Memory-Bound Pursuit

Integrating Memory into Cross-Situational Word Learning Models

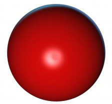
Christine Soh[1] and Charles Yang[1]

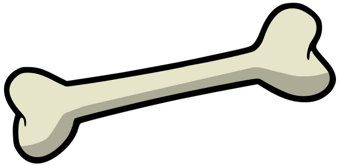
[1] University of Pennsylvania

CogSci 2021

Cross-situational word learning

- Early word learning: *referential ambiguity*
  - “Dog”



- Learning happens across co-occurrences of words and objects

(Quine 1960; Yu & Smith, 2007; Medina et al 2011; Trueswell et al 2013; i.a.)

Big questions

- How do we account for:
  - Development?
  - Individual differences?
  - Homophone-learning?
  - Effects of different orders of presentation (massed vs. interleaved)?

(Vlach & Debrock, 2017; Adams & Gathercole, 2000; Vlach, 2019; Wojcik, 2013; Benitez et al 2010; i.a.)

The proposal: Integrating memory

- Gap in cross-situational word learning models
  - Word learning involves many cognitive processes: memory, attention, language learning
  - Current word-learning models focus on disambiguation and encoding of new words after a brief exposure
- Mechanism for memory to complement word learning models
- Systematically evaluate performance across a wide range of behavioral studies

(Vlach & Debrock, 2017; Adams & Gathercole, 2000; Vlach, 2019; Wojcik, 2013; a.o)
Code can be found at: https://github.com/christinesoh/memory-bound-pursuit

The base: Pursuit (Stevens et al 2017)

- A more robust variant of Propose but Verify (Trueswell et al 2013)
  - “Pursues” the best hypothesis at any given time, but maintains disconfirmed hypotheses and counts of referential success
  - Lexicon created at end, based on a threshold parameter
  - Memory coded as recall parameter (=0.75)

Pursuit (Stevens et al 2017)

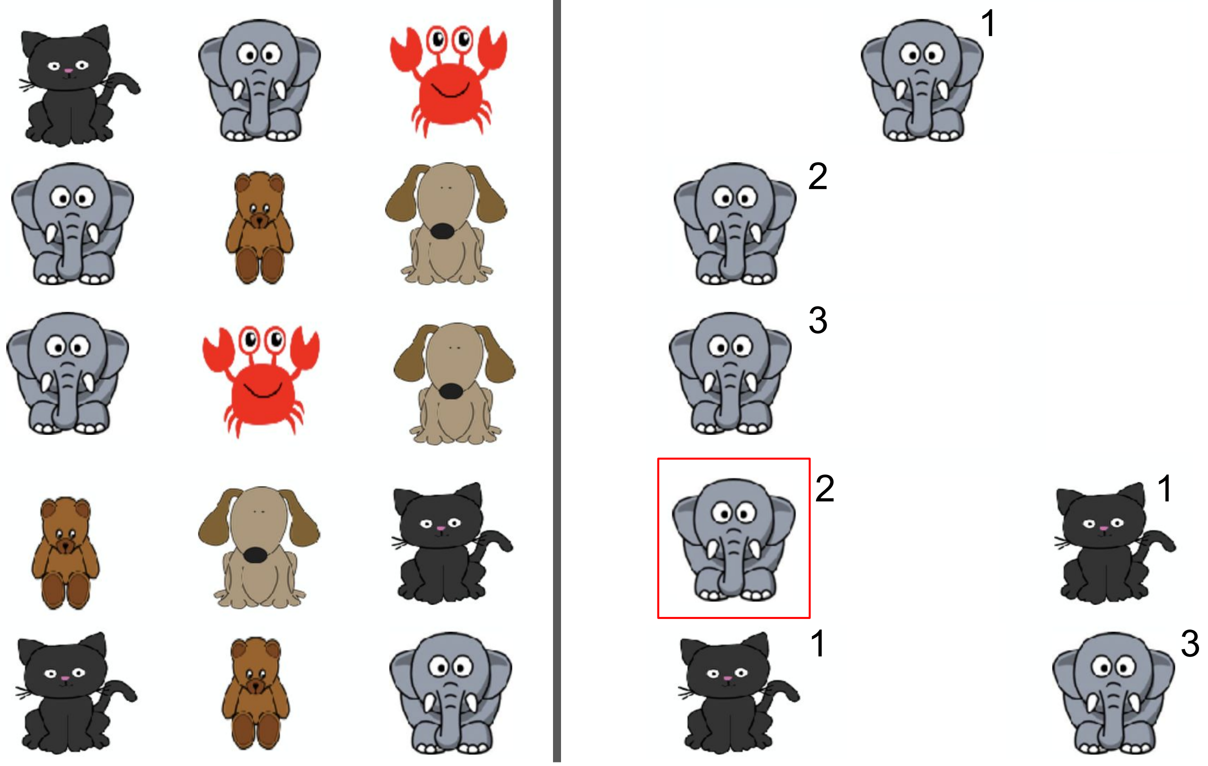
“Mipen”

