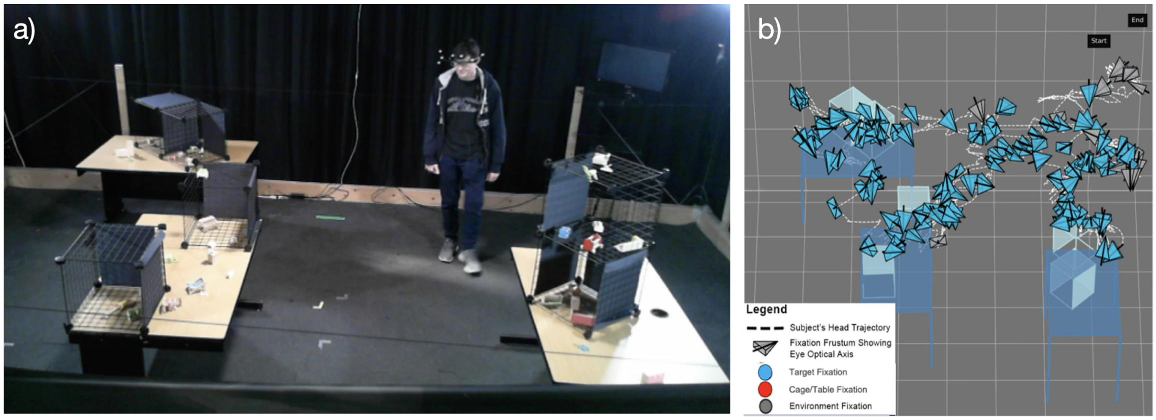

Figure 6: a) The setup and subject for a visual search trial (target-absent) (Wu & Tsotsos, 2025). b) 3D scanpath recorded from this trial. Note the large viewpoint changes and numerous fixations needed to confirm that the target is absent.

cally part of many past cognitive architectures and computer vision systems. ST provides several human attention mechanisms including priming (global bias), cueing (local bias), foveal selection, surround suppression in spatial, feature, and object dimensions (to improve signal-to-noise), binding and localization (via top-down search), peripheral saliency, eye movement control, attentional sample for working memory, and more (Tsotsos, 2011, 2022). The Perception component, directed by the Executive, depends on ST operating within its visual processing hierarchy. This corresponds to arrow 3 in Figure 3.

How does the agent exploit what is known about its world, decide how to act, and understand the outcomes of its actions? In order to be somewhat concrete about these questions, some context is helpful. STAR will be tested on a 3D physical visual search task. Specifically, it is the task described in Wu & Tsotsos (2025), the first documented experimental study that details human active observer behavior when searching in a real 3D world. There, a human experiment was conducted by building recording and analysis tools (described in Solbach & Tsotsos (2021)) and deploying them in a physical setup containing several tables, each having a few cages (as scaffolds) containing a number of toy objects (set sizes of 20 to 60) in varying 3D poses. A human subject was shown a target object in a canonical pose, then asked to find it. We recorded the subject's 6DOF head motion and fixation gaze changes as they searched for a given target. See Figure 6.

This is a classical visual search task but fully in a static three-dimensional world, with an active observer. The data from this, and from a related experiment where active subjects judged sameness of 2 given 3D objects Solbach & Tsotsos (2023) document the quality of behavior we seek for SA. The exact same setup is the setting for SA except that the tables will be removed and the entire setup placed on the floor, so that the robot dog can actively search the space. The goal is not to only be inspired by human behavior but to explain it.

With this context, the questions posed can be addressed. First, **how does an agent exploit what is known about its world?** It is beyond the scope of this paper to cover this breadth and we focus on our novel aspects. Most of the STAR functionality will be enabled by Cognitive Programs

(CPs) (Tsotsos, 2013; Tsotsos & Kruijne, 2014). This approach updates and expands visual routines originally proposed by Ullman (1983). The idea of CPs is to provide the following:

- a set of primitive actions that specify basic operations on the visual hierarchy, memory and other components of STAR;
- a language for providing the model with task or instructions;
- mechanisms for decomposing the task into primitive elements;
- mechanisms for chaining primitive actions to complete the task, set up expectations for the effects of actions, track its execution, determine success or failure, and learn from the experience.

