

10 Cognitive Architectures

David Vernon

10.1 Introduction

As the definition of cognitive robotics in chapter 1 makes clear, the field draws on several disciplines, including robotics, artificial intelligence, and cognitive science. Its goal is to design an integrated cognitive system that combines a range of abilities, such as sensorimotor behaviors, knowledge-based reasoning, and social skills, in the form of an intelligent robot. Its foundations in systems engineering and cognitive science coalesce in a single concept: a cognitive architecture.

From the perspective of systems engineering, a cognitive architecture mirrors the system architecture, using the power of abstraction to render the modeling, specification, and design of a complete complex system tractable.

From the perspective of cognitive science, in which the term “cognitive architecture” originates (Newell 1990), the concept of a cognitive architecture is the result of over sixty years of research. To understand what it means from this perspective requires us to first familiarize ourselves with the roots of cognitive science and the different paradigms that exist within that discipline. In turn, this will allow us to understand the different types of cognitive architecture and the role a cognitive architecture plays in cognitive science, in general, and cognitive robotics, in particular.

With this understanding in place, we review the key attributes of a cognitive architecture before surveying the core cognitive abilities of the many cognitive architectures that exist today. We examine two cognitive architectures in some detail to highlight the way these abilities are realized in cognitive robots. We finish by exploring what the future might hold for cognitive architectures and the challenges that remain.

10.2 The Foundations of Cognitive Science

Cognitive science embraces neuroscience, cognitive psychology, linguistics, epistemology, philosophy, and artificial intelligence, among other disciplines. Its primary goal is to explain the underlying processes of human cognition, ideally in the form of a model that can be replicated in artificial agents. It has its roots in cybernetics in the early 1940s (Wiener 1948) but appears as a formal discipline referred to as cognitivism in the late

1950s. Cognitivism built on the logical foundations laid by the early cyberneticians and exploited the computer as a literal metaphor for cognitive function and operation, using symbolic information processing as its core model of cognition. Cybernetics also gave rise to the alternative emergent systems approach, which recognized the importance of self-organization in cognitive processes, eventually embracing connectionism, dynamical systems theory, and the enactive perspective on cognitive science. Hybrid systems attempt to combine the cognitivist and emergent paradigms to varying degrees, quite often ignoring some of the incompatible assumptions that the cognitivist and the emergent paradigms make about the fundamental nature of cognition (Vernon 2014).

10.2.1 The Cognitivist Paradigm of Cognitive Science

The cognitivist paradigm, which embraces artificial intelligence (AI), dates from a conference held at Dartmouth College, New Hampshire, in July and August 1956. It was attended by Allen Newell, Herbert Simon, John McCarthy, Marvin Minsky, and Claude Shannon, among others, all of whom exerted a very significant influence on the development of AI over the next half century.

The essential position of cognitivism is that cognition is achieved by computations performed on internal symbolic knowledge representations in a process whereby information about the world is taken in through the senses, filtered by perceptual processes to generate descriptions that abstract away irrelevant data, represented in symbolic form, and reasoned about to infer what is required to perform some task and achieve some goal. In the cognitivist paradigm, any physical platform that supports the performance of the required symbolic computation will suffice. In other words, the physical realization of the computational model is inconsequential to the model. The principled decoupling of computational operation from the physical platform that supports these computations is referred to as computational functionalism (Piccinini 2010). Allen Newell made several landmark contributions to the establishment of practical cognitivist systems: in the early 1980s with his introduction of the concept of a knowledge-level system, the maximum rationality hypothesis, and the principle of rationality (Newell 1982); in the mid-1980s with the development of the Soar cognitive architecture for general intelligence (along with John Laird and Paul Rosenbloom; Laird, Newell, and Rosenbloom 1987); and in 1990 with the concept of a unified theory of cognition (Newell 1990)—that is, a theory that covers a broad range of cognitive issues, such as attention, memory, problem-solving, decision-making, and learning from several aspects, including psychology, neuroscience, and computer science.

10.2.2 The Emergent Paradigm of Cognitive Science

In the emergent paradigm, cognition is one of the processes by which an autonomous system maintains its autonomy. Through cognition, the system constructs its reality—its world and the meaning of its perceptions and actions—as a result of its operation in that world. This process of making sense of its environmental interactions is one of the foundations of a branch of cognitive science called *enaction* (Stewart, Gapenne, and Di Paolo 2010; Vernon 2010). Cognition is also the means by which the system prepares for interaction that may be necessary in the future. Thus, cognition is intrinsically linked with the ability of an agent to act prospectively. As such, many emergent approaches focus on the acquisition of anticipatory skills rather than knowledge, asserting that processes that guide

action and improve the capacity to guide action form the root capacity of all intelligent systems (Christensen and Hooker 2000). As a result, in contrast to cognitivism, emergent approaches are necessarily embodied, and the physical form of the agent's body plays a causal role in the cognitive process. Together, the body and the brain form the basis of a cognitive system, and they do so in the context of a structured environmental niche to which the body is adapted. Because of this, cognition in the emergent paradigm is sometimes referred to as *embodied cognition*, although some emergent approaches make even stronger assertions about the nature of cognition. The emergent paradigm typically exploits connectionism or dynamical systems theory. In general, connectionist systems correspond to models at a lower level of abstraction, dynamical systems to a higher level. They are sometimes referred to as subsymbolic processes.

10.2.3 Hybrid Systems

Hybrid systems are attempts to exploit both the cognitivist and emergent paradigms of cognitive science. They exploit symbolic knowledge to represent the agent's world and logical rule-based systems to reason with this knowledge to pursue tasks and achieve goals. At the same time, they typically use emergent models of perception and action to explore the world and construct this knowledge. Hybrid systems use both symbolic and subsymbolic representations. The latter are constructed using subsymbolic connectionist processes as the system interacts with and explores the world. So, instead of a designer programming in all the necessary knowledge, objects and events in the world can be represented by observed correspondences between sensed perceptions, agent actions, and sensed outcomes. Thus, as with an emergent system, a hybrid system's ability to understand the external world is dependent on its ability to flexibly interact with it. Interaction becomes an organizing mechanism that establishes a learned association between perception and action. For a detailed comparison of cognitivist, emergent, and hybrid paradigms of cognitive science, see Vernon, Metta, and Sandini (2007b) and Vernon (2014).

10.3 The Types of Cognitive Architecture

A cognitive architecture is a software framework that integrates all the elements required for a system to exhibit the attributes considered to be characteristic of a cognitive agent. Just what these elements are is open to interpretation, but as we will see, there is common ground in the identification of core cognitive abilities in these interpretations—for example, perception, action, learning, adaptation, anticipation, motivation, autonomy, internal simulation, attention, action selection, memory, reasoning, and metareasoning (Vernon 2014; Vernon, von Hofsten, and Fadiga 2016; Kotseruba and Tsotsos 2020).

Furthermore, a cognitive architecture determines the overall structure and organization of a cognitive system, including the component parts or modules (Sun 2004), the relations between these modules, and the essential algorithmic and representational details within them (Langley 2006). The architecture specifies the formalisms for knowledge representations and the types of memories used to store them, the processes that act upon that knowledge, and the learning mechanisms that acquire it. For cognitivist and hybrid approaches, a cognitive architecture also provides a way of programming the system so that domain and task knowledge can be embedded in the system.

A cognitive architecture makes explicit the set of assumptions upon which that cognitive model is founded. These assumptions are typically derived from several sources: biological or psychological data, philosophical arguments, or working hypotheses inspired by work in different disciplines such as neurophysiology, psychology, and artificial intelligence. Once it has been created, a cognitive architecture also provides a framework for developing the ideas and assumptions encapsulated in the architecture.

There are three different types of cognitive architecture, each derived from the three paradigms of cognitive science: the cognitivist, the emergent, and the hybrid. Cognitivist cognitive architectures are often referred to as symbolic cognitive architectures (Kotseruba and Tsotsos 2020). It is noteworthy that the term “cognitive architecture” itself is due to Allen Newell and his colleagues in their work on unified theories of cognition (Newell 1990). Consequently, for cognitivism a cognitive architecture represents any attempt to create a unified theory of cognition. The cognitive architectures Soar (Laird, Newell, and Rosenbloom 1987; Laird 2009, 2012), ACT-R (Anderson 1996; Anderson et al. 2004), and CLARION (Sun 2007, 2016) are archetypal candidate unified theories of cognition, all of which are classified as hybrid cognitive architectures in the survey by Kotseruba and Tsotsos (2020).

