

A Multi-Domain Evaluation of Scaling in a General Episodic Memory

Nate Derbinsky, Justin Li, John E. Laird

University of Michigan



24 July 2012



support from

AAAI 2012 - Toronto

Motivation

Prior work has provided evidence that cognitive systems with **episodic memory** can be...

- more capable in problem solving
- better able to account for human psychological phenomena
- more believable as virtual characters and companions

Limitations of current approaches:

- Restricted representation
- Task-specific processing
- **Little evaluation of real-time use in long-term tasks**

Multi-Domain Evaluation of Scaling

- Episodic memory in Soar [Laird, 2012]
 - Relational representation, task-independent integration
- Existing agents from many diverse tasks (49)
 - Linguistics, planning, games, robotics
- Long agent runs
 - Hours-days [cognitive] RT
(10,000s – 100,000,000s episodes)
- Evaluate at each n episodes
 - Memory consumption
 - Reactivity for >100 task-relevant cues
 - Maximum time for cue matching <? 50 msec.

Cognitive Capabilities

Virtual Sensing
Detecting Repetition
Action Modeling
Environmental Modeling
Explaining Behavior
Managing LT Goals
Predicting Success/Failure

Outline

1. Overview of episodic memory in Soar
2. Evaluation domains^{*}
 - a) Word sense disambiguation
 - b) Video games & robotics

^{*}See paper/poster for results of using Soar's episodic memory for repeated-state detection in 44 PDDL planning domains.

Episodic Memory in Soar

Problem Formulation

Representation

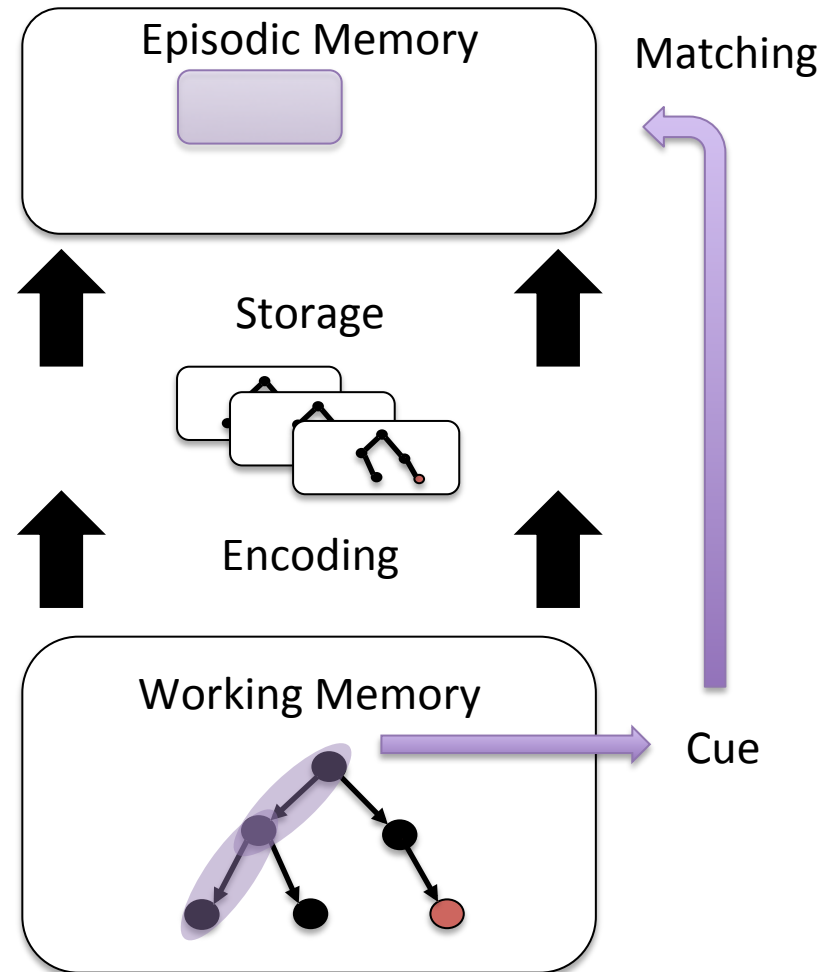
- Episode: connected di-graph
- Store: temporal sequence

