AI Architectures

Architectural Thinking for Intelligent Systems

Winter 2019/2020

Prof. Dr. habil.Jana Koehler

Agenda

- AI agent architectures
- The AI agent model
- Shakey's software architecture
- Belief Desire Intention (BDI) architecture
- Subsumption architecture
- Soar cognitive architecture

Concept of an Agent in Artificial Intelligence

(1) Ability to perceive environment

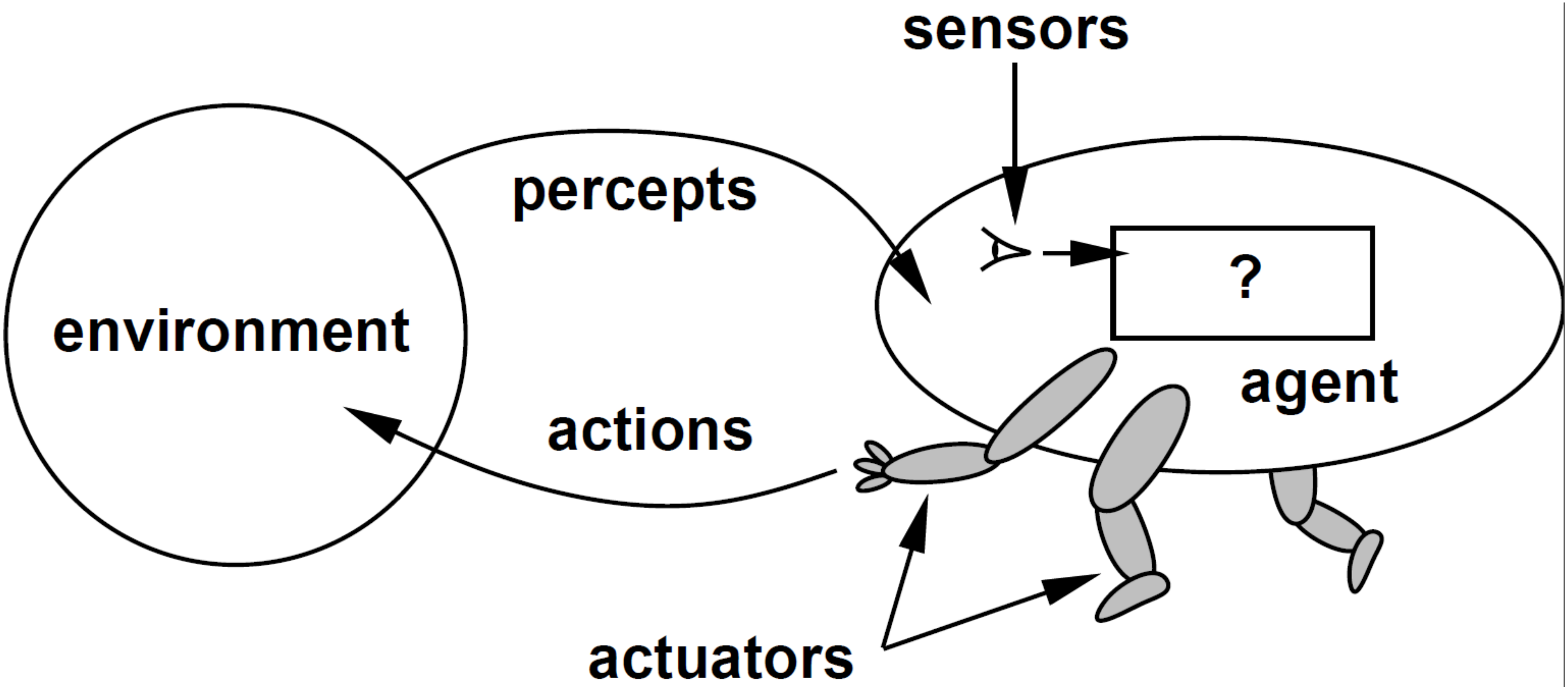
(2) Observations used to make Decisions

(3) Decisions will result in actions