10.3.1 The Cognitivist Perspective on Cognitive Architecture

In the cognitivist paradigm, the focus in a cognitive architecture is on the aspects of cognition that are constant over time and that are independent of the task (Ritter and Young 2001; Langley, Laird, and Rogers 2009). A cognitivist cognitive architecture is a generic computational model that is neither domain-specific nor task-specific, and it needs to be provided with knowledge to perform any given task. This combination of a given cognitive architecture and a particular knowledge set is generally referred to as a *cognitive model*. In many cognitivist systems, much of the knowledge incorporated into the model is normally provided by the designer, and often this knowledge is highly crafted, possibly drawing on years of experience working in the problem domain. Machine learning is increasingly used to augment and adapt this knowledge.

10.3.2 The Emergent Perspective on Cognitive Architecture

Emergent approaches to cognition focus on the development of the agent from a primitive state to a fully cognitive state over its lifetime. As such, an emergent cognitive architecture is the initial state from which an agent subsequently develops. Development requires exposure to an environment that is conducive to development, one in which there is sufficient regularity to allow the system to build a sense of understanding of the world around it but not excessive variety that would overwhelm an agent that has inherent limitations on the speed with which it can develop. Thus, emergent cognition has two aspects, architecture and gradually acquired experience, mirroring the two aspects of a cognitivist cognitive architecture: architecture and knowledge. These two aspects of emergent cognition are referred to as phylogeny and ontogeny (or ontogenesis), the latter being the interactions and experiences that a developing cognitive system is exposed to as it acquires an increasing degree of cognitive capability. Since the emergent paradigm holds that the physical system—the body—is also a part of the cognitive process, an emergent cognitive archi-

ecture should reflect in some way the structure and capabilities—the morphology—of the physical body in which it is embedded.

10.3.3 The Hybrid Perspective on Cognitive Architecture

As we have noted, hybrid systems endeavor to have the best of both worlds, combining the strengths of the cognitivist and emergent approaches. Most hybrid systems focus on integrating symbolic and subsymbolic (usually connectionist) processing.

Hybrid cognitive architectures are the most prevalent type. The survey by Kotseruba and Tsotsos (2020) lists twenty-two symbolic (i.e., cognitivist) cognitive architectures, fourteen emergent, and forty-eight hybrid, thirty-eight of which are fully integrated.

10.4 Desirable Characteristics of a Cognitive Architecture

If a cognitive architecture is intended to be a unified theory of cognition, as most cognitivist cognitive architectures are, then it should exhibit certain desirable attributes—*desiderata*—including ecological realism, bioevolutionary realism, cognitive realism, and eclecticism of methodologies and techniques, as well as several behavioral characteristics (Sun 2004). Ecological realism means that a cognitive architecture should focus on allowing the cognitive system to operate in its natural environment, engaging in everyday activities and dealing with many concurrent and often conflicting goals with many environmental contingencies. Bioevolutionary realism means that a cognitive model of human intelligence should be reducible to a model of animal intelligence. Cognitive realism means that a cognitive architecture should attempt to capture the essential characteristics of human cognition from the perspectives of psychology, neuroscience, and philosophy. Eclecticism of methodologies and techniques means that new models should draw on, subsume, or supersede older models. Most cognitive architectures for cognitive robotics are not intended to be a unified theory of cognition, and consequently, these attributes can be addressed only to the extent that they are useful from a robotics perspective.

In the emergent paradigm of cognitive science, development is the process whereby a cognitive agent 1) expands its repertoire of action capabilities and 2) extends the time horizon of its ability to anticipate events in its world, including the need to act, the outcome of selected actions, the intentions of other cognitive agents, and the outcome of their actions (Vernon 2010). These considerations give rise to an additional set of *desiderata* for developmental cognitive architectures (Vernon, von Hofsten, and Fadiga 2016), including the need for a value system to determine the goals of actions and provide the drive for achieving them (Oudeyer, Kaplan, and Hafner 2007; Merrick 2010) along with exploratory and social motives (Piaget 1954; Vygotsky 1978; Lindblom 2015) to modulate behavior and select actions (Edelman 2006). The adaptation inherent in development is dependent on learning. A developmental cognitive architecture needs to have at least three different modes of learning: supervised learning, reinforcement learning, and unsupervised learning (Doya 1999). It also requires some mechanism to simulate future events (Seligman et al. 2013), to simulate the execution of actions and the likely outcome of those actions (Hesslow 2002, 2012), and to take alternative perspectives, including those of other agents (Schacter, Addis, and Buckner 2008).

10.5 Surveys of Cognitive Architectures

While several surveys of cognitive architectures have been published over the past ten or so years (Vernon, Metta, and Sandini 2007b; Duch, Oentaryo, and Pasquier 2008; Samsnovich 2010; Thórisson and Helgasson 2012), the recent survey by Kotseruba and Tsotsos (2020) is by far the most comprehensive. It targets eighty-four cognitive architectures, estimating that approximately three hundred cognitive architectures have been developed and that approximately one-third are currently active. The most often cited cognitive architectures are ACT-R (Anderson 1996; Anderson et al. 2004), Soar (Laird, Newell, and Rosenbloom 1987; Laird 2012), CLARION (Sun 2007, 2016), ICARUS (Langley 2006; Langley and Choi 1999 [2006]), EPIC (Kieras and Meyer 1997), and LIDA (Franklin et al. 2007, 2014). The majority of cognitive architectures focus on modeling human cognition.

Despite its comprehensive coverage, almost inevitably the Kotseruba and Tsotsos survey is not complete. For example, it omits the CRAM cognitive architecture (Beetz et al. 2010; Mösenlechner 2016), possibly because the CRAM literature refers to a cognitive robot abstract machine and to cognition-enabled robotics, rather than a cognitive architecture. Later in the chapter, we use CRAM as one of our two examples of cognitive architectures for cognitive robotics. Nevertheless, the survey provides a peerless basis on which to compare and contrast existing cognitive architectures by addressing the extent to which they exhibit core cognitive abilities, and we will refer to it throughout this section.

10.5.1 Comparing Cognitive Architectures

Despite efforts to establish an agreed set of criteria for comparing and evaluating cognitive architectures based on desirable characteristics such as Sun's (2004) desiderata and Newell's (1990, 1992) functional criteria, disagreements persist regarding the research goals, structure, operation, and application of cognitive architectures. Because of this, and in the absence of a clear definition and general theory of cognition, not to mention difficulties in defining intelligence, Kotseruba and Tsotsos adopt a pragmatic approach, treating intelligence as a set of system competences and behaviors. Thus, rather than summarize and review each cognitive architecture individually, Kotseruba and Tsotsos address seven core cognitive abilities—perception, attention mechanisms, action selection, memory, learning, reasoning, and metareasoning—and discuss the degree to which the eighty-four architectures surveyed exhibit these abilities. Significantly, they don't include anticipation (i.e., prospection) as a core cognitive ability as others do, both in surveys of cognitive architectures (Vernon, Metta, and Sandini 2007b) or in the cognitive science literature (Atance and O'Neill 2001; Gilbert and Wilson 2007; Schacter et al. 2012; Seligman et al. 2013). On the other hand, they do include attention, reasoning, and metacognition, three pivotal abilities that have often been omitted from other surveys. We summarize these seven core cognitive abilities in the following, adding, for completeness, a short note on the central role played by anticipation (i.e., prospection) in cognition and cognitive architectures.

10.5.2 Core Cognitive Abilities

Perception

Perception is a process that transforms raw input into the system's internal representation. Vision is the most commonly implemented sensory modality, but more than half of the

cognitive architectures surveyed use simulated visual input rather than transforming the raw sensory data. In general, symbolic cognitive architectures tend to have limited perceptual abilities, and therefore they rely on direct simulated data input. Audition is less commonly found in cognitive architectures, while touch, smell, and proprioception are rarely implemented with any fidelity. Most architectures use only two modalities simultaneously—for example, vision and audition or vision and range data (e.g., from Lidar sensors). Only a few architectures aim for biological fidelity in perception. For the most part, cognitive architectures ignore crossmodal interaction and adopt a modular approach when dealing with sensory modalities, despite its importance in developmental robotics (Cangelosi and Schlesinger 2015).

Attention

Attention is a process that reduces the information a cognitive system has to process, selecting relevant information and filtering out irrelevant information from sensory data. Kotseruba and Tsotsos refer to three classes of information reduction mechanism (Tsotsos 2011): *selection*, *restriction*, and *suppression*. Selection mechanisms choose one entity from many—for example, gaze and viewpoint selection, restriction mechanisms choose some entities from many, and suppression mechanisms suppress some entities from many. The restrictive mechanism reduces the search space by priming—that is, preparing the visual system for input based on task requirements, exogenous motivations (e.g., domain knowledge), exogenous cues (external stimuli), exogenous tasks (restricting attention to objects relevant to the task), and visual field (limiting the visual field). The suppression mechanisms include feature or spatial inhibition, task-irrelevant stimuli suppression, negative priming, and location or object inhibition of return to bias the agent returning attention to previously attended locations. The most frequently implemented mechanisms of attention are selection and restriction, with only a few cognitive architectures implementing a suppression mechanism. Kotseruba and Tsotsos note that visual attention is largely overlooked in cognitive architectures, with exceptions including the ISAC (Kawamura et al. 2008) and iCub cognitive architectures (Vernon, Sandini, and Metta 2007a).

