Importance of Action History in Decision Making and Reinforcement Learning

Yongjia Wang

John E. Laird

Outline

- Motivations
- T-maze task
  - Soar-RL Model (using action history)
  - Compare with an ACT-R model

Motivations

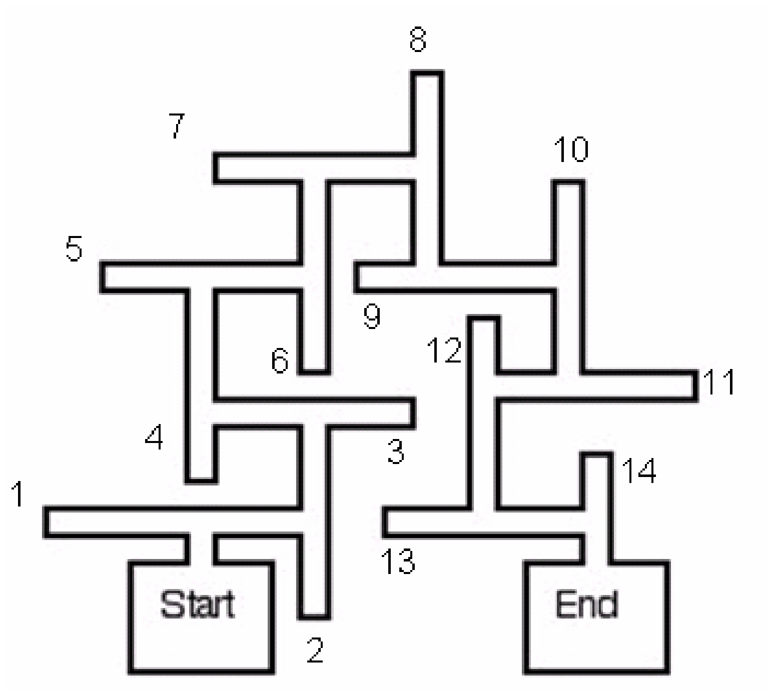
- Functional characterization
  - Use temporal sequence representation in the context of reinforcement learning
  - Explore Soar-RL
- Cognitive modeling
  - Testing hypothesis - compare simulation results with experimental data
  - Compare different models

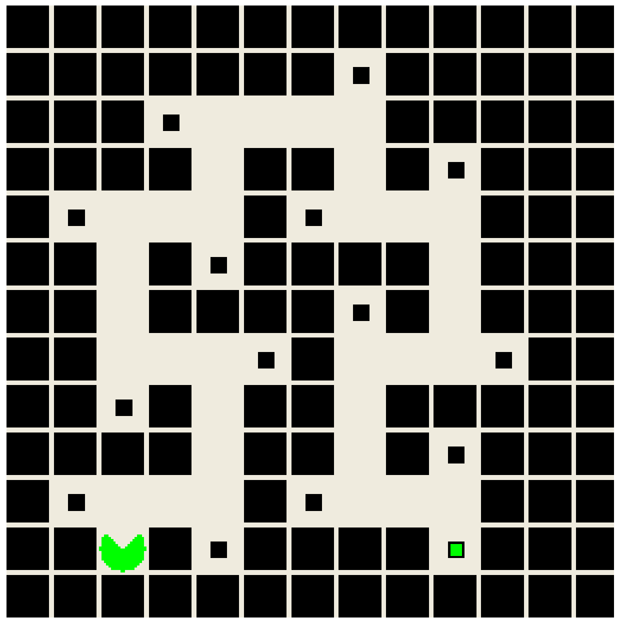
Outline

- Motivations
- T-maze task
  - Soar-RL Model (using action history)
  - Compare with an ACT-R model

T-maze Task

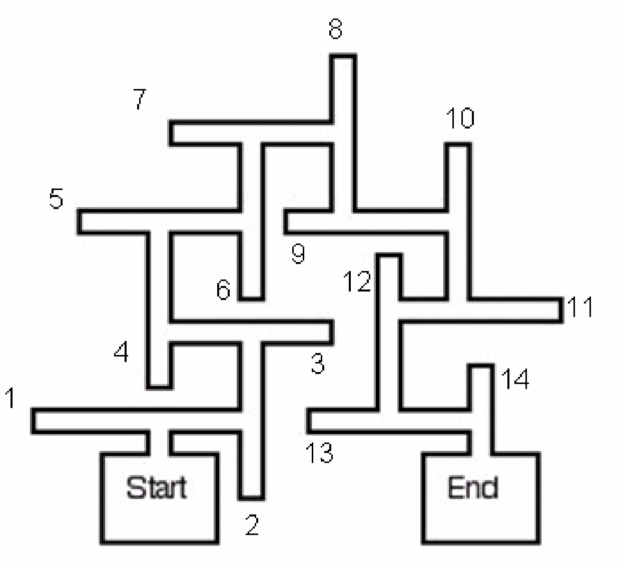
(Tolman & Honzik 1930)