“Mipen”

“Mipen”

“Mipen”

“Mipen”



Incorporating Memory

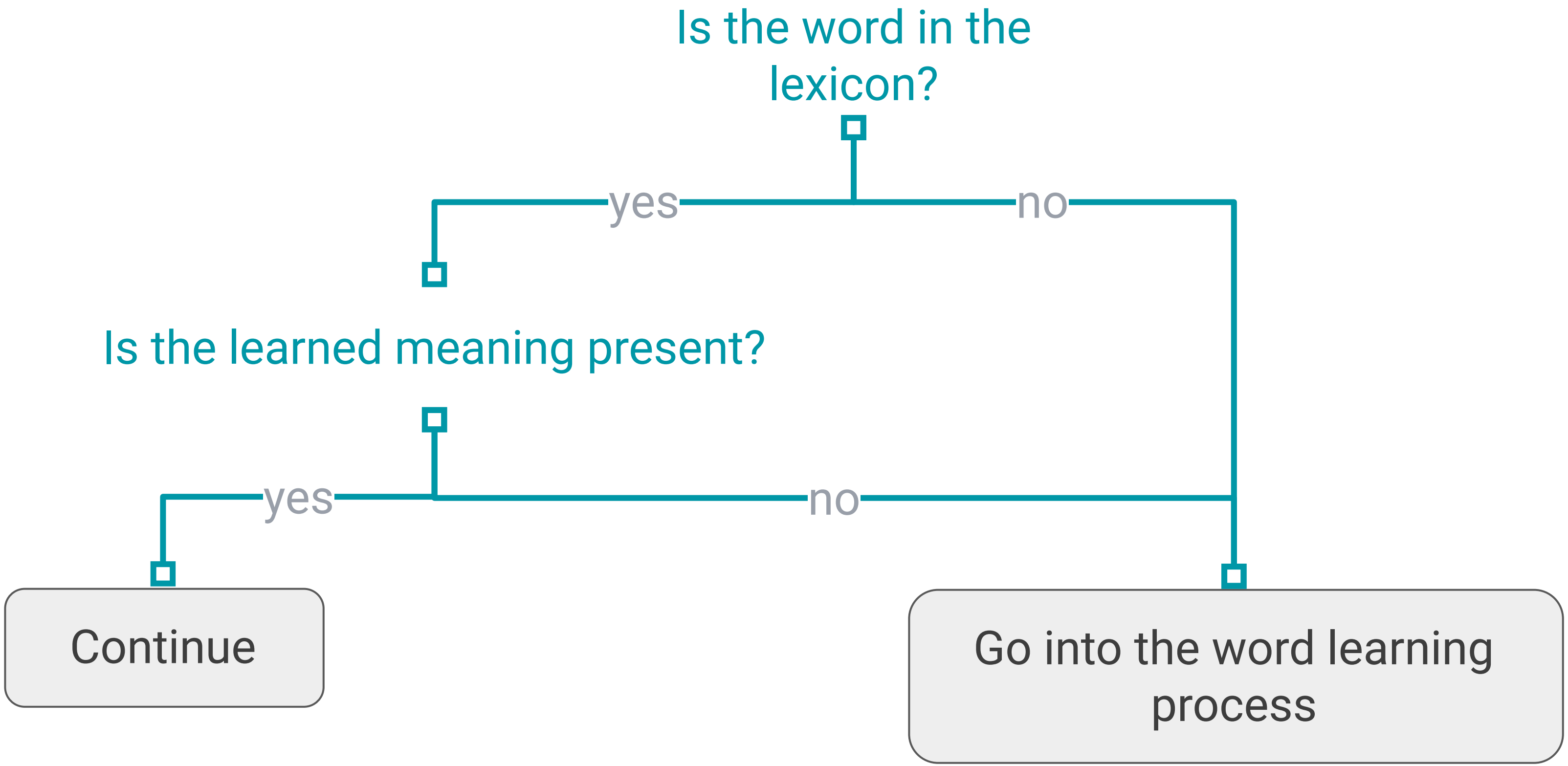
Atkinson-Shiffrin model (modal model): separate into components

