The Adaptron Cognitive Architecture and Binary Neurons as a General–Purpose Representation

Brett N. Martensen (www.adaptroninc.com)
Adaptron Inc., Suite 302 – 40 Boteler St.
Ottawa, ON, Canada, K1N 9C8

Abstract

Adaptron is a cognitive architecture that is designed to control the intelligent behaviour of robots. It uses compositional hierarchies of binary neurons (binons) as its representational system. Binons are general-purpose, relational and functional nodes for representing knowledge, concepts and abilities. Adaptron satisfies many of the important requirements for artificial general intelligence. These requirements include purposeful, grounded, autonomous, general-purpose, scalable and reliable. Experiments have shown that this architecture is effective for the recognition of handwritten digits and Morse code as well as for the control of a simulated robot in a maze environment.

Adaptron interacts with its environment via senses and action devices. As it learns, it builds up integrated perception–action hierarchies of binons to represent its experiences. This mental model of its world is then used for thinking and rehearsing action outcomes. It is also used to control mental operations such as paying attention, selecting actions and reasoning. Binons are used to perform all of these operations.

A binon is a simple deterministic artificial neural node that represents a relationship. It contains an integer value used to help represent things and their relationships. It has links to two lower nodes and it is reused by zero or more upper nodes. Binons are general-purpose components that interact with each other like objects in object-oriented software. There are currently four types of binons.

- Name binons represent the category names for all types of things.
- Value binons are used to represent sense independent property values such as position, intensity, time and quantity plus properties derived from them.
- Entity binons represent types of things such as properties, objects, events and actions. An entity binon is a combination of a name and a value binon.
- Control binons are used to learn, manage and repeat behavioural and mental processes.

Binons can also be subdivided based on what role they play.

- Perception binons are used in recognition and prediction.

- Action and expectation binons are used for behavioural control. They are equivalent to command neurons in neuroscience, production rules in cognitive science, or the forward and inverse models in motor control.
- There are mental operation binons to focus attention, perform reasoning and initiate actions.

Adaptron starts with no knowledge or abilities. New binons are continuously created and integrated with existing ones to represent everything it learns. The resulting network is an overlapping composition of binary hierarchies. Learning takes place in five stages: reflexes, babbling, reuse, practice and automaticity. Novel experiences result from reflexes and babbling. These experiences become familiar and are learnt through reuse. They become more reliable through practice and can be performed as automatic habits. This is consistent with the dual process theory of cognition.

Table of Contents

Chapter 1 – Introduction

1.0 Introduction

I began thinking about artificial intelligence (AI) in 1968 during my second year of a B.Sc. program in physics at the University of British Columbia. I was sitting in the UBC computer science building trying to debug a FORTRAN program. Those were the days when all the software was submitted on decks of punch cards. UBC had a new IBM 360/67 mainframe computer. After getting help from a teaching assistant because I had spelt “integer” incorrectly (“interger”), I realized that the IF-THEN-ELSE statements were just like what the mind did when thinking. I started documenting my thoughts at the age of 17. I took up software engineering as a career after obtaining my M.Sc. in computer science from Queen’s University in Ontario, Canada in 1976.

Here it is 53 years later and AI has phased in and out of popularity, sophistication and success. When artificial intelligence research began, the idea was to develop software that could behave, think and reason like humans. I took on this original vision and my dedication to this objective has not wavered. However, because such general-purpose intelligent systems are so difficult to develop the AI field split into many separate sub-fields with techniques that are best suited for particular domains. For example:

- Symbolic AI applies logic programming, production rules and semantic nets in applications such as expert systems.
- Reinforcement learning has been effectively applied in game playing, from Chess, to Go and video games.
- Natural language processing has been most successful for language understanding and production.
- Evolutionary algorithms, which include genetic algorithms, use natural selection to find optimal solutions to problems.
- The AI sub-field of machine learning, which includes deep learning, has had great success recently in image and speech recognition. Today, it has become the latest high-tech tool of industry for finding and using patterns in large data sets.

There are now a growing number of people and organizations working to integrate these and other areas to achieve the original AI vision.

This book describes Adaptron and Binons (<u>Bin</u>ary neur<u>ons</u>). Adaptron is a cognitive architecture designed to control the intelligent behaviour of robots. It uses compositional hierarchies of binons as its representational system. Binons are general-purpose relational and functional nodes for representing knowledge, concepts and abilities. This book describes version 1.0 of Adaptron. As a software application, it satisfies many, but not all of the important requirements for artificial general intelligence (AGI). Many highlighted subject areas require further research. This is an ongoing project.

Adaptron is inspired by cognitive science, computer science, neuroscience, psychological and psychophysical theories and results. It is not inspired by many of the well-known machine learning or artificial intelligent approaches such as autoencoders, back-propagation, compression, fuzzy logic, genetic algorithms, kernels, Markov chains, transformers, high-dimensional vectors or neural networks such as Bayesian, convolutional, generative adversarial, long short-term memory, recurrent or self-organizing map networks.

The book is structured along the same lines as the artifacts produced in the development of software applications. Each chapter corresponds to a different artifact from the software development life cycle. These artifacts are the:

- Case study,
- General requirements,
- Solution requirements,
- Test plans and test cases,
- Software architecture,
- Detailed design and
- Software implementation

This breakdown is similar to the three levels of specification for information-processing systems as described by David Marr (Peebles and Cooper, 2015). Marr's three levels are the computational, algorithmic / representational, and implementation levels. In software development, these three correspond to solution requirements, software architecture/detailed design and software implementation. They also correspond to conceptual, logical and physical specification.

- Solution requirements are conceptual. They describe the problem from the three views of <u>what</u> a solution needs to do, <u>what</u> it needs to know and <u>what</u> interfaces it needs to use.
- Software architecture/detailed designs are logical. They describe <u>how</u> a solution should do what it does (the processes/algorithms), <u>how</u> it stores what it knows (the data structures) and <u>how</u> it should interface to its environment.
- Software implementation is physical. It describes a solution in software.

There is a rule of thumb in software development that the first version of a deliverable that describes anything complex is only about 60% correct or complete. The remaining 40% is either incorrect or missing. The rule of thumb goes on to say that when we iteratively revise software deliverables we correct approximately 60% of the remainder. This means the second version is about 85% complete and a third iteration usually achieves a 95% or greater correct and complete status. It is impossible to reach 100% for anything complex. The chapters in this book are no exception to the rule.

This introductory chapter contains an executive summary for those who do not wish to read all the details about Adaptron and binons. It also provides definitions of some key terms as used in the book. The case for AGI in Chapter 2 describes the problem and proposes Adaptron as a possible solution. The general AGI requirements in Chapter 3 provide the context / environment in which the solution must operate. In the case of Adaptron, this is the real world. It also lists some of Adaptron's desirable features such as intelligent and domain independent. The 13 solution requirements in Chapter 4 are far more detailed. They consist of the functional requirements for learning how to behave and reason plus the non-functional requirements. This chapter highlights what is important and what is not so important. Chapter 5 describes test plans and test cases. It mentions the Turing test (Turing, 1950) but focuses more on other useful testing approaches. The software architecture in Chapter 6 describes the Adaptron cognitive architecture. It describes how the architecture addresses the solution requirements using its dual subsystem architecture. The detailed design in Chapter 7 describes the structure of binons. It also describes how they are learnt and used to control behaviour and reasoning. This includes the processes of perception, action and mental concentration. Throughout the chapter, there are Principles of OPeration (POPs) for the structuring and functioning of binons. Implementation considerations are covered in Chapter 8 and the first experimental results are presented in Chapter 9. Chapter 10 concludes with a discussion of the limitations of Adaptron and binons. It also discusses outstanding questions and possible future directions for research.

1.1 Executive Summary

This book specifies the architecture and design for a software application that can be used in robots to allow them to learn and act intelligently. It is not intended to entertain. Instead, it is meant to be educational, technical and detailed. Like this book, this summary is not meant for the average reader. It is rich in the terminology of computer science, artificial intelligence and cognitive science.

<u>Adaptron</u>

Adaptron is a cognitive architecture designed for use in artificial general intelligent (AGI) agents such as robots. The objective is to build intelligent agents that will assist humanity by performing tasks that will help us achieve our goals. They will learn to communicate with us and interact with the world to humanity's benefit. They will think about our problems and goals and suggest solutions to address them. They will become intelligent assistants able to augment our human abilities.

<u>The case for AGI</u>

Just like any technology, as simple as a hammer or sophisticated as atomic energy, AGI agents can be used for good or evil. Benefits such as expanding the human knowledge base and providing fresh insights and solutions in areas such as medicine, culture, finance, law and technology are some examples. Many of humanity's problems may be solved through the intelligent application of resources. However, some might say the risks of AGI agents outweigh the rewards. Such risks include, loss of jobs, losing control of the agents and their abuse by some to the detriment of others.

<u>General requirements</u>

As software, Adaptron is designed to work inside an agent's body. It interacts with an environment by monitoring senses and controlling action devices. To accomplish its goals successfully, Adaptron needs to operate in worlds that are safe, systematic and dynamic to an acceptable degree. General features of Adaptron include domain independent, reliable, able to continuously learn and reason. These and many more are detailed in the solution requirements.

<u>Solution requirements</u>

Adaptron satisfies many of the important requirements for artificial general intelligence. These requirements are purposeful, grounded, autonomous, controlled and safe, general-purpose, adaptable, transparent, scalable, efficient, robust, reliable and able to behave and reason. There are other human features, that could be considered important for an AGI agent, but they are out of scope in the current version of Adaptron. Examples include two and three-dimensional perception, emotions and forgetting. Adaptron's current goals are built into its architecture. They are solely based on intrinsic motivation. These goals are the pursuit of novelty, avoidance of boredom, achieving reliability in perception and action and minimizing uncertainty. These goals result in curiosity and practice. Curiosity produces exploratory behaviour and practice improves the reliability and certainty of behaviour. Goals based on extrinsic motivation, those that are pleasant and rewarding or those that are unpleasant and punishing, which result in exploitation behaviour, are not included in Adaptron's current design.

Testing

All software needs to be tested to make sure it meets the requirements. Many approaches for testing AGI agents are discussed in Chapter 5. The discussion mentions the Turing test but the major focus is on using simulated worlds as better test environments. Once validated, AGI software can be embedded in robots for testing in the real world.

Software architecture

Adaptron's architecture is divided into two subsystems, the mental and behavioural ones as shown in Figure 1.1. Both subsystems contain artificial neural networks of binary neurons (binons). The subsystems are structured as compositional hierarchies and they function in a similar way. They mainly differ in their sources of information and interface devices. The behavioural subsystem contains integrated multi-layer perception–action hierarchies that retain experiences and control action performance. They interact with the environment via senses and action devices. The subsystem builds up a model of the world as Adaptron experiences it. It is a memory of what has been perceived and what has been done. The perception process senses stimuli from the sensors to recognize objects and events while the action process performs action sequences and produces them on action devices. Adaptron can be configured to handle a wide variety of sensor and actuator types.

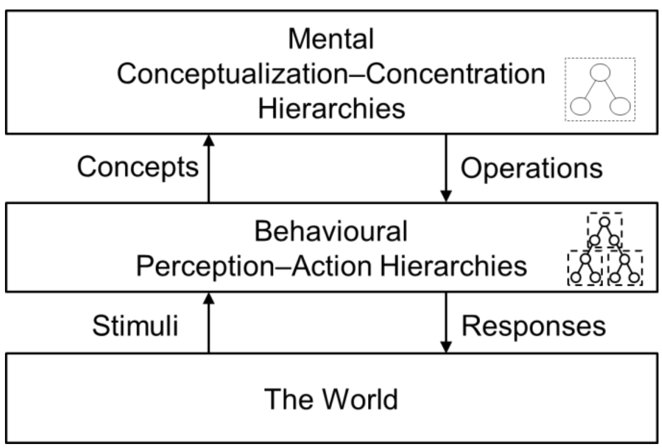

Figure 1.1 – Adaptron's software architecture.

The mental subsystem contains binon hierarchies that learn to reason and direct behaviour. It interacts with the binons in the behavioural subsystem. It recalls memories from the behavioural subsystem as concepts for reasoning and imagining new ones. This includes determining the outcomes of actions. It then performs mental operations on the behavioural subsystem for selecting and enabling action sequences while paying attention to the results.

Detailed design - Binons

Binons (binary neurons) are the software components that make up Adaptron's two subsystems. They are general-purpose representational components that interact as objects in an object-oriented fashion. Binons are combined to form multi-layer compositional hierarchies to represent the knowledge, abilities and concepts of an agent. The perception and action hierarchies are tightly integrated in parallel at all levels of complexity. Senses and action devices are at the bottom of this structure. Activation of perception binons is feedforward, up the hierarchies and action performance is in the opposite direction, down the hierarchies. There are 15 Principles of OPeration (POPs) that govern how binons work and how they are structurally related.

