FEELING AS INTRINSIC REWARD

Bob Marinier

John Laird

27th Soar Workshop

May 24, 2007

OVERVIEW

- What are feelings good for?
  - Intuitively, feelings should serve as a reward signal that can be used by reinforcement learning
- Outline
  - Emotion theory
  - Agency theory
  - Domain description
  - Features and results

APPRAISAL THEORIES

- Appraisal theories postulate a set of dimensions that a person uses to evaluate a situation with respect to a goal
  - Typical dimensions include:
    - Novelty
    - Goal relevance
    - Goal conduciveness
    - Causal agent/motive
    - Outcome probability
    - Discrepancy from expectation
    - Power/control
- Regions of appraisal space map onto emotions
  - Examples:
    - high relevance + low conduciveness + other agent = anger
    - high relevance + high conduciveness = joy

EMOTION, MOOD AND FEELING

- Emotion is about current situation
- Mood provides historical context
- Feeling is what agent actually (internally) perceives

- Emotion = result of current appraisals
- Mood = “average” of recent emotions
- Feeling = Emotion “+” Mood

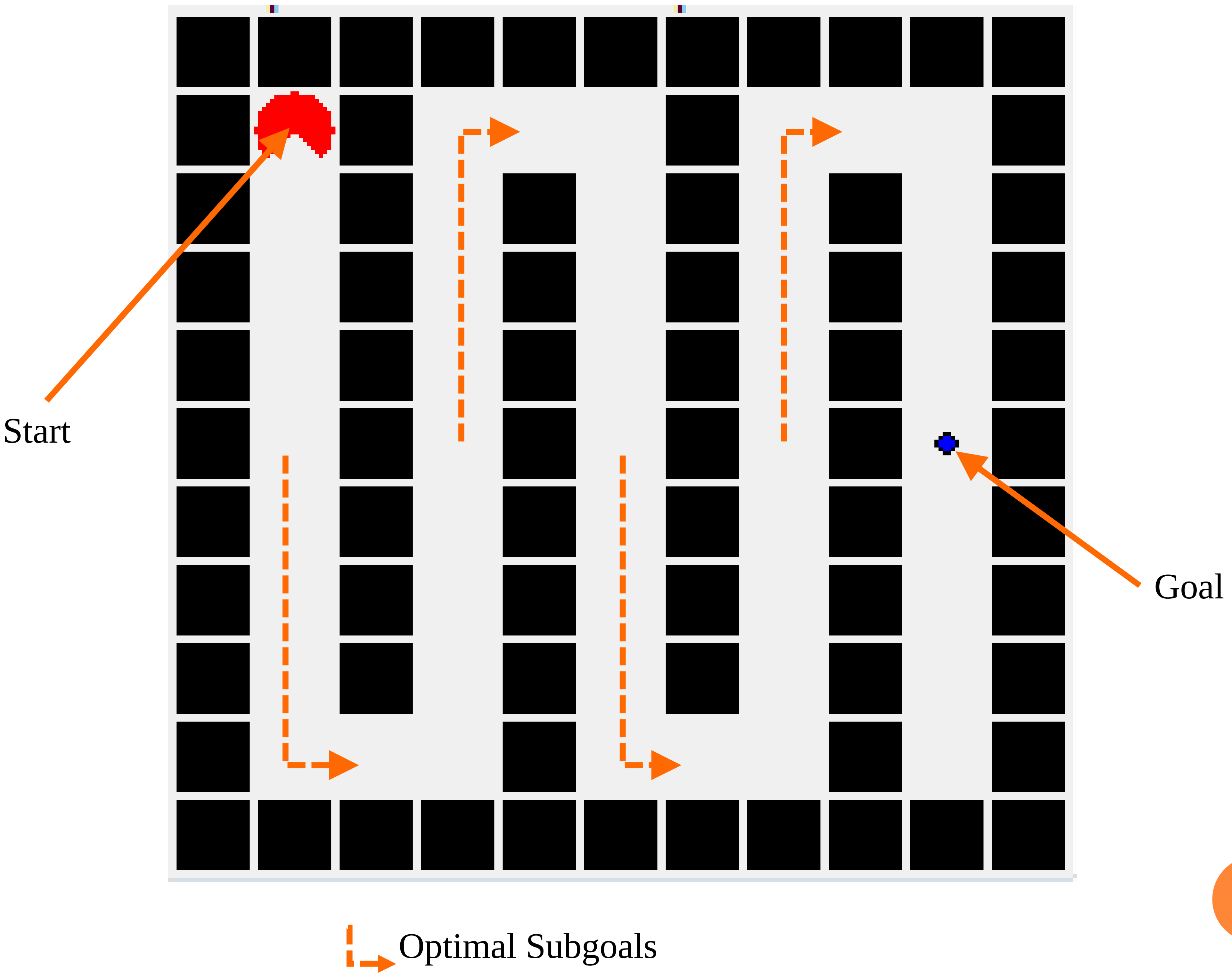
- Feeling intensity = “surprise factor” * average of appraisals
  - Surprise factor based on Outcome Probability and Discrepancy from Expectation