- *Lexicon* : learned word-meaning pairs
- *Memory buffer* : set of words being learned in working memory, their hypothesized meanings, and corresponding association scores
  - Incorporating idea of “rehearsal”: updated with each encounter, moved to the top of the buffer
  - When full, forget a random word* and its hypothesized meanings
  - Simplest mechanism

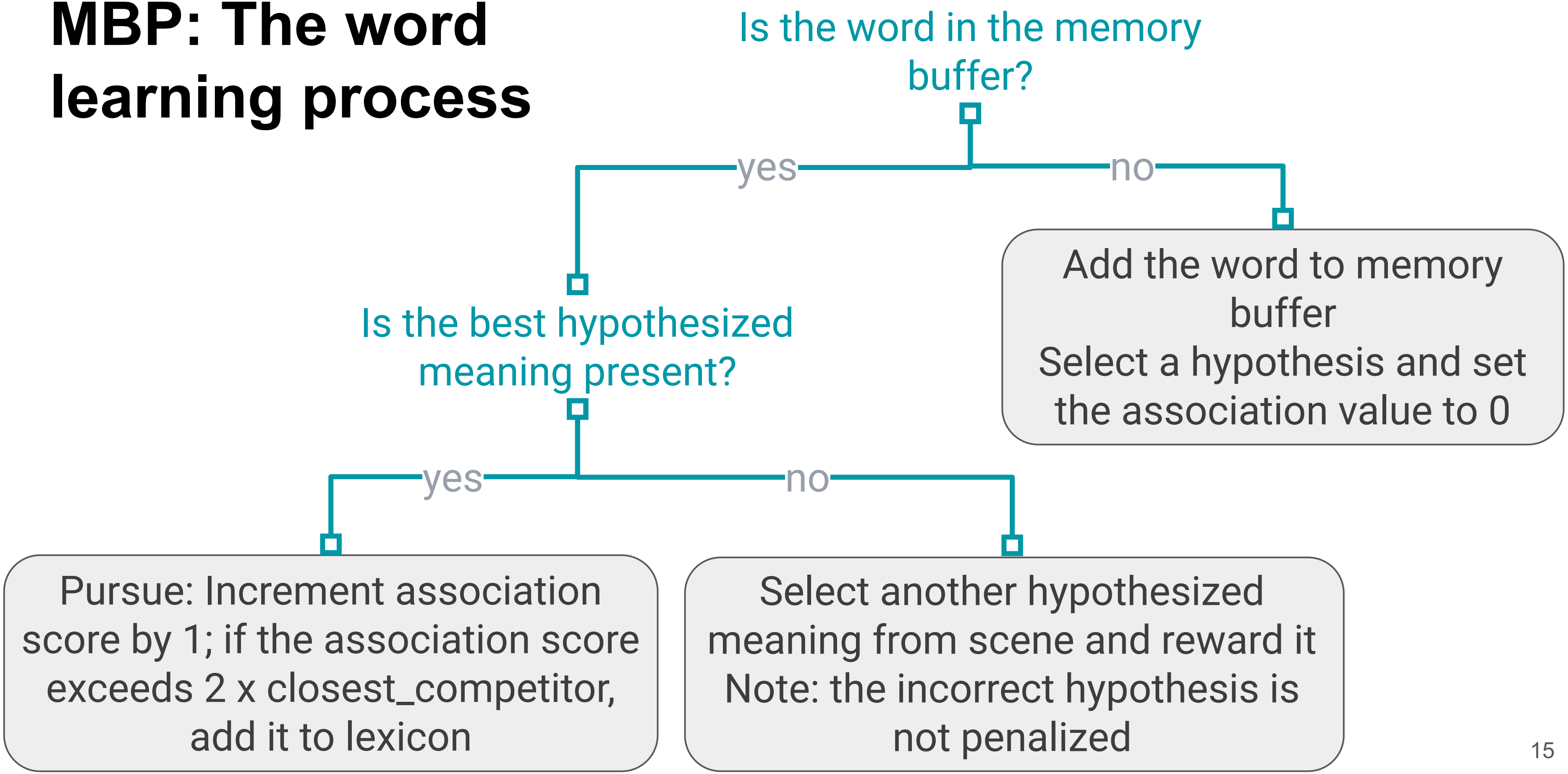
* The most recent development in the model

(Atkinson & Shriffin 1968; Ebbinghaus 1913; i.a.)

