# Introduction to Cognitive Robotics

## Module 7: Cognitive Architectures

## Lecture 1: Role & requirements; desirable characteristics; core cognitive abilities

David Vernon
Carnegie Mellon University Africa

www.vernon.eu

# What are the characteristics of a cognitive agent?

The chief characteristic of a cognitive agent is the ability to act effectively in a world that is uncertain, under-specified, dynamic, possibly cooperating with other cognitive agents

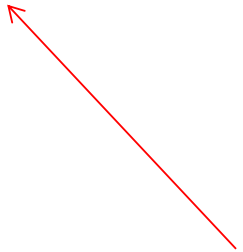
To achieve goals adaptively and robustly in these circumstances requires a complex system that can

- Construct models of the way the world works,
- Use them to guide actions prospectively, and
- Update them dynamically as the system continually learns through its interactions

A cognitive architecture is the way we specify what is required to achieve this.

# What is a cognitive architecture?

A cognitive architecture is a software framework that integrates all the elements required for a system to exhibit the characteristic attributes of a cognitive agent



The design of a cognitive architecture requires the specification of the formalisms for all the processes and knowledge representations used by that framework

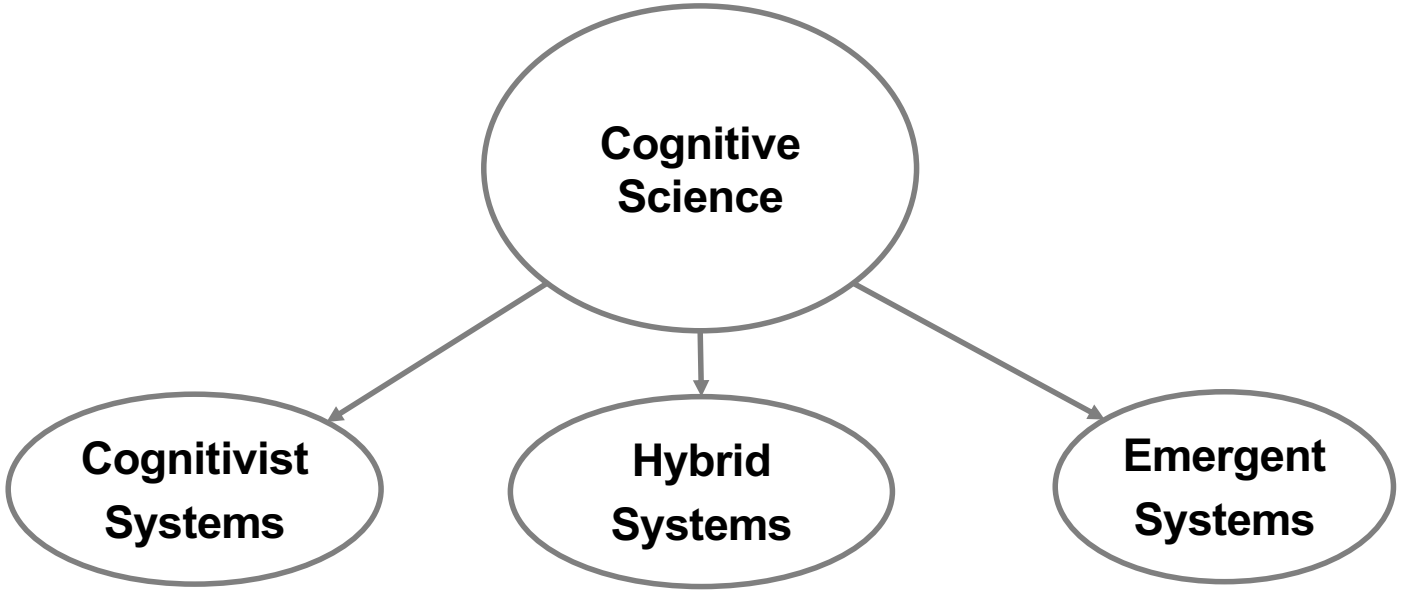
# How does a cognitive architecture work?

A cognitive architecture integrates the core cognitive abilities so that these abilities can be dynamically coordinated

