Emotion-Driven Learning in a Complex Environment

Bob Marinier
University of Michigan
28th Soar Workshop

Introduction

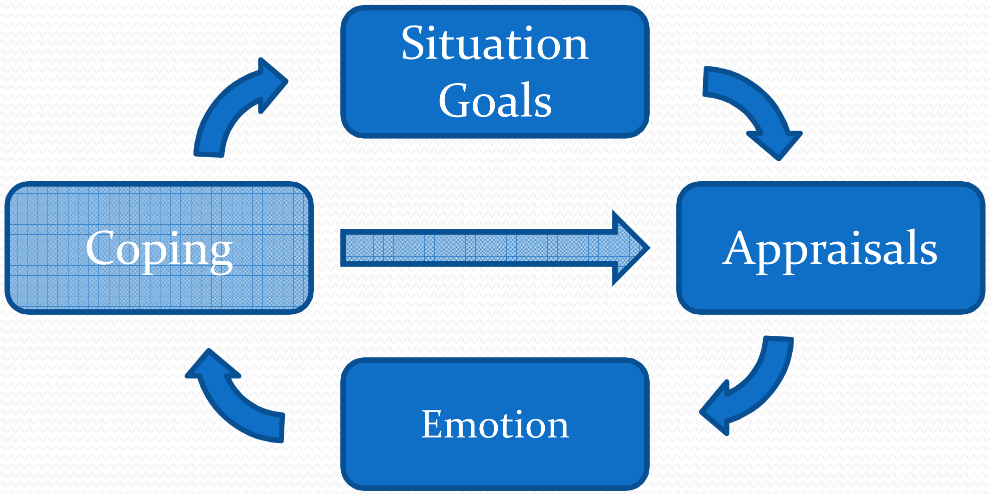
- Exploring
  - How to integrate emotion and cognition?
    - Emotion provides data, cognition provides process
  - Functional benefits of emotion?
    - Use emotion to drive reinforcement learning
- Goals
  - Does the system scale to a complex (continuous time/space) environment?
  - Does each appraisal influence behavior and learning?

Outline

- Background
- Clean House Domain
- Evaluation
- Conclusion

Appraisal Theories of Emotion

- A situation is evaluated along a number of *appraisal dimensions*, many of which relate the situation to current goals
  - Novelty, goal relevance, goal conduciveness, expectedness, causal agency, etc.
- Result of *appraisals* determines *emotion*
- Emotion can then be *coped* with (via internal or external actions)



Appraisals to Emotions (Scherer 2001)

	Joy	Fear	Anger
Suddenness	High/medium	High	High
Unpredictability	High	High	High
Intrinsic pleasantness		Low	
Goal/need relevance	High	High	High
Cause: agent		Other/nature	Other
Cause: motive	Chance/intentional		Intentional
Outcome probability	Very high	High	Very high
Discrepancy from expectation		High	High
Conduciveness	Very high	Low	Low
Control			High
Power		Very low	High

- Why these dimensions?
- What is the functional purpose?

Newell's Abstract Functional Operations a.k.a. PEACTIDM (Newell 1990)

- Allen Newell defined a set of computational Abstract Functional Operations that are *necessary and sufficient* for immediate behavior in humans and complete agents

Perceive	Obtain raw perception
Encode	Create domain-independent representation
Attend	Choose stimulus to process
Comprehend	Generate structures that relate stimulus to tasks and can be used to inform behavior
Task	Perform task maintenance
Intend	Choose an action, create prediction
Decode	Decompose action into motor commands
Motor	Execute motor commands

Newell's Abstract Functional Operations a.k.a. PEACTIDM (Newell 1990)

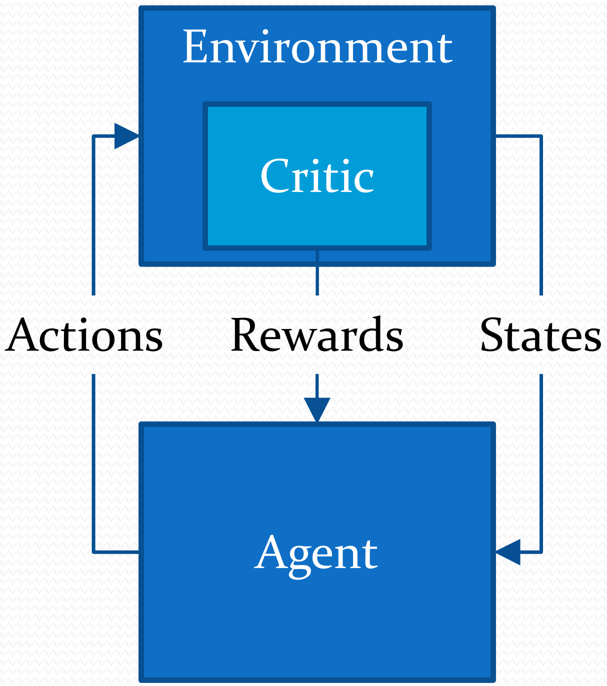
- …but how these actually work was not clear.