THEORY OF AGENCY: NAFO

- Agent is organized around Newell's Abstract Functional Operations (NAFO)
  - Perceive: Input phase
  - Encode
    - Elaborations create general event structures
    - Novelty appraisals generated
  - Attend
    - Choose which event to process next
    - Enable complete appraisal generation
  - Comprehend: Generate complete appraisals
  - Intend
    - Determine motor actions
    - Create prediction
  - Decode, Motor: Output phase, environment processing
  - Task: Create and manages goals/subgoals

DOMAIN

- Wayfinding in Eaters
  - Agent must find path from starting point to goal
  - Appraisal heuristic: Dynamic difference reduction
    - Moving directly towards goal is good
    - If can’t move directly towards goal, can create a subgoal
    - Movement in subgoal not as good as movement in main goal
  - Simple episodic memory
    - Agent has some idea of whether its getting closer to the goal
      - Classifies a subgoal as good or bad


Start
Goal
Optimal Subgoals

RL DESCRIPTION OF DOMAIN

- Learns which functional operations to execute to get to goal with highest reward
- Encoded event: Direction + on path + passable (+ goal)
- Actions
- Attend to an event (then Ignore or Intend)
  - State: Event + good/bad subgoal
- Create a subgoal
  - State: All 4 events + subgoal
- Retrieve a supergoal
  - State: All 4 events
- Reward = (feeling intensity)*Valence(goal conduciveness)
- This is hard
- States are only partially observable
- Non-Markovian (history matters): Recent states influence reward via mood

