The Importance of Architecture for Achieving Human-level AI

John Laird

University of Michigan

June 17, 2005

25th Soar Workshop

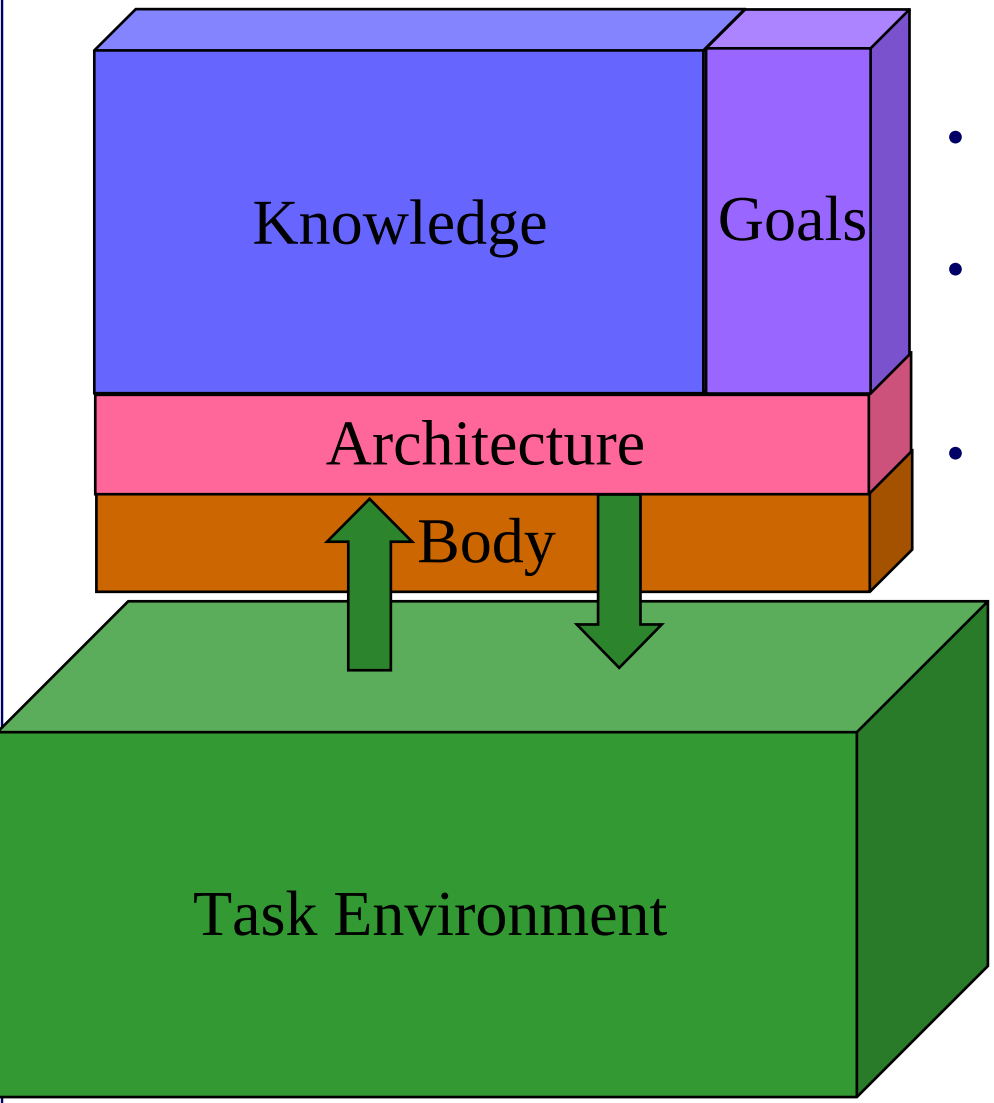
laird@umich.edu

Requirements for Human-Level AI

- Behave flexibly as a function of the environment
- Exhibit adaptive (rational, goal-oriented) behavior
- Operate in real time
- Operate in a rich, complex, detailed environment
  - Perceive an immense amount of changing detail
  - Use vast amounts of knowledge
  - Control a motor system of many degrees of freedom
- Use symbols and abstractions
- Use language, both natural and artificial
- Learn from the environment and from experience
- Live autonomously within a social community
- Exhibit self-awareness and a sense of self
- *All of these can apply to almost any task*

