Memory as a computational constraint in cross-situational word learning

Christine Soh Yue (csohyue@sas.upenn.edu)
Department of Linguistics, 3401-C Walnut Street
Philadelphia, PA 19104 USA

Alexander LaTourrette (alatour@sas.upenn.edu)
Department of Psychology, 425 S. University Ave.
Philadelphia, PA 19104 USA

Charles Yang (charles.yang@ling.upenn.edu)
Departments of Linguistics, Computer Science, and Psychology, 3401-C Walnut Street
Philadelphia, PA 19104 USA

John Trueswell (trueswel@psych.upenn.edu)
Department of Psychology, 425 S. University Ave.
Philadelphia, PA 19104 USA

Abstract

A central challenge of cross-situational word learning is retaining word-referent mappings across exposures. We evaluate Memory-Bound Pursuit (MBP), a hypothesis-testing model of cross-situational word-learning which aims to account for learners' memory constraints via a single parameter targeting the number of words that can be learned concurrently. Here, we show that by varying this parameter with age, MBP can capture both children's and adults' cross-situational word-learning success under varying levels of ambiguity. We also present new experimental findings supporting novel predictions made by MBP about the retention of word-referent mappings across intervening exposures. These findings suggest that MBP provides a strong baseline model of cross-situational word learning, capturing both developmental trends and experimental evidence of memory limitations for word learning.

Keywords: memory; word learning; language acquisition; statistical learning; cross situational word learning

Introduction

Word learning presents a substantial challenge, with factors such as language abilities, attention, and memory affecting word learners' success. Research suggests a key component of word learning is resolving ambiguity across multiple utterances of a word: cross-situational word learning. Both adult and child learners track the co-occurrences of words and referential meanings across exposures to correctly infer word-object mappings (Yu & Smith, 2007; Trueswell et al., 2013).

However, a critical part of cross-situational word learning is the retention of previously formed semantic mappings (Vlach, 2019). While word-learning models sometimes address retention constraints by adding a parameter for forgetting mappings (Kachergis et al., 2012; Stevens et al., 2017; i.a.), most models focus primarily on the mechanisms of tracking and updating the word-meaning associations. Here, we evaluate a memory-limited model of cross-situational word learning[1] (Soh & Yang, 2021) that includes a motivated parameter capturing memory constraints on word learning: a limit on the number of words a learner can be learning concurrently in a local temporal context.

[1] https://github.com/csohyue/memory_bound_pursuit

Drawing on the modal model of memory (Atkinson & Shiffrin, 1968; Healy & McNamara, 1996), the Pursuit word learning model presented in Stevens et al. (2017) was amended to account for these memory constraints. More specifically, Soh and Yang (2021) implemented two distinct ways of storing hypothesized word meanings: the *memory buffer*, which stores a limited number of words (labels) and their hypothesized meanings, and the *lexicon*, where word-meaning pairs are permanently stored once sufficiently confirmed. Crucially, the memory buffer is finite and quite small: only a limited number of words can be learned concurrently. The lexicon, in contrast, is unbounded, containing all sufficiently confirmed semantic mappings. This model is dubbed Memory-Bound Pursuit (MBP). See Figure 1 for details on the learning algorithm. MBP learns by storing word-meaning hypotheses in the size-limited memory buffer and moving them to the lexicon only if the hypothesized meaning dominates its competitor meanings, a process described in more detail below. The model contains a single, psychologically motivated parameter: the size of memory buffer, with no additional tuned parameters. Here we present an overview of this model, test the model by simulating two studies previously conducted with children and adults, and finally test several novel predictions of the model by conducting a new cross-situational word-learning experiment with adults. Our findings show that MBP captures developmental differences in cross-situational word learning with a single parameter and successfully predicts adult performance in a new cross-situational word learning experiment.

