GALIS

Gated Attractors Learning Instruction Sequences

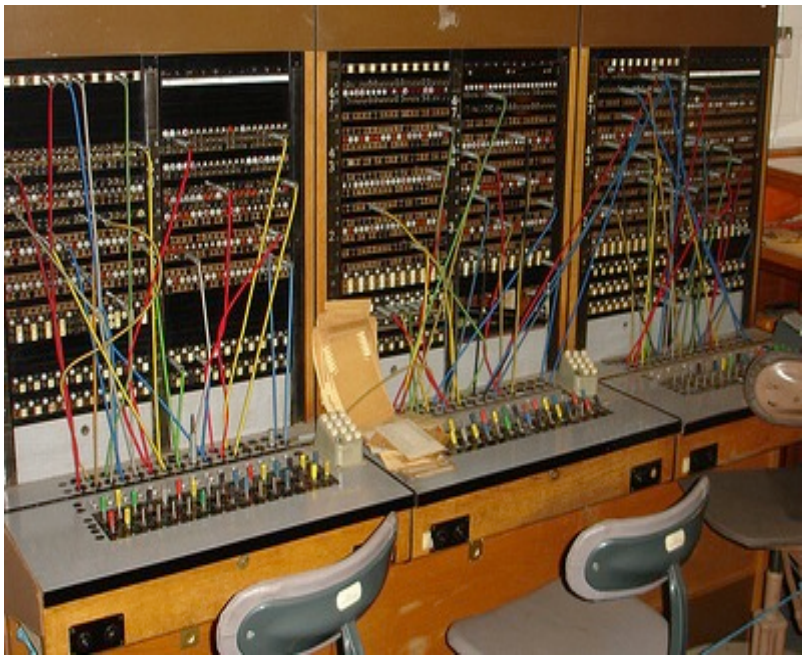
Jared Sylvester

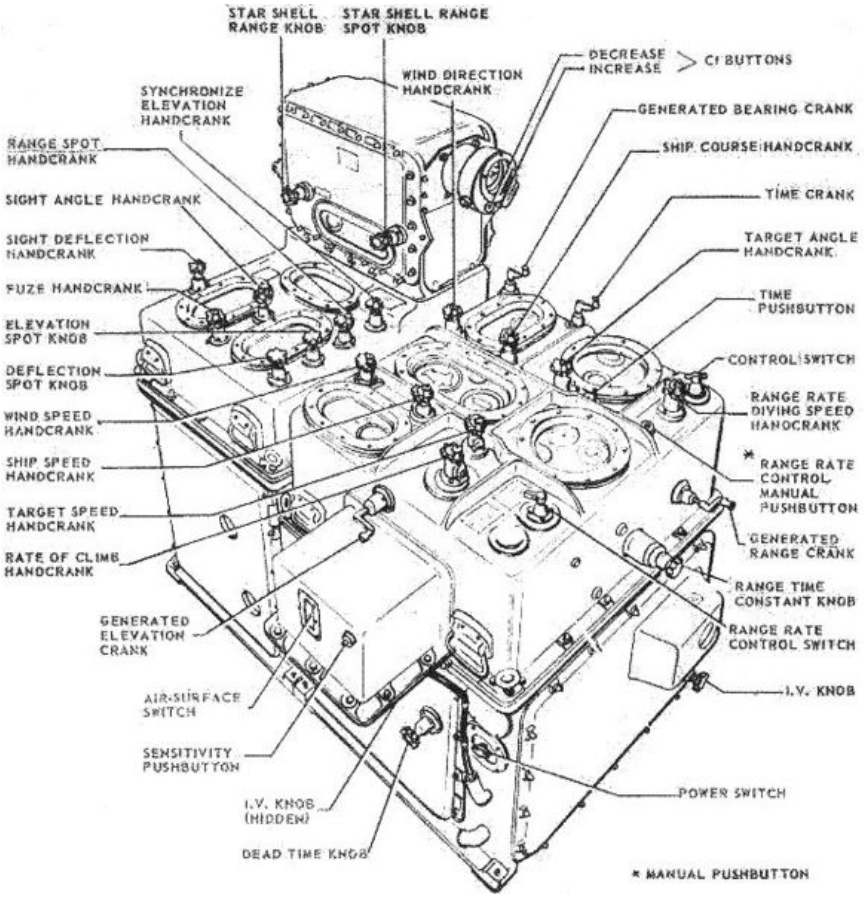
7 May 2013

Computer Science
UNIVERSITY OF MARYLAND

Motivation

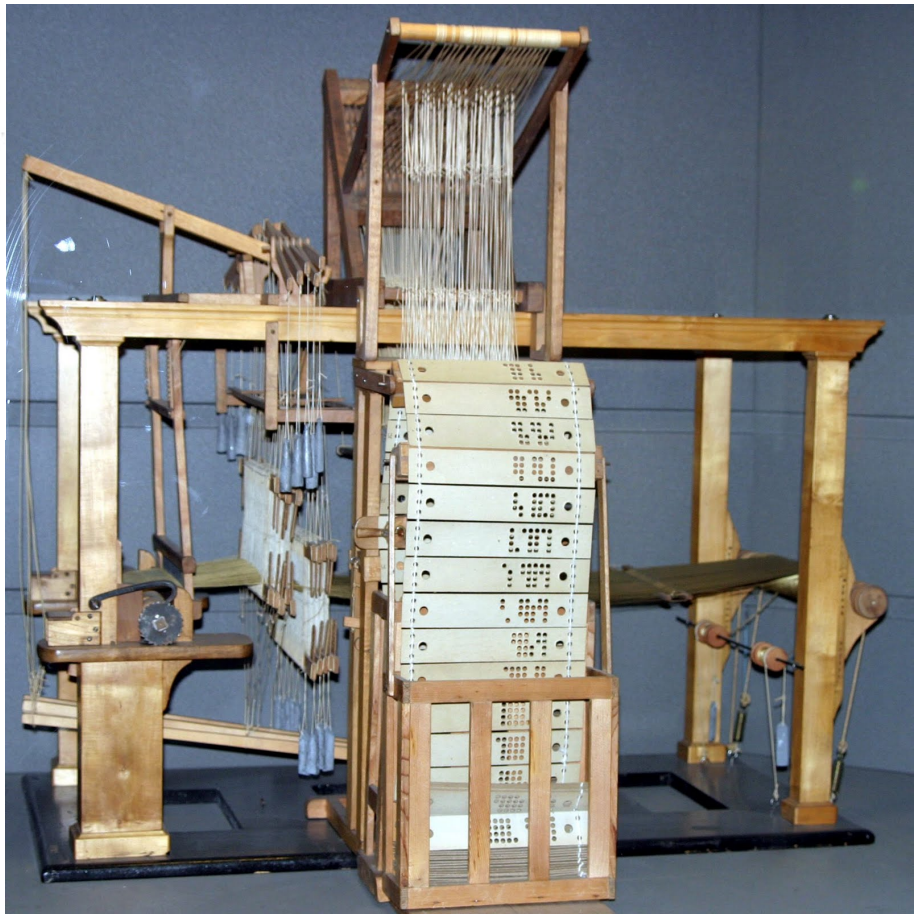
- Gap between neural & symbolic AI systems
  - Neural: perception, motor control, …
  - Symbolic: planning, goals, rules, …
- Neural systems are too "hard-wired"
  - Behavior is baked into architecture
  - New problems require entirely new systems