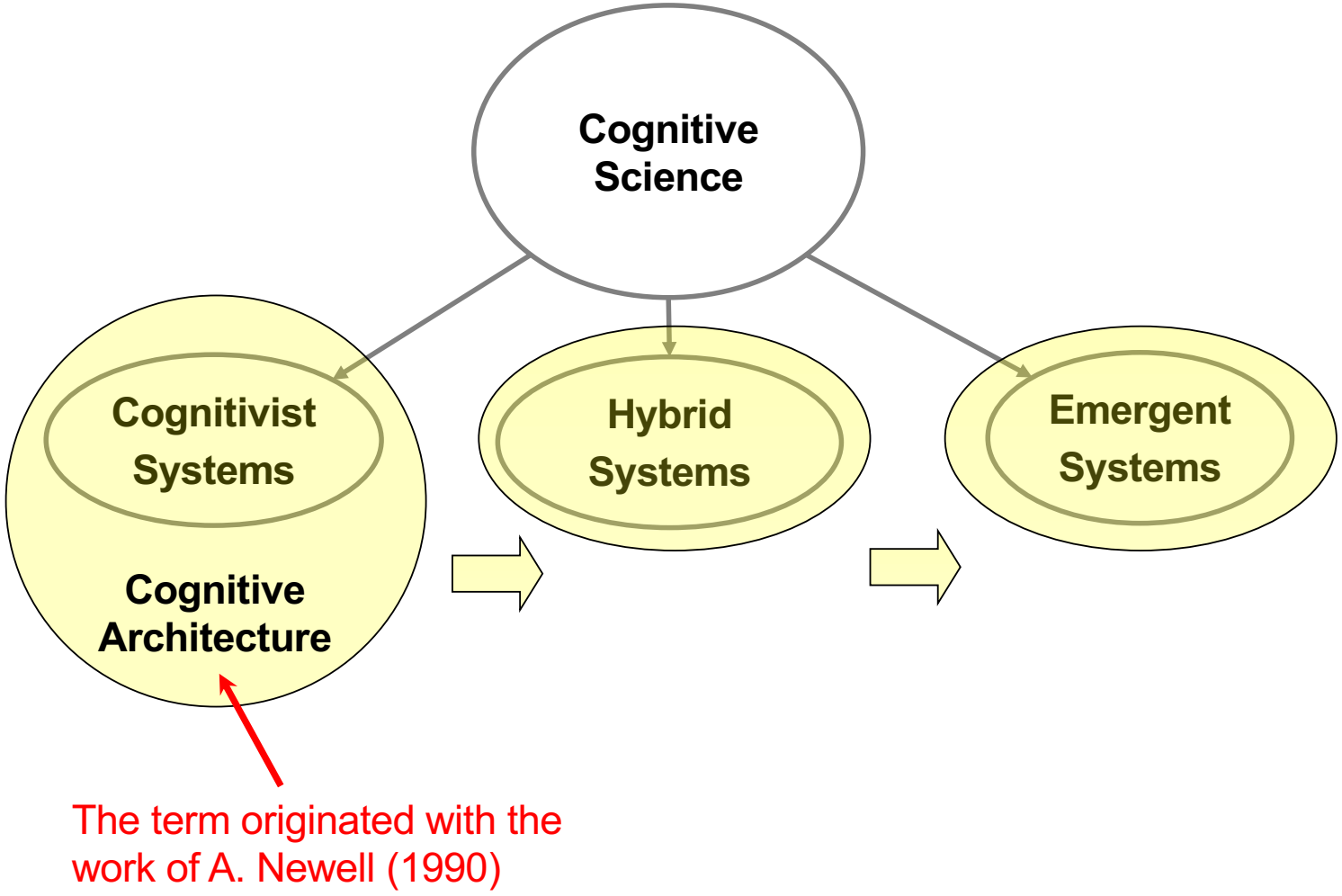
Perception
Attention
Action selection
Memory
Learning
Reasoning
Meta-reasoning
Prospection

Allowing the agent to exhibit flexible context-sensitive behaviour, prospectively selecting and controlling the actions that are required to achieve given goals

A cognitive architecture should also be able to develop autonomously so that its performance improves over time with experience




There are three paradigms of cognitive science


Cognitive Science
Cognitivist Systems
Cognitive Architecture
Hybrid Systems
Emergent Systems
The term originated with the work of A. Newell (1990)

# Cognitivist Cognitive Architecture

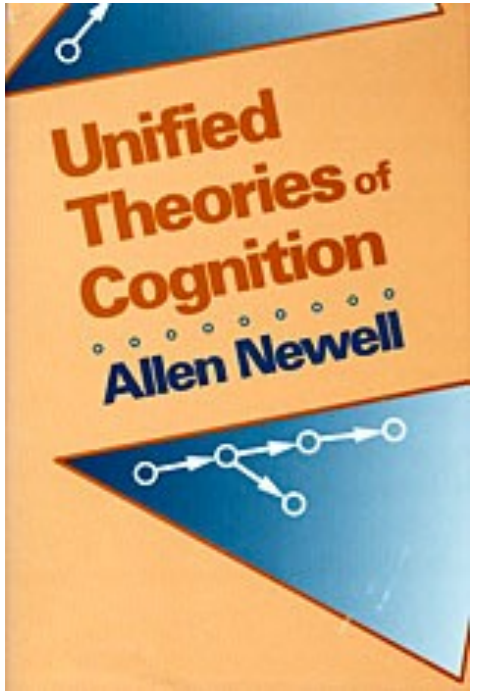
Attempts to create Unified Theories of Cognition (UTC)