Action Selection

Action selection determines what the agent should do next. There are two major approaches: planning and dynamic action selection. Planning, using traditional AI techniques, determines a sequence of steps to reach a certain goal or solve a problem prior to execution of the plan. Dynamic action selection involves the selection of one action based on knowledge at the time, typically using winner-take-all, probabilistic, or predefined order selection mechanisms. The criteria for selection include relevance, utility (in the sense of expected contribution to the current goal), and internal functions—for example, transient emotion, drives, or internal mechanisms, including basic physiological needs and high-level social drives and personality traits that bias or modulate the action selection rather than directly determining the next behavior. Planning, prevalent in symbolic architectures and in hybrid architectures but also found in emergent architectures, is often augmented with dynamic action selection mechanisms to improve the capability for adaptivity to environmental changes.

Memory

Kotseruba and Tsotsos identify six types of memory in cognitive architectures: short-term sensory memory and working memory and long-term episodic, semantic, procedural, and

global memory. Sensory memory is a very short-term buffer that stores several recent percepts and has a decay rate in the region of tens of milliseconds for visual data, longer for aural data. Working memory is temporary storage for percepts and information related to the current task and is frequently associated with the current focus of attention. It is critical for attention, reasoning, and learning.

Episodic memory (Tulving 1972, 1984) plays a key role in the anticipatory aspect of cognition. It refers to specific instances in the agent's experience, while semantic memory refers to general knowledge about the agent's world that may be independent of the agent's specific experience: knowledge of general facts about objects and concepts and the relationships between those objects. In symbolic cognitive architectures, semantic memory is often represented as a graph-like ontology network, the nodes being the concepts and the links the relationships. In emergent cognitive architectures, semantic memory is typically represented by a pattern of activity in a connectionist network.

Episodic and semantic memory are collectively known as declarative memory. Declarative memory captures knowledge, while procedural memory captures skills, equipping an agent to "know that" and "know how," respectively (Ryle 1949).

In symbolic production systems, procedural knowledge is the knowledge of how to carry out some task, represented by a set of if-then rules preprogrammed or learned for a particular domain. In emergent systems, procedural memory may comprise sequences of state-action pairs or perceptuomotor associations.

Global memory is reserved for cognitive architectures that don't draw the type-duration distinction and use a unified global structure for all knowledge.

Learning

Learning refers to an ability for a system to improve its performance over time through the acquisition of knowledge or skill. Two types of learning can be distinguished: declarative and nondeclarative. Declarative learning is concerned with explicit knowledge acquisition, while nondeclarative learning focuses on perceptual, procedural, associative, and nonassociative learning.

Of the eighty-four cognitive architectures surveyed by Kotseruba and Tsotsos, nineteen—mostly symbolic and hybrid—do not implement learning of any type.

Declarative learning can take several forms. In production systems, new declarative knowledge—facts about the world—are learned when either a fact or a rule is added to declarative memory—for example, after completing a goal or resolving an impasse. Thus, new symbolic knowledge is learned when local inference rules are applied to existing knowledge to obtain new knowledge, encapsulated in what is referred to as a *chunk*. In emergent and hybrid cognitive architectures, declarative learning often takes the form of the association of perceptual features with the identity of objects.

Perceptual learning refers to learning about the environment from perceptual data: uncovering perceptual patterns, constructing associations between percepts, and inferring knowledge about the environment—for example, its spatial organization.

Procedural learning refers to learning skills by repetitive practice until the skill becomes automatic. Note that this view of procedural learning entails a different view of what constitutes procedural knowledge compared with procedural knowledge in cognitivist production systems.

Associative learning is used to refer to the process of improving decision-making through the influence of reward and punishment. Reinforcement learning is often used as a computational model of associative learning, including variants such as temporal difference learning, Q-learning, and Hebbian learning. Nearly half the cognitive architectures surveyed use reinforcement learning to implement associative learning. Since reinforcement learning can be used with many forms of representation, it is used in all types of cognitive architecture: symbolic, emergent, and hybrid. In symbolic (i.e., cognitivist) cognitive architectures, reinforcement learning facilitates adaptation by weighting the importance of beliefs and actions based on the outcome of their use. In hybrid and emergent cognitive architectures, reinforcement learning also facilitates adaptation, but in these cases by establishing weighted associations between states and actions. This often takes the form of an initial phase of motor babbling—that is, performing random movements and observing their sensory outcome, followed by a learning phase to establish stable patterns known as sensorimotor contingencies.

Nonassociative learning refers to an often gradual adjustment of the weighting or importance of a single system entity, rather than an associative linking between two or more entities—for example, the gradual reduction of the strength of a response to some stimulus or pattern of system activity that is repeatedly presented. This is known as habituation. Sensitization has the opposite effect, such as a gradual increase in the strength of response to some repeated stimulus or activity.

Kotseruba and Tsotsos note that, surprisingly, deep learning does not yet feature strongly in cognitive architectures, but it is likely to play an important role in the future. We return to this topic in section 10.7.

Reasoning

Reasoning is the ability to logically and systematically process knowledge, typically to infer conclusions. The three classical forms of logical inference are deduction, induction, and abduction. In the context of cognitive architectures, reasoning focusses on the practical objective of finding the next (best) action to perform. Cognitive architectures typically aim to facilitate human-level intelligence, but they do not necessarily try to model the processes of human reasoning. Those that do include ACT-R (Anderson 1996; Anderson et al. 2004), Soar (Laird et al. 1987; Laird 2009, 2012), and CLARION (Sun 2007, 2016). Many emergent cognitive architectures do not address reasoning, even if they are capable of facilitating complex behavior. Some emergent cognitive architectures, such as SPA (Eliasmith et al. 2012), effect symbolic reasoning using neural architectures, raising the possibility that it might not be necessary to introduce a hard distinction between symbolic cognition and subsymbolic cognition.

Metacognition

Metacognition refers to a cognitive system's ability to monitor its internal cognitive processes and reason about them, acquiring data about the internal operation and status of the cognitive system—for example, availability of internal resources, confidence values during task execution, and sometimes generating temporal traces of activity during task execution. Approximately one-third of the eighty-four cognitive architectures surveyed by Kotseruba and Tsotsos have a metacognition element. These are mainly symbolic cognitive

architectures and hybrid cognitive architectures with a strong component of symbolic processing. Metacognition is needed for social cognition, especially if the cognitive architecture is to form a *theory of mind*, also known as perspective taking—that is, the ability to infer the cognitive states of other agents with which it is interacting, predicting their behavior, and acting appropriately. Very few cognitive architectures support this ability. Kotseruba and Tsotsos note only two: Sigma (Rosenbloom, Demski, and Ustun 2016) and PolyScheme (Trafton et al. 2005).

Prospection

Although the core cognitive abilities identified by Kotseruba and Tsotsos do not include prospection, it plays such a central role in cognition that we include it here for completeness.

Prospection—the capacity to anticipate the future—is one of the hallmark attributes of cognition. It also lies at the heart of the other core characteristics of a cognitive agent: autonomy, perception, action, learning, and adaptation (Vernon 2014). It facilitates autonomy and the ability to cope with adversarial conditions by allowing the agent to prepare to act. It is also involved in constitutive autonomy (Froese, Virgo, and Izquierdo 2007), predictively adjusting internal system processes through allostasis (Sterling 2012). It facilitates perception through expectation-driven attentional processes (Borji, Sihite, and Itti 2014). Attention, in turn, facilitates predictive control of, for example, gaze (Flanagan and Johansson 2003) and the prediction of the consequences of actions (Flanagan et al. 2013). In general, anticipation is central to action since actions are goal directed and guided by prospective information (von Hofsten 2009): a cognitive agent continually anticipates the need to act, and it anticipates the outcome of those actions (Vernon, von Hofsten, and Fadiga 2011). Prospection also lies at the heart of learning, for learned models are used both for prediction and explanation. Finally, adaptivity arises in cognitive agents when the learned models fail to produce accurate or reliable predictions.

Consensus is emerging that internal simulation plays a key role in prospection (Svensson, Lindblom, and Ziemke 2007; Mohan, Bhat, and Morasso 2018). However, there is less agreement about the manner in which internal simulation is accomplished. Some cognitive architectures opt for an explicit module in the architecture (e.g., Kawamura et al. 2008; Beetz et al. 2010; Kunze and Beetz 2017), while in others it is a covert mode of operation, with internal simulation effected by the same subsystems as those responsible for sensorimotor-mediated action but using covert, internally generated endogenous sensorimotor signals rather than exogenous sensorimotor signals (e.g., Demiris and Khadhour 2006; Shanahan 2006).