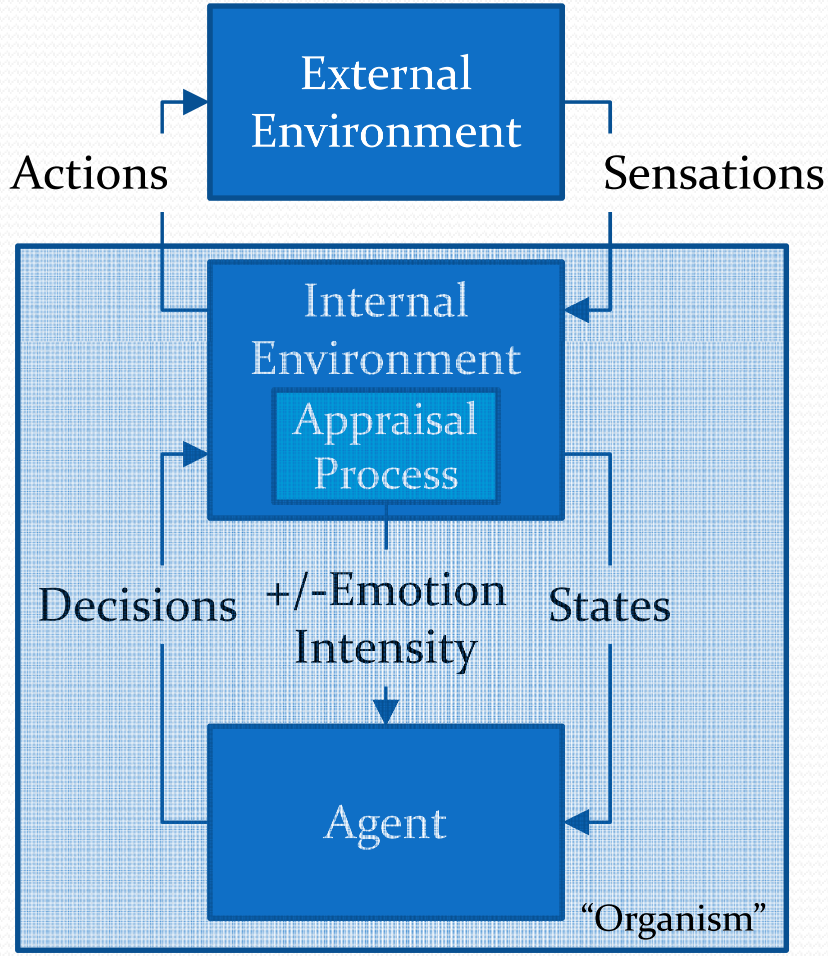
Perceive	What information is generated?
Encode	What information is generated?
Attend	What information is required?
Comprehend	What information is required and generated?
Task	What information is required?
Intend	What information is required?

PEACTIDM and Appraisal (Marinier & Laird 2006)

	Generated By	Required By
Suddenness	Perceive	Attend
Unpredictability	Encode	
Intrinsic pleasantness		
Goal relevance		
Causal agent	Comprehend	Comprehend, Task, Intend
Causal motive		
Outcome probability		
Discrepancy from expectation		
Goal/need conduciveness		
Control		
Power		

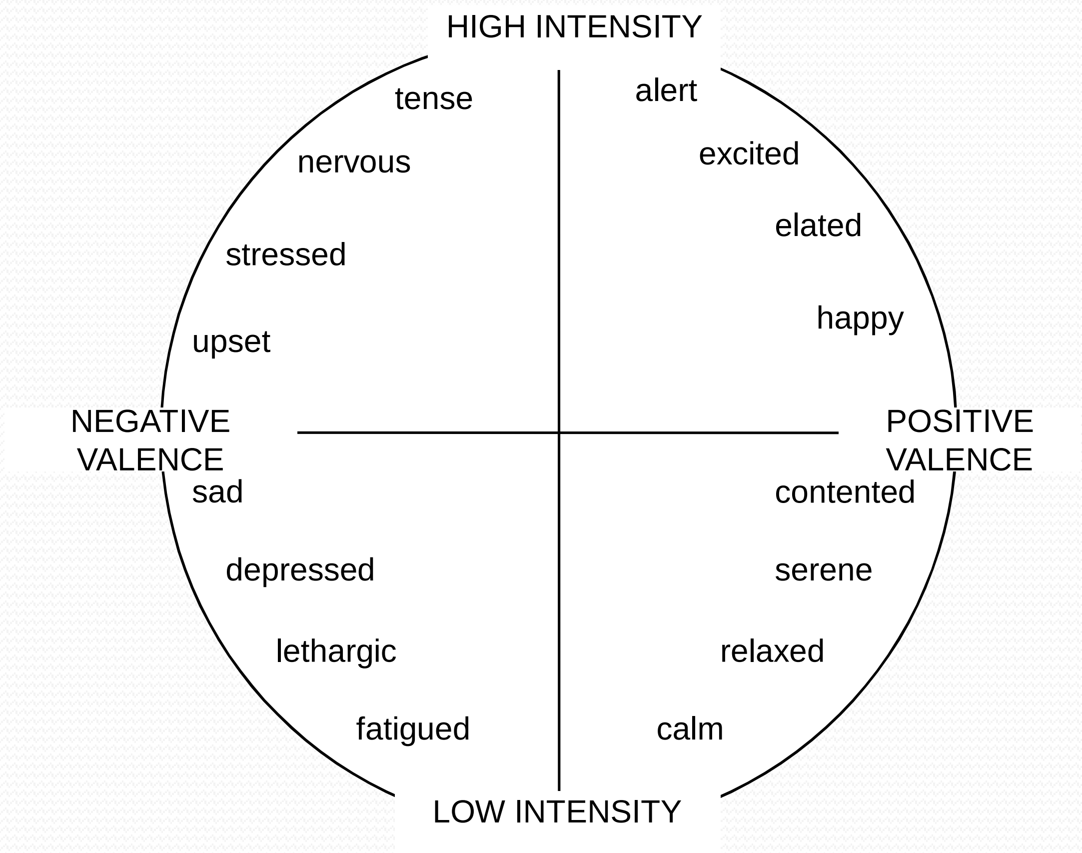
Intrinsically Motivated Reinforcement Learning (Sutton & Barto 1998; Singh et al. 2004)