UTCs cover a broad range of cognitive issues

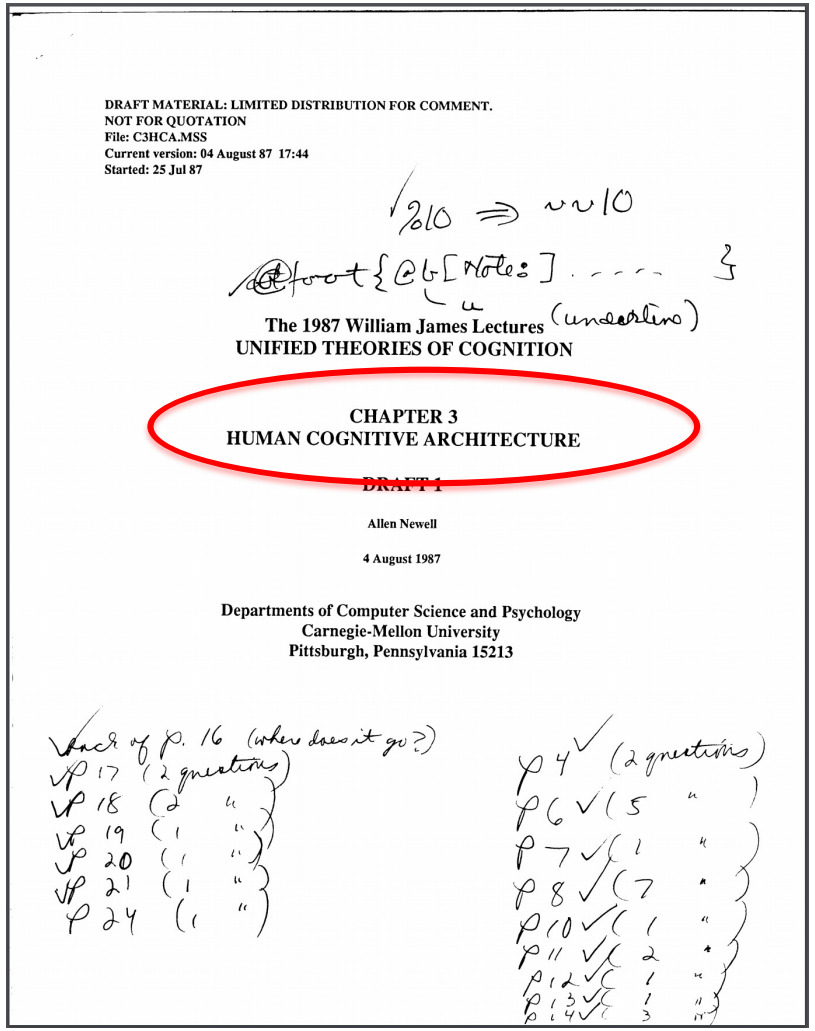
- Attention
- Memory
- Problem solving
- Decision making
- Learning

from several aspects

- Psychology
- Neuroscience
- Computer Science




https://www.hup.harvard.edu/catalog.php?isbn=9780674921016



DRAFT MATERIAL: LIMITED DISTRIBUTION FOR COMMENT.
NOT FOR QUOTATION
File: C3HCA.MSS
Current version: 04 August 87 17:44
Started: 25 Jul 87

The 1987 William James Lectures (underline)
UNIFIED THEORIES OF COGNITION

CHAPTER 3
HUMAN COGNITIVE ARCHITECTURE

DRAFT 1

Allen Newell

4 August 1987

Departments of Computer Science and Psychology
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

http://digitalcollections.library.cmu.edu/awweb/awarchive?type=file&item=352120

# Cognitivist Cognitive Architecture

An encapsulation of a scientific hypothesis about those aspects of human cognition that are

- relatively constant over time and
- relatively independent of task

(Ritter and Young 2001)

# Cognitivist Cognitive Architecture

- Generic computational model:
  - Not domain-specific
  - Not task-specific
- Knowledge provides the required specificity:

> Cognitive Architecture + Knowledge = Cognitive Model

  - Lehman et al. (1998) put it slightly differently:

BEHAVIOR = ARCHITECTURE x CONTENT

# Cognitivist Cognitive Architecture

Knowledge is typically:

- Determined by the designer (explicitly or implicitly)
- Adapted and augmented by machine learning techniques

# Cognitivist Cognitive Architecture

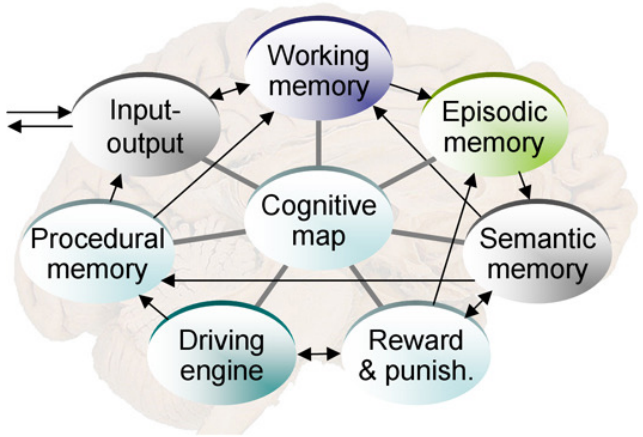
## Overall structure and organization of a cognitive system

- Essential modules
- Essential relations between these modules
- Essential algorithmic and representational details in each module

[Sun 2007]

Commitment to formalisms for representation and processes

(Langley 2005, Langley 2006, Langley et al. 2009)




GMU BICA Architecture (Samsonovich 2010)

# Emergent Cognitive Architecture

Emergent approaches focus on development

- From a primitive state
- To fully cognitive state, over the system's lifetime

https://childmaltreatmentresearchblog.wordpress.com/about/

# Emergent Cognitive Architecture

- Two different views of development
  - Individual
  - Social
- Two complementary theories of cognitive development

Jean Piaget
1896–1980

https://en.wikipedia.org/wiki/Jean_Piaget

Lev Vygotsky
1896–1934

https://en.wikipedia.org/wiki/Lev_Vygotsky

# Emergent Cognitive Architecture

The cognitive architecture is the system's phylogenetic configuration

- The basis for ontogenesis: growth and development
  - Innate skills
  - Core knowledge (cf. E. Spelke)