(4) Decision must be RATIONAL

    (1) best possible action the agent can take

The AI Agent Model



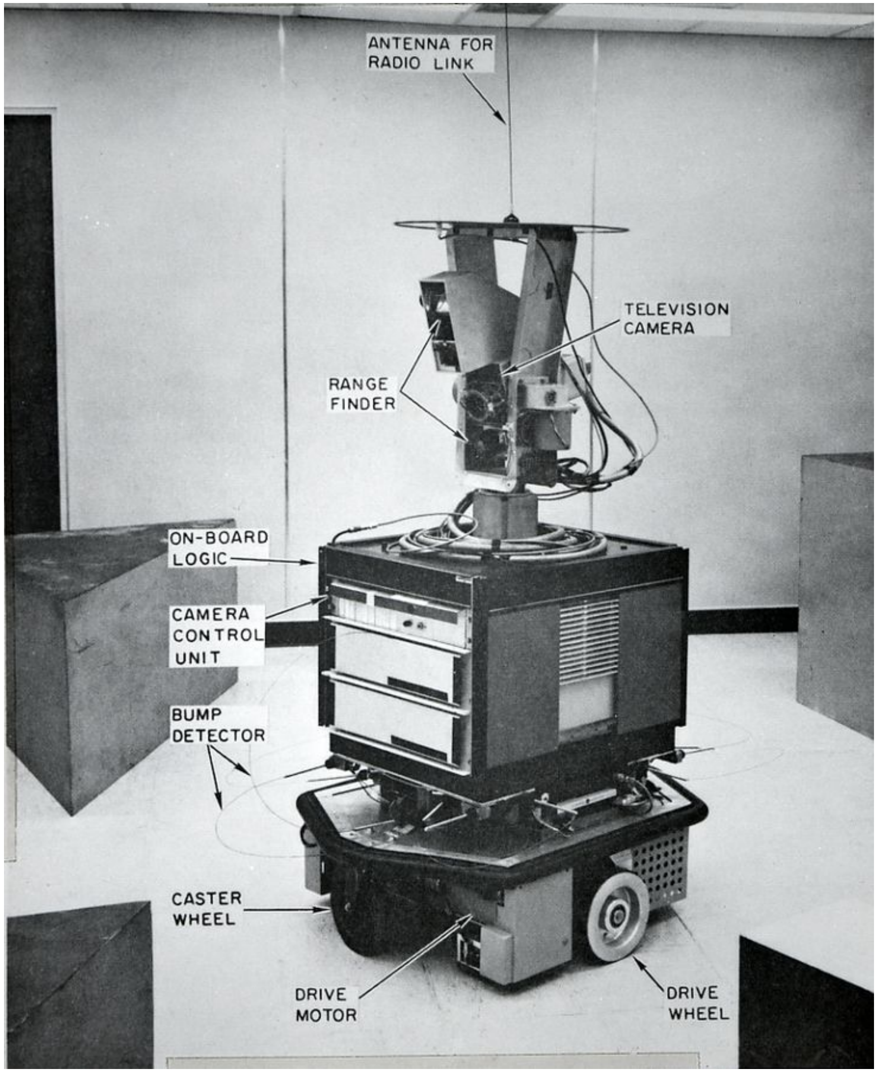
Russel & Norvig

Basic Ideas in AI Agent Architectures

- Layering
- State-based control
- Deliberation
  - agents need to think
- Emergent Behavior
  - complex, intelligent behavior results from less complex, lower-level behaviors
- Embodiment
  - to successfully act in our world, the agent needs to have a body and physical awareness