The exploit part of STAR's cognitive cycle is triggered by the presentation of a user task. In terms of task, STAR is distinct as it will be applied to solving real-world perceptual problems in 3D that exercise and test the full breadth of visual attention functionality of ST. Macmillan & Creelman (2005) give an extensive picture of psychophysical tests and there are many that seem unexplored computationally (specifically, the tasks of absolute identification, ABX or match-to- sample, categorization, classification, correspondence, detection, detection with uncertainty, discrimination, fixed discrimination, m-alternative forced choice, multiple-look experiment, oddity, rating experiment, recognition, roving discrimination, same-different experiment, simultaneous detection and identification). Also, these experiments have a huge history of results in the literature, however, the vast majority of studies consider these tasks in more passive and 2D settings. Our goal is for STAR to live in a real 3D world as an active visual problem-solving actor (Bajcsy et al., 2018). The kinds of visual problems we are considering are the same, or extensions of, the visual problems first considered by Visual Routines (Ullman, 1983). Visual routines (VR) are composed of sequences of elemental operations. Routines for different properties and relations share elemental operations. Using a fixed set of basic operations, the visual system can assemble different routines to extract an unbounded variety of shape properties and spatial relations. The basic operations Ullman described include include shifting of the processing focus, indexing to an odd-man-out location, bounded activation, boundary tracing, and marking.

The problem of assembling operations into meaningful visual routines is only abstractly mentioned by Ullman and remains a central task for our research and cognitive science in general (Tsotsos et al., 2021). The idea of dynamic composition of behaviors for tasks is the foundation of means-ends analysis (MEA) Newell et al. (1959). However, many aspects of both VR and MEA proposals have become somewhat dated due to decades of knowledge gained by research in computational and cognitive science. Ullman's VRs employ only selection in a saliency map for attention, not the breadth of attentional mechanisms known to exist. MEA seems to assume that the sensing mechanism is fixed whereas in ST, as in humans, vision tunes processing depending on expectations. The tuning not only adds efficiency but it also improves signal-to-noise and often disentangles visual representations to make what is seen clear. Although MEA methods do indeed track expectations with perception, they appear independent, seemingly following a Marr-style processing regimen as did Ullman, which now is known to not represent human vision. Our Cognitive Programs are proposed as a modernized version of both.

Given a task, SA's Thinking component will search Memory for the Cognitive Programs suitable for the requested Task. If there is one, it is passed on to the Behavior component. If there is no CP, then one is assembled using existing ones or reasoned options (in a trial-and-error manner). The

Behavior component is tasked with taking a candidate CP, in its generic form, and tuning it so that it can execute in the current world and agent state. Behavior also makes predictions about the effects of any actions the CP will take.

Second, **how does an agent decide how to act?** As the brief description of CPs suggests, it is assumed that actions to be taken towards a task are specified in the CP. A CP is proposed as the sequence of actions required to solve the given task. CP's however, are maintained in memory in a generic state; we term them CP methods. CP methods thus require adaptation or tuning to the current state of the world and the current state of the agent. This tuning is the task of Behavior, to turn a generic sequence of actions into a sequence of actionable behaviors, the distinction being that behaviors act on the real world while methods cannot. In order to tune, Behavior must have access to Memory where state information is stored. It may also be the case that some aspect of a CP method requires tuning of some variable who value is unknown. Behavior can then request that Executive tune Perception for the observation of that variable. The output of Behavior is a tuned CP method, termed a script, ready for execution and passed off to Action. Action then sequences the steps to the appropriate Sensing & Motility elements.

Finally, **how does an agent understand the outcomes of its actions?** Intelligent agents need to be maximally sensitive to the task-relevant while remaining vigilant about the task irrelevant, and thus some connection between perception, their mission and task/world knowledge must exist. One such connection could be an effective use of inductive reasoning. Inductive reasoning takes specific information (premises) and makes a broader generalization (conclusion) that is considered probable. The only way to know is to test the conclusion; a passive sensing strategy could only do this by accident. Passive sensing thus impedes the use of any form of inductive reasoning. Part of the duties of the Behavior component is to develop the set of time-ordered expectations for the selected CP. These expectations must be communicated to the Executive and Perception, in sequence, so that the Executive can prepare the Perception system to be tuned in such a way as to be maximally able to observe those expected changes. For example, if the camera system is not looking at the right part of the world, it will miss an expected action. (Arrows 1, 2, 4, and 6 in Figure 3.)