Encoding/Storage

- Automatic
- No dynamics

Cue Matching

- Cue: acyclic graph
- Semantics: desired features in context
- Find the most recent episode that shares the most leaf nodes in common with the cue



Episodic Memory in Soar

Algorithmic Overview

Storage (*only process changes*)

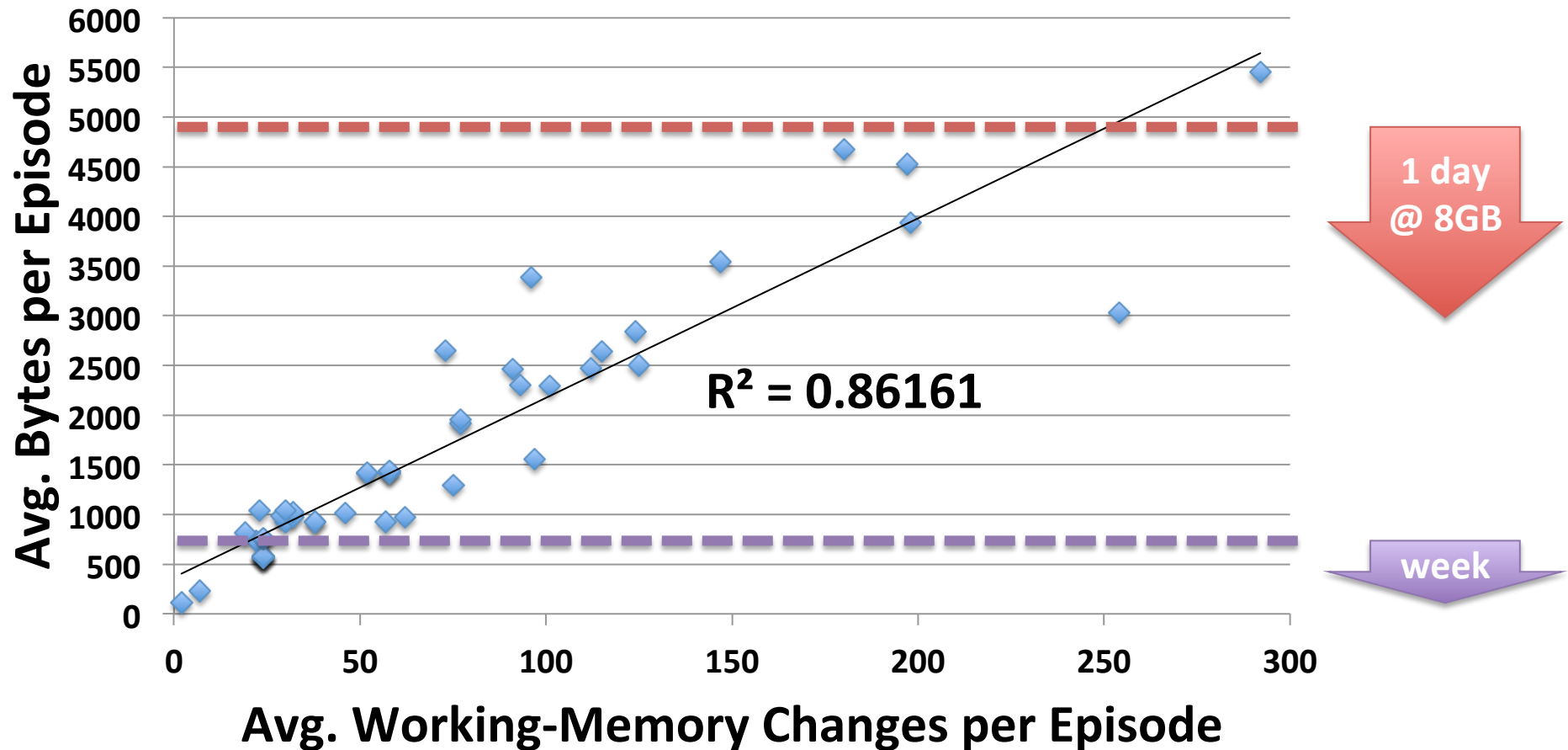
- Capture WM-changes (Δ 's) as temporal intervals