10.5.3 Applications

Kotseruba and Tsotsos identify more than nine hundred projects that use one of the eighty-four cognitive architectures surveyed. They identify ten classes of application: psychological experiments, robotics, human performance modeling, human-robot and human-computer interaction, natural language processing, categorization and clustering, computer vision, games and puzzles, virtual agents, and miscellaneous projects that don't fall into the other nine classes. Robotics applications account for nearly a quarter of all applications, mainly for navigation and obstacle avoidance, fetch and carry tasks, object localization, and object manipulation.

10.6 Example Cognitive Architectures

To highlight the issues we have covered so far, in this section we examine two examples of cognitive architectures that focus specifically on cognitive robotics: CRAM (Betz, Mösenlechner, and Tenorth 2010; Mösenlechner 2016), a knowledge-based reasoning architecture, and ISAC (Kawamura et al. 2008), an architecture built from communicating software agents and memory subsystems.

10.6.1 The CRAM Cognitive Architecture

CRAM stands for cognitive robot abstract machine. It is a hybrid cognitive architecture, first introduced in 2010 (Betz, Mösenlechner, and Tenorth 2010). Since then it has developed significantly, building on the original basis for the architecture: the achievement of cognition-enabled robot manipulation in everyday situations, carrying out goal-directed tasks that need only be vaguely defined using underdetermined robot action plans specified in abstract terms. The vagueness is resolved at runtime by reasoning: querying knowledge bases and combining the resultant knowledge with information about the current state of the robot's environment acquired through perception, inferring the concrete actions that need to be performed to achieve the goal, and adapting them at runtime, as necessary. For example, figure 10.1 shows a PR2 robot setting a table during a demonstration of CRAM-based robot manipulation at the Everyday Activity Science and Engineering interdisciplinary research center (<https://ease-crc.org/>).

CRAM—see figure 10.2—comprises five core elements: 1) the CRAM Plan Language (CPL) executive; 2) a suite of knowledge bases and associated reasoning mechanisms, collectively referred to as KnowRob2 (Betz et al. 2018); 3) a perception executive; 4) an action executive; and 5) a metacognitive reasoning system. Several publications document



Figure 10.1

A PR2 robot setting a table during a demonstration of cognition-enabled robot manipulation using the CRAM. *Source:* Courtesy of the EASE interdisciplinary research center at the University of Bremen, Germany.

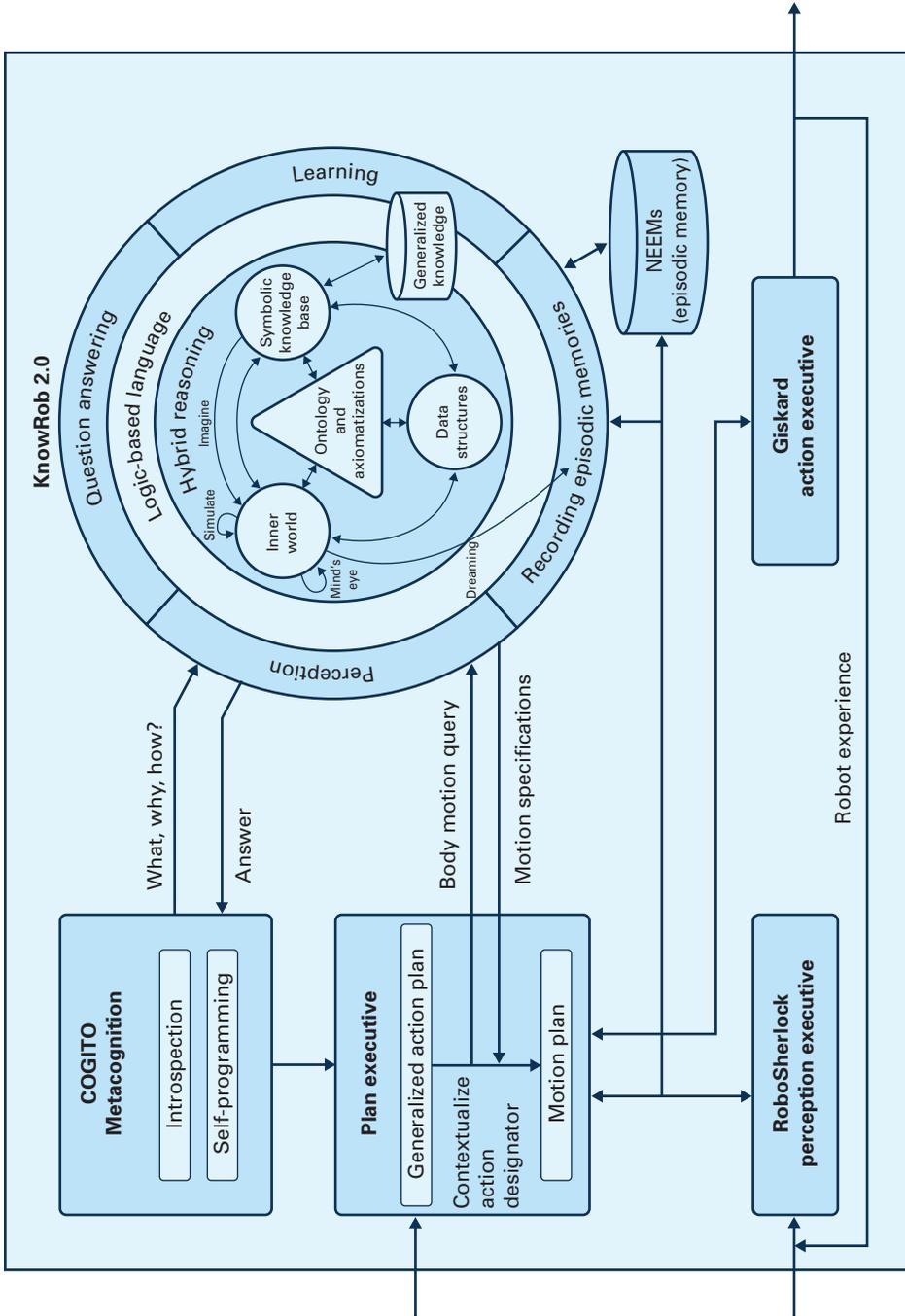


Figure 10.2 The CRAM cognitive architecture. *Source:* Courtesy of the EASE interdisciplinary research center at the University of Bremen, Germany.

the development of CRAM over the past ten years, a small subset of which includes Winkler et al. (2012), Tenorth and Beetz (2013), Beetz et al. (2015), and Kunze and Beetz (2017).

The CRAM Plan Language (CPL) executive is an extension of the Lisp programming language. It represents all the key aspects of a plan as persistent first-class objects in first-order logic. Thus, CRAM can reason about its plans, even at runtime. This is particularly relevant in the metacognition system. Plans specify how the robot should respond to sensory events, changes in belief states, and detected failures in plans. All these aspects of a plan can be queried, inspected, and reasoned about. A plan comprises a set of abstract plan designators for actions, objects, locations, and motions—that is, elementary movements. Designators are effectively placeholders and require runtime resolution based on the current context of the task action. Designator resolution is accomplished either by querying a priori knowledge embedded in the plan, by querying knowledge in the KnowRob2 knowledge base, or by accessing sensorimotor data through the perception executive. All plans have a similar generic structure, as shown below. The terms prefixed with a question mark are resolved at runtime based on the current state of the robot and the environment.

```
(par
  (perform
    (desig: an action
      (type picking-up)
      (arm ?grasping-arm)
      (grasp left-side)
      (object ?perceived-object ))
    . . .
  )
)
```

The KnowRob2 knowledge base is a knowledge representation and reasoning framework for robotic agents (cf. chapter 21). It is implemented in Prolog, and it is exposed as a conventional first-order time interval logic knowledge base. However, many logic expressions are constructed on demand from sensorimotor data computed in real time. It provides the background common sense intuitive-physics knowledge required by the CPL executive to implement its goal-directed underdetermined task plans—for example, how to grasp an object, depending on the object’s shape, weight, softness, and other properties; how it must be held while moving it—for example, upright to avoid spilling its contents; and where the object is normally located. Some knowledge is specified a priori, some is derived from experience, and some is the result of the simulated execution of candidate actions using a high-fidelity virtual reality physics engine simulator. All knowledge is represented by a first-order time interval logic expression and reasoned about, as needed.

KnowRob2 comprises five core elements embedded in a hybrid (i.e., multiformalism) reasoning shell, exposed through a logic-based language layer to an interface shell that provides perception, question answering, experience acquisition, and knowledge learning. The five elements are 1) a central set of knowledge ontologies and axiomatizations; 2) an episodic memory knowledge base encapsulating the robot’s experiences, represented in both subsymbolic form and in generalized symbolic form; 3) an inner-world knowledge base and virtual reality physics engine simulator; 4) a logic knowledge base with abstracted symbolic

sensor and action data, logical axioms, and inference rules; and 5) a virtual knowledge base comprising a set of data structures for parameterized motion control and path planning.