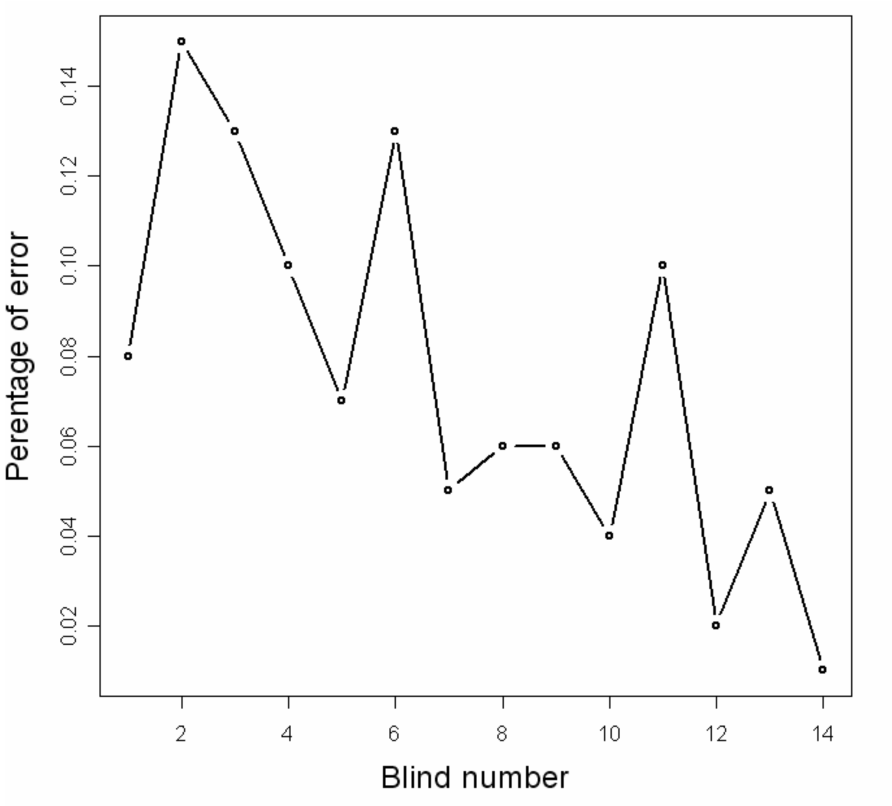




Experiment Result

(Tolman & Honzik 1930)