- A structure in which to embed mechanisms for
  - Perception
  - Action
  - Adaptation
  - Anticipation
  - Motivation
  - … Development of all these

# Emergent Cognitive Architecture

Strong focus on

- Autonomy-preserving, anticipatory, adaptive skill construction
- The morphology of the physical body in which the architecture is embedded

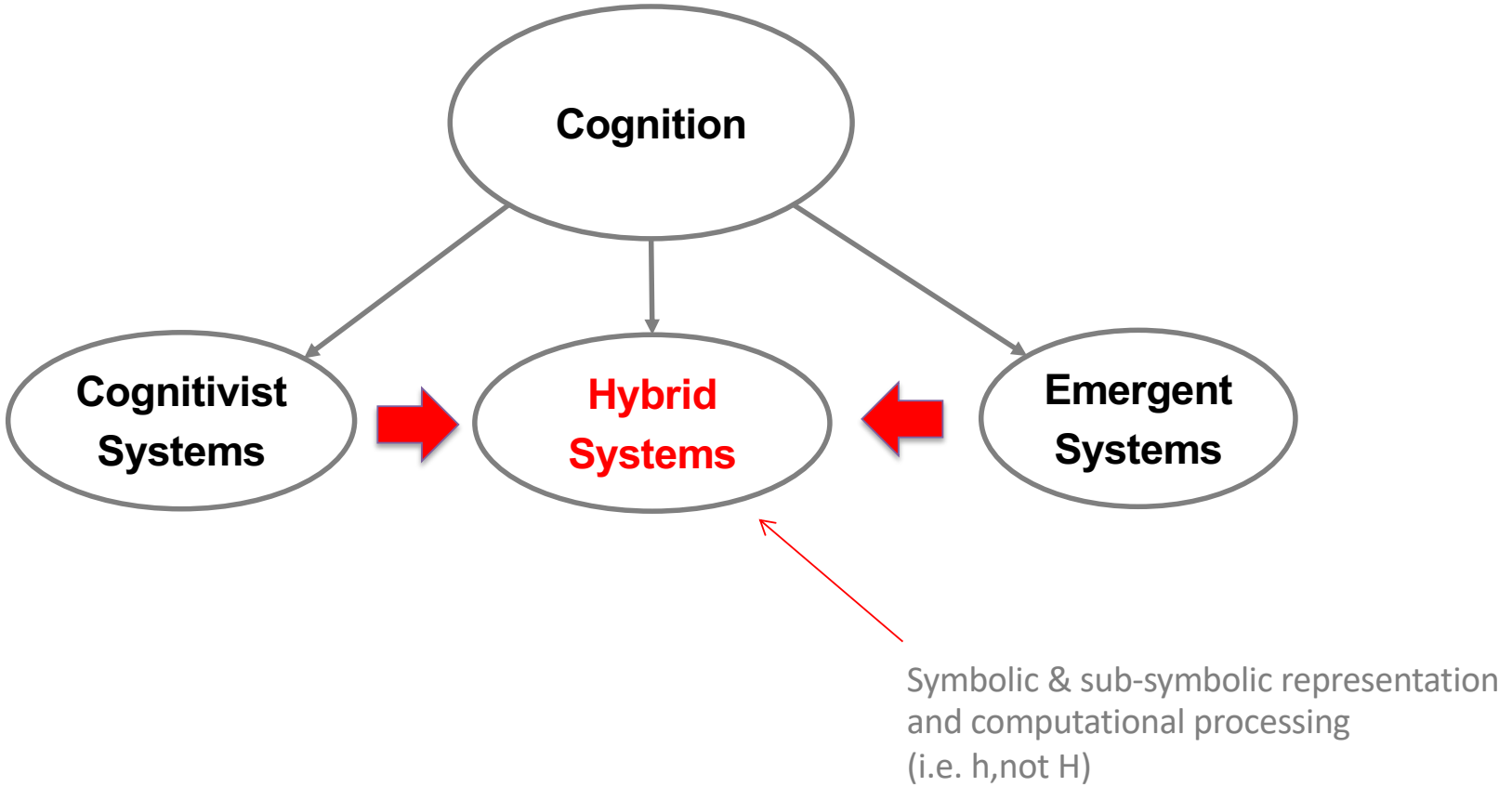
# Emergent Cognitive Architecture

The emergent approach rejects:

- Dualism between mind and body
- Functionalism that treats cognitive mechanisms independently of the physical platform
  - Computational functionalism
  - Robotic functionalism

  (Ziemke, 2016)