Specifically, the Executive provides three kinds of preparation for Perception: it directs the choice of imaging geometry and sensor parameters to apply for the next scene sample to be acquired; it tunes the visual hierarchy, where possible, to be most receptive to the kinds of visual information needed for the task at that moment; and, it biases the interpretation process for the categorical classes expected from the current sampling action. As expectations are confirmed by observation, the CP moves along. When denied, the CP choice must be re-considered. The observed mis-match with expectations serves as a feedback signal for learning how that CP performs and to guide re-planning of an alternate one. In this way, by setting up expectations, optimizing for their visual confirmation, re-setting when not confirmed, and recording this series of events, the SA may be thought to understand the outcomes of its actions. (Arrows 3 and 6 in Figure 3.)

How is knowledge and experience about the world and behavior captured and remembered? STAR proposes a novel learning paradigm, motivated by current developmental science. Any one who has raised children, or even just observed them, knows that children become very adept from a very young age for visuospatial tasks similar to those on which we are focused. Thus the question of starting point arises: what is innate at birth? Although debate on this abounds, the

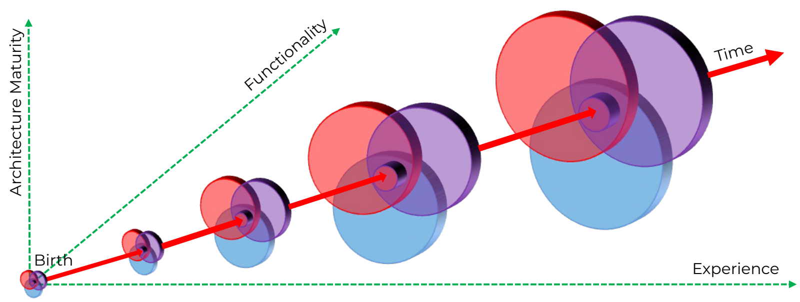

Figure 7: STAR's Learning Paradigm—Adaptive Developmental Bootstrapping over time. From "birth", STAR matures, accumulates experience, and converts it into new functionalities.

developmental literature suggests a possible answer: the process of developmental bootstrapping Aslin et al. (2023). Developmental bootstrapping considers that a low-level system, mediated by subcortical mechanisms, is modulated by a higher-level cortical system that operates in tandem with the subcortical system. The innate capabilities of the low-level system are then refined and matured under guidance of the higher-level system.

This basic idea is complicated by the fact that the visual system machinery and and physical substrate mature with age. Siu & Murphy (2018) provide a nice summary of the many neuroanatomical, physiological and functional changes that occur during a lifetime, both as they develop and then as they degrade. Taken together, with additional inspiration from Gopnik et al. (2017) regarding how behavior moves from exploration to exploitation with age (note the roles these two basic dimensions play throughout the STAR description), STAR proposes the Active Developmental Bootstrapping learning paradigm (Figure 7). Figure 7 uses the basic agent drawing of Figure 5 to illustrate 3 important elements: 1. STAR's birth explicitly includes a small set of innate abilities—the agent disks are smallest; 2. STAR's abilities grow with time—the disks grow in ability over time; 3. Growth is determined by a combination of experiential learning, architectural maturity and functional expansion—the three axes that underpin Active Developmental Bootstrapping. (Figure 3 arrow 3.)