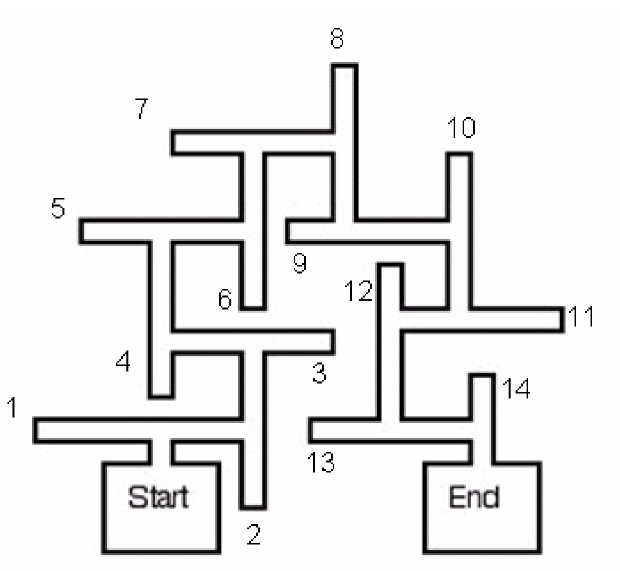


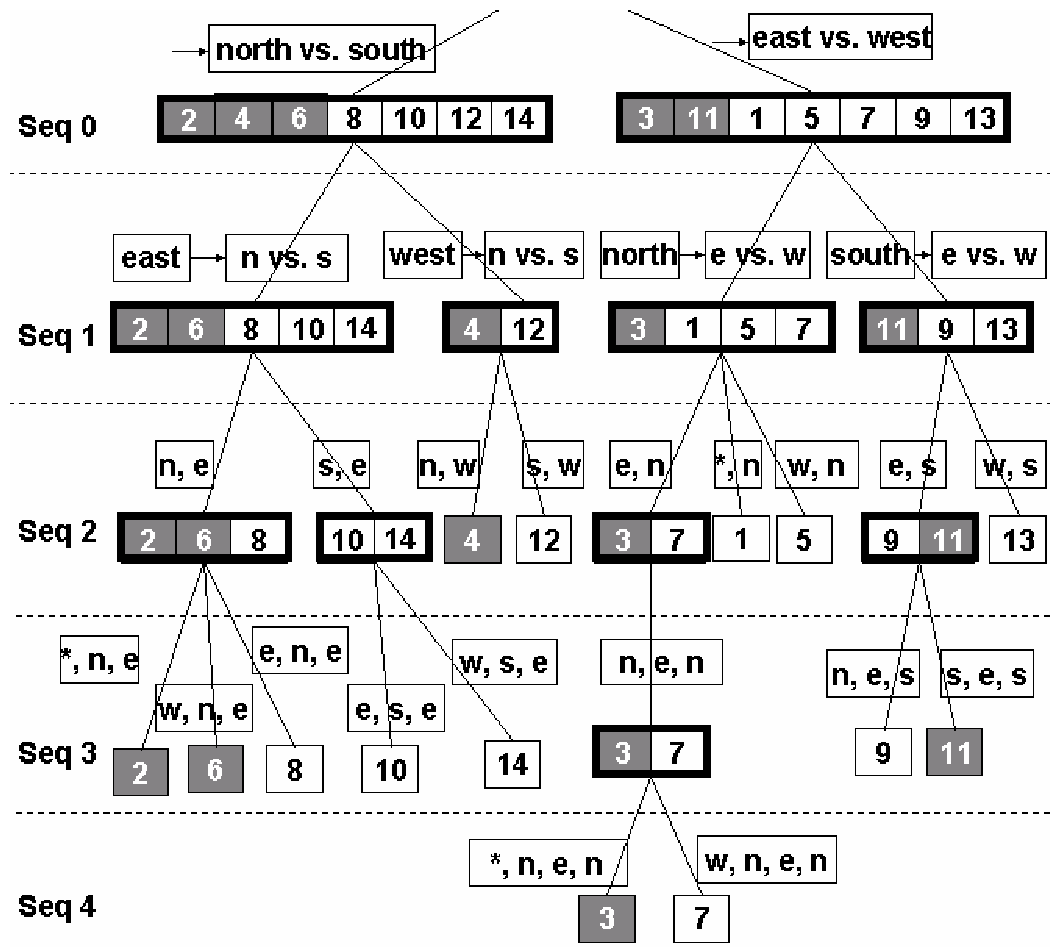


Outline

- Motivations
- T-maze task
  - Soar-RL Model (using action history)
  - Compare with an ACT-R model

Task Constraints





Soar-RL Representation

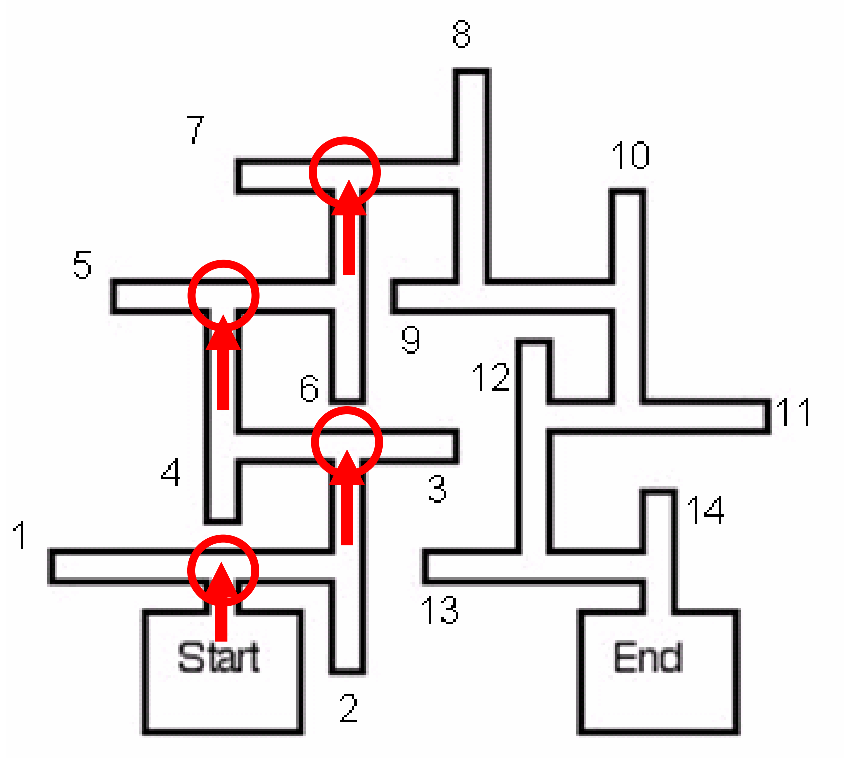
Rules at the level of seq 1

```
sp {Soar-RL-1
   (state <s> ^action-history-sequence <ahs>
              ^operator <o> +)
   (<ahs> ^previous-1 north)
   (<o> ^name move
        ^direction east)
-->
   (<s> ^operator <o> = 3.0)

}

sp {Soar-RL-2
   (state <s> ^action-history-sequence <ahs>
              ^operator <o> +)
   (<ahs> ^previous-1 north)
   (<o> ^name move
        ^direction west)
-->
   (<s> ^operator <o> = -3.0)

}
```