Shakey Project Team in 1983

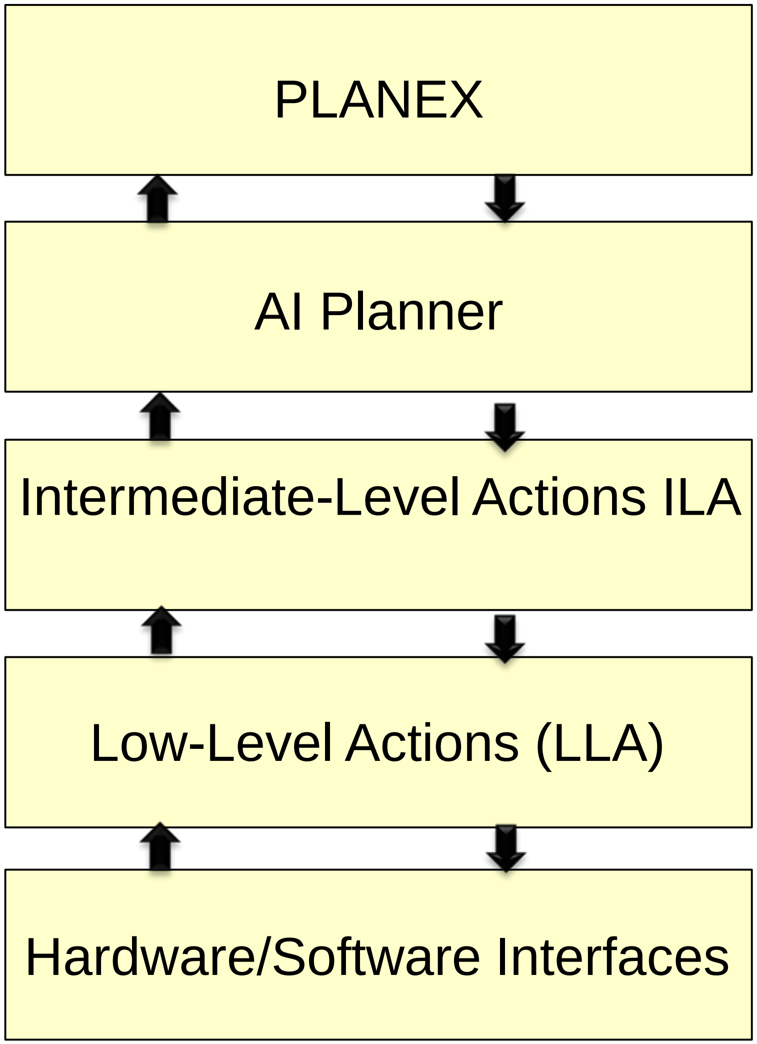
Shakey Hardware and Major Achievements



- Layered Robot Control
- Line Detection in Images
- A* Search Algorithm
- STRIPS / AI Planning

IEEE Milestone Award Feb 16, 2017

Shakey's Architecture



Plan execution
receive goal, execute plan, monitor execution

AI Planner (STRIPS)
construct sequences of ILAs to achieve goals
a plan is a macro operation (MACROP)

Library of packaged LLAs
«instinctive» abilities of the robot
«push» «goto», sensing actions

basic actions the robot can execute «roll» «tilt»
low-level sensing

elementary physical capabilites
start motors, turn wheels, send data, …
connection to PDP-15/PDP-10

Low-Level Actions (LLA)

- Actions: roll(forward, backward), turn, tilt/pan (head movements), focus/iris (camera), shoot (take pictures), range (get rangefinder measure),
- Tactile sensors: 7 catwhiskers, 1 pushbar
- AT, THETA: position and bearing of robot in global coordinate system, statistical uncertainties in DAT, DTHETA
- Notion of conflicting actions: an action needs to be completed before another conflicting action can start
  - move of the robot head is in conflict with taking pictures
  - robot will not move when a catwhisker is engaged

Intermediate-Level Actions (ILA)

- Set of carefully defined subroutines that are available for problem solving
- Main challenge: detecting and recovering from errors

GoToAdjacentRoom(r1 <<roomFrom>>, d <<door>>, r2 <<roomTo>>)

Label	Predicate	Actions	Control
1	not room(r1)	-	return
2	room(r2)	-	return
3	true	set(s, doorstatus(d))	4
4	infrontof(d) & s=open	bumblethrough(r1,d,r2)	2
	near(d) & s=open	align(r1,d,r2)	
	near(d) & s=unknown	doorpic(d)	3
	s=closed		return
	true	navto(nearpt(r1,d))	4

another ILA

function to compute goal position

Escape from Infinite Action Loops

- Navto is tried as long as the robot is not near the door

- Use a separate monitor program that is aware of the complete state of the system (and the world!) and that can decide if an action was successful and has brought the robot closer to its goal

- simpler: each ILA keeps a status record
  - how often an action has been taken?
  - if threshold is exceeded, conclude that no progress is made and return control

Shakey's World Model

- 5 objects: doors, walls, rooms, objects, the robot
- atomic predicates are used to describe these entities

- type(object «object») — *type(o1 object)*
- name(object name) — *name(o1 box1)*
- at(object number number) — *at(o1 3.1 5.2)*
- inroom(object inroom) — *inroom(o1 r1)*
- shape(object shape) — *shape(o1 wedge)*
- radius(object number) — *radius(o1 3.1)*

- conjunctions/disjunctions and universal quantification
  - *all objects are of shape wedge and in room A or B*

Operations on the World Model

- ASSERT – add a new statement to the model
- FETCH – query the model FETCH(INROOM $ R1)
  - returns all objects which are in room R1
- DELETE – delete statements matching a formula
- REPLACE (combination of ASSERT and DELETE)

- Allows the robot to keep the internal model of the world in sync with its sensory inputs and action effects

Famous Built-In Assumptions

- **Closed-World Assumption**
  - All facts that cannot be deduced from a knowledge base are assumed to be false.
- **Unique-Name Assumption**
  - Different names always refer to different entities in the world unless it can be deduced otherwise.