(Adapted from Newell, Rosenbloom, Laird, 1989)

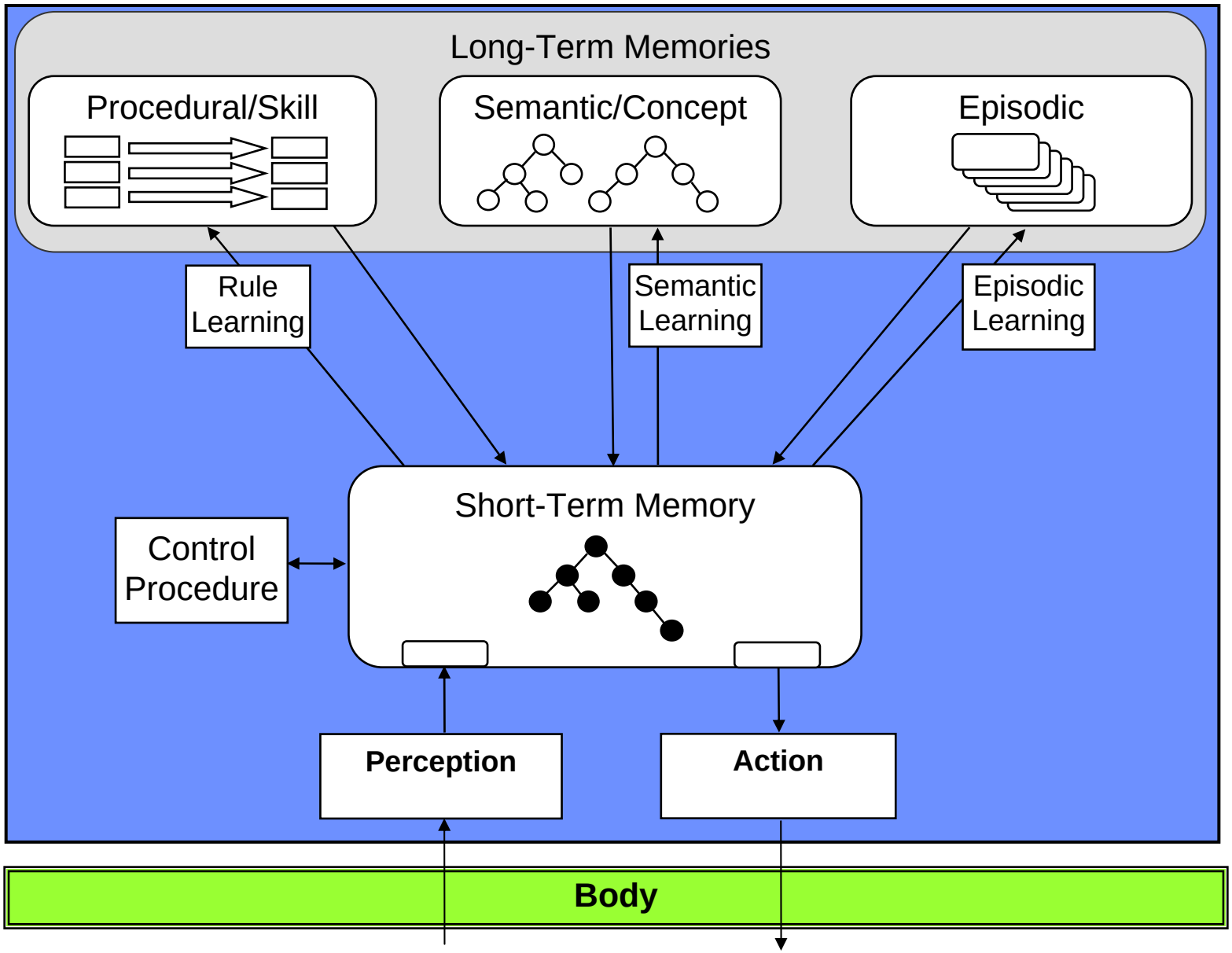
Human-level AI



Architecture = Structure

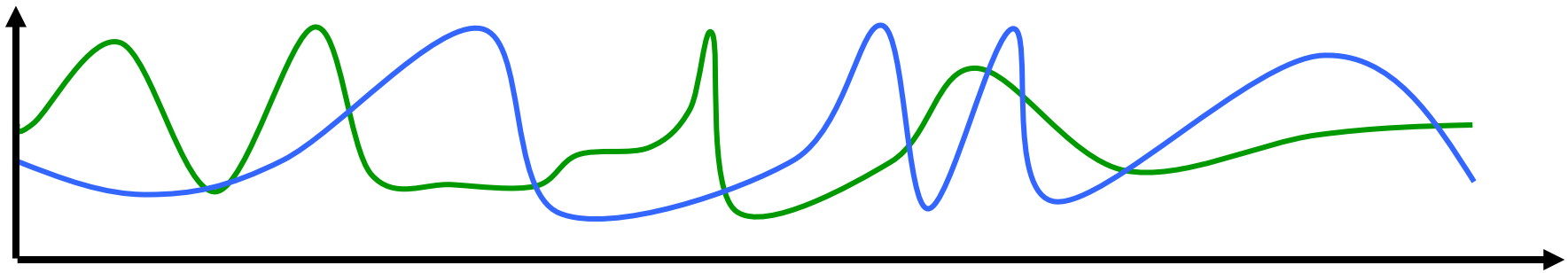
- Fixed mechanisms underlying cognition
  - Memories, processing elements, control
- Purpose:
  - Bring all relevant knowledge to bear in select actions to achieve goals
- Examples:
  - Soar, ACT-R, EPIC, ICARUS, 3T, CLARION, dMARS, CAPS, …

Generic Architecture



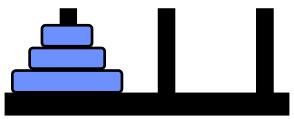
Why Architecture Matters

- Architecture determines:
  - The complexity profile of an agent's computations
  - The primitive units of reasoning/deliberation
  - The primitive units of knowledge
  - What is fixed and unchanging vs. what is programmed/learned
- Architecture provides:
  - The building blocks for creating a complete agent