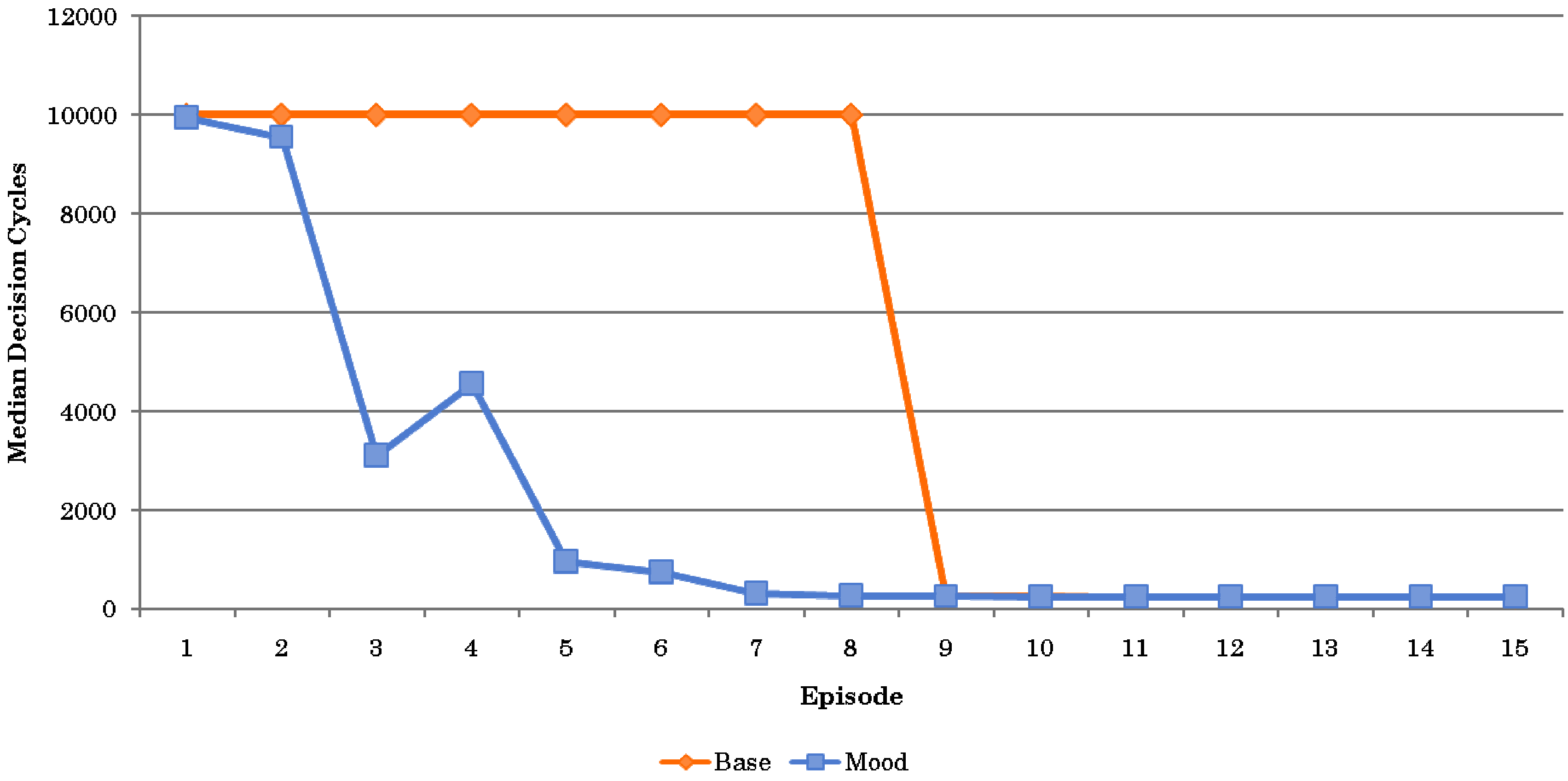
METHODOLOGY

- 15 episodes, 50 trials/episode
- Episode cutoff after 10000 steps
- Reporting median
- Action selection: epsilon greedy
- Exploration rate starts at 10%, decreases to 0 in 11^{th} episode
- Learning rate starts at 0.3, decreases to 0 in 15^{th} episode
- Agent keeps cognitive map across episodes (where applicable)