MBP: Checking the lexicon



MBP: The word learning process



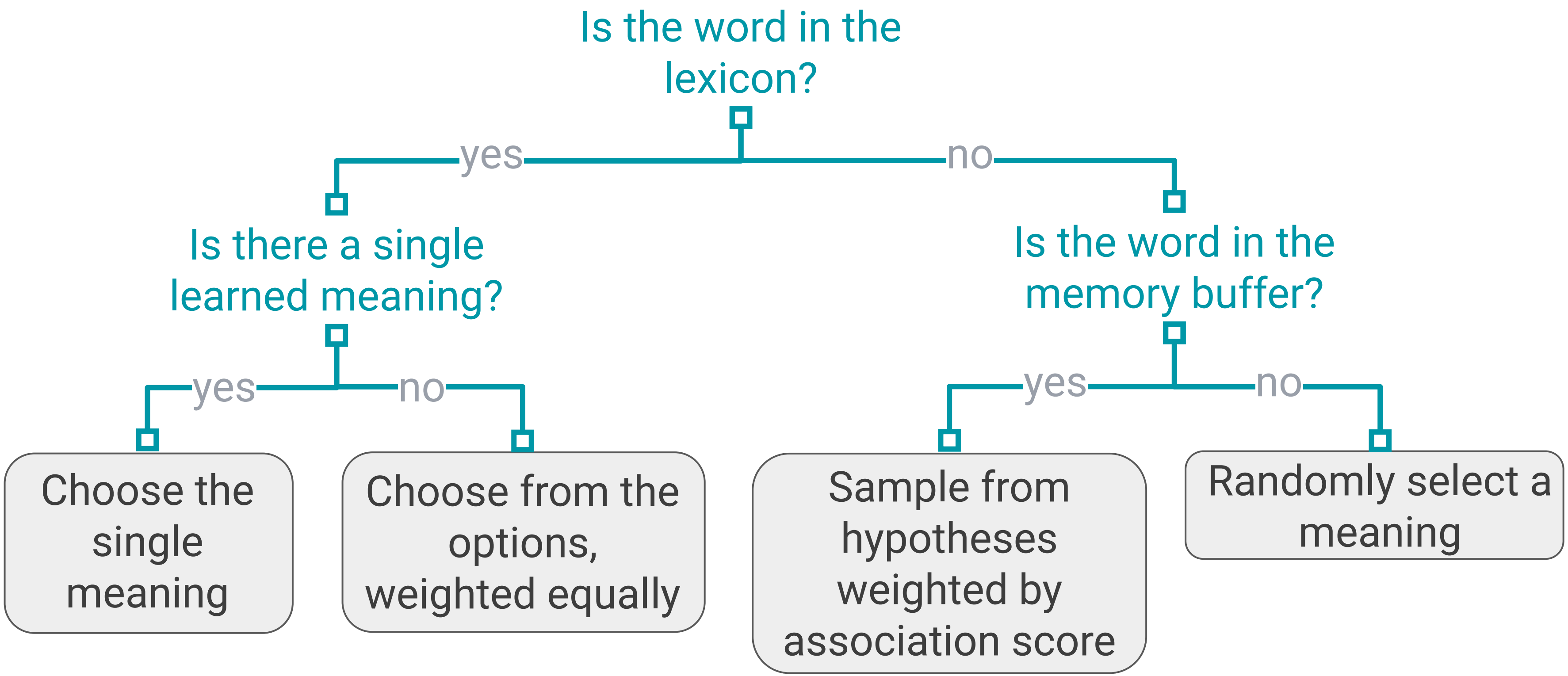
MBP

- Key differences from the original model:
    - Limited memory as an independently justified parameter
        - **Memory** parameter that constrains the size of the memory buffer
    - Threshold: in comparison to closest competitor rather than a tuned value
    - Lexicon is incrementally built rather than compiled at end
- Learns a hypothesized meaning for a word only if the word stays in the memory buffer long enough for it to dominate other competitor meanings