Cue Matching (*reverse walk of cue-relevant Δ 's*)

- 2-phase search
 - Only graph-match episodes that have all cue features *independently*
- Only evaluate episodes in which cue features change
- Incrementally re-score episodes

Episodic Memory in Soar

Storage Characterization



Episodic Memory in Soar

Cue-Matching Characterization

Assumptions

- Few changes per episode (*temporal contiguity*)
- Representational re-use (*structural regularity*)
- Small cue

Scaling

- Search distance (# changes to walk)
 - *Temporal Selectivity*: how often does a feature change
 - *Feature Co-Occurrence*: how often do features co-occur within a single episode
- Episode scoring
 - *Structural Selectivity*: how many ways can a cue feature match an episode feature

Word Sense Disambiguation

Experimental Setup

- Input: <“word”, POS>; Output: sense #; Result
 - SemCor (~900K eps/exposure) x 5 (~4.5M episodes)
- Agent
 - Maintain context as n-gram: $\langle w_{t-1}, w_{t-2}, \dots w_{t-n} \rangle$
 - Query episodic memory for context
 - If success, examine prior result, output
 - If failure, *null*

<u>Accuracy</u>	Exposure #1	Exposure #2
2-gram	14.57%	92.82%
3-gram	2.32%	99.47%

Word Sense Disambiguation

Results

Storage

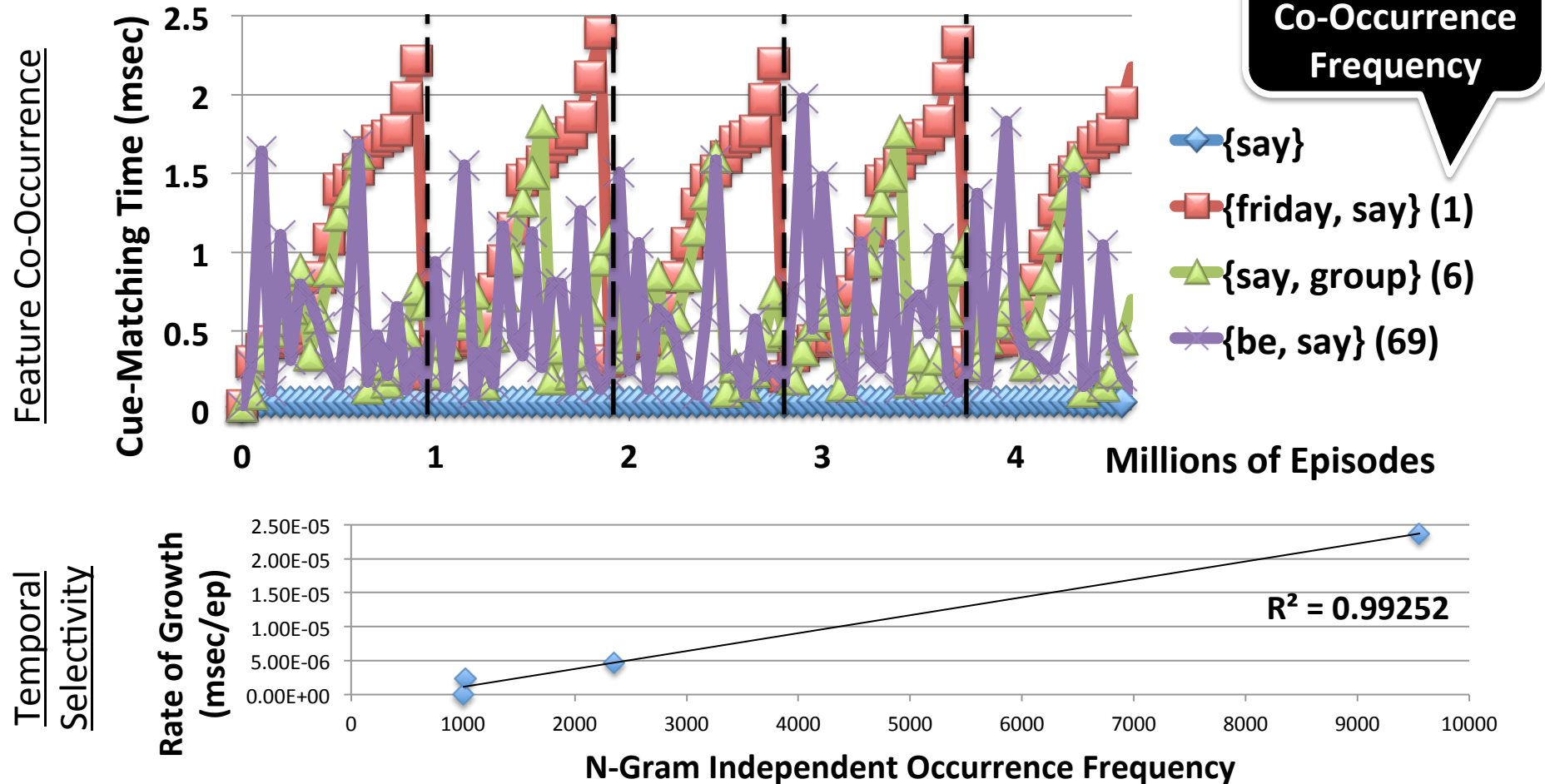
- Avg. 234 bytes/episode
- Max. <1 msec.