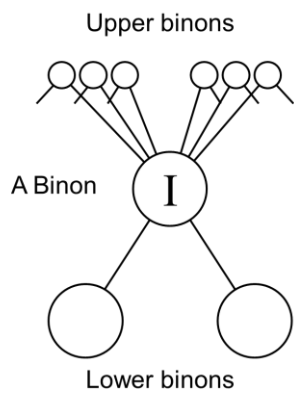

Figure 1.2 – Binons

As illustrated in Figure 1.2, structurally, a binon is a simple deterministic node that represents a relationship. It contains an integer value used to help represent things and their relationships. It is linked to two lower nodes and may be reused and shared by zero or more upper nodes. The resulting network is an overlapping composition of binary hierarchies. There are no weights on the network links. Binons are novel when first created and become familiar on their first activation. They must be familiar before they can be linked together to form upper nodes. Binon hierarchies grow from the bottom up as new things are learnt. Binons are either spatial or temporal. Spatial binons represent things whose parts or properties occur simultaneously whereas temporal binons represent things whose features occur sequentially or represent controllers for sequencing acts and thoughts. Temporal binons represent causal relationships.

<u>Types of Binons</u>
The four types of binons are name, value, entity and control binons.

- Name binons represent the category names for all types of things.
- Value binons are used to represent amodal (i.e. not sense or action device specific) property values of things such as the core properties of position, intensity, time and quantity plus properties derived from them.
- Entity binons represent types of things such as properties, objects, events and actions. An entity binon is a combination of a name and a value binon.
- Control binons are used to learn, manage and repeat behavioural and mental processes.

<u>Roles played by binons</u>

As a general-purpose representation mechanism, binons can be used to capture the core properties experienced from sensors and those needed to drive actuators. Binons can also be used to represent properties derived from the core properties, such as distance, size, duration, speed, delay, density, frequency, expansion etc. These can be combined to represent objects and events experienced on multiple senses and acts performed on multiple action devices.

Action control and expectation binons are behavioural control binons. Action control binons associate trigger situations with acts that can be performed (i.e. affordances in cognitive science). Expectation binons associate actions with their perceptual results. The relationships between context and resulting situations are captured in temporal prediction binons. Together these three binons represent the forward and inverse models in motor control. This is called a simple action habit in Adaptron. Simple action habits are grouped into binary hierarchies to represent tasks. Lower level action habits are reused and shared by more complex, higher-level tasks.

Adaptron recognizes the difference between events it caused (called achievable events) and those it did not cause (called incidental events). Events caused by others are regarded as incidental by Adaptron because of its egocentric perspective. Sequences of events are represented as prediction binons. They allow it to make decisions about what actions to perform by recalling achievable events that have interesting results. It can then activate action habits that will cause these events.

<u>Learning</u>
Learning of behaviour and reasoning in Adaptron takes place in five stages: reflexes, babbling, reuse, practice and automaticity. Reflexes and action babbling cause the performance of involuntary acts and the orienting of attention. Such acts are reused when Adaptron is bored because nothing new is happening and then practiced until they reliably reproduce the same results. They are then habits that can be done automatically. Once habits are activated, they continue operating without the need for conscious attention. An action habit continues as long as its intermediate results contain the triggers for subsequent steps. Perception also becomes automatic as soon as objects and events become familiar. This is consistent with the dual process theory of cognition, which distinguishes between explicit/reasoning (system 2) and implicit/automatic (system 1) processes.

<u>Mental operations</u>
Mental operation and control binons are in the mental subsystem. They focus attention, initiate actions and perform reasoning. Focusing attention on the senses is done by the priming operation while initiating action is done by the enabling operation. Enabling is done on an interesting result, which then activates all the action habits that can achieve that result via their expectation binons. Reasoning is a search process that focusses attention on memories to determine the possible consequences of acts before doing them. It is done by a series of recall operations. Recall uses the behavioural model of the world to produce real and imaginary concepts. As in perception, the process of conceptualization of real and imaginary concepts becomes a habit.

Attention and mental processes

The attention mechanism provides a single point of focus for the awareness of things as recognized on the senses or the awareness of concepts in memory. Attention can be distracted during reasoning by unexpected stimuli. Focusing of attention provides for metacognitive feelings to be experienced. They are used for learning and controlling mental processes. Examples of metacognitive feelings are the tip of the tongue feeling, the knowledge that you do not know something, knowing that you can or cannot do something, and the feelings of success or failure in reaching a goal. Mental processes search memories for novel and interesting experiences and initiate actions to re-experience them. By recalling specific subjects and the associated metacognitive feelings, Adaptron is able to perform deductive reasoning and problem solving.

Implementation

Although binons are designed using an object-oriented software approach, not all of its implementation is object-oriented. To identify the properties of percepts in the perceptual field a short-term memory (STM) is used. This is because the same recognition binons may repeat and/or be found at multiple locations on the sensory arrays.

Empirical evidence

A number of applications have been developed using the Adaptron architecture.

- The Perceptra application is proof that binons can be used to recognize and classify handwritten digits.
- The Morse code application shows that binons can be used for prediction.
- The Smarty robot simulation is proof that binons can be used to represent perception–action associations and control a robot in a maze environment.

Next steps include changing the Smarty test environment by adding objects with dynamic behaviour and applying the complete set of principles of operation. The use of binons for reasoning has yet to be implemented or experimentally proven.

Conclusions

Obviously, more experiments are required to show that the complete Adaptron cognitive architecture is feasible. Its design will evolve as additional, possibly simpler, more efficient and general principles of operation are discovered. There remain many unanswered design questions.

1.2 Definitions

In software development, the requirements and designs not only need to be correct and complete but also unambiguous. That means that the terminology used must be clearly defined, consistently used and understandable. This section provides definitions for many of the terms as used in Adaptron's description. For reference purposes, useful definitions of underlined terminology can be found in Wikipedia (www.wikipedia.org). Note that not all terms used are consistent with commonly understood meanings.

Artificial General Intelligence

The primary objective of Adaptron is to animate an Artificial General Intelligent (AGI) agent (Goertzel, 2014). As its name suggests, an AGI agent is a general-purpose agent able to learn and perform tasks that humans can do. It is also known as "Strong Artificial Intelligence (AI)". On a hypothetical line of increasing intelligence in Figure 1.3 an AGI agent is more intelligent than non-human animals but less intelligent than an Artificial Super Intelligence (ASI) (Bostrom, 2014; Legg, 2008). There is no clear boundary between these categories. At the top of this intelligence range is Universal Intelligence (AIXI) (Legg and Hutter, 2007). It is not expected that all of Adaptron's knowledge and abilities will lie within the AGI range. Some will be below AGI and some will be in the ASI range. The objective is to have a majority of its capabilities in the human range, but it is not expected to be an exact duplicate of human intelligence.

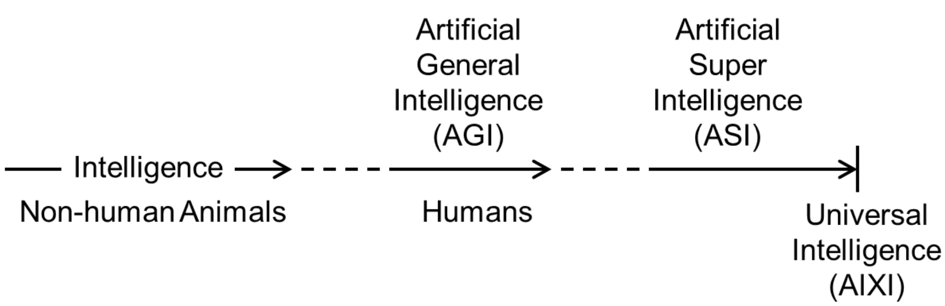

Figure 1.3 – Ranges of intelligence

The evolution of Artificial Intelligence (AI)

The evolution of artificial intelligence is following a similar course to that of human flight. Birds flew long before flying machines did and humans have acted intelligently long before computers existed. This evolution can be divided into five simple phases: imagination, simulation, specialization, reproduction, and refinement. Right now, AI is in the third phase of this evolution: specialization.

Imagination phase

The imagination phase is made up of myths and legends. The story of Icarus is the most well-known one for human flight. In artificial intelligence the Greek myth of Talos is probably the oldest. More modern AI stories include Frankenstein and Rossum's Universal Robots.

Simulation phase

In the simulation phase of evolution, the initial attempts at human flight and artificial intelligence involved the design and/or building of something that might work by directly recreating the phenomena. In flight there were attempts like tower jumping with attached wings and designs like Leonardo da Vinci's ornithopter. Many of the last unsuccessful attempts at heavier than air human flight in the 19th century were attempted using steam engines as the power source. In AI there were attempts to build automata that appeared intelligent such as Vaucanson's mechanical duck, Pierre Jaquet-Droz's automata. Charles Babbage's analytical engine was one of the first attempts at building a general purpose computing device.

Specialization phase

The specialization phase of evolution is made up of successful attempts to duplicate one or more aspects of the phenomena but not put it all together in an integrated and working whole. This phase also includes the discovery of important principles and theories. For example, Sir George Cayley described the principle of how an air foil produces lift and went on to build a manned glider based on this principle. Successful heavier than air aviation has been achieved in many special forms. Examples included kites, gliders, hot-air balloons, airships and rockets, all of which work better than birds but only in narrow ways. They might fly faster than birds but not necessarily further or they may fly further but necessarily higher. Philosophers and scientists have addressed theories of intelligence for centuries; however the most recent pre-computer example is that of the Turing-machine. The advent of functioning computers accelerated AI theories and the design and development of successful AI applications. We now have machine learning techniques such as deep learning, transformers, reinforcement learning, genetic algorithms, recurrent neural networks and other you-name-it techniques, all of which work as well as or better than humans but only in very narrow domains.

Reproduction phase

The reproduction phase of evolution is characterized by the development of a system that meets the basic requirements. In the case of flight the basic requirements were to achieve sustained, powered human flight. The goal was just to get it to work, not to be any better than birds. The Wright brothers are well known for achieving this requirement. In artificial intelligence, this phase will be met when software is able to learn and perform intelligent behaviour and reasoning. But more specifically, it will be general purpose, able to adapt its behaviour and reasoning across a broad set of domains. Just as the Wright flyer could not fly as high or as fast as birds, so too the first AGI agent is not expected to behave and reason as well as humans. For example, it may only have the communication skills of a dog.

Refinement phase

The refinement phase of evolution is one in which capabilities are improved and new ones added to make the system work as well as and then exceed avian flight or human intelligence. In flight, features such as ailerons and jet engines have allowed us to reach "super-bird" performance. More recent aviation advances have been achieved by using winglets and composite materials. The same is bound to happen in the field of AGI at which point the artificial super intelligence (ASI) level will be achieved.

Representational Systems

Adaptron is a mental representational system (Gallistel, 2001). The design of an effective representational system is the defining challenge for a cognitive architecture. The general definition of a system is something composed of parts. The parts are combined and function together as a whole. A system interacts with its environment via its interfaces, which are input and output devices. More specifically, in artificial intelligence and cognitive science an agent's representational system is a system that captures a model of its world (Tonneau, 2011). It represents this model in the form of knowledge and abilities that it gains from experience while interacting with its environment. Knowledge and abilities are both spatial and temporal. They can be used to reliably predict and control behavioural and mental processes (Conant and Ashby, 1970). In Adaptron, the term "behaviour" encompasses both perception and action but not reasoning or thinking. The subject of behavioural and mental representations in cognitive science has a long and sometimes controversial history (Mirolli, 2012).

1.2.1 Cognitive Architectures

A cognitive architecture (Thagard, 2012) is a man-made representational system (Doumas and Hummel, 2005) built to achieve specific goals. These goals are to model intelligent behaviour and mental processes. Cognitive architectures have primarily been used to understand the human mind by creating, predicting or reproducing psychological or neurological theories, behavioural data, brain regions and activity from fMRI data (Andrea Stocco, Catherine Sibert, Zoe Steine-Hanson, Natalie Koh, John E Laird, Christian J Lebiere, Paul Rosenbloom
Analysis of the human connectome data supports the notion of a "Common Model of Cognition" for human and human-like intelligence across domains
Neuroimage. 2021 Apr 7;235:118035.
). Although such theories and facts have served as inspiration for the design of Adaptron, it is not currently designed for this purpose.