USN Mk I
Fire Control Computer
1936–1969

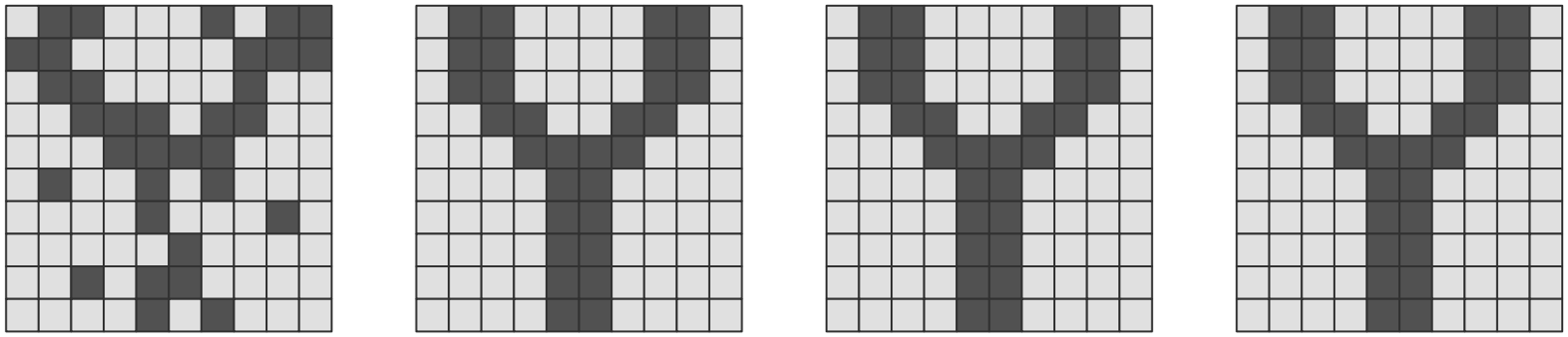
Jacquard Loom
1801–



Goal

- Generalizable model of cognitive control
  - Learned, not hard-wired into network structure
  - Base behavior on memory contents
- Two type of memory/learning:
  - Memory of perceptual stimuli
  - Memory of task procedures
- Biological inspiration:
  - Network of regions, recurrent attractor nets, gating, distributed representations, Hebbian learning

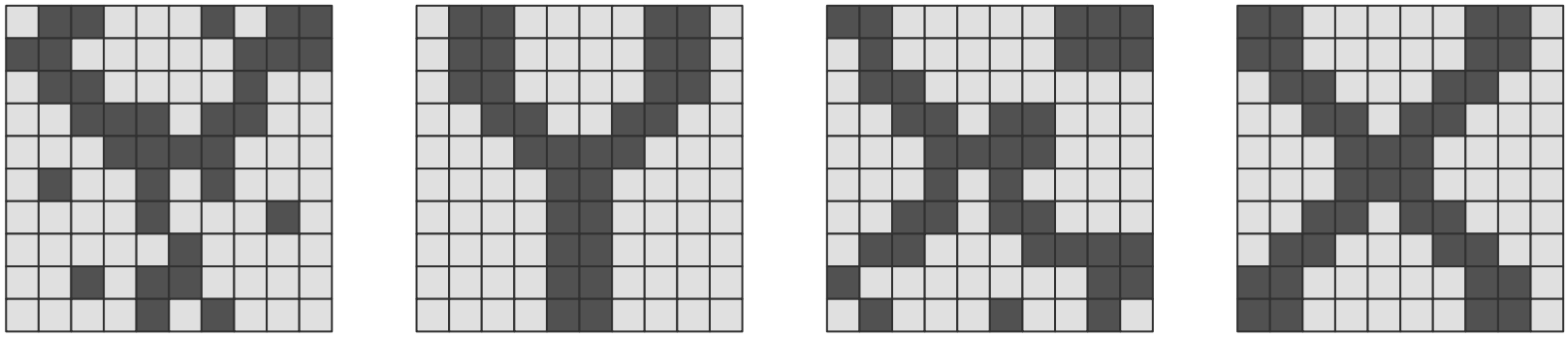
Attractor Net Memories



$$w_{ij}^t = (1 - k_D) w_{ij}^{t-1} + \frac{1}{N} a_i^t a_j^t (1 - \delta_{ij})$$

- Stored patterns are attractors
  - Form auto-associative memory
- But fixed-point attractors
  - Network gets “stuck” in attractor basin

Sequential Attractor Nets

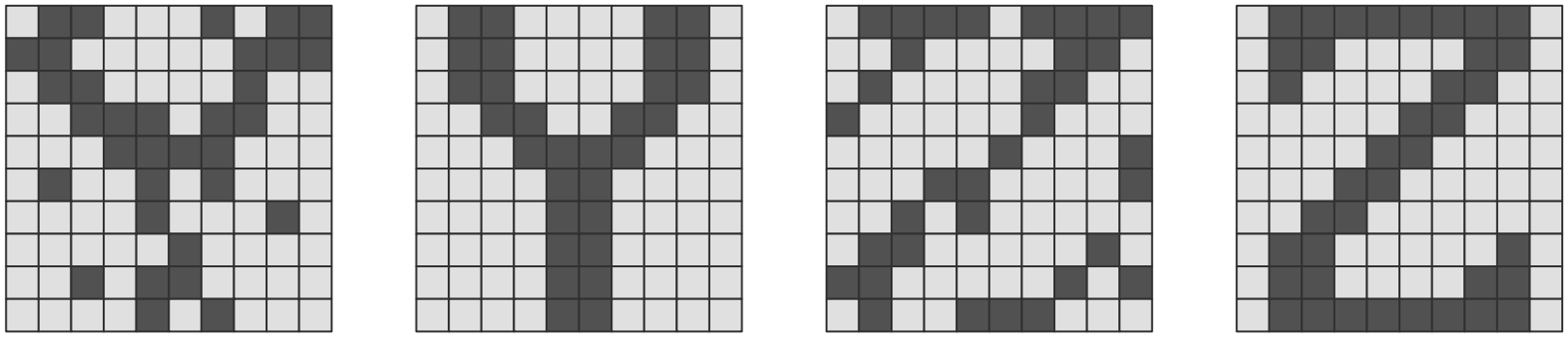


- Dynamic thresholds
  - Increase when node's state remains unchanged
  - Harder for node to stay in the same state

Reggia, Sylvester, Weems & Bunting "A simple oscillatory short-term memory model." BICA 2009.

Winder, Reggia, Weems & Bunting. "An oscillatory Hebbian network model of short-term memory." Neural Computation, 2009.

Ordered Sequential Attractors



- Asymmetric weights
  - Correlate activity with other nodes' previous activity

$$v_{ij}^{t} = (1 - k_D)v_{ij}^{t-1} + \frac{1}{N}a_i^t a_j^{t-1}$$

  - Network transitions between attractors in order

Sylvester, Reggia, Weems & Bunting. "A temporally asymmetric Hebbian network for sequential working memory." ICCM 2010.

Adding Cognitive Control

- Modeled Running Memory Span task
  - Can match human behavioral results
  - But all control was exogenous
- For internal control, use multiple networks
  - Network of attractor networks
  - Controlled by gating
  - Learn processing of sequences

Control Mechanism

- Built around attractor networks
- Trained prior to task beginning
- Directs the model by operating gates
- Core is "Instruction Sequence Memory"