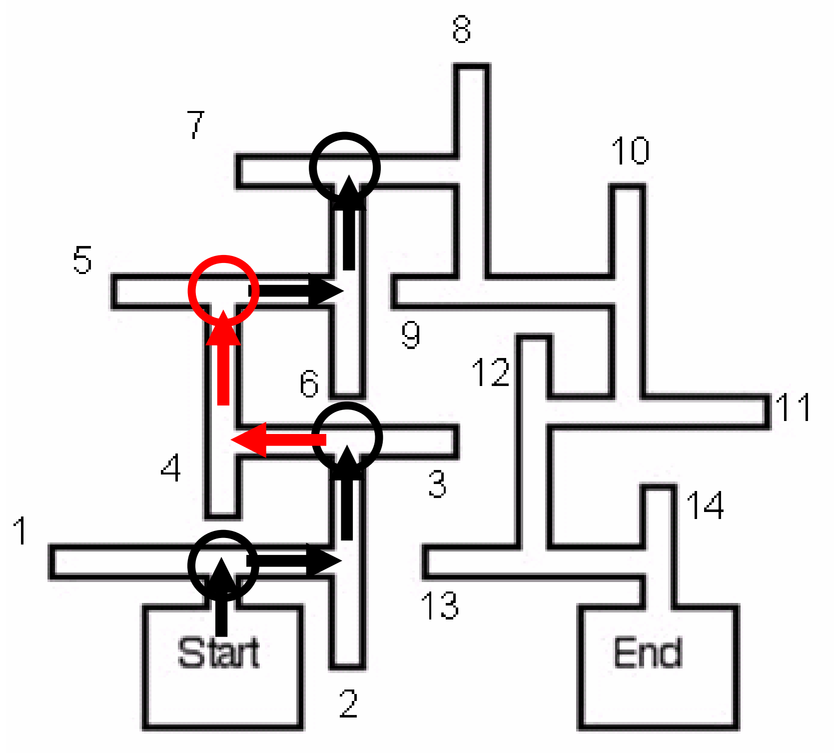


Soar-RL Representation

Rules at the level of seq 2

```
sp {Soar-RL-3
   (state <s> ^action-history-sequence <ahs>
              ^operator <o> +)
   (<ahs> ^previous-1 north
          ^previous-2 west)
   (<o> ^name move
        ^direction east)
-->
   (<s> ^operator <o> = 5.0)
}

sp {Soar-RL-4
   (state <s> ^action-history-sequence <ahs>
              ^operator <o> +)
   (<ahs> ^previous-1 north
          ^previous-2 west)
   (<o> ^name move
        ^direction west)
-->
   (<s> ^operator <o> = -5.0)
}
```