STAR will assume a 2-year-old child as starting point whose innate capabilities include: eye, head and body movements, binocular fusion, stereopsis, feedforward and feedback connections within visual processes, with horizontal and intra-area connectivity (Siu & Murphy, 2018). It is assumed that these will be encoded as Cognitive Programs with movement functionality for our binocular head and a basic visual hierarchy. With this starting point we can imbue our agent with substantial visual talent. Understandably, this starting point deserves debate and skepticism; it is the hypothesis we wish to examine and certainly more plausible that the null or random starting points of other learning strategies. This set of basic behaviors permit the agent to explore its world and to complete simple visual tasks. Explore here may be as simple as noting areas of interest via visual saliency and taking a look. Within ST is a broadly used saliency model AIM (Bruce & Tsotsos, 2005). We assume that such a saliency detecting ability is innate. Any directed view of a scene involves foveation and thus selection of where to foveate using some criterion, for example, a visual onset or offset, another kind of attention. Another basic behavior might be trying to see what is beyond the edge of the current visual image. This requires a change in viewpoint, another topic where attention plays a role and we have existing well-tested past work (Ye & Tsotsos, 1999). With each exploration action, some bit of new knowledge is added to memory. With each successful action, the corresponding CP is strengthened. With each error, the CP can be modified by trial-and-error or by advice from the teacher. With each new exploration, a CP may be enhanced to slightly greater functionality, and depending on the tasks and circumstances it faces, developing an increasingly more complete set of behaviors with experience. Importantly, with each subsequent use of an existing CP, it may be streamlined to be more efficient. This was a surprising observation in Solbach & Tsotsos (2023) and Wu & Tsotsos (2025). Accuracy was high from the first trial and stayed high across trials. What improved over trials was efficiency in terms of fewer eye fixations, less head motion and less body translation (for these experiments, subjects were given no feedback after trials). These are among the roles of what Aslin et al. (2023) call the "higher level system" that modulated and improved the lower level. This description does not explicitly take into account the Architecture Maturation dimension of Figure 7, in part because the goals are difficult enough. However, if our plan is successful we will consider one specific aspect, namely, the maturation of long-distance top-down feedback (Siu & Murphy (2018), place full maturation of intracortical myelin at age 35) and the development of ST's surround suppression mechanism for which we have explicit data (Wong-Kee-You et al. (2019), show that the suppressive surround of ST is not present in children until about 8 years old and matures by 18 years old). (Arrow 5 in Figure 3).

4.3 Current Status of STAR

Finally, we summarize the current status of STAR. The genesis of the whole enterprise is the Selective Tuning model and it has seen many implementations and tests over the 4 decades since its appearance (some are shown in (Tsotsos, 2011)). A new implementation is underway.

A second key ingredient comes from Ullman's Visual Routines concept. Although we needed to modernize this as noted earlier into Cognitive Program, the implementation of CPs has been delayed mostly due to lack of clarity as to what they should encode. This problem has recently been resolved via our human experimental work which was the first to document how human subjects solve visual problems in an active manner in 3D environments, as mentioned earlier. Those experimental results make clear that the tasks, although visual in nature as problems to solve, involve a great deal of navigation and viewpoint change as well. These are currently being considered in the Cognitive Programs setting. Mean-Ends Analysis as a problem-solving framework also is important for STAR but it too required some modernization. We have yet to test these changes.

The embodiment has recently been completed, specifically, a human-like robotic binocular camera system. The connection between attention and eye fixation control, critical to the use of the robot head, has an effective implementation (Wloka et al., 2018). Other aspects of attentive and active behavior such as perceptual hierarchies for objects shapes (Mehrani & Tsotsos, 2023), figure-ground segregation (Mehrani & Tsotsos, 2021), color (Mehrani et al., 2020), and motion (Tsotsos et al., 2005), saliency (AIM, Bruce & Tsotsos (2005)), active sensor planning for search (Ye & Tsotsos, 1999) all have successful implementations, as cited earlier, but have not yet been tailored to the STAR architecture. The medium of communication between the world and the agent, Imperative

English, is newly re-implemented (Kunić, 2017). Further research on the components shown in Figures 4 and 5 is in progress.

4.4 STAR in the Core Model Representation

An alternative view of STAR using the Core Model representation (Figure 8) allows comparisons with other architectures. While STAR has similar components and connections, it focuses on the historically underrepresented areas related to perception, attention, learning, motor feedback, and error correction (highlighted in red in Figure 8). Specifically, with respect to the 6 highlighted (numbers in red boxes) topics in Figure 3, a summary of what STAR contributes follows.

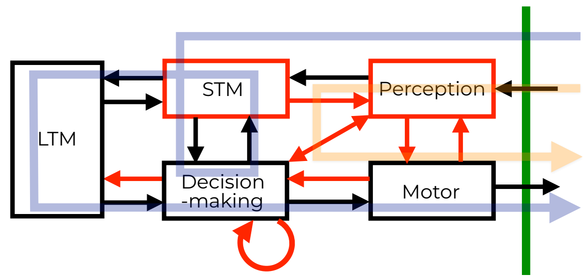

Figure 8: STAR in the Core Model representation. Deliberate and reactive cognitive cycles are shown as blue and orange arrows, respectively. Arrows and modules highlighted in red are areas of focus in STAR that are underrepresented in other architectures (see Figure 3).