A cognitive architecture is an essential part of any intelligent agent that is expected to achieve artificial general intelligence. There are a wide variety (Bengio et al., 2013) and a long history (McCarthy, 1987) of representational systems that have been used in the development of cognitive architectures (Dong and Franklin, 2014; Duch et al., 2008; Kotseruba and Tsotsos, 2018; Langley, 2017; Samsonovich, 2010). This section is not meant to be a comparison of cognitive architectures. However, it is interesting to note the variety of representational systems that have been used in them (Peebles, 2017). Some examples of cognitive architectures are:

- ACT-R – Adaptive Control of Thought – Rational (Stewart and West, 2006),
- APNN – Associative-Projective Neural Networks (Rachkovskij et al., 2013),
- CHREST – Chunk Hierarchy and Retrieval Structures (Gobert and Lane, 2017),
- CLARION – Connectionist Learning with Adaptive Rule Induction ON-line (Sun, 2006; Sun et al., 2006),
- EPIC – Executive Process-Interactive Control (Kieras, 2004),

- ICARUS (Choia and Langley, 2018),
- LIDA – Learning Intelligent Decision Agent (Faghihi and Franklin, 2012),
- NARS – Non-Axiomatic Reasoning System (Hammer et al., 2016),
- Sigma (Rosenbloom and Demski, 2016), and
- Soar (Laird, 2012).

ACT-R, EPIC and Soar (Laird et al., 2017, Kieras and Meyer, 1997) use production rules as their representational component. Production rules are expressions with a conditional and action part. If the condition is true then the action is performed. Production rules are applied to symbolic structures and information is kept in a working memory.

APNN is a cognitive architecture design:

> "… that works with an original scheme of sparse binary distributed representations to construct world models of varied complexity required for both task-specific and more general cognitive modeling. APNNs provide scalability and flexibility due to a number of design features. Internal representations of APNNs are sparse binary vectors of fixed dimensionality for items of various complexity and generality. Representations of input scalars, vectors, or compositional relational structures are constructed on-the-fly, so that similar items produce representations similar in terms of vector dot-products" (Rachkovskij et al., 2013).

CHREST uses a long-term memory (LTM) of chunks to symbolically represent information obtained from verbal and visual input/output units. A discrimination network is used to index into and retrieve information from the LTM. It is based on the concepts of limited attention and a limited short-term memory (STM).

CLARION uses a dual representation for symbolic (explicit) and sub-symbolic (implicit) data.

> "Symbolic knowledge is captured with data structures called rules and chunks, while sub-symbolic knowledge is encoded in connectionist networks" (Thórisson and Helgasson, 2012).

ICARUS uses a variety of data structures to represent concepts, percepts, skills, beliefs and goals. Conceptual and skill memories contain a set of hierarchically organized logical rules represented as clauses. Each clause has a head containing the concept's name and arguments or the skill's objective. Its body describes the conditions under which the concept or skill applies. For a skill, the clause's body describes the ordered actions or sub-goals needed to achieve the skill's goal (Stracuzzi et al., 2009).

LIDA (Franklin et al., 2016) uses a general-purpose nodes and links representation. Nodes represent features, objects, feelings, actions, events, categories etc. Links are the different kinds of relationships between the nodes such as features-of, causation, category membership etc. Vector LIDA (Agrawal et al., 2018) is a major overhaul of the original representation system in LIDA. Vectors are an ordered set of integer values that represent something in a multidimensional space. LIDA divides memory into separate modules for spatial, perceptual associative, transient episodic, declarative, sensory motor and procedural memory.

NARS uses the Narsese language to represent goal, question and belief statements. It is a term-oriented formal language. A set of inference rules for reasoning under uncertainty is used to process the statements (Thórisson and Helgasson, 2012).

Sigma uses a graphical model to represent knowledge in a general-purpose unified cognitive architecture. Nodes in the graph represent variables for functions and the factors into which functions are decomposed. Variable nodes are connected to the factor nodes that use them.

> "Graphical models in general provide an efficient means of computing with complex multivariate functions by decomposing them into products of simpler functions and then translating them into graphs." (Rosenbloom and Demski, 2016)

Adaptron is a cognitive architecture that uses hierarchies of binary neurons (binons) as its representational system. Binons share many features with other representational approaches found in:

- Structured connectionist models (Shastri, 2003),
- Symbolic connectionist models such as LISA – Learning and Inference with Schemas and Analogies (Hummel and Holyoak, 1997 and 2003) and its descendant DORA – Discovery of Relations by Analogy (Doumas et al., 2008; Doumas and Martin, 2018),
- Perceptual symbol systems (Barsalou, 1999),
- Schema Theory (schemata have representational properties required for perception–action coupling) (Arbib, 2003),
- Language of Thought (it describes the properties of representations as mental states required for concepts in thinking) (Schneider, 2009; Schneider, 2011),
- Neuro-symbolic approaches (integration of neural machine learning with symbolic knowledge representation and reasoning) (D'Avila Garcez et al., 2019; Besold, et al 2017; Garnelo et al., 2016;) and
- Formal concept analysis (Aswani Kumar et al, 2015; Priss, 2019)

It is most likely that the best cognitive architecture for an AGI agent will incorporate all or part of these approaches in one form or the other (Barsalou, 2012; Domingos, 2015).

1.2.2 Binons and other things

A binon is an artificial neural node (artificial neuron) that represents a relationship between its two lower binons (Halford et al., 2010), as illustrated in the Figure 1.4 binon structure diagram. It is the simplest structure that can be used to represent a relationship. It is reused and shared by many upper binons to represent relationships that are more complex. At first glance they may be mistaken for an AND gate in digital electronics but they are far more than that. They contain an integer value that captures the relationship property between the two lower binons. They are discrete, uniquely identifiable representations and therefore symbolic. They are mental representations or mental symbols. When grounded on sensors and actuators and then combined they can represent all manner of things and their types, such as values, properties, percepts, objects, events, acts, thoughts and relationships between them. Binons are also functional components that perform functions when stimulated. That stimulation is not always from lower binons to upper binons. For action, stimulation flows top down and for controlling behaviour; it flows left to right or vice versa. Therefore, they are quite different from AND gates and biological neurons. They are created, learnt and performed. They are mechanisms that are programmed to function as perception, action and mental processes. Because they have identities, values and functionality, they are similar to objects, attributes, and classes in object-oriented software.

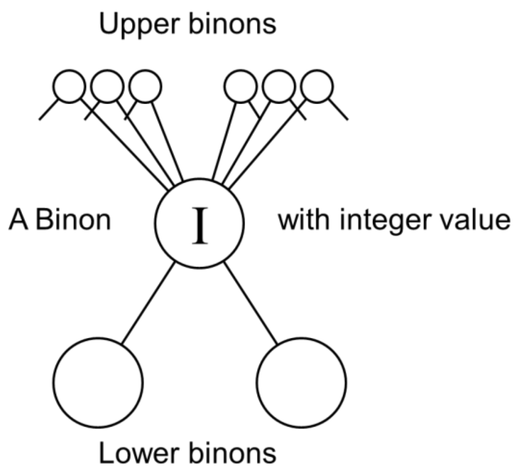

Figure 1.4 – Binon structure diagram

Things

Given that binons can represent all kinds of things, it is important to define the term "thing". Throughout this book, I use the words thing, something, nothing and anything. Things include everything in the world that is possible to know, imagine or talk about. This includes tangible and abstract entities. Things include primitive things such as values, properties, attributes, parameters, characteristics, categories and features. Visible and touchable objects are things, as are tastes, odors and sounds. Things include instantaneous events that occur and actions that have and can be done. They include time, feelings, concepts, mental representations and imaginary things. Things also include processes performed, whether mental, mechanical, electrical or chemical.

Terminology

The correspondence between terms used for things in the real world and the terms used for the mental representations (binons) in Adaptron is captured in Table 1. Words we use in normal conversation are often ambiguous. For example, the term "event" can be understood two ways, either as an instantaneous event like a flash of light or having duration like a birthday party. In this book, events are of the instantaneous kind. Instantaneous events occur at the boundaries of events that have duration. Events are like edges in time.

Things in the world	Represented as binons
Attributes, characteristics, features	Properties
Objects, parts, sounds, odors, events, etc.	Percepts
Event sequences, sentences	Perception sequences
Actions, movements, speech, etc.	Acts
Skills, abilities, tasks	Action habits
Thoughts, Ideas	Concepts
Mental processes	Operations

Table 1. Things in the world and Adaptron's corresponding mental representations

Properties, objects and events

It is useful to divide things into properties, objects and events because we experience properties, interact with objects and participate in events.

- Properties have values that are measured or set. They do not exist in or require any space because they are pieces of information. However, they do exist in time.
- Objects on the other hand exist in space and time. They are spatial things that last.
- Events are changes that happen to objects. They only exist in time and are therefore temporal things.

The only way we can identify objects and events is via the properties they have. More than one property is required to describe them. Objects may take up space because they are solid, liquid, and gaseous or plasma, but they do not need to be composed of atoms or particles. For example, electric and magnetic fields are objects. Objects take up time and have a lifecycle that is composed of events. They are created, live and may be destroyed. During their life, they may be in different states, which are described by their properties. Events occur when the states of objects change. Objects may move in space. Examples are pressure waves such as sounds and ripples on a pond. Objects may have structure and be composed of parts, which are also objects, such as a car is composed of doors, windows, wheels and an engine. Alternatively, objects may be composed of the same substance throughout such as glass, clay, air, water, quartz or mercury. They may be amorphous or crystalline. These kinds of objects invariably have shapes and surfaces that are recognized by their contours and shading properties. The shapes of objects can change but they may not have a well-defined boundary, for example, clouds are objects. Organizational units such as a company are also objects. They are human engineered collections of objects that function together as a system as defined earlier. We may not always know where in space an object exists. Abstract objects are a good example. They are the product of human imagination and exist in our minds. Concepts are also objects. They are mental representations that we name and can identify because of their properties.

Types of things

However, properties, objects and events (i.e. all things) are identifiable and belong to categories (also known as types, kinds or classes). Objects, as well as events are of the same type when they have common features. In Adaptron features are properties, parts and relations. Objects and events can be described based on the values of their properties (e.g. size), the parts of which they are composed (e.g. a car has doors) and/or the structural and causal relations they have with other objects and events (e.g. a chair is beside a table and animals eat food). Features are combined to form patterns. A pattern is a structural arrangement formed from the relationships between things. Patterns are groupings or combinations of things that are interrelated and often repeat. In Adaptron, patterns of properties are used to represent objects and events. Actions are a particular type of event. They are events caused by agents. In addition, a process is a sequence of actions that transforms the state of objects. This means that a process is a sequence of events.

Experience

The word experience is used extensively in this book. An experience is composed of the stimuli detected by Adaptron during a period of time that is long enough for it to detect properties and identify objects and events. An experience consists of transient information. A situation is part of an experience. It is the sum of Adaptron's stimuli, which are perceived at a particular time and are constant for a given duration. Types of situations include context situations that exist before performing acts and resulting situations that exist after acts have been done.

Actions and goals

Note that an action, as an outward flowing response, is not part of an experience. The experience only includes an action's context and resulting situations. This combination of situations provides for action perception and recognition. A resulting situation is the final product of an action. It is also called the outcome of an action. However, the outcome is not the goal of an action. The goal of an action is to achieve a change in state of one or more objects. This means goals are achievable events. To make this point clearer, consider a banana. A banana is not the goal of an action. Instead, "having a banana" is a goal. The verb "having" is necessary because it indicates the action that is needed to change the state of the banana from "not having" to "having". Similarly, a clean car is not a goal. The goal is to change the car's state from dirty to clean by cleaning it. A clean car is the result of the cleaning action.

Percepts

A percept in Adaptron is a grounded mental representation of a type of object or event formed or identified in the process of sensing, recognizing and encoding stimuli (i.e. the process of perception). Note that a percept represents a type of object or event, not an actual specific object or event. A percept may represent something as simple as a sensor specific property or something as complex as any combination of or relationship between these properties. Examples of sensor specific properties of object's or event's include: position, intensity, colour, time, duration or quantity, and types of bodily properties, such as, pain, hunger and thirst. Examples of things represented by combinations of properties include types of physical objects, motions, event sequences, perceived actions, sounds, or words etc. In Adaptron, percepts are divided into its incidental and achievable percepts. Its incidental percepts are those that are not the result of its own actions. They may happen as part of nature or be generated by another agent's actions. Adaptron's achievable percepts are the results of events caused by its own actions. This means that an event caused by the action of another agent is an incidental percept from Adaptron's egocentric perspective.

Perception

When I first started working in the field of artificial intelligence, I thought perception meant visual observation. However, I soon learnt that perception includes all manner of stimulus recognition. It can be divided into exteroception (external perception) and interoception (internal perception). Exteroception includes vision for seeing, audition for listening, haptic and tactile perception for touch, gustation for taste and olfaction for smell. Interoception includes chronoception for time perception, proprioception for muscles and body configuration, balance and spatial orientation from the vestibular system, nociception for pain and visceroception as the perception of the internal organs such as the heart, lungs, bladder, stomach, and bones.