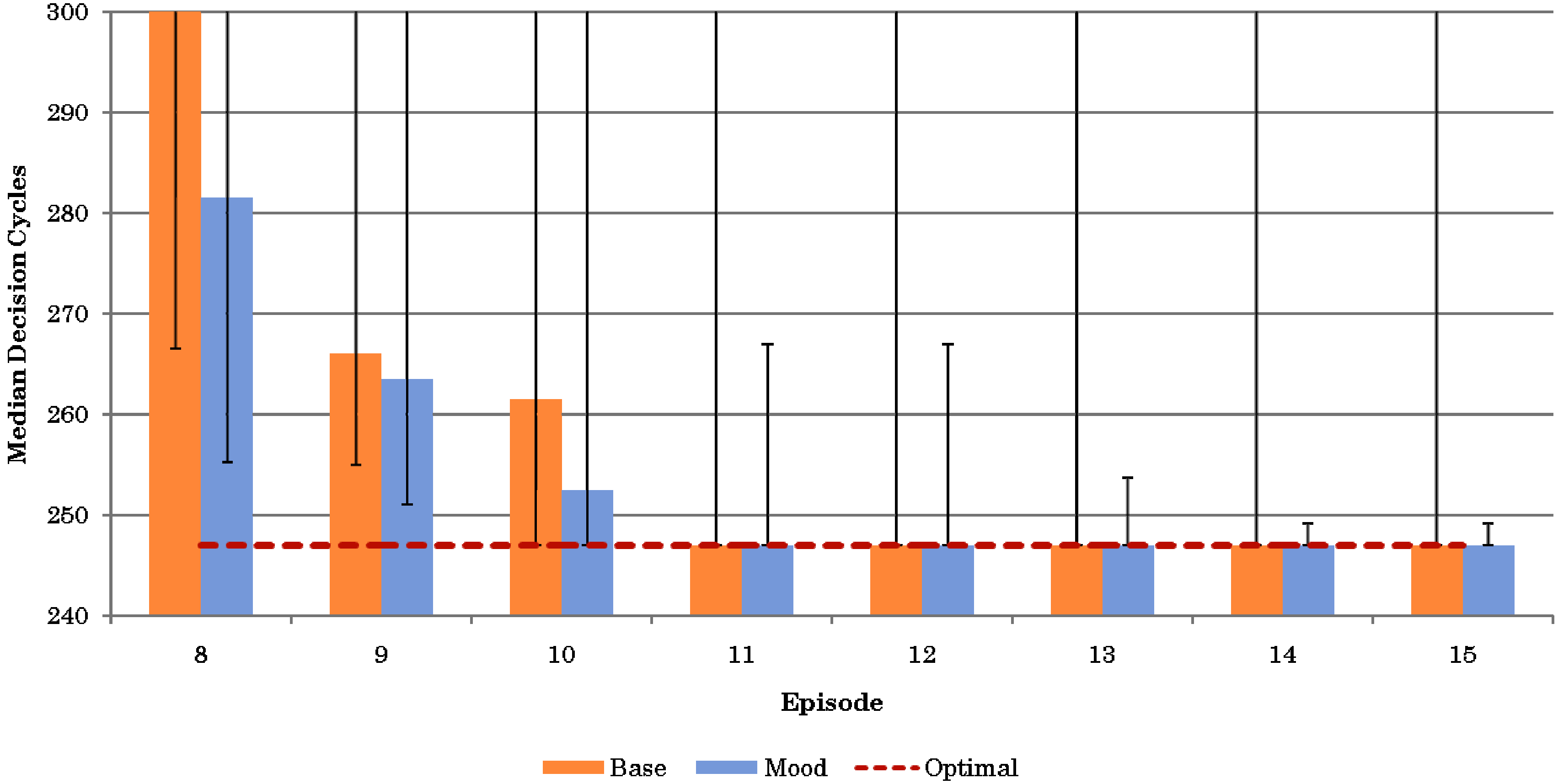
RESULTS: VARIABLES TESTED

- Mood (on/off)
- Cognitive map (on/off)
  - A way to learn about space as a whole
- Dynamic learning (on/off)
  - A way for the agent to adjust its learning rate