Physical Symbol System Hypothesis - Alan Newell, Herbert Simon

- "A physical symbol system has the ***necessary*** and ***sufficient*** means for general intelligent action" (1976,1980)

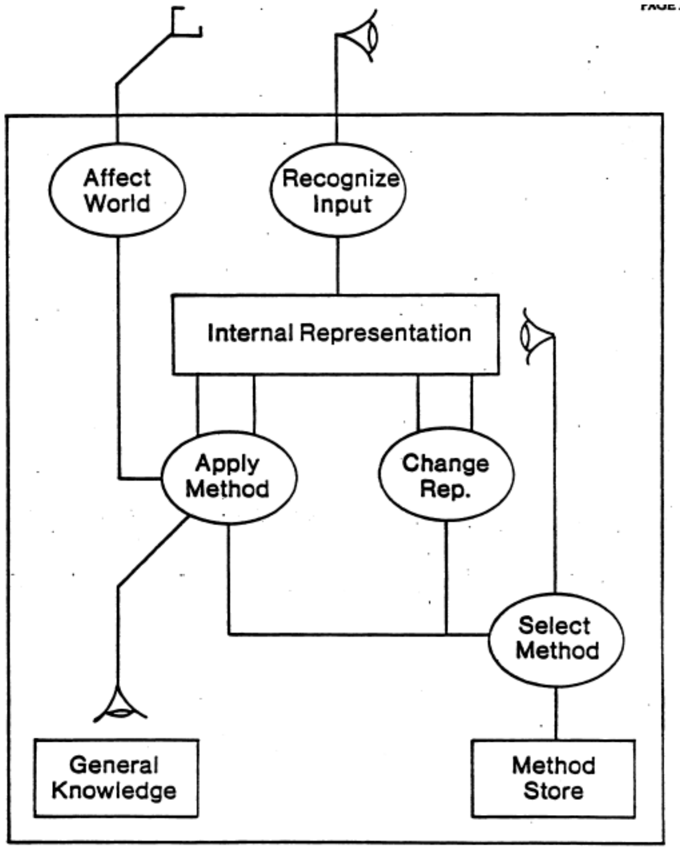
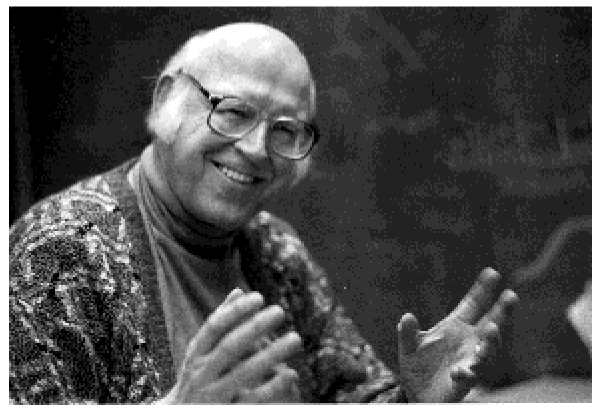


Figure 2-1: Functional diagram of general intelligent agent (after Newell & Simon, 1972)



Does Intelligence require knowledge?

What is knowledge?

Can knowledge be represented in logic?

Belief Desire Intention (Michael Bratman 1987)

- Beliefs: what the agent thinks the world is like
- Desire: what the agent wants (can be contradictory)
- Intention: what goals the agent pursues

Strong Consistency Principle

If A intends to φ, then, if A is not criticizably irrational, this intention must be able to be put together with the rest of my intentions into a plan which is consistent with my beliefs.