Acts

An act in Adaptron is a grounded mental representation of a response formed or executed in the process of actuating and performing actions. Note that an act is not a type of thing; it is an actual thing with specific property values. It may represent something as simple as the intensity value to be set at a specific position on a type of actuator of an action device or as complex as a combination of intensity values to be set across multiple types of actuators on multiple action devices. Depending on the type of device and actuator, the intensity may represent force, temperature, speed, loudness or brightness etc. An act may specify a movement, a projection on a display, a sound as produced during speech or the control information required for producing something such as adrenaline etc.

Concepts

A concept in Adaptron is a percept that has been recalled in the process of reasoning. Concepts are often referred to as mental images, ideas or thoughts. Mental images may be visual, auditory or olfactory etc. Real concepts are based on percepts that have been experienced via the senses while imaginary ones have been produced by combining real ones during reasoning.

1.3 Acknowledgements

I am grateful to the following for their helpful comments on early versions of this book: Michael Roche, and [other names here]. In addition, I thank those who analyzed my work and helped me to think through the subjects. They are [names here]. Above all, I am grateful for the many brilliant scientists that have studied and written about subjects related to the design of artificial general intelligence. Their papers and articles are referenced throughout this book. However, special thanks go to my wife, Ann Ketcheson, who has shown unending patience and provided me with invaluable suggestions.

Chapter 2 – The Case for AGI

Chapter 3 – General Requirements

Chapter 4 – The Solution Requirements

Chapter 5 – Testing

Chapter 6 – The Software Architecture

Chapter 7 – Detailed Software Design

Chapter 8 – Implementation

Chapter 9 – Experimental Results

Chapter 10 – Conclusions

References

Abbott, J. T., Austerweil, J. L., and Griffiths, T. L. 2015. Random walks on semantic networks can resemble optimal foraging. *Psychological Review.* 122(3):558–569.

Agrawal, P., Franklin, S. and Snaider, J. 2018. Sensory Memory for Grounded Representations in a Cognitive Architecture. *Sixth Annual Conference on Advances in Cognitive Systems (ACS'18)*. Poster Collection 1–18.

Ahmed, O., Träuble, F., Goyal, A., Neitz, A., Wüthrich, M., Bengio, Y., Schölkopf, B. and Bauer, S. 2020. CausalWorld: A Robotic Manipulation Benchmark for Causal Structure and Transfer Learning. *arXiv:2010.04296 [cs.RO]*

Albus, J. S. 1996. The Engineering of Mind. *From Animals to Animats 4 – Proceedings of the fourth International Conference on Simulation of Adaptive Behavior (SAB'96)* 23-32.

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J. and Mané, D. 2016. *Concrete Problems in AI Safety.* arXiv:1606.06565v2 [cs.AI]

Anderson, J. R. 1978. Arguments concerning representations for mental imagery. *Psychological Review*. 85(4):249–277.

Anderson, J. R. and Lebiere, C. 2003. The Newell Test for a Theory of Cognition. *Behavioral and Brain Sciences*. 26(5);587-601.

Andrews, G. and Halford, G. S. 2002. A cognitive complexity metric applied to cognitive development. *Cognitive Psychology.* 45:153–219.

Arbib M. A. 2003. Schema Theory. In Arbib, M. A. (Ed.), *The Handbook of Brain Theory and Neural Networks – 2nd ed.* Cambridge, Massachusetts: MIT Press. 993-998.

Arbib, M. A., 1981. Perceptual structures and distributed motor control, in *Handbook of Physiology—The Nervous System II. Motor Control* (V. B. Brooks, Ed.), Bethesda, MD: American Physiological Society. 1449–1480.

Aswani Kumar, C., Ishwarya, M. S. and Loo, C. K. 2015. Formal concept analysis approach to cognitive functionalities of bidirectional associative memory. *Biologically Inspired Cognitive Architectures*. 12:20-33.

Balci, F., & Gallistel, C. R. 2006. Cross-domain transfer of quantitative discriminations: Is it all a matter of proportion? *Psychonomic Bulletin & Review*. 13(4):636–642.

Baldassarre, G. 2011. What are intrinsic motivations? A biological perspective. *2011 IEEE International Conference on Development and Learning (ICDL)*, Frankfurt am Main, 1-8.
Baldassarre, G. and Mirolli, M. 1998. *Intrinsically Motivated Learning in Natural and Artificial Systems*. Springer, Publishers.

Baldassarre, G., Lord, W., Granato, G. and Santucci, V. G. 2019. An Embodied Agent Learning Affordances with Intrinsic Motivations and Solving Extrinsic Tasks with Attention and One-Step Planning. *Frontiers in Neurorobotics*. 13(45):1-26.

Balleine, B. W. and Dickinson, A. 1998. Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology*, 37(4-5):407–419.

Barsalou, L. W. 2003. Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences.* 358(1435):1177–1187.

Barsalou, L. W. 2009. Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Science*. 364(1521):1281-1289.

Barsalou, L. W. 2010. Grounded Cognition: Past, Present, and Future. *Topics in Cognitive Science* 2(4):716-724.

Barsalou, L. W. 2012. The human conceptual system. In M. J. Spivey, K. McRae, & M. F. Joanisse (Eds.), *The Cambridge Handbook of Psycholinguistics.* 12:239–258. Cambridge University Press.

Barsalou, L.W., 1999. Perceptual symbol systems. *Behavioral and Brain Sciences*. 22, 577–660.

Barsalou, L.W., 2008. Grounded Cognition. *Annual Review of Psychology.* 59, 617–645.

Barto, A. G. 2003. Reinforcement Learning in Motor Control. In Arbib, M. A. (Ed.), *The Handbook of Brain Theory and Neural Networks – 2nd ed.* Cambridge, Massachusetts: MIT Press. 968-972.

Barto, A. G., Mirolli, M. and Baldassarre. G. 2013. Novelty or Surprise? *Frontiers in Psychology*. 4(907):1-15.

Barto, A. G., Singh, S, and Chentanez, N. 2004. Intrinsically Motivated Learning of Hierarchical Collections of Skills. in *Proceedings of the 3rd International Conference on Developmental Learning (ICDL '04)*., LaJolla, CA, 112–119.

Bengio, Y. 2017. The Consciousness Prior. *Computing Research Repository (CoRR).* arXiv:1709.08568v2 [cs:LG]

Bengio, Y. 2018. From Deep Learning of Disentangled Representations to Higher-level Cognition. *Microsoft Research*. 1:07:31.

Bengio, Y., Courville, A. and Vincent, P. 2013. Representation Learning: A Review and New Perspectives. IEEE Transactions on Software Engineering 35(8):1798-1828.

Bergen, B. K. 2012. *Louder Than Words: The New Science of How the Mind Makes Meaning*. New York, NY: Basic Books.

Besold, T. R., d'Avila Garcez, A. S., Bader, S., Bowman, H., Domingos, P., Hitzler, P., Kuehnberger, K.-U., Lamb, L. C., Lowd, D., Lima, P. M. V., de Penning, L., Pinkas, G., Poon, H., and Zaverucha, G. 2017. Neural-Symbolic Learning and Reasoning: A Survey and Interpretation. *Computing Research Repository (CoRR).* arXiv:1711.03902v1 [cs.AI]

Beyret, B., Hernández-Orallo, J., Cheke, L., Halina, M., Shanahan, M. and Crosby, M. 2019. The Animal-AI Environment: Training and Testing Animal-Like Artificial Cognition. *arXiv:1909.07483v2 [cs.LG]*

Boney, R., Sainio, J., Kaivola, M., Solin, A. and Kannala, J. 2020. RealAnt: An Open-Source Low-Cost Quadruped for Research in Real-World Reinforcement Learning. *arXiv:2011.03085 [cs.RO]*

Bonn, C. D. and Cantlon, J. F. 2012. The origins and structure of quantitative concepts, *Cognitive Neuropsychology*, 29(1-2):149-173.

Bonn, C. D. and Cantlon, J. F. 2017. Spontaneous, modality-general abstraction of a ratio scale. *Cognition*. 169:36–45.

Bostrom, N. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Botvinick, M., & Plaut, D. C. 2004. Doing without Schema Hierarchies: A Recurrent Connectionist Approach to Normal and Impaired Routine Sequential Action. *Psychological Review*. 111(2), 395–429.

Bringsjord, S. and Licato, J. 2012. Psychometric Artificial General Intelligence: The Piaget-MacGuyver Room. In: Wang P., Goertzel B. (eds) Theoretical Foundations of Artificial General Intelligence. Atlantis Thinking Machines, 4:25-48. Atlantis Press, Paris.

Burda, Y., Edwards, H., Pathak, D., Storkey, A., Darrell, T. and Efros, A. A. 2018. Large-scale study of curiosity-driven learning. *arXiv:1808.04355v1 [cs.LG]*.

Carpenter, G. A. and Grossberg, S. 2003. Adaptive resonance theory. In Arbib, M. A. (Ed.), *The Handbook of Brain Theory and Neural Networks – 2nd ed.* Cambridge, Massachusetts: MIT Press. 87-90.

Carruthers, P. 2018. The Causes and Contents of Inner Speech. In Langland-Hassan, P. and Vicente, A. (Eds.), *Inner Speech: New Voices*. Oxford University Press.

Casati, R. and Varzi, A. C. 1994. *Holes and Other Superficialities*. Cambridge, Mass. MIT Press.

Casati, R. and Varzi, A. C. 1999. *Parts and Places: The Structures of Spatial Representation*. Cambridge, Mass. MIT Press.

Catmur, C., Walsh, V. and Heyes, C. 2009. Associative sequence learning: the role of experience in the development of imitation and the mirror system. Philosophical Transactions of the Royal Society, B, Biological Science. 364(1528):2369-2380.

Cepelewicz, J. 2018. To Make Sense of the Present, Brains May Predict the Future. *Quanta Magazine.* July 10th, 2018.

Chandrasekaran, B. 2006. Multimodal Cognitive Architecture - Making Perception More Central to Intelligent Behavior*. Proceedings of the AAAI National Conference on Artificial Intelligence (AAAI'06),* 1508-1512.

Choia, D. and Langley, P. 2018. Evolution of the Icarus Cognitive Architecture. *Cognitive Systems Research*. (48)5:25-38.

Chollet, F. 2019. On the Measure of Intelligence. arXiv:1911.01547v2 [cs.AI].

Clark, A. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*. 36:181–253

Cohn, A. G. and Renz, J. 2007. Qualitative Spatial Representation and Reasoning. In Harmlen, V. E., Lifschitz, V. and Poster, B. (Eds.), *Handbook of knowledge representation*. Amsterdam: Elsevier. 13:551-596.

Conant, R. C. and Ashby, W. R. 1970. Every Good Regulator of a System Must Be a Model of That System. *International Journal of Systems Science*. 1(2):89-97

Cooper R., and Glasspool D. 2001. Learning Action Affordances and Action Schemas. In: French R.M., Sougné J.P. (eds) *Connectionist Models of Learning, Development and Evolution. Perspectives in Neural Computing*. Springer, London 133-142.

Cooper, R. P. 2010. Forward and inverse models in motor control and cognitive control. *Proceedings of the International Symposium on AI-Inspired Biology*, Jackie Chappell, Susannah Thorpe, Nick Hawes and Aaron Sloman (Eds.), at the AISB 2010 convention, De Montfort University, Leicester, UK

Cooper, R. P., and Shallice, T. 2000. Contention scheduling and the control of routine activities. *Cognitive neuropsychology*. 17(4):297-338.

Cooper, R. P., and Shallice, T. 2006. Hierarchical schemas and goals in the control of sequential behavior. *Psychological Review*. 113(4):887–916.

Cooper, R. P., Ruh, N., & Mareschal, D. 2014. The Goal Circuit Model: A Hierarchical Multi-Route Model of the Acquisition and Control of Routine Sequential Action in Humans. *Cognitive Science*, 38(2):244-274.

Cowan, N. 2000. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences.* 24:87–185.

Craik, K. 1943. *The nature of explanation*. Cambridge: Cambridge University Press. 5:50-61

D'Avila Garcez, A., Gori, M., Lamb, L. C., Serafini, L., Spranger, M. and Tran, S. N. 2019. Neural-Symbolic Computing: An Effective Methodology for Principled Integration of Machine Learning and Reasoning. *Journal of Applied Logics - IfCoLog Journal of Logics and their Applications* 6(4):611-631.

Damen, D., Doughty, H., Farinella, G. M., Fidler, S., Furnari, A., Kazakos, E., Moltisanti, D., Munro, J., Perrett, T., Price, W. and Wray, M. 2018. Scaling Egocentric Vision: The EPIC-KITCHENS Dataset. *arXiv:1804.02748v2 [cs.CV]*

De Houwer, J., Barnes-Holmes, D. and Moors, A. 2013. What is learning? On the nature and merits of a functional definition of learning. *Psychonomic Bulletin & Review*. 20:631–642.

de Wit, S., and Dickinson, A. 2009. Associative theories of goal-directed behaviour: A case for animal-human translational models. *Psychological Research.* 73(4).