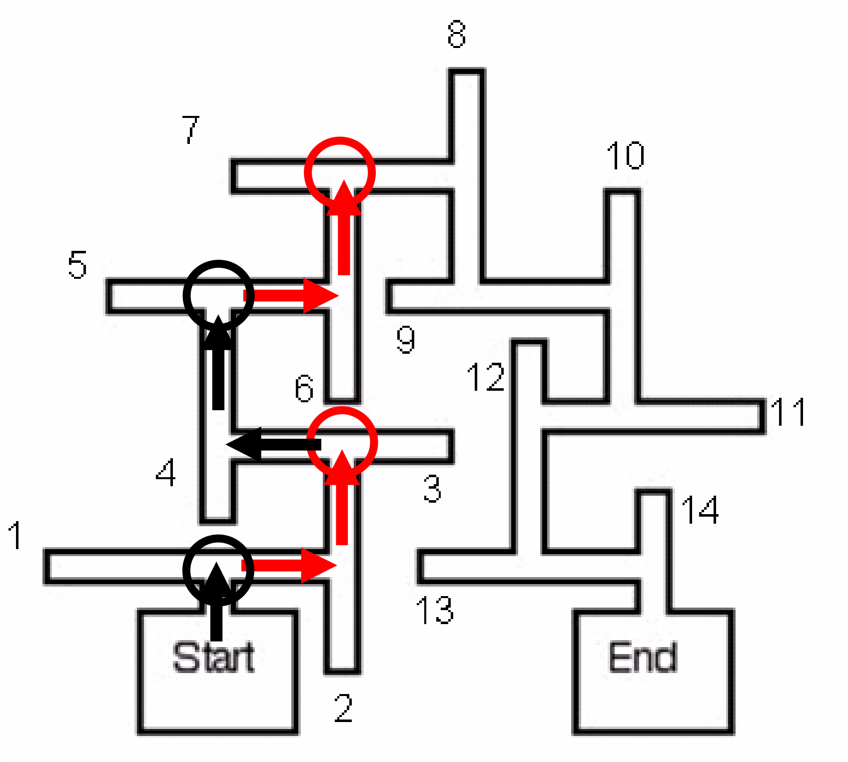


Soar-RL Representation

Rules at the level of seq 2

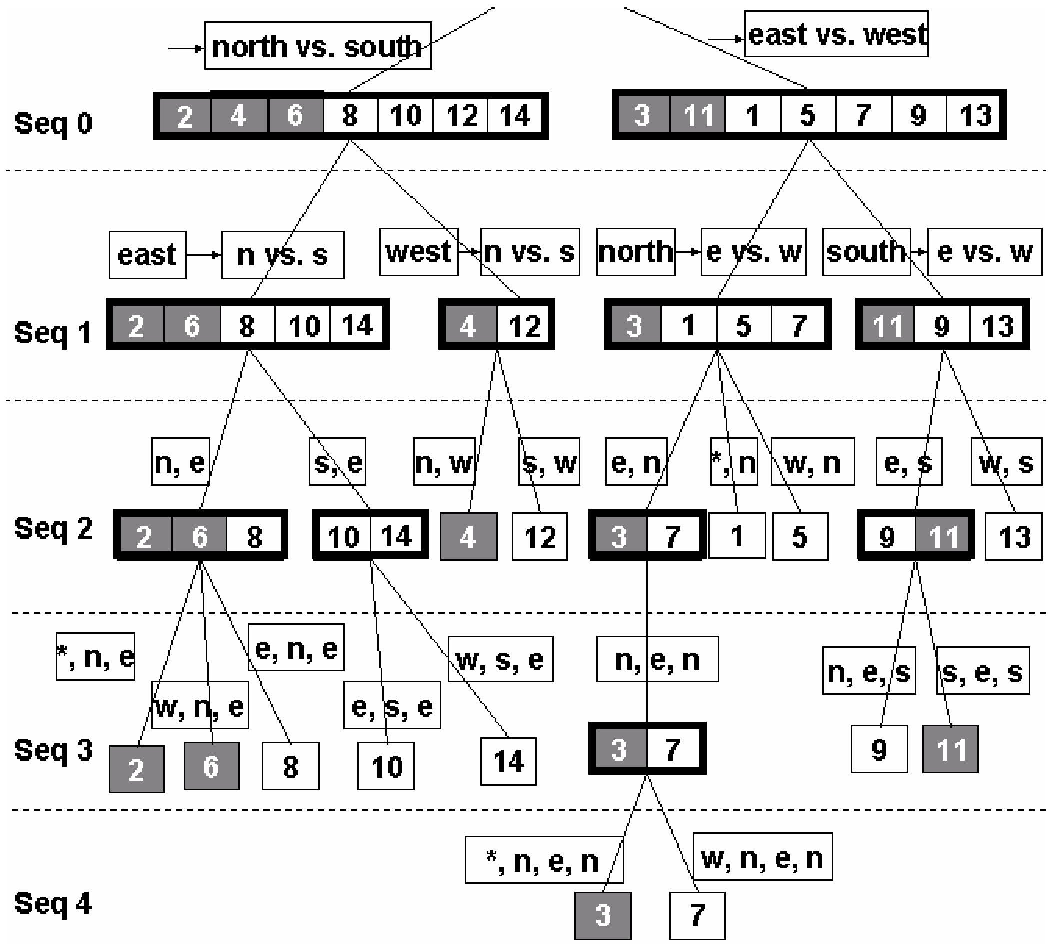
```
sp {Soar-RL-5
   (state <s> ^action-history-sequence <ahs>
              ^operator <o> +)
   (<ahs> ^previous-1 north
          ^previous-2 east)
   (<o> ^name move
        ^direction east)
-->
   (<s> ^operator <o> = 1.0)
}

sp {Soar-RL-6
   (state <s> ^action-history-sequence <ahs>
              ^operator <o> +)
   (<ahs> ^previous-1 north
          ^previous-2 east)
   (<o> ^name move
        ^direction west)
-->
   (<s> ^operator <o> = -0.5)
}
```