The knowledge ontologies and associated axiomatizations provide structured representation of the knowledge about the robot and its environment. There is a core ontology and additional special-purpose, application-specific ontologies. The core ontology defines the robot configuration, object configurations, robot actions, tasks, activities and behaviors, environment configuration, and situational context. The axioms identify roles that objects can play—for example, a mug is a cylindrical vessel, with a handle, that can be used as a receptacle from which its contents can be drunk, mixed, or poured.

One of the main distinguishing aspects of KnowRob2 is its focus on episodic memory. This is an autobiographical memory of the robot's experience as it had carried out tasks in the past. These are organized as NEEMS—narrative-enabled episodic memories—a concept introduced by the KnowRob2 designers. A NEEM comprises an experience and a narrative. The experience is a detailed low-level recording of a certain episode, such as records of poses and percepts based on exteroceptive and proprioceptive sensory data. It also includes control signals. This is unusual because motor aspects of memory are normally stored in procedural memory. Thus, CRAM episodic memory, and NEEMS in particular, generalizes the concept of an episode to include procedural elements. The narrative is an abstract symbolic description of the tasks, the context, the intended goals, and the observed effects (Beetz et al. 2018). KnowRob2 episodic memory, in representing procedural knowledge as declarative knowledge, allows it to be reasoned about. The episodic knowledge base provides the basis for answers to queries such as what actions were performed by the robot, when it performed them, how it performed them, why they were performed, whether or not they were successful, what the robot perceived while performing them, and what the robot believed when it performed them. The extraction of the generalized symbolic knowledge from NEEMS is facilitated by an interface to the Weka machine-learning framework (Holmes, Donkin, and Witten 1994).

The inner-world knowledge base facilitates geometric reasoning using a high-quality virtual reality system and physics engine. This allows KnowRob2 to simulate the outcome of candidate action and to establish the feasibility of that action. It provides symbolic names and properties for each entity, and it can infer background knowledge—for example, where an object is stored. The inner-world knowledge base serves two roles: as a representation of the belief state of the robot about itself and the world and as a reasoning mechanism for determining the outcome of candidate actions. Thus, it encapsulates two types of knowledge: current beliefs about robots and the world and the projected internal simulation of future states. It also acts as a learning mechanism, generating episodic memories off-line, effectively dreaming while physically inactive, and running simulations of activities with varying control parameters. These are recorded and transferred to the episodic-memory knowledge base.

The logic knowledge base provides information about the entities in the robot's environment, including objects, object parts, object articulation models, and environments composed of objects, actions, and events. It uses an entity description language that allows partial descriptions of entities in terms of both symbolic and subsymbolic properties.

The virtual knowledge base provides computable predicates that facilitate the integration of nonsymbolic data into the reasoning process, allowing symbolic queries of non-

symbolic data. This allows run-time sensorimotor states to be integrated into the knowledge base at run-time and to be used in reasoning in the same way as symbolic knowledge.

KnowRob2 provides a logic-based language interface that allows the hybrid reasoning shell to be exposed as a purely symbolic knowledge base even though internally it uses multiple symbolic and subsymbolic representations and reasoning formalisms. In this way, KnowRob2 can be treated by the CPL executive (and other systems through its OpenEASE interface; Beetz et al. 2015) as a symbolic, object-oriented query system in which entities can be retrieved by providing partial descriptions of them using the entity predicate. This allows KnowRob2 to appear as a “Siri for robots” (Beetz 2018)—that is, as a query and response oracle. Consequently, during task execution there is an ongoing dialogue between the CPL executive and KnowRob2, in which the CPL executive presents a series of under-determined queries, and KnowRob2 provides the corresponding responses, allowing the CPL executive to carry out the task using the action executive.

The action executive controls the robot by mapping parameterized actions (as requested by the CPL executive) to adaptive trajectories in real time.

Sensory information is available to the CPL executive either directly from the perception executive or indirectly through KnowRob2 by means of the virtual knowledge base and the associated computable predicates.

The metacognition subsystem allows CRAM to reason about plans and exploit transformational learning and planning to improve them in two complementary ways: by specialization using pragmatic everyday activity manifolds (PEAMs) and by generalization through metacognitive induction. This is possible because, as we noted above, the plans themselves are represented as first-class objects in first-order logic. PEAMs capture the subspace of motions necessary to carry out an action successfully by exploiting the constraints that knowledge of everyday activities and the environment bring to bear, rendering tractable by specialization the solution of problems that in their full generality are intractable. Generalization through metacognitive induction complements the PEAM solution strategy by exploring patterns among actions plans, seeking ways to transform them either by carrying out the action in a more efficient and effective manner or by accomplishing the outcome of the action in a different way.

10.6.2 ISAC

ISAC—intelligent soft arm control—is a hybrid cognitive architecture for an upper-torso humanoid robot also called ISAC (Kawamura et al. 2008). From a software engineering perspective, ISAC is constructed from an integrated collection of software agents and associated memories. Agents encapsulate all aspects of a component of the architecture, operate asynchronously (i.e., without a shared clock to keep the processing of all agents locked in step with each other), and communicate with each other by passing messages.

As shown in figure 10.3, the multiagent ISAC cognitive architecture comprises Activator Agents for motion control, Perceptual Agents, and a First-Order Response Agent (FRA) to effect reactive perception-action control. It has three memory systems: short-term memory (STM), long-term memory (LTM), and a working memory system (WMS).

STM has a robot-centered spatiotemporal memory of the perceptual events currently being experienced. This is called a Sensory EgoSphere (SES), and it is a discrete representation of what is happening around the robot, represented by a geodesic sphere indexed

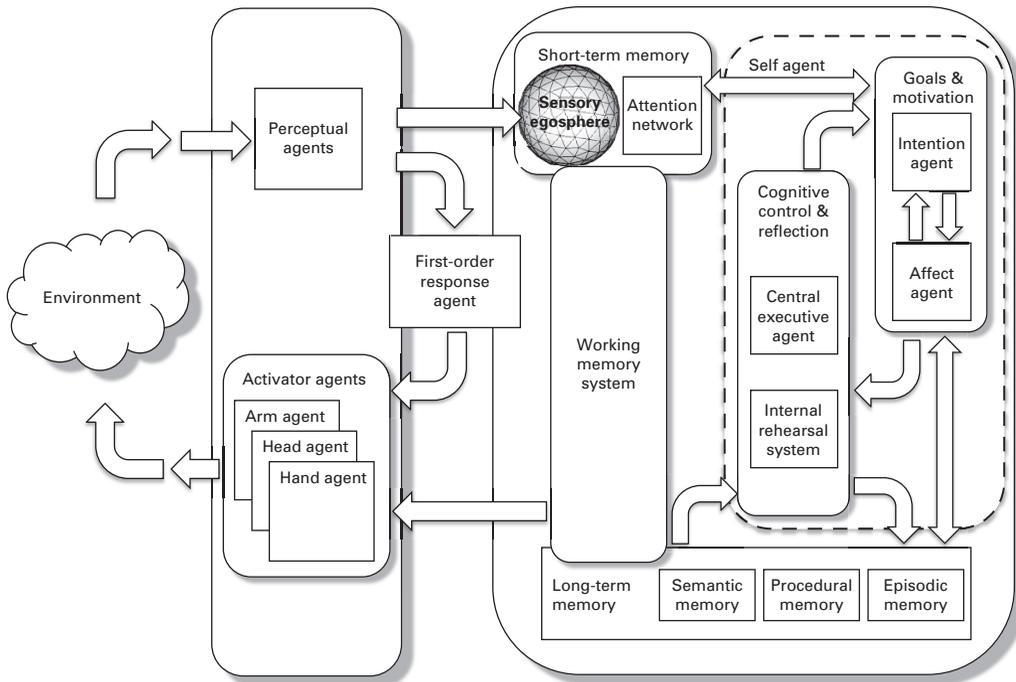


Figure 10.3
The ISAC cognitive architecture.

by two angles: horizontal (azimuth) and vertical (elevation). STM also has an Attention Network that determines the perceptual events that are most relevant and then directs the robot's attention to them.

LTM stores information about the robot's learned skills and past experiences. LTM is made up of semantic, episodic, and procedural memory. Together, the semantic memory and episodic memory make up the robot's declarative memory of the facts it knows. On the other hand, procedural memory stores representations of the motions the robot can perform.

ISAC's episodic memory abstracts past experiences and creates links or associations between them. It has multiple layers. At the lowest level, an episodic experience contains information about the external situation (i.e., task-relevant percepts from the SES), goals, emotions (in this case, internal evaluation of the perceived situation), actions, and outcomes that arise from actions and the valuations of these outcomes (e.g., how close they are to the desired goal state and any reward received as a result). Episodes are connected by links that encapsulate behaviors: transitions from one episode to another. Higher layers abstract away specific details and create links based on the transitions at lower levels. This multilayered approach allows for the efficient matching and retrieval of memories.