Calculating Reward

- Emotions can be described in terms of intensity and valence, as in a circumplex model:



Adapted from Feldman Barrett & Russell (1998)

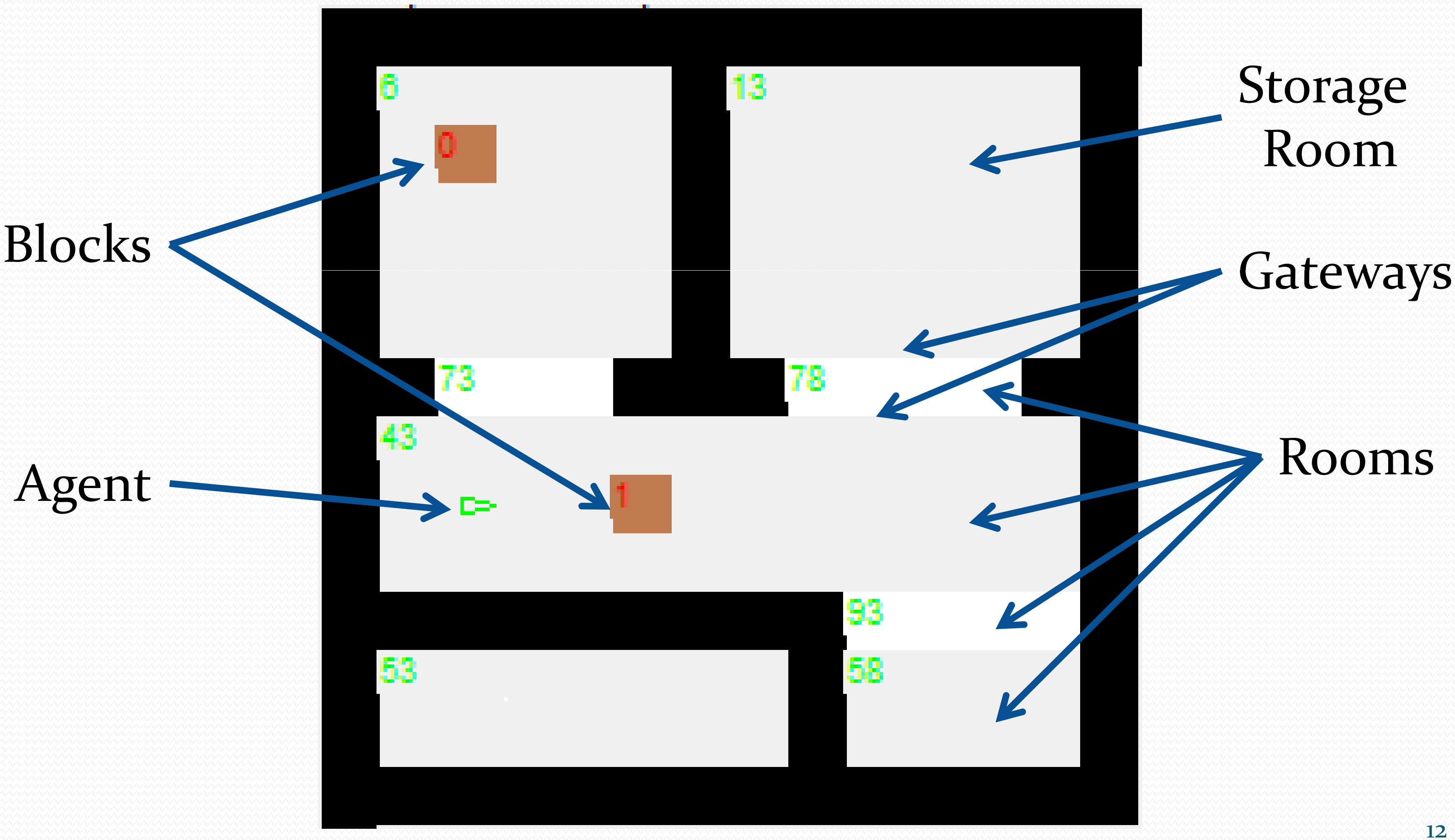
Calculating Reward

- Reward = Intensity * Valence