Memory-Bound Global model (MGX)

- Based on the global model of cross-situational learning (Fazly et al 2010)
- Co-occurrences tabulated across all learning instances, using reinforcement learning style probabilistic adjustment
- Limit the number of words the model can learn concurrently, builds lexicon incrementally
- Parameters tuned to the CHILDES Rollins corpus

Retrieval



Is the word in the lexicon?
yes
no
Is there a single learned meaning?
Is the word in the memory buffer?
yes
no
yes
no
Choose the single meaning
Choose from the options, weighted equally
Sample from hypotheses weighted by association score
Randomly select a meaning

Testing Learning Models on Experiments

- Variation introduced by the difference in memory development
  - Normal distribution with a standard deviation of 1
  - Mean memory size = 10
- Models have stochastic behavior, both from the sampling of the memory buffer size and from the hypothesis selection
  - Average across 300 runs (as in Stevens et al 2017)
- Run on Yu & Smith (2007), Trueswell et al (2013), Koehne et al (2013)

(Miller, 1956)

Takeaways

- Integrating memory...
    - Maintains performance of Pursuit
    - Improves global models: MGX captures more of the trends of the experimental data

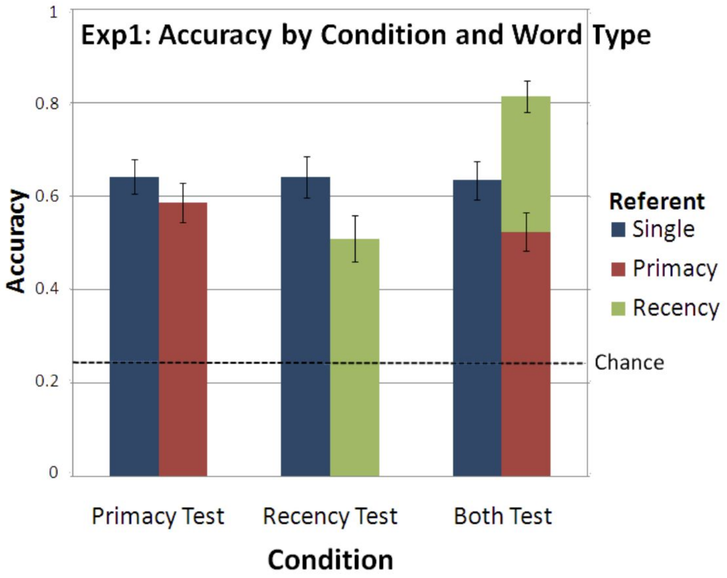
Yu & Yurovsky 2008: Homophone learning

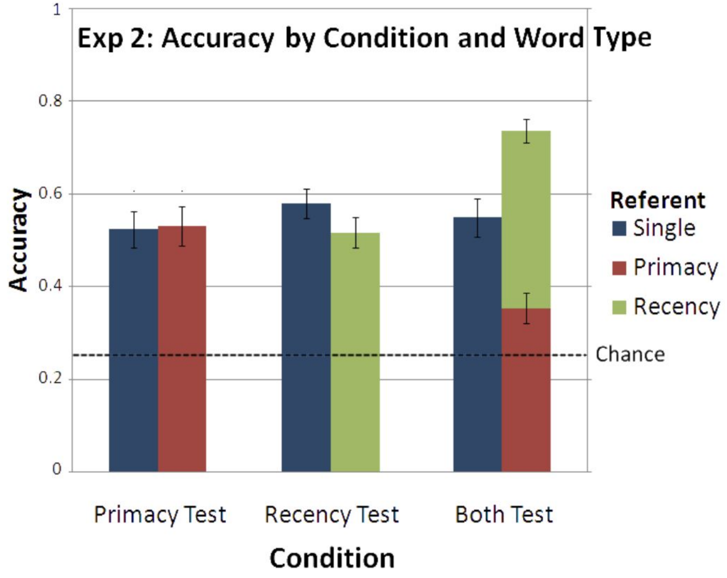
- Complex CSWL task: learning homophones and single meaning words
- Two conditions of the presentation of two meanings of homophones: massed vs. interleaved
- Takeaway results:
    - CSWL can facilitate the learning of homophones
    - Massed condition results in better learning of the first meaning over the second
    - Interleaved condition results in decreased performance, *even on single meaning words*

(Yu & Yurovsky 2008, Markman & Wachtel 1988; i.a.)