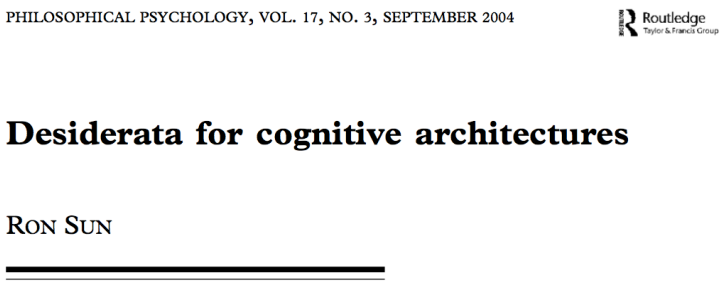
# Hybrid Cognitive Architecture



# Desirable Characteristics of a Cognitive Architecture

# Desirable Characteristics

- Realism
- Behavioral Characteristics
- Cognitive Characteristics



PHILOSOPHICAL PSYCHOLOGY, VOL. 17, NO. 3, SEPTEMBER 2004

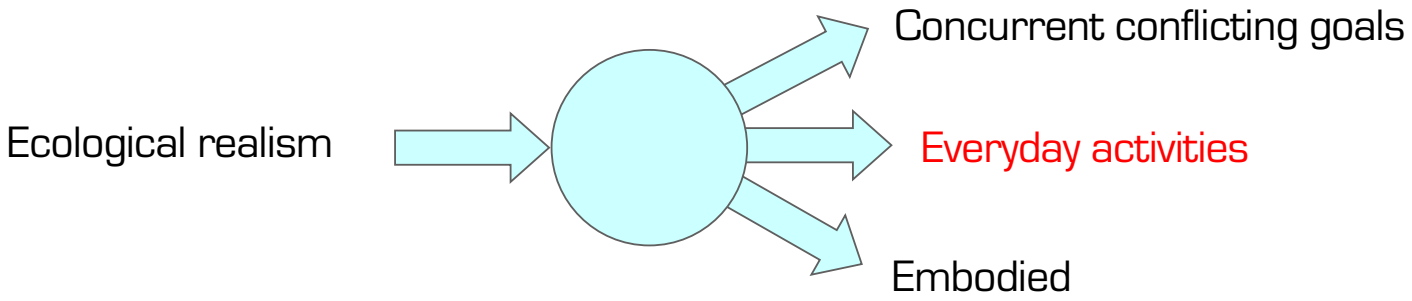
Routledge
Taylor & Francis Group

**Desiderata for cognitive architectures**

RON SUN

- Functional Capabilities (Sun 2004; Langley et al. 2009)
- Development (Krichmar 2012; Vernon et al. 2016)
- Dynamics

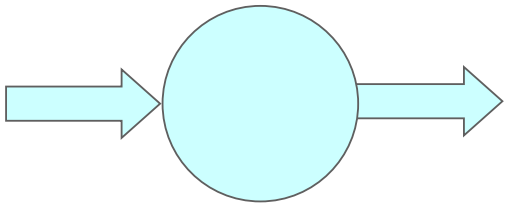
# Desirable Characteristics




(Sun 2004)

# Desirable Characteristics

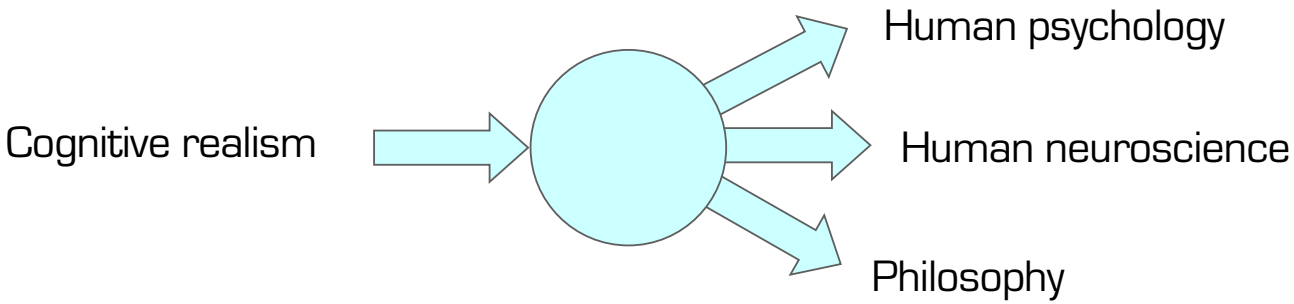
Bio-evolutionary realism



Human intelligence reducible to model of animal intelligence

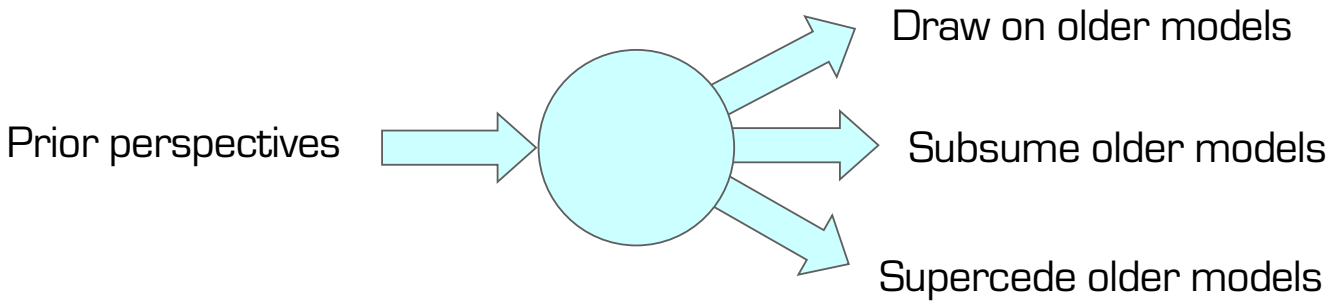
(Sun 2004)

# Desirable Characteristics




(Sun 2004)

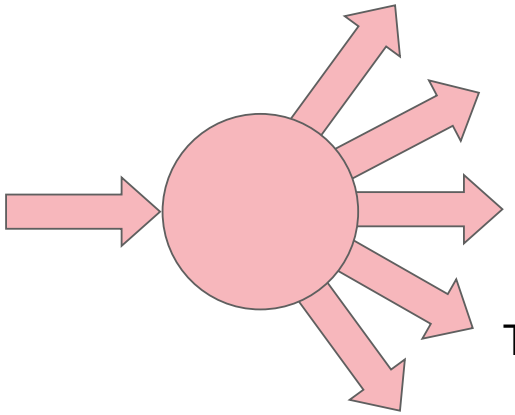
# Desirable Characteristics




(Sun 2004)

# Desirable Characteristics

Behavioural Characteristics
[Sun 2004]



Act & React ...