RESULTS: MOOD



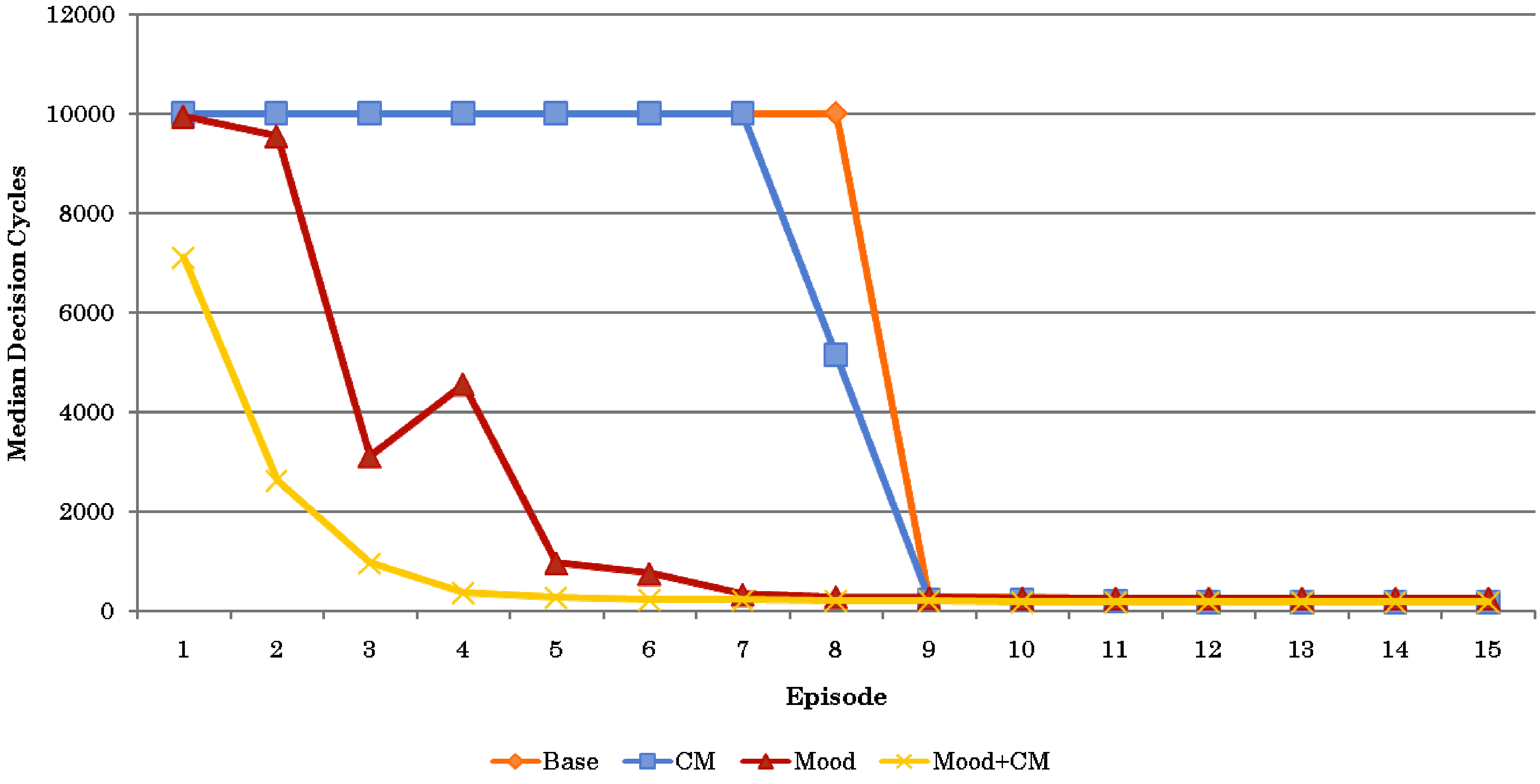
RESULTS: MOOD



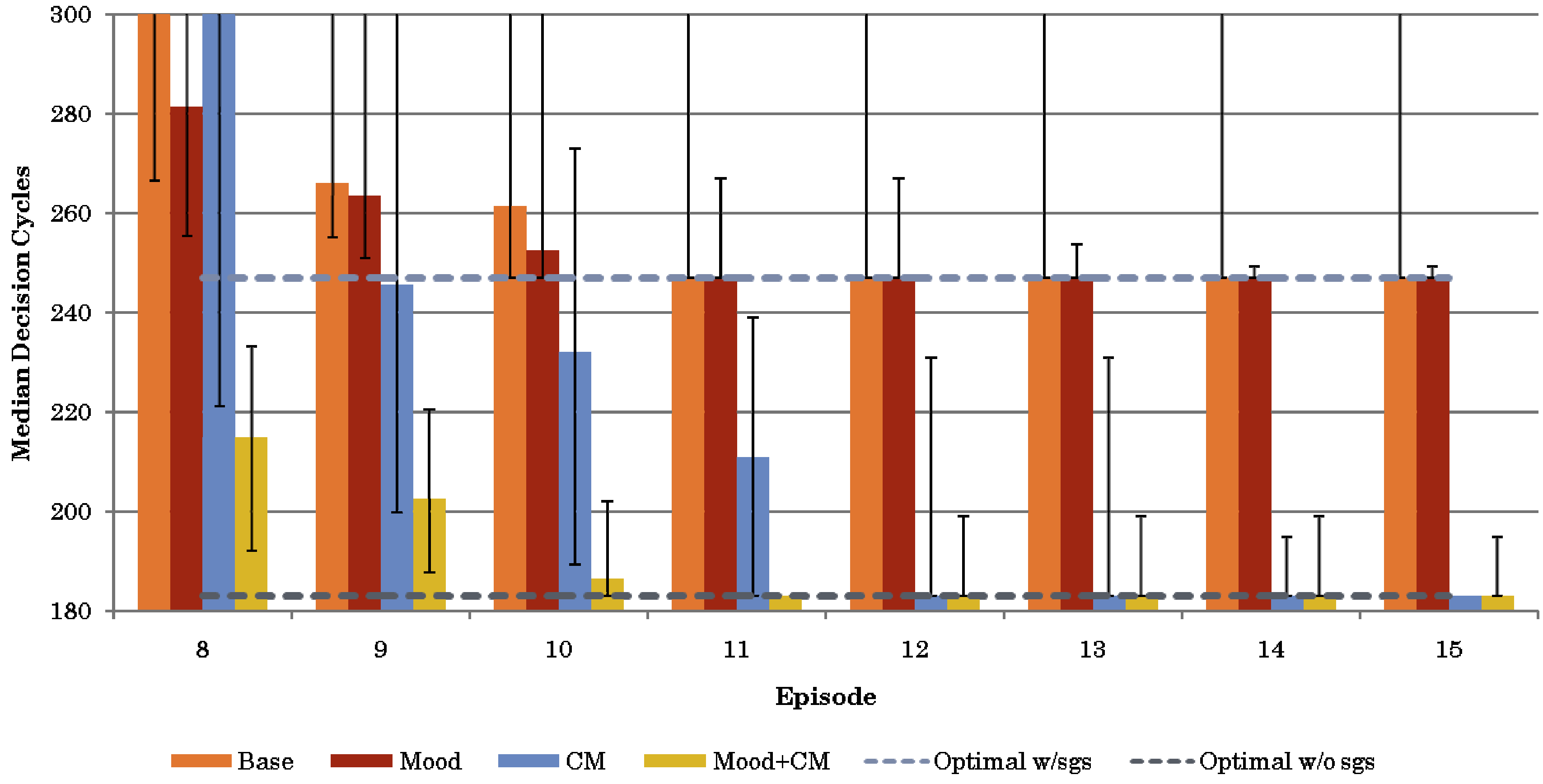
COGNITIVE MAPS

- Using pure RL, agent will never learn overall space
  - To maximize reward, always has to create subgoals
- A cognitive map is a landmark-based map of problem space
  - Allows agent to “see” direct route to goal (i.e. states are encoded as on path)
  - Allows agent to skip subgoals
- Cognitive map is used to make predictions based on experience
  - This makes the RL problem nonstationary
    - As the agent gets better at predicting events, the reward it gets goes down