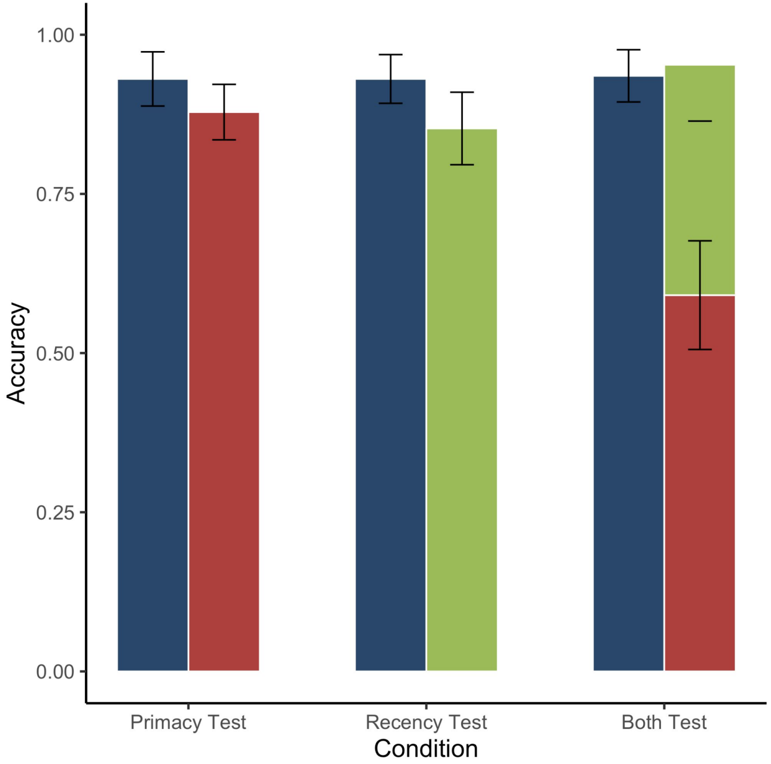
Yu & Yurovsky 2008: Homophone learning

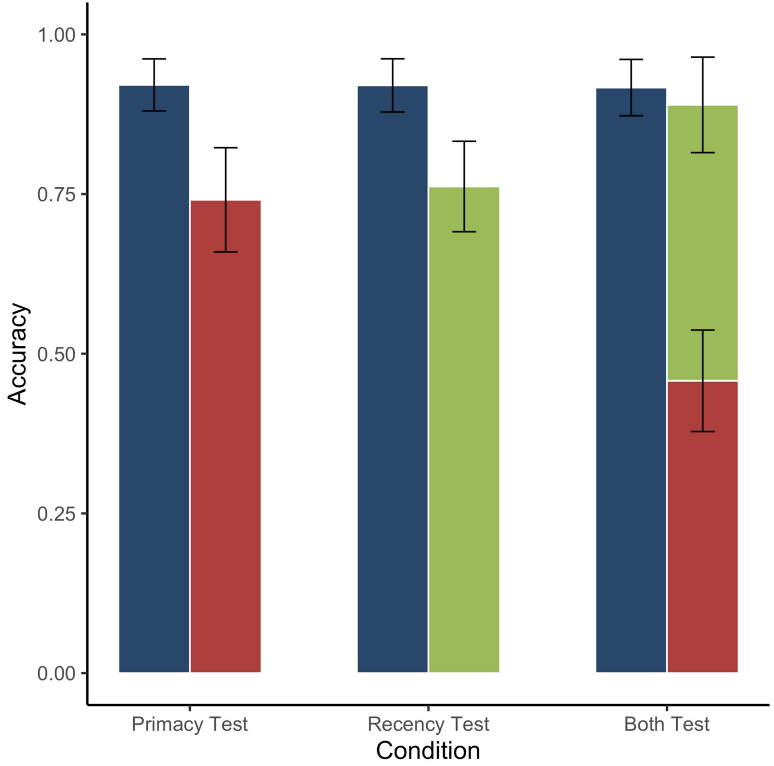
- Results (Exp1 = massed, Exp2 = interleaved)