Control Mechanism

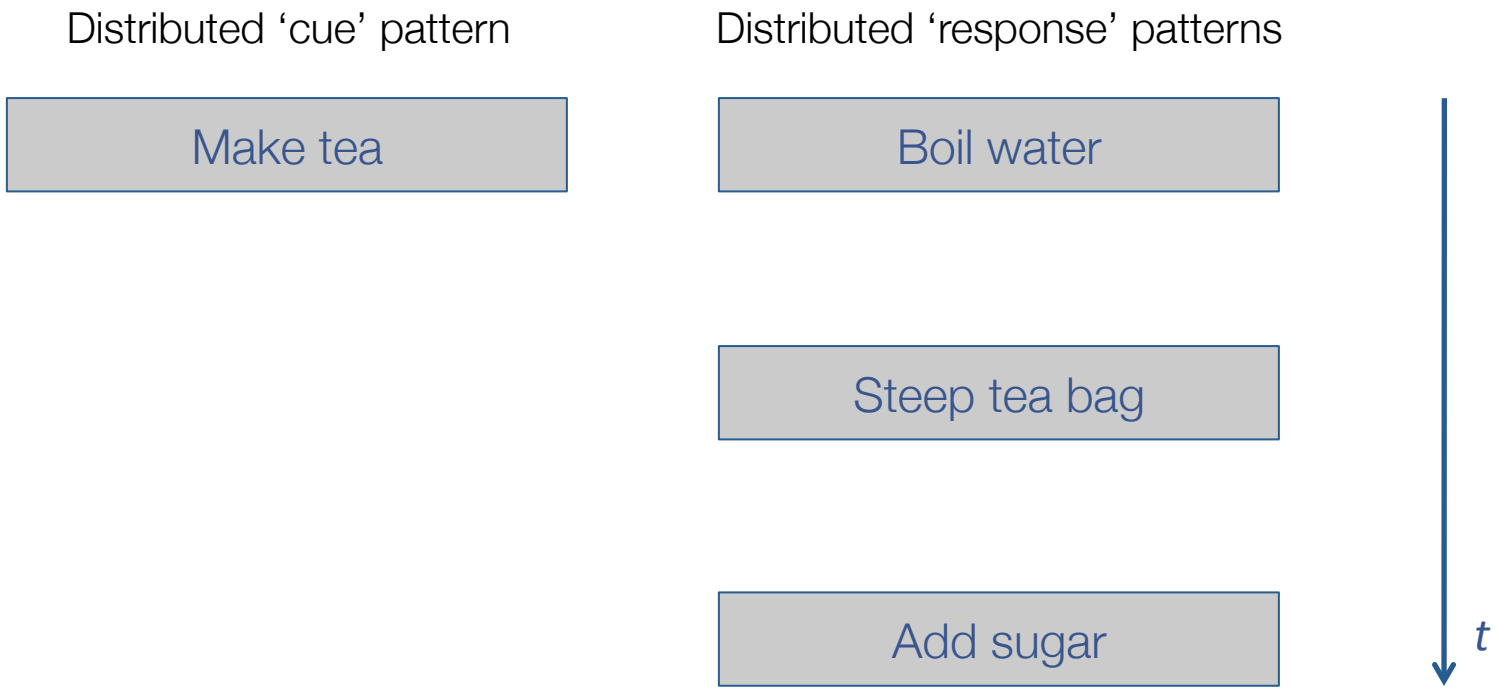
- Built around attractor networks
- Trained prior to task beginning
- Directs the model by operating gates
- Core is "Instruction Sequence Memory"
  - Stores sequence of steps to do subtasks
  - Multiple sequences stored simultaneously
  - Divided into cue & response sections