Dean, J. and Cruse, H. 2003. Motor Pattern Generation. In Arbib, M. A. (Ed.), *The Handbook of Brain Theory and Neural Networks – 2nd ed*. Cambridge, Massachusetts: MIT Press. 696-701.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K. and Fei-Fei, L. 2009. ImageNet: A Large-Scale Hierarchical Image Database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009), 248-255.

Dezfouli, A. and Balleine, B. W. 2012. Habits, action sequences, and reinforcement learning. *European Journal of Neuroscience*. 35(7):1036–1051.

DiCarlo, J. J. and Cox, D. D. 2007. Untangling invariant object recognition. *Trends in Cognitive Sciences*. 11(8):333-341.

Dickinson, A. 1985. Actions and habits: The development of behavioral autonomy. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, 308(1135):67-78.

Dolan, R, J and Dayan, P. 2013. Goals and habits in the brain. *Neuron*. 80(2):312–325.

Domingos, P. 2015. *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. Basic Books, NY.

Dong, D. and Franklin, S. 2014. The Action Execution Process Implemented in Different Cognitive Architectures: A Review. *The Journal of Artificial General Intelligence.* 5(1):47-66.

Doumas, L. A. A. and Hummel, J. E. 2005. Approaches to Modeling Human Mental Representations: What Works, What Doesn't, and Why. In Holyoak, K. and Morrison, R. (Eds.), *The Cambridge handbook of thinking and reasoning*. Cambridge University Press. Ch. 4:73-91.

Doumas, L. A. A., & Martin, A. E. 2018. Learning structured representations from experience. In K. D. Federmeier (Ed.), *The psychology of learning and motivation*. 165–203. Elsevier Academic Press.

Doumas, L. A. A., Hummel, J. E., and Sandhofer, C. M. 2008. A theory of the discovery and predication of relational concepts. *Psychological Review*, 115(1), 1–43.

Duch, W., Oentaryo, R. J. and Pasquier, M. 2008. Cognitive Architectures: Where do we go from here? *Proceedings of the First AGI Conference (AGI'08).* 122-136.

Eberding, L. M., Thórisson, K. R. Sheikhlar, A. and Andrason, S. P. 2020. SAGE: Task-Environment Platform for Evaluating a Broad Range of AI Learners. *Artificial General Intelligence*, Springer International Publishing

Eckstein, M. K. and Collins, A. G. E. 2021. How the Mind Creates Structure: Hierarchical Learning of Action Sequences. *Proceedings of the Annual Meeting of the Cognitive Science Society*. 43.

Eckstein, M. P. 2011. Visual search: A retrospective. *Journal of Vision.* 11(5):14;1–36.

Elman, J. L. 1993. Learning and development in neural networks: the importance of starting small. *Cognition* 48:71-99.

Evans, J. St. B. T. 2003. In two minds: dual-process accounts of reasoning. Trends in Cognitive Sciences 7(10):454-459.

Evans, J. St. B. T. 2005. Deductive Reasoning. In Holyoak, K. and Morrison, R. (Eds.), *The Cambridge handbook of thinking and reasoning*. Cambridge University Press. Ch. 8:169-184.

Evans, J. St. B. T. and Stanovich, K. E. 2013. Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*. 8(3):223-241.

Everitt, T., Lea, G. and Hutter, M. 2018. AGI Safety Literature Review. arXiv:1805.01109v2 [cs.AI]

Ewert, J. 2003. Command Neurons and Command Systems. In Arbib, M. A. (Ed.), *The Handbook of Brain Theory and Neural Networks – 2nd ed*. Cambridge, Massachusetts: MIT Press. 233-238.

Faghihi, U. and Franklin, S. 2012. The LIDA model as a foundational architecture for AGI. In:Wang, P. and Goertzel, B. (Eds.), *Theoretical Foundations of Artificial General Intelligence.* Atlantis thinking machines, 4:105–123.

Fagioli, S., Hommel, B. and Schubotz, R. I. 2005. Intentional control of attention: action planning primes action-related stimulus dimensions. *Psychological Research* 71(1):22-29.

Feigenson, L., Dehaene, S. and Spelke, E. 2004. Core systems of number. *Trends in Cognitive Sciences*. 8(7):307-314.

Feix, T., Romero, J., Schmiedmayer, H., Dollar, A. M. and Kragic, D. 2016. The GRASP Taxonomy of Human Grasp Types. *IEEE Transactions on Human-Machine Systems*. 46(1):66-77.

Fidler, S. and Leonardis, A. 2007. Towards Scalable Representations of Object Categories: Learning a Hierarchy of Parts. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007)*.

Fidler, S., Boben, M. & Leonardis, A. 2009. Learning Hierarchical Compositional Representations of Object Structure. In Dickinson, S., Leonardis, A., Schiele, B. and Tarr, M. (Eds.), *Object Categorization: Computer and Human Vision Perspectives*. New York: Cambridge University Press.

Flavell, J.H. 1976. Metacognitive Aspects of Problem Solving. In: Resnick, L.B., Ed., *The Nature of Intelligence*, Lawrence Erlbaum Associates, Hillsdale, 231-235.

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.

Frégnac, Y. 1995. Hebbian synaptic plasticity: comparative and developmental aspects. In Arbib M. A. (Ed.), *The Handbook of Brain Theory and Neural Networks – 1st ed.* Cambridge, Massachusetts: MIT Press. 459-464.

Friston, K. J. 2009. The free-energy principle: a rough guide to the brain? *Trends in Cognitive Science* 13(7):293-301.

Friston, K. J. 2010. The free-energy principle: a unified brain theory? *Nature Reviews. Neuroscience* 11(2):127-138.

Fu, W.-T., and Anderson, J. R. 2006. From recurrent choice to skill learning: A reinforcement-learning model. *Journal of Experimental Psychology: General*. 135(2):184-206.

Fuster, J. M. 2003a. *Cortex and Mind, Unifying Cognition*. Oxford University Press, New York.

Fuster, J. M. 2003b. Prefrontal Cortex in Temporal Organization of Action. In Arbib, M. A. (Ed.), *The Handbook of Brain Theory and Neural Networks – 2nd ed.* Cambridge, Massachusetts: MIT Press. 905-910.

Gallistel, C. R. 1980. *The Organization of Action: A New Synthesis*. Hillsdale, N.J.: Erlbaum.

Gallistel, C. R. 1990. *The Organization of Learning*. Cambridge, Massachusetts: MIT Press.

Gallistel, C. R. 2001. Mental Representations, psychology of. *International Encyclopedia of the Social and Behavioral Sciences 9691-9695.*

Gallistel, C. R. 2011. Mental Magnitudes. In Dehaene S. and Brannon, E. (Eds.), *Space, Time and Number in the Brain: Searching for the Foundations of Mathematical Thought*. Ch. 1:3-12. Elsevier Academic Press.

Gallistel, C. R. and Gelman, R. 2000. Non-verbal numerical cognition: from reals to integers. *Trends in Cognitive Sciences*. 4(12):59-65.

Gallistel, C. R. and Gibbon, J. 2000. Time, Rate and Conditioning. *Psychological Review*. 107(2):289-344.

Gallistel, C.R. and Gelman, R. 2005. Mathematical Cognition. In Holyoak, K. and Morrison, R. (Eds.), *The Cambridge handbook of thinking and reasoning*. Cambridge University Press. Ch. 23:559-588.

Garnelo, M., Arulkumaran, K. and Shanahan, M. 2016. Towards Deep Symbolic Reinforcement Learning. *Computing Research Repository (CoRR).* arXiv:1609.05518v2 [cs.AI]

Gatsoulis, Y., Kerr, E., Condell, J. V., Siddique, N. H. and McGinnity, T. M. 2010. Novelty detection for cumulative learning. *Proceedings of Towards Autonomous Robotic Systems (TAROS'10),* 62-67.

Gawronski, B. and Creighton, L. A. 2013. Dual Process Theories. In Carlston, D. E. (Ed.) *The Oxford handbook of social cognition*. New York, NY: Oxford University Press. Ch. 14:282-312.

Gay, S. L., Mille, A., Georgeon, O. L. and Dutech, A. 2017. Autonomous construction and exploitation of a spatial memory by a self-motivated agent. *Cognitive Systems Research*. 41:1–35

Gentner, D. and Smith, L. A. 2013. Analogical learning and reasoning. In D. Reisberg (Ed.), *The Oxford handbook of Cognitive Psychology.* New York, NY: Oxford University Press. 668-681.

Gentner, D. R., Grudin, J. and Conway, E. 1980. *Finger Movements in Transcription Typing*. La Jolla, CA: University of California, San Diego, Center for Human Information Processing (Technical Report 8001).

Georgeff, M., Pell, B., Pollack, M., Tambe, M., Wooldridge, M. 1999. The Belief-Desire-Intention Model of Agency. In: Müller J.P., Rao A.S., and Singh M.P. (Eds.), *Proceedings of the 5th International Workshop on Intelligent Agents V: Agent Theories, Architectures, and Languages*. ATAL 1998. Lecture Notes in Computer Science, vol 1555. Springer, Berlin, Heidelberg. 1-10.

Georgeon, O. L. & Aha, D. W. 2013. The Radical Interactionism Conceptual Commitment. Journal of Artificial General Intelligence. 4(2);31-36.

Georgeon, O. L. and Marshall, J. B. 2012. The Small Loop Problem: A Challenge for artificial emergent cognition. In proceedings of *Biologically Inspired Cognitive Architectures*.:137-144.

Georgeon, O. L. and Marshall, J. B. 2013. Demonstrating sensemaking emergence in artificial agents: A method and an example. International Journal of Machine Consciousness, 2(5):131-144.

Georgeon, O. L. and Ritter, F. E. 2012. An intrinsically-motivated schema mechanism to model and simulate emergent cognition. *Cognitive Systems Research* 15-16:73-92.

Georgeon, O. L., Marshall, J. B. and Gay, S. 2012. Interactional Motivation in artificial systems: Between extrinsic and intrinsic motivation. In Proceedings of the 2nd international conference on development and learning and on epigenetic robotics (EPIROB2012), San Diego. :1–2. Development and Learning and Epigenetic Robotics (ICDL)

Gibson, J. J. 1979. *The Ecological Approach to Visual Perception*. Boston, MA: Houghton Mifflin.

Gobert, F. and Lane, P. C. R. 2017. Constructing a Standard Model: Lessons from CHREST. *Proceedings of the 2017 AAAI Fall Symposium Series*.

Goertzel, B. 2012. What counts as a conscious thinking machine? New Scientist. 2881.

Goertzel, B. 2014. Artificial General Intelligence - Concept, State of the Art, and Future Prospects. *Journal of Artificial General Intelligence.* 5(1):1-46.

Goldberg, D., Josyula, D. and Perlis, D. 2020. Quotation for Real-Time Metacognition. Proceedings of the Eighth Annual Conference on Advances in Cognitive Systems.

Goldstone, R. L. 1994. The role of similarity in categorization: providing a groundwork. *Cognition*. 52(2):125-57.

Goldstone, R. L. and Barsalou, L.W. 1998. Reuniting perception and conception. *Cognition*. 65:231–262.

Goldstone, R. L. and Son, J. Y. 2005. Similarity. In Holyoak, K. and Morrison, R. (Eds.), *The Cambridge handbook of thinking and reasoning*. Cambridge University Press. Ch. 2:13-36.

Goldstone, R. L. and Son, J. Y. 2012. Similarity. In K. J. Holyoak & R. G. Morrison (Eds.), *Oxford library of psychology. The Oxford handbook of thinking and reasoning.* New York, NY: Oxford University Press. Ch 10:155–176.

Goodale, M. A. and Milner, A. D. 1992. Separate visual pathways for perception and action. *Trends in Neuroscience*. 15(1):20–25.

Gopnik, A. 2020. Childhood as a solution to explore-exploit tensions. *Philosophical Transactions, Royal Society B, Biological Science*. 375(1803):20190502.

Gottlieb, J., Oudeyer, P-Y., Lopes, M. and Baranes, A. 2013. Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in Cognitive Sciences*. 17(11):585-593.

Granger, R. 2006. Engines of the brain: The computational instruction set of human cognition. AI Magazine 27:15-32

Granit, R. 1955. *Receptors and sensory perception*. Yale: University Press; London: Cumberlege.

Graziano, M. S. A., Taylor, C. S. R. and Moore, T. 2002. Complex Movements Evoked by Microstimulation of Precentral Cortex. *Neuron*. 34(5):841–851.