- Sensorimotor integration (arrow 1 for Motor to Perception and arrow 2 for Perception to Motor Communication). In STAR there is a tight coupling between these. Perception tunes the motor dimensions of achieving a new view of the world and the novel designed binocular camera system that STAR embodies is designed to expect and deploy such tuning. The motor system provides proprioceptive feedback on all actions, specifically for all eye, head, neck and body movements on the embodiment.
- Priming, top-down attention, active perception (arrow 3). The primary purpose of STAR is to examine the role of these attentive components during visual problem-solving. The Selective Tuning model provides for priming and top-down attention, the AIM model provides for saliency, the data we collected experimentally documents the important roles played by navigating through a search environment and adjusting viewpoints as needed and STAR is designed to include those roles.
- Motor feedback to Decision-making (arrow 4). As mentioned above, the motor system provides proprioceptive feedback for motor actions. This feedback must first be interpreted, i.e., the raw motor, IMU or force feedback signals, just like raw image signals, require classification or abstraction in order to be useful. The Perception system of STAR is responsible for this and then can pass them to the Executive function for further analysis.
- Learning (arrow 5). One of the novel contributions of STAR is the proposed Adaptive Developmental Learning framework. In the CORE model learning appears as arrows from short-term memory and decision-making to long-term memory. In STAR, learning requires a discovery step as well. It is not enough to think that anything in short-term memory will wind up in long-term. Short-term memory also has a role as working memory - temporary storage for items to be used again in problem-solving. They may or may not be important for LTM, but only the Executive makes that decision. Discovery itself might involve perception and/or action as well, to con-

firm novelty of items being considered for LTM. This is a more sophisticated concept than most learning methods and is currently under development.

- Metacognition (arrow 6). This is covered in the paragraph above that deals with how the agent understands its actions. The loop in the Core Model figure, as far as STAR is concerned, represents the interplay among generation of expected outcomes of actions, tuning the perception system to check those outcomes, determining if the outcomes sufficiently satisfy the expectations, and deciding on the next course of action.

5. Conclusion

The cognitive cycle is a key component of any cognitive architecture determining its operation. Even though implemented cognitive architectures necessarily have some specification of cognitive cycle, it is often difficult to extract from literature and code. As a result, there are few comparisons of architectures at this level of abstraction. In this paper, we introduced a Core Model framework towards making such analysis possible. We defined core modules and connections among them, applied this template to 43 existing architectures, and combined the results. This revealed a number of issues, including lack of active perception and learning through feedback and self-observation. We then presented our STAR architecture as a pathway to solving these issues which we aim to achieve by relying on the established Selective Tuning model of attention, a novel Adaptive Developmental Bootstrapping paradigm, and a set of 3D vision tasks with ample human data. STAR is a computational probe into the roles that active vision and attention play in an embodied intelligent agent operating in a real 3D world. These roles are evident in perception, in testing expectations, in viewpoint determination, in novelty detection, in every kind of visual task SA may be asked to complete, and more. Exactly how each of these may be detailed is currently under study as is how the full STAR architecture may be developed.

References

Albus, J. S. (1971). A theory of cerebellar function. *Mathematical Biosciences*, *10*, 25–61.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*, 1036–1060.

Arbib, M. A. (1981). Perceptual structures and distributed motor control. In V. Brooks (Ed.), *Handbook of Physiology—The nervous System II. Motor Control*, 1449–1480.

Arrabales, R., Ledezma Espino, A. I., & Sanchis de Miguel, M. A. (2009). CERA-CRANIUM: A test bed for machine consciousness research. *Proceedings of the International Workshop on Machine Consciousness*.

Aslin, R. N., Fox, N. A., Lewkowicz, D. J., Maurer, D., Nelson, C. A., & von Hofsten, C. (2023). Multiple pathways to developmental continuity in infant cognition. *Trends in Cognitive Sciences*, *27*, 692–693.

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge University Press.

Bach, J. (2009). *Principles of Synthetic Intelligence PSI: An Architecture of Motivated Cognition*. Oxford University Press.

Bajcsy, R., Aloimonos, Y., & Tsotsos, J. K. (2018). Revisiting active perception. *Autonomous Robots*, *42*, 177–196.

Bartsch, M. V., Merkel, C., Strumpf, H., Schoenfeld, M. A., Tsotsos, J. K., & Hopf, J.-M. (2023). A cortical zoom-in operation underlies covert shifts of visual spatial attention. *Science Advances*, *9*, eade7996.

Bellas, F., Duro, R. J., Faiña, A., & Souto, D. (2010). Multilevel Darwinist Brain (MDB): Artificial evolution in a cognitive architecture for real robots. *IEEE Transactions on Autonomous Mental Development*, *2*, 340–354.