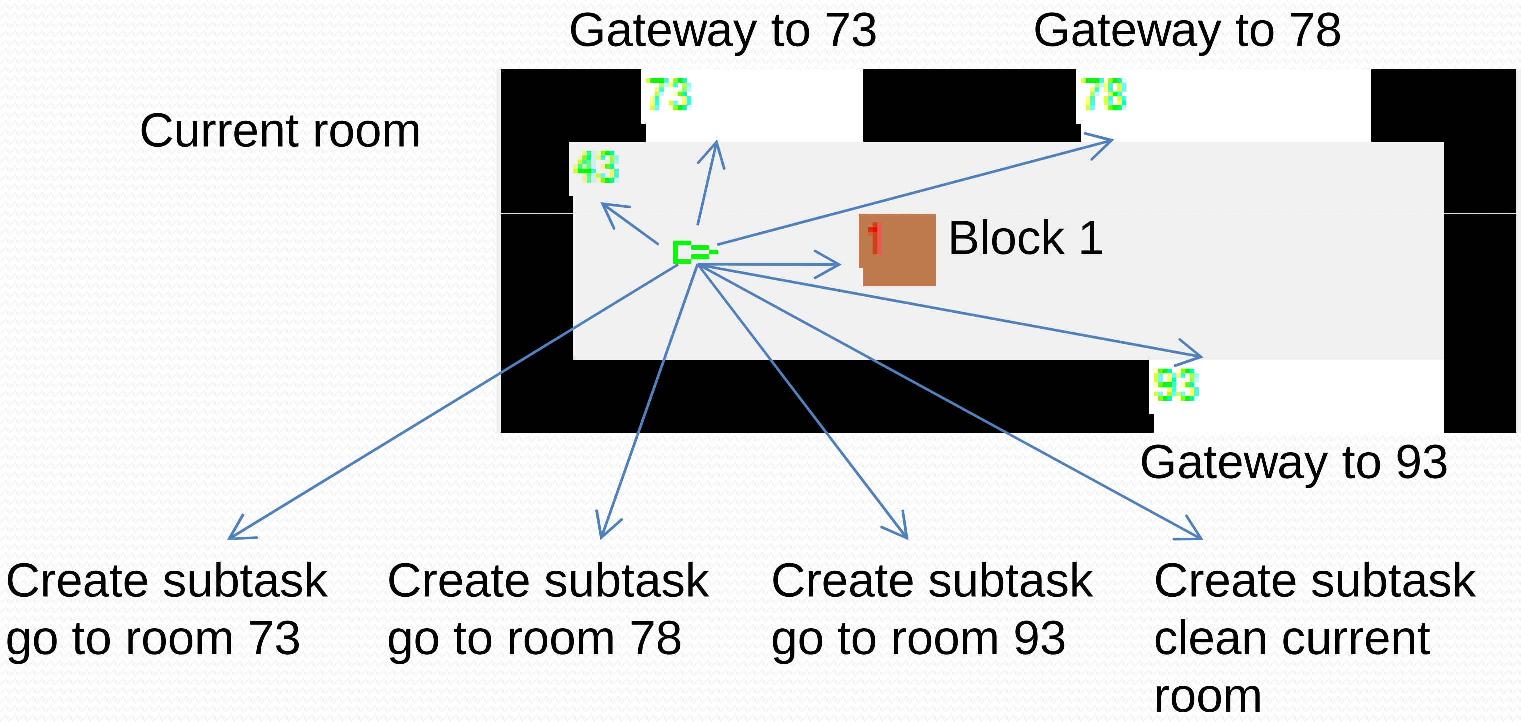
- Intensity = "Surprise Factor" * (average of other appraisals)
  - "Surprise Factor" determined by Outcome Probability and Discrepancy from Expectation appraisals

- Valence = Average of valenced appraisals
  - Conduciveness, Intrinsic Pleasantness