The final utility value for a state-action pair is the sum of matched rules from all specificity levels

General-to-Specific Reinforcement Learning



Q Value and Action Probability

$$P_i = \frac{e^{Q(s,O_i)/Temperature}}{\sum_i e^{Q(s,O_i)/Temperature}}$$

Pi: probability of choosing operator i

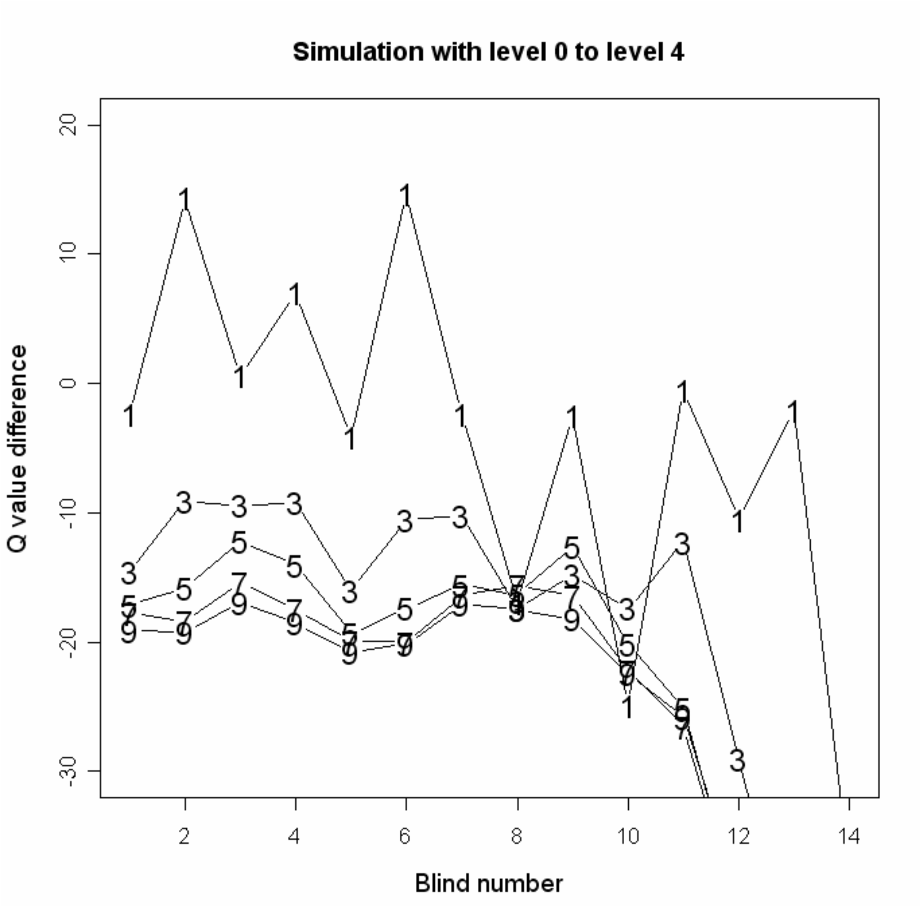
Q Value and Action Probability

$$P_i = \frac{e^{Q(s,O_i)/Temperature}}{\sum_i e^{Q(s,O_i)/Temperature}}$$

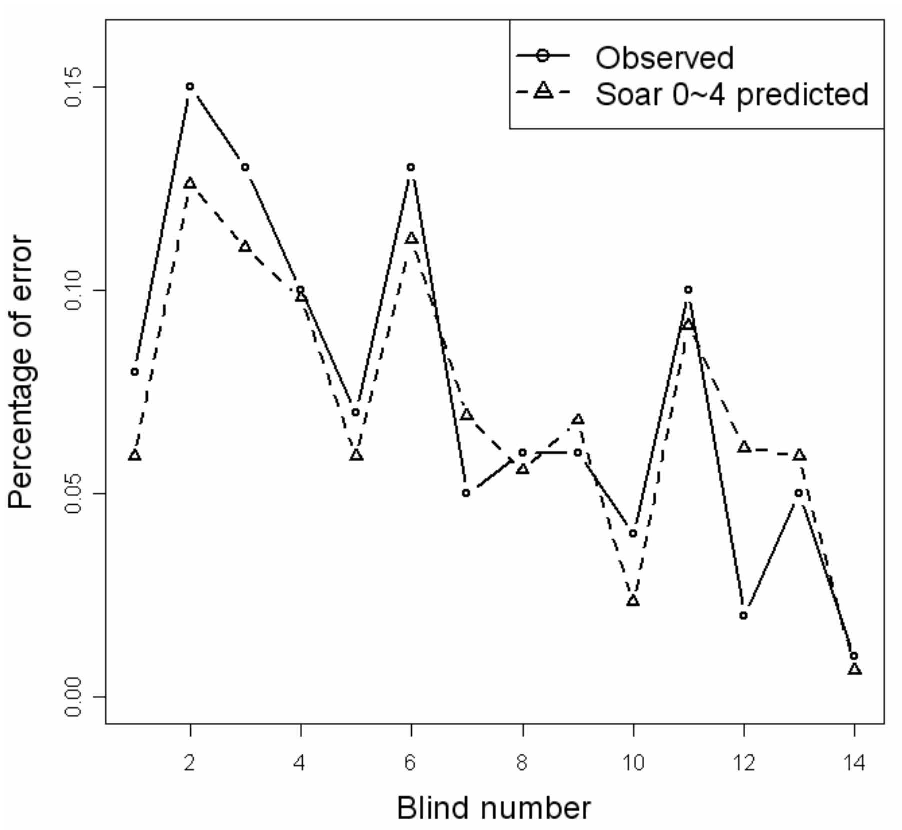
Pi: probability of choosing operator i

For 2 choices: $$P_1 = \frac{e^{Q(s,O_1)-Q(s,O_2)/Temperature}}{1 + e^{Q(s,O_1)-Q(s,O_2)/Temperature}}$$