Benjamin, D. P., Lyons, D., & Lonsdale, D. (2004). ADAPT: A cognitive architecture for robotics. *International Conference on Cognitive Modeling* (pp. 337–338).

Bonasso, P. R., Firby, R. J., Gat, E., Kortenkamp, D., Miller, D. P., & Slack, M. G. (1997). Experiences with an architecture for intelligent, reactive agents. *Journal of Experimental & Theoretical Artificial Intelligence*, *9*, 237–256.

Breazeal, C. (2003). Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, *59*, 119–155.

Bresina, J., & Drummond, M. (1990). Integrating planning and reaction: A preliminary report. *Proceedings of the AAAI Spring Symposium on Planning in Uncertain, Unpredictable or Changing Environments*.

Bridewell, W., & Bello, P. (2016). A theory of attention for cognitive systems. *Proceedings of the Annual Conference on Advances in Cognitive Systems* (pp. 1–16).

Brooks, R. (1986). A robust layered control system for a mobile robot. *Journal on Robotics and Automation*, *2*, 14–23.

Brooks, R. A., Breazeal, C., Marjanović, M., Scassellati, B., & Williamson, M. M. (1999). The Cog project: Building a humanoid robot. In C. Nehaniv (Ed.), *Computation for metaphors, analogy, and agents*, 52–87.

Bruce, N., & Tsotsos, J. (2005). Saliency based on information maximization. *Advances in neural information processing systems*, *18*.

Bustos, P., Manso, L. J., Bandera, A. J., Bandera, J. P., Garcia-Varea, I., & Martinez-Gomez, J. (2019). The CORTEX cognitive robotics architecture: Use cases. *Cognitive Systems Research*, *55*, 107–123.

Carpenter, G. A., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, *37*, 54–115.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision research*, *51*, 1484–1525.

Cassimatis, N., Bignoli, P., Bugajska, M., Dugas, S., Kurup, U., Murugesan, A., & Bello, P. (2009). An architecture for adaptive algorithmic hybrids. *IEEE Transactions on Systems, Man, and Cy-*

bernetics, Part B, *40*, 903–914.

Choi, D., & Langley, P. (2018). Evolution of the ICARUS cognitive architecture. *Cognitive Systems Research*, *48*, 25–38.

Edelman, G. M. (2007). Learning in and from Brain-Based Devices. *Science*, *318*, 1103–1105.

Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., & Rasmussen, D. (2012). A large-scale model of the functioning brain. *Science*, *338*, 1202–1205.

Engelmore, R. (1988). *Blackboard systems*. Addison-Wesley.

Epstein, S. L., Freuder, E. C., Wallace, R., Morozov, A., & Samuels, B. (2002). The adaptive constraint engine. *Proceedings of the International Conference on Principles and Practice of Constraint Programming* (pp. 525–540).

Franklin, S., Madl, T., Strain, S., Faghihi, U., Dong, D., Kugele, S., Snaider, J., Agrawal, P., & Chen, S. (2016). A LIDA cognitive model tutorial. *Biologically Inspired Cognitive Architectures*, *16*, 105–130.

Gat, E. (1998). On three-layer architectures. In D. Kortenkamp, P. R. Bonasso, & R. Murphy (Eds.), *Artificial intelligence and mobile robots: Case studies of successful robot systems*, 195–210.

Georgeff, M. P., & Lansky, A. L. (1987). Reactive reasoning and planning. *Proceedings of the AAAI Conference on Artificial Intelligence* (pp. 677–682).

Gobet, F., & Lane, P. (2012). Chunking mechanisms and learning. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning*, 541–544.

Gobet, F., Lane, P. C., Croker, S., Cheng, P. C., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, *5*, 236–243.

Goertzel, B., Ke, S., Lian, R., O'Neill, J., Sadeghi, K., Wang, D., Watkins, O., & Yu, G. (2013). The CogPrime architecture for embodied artificial general intelligence. *IEEE Symposium on Computational Intelligence for Human-Like Intelligence* (pp. 60–67).

Gopnik, A., O'Grady, S., Lucas, C. G., Griffiths, T. L., Wente, A., Bridgers, S., Aboody, R., Fung, H., & Dahl, R. E. (2017). Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood. *Proceedings of the National Academy of Sciences*, *114*, 7892–7899.