WMS, inspired by neuroscience models of brain function, temporarily stores information that is related to the task currently being executed. It forms a type of cache memory for STM, and the information it stores, called chunks, encapsulates expectations of future reward that are learned using a neural network.

Cognitive behavior is the responsibility of a Central Executive Agent (CEA) and an Internal Rehearsal System (IRS), a system that simulates the effects of possible actions. Together with a Goals and Motivation subsystem comprising an Intention Agent and an

Affect Agent, the CEA and the IRS form a compound agent called the Self Agent that, along with the FRA, makes decisions and acts according to the current situation and ISAC's internal states. The CEA is responsible for cognitive control, invoking the skills required to perform some given task on the basis of the current focus of attention and past experiences. The goals are provided by the Intention Agent. Decision-making is modulated by the Affect Agent.

ISAC works in the following way. Normally, the FRA produces reactive responses to sensory triggers. However, it is also responsible for executing tasks. When a task is assigned by a human, the FRA retrieves the skill from procedural memory in the LTM that corresponds to the skill described in the task information. It then places it in the WMS as chunks along with the current percept. The Activator Agent then executes it, suspending execution whenever a reactive response is required. If the FRA finds no matching skill for the task, the CEA takes over, recalling from episodic memory past experiences and behaviors that contain information similar to the current task. One behavior-percept pair is selected, based on the current percept in the SES, its relevance, and the likelihood of successful execution as determined by internal simulation in the IRS. This is then placed in working memory, and the Activator Agent executes the action.

10.7 Future Prospects

The design and implementation of a cognitive architecture is a daunting undertaking. This is evident when you consider that contemporary cognitive architectures such as Soar (Laird, Newell, and Rosenbloom 1987; Laird 2009, 2012), ACT-R (Anderson et al. 2004; Anderson 1996), CLARION (Sun 2007, 2016), and CRAM (Beetz et al. 2010; Mösenlechner 2016) have taken ten to twenty years or more to develop and are all still being developed further.¹ In an effort to consolidate cognitive architecture research, the cognitive science community has launched an exercise to identify the key design features shared by the most prominent cognitive architectures, with the goal of creating a common model of cognition (Laird, Lebiere, and Rosenbloom 2017) and promoting more cohesive development and achieving greater progress. In any case, progress will depend on the thorough evaluation of cognitive architectures in diverse, challenging, realistic environments (Kotseruba and Tsotsos 2020) consistent with human-level intelligence, such as the CRAM cognitive architecture targets in everyday activity science and engineering (EASE).

There is a need for more realistic perceptual capabilities that can operate in adverse conditions with noise and uncertainty, using context to improve performance. Almost half the cognitive architectures surveyed by Kotseruba and Tsotsos do not implement any visual perception or other sensory modalities. For example, audition, touch, and olfaction are typically addressed in a trivial manner (Kotseruba and Tsotsos 2020).

Cognitive architectures also need to facilitate more natural communication with humans to infer their intentions and emotional states; engage in adaptive, personalized interaction; read body language, such as gestures and facial expressions; engage in natural turn-taking; and facilitate human-robot joint action. Examples of cognitive architectures that focus on these aspects of cognitive human-robot interaction include Lemaignan et al. (2017); Sandini et al. (2018); Tanevska et al. (2019).

Computational models of episodic memory have not received significant attention, especially for lifelong learning, despite the fact that its existence and importance has been

widely recognized (Kotseruba and Tsotsos 2020). Notable exceptions include the CRAM cognitive architecture (Beetz, Mösenlechner, and Tenorth 2010; Mösenlechner 2016) and the iCub neural framework for episodic memory (Mohan, Sandini, and Morasso 2014).

Deep learning (Schmidhuber 2015; Goodfellow, Bengio, and Courville 2016) has not yet made a significant impact on cognitive architectures (Kotseruba and Tsotsos 2020). This will almost certainly change, giving rise to new architectural requirements—for example, deep developmental robotics architectures (Sigaud and Droniou 2016) and a reconciliation of deep learning with symbolic artificial intelligence (Garnelo and Shanahan 2019). One of the main advantages of deep learning is its ability to produce end-to-end systems—that is, systems that map directly from an input space to an output space, such as pixels-to-classes in computer vision. In robotics, the situation is different: end-to-end systems must map from pixels (and other sensory stimuli) to torques in a dynamic interactive environment. Supervised deep learning based on static data sets is not viable in these circumstances. However, deep reinforcement learning (Arulkumaran et al. 2017; Li 2018) is capable of learning end-to-end robot control or action policies. This form of learning is typically implemented using simulators and may not be feasible on physical robots. Sünderhauf et al. (2018) estimate that it would take fifty-three days to accomplish a deep reinforcement learning exercise that currently takes twenty-four hours using simulation. They suggest that there is also a reality gap between simulation and the real world that limits the usefulness of simulation-based deep reinforcement learning, and they discuss a solution based on transfer learning, initially learning in the simulated environment, freezing the network weights, and then continuing the learning with the physical robot. On the other hand, results using photorealistic simulations to support reasoning in cognition-enabled robots (Beetz et al. 2018; Mania and Beetz 2019) suggest that the reality gap may not be significant and that the simulation approach may be plausible.

10.8 Conclusion

A cognitive architecture captures both the abstract conceptual form and the details of functional operation, focusing on inner cohesion and self-contained completeness. This means that all of the mechanisms required for cognition fall under the compass of a cognitive architecture, including perception, attention, action, control, learning, reasoning, memory, adaptivity, and anticipation. Thus, cognition, as a process, and a cognitive architecture, as a framework, embrace all of the elements required for effective action. A cognitive architecture specifies the system components and the way these components are dynamically related as a whole. It provides both an abstract model of cognitive behavior and a sufficient basis for a software instantiation of that model (Lieto et al. 2018). Despite the magnitude of the task, the design and implementation of an appropriate cognitive architecture remains an indispensable step in the creation of a cognitive robot.

Additional Reading and Resources

- To delve more deeply into the field of cognitive architectures, you might begin by reading the review by Kotseruba and Tsotsos (2020) and referring to the companion

website: http://jtl.lassonde.yorku.ca/project/cognitive_architectures_survey/index.html. The review does not focus specifically on robot cognitive architectures but provides a contemporary and comprehensive overview of the field, nonetheless.

- Appendix A of Vernon, von Hofsten, and Fadiga (2011), summarizing the operation of twenty cognitive architectures: Vernon, D., C. von Hofsten, and L. Fadiga. 2011. *A Roadmap for Cognitive Development in Humanoid Robots*. In Vol. 11, *Cognitive Systems Monographs*. Berlin: Springer. http://www.vernon.eu/COSMOS_CAs.pdf.
- The Introduction to Cognitive Robotics course (www.cognitiverobotics.net) has several lectures devoted to cognitive architectures, in general, and to the CRAM cognitive architecture summarized in section 10.6.1, in particular, expanding on the material in the online CRAM tutorials: <http://cram-system.org/tutorials>.
- Software is available online for, for example, the CRAM cognitive architecture: <http://cram-system.org>; the openEASE software components for cognition-enabled control of robotic agents: <https://ease-crc.org/open-ease/>; and the iCub cognitive robot platform: <http://www.icub.org>.
- Instructions on how to access, download, and install the CRAM software is included in the Introduction to Cognitive Robotics course and on the CRAM website (<http://cram-system.org/installation>), along with practical exercises to help you get started.
- For other software resources, refer to the “Resources” page on the IEEE Technical Committee for Cognitive Robotics site: <http://www.ieee-coro.org>.

Note

1. The average age of cognitive architecture projects in the survey by Kotseruba and Tsotsos (2020) is approximately fifteen years.