  - A framework for integrating multiple capabilities
- Architecture is an attempt to capture/formalize regularities
  - Forces the theorist to be consistent across tasks

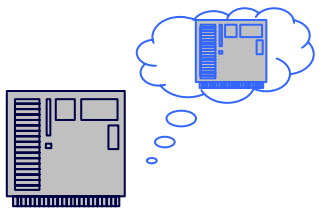


Architecture-based Research

- Pick subset of desired capabilities, performance, and behavior
- Analyze computational requirements
- Design and implement architecture
- Build a variety of agents that stress capabilities
- Evaluate agents and architecture
- Expand desired set of capabilities, performance, behaviors



Towers of Hanoi



R1-Soar

Hero-Soar

TacAir-Soar
& RWA-Soar

Soar MOUTbot

Utility of a Research Strategy?

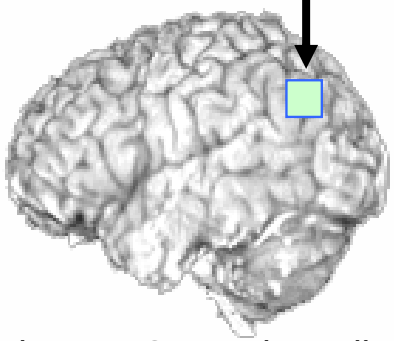
- Efficient at achieving research goal
- Focuses research on critical issues
- Supports incremental progress, results, and evaluation