Instruction Sequence Memory

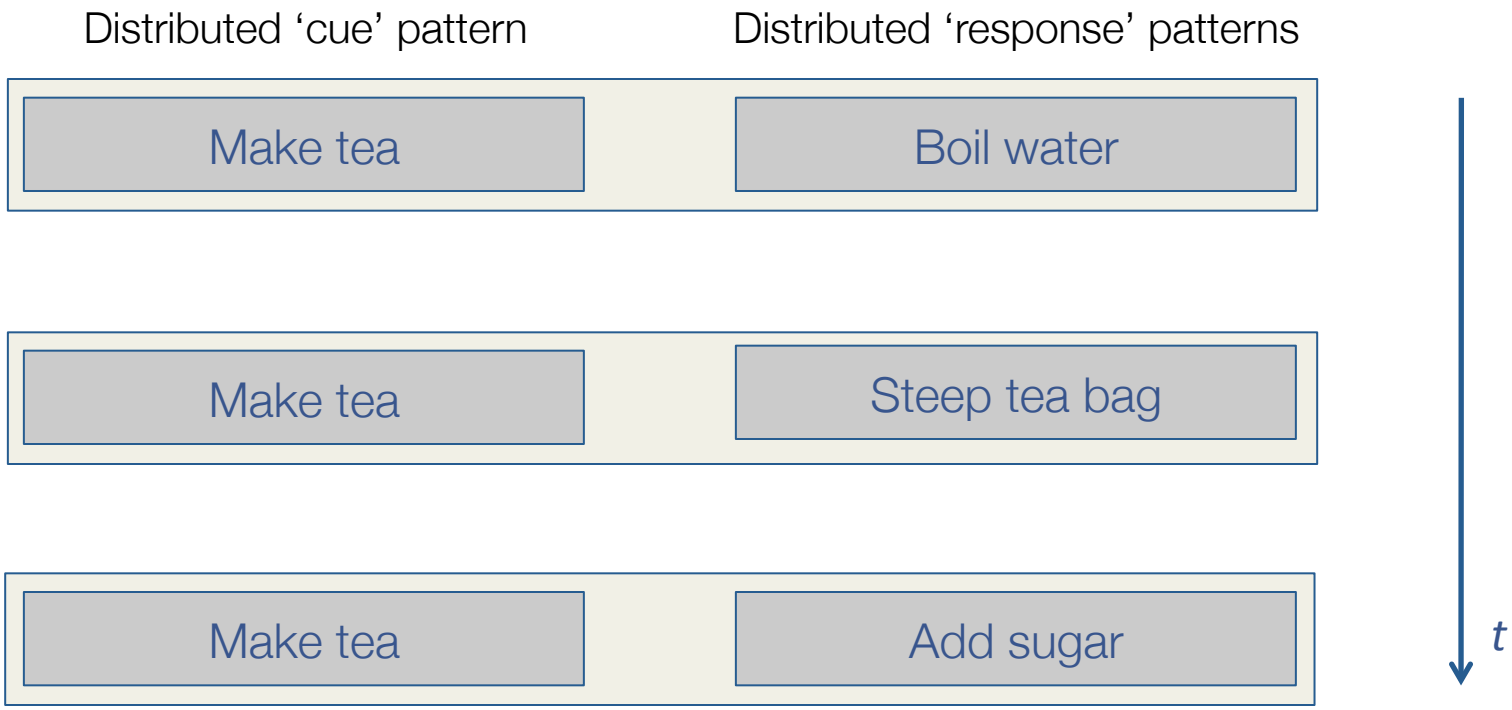
Distributed ‘cue’ pattern

Make tea

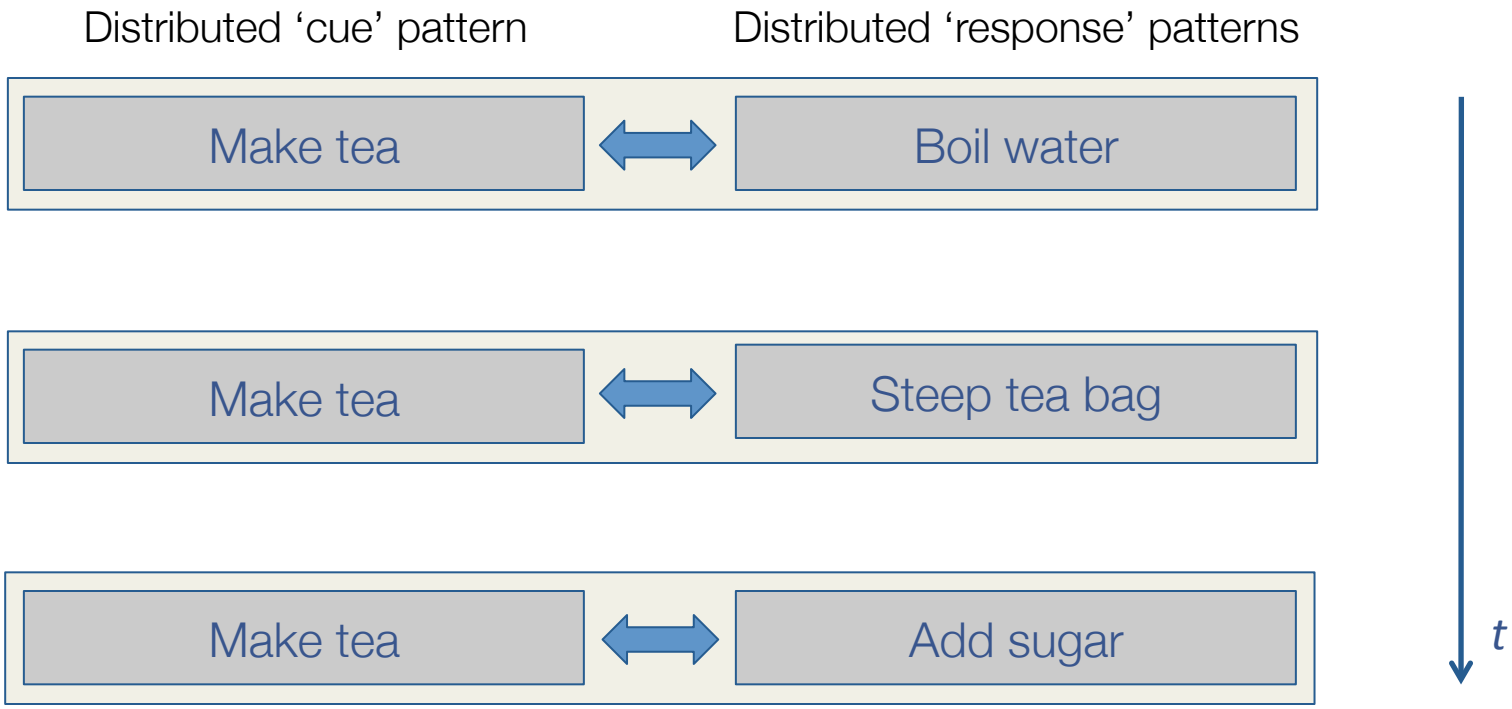
Instruction Sequence Memory



Instruction Sequence Memory

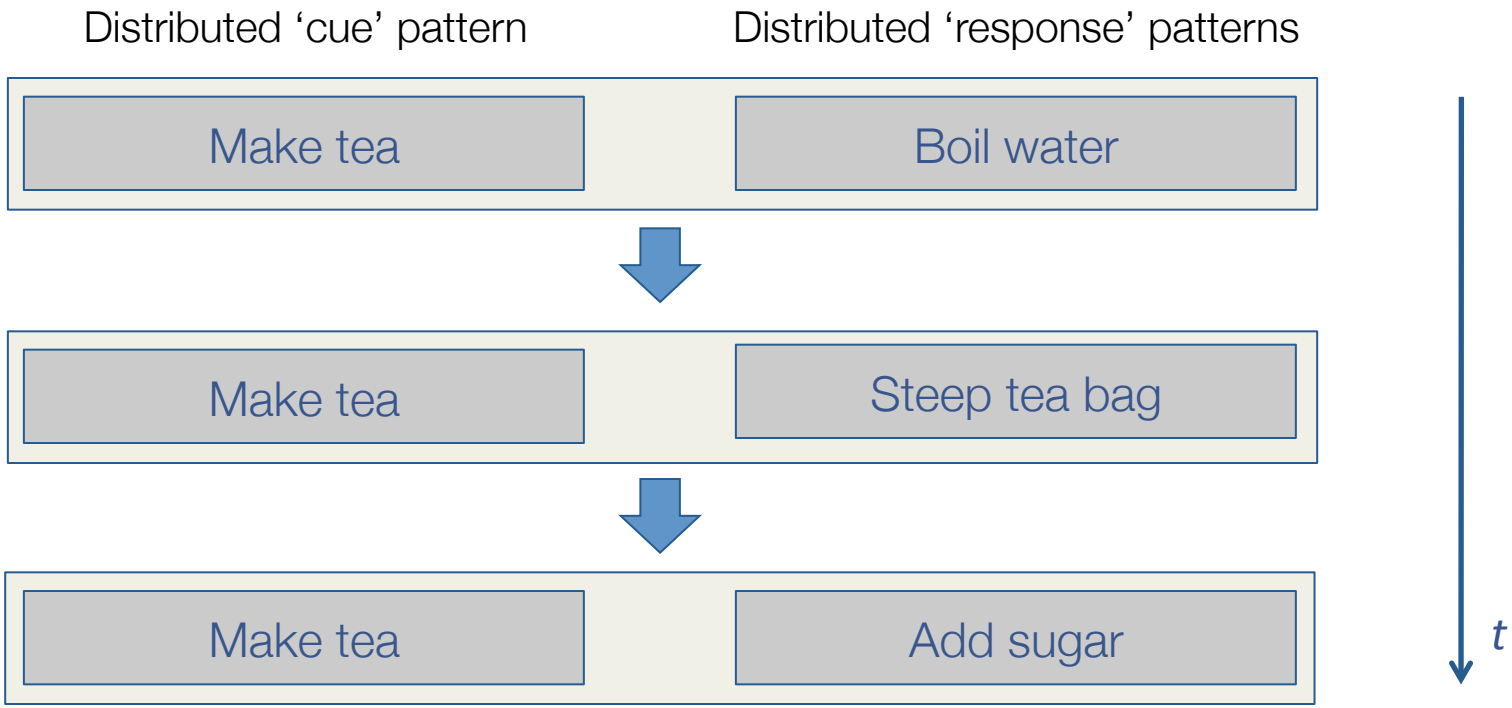


Instruction Sequence Memory



$$w_{ij}^{t} = (1 - k_{\mathrm{CTRL}})w_{ij}^{t-1} + \frac{1}{N}a_i^t a_j^t(1 - \delta_{ij})$$

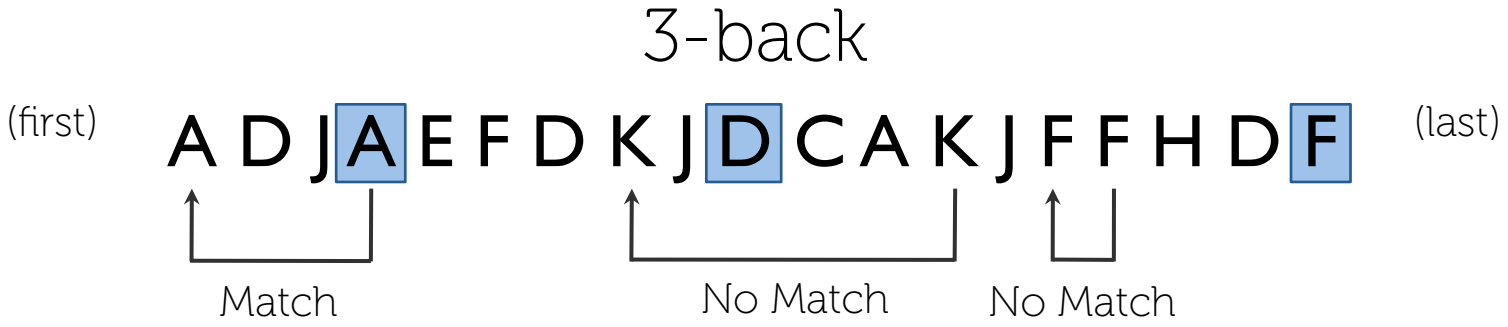
Instruction Sequence Memory



$$v_{ij}^{t} = (1 - k_{\mathrm{CTRL}})v_{ij}^{t-1} + \frac{1}{N}a_i^t a_j^{t-1}$$