References

- Anderson, John R. 1996. “ACT: A Simple Theory of Complex Cognition.” *American Psychologist* 51 (4): 355.
- Anderson, John R., Daniel Bothell, Michael D. Byrne, Scott Douglass, Christian Lebiere, and Yulin Qin. 2004. “An Integrated Theory of the Mind.” *Psychological Review* 111 (4): 1036.
- Arulkumaran, Kai, Marc Peter Deisenroth, Miles Brundage, and Anil Anthony Bharath. 2017. “Deep Reinforcement Learning: A Brief Survey.” *IEEE Signal Processing Magazine* 34 (6): 26–38.
- Atance, Cristina M., and Daniela K. O’Neill. 2001. “Episodic Future Thinking.” *Trends in Cognitive Sciences* 5 (12): 533–539.
- Beetz, Michael. 2018. Personal Communication.
- Beetz, Michael, Daniel Beßler, Andrei Haidu, Mihai Pomarlan, Asil Kaan Bozcuoğlu, and Georg Bartels. 2018. “KnowRob 2.0—a 2nd Generation Knowledge Processing Framework for Cognition-Enabled Robotic Agents.” In *2018 IEEE International Conference on Robotics and Automation*, 512–519. New York: IEEE.
- Beetz, Michael, Dominik Jain, Lorenz Mösenlechner, and Moritz Tenorth. 2010. “Towards Performing Everyday Manipulation Activities.” *Robotics and Autonomous Systems* 58 (9): 1085–1095.
- Beetz, Michael, Lorenz Mösenlechner, and Moritz Tenorth. 2010. “CRAM—a Cognitive Robot Abstract Machine for Everyday Manipulation in Human Environments.” In *Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1012–1017. New York: IEEE.
- Beetz, Michael, Moritz Tenorth, and Jan Winkler. 2015. “Open-EASE—a Knowledge Processing Service for Robots and Robotics/AI Researchers.” In *2015 IEEE International Conference on Robotics and Automation*, 1983–1990. New York: IEEE.

- Borji, Ali, Dicky N. Sihite, and Laurent Itti. 2013. "What/Where to Look Next? Modeling Top-Down Visual Attention in Complex Interactive Environments." *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 44 (5): 523–538.
- Cangelosi, Angelo, and Matthew Schlesinger. 2015. *Developmental Robotics: From Babies to Robots*. Cambridge, MA: MIT Press.
- Christensen, Wayne D., and Clifford Alan Hooker. 2000. "An Interactivist-Constructivist Approach to Intelligence: Self-Directed Anticipative Learning." *Philosophical Psychology* 13 (1): 5–45.
- Demiris, Yiannis, and Bassam Khadhouri. 2006. "Hierarchical Attentive Multiple Models for Execution and Recognition of Actions." *Robotics and Autonomous Systems* 54 (5): 361–369.
- Doya, Kenji. 1999. "What Are the Computations of the Cerebellum, the Basal Ganglia and the Cerebral Cortex?" *Neural Networks* 12 (7–8): 961–974.
- Duch, Włodzisław, Richard Jayadi Oentaryo, and Michel Pasquier. 2008. "Cognitive Architectures: Where Do We Go from Here?" In *Artificial General Intelligence 2008, Proceedings of the First AGI Conference, AGI 2008, March 1–3, 2008, University of Memphis, Memphis, TN, USA*, 122–136. *Frontiers in Artificial Intelligence and Applications* 171. Amsterdam: IOS Press.
- EASE. n.d. Everyday Activity Science and Engineering. Accessed August 21, 2021. <https://ease-crc.org/>.
- Edelman, Gerald M. 2006. *Second Nature: Brain Science and Human Knowledge*. New Haven, CT: Yale University Press.
- Eliasmith, Chris, Terrence C. Stewart, Xuan Choo, Trevor Bekolay, Travis Dewolf, Yichuan Tang, and Daniel Rasmussen. 2012. "A Large-Scale Model of the Functioning Brain." *Science* 338 (6111): 1202–1205.
- Flanagan, J. Randall, and Roland S. Johansson. 2003. "Action Plans Used in Action Observation." *Nature* 424 (6950): 769–771.
- Flanagan, J. Randall, Gerben Rotman, Andreas F. Reichelt, and Roland S. Johansson. 2013. "The Role of Observers' Gaze Behavior When Watching Object Manipulation Tasks: Predicting and Evaluating the Consequences of Action." *Philosophical Transactions of the Royal Society B: Biological Sciences* 368 (1628): 20130063.
- Franklin, Stan, Tamas Madl, Sidney D'Mello, and Javier Snaider. 2014. "LIDA: A Systems-Level Architecture for Cognition, Emotion, and Learning." *IEEE Transactions on Autonomous Mental Development* 6 (1): 19–41.
- Franklin, Stan, Uma Ramamurthy, Sidney K. D'Mello, Lee Mccauley, Aregahegn Negatu, Rodrigo L. Silva, and Vivek Datla. 2007. "LIDA: A Computational Model of Global Workspace Theory and Developmental Learning." In *AAAI Fall Symposium on AI and Consciousness: Theoretical Foundations*, 61–66. Menlo Park, CA: AAAI Press.
- Froese, Tom, Nathaniel Virgo, and Eduardo Izquierdo. 2007. "Autonomy: A Review and a Reappraisal." In Vol. 4648, *Proceedings of the 9th European Conference on Artificial Life: Advances in Artificial Life*, edited by F. Almeida E. Costa, L. Rocha, E. Costa, I. Harvey, and A. Coutinho, 455–465. Berlin: Springer.
- Garnelo, Marta, and Murray Shanahan. 2019. "Reconciling Deep Learning with Symbolic Artificial Intelligence: Representing Objects and Relations." *Current Opinion in Behavioral Sciences* 29:17–23.
- Gilbert, Daniel T., and Timothy D. Wilson. 2007. "Prospection: Experiencing the Future." *Science* 317 (5843): 1351–1354.
- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. Cambridge, MA: MIT Press.
- Hesslow, Germund. 2002. "Conscious Thought as Simulation of Behavior and Perception." *Trends in Cognitive Sciences* 6 (6): 242–247.
- Hesslow, Germund. 2012. "The Current Status of the Simulation Theory of Cognition." *Brain Research* 1428:71–79.
- Holmes, Geoffrey, Andrew Donkin, and Ian H. Witten. 1994. "Weka: A Machine Learning Workbench." In *Proceedings of ANZIS'94—Australian New Zealand Intelligent Information Systems Conference*, 357–361. New York: IEEE.
- Kawamura, Kazuhiko, Stephen M. Gordon, Palis Ratanaswasd, Erdem Erdemir, and Joseph F. Hall. 2008. "Implementation of Cognitive Control for a Humanoid Robot." *International Journal of Humanoid Robotics* 5 (4): 547–586.
- Kieras, Davis E., and Davis E. Meyer. 1997. "An Overview of the EPIC Architecture for Cognition and Performance with Application to Human-Computer Interaction." *Human-Computer Interaction* 12 (4): 391–438.
- Kotseruba, Iuliia, and John K. Tsotsos. 2020. "40 Years of Cognitive Architectures: Core Cognitive Abilities and Practical Applications." *Artificial Intelligence Review* 53 (1): 17–94.
- Kunze, Lars, and Michael Beetz. 2017. "Envisioning the Qualitative Effects of Robot Manipulation Actions Using Simulation-Based Projections." *Artificial Intelligence* 247:352–380.
- Laird, John E. 2009. "Toward Cognitive Robotics." In Vol. 7332, *Proceedings of the SPIE—Unmanned Systems Technology XI*. P. 73320Z. Bellingham, WA: International Society for Optics and Photonics.
- Laird, John E. 2012. *The Soar Cognitive Architecture*. Cambridge, MA: MIT Press.