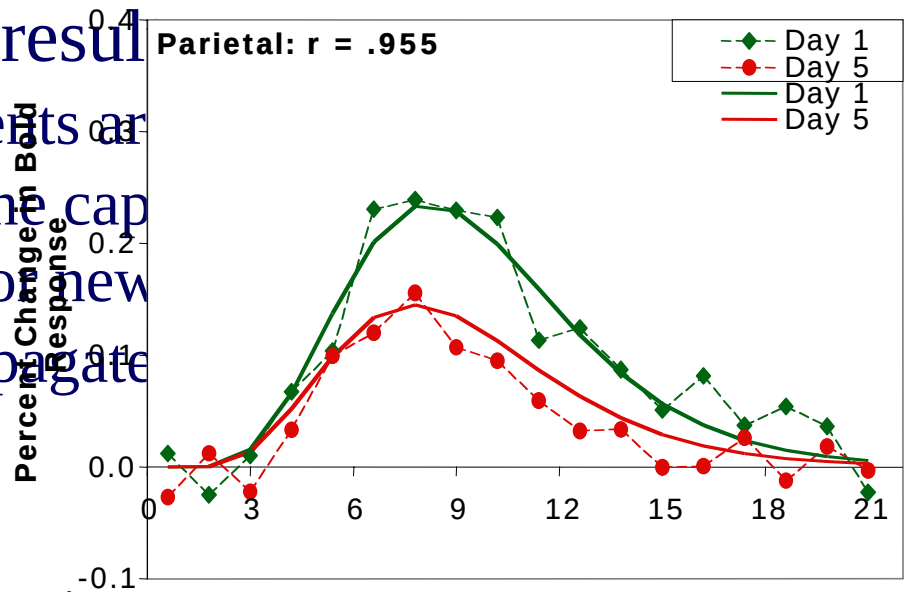
Efficient at Achieving Research Goal

- Supports parallel exploration of solution space
  - Alternative architectures
  - Research can be decomposed into architecture and knowledge
- Integrates research results from all available sources
  - Applications, AI, psychology, neuro-science
- Supports accumulation of resul
  - New architectural components ar
  - Constraints derived from one cap
  - Architec

3x-5=7

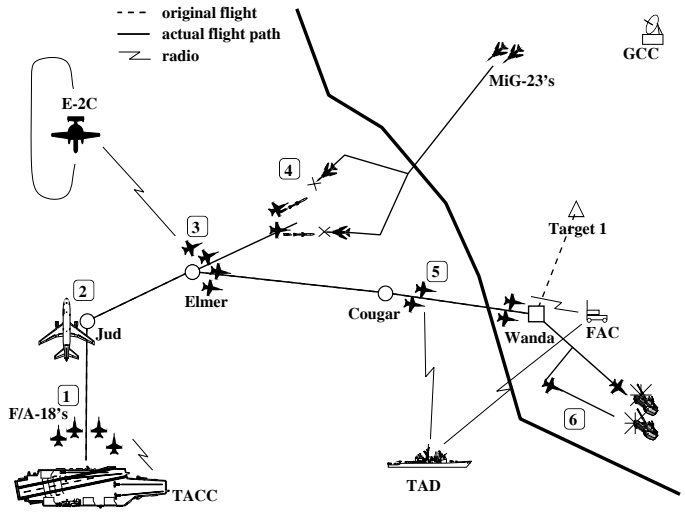
Parietal/Imaginal: BA 39/40
[ACT-R Problem State]





Focuses Research on Critical Issues: Creating Complete Agents

- Coarse-grain integration
  - Connecting all capabilities, from perception to action
- Fine-grain integration of capabilities/knowledge
  - Dynamic intermixing of perception, situational assessment, planning, language, reaction, …
- Ubiquitous learning that
  - is not deliberately cared for and controlled
  - is incremental and real-time
  - doesn't interfere with reasoning
  - *impacts everything an agent does*
- Long-term existence
  - Scaling to tasks employing large bodies of knowledge
  - Behave for hours or days, not minutes
  - Generation of goals, drives, internal rewards, …



Supports Incremental Progress, Results, and Evaluation

- Has useful intermediate results
  - Can build useful end-to-end systems today, even if approximations to human-level intelligence
- Supports evaluation of incremental progress
  - Capabilities of agents developed with architecture
    - Ability to meet requirements for human-level behavior
  - Separates architecture from knowledge, goals, environment
    - Amount of knowledge required to achieve a level of performance
    - Competence on complete tasks with given knowledge
    - Breadth of knowledge that can be encoded/used
    - Breadth and difficulty of goals that can be attempted
  - Comparison to human behavior
    - Match behavior, reaction time, error rates, …

Concluding Remarks