Memory Bound Pursuit

Unlike previous cross-situational word learning models, MBP has only a single parameter: the buffer's size. While learners of all ages have large lexicons of established meanings, we suggest memory buffer size increases with age. Individual variation at each age is modeled by sampling from a normal distribution with a standard deviation of 1. This distribution is centered at 4 for 5- to 7-year-old children and

10 for adults, which is within the range of previous estimates of human processing capacity. These values cover a range of published findings (see Soh & Yang, 2021 for more details), including the studies simulated below. A key contribution of this model is that by changing only one parameter–the memory buffer size–the model captures cross-situational word learning performance across different age groups.

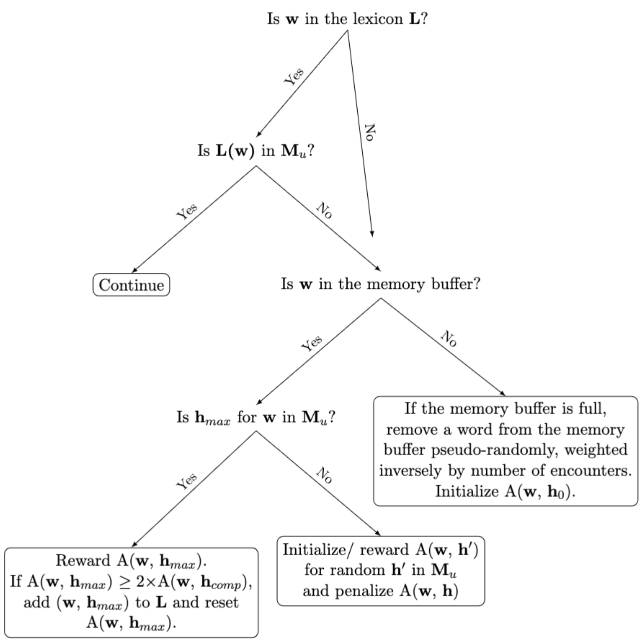

Figure 1: A decision tree for MBP in the process of updating the association strengths in the memory buffer. $\mathbf{w}$ is the word being learned, $\mathbf{L}$ the learned lexicon, $\mathbf{M}_u$ the set of referents present in the utterance, $\mathbf{h}_{max}$ the best hypothesis for a word, $\mathbf{h}_{comp}$ the next best hypothesis, $\mathbf{A}$ the set of association scores for words and their hypothesized meanings, and $\mathbf{h}_0$ is the initial hypothesis selected by a principle of mutual exclusion.

The basic schema of the model is shown in Figure 1. MBP is a hypothesis-testing model, so as it learns, the model considers only its single best hypothesized referent. The learning occurs only in the memory buffer, which is limited in size, and when a word's meaning is "successfully learned" (i.e. the best meaning exceeds twice the score of its closest competitor), the word-meaning pair enters the lexicon. Words and corresponding meanings are "forgotten" (i.e. removed) from the buffer when it reaches capacity. A word is learned if it can stay in memory long enough for a meaning to receive sufficient confirmation, relative to competitors.

Simulating prior experimental findings

First, we compare cross-situational word learning performance of MBP for adults and children by simulating two published studies (Yu & Smith, 2007; Suanda et al., 2014). Both studies asked participants to learn words across multiple exposures with varying levels of referential ambiguity. This training was followed immediately by a test, asking participants to select each word's target referent from among a set of other previously seen objects: thus, we assume that the content of the memory buffer is still accessible to the participant. In a test trial, MBP first checks its lexicon: if the word is in the lexicon and the learned referent is a possible option, it selects the learned referent. Next, the model checks its memory buffer: if the word is in the memory buffer, then it samples from the options weighted by the association value, following Luce's choice axiom. Finally, if the word in question is neither in the lexicon nor in the memory buffer, the model selects randomly from the options provided.

Because MBP randomly selects its hypotheses and the multiple choice selection has stochastic behavior, the model was run 300 times with accuracy averaged across the runs. We then compared the model's average performance to see if it lies within the human 95% confidence intervals.This value is stable across simulated sample sizes above 300 runs and offers a conservative measure of model performance.