Simple conceptual schemas

Simple weighing of alternatives

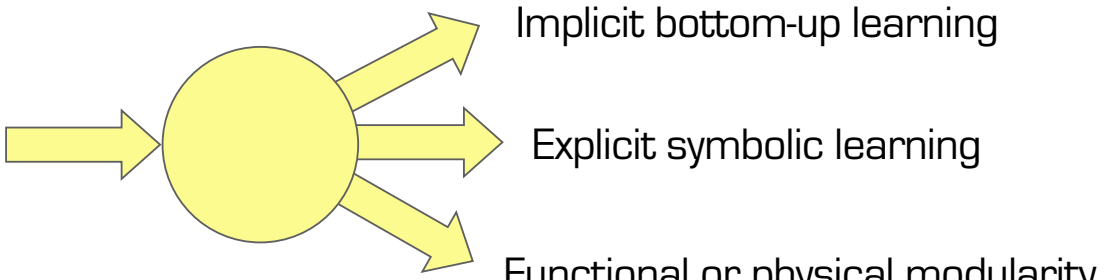
Temporal sequence of actions

Gradually-learned routine behaviours
... trial-and-error adaptation

(Sun 2004)

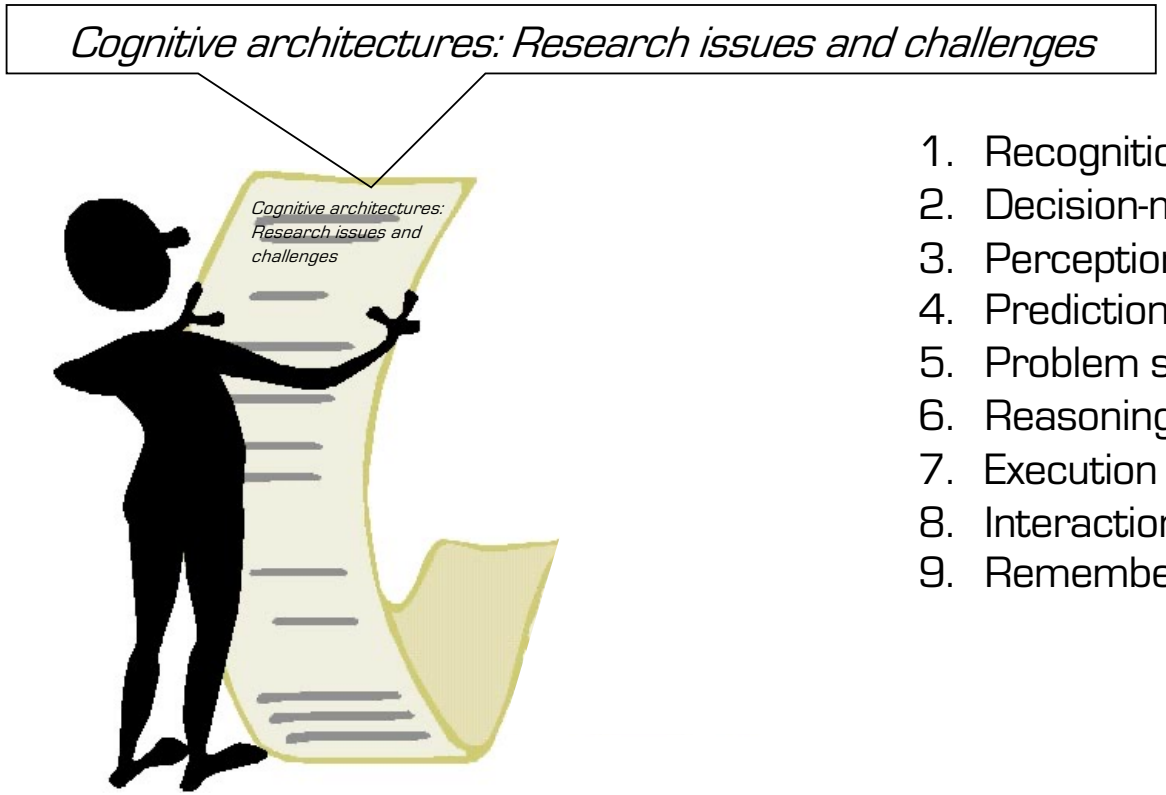
# Desirable Characteristics

Cognitive Characteristics
[Sun 2007]




(Sun 2004)

# Desirable Characteristics




[Langley et al. 2009]