Cue Matching

- All 1-, 2-, and 3-gram cues reactive (<50 msec.)
- 0.2% of 4-grams exceed 50 msec.

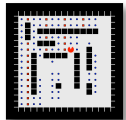
Word Sense Disambiguation

N-gram Cue-Matching Scaling



Video Games & Mobile Robotics

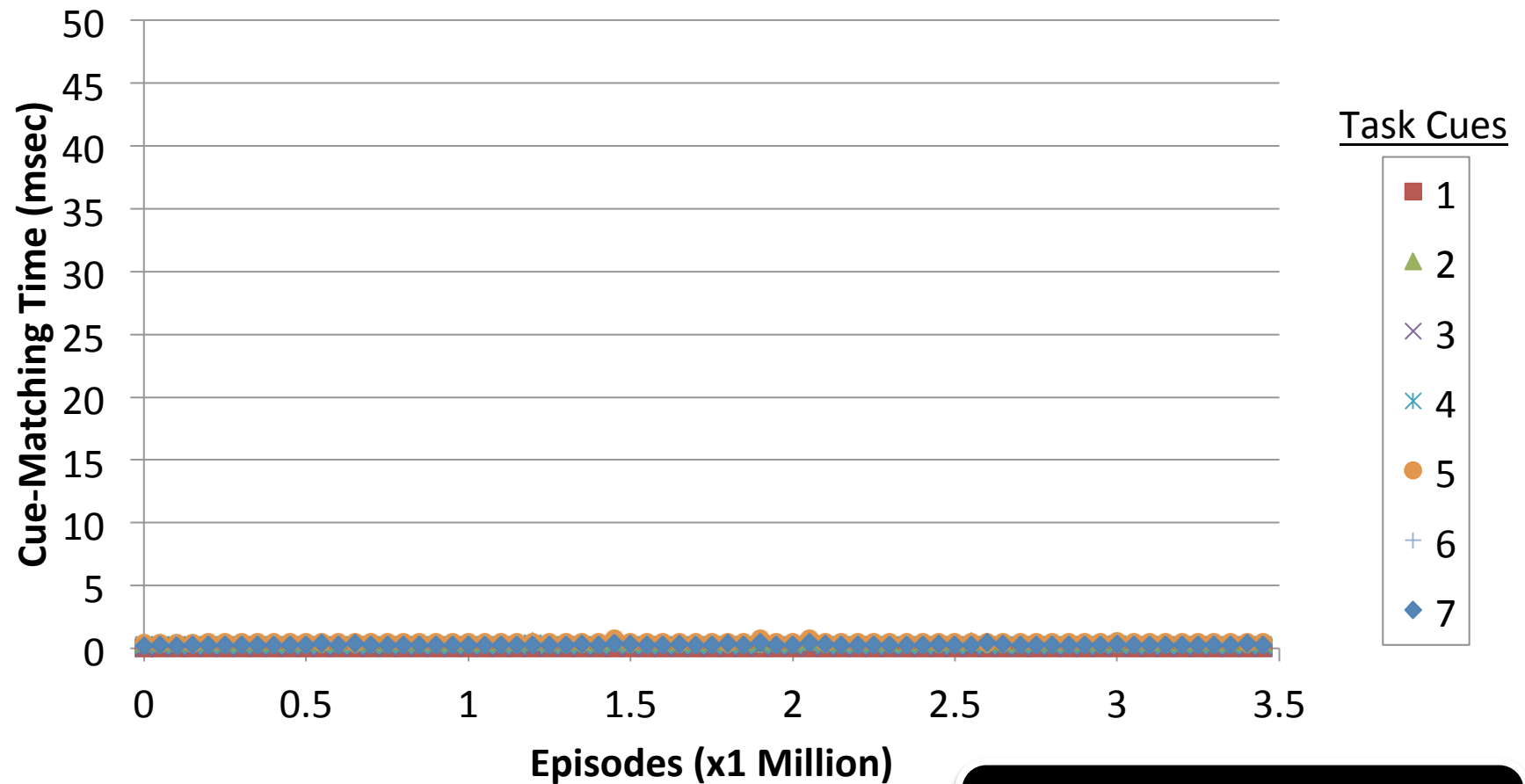
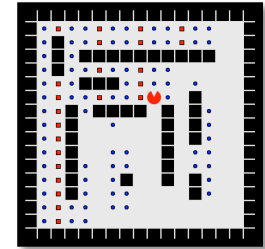
Experimental Setup



Domain	Agent	Duration	Eval. Freq.
TankSoar	<i>mapping-bot</i>	3.5M	50K
Eaters	<i>advanced-move</i>	3.5M	50K
Infinite Mario	[Mohan & Laird '11]	3.5M	50K
Mobile Robotics	[Laird, Derbinsky & Voigt '11]	108M	300K