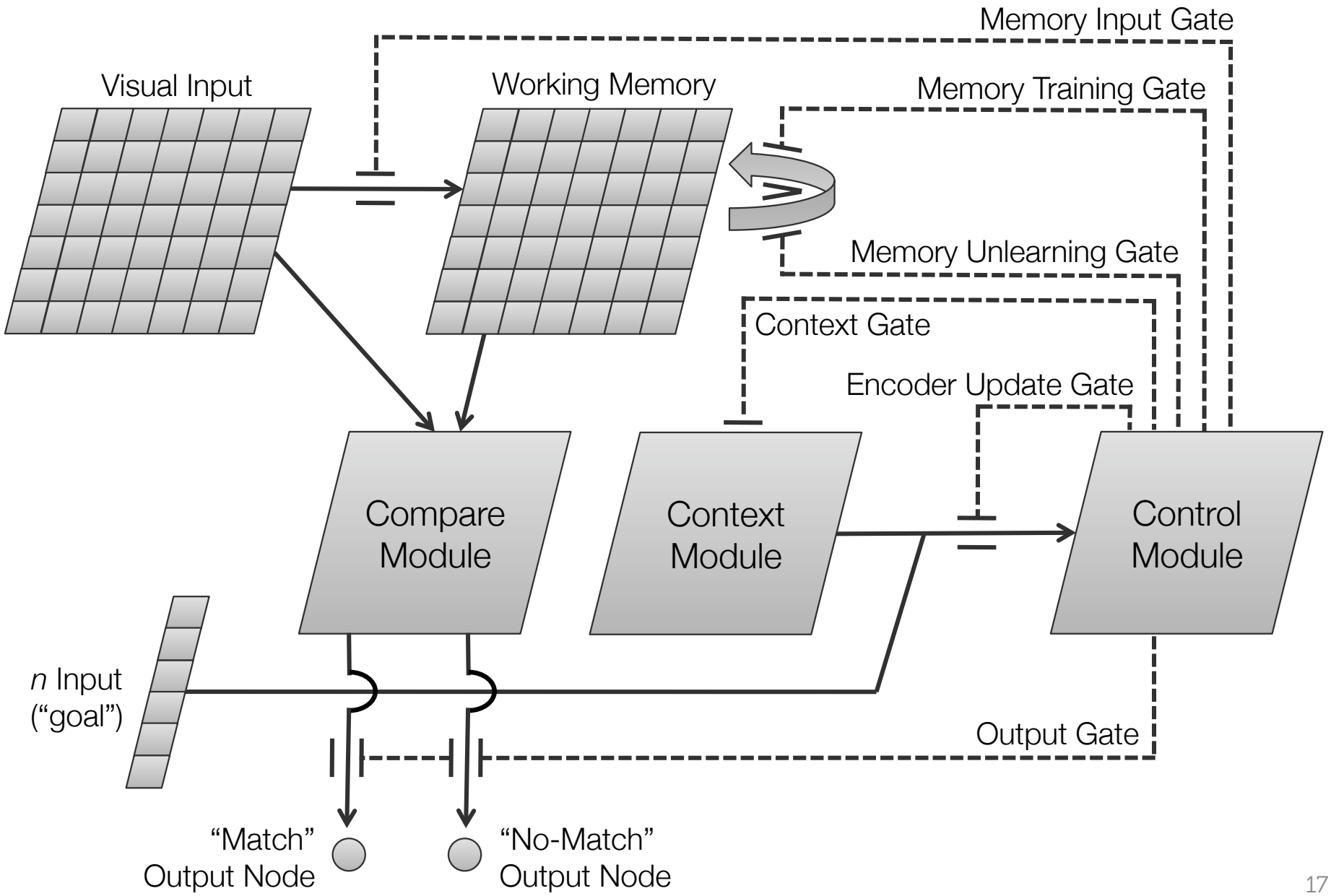
n-Back

- Given sequence of inputs…
  …does most recent input match input *n* steps ago?
- Must maintain sequence in WM; make judgments

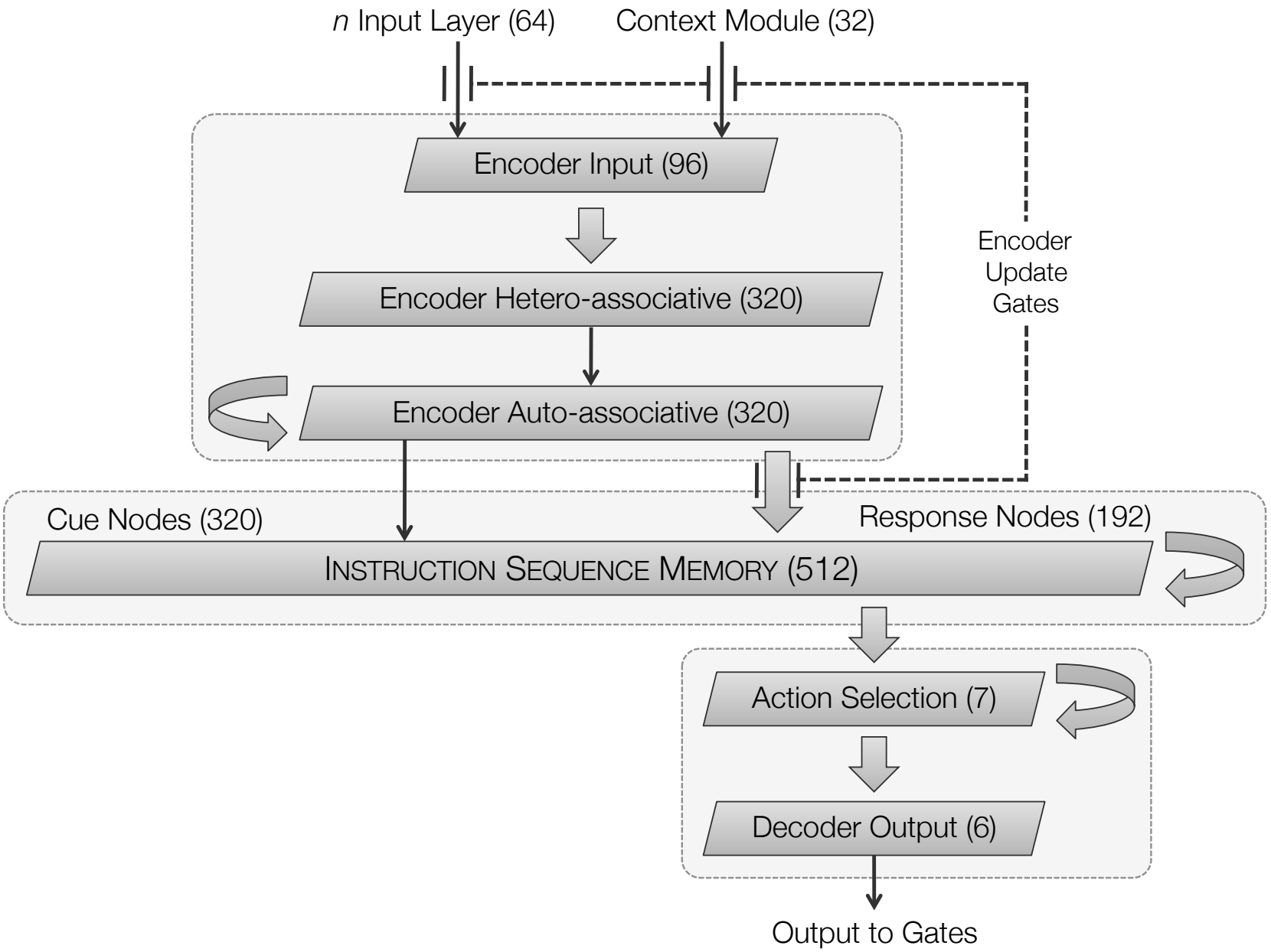


- GALIS model learns *n*=1,2,3,4,5
  - Learns all five without knowing which it will perform
  - Version determined by input patterns only