Yu & Smith (2007)

In the first study that we simulate, adult participants were exposed to learning trials with 2, 3, or 4 nonce words and the corresponding number of referents (Yu & Smith, 2007). With increased ambiguity, participants' accuracy decreased. This experiment was run with adult participants; accordingly, MBP was set to have a mean buffer size of 10.

Experimental Results As expected, MBP captures the main result: increased ambiguity leads to decreased accuracy. Moreover, MBP also has overlapping confidence intervals with the reported results: see Table 2. This indicates the MBP model accurately captures human behavior.

Table 1: Yu and Smith (2007) experimental simulation results

	4x4	3x3	2x2
Reported	0.53	0.76	0.89
	(0.37-0.69)	(0.62–0.90)	(0.79–0.99)
MBP	**0.52**	**0.77**	**0.96**

Note. The bold indicates that the average performance lies within 95% confidence interval (CI) of the reported human results.

Suanda et al. (2014)

Next, we simulate a child study, which was conducted to examine the role of contextual ambiguity in children's word learning (Suanda, Mugwanya, & Namy, 2014). Children were asked to learn 8 nonce words, and like the previous experiment, there were three degrees of ambiguity. Here, ambiguity was captured by the contextual diversity, i.e. the number of different sets of stimuli that a word–object pairing co-occurred with across learning trials. The central finding was that increased contextual diversity (and thus decreased ambiguity across trials) improved cross-situational word learning.

Experimental Results MBP again captures the main result: decreased contextual diversity (and increased ambiguity) leads to lower accuracy (Table 1). By setting the average size of the memory buffer to 4 (rather than 10 for adults), MBP begins to approximate child learning behavior. It predicts a sizable difference in performance between children and adults here, which is not experimentally verified but is a likely adult baseline.

Table 2: Suanda et al. (2014) experimental simulation results

	HighCD	MedCD	LowCD
Reported	0.48	0.39	0.34
	(0.44-0.52)	(0.35-0.43)	(0.31-0.37)
MBP - 4	**0.52**	0.49	0.47
MBP - 7	0.83	0.79	0.75

Note. The bold indicates that average performance lies within the 95% CI of the reported human results.

Discussion

These results suggest that MBP is a compelling baseline model for word learning. While MBP does not completely capture the exact means child participants, the trends are captured. Moreover, the 95% CI of the model runs overlaps with human CI's, suggesting that the results are within the variation expected by MBP. While additional apparatus will no doubt improve the model's empirical coverage, its success in accounting for results across different age groups by altering a single, psychologically motivated parameter is remarkable. By changing only the size of learners' memory buffer, MBP can accurately predict children and adults' performance in multiple cross-situational word learning tasks, capturing the effects of referential ambiguity and of developmental differences.

Testing novel predictions

The Memory-Bound Pursuit model also makes unique novel predictions regarding the effect of confirmation on retention. The model separates the size-limited "memory buffer," where words and their meaning hypotheses are stored, from the size-unlimited "lexicon," where words with hypothesized meanings that dominate competitor meanings are stored. Items in the buffer may be removed as new items enter whereas items in the lexicon are retained through other learning exposures. Here, we conducted a new experiment testing this distinction by manipulating whether a particular word-referent mapping can enter the lexicon or must instead remain in the memory buffer. Critically, these manipulated learning exposures occurred either before or after an intervening "flush," a large set of learning exposures that were predicted to remove most mappings from the buffer but not the lexicon. We then compared performance for words stored in either the memory buffer or the lexicon, before and after the intervening flush.

Subjects

Adults (N=80; 40 per condition) were recruited on Prolific and participated online through PCIbex (Zehr & Schwarz, 2018). All were native speakers of English and provided informed consent. Participation lasted 10-20 minutes.

Stimuli

For visual stimuli, we chose four different but readily identifiable photographic images as referents for 90 common basic-level object labels, resulting in 360 images. Three of the images were used for learning, and one for testing. We had 18 nonce words, which followed English phonological rules.