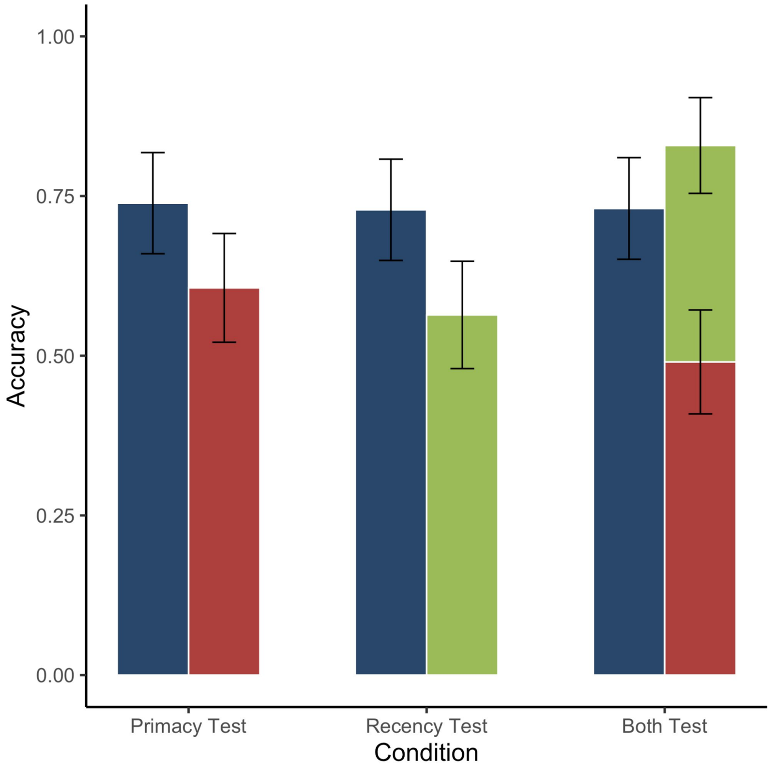


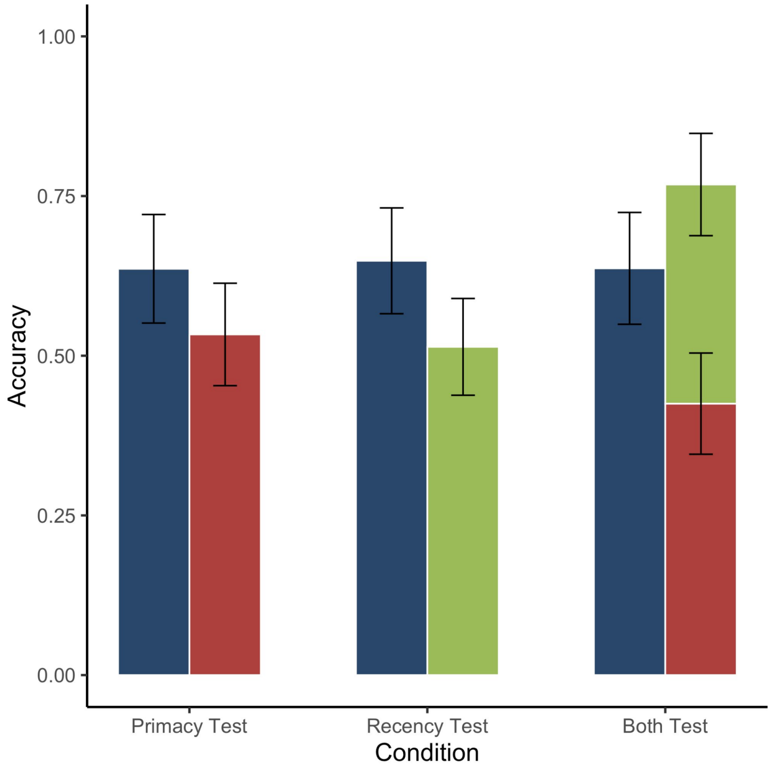
Yu & Yurovsky (2008): MGX





Yu & Yurovsky (2008): MBP





Results

- Support for the story that the decrease in performance results from increased memory load
- Memory constraint also allows for homophone learning / violation of mutual exclusion, though MBP outperforms MGX
- MBP provides a more cohesive account of learning in this complex task than Pursuit

Thank you