1. Recognition & categorization
2. Decision-making & choice
3. Perception & situation assessment
4. Prediction & monitoring
5. Problem solving & planning
6. Reasoning & belief maintenance
7. Execution & action
8. Interaction & communication
9. Remembering, reflection, & learning

# Desirable Characteristics

*The importance of cognitive architectures …*


(Sun 2007)

1. Perception
2. Categorization
3. Multiple representations
4. Multiple types of memory
5. Decision making
6. Reasoning
7. Planning
8. Problem solving
9. Meta-cognition
10. Communication
11. Action control and execution
12. Several types of learning

The importance of the interconnectivity between these processes

# Desirable Characteristics

PHILOSOPHICAL PSYCHOLOGY, VOL. 17, NO. 3, SEPTEMBER 2004

Routledge
Taylor & Francis Group

**Desiderata for cognitive architectures**

RON SUN

ELSEVIER

Biologically Inspired Cognitive Architectures

Volume 18, October 2016, Pages 116–127


Research article

**Desiderata for developmental cognitive architectures**

David Vernon[a], Claes von Hofsten[b], Luciano Fadiga[c, d]

Show more

http://dx.doi.org/10.1016/j.bica.2016.10.004

Get rights and content

# Desirable Characteristics

Desideratum 1. Value systems and motives

Desideratum 2. Physical embodiment

Desideratum 3. Sensorimotor contingencies

Desideratum 4. Perception

Desideratum 5. Attention

Desideratum 6. Prospective action

Desideratum 7. Declarative and procedural memory

Desideratum 8. Multiple modes of learning

Desideratum 9. Internal simulation

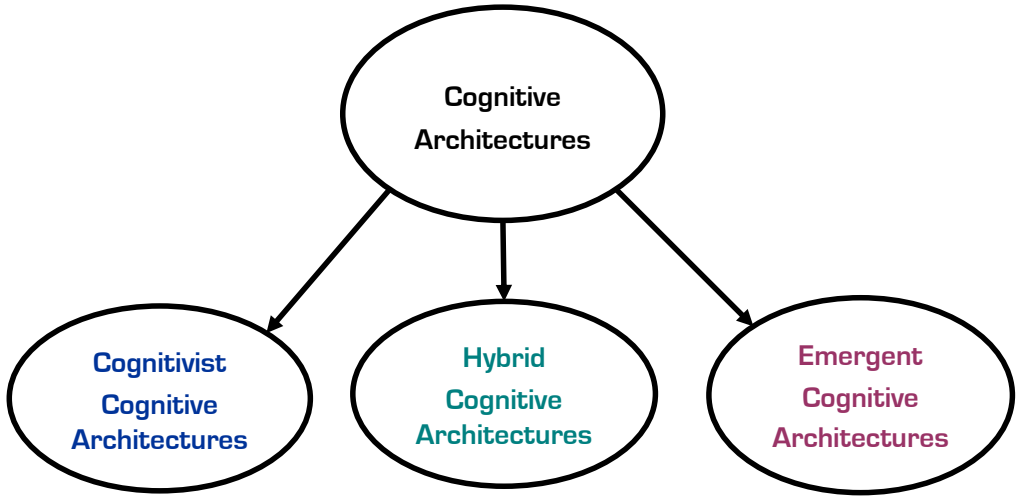
Desideratum 10. Constitutive autonomy

# Facets of a Cognitive Architecture

- Component functionality
- Component interconnectivity
- System dynamics

## Organizational decomposition

- Explicit inter-connectivity
- Representational formalism
- Algorithmic formalism




**Framework in which to embed knowledge**

- Memories
- Formalisms for learning
- Programming mechanism

**Phylogeny - basis for development**

- Innate skills & core knowledge
- Memories
- Formalism for autonomy
- Formalism for development

# Core Cognitive Abilities

1. Perception
2. Attention
3. Action selection
4. Memory
5. Learning
6. Reasoning
7. Metacognition
8. Prospection

I. Kotseruba and J. Tsotsos. 40 years of cognitive architectures: core cognitive abilities and practical applications. Artificial Intelligence Review, Vol. 53, No. 1, pp. 17-94, 2020.

Not included in [Kotseruba and Tsotsos 2020]

# Core Cognitive Abilities

1. Perception
2. Attention
3. Action selection
4. Memory
5. Learning
6. Reasoning
7. Metacognition
8. Prospection

Makes use of many sensory modalities, e.g. vision, audition, and haptic (tactile and kinesthetic)

Perception transforms raw input into the system's internal representation

# Core Cognitive Abilities

1. Perception
2. Attention
3. Action selection
4. Memory
5. Learning
6. Reasoning
7. Metacognition
8. Prospection

Reduces the information a cognitive system has to process by selecting relevant information & filtering out irrelevant information

Selective mechanisms

Choose one entity from many, e.g. gaze & viewpoint

Restrictive mechanisms

Choose some entities from many:
Priming what to look for or where to look for it

Suppressive mechanisms

Suppress some entities from many
i.e. features, objects, or locations that are not relevant

[Kotseruba and Tsotsos 2020]

# Core Cognitive Abilities

1. Perception
2. Attention
3. Action selection
4. Memory
5. Learning
6. Reasoning
7. Metacognition
8. Prospection

Determines what the agent should do next

Planning

Determines a sequence of steps to read a certain goal prior to execution of the plan

Dynamic action selection

Selection of one action based on knowledge at the time, typically using winner-take-all, probabilistic, or pre-defined order selection mechanisms