Experimental Design

Subjects were exposed to 18 nonce words, each occurring in 3 trials, resulting in a total of 54 trials during learning. There were four blocks of learning trials, with the target words presented in Blocks 2 and 4 (see Table 3). In the first block, participants saw 10 trials, each featuring a different nonce word. This "warm-up" block was designed to saturate participants' memory buffer. In the second block, participants learned their first set of target words, with 3 exposures to each of 4 nonce words, all interleaved. The third block consisted of the intervening "flush," with exposures to the same words as in the warm-up phase, with 2 exposures per word, interleaved across words. Finally, participants learned their second set of target words in Block 4, receiving 3 exposures to each of 4 new nonce words, interleaved, just as in Block 2.

Table 3: Experimental design

Block		**# Nonce words**	**# Exposures**
1	Warm-up	10	1
2	**Target Pre-flush**	**4**	**3**
3	Intervening Flush	10	2
4	**Target Post-flush**	**4**	**3**
5	Testing	18	–

The visual display for each trial consisted of four referent images arranged in a rectangle and accompanied by a pre-recorded labeling utterance from a native English female speaker (e.g., "Look, it's a dax! Click on the dax!"). Over the course of the experiment, each nonce word appeared with exactly one object (the target) 3 times. Each nonce word appeared with another object 2 times, and the word was paired with every other object no more than once (see Figure 3).

At test, the blocks were tested in reverse order, with the post-flush target block first, ensuring that participants would still have access to any post-flush words encoded in the memory buffer. Next, participants were tested on words from the intervening flush block, and lastly, words from the pre-flush target block. Within blocks, the words were shown in the same order at learning and test. At test, participants selected from nine referents.

Conditions Participants were assigned to either the Same-First or Switch-First condition. The only difference between

(a) Same-First condition

(b) Switch-First condition

Figure 2: Sample trials in each condition. The green circles mark the target object, e.g. the book. The red squares mark the participant's selections. In the Same-First condition (a), the selected referent on Exposure 1 becomes the target and is present on all subsequent exposures. In the Switch-First condition (b), two of the three un-selected referent objects are present in Exposure 2 (the sandwich and the wrench). The selection at Exposure 2 determines the target object.

conditions was whether the target word trials in Blocks 2 and 4 afforded participants the opportunity to sufficiently confirm their hypothesized meanings–and, theoretically, move these words from the memory buffer into the lexicon.

In the Same-First condition, the referent object selected by the participant on a word's first exposure became the target object and was therefore present and available for confirmation on the second and third exposures–providing the opportunity for it to move into the lexicon. In the Switch-First condition, two of the three un-selected referent objects from the first exposure were shown in the second exposure, providing participants with a 50% chance of selecting an object present on the prior exposure. If either of these objects was selected, that object become the "target" on the third exposure and test. If neither was selected on Exposure 2, the "target" was arbitrarily assigned as one of those two objects. Thus, just as in the Same-First condition, the target was always present on all three exposures in the Switch-First condition, but now, it could be selected at most twice, ensuring it did not move into MBP's lexicon. In both groups, the 10 nonce words in the warm-up and intervening flush had a switch-first manipulation, with participants' first selection always being incorrect.

Predictions

Learning Phase Given that MBP is built with a base of the hypothesis-testing model Pursuit (Stevens et al., 2017), MBP predicts hypothesis-testing behavior at learning (cf. Trueswell et al., 2013). There are two key signatures of this hypothesis-testing behavior. First, when the target object has not been selected before, participants will select it at chance levels, even when it has previously co-occurred with the word. Second, when the target object has been selected on a prior trial, participants are very likely to select it again: assuming that the word is maintained in the memory buffer, the model will always choose the previously selected referent as its best hypothesis.