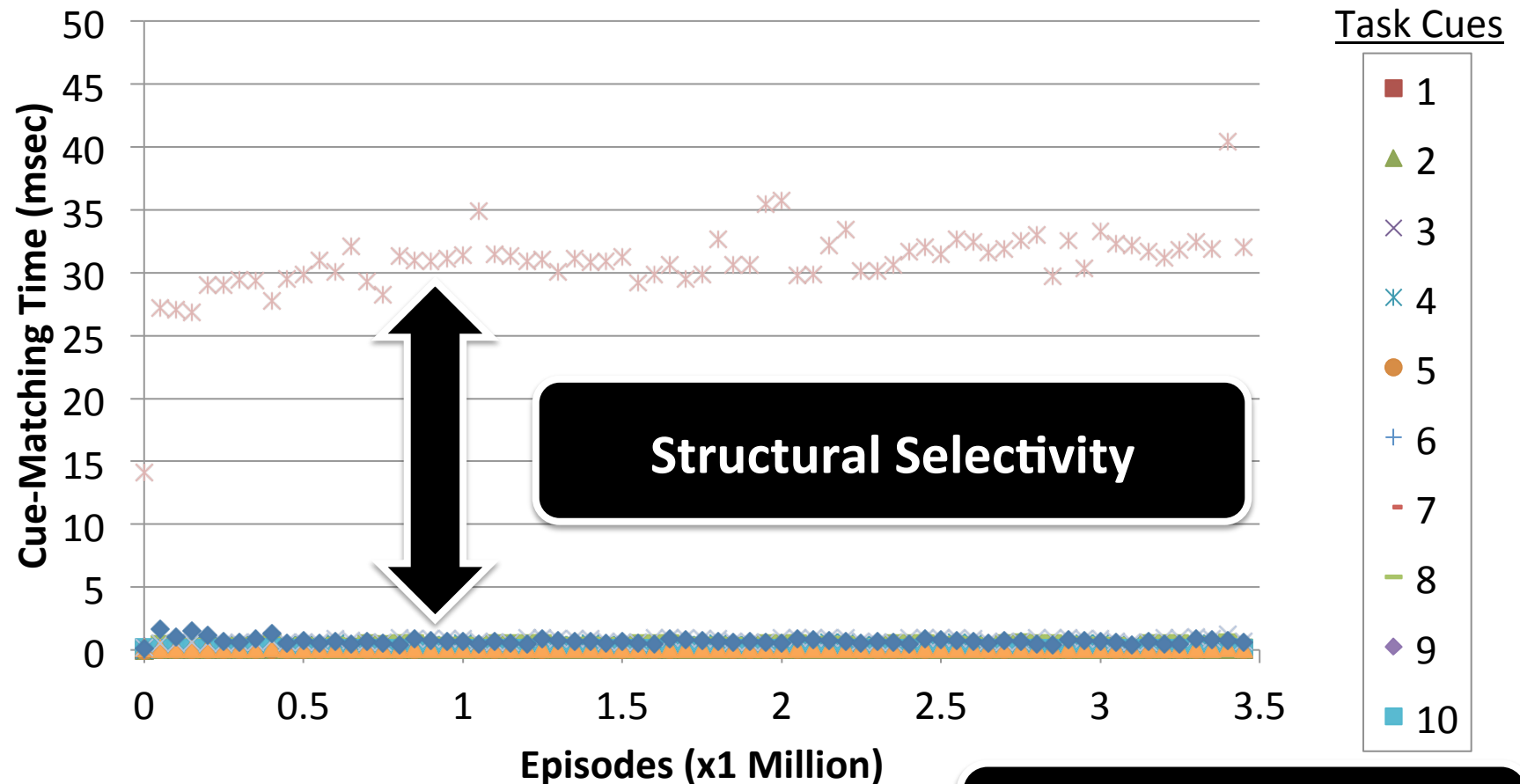
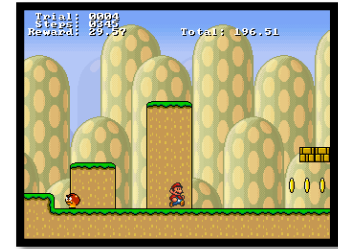
- Hand-coded cues (per domain)