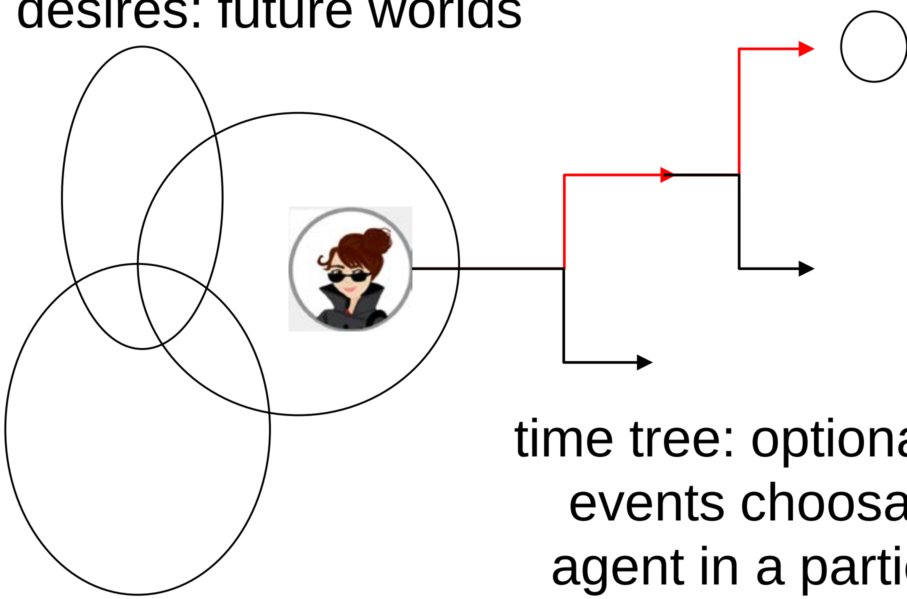
Weak Consistency Principle

If A intends to φ, then, if A is not criticizably irrational, this intention must be consistent with the rest of my intentions.

BDI Architecture (Anand Rao und Michael Georgeff 1995)

current possible worlds
(interpreting beliefs)
desires: future worlds

intention: the future path the agent chooses to follow



goals: suitable to achieve situations the agent desires to be in

time tree: optional courses of events choosable by the agent in a particular world

Strong Realism: The agent beliefs she can optionally achieve her goals by carefully choosing the events (actions) she executes.

Axioms for BDI-Based Agents

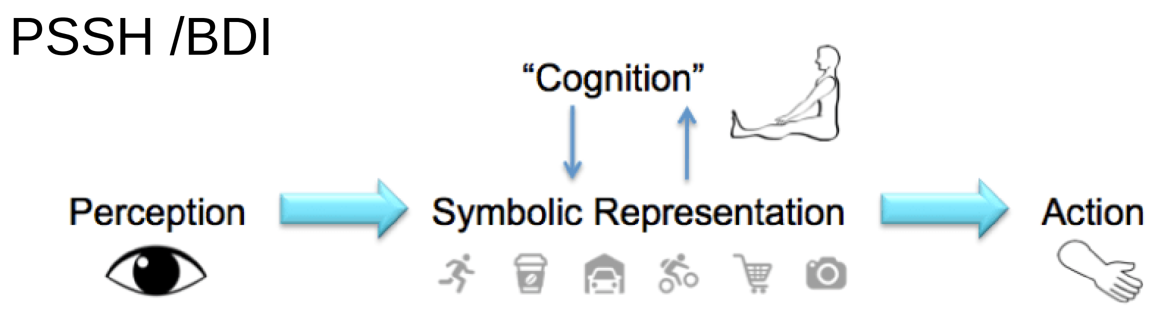
- If the agent adopts a goal, she believes this goal.
  - there is at least one path to a possible world where the goal is true
  - the agent does not need to believe it will ever reach this world
- If the agent intends something, it is also a goal.
- The agent is committed to try execute actions achieving its intentions.
  - does not imply that these actions will succeed
  - the agent can also execute other actions
- The agent is aware of all primitive events occurring in the world.
- The agent will finally abandon an intention.
  - a *blindly committed agent* will only abandon an intention when she believes she has achieved this intention