Clean House Domain



Stimuli in the Environment



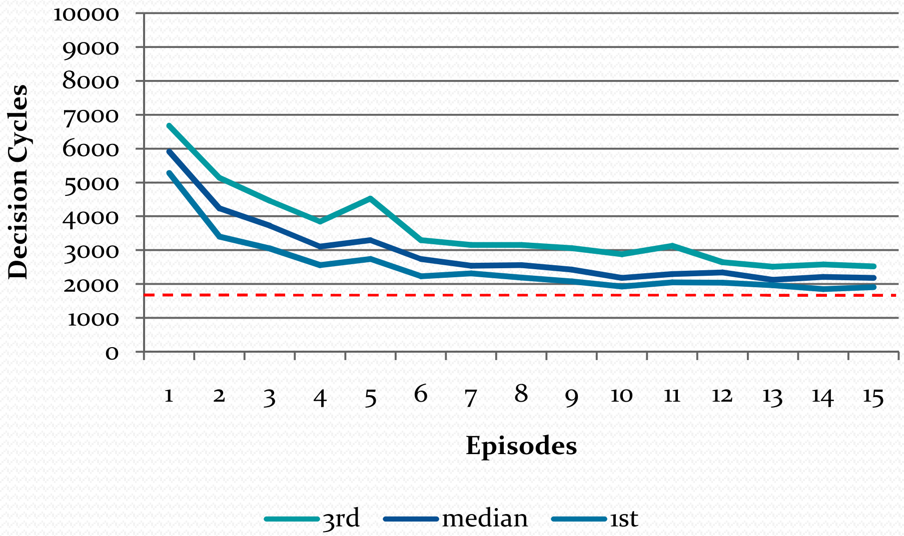
Adapting to a Continuous Environment

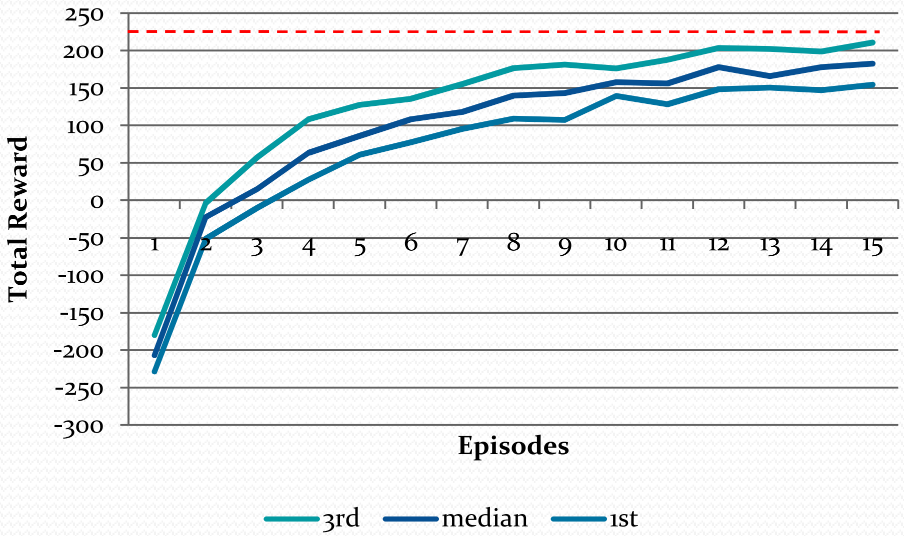
- Temporally extended actions
  - Emotion only active until Intend starts
    - Prevents temporally-extended actions from dominating reward
- Temporally separated states
  - Soar-RL can jump over gaps (operators with no associated RL value), discounting rewards based on size of gap (decision cycles)

Learning

- In this domain, the agent is only learning what to Attend to (including Tasking)
  - Not learning what action to take
- Goal: What is the impact of various appraisals?
  - Disabled most and developed a few
    - Conduciveness
    - Discrepancy from Expectation and Outcome Probability
    - Goal Relevance
    - Intrinsic Pleasantness

Conduciveness



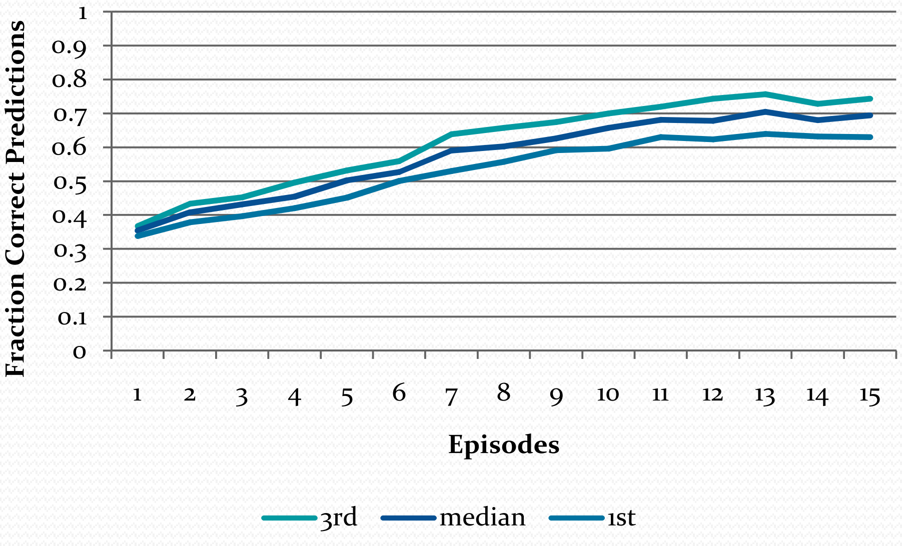


Final Episode Failures	Trial Failures	Total Failures
6%	24%	7.6%

Outcome Probability and Discrepancy from Expectation

- Gave the agent ability to learn task model
  - Similar to episodic and semantic memory
  - Records which stimuli occur in sequence
  - “Strength” of links indicate which sequences are most frequent/recent
- Used to make predictions and determine Outcome Probability
  - Predictions used to determine Discrepancy from Expectation
- As agent settles on some behavior, prediction accuracy should increase → lower “surprise factor” → lower intensity → lower reward