- Laird, John E., Christian Lebiere, and Paul S. Rosenbloom. 2017. "A Standard Model of the Mind: Toward a Common Computational Framework across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics." *AI Magazine* 38 (4): 13–26.
- Laird, John E., Allen Newell, and Paul S. Rosenbloom. 1987. "Soar: An Architecture for General Intelligence." *Artificial Intelligence* 33:1–64.
- Langley, Pat. 2006. "Cognitive Architectures and General Intelligent Systems." *AI Magazine* 27 (2): 33–33.
- Langley, Pat, and Dongkyu Choi. 1999 [2006]. "A Unified Cognitive Architecture for Physical Agents." In *Proceedings of the National Conference on Artificial Intelligence*, Vol. 21, no. 2, 1469. Menlo Park, CA: AAAI Press; Cambridge, MA: MIT Press.
- Langley, Pat, John E. Laird, and Seth Rogers. 2009. "Cognitive Architectures: Research Issues and Challenges." *Cognitive Systems Research* 10 (2): 141–160.
- Lemaignan, Séverin, Mathieu Warnier, E. Akin Sisbot, Aurélie Clodic, and Rachid Alami. 2017. "Artificial Cognition for Social Human–Robot Interaction: An Implementation." *Artificial Intelligence* 247:45–69.
- Li, Yuxi. 2018. "Deep Reinforcement Learning." ArXiv preprint: 1801.06339v1.
- Lieto, Antonio, Mehul Bhatt, Alessandro Oltramari, and David Vernon. 2018. "The Role of Cognitive Architectures in General Artificial Intelligence." *Cognitive Systems Research* 48:1–3.
- Lindblom, Jessica. 2015. *Embodied Social Cognition*. Vol. 26. Berlin: Springer.
- Mania, Patrick, and Michael Beetz. 2019. "A Framework for Self-Training Perceptual Agents in Simulated Photorealistic Environments." In *2019 International Conference on Robotics and Automation*, 4396–4402. New York: IEEE.
- Merrick, Kathryn Elizabeth. 2010. "A Comparative Study of Value Systems for Self-Motivated Exploration and Learning by Robots." *IEEE Transactions on Autonomous Mental Development* 2 (2): 119–131.
- Mohan, Vishwanathan, Ajaz Bhat, and Pietro Morasso. 2019. "Muscleless Motor Synergies and Actions without Movements: From Motor Neuroscience to Cognitive Robotics." *Physics of Life Reviews* 30:89–111.
- Mohan, Vishwanathan, Giulio Sandini, and Pietro Morasso. 2014. "A Neural Framework for Organization and Flexible Utilization of Episodic Memory in Cumulatively Learning Baby Humanoids." *Neural Computation* 26 (12): 2692–2734.
- Mösenlechner, Lorenz. 2016. "The Cognitive Robot Abstract Machine: A Framework for Cognitive Robotics." PhD thesis, Technical University of Munich.
- Newell, Allen. 1982. "The Knowledge Level." *Artificial Intelligence* 18 (1): 87–127.
- Newell, Allen. 1990. *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.
- Newell, Allen. 1992. "Précis of Unified Theories of Cognition." *Behavioral and Brain Sciences* 15:425–492.
- Oudeyer, Pierre-Yves, Frédéric Kaplan, and Verena V. Hafner. 2007. "Intrinsic Motivation Systems for Autonomous Mental Development." *IEEE Transactions on Evolutionary Computation* 11 (2): 265–286.
- Piaget, Jean. 1954. *The Construction of Reality in the Child*. New York: Basic Books.
- Piccinini, Gualtiero. 2010. "The Mind as Neural Software? Understanding Functionalism, Computationalism, and Computational Functionalism." *Philosophy and Phenomenological Research* 81 (2): 269–311.
- Ritter, Frank E., and Richard M. Young. 2001. "Embodied Models as Simulated Users: Introduction to This Special Issue on Using Cognitive Models to Improve Interface Design." *International Journal of Human-Computer Studies* 55 (1): 1–14.
- Rosenbloom, Paul S., Abram Demski, and Volkan Ustun. 2016. "The Sigma Cognitive Architecture and System: Towards Functionally Elegant Grand Unification." *Journal of Artificial General Intelligence* 7 (1): 1–103.
- Ryle, G. 1949. *The Concept of Mind*. London: Hutchinson's University Library.
- Samsonovich, Alexei V. 2010. "Toward a Unified Catalog of Implemented Cognitive Architectures." In *Proceedings of the Conference on Biologically Inspired Cognitive Architectures*, 195–244. Amsterdam: IOS Press.
- Sandini, Giulio, Vishwanathan Mohan, Alessandra Sciutti, and Pietro Morasso. 2018. "Social Cognition for Human-Robot Symbiosis—Challenges and Building Blocks." *Frontiers in Neurobotics* 12:34.
- Schacter, Daniel L., Donna Rose Addis, and Randy L. Buckner. 2008. "Episodic Simulation of Future Events: Concepts, Data, and Applications." *Annals of the New York Academy of Sciences* 1124:39–60.
- Schacter, Daniel L., Donna Rose Addis, Demis Hassabis, Victoria C. Martin, R. Nathan Spreng, and Karl K. Szpunar. 2012. "The Future of Memory: Remembering, Imagining, and the Brain." *Neuron* 76 (4): 677–694.
- Schmidhuber, Jürgen. 2015. "Deep Learning in Neural Networks: An Overview." *Neural Networks* 61:85–117.
- Seligman, Martin E. P., Peter Railton, Roy F. Baumeister, and Chandra Sripada. 2013. "Navigating into the Future or Driven by the Past." *Perspectives on Psychological Science* 8 (2): 119–141.

- Shanahan, Murray. 2006. "A Cognitive Architecture That Combines Internal Simulation with a Global Work-space." *Consciousness and Cognition* 15 (2): 433–449.
- Sigaud, Olivier, and Alain Droniou. 2016. "Towards Deep Developmental Learning." *IEEE Transactions on Cognitive and Developmental Systems* 8 (2): 99–114.
- Sterling, Peter. 2012. "Allostasis: A Model of Predictive Regulation." *Physiology and Behavior* 106 (1): 5–15.
- Stewart, John, Olivier Gapenne, and Ezequiel A. Di Paolo. 2010. *Enaction: Toward a New Paradigm for Cognitive Science*. Cambridge, MA: MIT Press.
- Sun, Ron. 2004. "Desiderata for Cognitive Architectures." *Philosophical Psychology* 17 (3): 341–373.
- Sun, Ron. 2007. "The Importance of Cognitive Architectures: An Analysis Based on CLARION." *Journal of Experimental and Theoretical Artificial Intelligence* 19 (2): 159–193.
- Sun, Ron. 2016. *Anatomy of the Mind: Exploring Psychological Mechanisms and Processes with the Clarion Cognitive Architecture*. Oxford: Oxford University Press.
- Sünderhauf, Niko, Oliver Brock, Walter Scheirer, Raia Hadsell, Dieter Fox, Jürgen Leitner, Ben Upcroft, et al. 2018. "The Limits and Potentials of Deep Learning for Robotics." *International Journal of Robotics Research* 37 (4–5): 405–420.
- Svensson, Henrik, Jessica Lindblom, and Tom Ziemke. 2007. "Making Sense of Embodied Cognition: Simulation Theories of Shared Neural Mechanisms for Sensorimotor and Cognitive Processes." In *Body, Language and Mind, Volume 1: Embodiment*, edited by T. Ziemke, J. Zlatev, and R. M. Frank, 241–269. Berlin: Mouton De Gruyter.
- Tanevska, Ana, Francesco Rea, Giulio Sandini, Lola Cañamero, and Alessandra Sciutti. 2019. "A Cognitive Architecture for Socially Adaptable Robots." In *2019 Joint IEEE 9th International Conference on Development and Learning and Epigenetic Robotics*, 195–200. New York: IEEE.
- Tenorth, Moritz, and Michael Beetz. 2013. "KnowRob: A Knowledge Processing Infrastructure for Cognition-Enabled Robots." *International Journal of Robotics Research* 32 (5): 566–590.
- Thórisson, Kristinn, and Helgi Helgasson. 2012. "Cognitive Architectures and Autonomy: A Comparative Review." *Journal of Artificial General Intelligence* 3 (2): 1–30.
- Trafton, J. Gregory, Nicholas L. Cassimatis, Magdalena D. Bugajska, Derek P. Brock, Farilee E. Mintz, and Alan C. Schultz. 2005. "Enabling Effective Human-Robot Interaction Using Perspective-Taking in Robots." *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans* 35 (4): 460–470.
- Tsotsos, John K. 2011. *A Computational Perspective on Visual Attention*. Cambridge, MA: MIT Press.
- Tulving, Endel. 1972. "Episodic and Semantic Memory." *Organization of Memory* 1:381–403.
- Tulving, Endel. 1984. "Précis of Elements of Episodic Memory." *Behavioral and Brain Sciences* 7 (2): 223–238.
- Vernon, David. 2010. "Enaction as a Conceptual Framework for Development in Cognitive Robotics." *Paladyn Journal of Behavioral Robotics* 1 (2): 89–98.
- Vernon, David. 2014. *Artificial Cognitive Systems—a Primer*. Cambridge, MA: MIT Press.
- Vernon, David, Giorgio Metta, and Giulio Sandini. 2007a. "The iCub Cognitive Architecture: Interactive Development in a Humanoid Robot." In *2007 IEEE 6th International Conference on Development and Learning*, 122–127. New York: IEEE.
- Vernon, David, Giorgio Metta, and Giulio Sandini. 2007b. "A Survey of Artificial Cognitive Systems: Implications for the Autonomous Development of Mental Capabilities in Computational Agents." *IEEE Transactions on Evolutionary Computation* 11 (2): 151–180.
- Vernon, David, Claes von Hofsten, and Luciano Fadiga. 2011. *A Roadmap for Cognitive Development in Humanoid Robots*. Vol. 11. Berlin: Springer Science and Business Media.
- Vernon, David, Claes von Hofsten, and Luciano Fadiga. 2016. "Desiderata for Developmental Cognitive Architectures." *Biologically Inspired Cognitive Architectures* 18:116–127.
- von Hofsten, Claes. 2009. "Action, the Foundation for Cognitive Development." *Scandinavian Journal of Psychology* 50:617–623.
- Vygotsky, Lev S. 1978. *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
- Wiener, Norbert. 1948. *Cybernetics: Or the Control and Communication in the Animal and the Machine*. New York: John Wiley and Sons.
- Winkler, Jan, Georg Bartels, Lorenz Mösenlechner, and Michael Beetz. 2012. "Knowledge Enabled High-Level Task Abstraction and Execution." In *First Annual Conference on Advances in Cognitive Systems* 2 (1): 131–148.