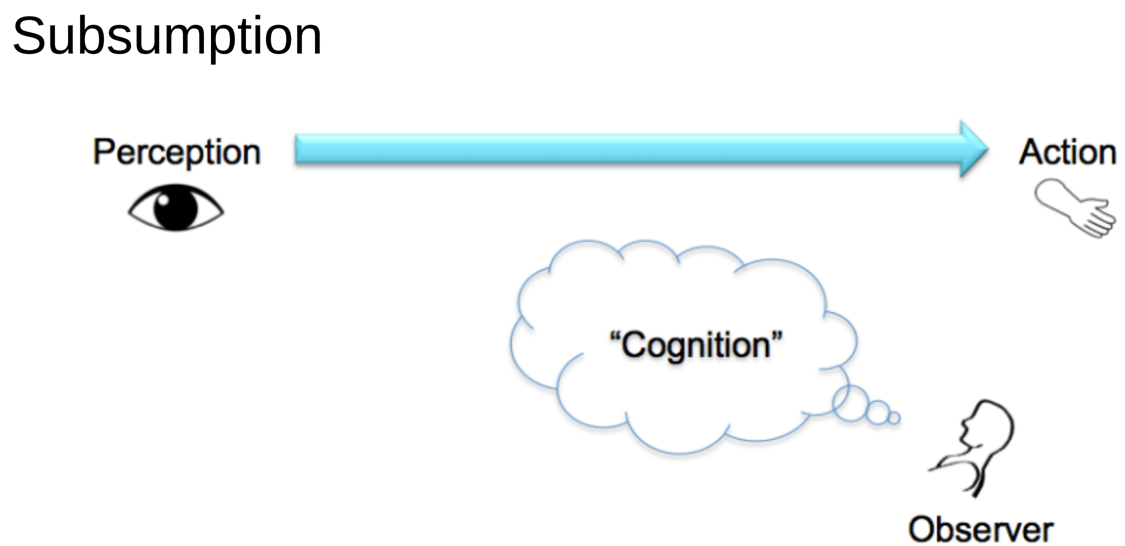
BDI Frameworks

- https://en.wikipedia.org/wiki/Belief%E2%80%93desire%E2%80%93intention_software_model

- Criticism (of the theoretical BDI framework)
  - Does not address the ability of the agent to learn
  - Do we need less of BDI or more?
  - No explicit representation of goals
  - No lookahead planning
- Practical implementations address these questions

Subsumption Architecture (Rodney Brooks, 1985)





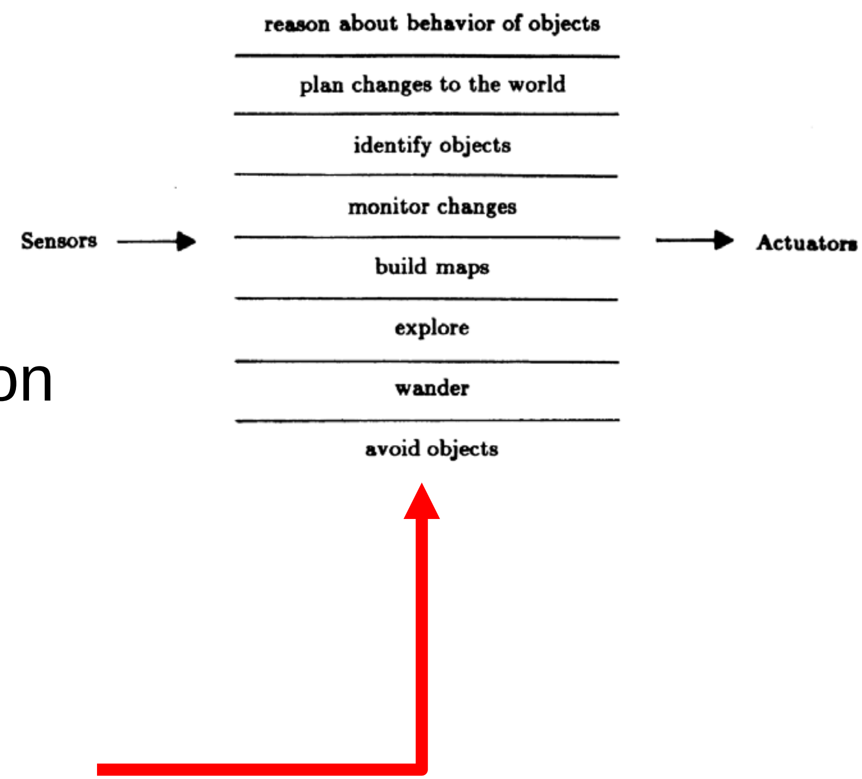
Behaviors of a Robot

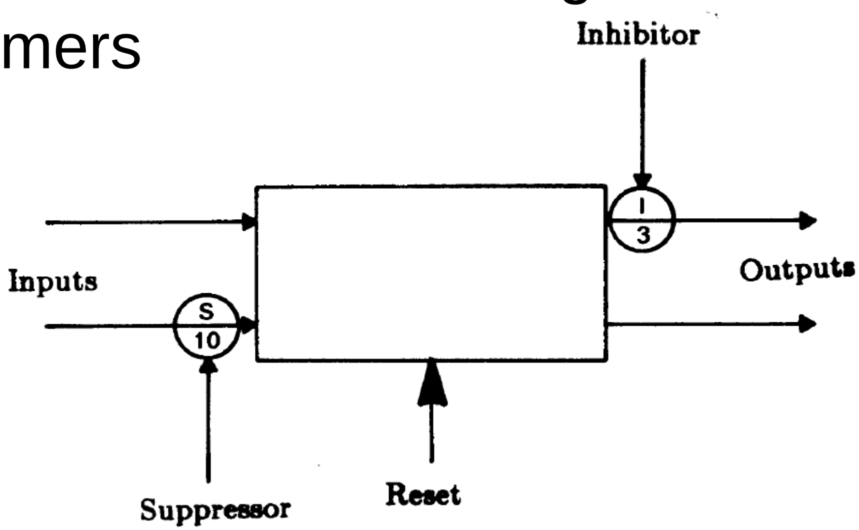
- Work your way up the evolutionary chain by layering on new behaviours