# Core Cognitive Abilities

1. Perception
2. Attention
3. Action selection
4. Memory
5. Learning
6. Reasoning
7. Metacognition
8. Prospection

Kotseruba and Tsotsos identify six types of memory

| | |
|---|---|
| Short-term sensory memory | recent percepts |
| Short-term working memory | information relevant to current task |
| Long-term episodic memory | key to anticipation; autobiographical |
| Long-term semantic memory | general knowledge about the world |
| Long-term procedural memory | motor skills |
| Long-term global memory | for architectures that don't draw a type-duration distinction: |

# Core Cognitive Abilities

1. Perception
2. Attention
3. Action selection
4. Memory
5. Learning
6. Reasoning
7. Metacognition
8. Prospection

The ability of a system to improve performance over time through the acquisition of knowledge or skill

| | |
|---|---|
| Declarative learning … | explicit knowledge acquisition |
| Non-declarative learning … | perceptual, procedural, associative, non-associative learning |

Supervised learning
Unsupervised learning
Reinforcement learning

# Core Cognitive Abilities

1. Perception
2. Attention
3. Action selection
4. Memory
5. Learning
6. Reasoning
7. Metacognition
8. Prospection

The ability to logically and systematically process knowledge, typically to infer conclusions.

Three classical forms of logical inference: deduction, induction, abduction.

Reasoning focusses on the practical objective of finding the next (best) action to perform

# Core Cognitive Abilities

1. Perception
2. Attention
3. Action selection
4. Memory
5. Learning
6. Reasoning
7. Metacognition
8. Prospection

Metacognition refers to the ability of a cognitive system has to monitor its internal cognitive processes, reason about them, and adapt them

Metacognition is needed for social cognition if the agent is to form a theory of mind, also known as perspective-taking, i.e. the ability to infer the cognitive states of other agents with which it is interacting

# Core Cognitive Abilities

1. Perception
2. Attention
3. Action selection
4. Memory
5. Learning
6. Reasoning
7. Metacognition
8. Prospection

Prospection – the capacity to anticipate the future – is one of the hallmark attributes of cognition

It lies at the heart of the other core characteristics of a cognitive agent: autonomy, perception, action, learning, and adaptation

[Vernon 2014]

It is central to action since actions are goal-directed and guided by prospective information

[von Hofsten 2009]

Internal simulation plays a key role in prospection

# Cognitive Architectures Forums


http://bicasociety.org/


## Biologically Inspired Cognitive Architectures

› Supports Open Access

Editor in Chief: A. Samsonovich

https://www.journals.elsevier.com/biologically-inspired-cognitive-architectures/


## Cognitive Systems Research

Editor-in-Chief: T. R. Besold

› View Editorial Board

› **CiteScore: 3.2** ⓘ **Impact Factor: 1.902** ⓘ

ISSN: 1389-0417

https://www.journals.elsevier.com/cognitive-systems-research

# Recommended Reading

D. Vernon, Artificial Cognitive Systems – A Primer, MIT Press, 2014; Chapter 3.

D. Vernon. "Cognitive Architectures" in Cognitive Robotics, A. Cangelosi and M. Asada (Eds.), MIT Press, in press.

A. Lieto, M. Bhatt, A. Oltramari, and D. Vernon, "The Role of Cognitive Architectures in General Artificial Intelligence", editorial for a special issue on "Cognitive Architectures for Artificial Minds", Cognitive Systems Research, Vol. 48, pp. 1-3, 2017.

# References


Krichmar, J. L. (2012). Design principles for biologically inspired cognitive architectures. Biologically Inspired Cognitive Architectures 1, 73–81.

Langley, P. (2005). An adaptive architecture for physical agents. In IEEE/WIC/ACM International Conference on Intelligent Agent Technology, Compiegne, France, pp. 18–25. IEEE Computer Society Press.

Langley, P. (2006). Cognitive architectures and general intelligent systems. AI Magazine 27(2), 33–44.

Langley, P., J. E. Laird, and S. Rogers (2009). Cognitive architectures: Research issues and challenges. Cognitive Systems Research 10 (2), 141– 160.

Lehman, J. F., J. E. Laird, and P. S. Rosenbloom (1998). A gentle intro- duction to soar, an architecture for human cognition. In S. Sternberg and D. Scarborough (Eds.), Invitation to Cognitive Science, Volume 4: Methods, Models, and Conceptual Issues. Cambridge, MA: MIT Press.

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

# References


Ritter, F. E. and R. M. Young (2001). Introduction to this special issue on using cognitive models to improve interface design. International Journal of Human-Computer Studies 55, 1–14.

Samsonovich, A. (2010). Toward a unified catalog of implemented cog- nitive architectures. In Proc. the Conference on Biologically Inspired Cognitive Architectures, pp. 195–244.

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

Sun, R. (2007). The importance of cognitive architectures: an analysis based on CLARION. Journal of Experimental & Theoretical Artificial Intelligence 19(2), 159–193.

Vernon, D., C. von Hofsten, and L. Fadiga (2016). Desiderata for developmental cognitive architectures. Biologically Inspired Cognitive Architectures 18, 116–127.

Ziemke, T. (2016). The body of knowledge: On the role of the living body in grounding embodied cognition. BioSystems 148, 4–11.