Greve, P.F. 2015. The role of prediction in mental processing: A process approach. *New Ideas in Psychology.* 39:45-52.

Grèzes, J., Armony, J. L., Rowe, J. and Passingham, R. E. 2003. *Activations related to “mirror” and “canonical” neurones in the human brain: an fMRI study.* NeuroImage. 18: 928–937

Grush, R. 2004. The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences.* 27:377–442.

Gupta, A., Vig, L. and Noelle, D. C. 2012. A neurocomputational approach to automaticity in motor skill learning. *Biologically Inspired Cognitive Architectures* 2:1–12.

Halford, G. S. and Wilson, W. H. 1980. A category theory approach to cognitive development. *Cognitive Psychology*. 12(3):356–411.

Halford, G. S., Baker, R., McCredden, J. E. and Bain, J. D. 2005. How Many Variables Can Humans Process? *American Psychological Society* 16(1).

Halford, G. S., Wilson, W. H. and Phillips S. 1998. Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences.* 21(6):803–831.

Halford, G. S., Wilson, W. H. and Phillips, S. 2010. Relational knowledge: The foundation of higher cognition. *Trends in Cognitive Sciences*, 14(11):497–505.

Hammer, P., Lofthouse, T. and Wang, P. 2016. The OpenNARS implementation of the Non-Axiomatic Reasoning System. *Proceedings of the Ninth AGI Conference (AGI’16).* 160-170.

Hampton, J. 1998. The Role of Similarity in How We Categorize The World. *International Workshop on Advances in Analogy Research: Integration of Theory and Data from the Cognitive, Computational, and Neural Sciences*. Edited by Keith Holyoak, Dedre Gentner, and Boidio Kokinov. NBU Series in Cognitive Science. New Bulgarian University, Sofia. 19-30

Hampton, J. A. 2001. The role of similarity in natural categorization. In U. Hahn & M. Ramscar (Eds.), *Similarity and categorization.* 13–28. Oxford University Press.

Hard, B. M., Tversky, B. & Lang, D. S. 2006. Making sense of abstract events: Building event schemas. *Memory & Cognition.* 34(6):1221-1235

Harnad, S. 1990. The Symbol Grounding Problem. *Physica D* 42:335-346.
Hawkins, J., Lewis, M., Klukas, M., Purdy S. and Ahmad, S. 2019. A Framework for Intelligence and Cortical Function Based on Grid Cells in the Neocortex. *Frontiers in Neural Circuits.* 12(121):1-14.

Hayes-Roth, B. 1995. An Architecture for Adaptive Intelligent Systems. *Artificial Intelligence* 72(1-2):329-365.

Hebb, D.O. 1949. *The Organization of Behavior: A Neuropsychological Theory.* New York, John Wiley & Sons.

Heider, F., & Simmel, M. 1944. An experimental study of apparent behavior. *American Journal of Psychology*, 57(2), 243-259.

Held, R. 1961. Exposure history as a factor in maintaining stability of perception and coordination. *Journal of Nervous and Mental Disease.* 132:26–32.

Helgason H.P., Nivel E. and Thórisson K.R. 2012. On Attention Mechanisms for AGI Architectures: A Design Proposal. In: Bach J., Goertzel B., Iklé M. (Eds) *Artificial General Intelligence. AGI 2012*. Lecture Notes in Computer Science, 7716:89-98. Springer, Berlin, Heidelberg.

Heller, J. 2014. On the Perception of Achromatic Colors: Wallach's Ratio Principle Revisited. *TPM* 21(4):371-388.

Hercus, R., Wong, K.-Y. and Ho, K.-F. 2013. Balancing of a simulated inverted pendulum using the NeuraBASE network model. *Artificial Neural Networks and Machine Learning – ICANN 2013, LNCS*, 8131:527-536.

Hernández-Orallo, J. 2016. Evaluation in artificial intelligence: From task-oriented to ability-oriented measurement. *Artificial Intelligence Review*. 48;397–447.

Hernández-Orallo, J. and Dowe, D. L. 2010. Measuring universal intelligence: Towards an anytime intelligence test. *Artificial Intelligence*. 174(18):1508-1539.

Hesslow, G. 2002. Conscious thought as simulation of behavior and perception. *Trends in Cognitive Sciences* 6(6):242-247.

Hesslow, G. 2012. Current status of the simulation theory of cognition. Brain Research, 1428:71-79.

Hexmoor, H. H. 1995. Representing and Learning Routine Activities. PhD dissertation, State University of New York at Buffalo.

Hommel, B. 2005. Perception in action: multiple roles of sensory information in action control. *Cognitive Processing* 6:3–14.

Hommel, B. 2007. Consciousness and Control, Not Identical Twins. *Journal of Consciousness Studies*, 14(1-2):155–176.

Hommel, B. 2009a. Action control according to TEC (theory of event coding). *Psychological Research.* 73:512–526.

Hommel, B. 2009b. Conscious and unconscious control of spatial action. In W.P. Banks (ed.), *Encyclopedia of consciousness*. 171-181. Oxford: Elsevier. And in *Reference Module in Neuroscience and Biobehavioral Psychology*, 2017, 1–8.

Hommel, B., Brown, S. B. R. E. and Nattkemper D. 2016. Human Action Control: From intentions to movements. Springer International Publishing AG, Switzerland.

Hommel, B., Müsseler, J., Aschersleben, G. and Prinz, W. 2001. The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences* 24:849-937.

Hong, H., Yamins, D. L. K., Majaj, N. J. and DiCarlo J. J. 2016. Explicit information for category-orthogonal object properties increases along the ventral stream. *Nature Neuroscience.* 19:613–622.

Hummel, J. E. and Biederman, I. 1992. Dynamic binding in a neural network for shape recognition. *Psychological Review*. 99(3):480-517.

Hummel, J. E. and Holyoak, K. J. 1997. Distributed Representations of Structure: A Theory of Analogical Access and Mapping. *Psychological Review*. 104(3):427–466.

Hummel, J. E. and Holyoak, K. J. 2003. A Symbolic-Connectionist Theory of Relational Inference and Generalization. *Psychological Review*. 110(2):220–264.

Hummel, J. E., Holyoak, K. J., Green, C., Doumas, L. A. A., Devnich, D., Kittur, K. and Kalar, D. J. 2004. A solution to the binding problem for compositional connectionism. In Simon D. Levy and Ross Gayler (Eds.), *Compositional Connectionism in Cognitive Science*. *Papers from the AAAI fall symposium*. 31-34. Menlo Park, CA: AAAI Press.

Isaac, A. M. C. 2013. Objective Similarity and Mental Representation. *Australasian Journal of Philosophy.* 91(4):683-704.

Izard, V. and Dehaene, S. 2008. Calibrating the mental number line. *Cognition*. 106(3):1221-1247.

Jaegle, A., Mehrpour, V. and Rust, N. 2019. Visual novelty, curiosity, and intrinsic reward in machine learning and the brain. *Computing Research Repository (CoRR).* arXiv:1901.02478 [q-bio.NC]

Jamone, L., Ugur, E., Cangelosi, A., Fadiga, L., Bernardino, A., Piater J. and Santos-Victor, J. 2018. Affordances in psychology, neuroscience and robotics: a survey. IEEE Transactions on Cognitive and Developmental Systems, 10(1):4-25.

Jeannerod, M. 1994. The Representing Brain: Neural Correlates of Motor Intention and Imagery. *Behavioral and Brain Sciences* 17(2):187-245.

Jeannerod, M. 1997. *The Cognitive Neuroscience of Action.* Wiley-Blackwell Publishers.

Jeannerod, M. 2006. *Motor Cognition: what actions tell the self*. Oxford University Press, Oxford, UK.

Johnson-Laird, P. N. 1998. Imagery, visualization, and thinking. In J. Hochberg (Ed.), *Handbook of perception and cognition (2nd ed.). Perception and Cognition at the Century's End.* 14:441-467. San Diego, CA: Academic Press.

Johnson-Laird, P. N. 2005. Mental Models and Thought. In Holyoak, K. and Morrison, R. (Eds.), *The Cambridge handbook of thinking and reasoning*. Cambridge University Press. Ch.9:185-208.

Johnston, B. 2010. The Toy Box Problem (and a Preliminary Solution). *Proceedings of the 3d Conference on Artificial General Intelligence*. 112-117

Kahneman, D. 2011. *Thinking, Fast and Slow*. Farrar, Straus and Giroux

Kieras, D. E. 2004. The EPIC architecture principles of operation.

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

Kilteni, K., Andersson, B. J., Houborg, C. and Ehrsson, H. H. 2018. Motor imagery involves predicting the sensory consequences of the imagined movement. *Nature Communications*. 9:1617

Kingo, O. 2008. The Concept of Concepts: Perceptual and Conceptual Categorization in Infancy under Scrutiny. *Journal of Anthropological Psychology* 19.

Klahr, D., Langley, P. and Neches, R. Eds. 1987. *Production system models of learning and development*. Cambridge, MA. MIT Press.

Knauff, M. 2013. *Space to reason: a spatial theory of human thought*. MIT Press, Cambridge, Massachusetts.

Kokinov, B. and French, R. M. 2003. Computational Models of Analogy-making. In Nadel, L. (Ed.) *Encyclopedia of Cognitive Science.* 1:113–118. London: Nature Publishing Group.

Kolers, P. A. 1964. The Illusion of Movement. *Scientific American* 211(4):98-106. Reprinted in Held, R. and Richards, W. (Eds.) *Perception: Mechanisms and Models: Readings from "Scientific American".* 1972, 316-323.

Kommrusch, S. 2020. Self-Supervised Learning for Multi-Goal Grid World: Comparing Leela and Deep Q Network. *Proceedings of Machine Learning Research*. 131:72-88.

Kosoy, E., Collins, J., Chan, D. M., Huang, S., Pathak, D., Agrawal, P., Canny, J., Gopnik, A. and Hamrick, J. B. 2020. Exploring exploration: Comparing children with RL agents in unified exploration environments. Published as a workshop paper at "Bridging AI and Cognitive Science" (ICLR 2020). *arXiv:2005.02880v2 [cs.AI].*

Kosslyn, S. M. 1994. *Image and Brain: The Resolution of the Imagery Debate*. Cambridge, MA: MIT Press.

Kosslyn, S. M. 2005. Mental images and the brain. *Cognitive Neuropsychology*. 22(3): 333–347.

Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Chabris, C. F., Rauch, S. L.,and Anderson, A. K. 1994. Identifying objects seen from different viewpoints A PET investigation. *Brain*. 117:1055-1071.

Kotseruba, I. and Tsotsos J. K. 2018. 40 years of cognitive architectures: core cognitive abilities and practical applications in *Artificial Intelligence Review.* 53, 17–94.

Kralik, J. D., Lee, J. H., Rosenbloom, P. S., Jackson, P. C., Epstein, S. L., Romero, O. J., Sanz, R., Larue, O., Schmidtke, H.R., Lee, S. W. and McGreggor, K. 2018. Metacognition for a Common Model of Cognition. *Procedia Computer Science* 145:730–739.

Krawczyk, D. C., Holyoak, K. J. and Hummel, J. E. 2004. Structural constraints and object similarity in analogical mapping and inference. *Thinking & Reasoning*. 10(1):85—104.

Kubovy, M. and van den Berg, M. 2008. The whole is equal to the sum of its parts: A probabilistic model of grouping by proximity and similarity in regular patterns. *Psychological Review*. 115(1):131–154.

Kubovy, M., Holcombe, A. O. and Wagemans, J. 1998. On the lawfulness of grouping by proximity. *Cognitive Psychology*, 35(1):71–98.

Kugele, S. and Franklin, S. 2020. General Intelligence Requires Autonomous, Cognitive, Intentional Agents. Proceedings of the Eighth Annual Conference on Advances in Cognitive Systems.

Kunde, W., Elsner, K. and Kiesel, A. 2007. No anticipation-no action: The role of anticipation in action and perception. *Cognitive Processing*. 8(2):71-78.

Laird, J. E. 2012. The soar cognitive architecture. *Artificial Intelligence and Simulation of Behaviour Quarterly*. 171(134):224–235.

Laird, J. E., Lebiere, C. and 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(4):13–26.

Lake, B. L., Salakhutdinov, R. and Tenenbaum, J. B. 2019. The Omniglot challenge: a 3-year progress report. *Current Opinion in Behavioral Sciences*. 29:97-104.

Land, E. H. 1959. Experiments in color vision. *Scientific American* 200(5):84-94. Reprinted in Held, R. and Richards, W. (Eds.) *Perception: Mechanisms and Models: Readings from "Scientific American".* 1972, 286:298.

Land, W. E., Volchenkov, D., Bläsing, B. E. and Schack, T. 2013. From action representation to action execution: exploring the links between cognitive and biomechanical levels of motor control. *Frontiers in Computational Neuroscience Hypothesis and Theory Article.* 7(127):1-14.