Data: Eaters



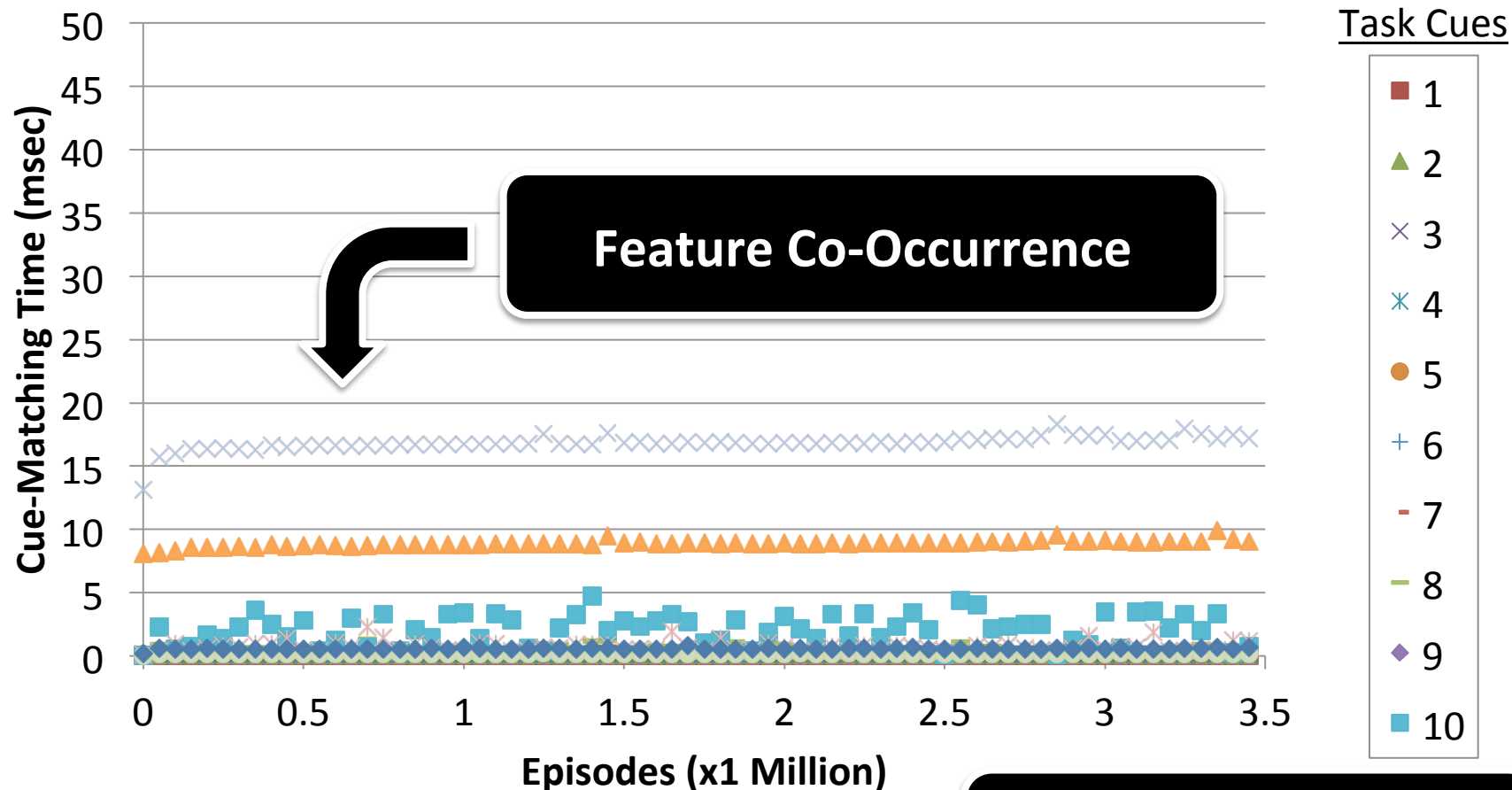
Storage
813 bytes/episode
< 2 msec.