Outcome Probability and Discrepancy from Expectation

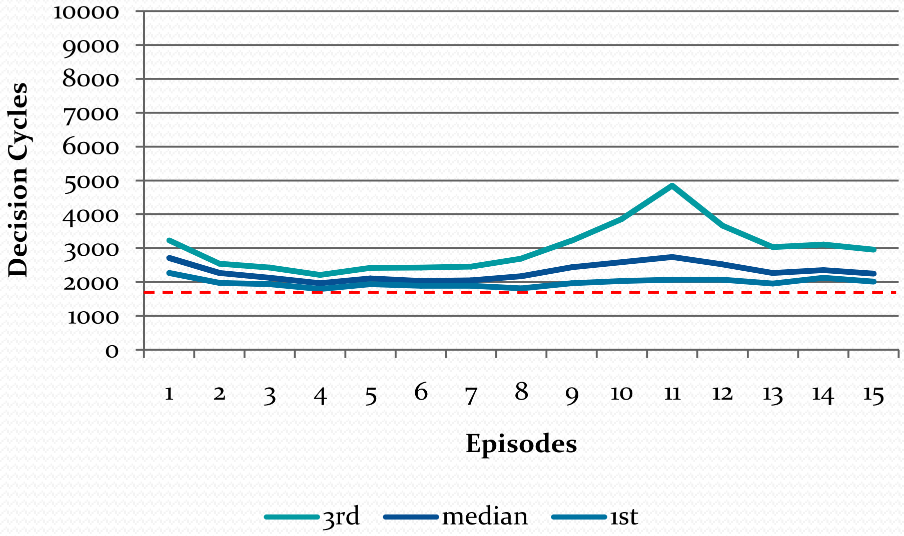


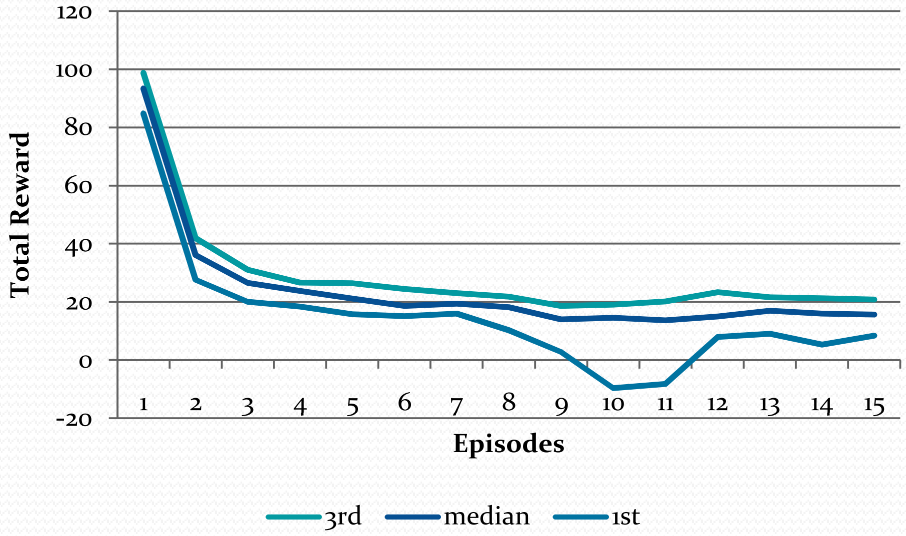
Final Episode Failures	Trial Failures	Total Failures
0%	0%	0%

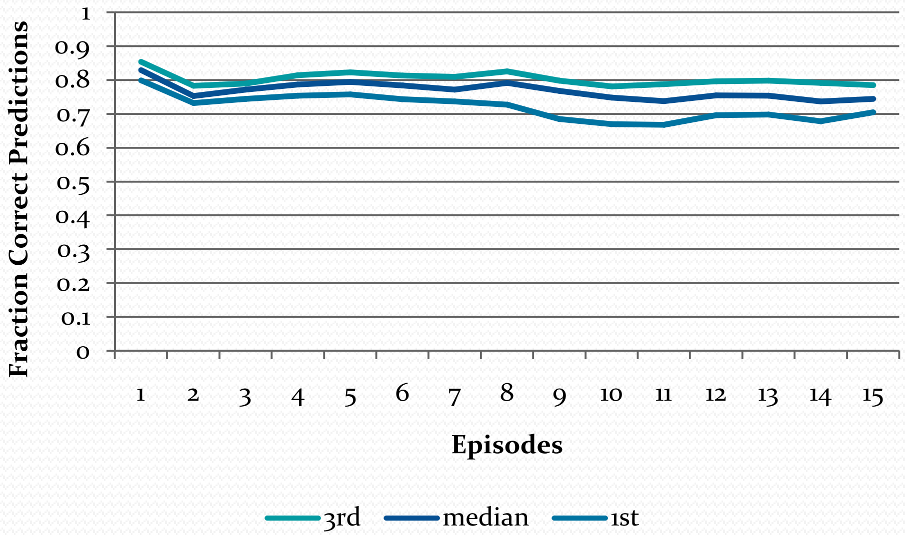
Goal Relevance

- Agent has knowledge about which stimuli are "on" or "off" the path to the goal
    - This determines the value of Goal Relevance
- The value of Goal Relevance for some stimulus is used to "boost" the RL value of that stimulus