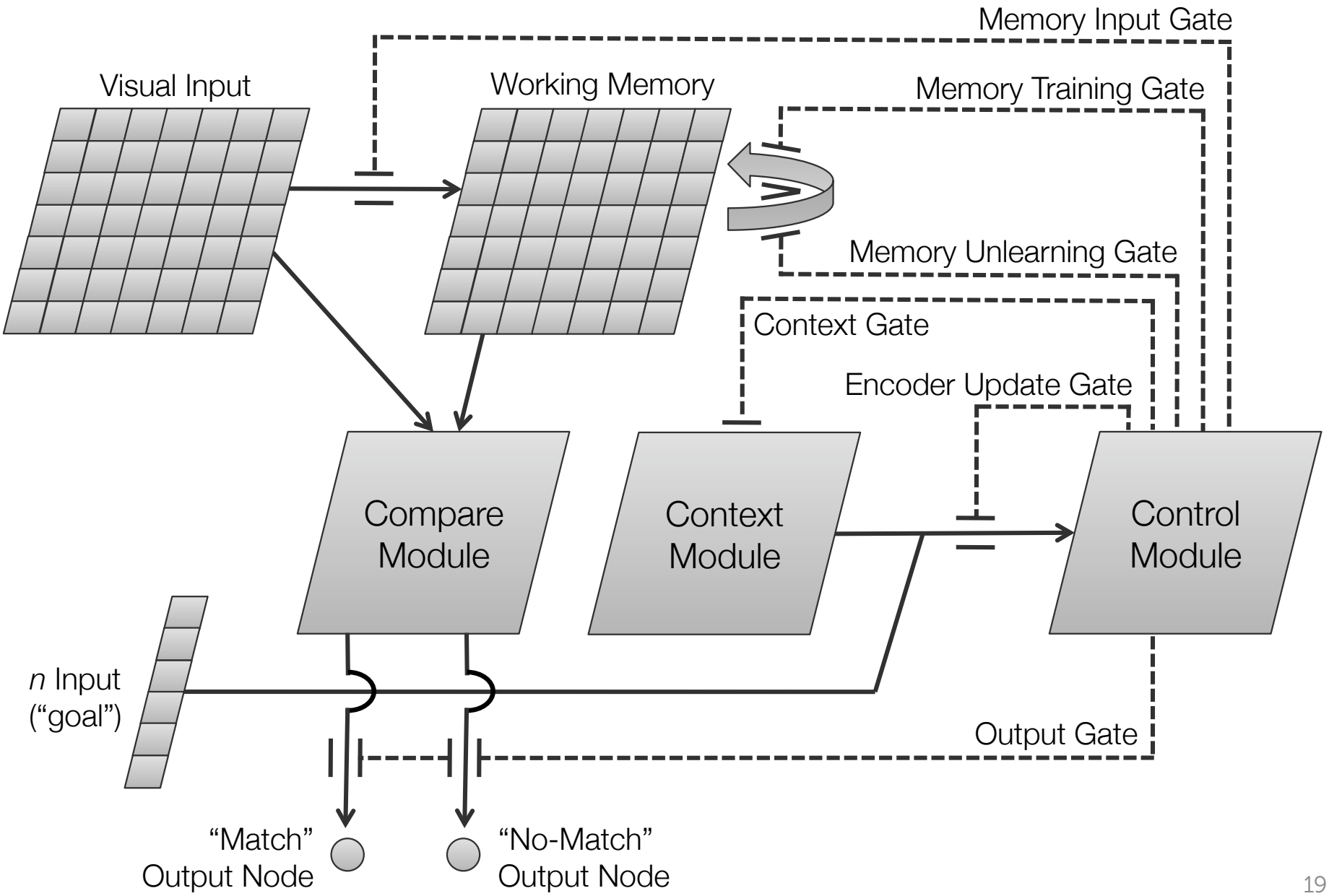
GALIS Architecture



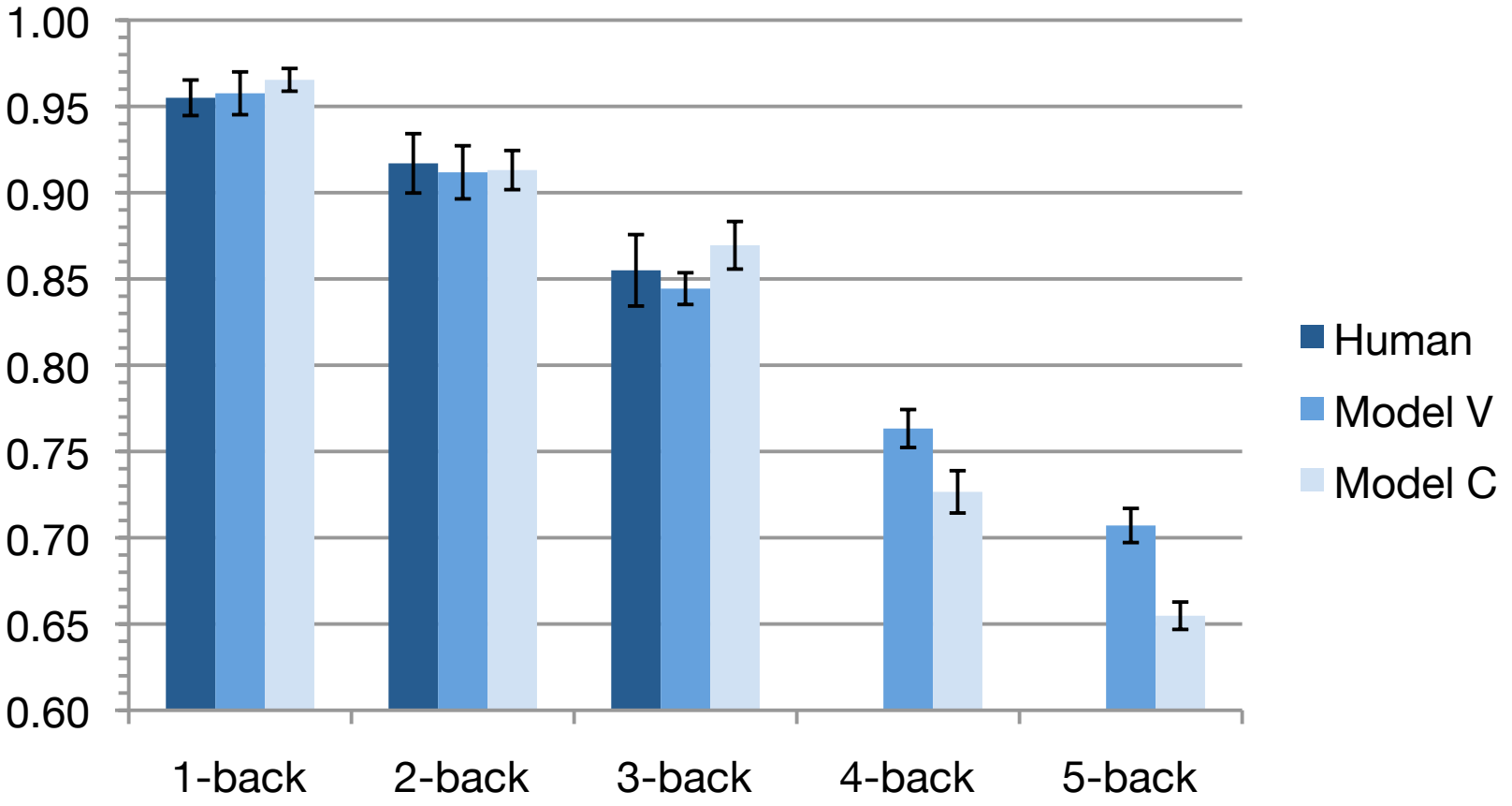
Controller


n Input Layer (64)
Context Module (32)
Encoder Input (96)
Encoder Hetero-associative (320)
Encoder Auto-associative (320)
Encoder Update Gates
Cue Nodes (320)
Response Nodes (192)
INSTRUCTION SEQUENCE MEMORY (512)
Action Selection (7)
Decoder Output (6)
Output to Gates