Data: Infinite Mario



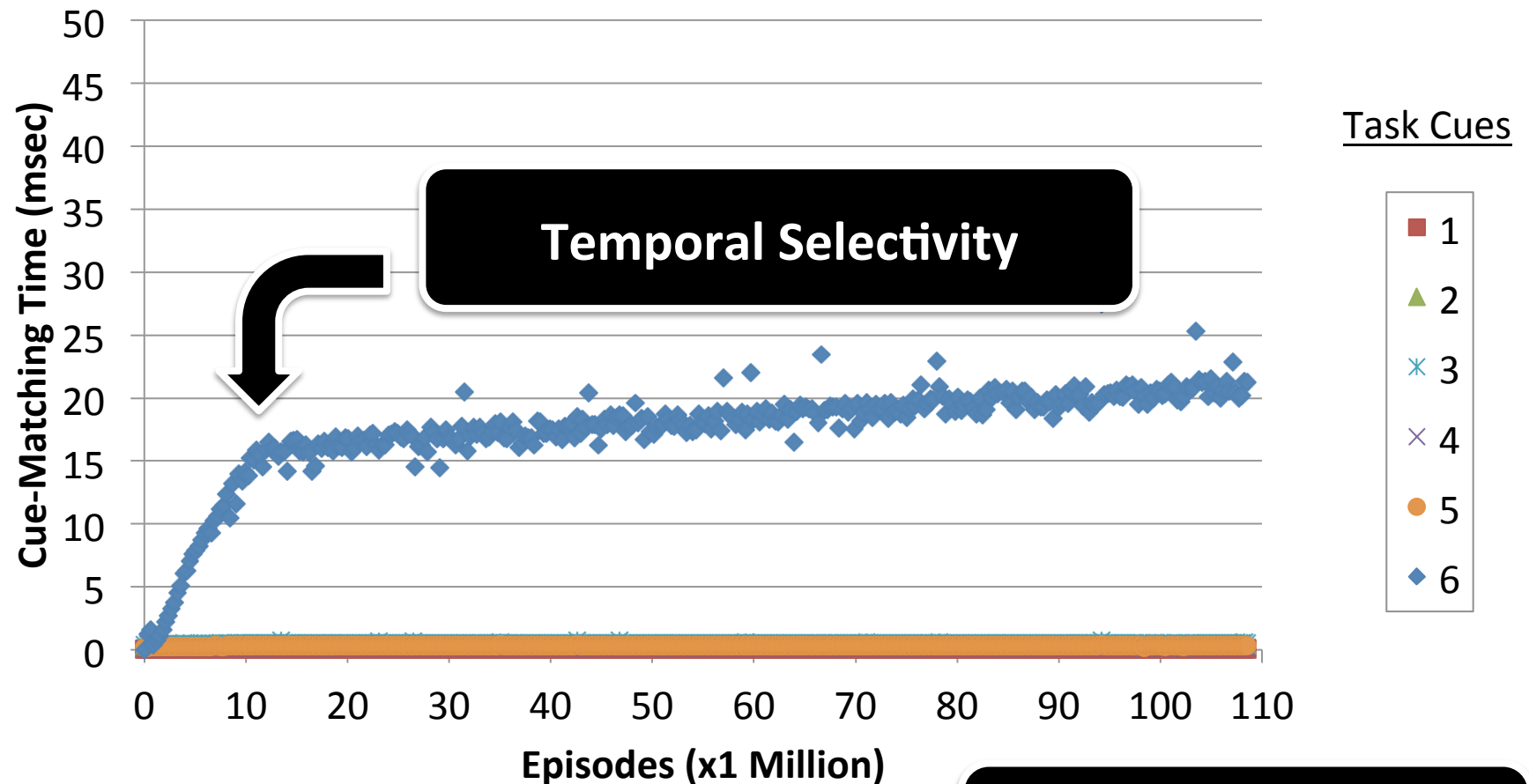
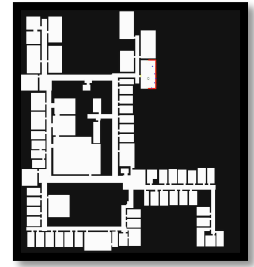
Storage
2646 bytes/episode
< 56 msec.

Data: TankSoar



Storage
1035 bytes/episode
< 19 msec.

Data: Mobile Robotics



Storage
113 bytes/episode
< 4 msec.

Summary of Results

Generality

- Evaluated numerous agents in 49 diverse problem domains
- Episodic cues to support a variety of cognitive capabilities

Reactivity

- <50 msec. storage for all tasks (except temporal discontinuity)
- <50 msec. cue matching for many cues



Scalability

- No growth in cue matching for many cues (days!)
 - Validated predictive performance models
- 0.11 – 5.5 kb/episode (days – months)

Thank You :)

Questions?