➢ Did not scale to really sophisticated machines

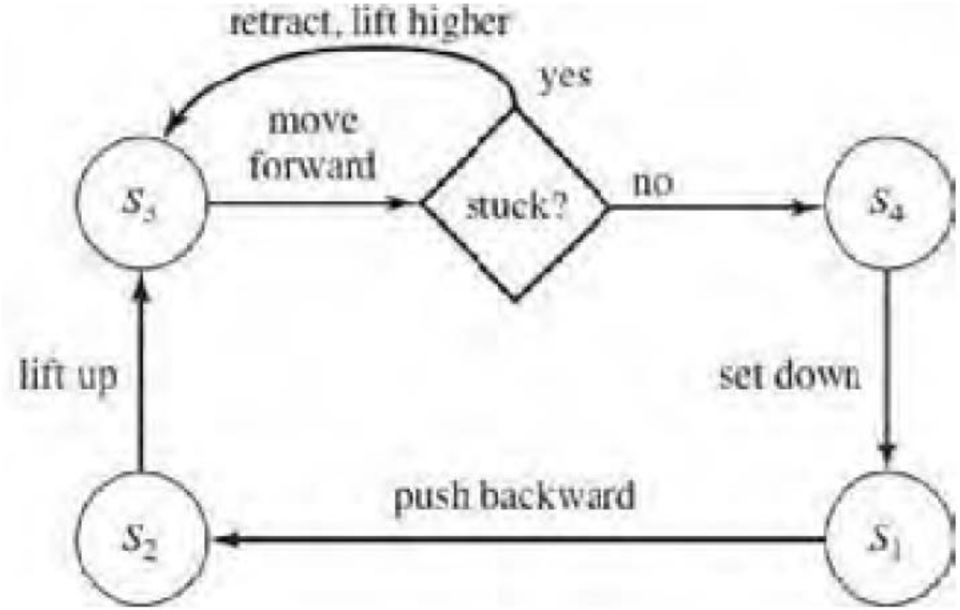
Augmented Finite State Machine AFSM Controllers

- Information messages from sensor readings or expired timers set values in memory
- Memory settings trigger state changes in the FSM
- AFSM can recombine information to create new messages or set timers





Genghis AFSM to control a single leg



Notice that this AFSM reacts to sensor feedback: if a leg is stuck during the forward swinging phase, it will be lifted increasingly higher.

57 AFSM in the final robot
http://people.csail.mit.edu/brooks/papers/AIM-1091.pdf

http://www.ai.mit.edu/projects/genghis/genghis.html

Soar – A General Cognitive Architecture (since 1987)

- Since 1987 - *State, Operator And Result*
- Artificial computational processes that act like human cognitive systems
  - artificial consciousness – which cannot only respond, but also think, perceive, and believe like a human
- Soar creates its own subgoals and learn continuously from its own experience
  - architecture framework and programming language
  - code, tutorials etc. at https://soar.eecs.umich.edu/