Goal Relevance Results



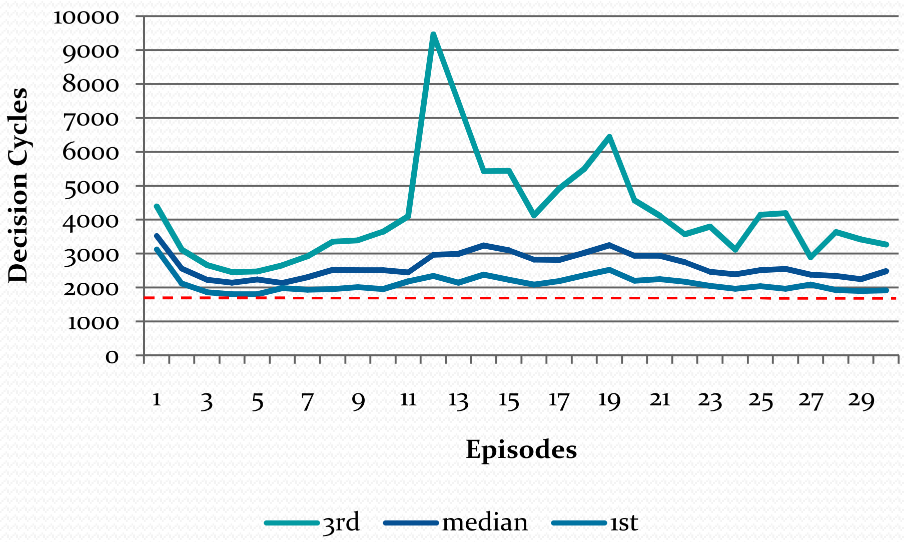


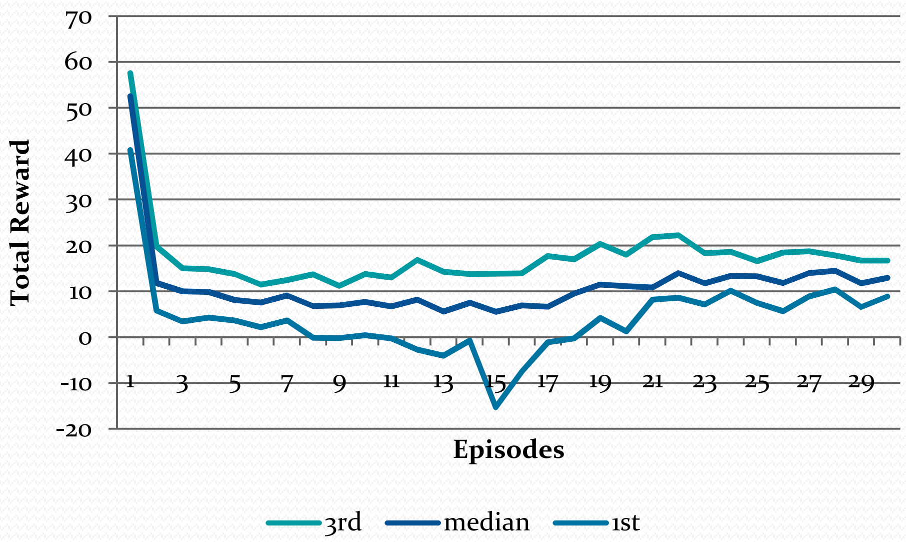


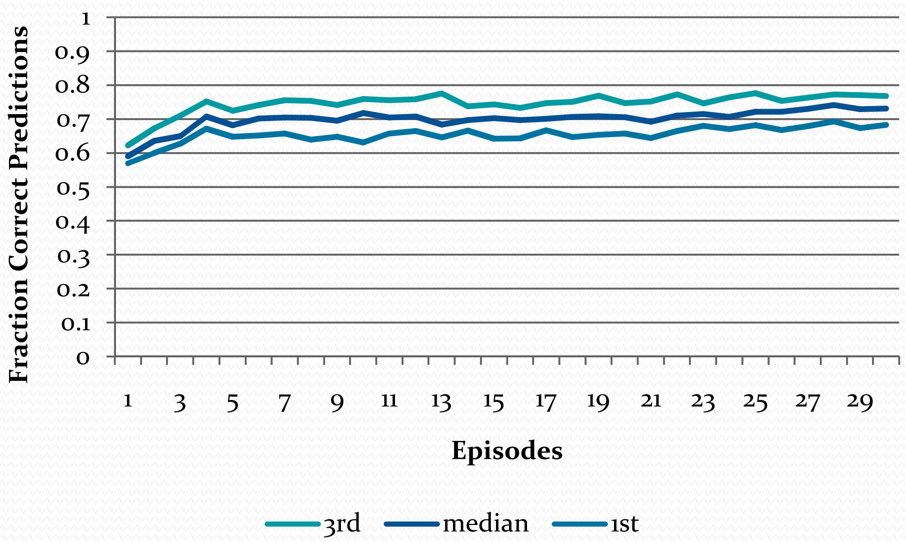
Knowledge Reduction

- Removed knowledge about:
  - How to get to non-adjacent rooms
  - How to put blocks down in storage room
- Path values in these cases are “unknown”
- Agent has to learn what to do