RESULTS: MOOD & COGNITIVE MAP



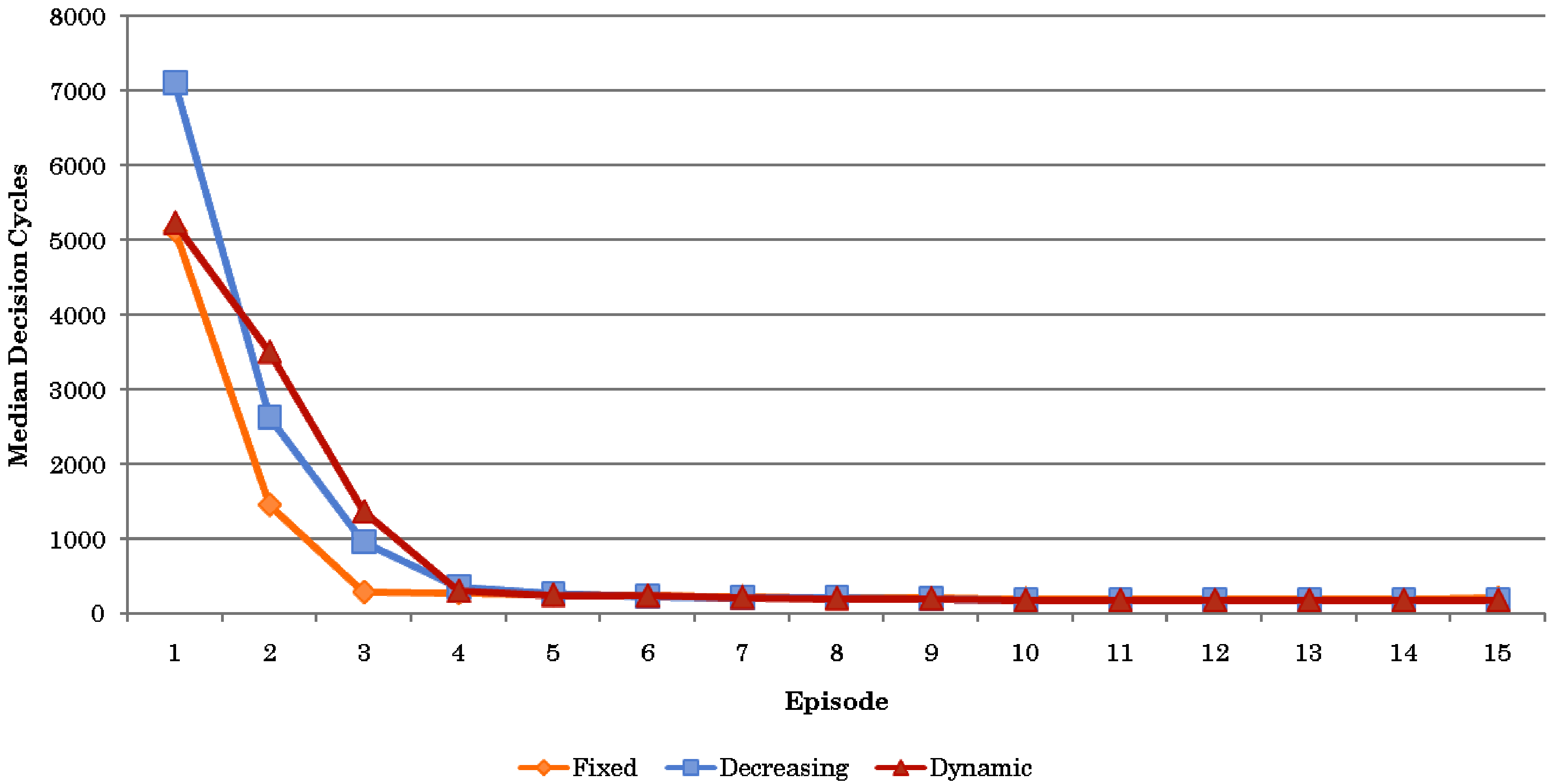
RESULTS: MOOD & COGNITIVE MAP



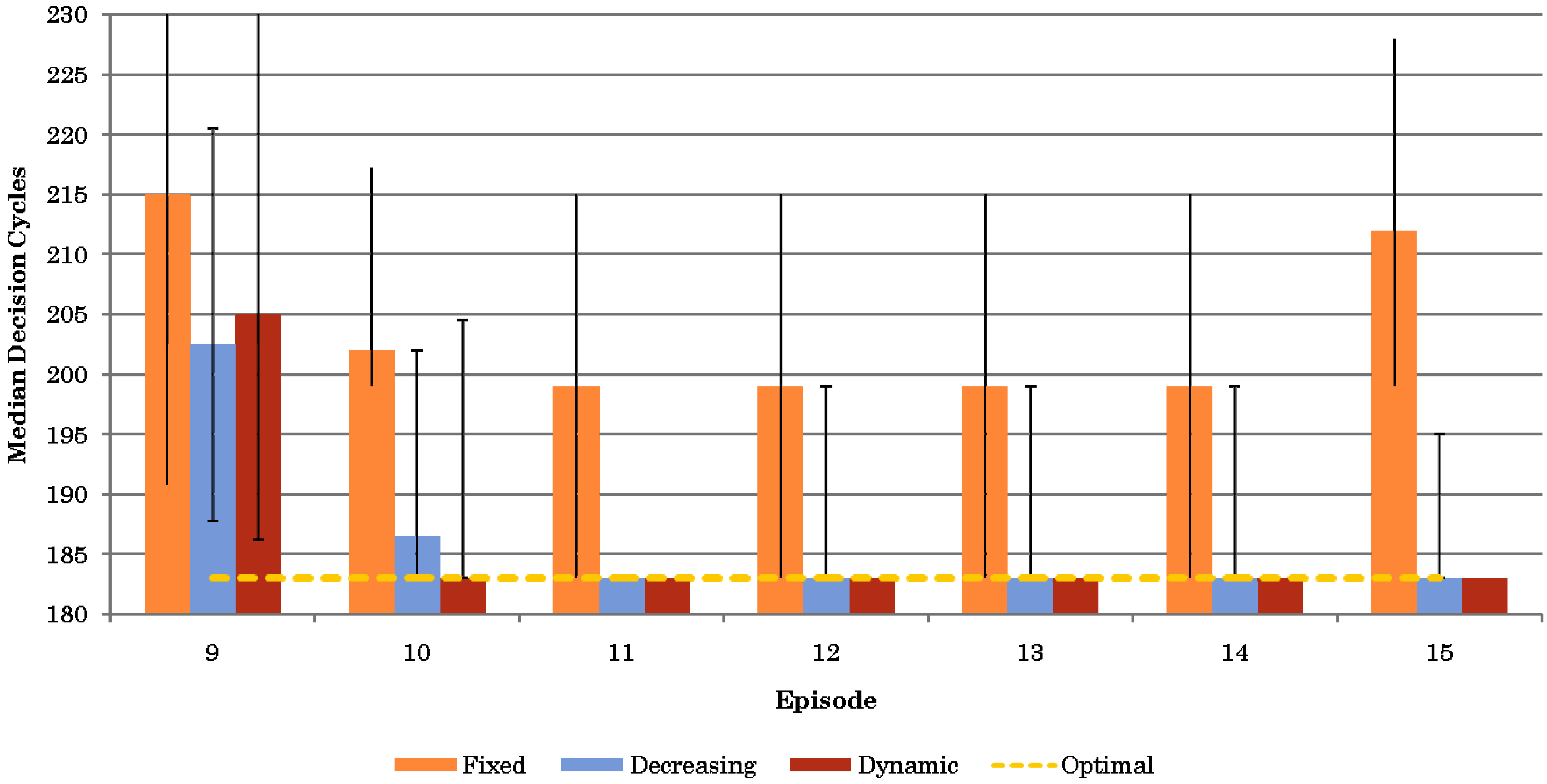
DYNAMIC LEARNING RATE

- RL typically decreases learning rate over time
  - Often required for convergence
  - No theory of source of decrease
  - Requires knowing the number of episodes in advance
- Idea: If agent is able to predict what will happen next, don't need to learn anything
  - Learning rate = feeling intensity
    - Since feeling intensity includes "surprise factor", an agent will only learn when it is "surprised"

RESULTS: LEARNING RATE (MOOD+CM)