Grossberg, S. (2021). *Conscious mind, resonant brain: How each brain makes a mind*. Oxford University Press.

Haikonen, P. O. (2007). *Robot brains: Circuits and systems for conscious machines*. Wiley.

Hayes-Roth, B. (1995). An architecture for adaptive intelligent systems. *Artificial Intelligence*, *72*, 329–365.

Hofstadter, D. R., & Mitchell, M. (1994). The Copycat project: A model of mental fluidity and analogy-making. In K. J. Holyoak & J. A. Barnde (Eds.), *Analogical connections*, 31–112. Ablex Publishing.

Huntsberger, T., Aghazarian, H., Howard, A., & Trotz, D. C. (2011). Stereo vision–based navigation for autonomous surface vessels. *Journal of Field Robotics*, *28*, 3–18.

Kawamura, K. (2023). A perspective on cognitive robot research and development. *International Journal of Humanoid Robotics*, (p. 2350023).

Kieras, D. E., Wakefield, G. H., Brungart, D. S., & Simpson, B. D. (2016). An EPIC cognitive-architectural account of spatial separation effects in two-channel listening tasks. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 686–690).

Kokinov, B. (1994). A hybrid model of reasoning by analogy. In K. Holyoak & J. Barnden (Eds.), *Advances in connectionist and neural computation theory*, volume 2, 247–318. Ablex.

Kotseruba, I., & Tsotsos, J. K. (2020). 40 years of cognitive architectures: Core cognitive abilities and practical applications. *Artificial Intelligence Review*, *53*, 17–94.

Kotseruba, I., & Tsotsos, J. K. (2025). *The computational evolution of cognitive architectures*. Oxford University Press.

Kunić, T. (2017). *Cognitive Program Compiler*. Master's thesis, York University.

Laird, J. E. (2022). An analysis and comparison of ACT-R and Soar. *Proceedings of the Annual Conference on Advances in Cognitive Systems*.

Laird, J. E., Lebiere, C., & Rosenbloom, P. S. (2017). A standard model of the mind: Toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. *AI Magazine*, *38*, 13–26.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. Lawrence Erlbaum Associates Publishers.

Maes, P. (1991). The agent network architecture (ANA). *ACM SIGART Bulletin*, *2*, 115–120.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. W. H. Freeman and Company.

Mehrani, P., Mouraviev, A., & Tsotsos, J. K. (2020). Multiplicative modulations enhance diversity of hue-selective cells. *Scientific Reports*, *10*, 8491.

Mehrani, P., & Tsotsos, J. K. (2021). Early recurrence enables figure border ownership. *Vision Research*, *186*, 23–33.

Mehrani, P., & Tsotsos, J. K. (2023). Self-attention in vision transformers performs perceptual grouping, not attention. *Frontiers in Computer Science*, *5*, 1178450.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part I. Basic mechanisms. *Psychological Review*, *104*, 3–65.

Minsky, M. (1986). *Society of mind*. Simon and Schuster.

Musliner, D. J., Durfee, E. H., & Shin, K. G. (1993). CIRCA: A cooperative intelligent real-time control architecture. *IEEE Transactions on Systems, Man, and Cybernetics*, *23*, 1561–1574.

Newell, A. (1990). *Unified theories of cognition*. Harvard University Press.

Newell, A., Shaw, J. C., & Simon, H. A. (1959). Report on a General Problem Solving Program. *International Conference on Information Processing*.

Rao, A. S., & Georgeff, M. P. (1995). BDI agents: From theory to practice. *International Conference on Multiagent Systems* (pp. 312–319).

Ritter, F. E., Bittner, J. L., Kase, S. E., Evertsz, R., Pedrotti, M., & Busetta, P. (2012). CoJACK: A high-level cognitive architecture with demonstrations of moderators, variability, and implications for situation awareness. *Biologically Inspired Cognitive Architectures*, *1*, 2–13.

Ritter, F. E., Tehranchi, F., & Oury, J. D. (2019). ACT-R: A cognitive architecture for modeling cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *10*, e1488.

Rohrer, B. (2012). BECCA: Reintegrating AI for natural world interaction. *AAAI Spring Symposium: Designing Intelligent Robots*.

Scheutz, M., Harris, J., & Schermerhorn, P. (2013). Systematic integration of cognitive and robotic architectures. *Proceedings of the Annual Conference on Advances in Cognitive Systems* (pp. 277–296).