Langley, P. 2017. Progress and Challenges in Research on Cognitive Architectures. *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17).* 4870-4876

Langley, P., Laird, J. E. and Rogers, S. 2009. Cognitive Architectures: Research Issues and Challenges. *Cognitive Systems Research* 10(2):141-160.

Lathrop S.D. 2008. Extending Cognitive Architectures with Spatial and Visual Imagery Mechanisms. PhD dissertation, University of Michigan

Leech, R., Mareschal, D. and Cooper, R. P. 2008. Analogy as relational priming: A developmental and computational perspective on the origins of a complex cognitive skill. *Behavioral and Brain Sciences*. 31(4):357–414.

Legg, S. 2008. *Machine Super Intelligence (PhD)*. Department of Informatics, University of Lugano. Retrieved September 19, 2014.

Legg, S. and Hutter, M. 2007. Universal Intelligence: A Definition of Machine Intelligence. arXiv:0712.3329v1 [cs.AI] and *Minds and machines*, 17(4):391–444.

Leonardis, A. and Fidler, S. 2007. Learning hierarchical representations of object categories for robot vision. *International Symposium on Robotics Research.* 66:99-110.

Leslie, A. M., Gallistel, C. R. and Gelman, R. 2019. Where integers come from. To appear in P. Carruthers (Ed.), *The Innate Mind: Foundations and Future*. Oxford: Oxford University Press.

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., and Dollár, P. 2014. Microsoft COCO: Common Objects in Context. In Proceedings of the European Conference on Computer Vision (ECCV 2014), 740-755.

Lorenz, D., Bereska, l., Milbich, T. and Ommer, B. 2019. Unsupervised Part-Based Disentangling of Object Shape and Appearance. *arXiv:1903.06946v3 [cs.CV]*

Lucas, C. G., Bridgers, S., Griffiths, T. L. and Gopnik, A. 2014. When children are better (or at least more open-minded) learners than adults: developmental differences in learning the forms of causal relationships. *Cognition*. 131(2):284-99.

Luong, M.-T. and Manning, C. D. 2016. Achieving Open Vocabulary Neural Machine Translation with Hybrid Word-Character Models. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 1054–1063.

Madl, T., Baars, B. J., and Franklin, S. 2011. The Timing of the Cognitive Cycle. *PLoS ONE* 6(4): e14803

Mahon, B. Z. and Caramazza, A. 2008. A Critical Look at the Embodied Cognition Hypothesis and a New Proposal for Grounding Conceptual Content. *Journal of Physiology-Paris* 102(1–3):59-70.

Mandler, J. M. 2000. Perceptual and Conceptual Processes in Infancy. *Journal of Cognition and Development* 1:3–36.

Mandler, J. M. 2004. *The Foundations of Mind*. Oxford University Press, New York

Marcus, G. 2018a. Deep Learning: A Critical Appraisal. *Computing Research Repository (CoRR).* arXiv:1801.00631v1 [cs.AI]

Marcus, G. 2018b. Innateness, AlphaZero, and Artificial Intelligence. arXiv:1801.05667v1 [cs.AI]

Martensen, B. N. 2013. Perceptra: A New Approach to Pattern Classification Using a Growing Network of Binary Neurons (Binons). In R. West & T. Stewart (Eds.), *Proceedings of the 12th International Conference on Cognitive Modeling (ICCM 2013)*.

Martensen, B. N. 2020. The Perception–Action Hierarchy and its Implementation Using Binons (Binary Neurons). Postproceedings of the 10th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2019 (Tenth Annual Meeting of the BICA Society), held August 15-19, 2019 in Seattle, Washington, USA. *Procedia Computer Science*. 169:489-500.

McCarthy, J. 1987. Generality in Artificial Intelligence. *Communications of the ACM* December 30(21):1030-1035.

McCulloch, W. S. and Pitts, W. 1943. A Logical Calculus of the Ideas Immanent in Nervous Activity. *The Bulletin of Mathematical Biophysics.* 5(4):115-133.

Medin, D. L. and Rips, L. J. 2005. Concepts and Categories: Memory, Meaning, and Metaphysics. In Holyoak, K. and Morrison, R. (Eds.), *The Cambridge handbook of thinking and reasoning*. Cambridge University Press. Ch. 3:37-72.

Meng, R., Matthews, P. G. and Toomarian, E. Y. 2019. The Relational SNARC: Spatial Representation of Nonsymbolic Ratios. *Cognitive Science* 43(8):1-17.

Miall, R.C. and Wolpert, D.M. 1996. Forward Models for Physiological Motor Control. *Neural Networks* 9(8):1265-1279.

Mirolli, M. 2012. Representations in dynamical embodied agents: re-analyzing a minimally cognitive model agent. *Cognitive Science*. 36(5):870-95.

Mitchell, M. 2020. On Crashing the Barrier of Meaning in AI. *AI Magazine.* 41(2):86–92.

Mountcastle, V. 1978. An organizing principle for cerebral function: the unit model and the distributed system. In Edelman, G. and Mountcastle, V. (Eds.), *The Mindful Brain: Cortical Organization and the Group-Selective Theory of Higher Brain Function*, (Cambridge, MA: MIT Press), 7–50.

Mountcastle, V. B., Motter, B. C., Steinmetz, M. A. and Duffy, C. J. 1984. Looking and seeing: the visual functions of the parietal lobe. In Edelman, G. M., Gall, W. E. and Cowan, W. M., eds. *Dynamic Aspects of Neocortical Function.* Wiley. 159-193.

Murray, J. D., Bernacchia, A., Freedman, D. J., Romo, R., Wallis, J. D., Cai, X., Padoa-Schioppa, C., Pasternak, T., Seo, H., Lee, D. and Wang, X-J. 2014. A hierarchy of intrinsic timescales across primate cortex. *Nature Neuroscience*. 17(12):1661–1663.

Nagashima, K., Morita, J. and Takeuchi, Y. 2020. Modeling intrinsic motivation in ACT-R: Focusing on the relation between pattern matching and intellectual curiosity. *18th International Conference on Cognitive Modeling*.

Nazarieh, M. 2016. A Brief History of Metacognition and Principles of Metacognitive Instruction in Learning. *BEST: Journal of Humanities, Arts, Medicine and Sciences (BEST: JHAMS)*. 2(2):61-64.

Neumann, O. 1984. Automatic processing: A review of recent findings and a plea for an old theory, In W. Prinz & A. F. Sanders, eds., *Cognition and Motor Processes*, Springer-Verlag, Berlin , pp. 255-293 .

Newell, A. 1973. Production systems: Models of control structures. In W. G. Chase (Ed.), *Visual Information Processing*. New York, Academic Press. 463-526.

Newell, A. and Simon, H. A. 1976. Computer Science as Empirical Inquiry: Symbols and Search. *Communications of the Association for Computing Machinery*. 19(3):113–126.

Newman, L. I. and Polk, T. A. 2007. The Emergence of Semantic Topography in a Neurally-Inspired Computational Model. *8th International Conference on Cognitive Modeling*. Edited by Lewis, R. L., Polk, T. A. and Laird, J. E.

Nguyen, T. N. and Gonzalez, C. 2020. Effects of Decision Complexity in Goal-seeking Gridworlds: A Comparison of Instance-Based Learning and Reinforcement Learning Agents. *18th International Conference on Cognitive Modeling*.

Nieder, A. 2011. The neural code for number. In Dehaene, S. and Brannon, E. (Eds.), *Space, Time and Number in the Brain: Searching for the Foundations of Mathematical Thought.* Ch.8:103–118. Elsevier Academic Press.

Nilsson, N. J. 2005. Human-Level Artificial Intelligence? Be Serious! *AI Magazine*. *26*(4):68-75

Nivel, E. and Thórisson, K. R. 2009. Self-Programming: Operationalizing Autonomy In B. Goertzel, P. Hitzler & M. Hutter (eds.), *Proceedings of the Second Conference on Artificial General Intelligence*. Atlantis Press (AGI-09) 150-155.

Norman, D. A. 1981. Categorization of Action Slips. *Psychological Review*. 88(1).

Norman, D. and Shallice, T. 1980. Attention to action: Willed and automatic control of behavior. *Center for Human Information Processing.* University of California, San Diego, La Jolla, California.

Norman, D. and Shallice, T. 1986. Attention to action: Willed and automatic control of behavior, in *Consciousness and Self-Regulation: Advances in Research and Theory*.

(Davidson, R. J., Schwartz, G. E. and Shapiro, D. (Eds.), New York: Plenum Press. 4:1–18.

Notterman, J. M., Filion, R. D. L. and Mandriota, F. J. 1971. Perception of Changes in Certain Exteroceptive Stimuli. *Science*. 173(4003):1206-1211.

O'Reilly, R. C., Russin, J. L., Zolfaghar, M. and Rohrlich, J. 2020. Deep Predictive Learning in Neocortex and Pulvinar. *arXiv:2006.14800 [q-bio.NC]*

Olds, J., & Milner, P. 1954. Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *Journal of Comparative and Physiological Psychology, 47*(6):419–427.

Oudeyer, P. 2018. Computational Theories of Curiosity-Driven Learning. *Computing Research Repository (CoRR).* arXiv:1802.10546v2 [cs.AI].

Oudeyer, P. and Kaplan, F. 2007. What is intrinsic motivation? A typology of computational approaches. *Frontiers in Neurorobotics*. 1(6):1-14.

Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., and Wermter, S. 2019. Continual Lifelong Learning with Neural Networks: A Review *arXiv:1802.07569v4 [cs.LG].* - Now in: *Neural Networks* 113:54-71

Pecher, D., Boot, I., & Van Dantzig, S. 2011. Abstract concepts: Sensory-motor grounding, metaphors, and beyond. In B. H. Ross (Ed.), *The psychology of learning and motivation: Vol. 54. The psychology of learning and motivation: Advances in research and theory*. 217–248. Elsevier Academic Press.

Peebles, D. and Cheng P, C.-H. 2017. Multiple Representations in Cognitive Architectures. *AAAI Fall Symposium.* "A Standard Model of the Mind".

Peebles, D. and Cooper, R. P. 2015. Thirty Years After Marr's Vision: Levels of Analysis in Cognitive Science. *Topics in Cognitive Science*. 7:187–190.

Perrett, D. I., Mistlin, A. J., Harries, M. H. and Chitty, A. J. 1990. Understanding the Visual Appearance and Consequence of Hand Actions. In Goodale, M. A., ed. *Vision and Action: The Control of Grasping*. Ablex. 163-180.

Pesek, M. 2018. Compositional hierarchical model for music information retrieval. PhD dissertation, University of Ljubljana

Pfeffer, A. and Lynn, S. K. 2018. Scruff: A Deep Probabilistic Cognitive Architecture for Predictive Processing. *Biologically Inspired Cognitive Architectures*. 245-*259*. Advances in Intelligent Systems and Computing, Vol 848. Springer, Cham.

Phillips, S. and Wilson, W. H. 2010. Categorial Compositionality: A Category Theory Explanation for the Systematicity of Human Cognition. *PLoS Computational Biology.* 6(7):e1000858

Phillips, S. and Wilson, W. H. 2011. Categorial Compositionality II: Universal Constructions and a General Theory of (Quasi-)Systematicity in Human Cognition. *PLoS Computational Biology*. 7(8):e1002102.

Piazza, M., Izard, V., Pinel, P., Le Bihan, D. and Dehaene, S. 2004. Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*. 44(3):547-555.

Pierella, C., Casadio, M., Mussa-Ivaldi, F. A. and Solla, S. A. 2019. The dynamics of motor learning through the formation of internal models. PLoS Computational Biology. 15(12):e1007118.

Portegys, T. 2010. A maze learning comparison of Elman, long short-term memory, and Mona neural networks. *Neural networks*. 23(2);306-313.

Portugal, R. D. and Svaiter, B. F. 2011. Weber-Fechner law and the optimality of the logarithmic scale. *Minds and Machines*. 12(1):73-81.

Powers, W. T. 1973. *Behavior: The control of perception*. Oxford, England: Aldine.
Prinz, W. 1990. A common coding approach to perception and action. In Neumann, O. and Prinz, W. (Eds.), *Relationships between perception and action* 167–201. Berlin: Springer.

Priss, U. 2019. Conceptual Alignment with Formal Concept Analysis. In: Cristea et al. (Eds.), *Supplementary Proceedings of ICFCA 2019 Conference and Workshops*. CEUR – 2738 Frankfurt, Germany, 14-27

Qin, H., Gong, R., Liu, X., Bai, X., Song, J. and Sebe, N. 2020. Binary Neural Networks: A Survey. *Pattern Recognition* (2020) 107281 *arXiv:2004.03333 [cs.NE]*

Quinn, P. C. and Eimas, P. D. 2000. The emergence of category representations during infancy: Are separate and conceptual processes required? *Journal of Cognition and Development* 1(1):55- 61.