John Laird

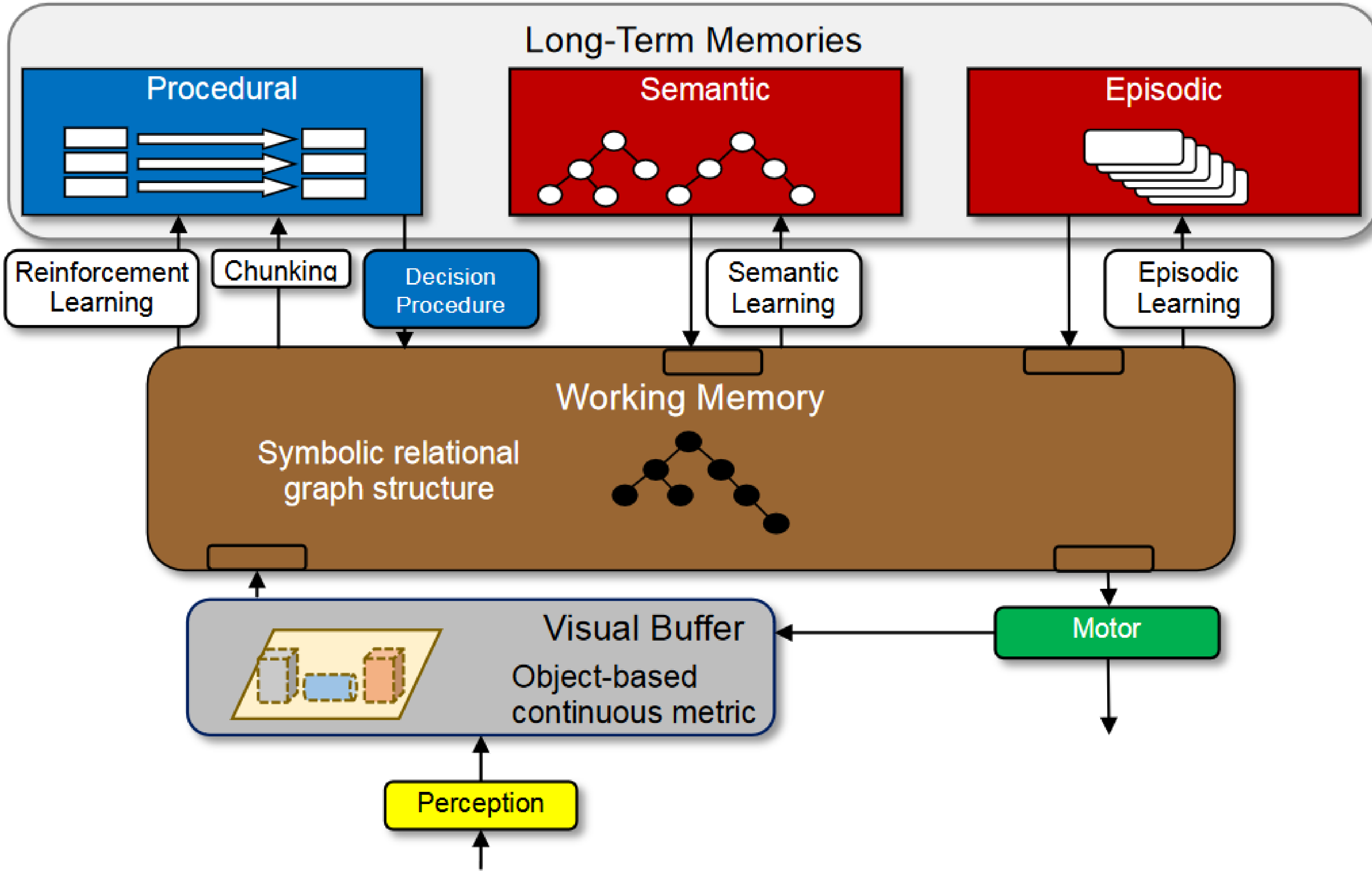

Allen Newell


Paul Rosenbloom

Fundamental Concepts

- All problem solving in Soar is regarded as a search through a problem space in which you apply an operator to a state to get a result
- *Long-term memory* for declarative (semantic), procedural, episodic knowledge
- *Preferences* guide decisions and contain control knowledge about the acceptability and desirability of actions
- *Goals* are determined when decision has come to an impasse – solve the impasse in a new context
- *Explanation-based chunking* summarizes solutions that achieved goals for the agent to learn
- *Spreading activation*: similar to neuronal firings in brains

Architecture



Summary

- Agent architectures rely on fundamental assumptions about cognition and intelligence
- Essential principles are layering, state-based control, memory, reasoning mechanisms, and learning
- BDI-like architectures include complex reasoning mechanisms and explicit knowledge models
- Subsumption-like architectures achieve emergent behavior by layers and networks of state-based control
- Cognitive architectures draw inspiration from humans and link reasoning with learning to achieve intelligence

- There is much more work to be done …

Working Questions

1. What is the rational agent as defined by AI?

2. Which key architectural styles and patterns do you recognize in the agent architectures we discussed?

3. In which form can infinite action loops occur in actions of your system? How can recognize and escape from these loops? How can you describe a world model for your behavior? Which properties and objects do you need?

4. How do you see the relationship between the world model in AI and domain-driven design?

Working Questions Ctd.

6. What is your opinion about the Physical Symbol System Hypothesis? Find arguments for and against the PSSH.

7. Do you think that intelligent agents should pursue their own goals?

8. Can a decision impasse happen in your system architecture?

9. Which role can Learning play in your system? What should the agent learn? How can the agent access the required information? How can you measure improvement?