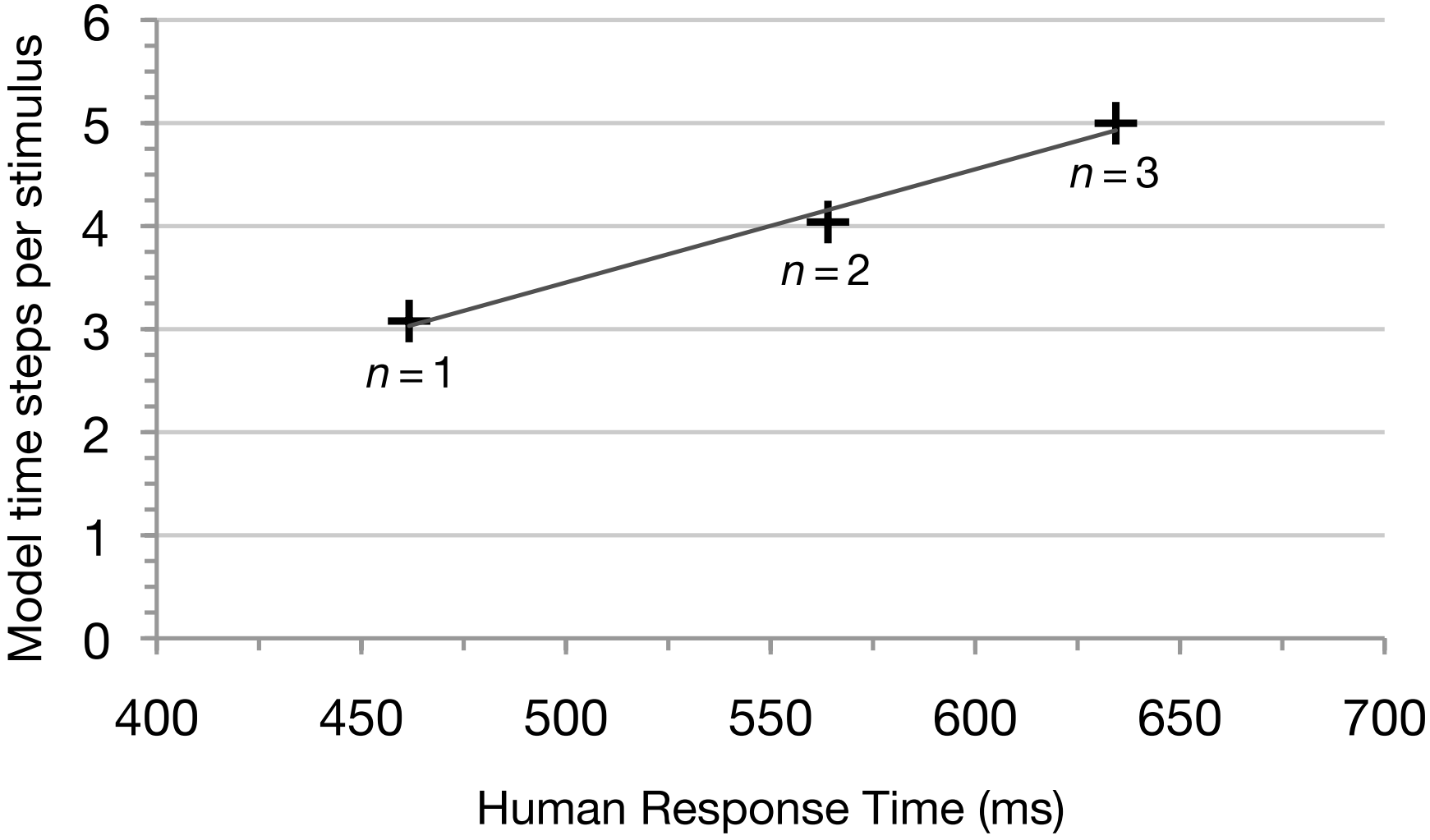
GALIS Architecture



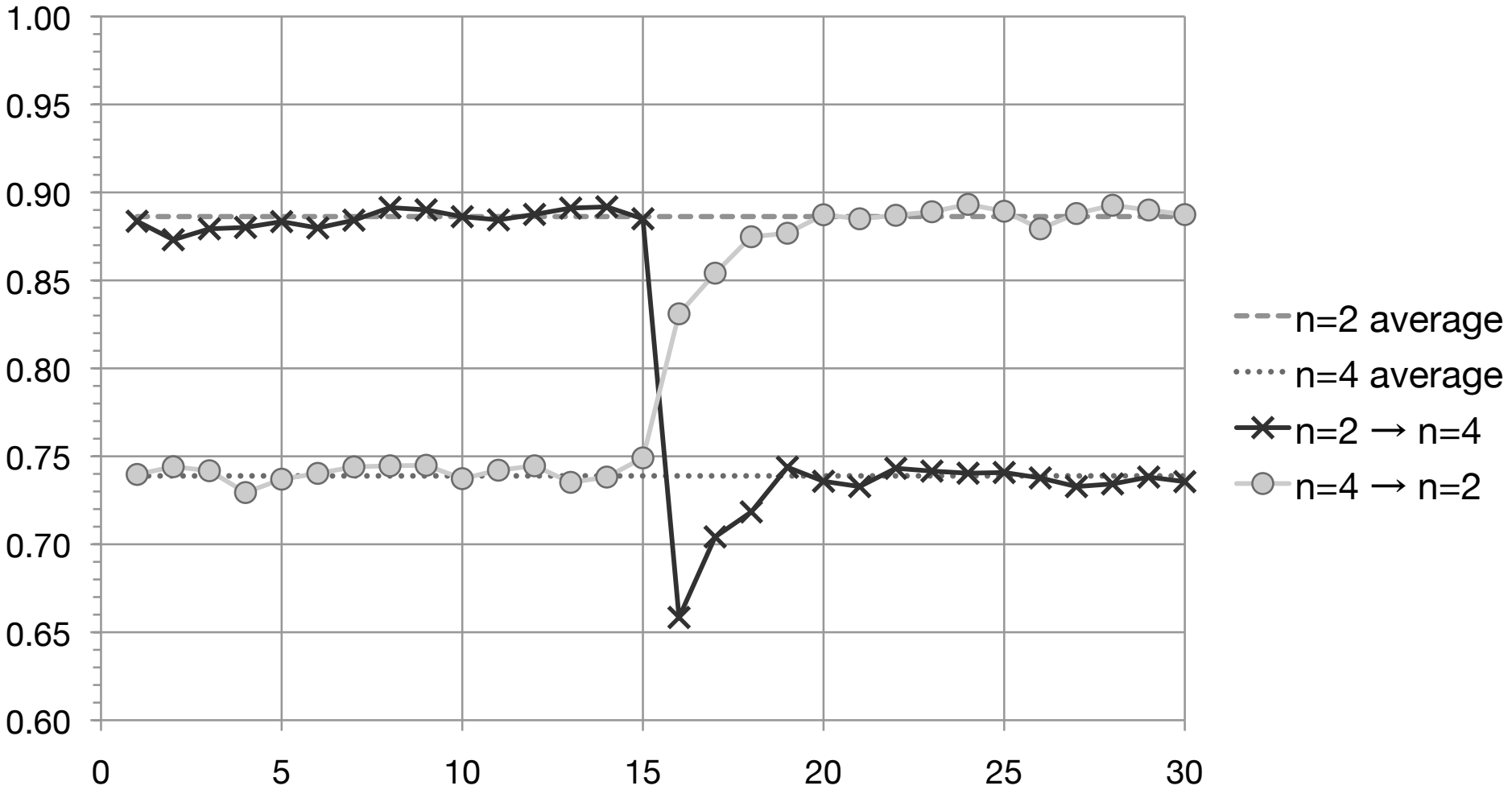
Human Comparison: Accuracy



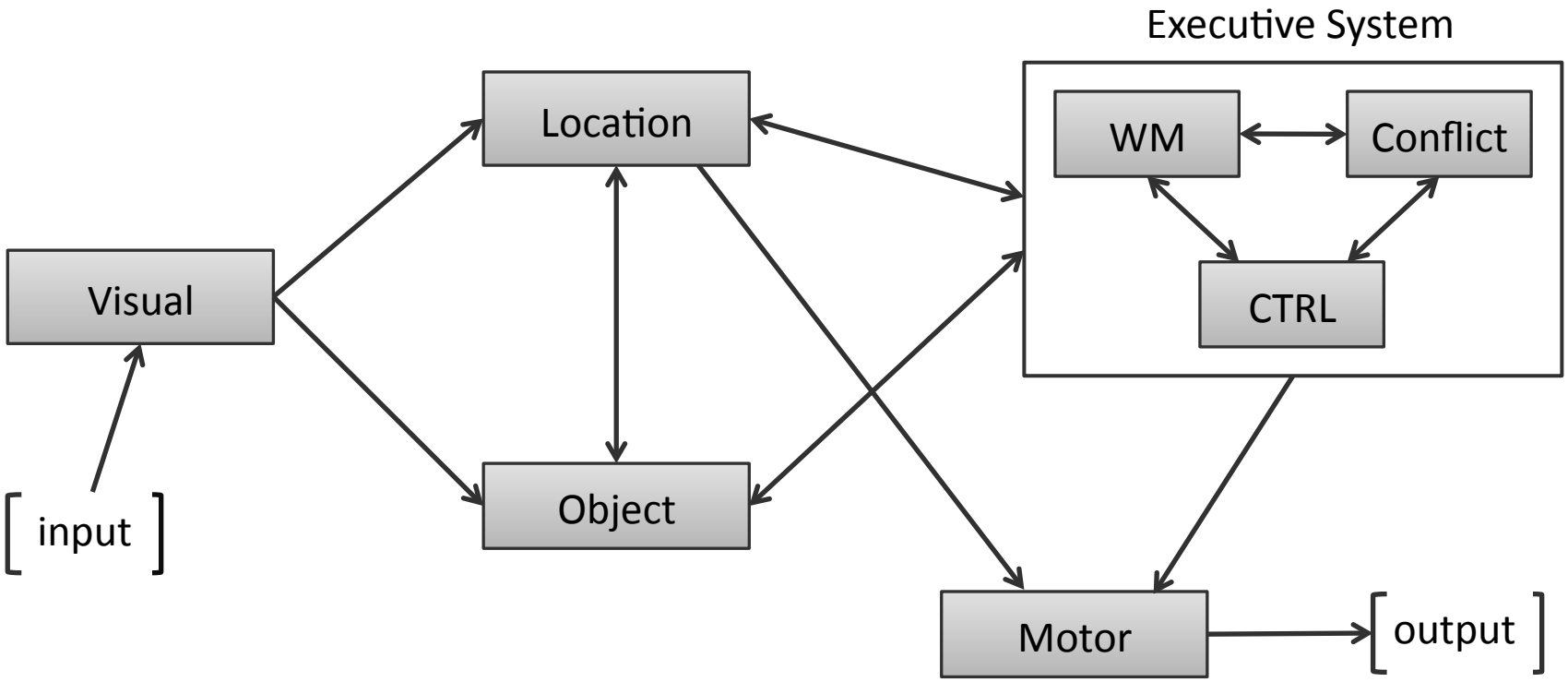
Human Comparison: Response Time



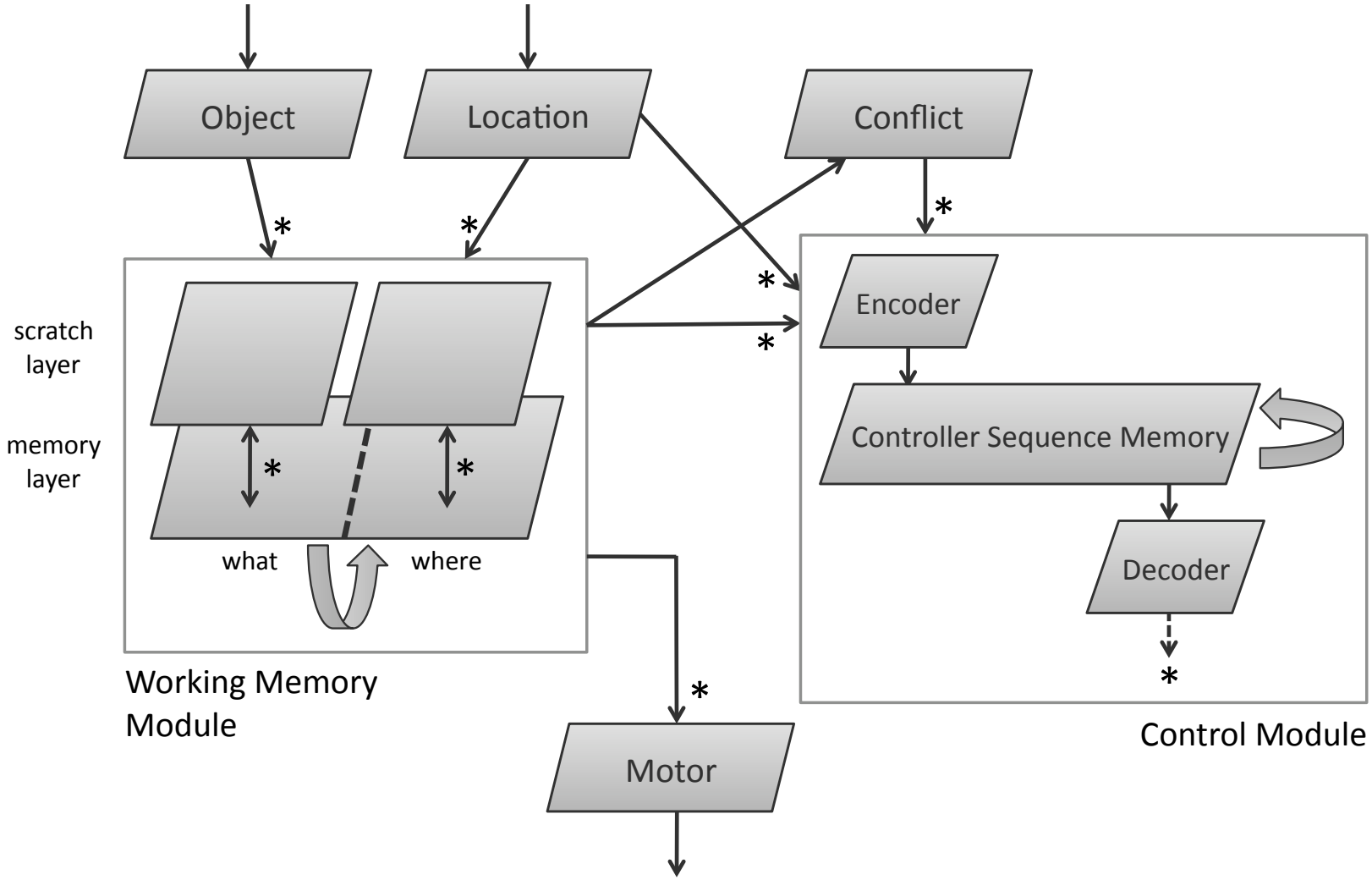
Changing *n*


1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0
5
10
15
20
25
30
n=2 average
n=4 average
n=2 → n=4
n=4 → n=2

Visuospatial Architecture



Visuospatial Architecture

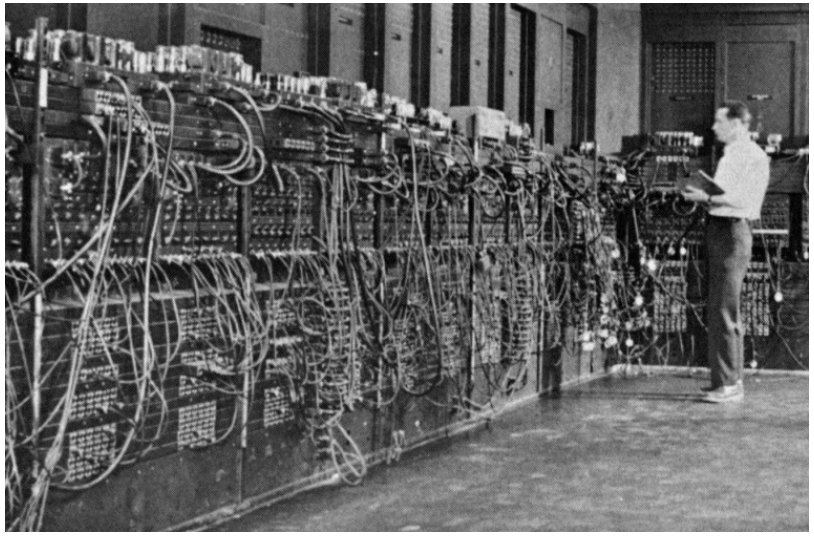