Raaijmakers, J. G., and Shiffrin, R. M. 1981. Search of associative memory. *Psychological Review*, 88(2), 93–134.

Rachkovskij, D. A., Kussul, E. M. and Baidyk, T. N. 2013. Building a world model with structure-sensitive sparse binary distributed representations. *Biologically Inspired Cognitive Architectures*. 3:64-86.

Rajpurkar, P., Zhang, J., Lopyrev, K. and Liang, P. 2016. Squad: 100,000+ Questions for Machine Comprehension of Text. arXiv preprint arXiv:1606.05250.

Ratterman, M. J. and Gentner, D. 1998. More evidence for a relational shift in the development of analogy: Children's performance on a causal-mapping task. *Cognitive Development*. 13(4):453-478.

Reitter, D., and Lebiere, C. 2010. A cognitive model of spatial path-planning. *Computational and Mathematical Organization Theory. 16*(3);220–245.

Reynolds, J. R., Zacks, J. M. and Braver, T. S. 2007. A Computational Model of Event Segmentation from Perceptual Prediction. *Cognitive Science* 31:613–643.

Richter, M., Lins, J. and Schöner, G. 2017. A Neural Dynamic Model Generates Descriptions of Object-Oriented Actions. *Topics in Cognitive Science,* 9:35–47.

Riehle, A. 2005. Preparation for Action: one of the Key Functions of Motor Cortex. *Motor Cortex in Voluntary Movements: A Distributed System for Distributed Functions* Ch8 CRC Press

Ring, M. 1992. Two Methods for Hierarchy Learning in Reinforcement Environments. *From Animals to Animats 2: Proceedings of the Second International Conference on Simulation of Adaptive Behavior (SAB'92)* 148-155.

Ring, M. B. 1994. Continual Learning in Reinforcement Environments. PhD dissertation, University of Texas at Austin.

Ring, M. B. 1997. CHILD: A First Step Towards Continual Learning. *Machine Learning*. 28:77–104.

Rips, L. J., Smith, E. E., & Medin, D. L. 2012. Concepts and categories: Memory, meaning, and metaphysics. In K. Holyoak & R. Morrison (Eds.), *The Oxford Handbook of Thinking and Reasoning*. 11:177-209.

Rolf, M. and Asada, M. 2015. What are goals? And if so, how many? *Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*. 332-339.

Rosenbaum, D. A., Kenny, S. B. and Derr, M. A. 1983. Hierarchical Control of Rapid Movement Sequences. *Journal of Experimental Psychology: Human Perception & Performance*. 9(1):86-102.

Rosenbloom, P. S. and Demski, A. 2016. The Sigma Cognitive Architecture and System: Towards Functionally Elegant Grand Unification. *Journal of Artificial General Intelligence.* 7(1):1-103.

Rumelhart, D. E., and Norman, D. A. 1982. Simulating a skilled typist: A study of skilled cognitive-motor performance. *Cognitive Science*, 6:1-36.

Russell, S. 2019. *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking

Rutkowska, J. C. 1994. Scaling Up Sensorimotor Systems: Constraints from Human Infancy. *Adaptive Behavior* 2(4):349-373.

Samsonovich, A. V. 2010. Toward a Unified Catalog of Implemented Cognitive Architectures. *Proceedings of the 2010 conference on Biologically Inspired Cognitive Architectures: Proceedings of the First Annual Meeting of the BICA Society*. 195–244.

Sandamirskaya, Y. and Burtsev, M. 2015. NARLE: Neurocognitive architecture for the autonomous task recognition, learning, and execution. *Biologically Inspired Cognitive Architectures.* 13:91-104

Sap, M., Le Bras, R., Allaway, E., Bhagavatula, C., Lourie, N., Rashkin, H., Roof, B., Smith, N. A. and Choi, Y. 2018. ATOMIC: An Atlas of Machine Commonsense for If-Then Reasoning. *Computing Research Repository (CoRR).* arXiv:1811.00146v3 [cs.CL].

Savinov, N., Raichuk, A., Marinier, R., Vincent, D., Pollefeys, M., Lillicrap, T. and Gelly, S. 2018. Episodic Curiosity Through Reachability. *arXiv:1810.02274v1 [cs.LG]*.

Schack, T. 2004. The Cognitive Architecture of Complex Movement. *International Journal of Sport and Exercise Psychology.* 2:403-438.

Schmidt, R. A. 1975. A schema theory of discrete motor skill learning. *Psychological Review.* 82(4):225-260.

Schmidt, R. A. 2003. Motor Schema Theory after 27 Years: Reflections and Implications for a New Theory. *Research quarterly for exercise and sport.* 74(4):366-75

Schneider, S. 2009. The language of thought. In Symons, J. and Calvo, P. (Eds.), *Routledge Companion to Philosophy of Psychology.* Routledge.

Schneider, S. 2011. *The language of thought: A New Philosophical Direction*. The MIT Press, Cambridge, Massachusetts.

Seegelke, C. and Schack, T. 2016. Cognitive Representation of Human Action: Theory, Applications, and Perspectives. *Frontiers in Public Health* 4(24):1-6.

Shallice, T. 2006. Contrasting domains in the control of action: The routine and the non-routine. In Y. Munakata & M. H. Johnson (Eds.), *Processes of change in brain and cognitive development. Attention and performance.* Oxford, UK: University Press. 21:3–29.

Shastri, L. 2003. Structured Connectionist Models. In Arbib, M. A. (Ed.), *The Handbook of Brain Theory and Neural Networks – 2nd ed*. Cambridge, Massachusetts: MIT Press. 1116-1119.

Silver, D., Yang, Q. and Li, L. 2013. Lifelong machine learning systems: Beyond learning algorithms. In *Proceedings of the AAAI Spring Symposium on Lifelong Machine Learning*. Stanford, CA.49–55.

Singh, S., Barto, A. G., Chentanez, N. 2005. Intrinsically motivated reinforcement learning. In: Saul, L.K., Weiss, Y., Bottou, L. (Eds.) *Advances in Neural Information Processing Systems (NIPS 2004)* 17:1281–1288.

Singh, S., Lewis, R., Barto, A. and Sorg, J. 2010. Intrinsically motivated reinforcement learning: An evolutionary perspective. I*EEE Transactions on Autonomous Mental Development*. 2(2);70–82.

Soares, N., Fallenstein, B., Yudkowsky,E. and Armstrong, S. 2015. *Corrigibility.* In AAAI Workshops: Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, AAAI Publications.

Standvoss, K., Quax, S. C. and van Gerven, M. A. J. 2020. Visual Attention Through Uncertainty Minimization in Recurrent Generative Models. bioRxiv.

Stevens, J. C. and Marks, L. E. 1965. Cross-modality matching of brightness and loudness. *Proceedings of the National Academy of Sciences*. 54 (2):407-411.

Stevens, S. S. 1966. Matching functions between loudness and ten other continua. *Perception & Psychophysics*. 1(1):5–8.

Stewart, T. C. and West, R. L. 2006. Deconstructing ACT-R. In *Proceedings of the Seventh International Conference on Cognitive Modeling.* 298-303.

Stracuzzi, D. J., Li, N., Cleveland, G. and Langley, P. 2009. Representing and Reasoning over Time in a Unified Cognitive Architecture. *Proceedings of the Thirty-First Annual Meeting of the Cognitive Science Society*. Amsterdam.

Sun, R. 2004. Desiderata for cognitive architectures. *Philosophical Psychology* 17(3):341-373.

Sun, R. 2006. The CLARION cognitive architecture: Extending cognitive modeling to social simulation In: Ron Sun (ed.), Cognition and Multi-Agent Interaction. Cambridge University Press, New York.

Sun, R., Zhang, X. and Mathews, R. 2006. Modeling meta-cognition in a cognitive architecture. *Cognitive Systems Research.* 7(4):327-338.

Sutherland, S. L., Cimpian, A., Leslie, S. and Gelman, S. A. 2015. Memory Errors Reveal a Bias to Spontaneously Generalize to Categories. *Cognitive Science* 39:1021–1046.

Taylor, P., Hobbs, J. N., Burroni, J. and Siegelmann, H. T. 2015. The global landscape of cognition - hierarchical aggregation as an organizational principle of human cortical networks and functions. *Scientific Reports* 5(18112):1-18.

Technical report. ADA437286, Michigan State University of East Lansing, Dept. of Computer Science.

Tenorth, M., Bandouch, J. and Beetz, M. 2009. The TUM Kitchen Data Set of Everyday Manipulation Activities for Motion Tracking and Action Recognition. *IEEE 12th International Conference on Computer Vision Workshops*, ICCV Workshops. 1089-1096.

Thagard, P. 2012. Cognitive architectures. In Frankish K. and Ramsay W. (Eds.), *The Cambridge handbook of cognitive science* 3:50-70. Cambridge: Cambridge University Press.

Thórisson, K. R. and Helgasson, H. P. 2012. Cognitive Architectures and Autonomy: A Comparative Review. *Journal of Artificial General Intelligence* 3(2):1-63.

Thórisson, K. R. and Nivel, E. 2009. Achieving Artificial General Intelligence Through Peewee Granularity. In *Proceedings of the 2nd Conference on Artificial General Intelligence*. 220-221.

Thórisson, K. R. and Talbot, A., 2018. Cumulative Learning with Causal-Relational Models. *Artificial General Intelligence 2018.* 227-238.

Tian, X, and Poeppel, D. 2012. Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation. *Frontiers in Human Neuroscience*. 6(314)

Tian, X. and Poeppel, D. 2010. Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in Psychology*. 1(166):1-23.

Tonneau, F. 2011. Metaphor and Truth - A Review of Representation Reconsidered by W.M. Ramsey. *Behavior and Philosophy*, 39/40:331-343.

Torrey, L. and Shavlik, J. 2009. Transfer Learning. *The Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*. IGI Global. 1:242-264.

Treisman, A. 2006. How the deployment of attention determines what we see. *Visual Cognition*. 14(4-8):411–443.

Turing, A. M. 1950. Computing machinery and Intelligence. *Mind – A quarterly review of Psychology and Philosophy*. 59(236):433-460.

Ugur E. and Piater, J. 2017. *Emergent Structuring of Interdependent Affordance Learning Tasks Using Intrinsic Motivation and Empirical Feature Selection*. IEEE Transactions on Cognitive and Developmental Systems. 9(4).

van den Oord, A., Li, Y. and Vinyals O. 2018. Representation Learning with Contrastive Predictive Coding. *Computing Research Repository (CoRR).* arXiv:1807.03748v2 [cs.LG].

Veksler, V. D., Gluck, K. A., Myers, C. W., Harris, J. and Mielke, T. 2104. Alleviating the curse of dimensionality – A psychologically-inspired approach. *Biologically Inspired Cognitive Architectures*. 10:51-60

Vernon, D., von Hofsten, C. and Fadiga, L. 2016. Desiderata for developmental cognitive architectures. *Biologically Inspired Cognitive Architectures* 18:116-127.

von Holst, E. 1954. Relations between the central nervous system and the peripheral organs. *British Journal of Animal Behavior.* 2:89–94.

Wallach, H. 1963. The perception of neutral colors. *Scientific American*, 208(1):107–116. Reprinted in Held, R. and Richards, W. (Eds.) *Perception: Mechanisms and Models: Readings from "Scientific American".* 1972, 278-285.

Waszak, F., Cardoso-Leite, P. and Hughes, G. 2012. Action effect anticipation: Neurophysiological basis and functional consequences. *Neuroscience and Biobehavioral Reviews* 36:943–959.

Weng, J. and Zhang, Y. 2005. Developmental Robots – A New Paradigm

West, R. L. and Conway-Smith, B. 2019. Put Feeling into Cognitive Models: A Computational Theory of Feeling. In Editor Stewart, T. C. *Proceedings of the 17th International Conference of Cognitive Modelling (ICCM 2019).* 295-300

Wilson, A. D. and Golonka, S. 2013. Embodied Cognition is Not What you Think it is. *Frontiers in Psychology*. 4(58).

Wortham, R. H. 2018. *Using Other Minds: Transparency as a Fundamental Design Consideration for Artificial Intelligent Systems*. PhD Dissertation. University of Bath.

Wozniak, S. G. 2007. Three minutes with Steve Wozniak. *ABC News*.

Yaworsky, P. 2018. A Model for General Intelligence. arXiv:1811.02546 [cs.AI]

Yu, Q., Tang, H., Tan, K. C. and Li, H. 2013. Rapid Feedforward Computation by Temporal Encoding and Learning with Spiking Neurons. *IEEE Transactions on Neural Networks and Learning Systems.* 24(10):1539-1552.