Schulz, M.-C., Bartsch, M. V., Merkel, C., Strumpf, H., Schoenfeld, M. A., & Hopf, J.-M. (2024). Parallel gain modulation mechanisms set the resolution of color selectivity in human visual cortex. *Science Advances*, *10*, eadm7385.

Shapiro, S. C., & Ismail, H. O. (2003). Anchoring in a grounded layered architecture with integrated reasoning. *Robotics and Autonomous Systems*, *43*, 97–108.

Simmons, R. G. (1994). Becoming increasingly reliable. *Proceedings of the Conference on Artificial Intelligence Planning Systems* (pp. 152–157).

Siu, C. R., & Murphy, K. M. (2018). The development of human visual cortex and clinical implications. *Eye and Brain*, (pp. 25–36).

Solbach, M. D., & Tsotsos, J. K. (2021). Tracking active observers in 3d visuo-cognitive tasks. *ACM Symposium on Eye Tracking Research and Applications* (pp. 1–3).

Solbach, M. D., & Tsotsos, J. K. (2023). The psychophysics of human three-dimensional active visuospatial problem-solving. *Scientific Reports*, *13*, 19967.

Sun, R. (2007). The importance of cognitive architectures: An analysis based on CLARION. *Journal of Experimental & Theoretical Artificial Intelligence*, *19*, 159–193.

Sun, R. (2020). Potential of full human–machine symbiosis through truly intelligent cognitive systems. *AI & Society*, *35*, 17–28.

Tecuci, G. (1991). A multistrategy learning approach to domain modeling and knowledge acquisition. *European Working Session on Learning* (pp. 14–32).

Thórisson, K. R. (1998). Real-time decision making in multimodal face-to-face communication. *Proceedings of the Second International Conference on Autonomous Agents* (pp. 16–23).

Tsotsos, J. K. (2011). *A computational perspective on visual attention*. MIT Press.

Tsotsos, J. K. (2013). Cognitive programs: Towards an executive controller for visual attention. *Vision Sciences Society Annual Meeting 2013*.

Tsotsos, J. K. (2017). Complexity level analysis revisited: What can 30 years of hindsight tell us about how the brain might represent visual information? *Frontiers in Psychology*, *8*, 1216.

Tsotsos, J. K. (2022). When we study the ability to attend, what exactly are we trying to understand? *Journal of Imaging*, *8*, 212.

Tsotsos, J. K., Abid, O., Kotseruba, I., & Solbach, M. D. (2021). On the control of attentional processes in vision. *Cortex*, *137*, 305–329.

Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial intelligence*, *78*, 507–545.

Tsotsos, J. K., & Kruijne, W. (2014). Cognitive programs: Software for attention's executive. *Frontiers in Psychology*, *5*, 1260.

Tsotsos, J. K., Liu, Y., Martinez-Trujillo, J. C., Pomplun, M., Simine, E., & Zhou, K. (2005). Attending to visual motion. *Computer Vision and Image Understanding*, *100*, 3–40.

Ullman, S. (1983). *Visual routines*. Technical Report 723, MIT.

Verschure, P. F. (2012). Distributed Adaptive Control: A theory of the mind, brain, body nexus. *Biologically Inspired Cognitive Architectures*, *1*, 55–72.

Wang, P. (2022). Intelligence: From definition to design. *International Workshop on Self-Supervised Learning* (pp. 35–47).

Weng, J. (2002). A theory for mentally developing robots. *Proceedings of the IEEE International Conference on Development and Learning* (pp. 131–140).

Wloka, C., Kotseruba, I., & Tsotsos, J. K. (2018). Active fixation control to predict saccade sequences. *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3184–3193).

Wong-Kee-You, A. M., Tsotsos, J. K., & Adler, S. A. (2019). Development of spatial suppression surrounding the focus of visual attention. *Journal of vision*, *19*, 9–9.

Wu, T. C., & Tsotsos, J. K. (2025). Real-world visual search goes beyond eye movements: Active searchers select 3d scene viewpoints too. *PLoS One*, *20*, e0319719.

Ye, Y., & Tsotsos, J. K. (1999). Sensor planning for 3d object search. *Computer Vision and Image Understanding*, *73*, 145–168.

Zachary, W. W., Ryder, J. M., & Hicinbothom, J. H. (1998). Cognitive task analysis and modeling of decision making in complex environments. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training*, 315–345. APA.