GR Knowledge Reduction Results



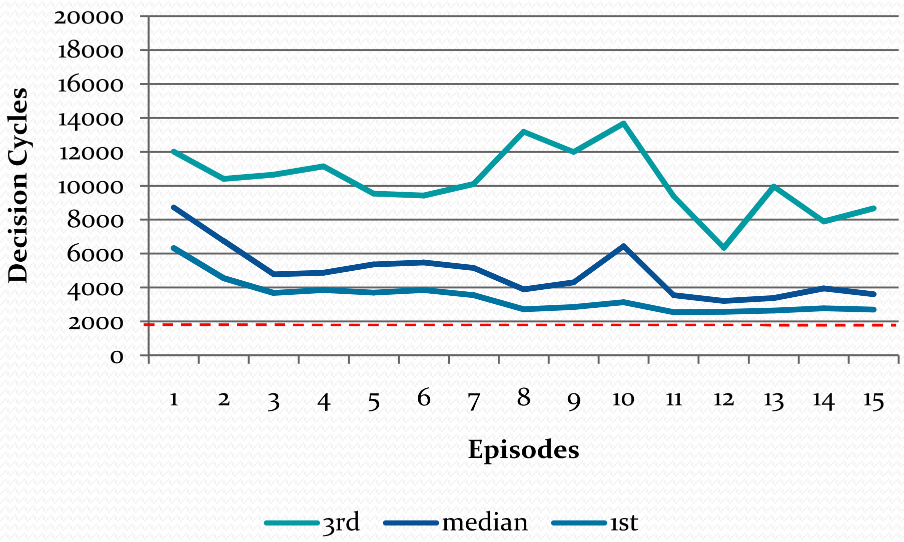


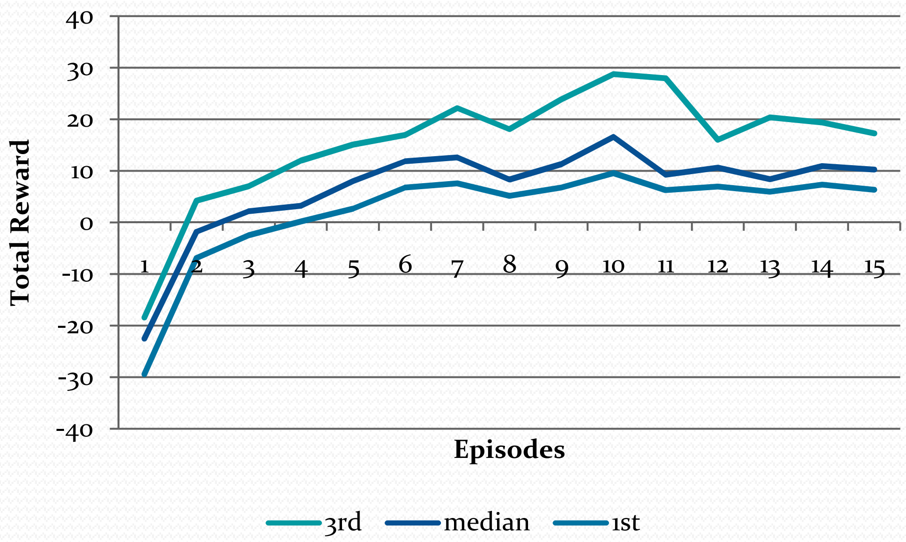


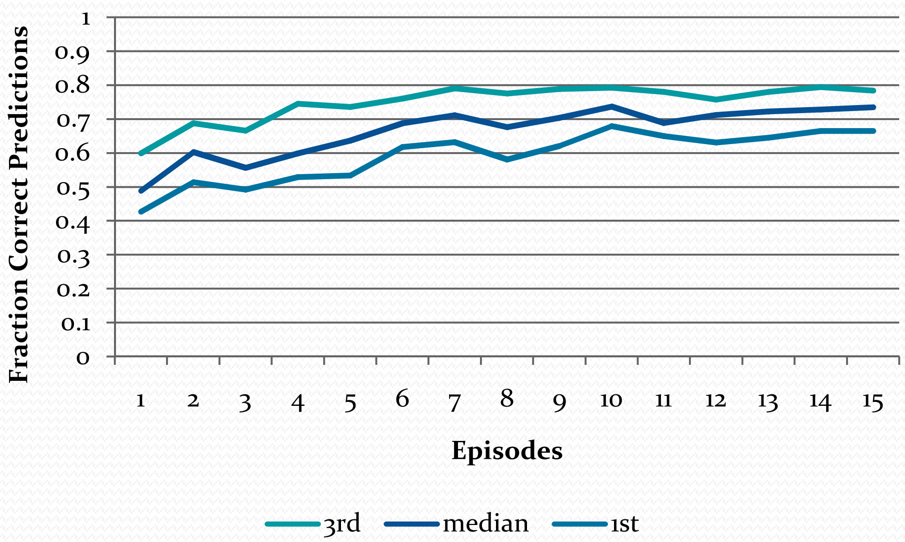
Intrinsic Pleasantness

- How attracted the agent is to a stimulus independent of the current goal
  - Influences valence and intensity
- Made blocks intrinsically pleasant
  - This is good because blocks need to be attended to get cleaned up
  - This is bad because agent may be distracted by blocks that have already been cleaned up
- Experiment done without Goal Relevance

Intrinsic Pleasantness Results







Summary

Conduciveness	Foundation to learning. Agent learns to perform the task better over time.
Outcome Probability, Discrepancy from Expectation	Introduced learned task model for generating predictions as basis for generating values for these appraisals. Agent learns to predict better over time. Also results in much improved failure rates.
Goal Relevance	Used to "boost" value of proposed Attend operators. Agent does extremely well (except for failures), to the point where it almost isn't learning, raising questions about the value of other appraisals. Knowledge about Goal Relevance was reduced, leading to more learning.
Intrinsic Pleasantness	Used to provide a task-independent bias on valence. Results are mixed, as expected, but agent generally learns to overcome problems.

Nuggets

- System scales to complex environment
- Learning works
- Each appraisal influenced behavior and learning

Coal

- Many appraisals apparently unnecessary for this task
- Performance is not perfect (some failures)
- Need to develop more complex appraisal models