Simulation Results



Changing of Q value difference
when learning with all levels of rules

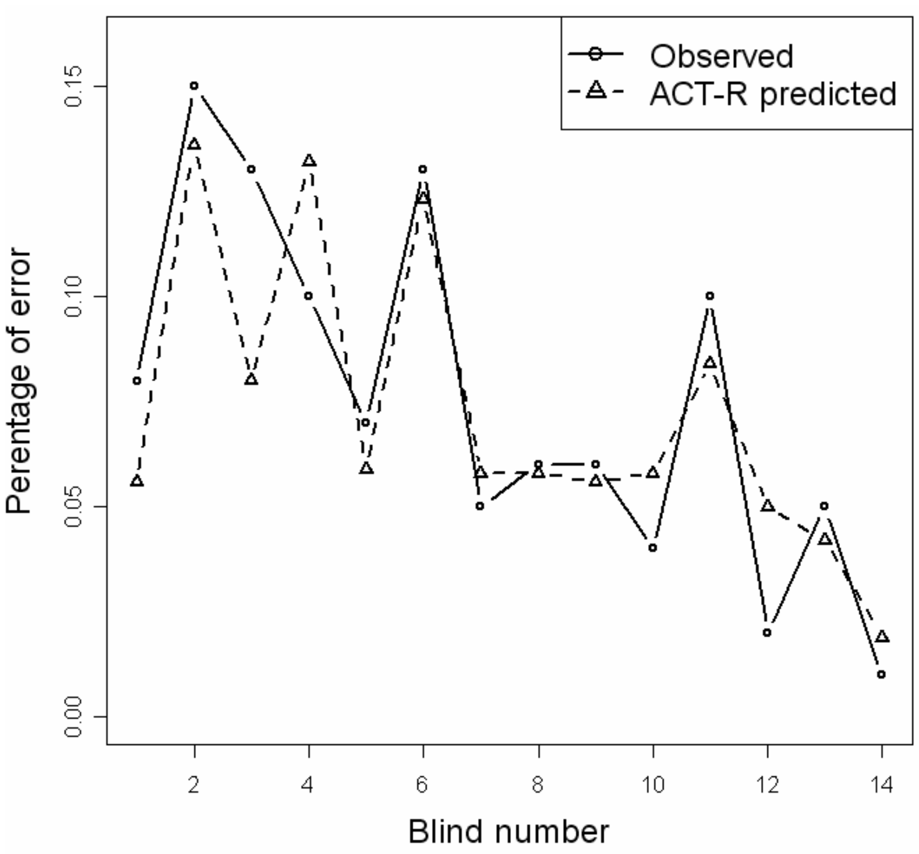


Percentage of error after 17 trials

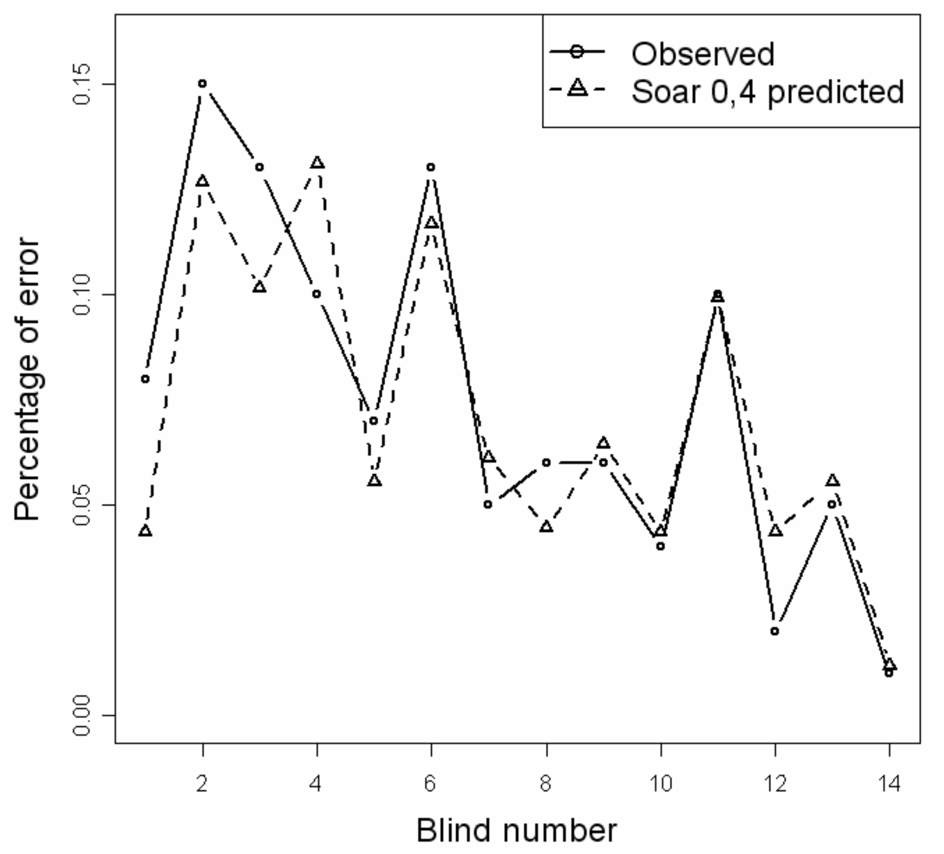
Outline

- Motivations
- T-maze task
  - Soar-RL Model (using action history)
  - Compare with an ACT-R model

Compare with an ACT-R Model



An ACT-R model with general rules and specific rules
(FU & Anderson 2006)



The equivalent Soar model with level 0 and level 4 rules – no intermediate levels

Correlation Matrix

	Observed	Soar0~4	Soar 0,4	ACT-R
Observed	-	0.91	0.89	0.86
Soar 0~4	-	-	-	0.86
Soar 0,4	-	-	-	0.95
ACT-R	-	-	-	-

Comparison

- ACT-R model
  - Can have prediction with correlation 0.95 by adjusting learning parameters (unpublished data)
  - Still cannot explain why error rate at blind 4 is much lower than at number 3 (which can be explained by having more intermediate levels)
  - Can have potentially more accurate predictions with the action history representation as in Soar
- Soar model
  - The exponential discount in Soar-RL results in poor match for later blinds (12,13,14). ACT-R uses a linear discount formula that better matches the data.
  - Can explain earlier blinds well, especially number 3 and number 4

Comparison

	Model Level	Architectural Level
ACT-R Model	Assume unique choice point labels	Single rule firing and independent updating, learn one rule per decision
Soar Model	Sequence of action history as state representation at different specificity levels	Parallel RL rule firing and updating - learn all levels simultaneously

Conclusions

- Soar-RL reinforcement learning mechanism naturally models general-to-specific learning
- The results suggest that rats use sequence of action history to discriminate situations

Nuggets and Coal

- Nuggets
  - Explored some applications of Soar-RL
  - Soar-RL model with action history sequence matches rats data well
- Coal
  - Still mismatch some data points
  - Confirmation of hypothesis is not very strong