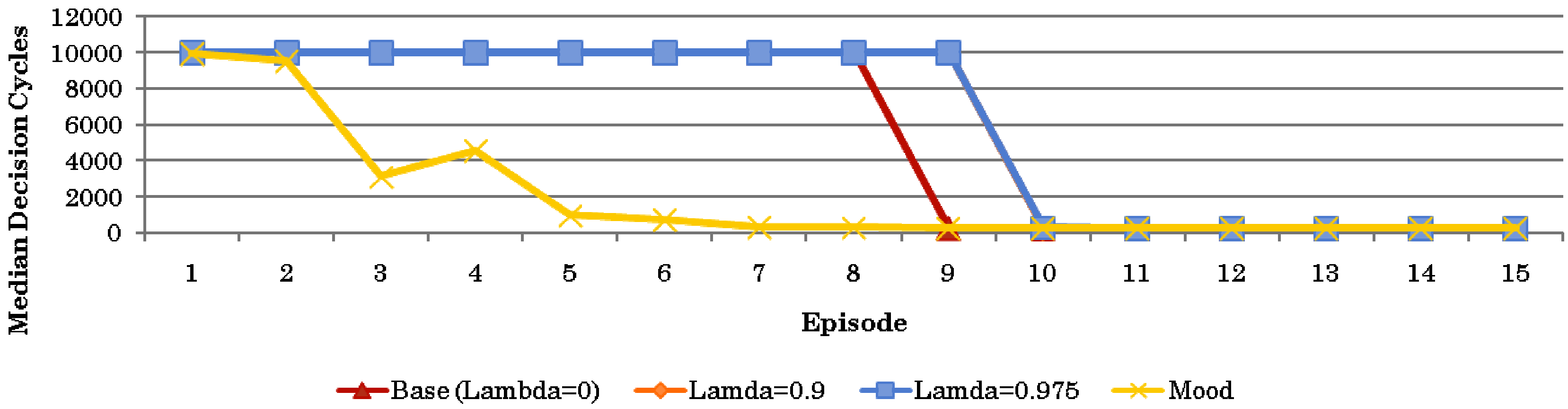
RESULTS: LEARNING RATE (MOOD+CM)

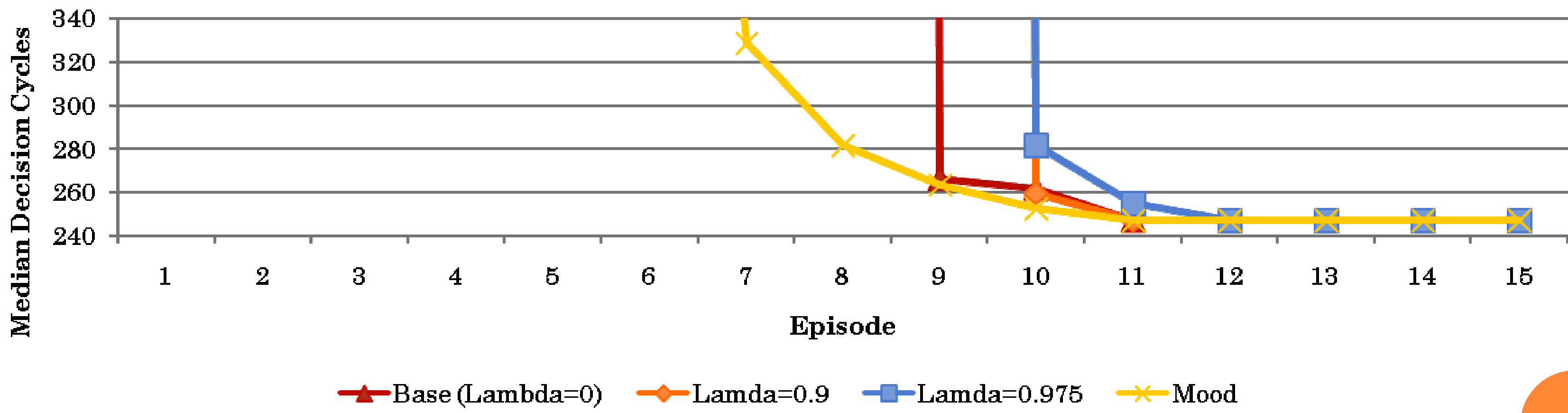


ELIGIBILITY TRACES

- Normal RL only backs up reward one step at a time
- Eligibility traces allow the agent to back up many steps, with decreasing influence determined by Lambda
  - Lambda=0: standard RL
  - Lambda=1: Monte-Carlo (all previous states equally influenced)
- Conceptually similar to mood
  - Mood: current reward influences reward (and hence value) for next states
  - Eligibility traces: current reward influences value for last states
- Will compare to mood, and in combination with mood

RESULTS: MOOD VS. ELIGIBILITY TRACES





DISCUSSION

- Agent learns fast!
  - Reward at every dc helps a lot
- Mood accelerates learning
  - Estimates the value of states in the absence of emotion
- Cognitive map improves performance
- Dynamic learning rate helps
- Agent needs a better episodic memory
  - Has a really hard time telling states apart in bad subgoals
    - Often gets stuck in looping behavior
    - Individual runs often regress in performance

- Nuggets
  - Links cognition, affect, and learning
  - Agent learns well
- Coal
  - Agent is fragile and (sometimes) inconsistent
  - Simple domain means (relatively) simple implementation
- Future work
  - Move to continuous, dynamic domain
    - Should reduce partial observability issues
    - Force increased sophistication appraisal, functional operations, cognitive map, episodic memory
  - Modulate exploration rate