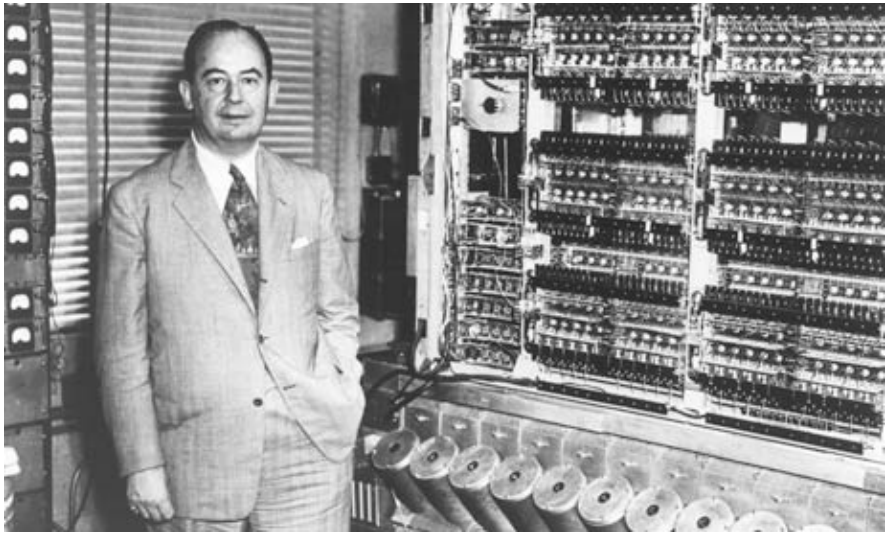
Symbolic / Sub-symbolic

- Attractor space is very high dimensional
  - Learning algorithms
  - Partial pattern matching
- Each attractor is a discrete symbol
- Gating also adds discreteness

Instruction vs. Construction

- Behavior based on memory contents
  not just architecture





- Can “program” a neural net
  - Now programs are hand-crafted by modeler
  - Store → improve → learn *ab initio*