This enables MBP to make predictions about the distributions of token counts for the number of times the target object was selected during the learng phase for the different conditions. In the Switch-First condition, MBP predicts that approximately half of the words will have 2 target-referent selections (i.e., over 3 exposures, learners will select the target twice). This is because after the first exposure, where the target object is guaranteed to go unselected, participants have a 50% (2/4) chance of selecting a target object on the second exposure. If they select the target on that exposure, MBP predicts they will also do so on the third exposure, assuming the word is maintained in the memory buffer. If they fail to select a target on the second exposure, MBP predicts they will select the target at chance rates (1/4) on the third exposure. In contrast, in the Same-First condition, MBP predicts most tokens will have 3 target-referent selections. This is because the second exposure always contains the referent object that was selected at the first exposure. Thus, assuming that the word is maintained in the memory buffer, as is expected approximately 75% of the time by MBP, participants should select their previous hypothesized referent.

In contrast, a global learner should be more robust in recognizing previously unselected referents in subsequent exposures. Specifically, in the Switch-First condition after the first exposure, where the target object is guaranteed to go unselected, 2 of the 4 objects should have a boosted association score given their co-occurrence with the word in the first exposure. Participants should therefore select between these 2 objects, rather than selecting at chance as a local learner would. In addition, when the global learner does select one of them, they should also select that object in the third exposure, assuming that learners select their best hypotheses at learning (as they do at test). Thus, in contrast to MBP, the global model predicts that most words should have 2 target-referent selections in the Switch-First condition (i.e., over 3 exposures, learners will select the target twice).

Test Phase This design tests the effects of both the condition manipulation (Same-First vs. Switch-First) and whether learning occurred before or after the intervening "flush." MBP predicts both these main effects and their interaction will be significant (Figure 3).

For the analysis of the results, we exclude items for which the target referent was never selected in learning–it is not expected for learners to retain a mapping they showed no evidence of forming. Additionally, for the Same-First condition, only items with a sufficient number of hypothesis confirmations (i.e. 3 target-referent selections) are included, as these are the only items predicted by MBP to be in the lexicon.

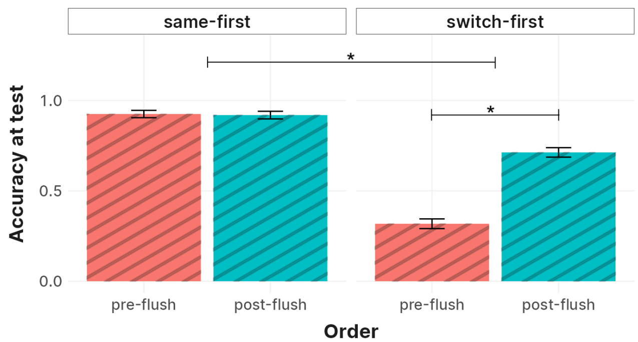

Figure 3: MBP-predicted performance on the experiment.

For the condition effect, we expect that words in the Same-First condition which are confirmed on every exposure should be sufficiently confirmed to enter the lexicon, yielding robust and highly accurate performance, while words in the Switch-First condition should remain in the memory buffer, yielding performance that is both more probabilistic and less robust to interference. The second main effect examines the impact of the intervening "flush." MBP predicts that a majority of the target words from the second block that were stored in the memory buffer will be removed during the flush, so we expect that performance on these words should be less accurate than words stored in the memory buffer but learned after the flush. Moreover, because the items in MBP's lexicon are more robust against the intervening flush, the difference between pre-flush and post-flush items in the Same-First condition is predicted to be smaller than the difference between the pre- and post-flush items in the Switch-First condition.

In contrast, the global learner predicts, at most, a significant effect of block. Given a strong enough memory decay, the items in the pre-flush condition will not be retained as well as those in the post-flush condition. However, the global learner predicts no effect of condition: the object counts co-occurring with a word in the two conditions are the same.

Results and Discussion

Here, we evaluate our empirical findings against the predictions generated by MBP. For the analysis of the results, like with model predictions, we exclude items for which the number of target referent selections did not reflect targeted behavior, as described above (i.e., Exposures 2 or 3). After this filtering, our analysis included 223 tokens for the Same-First condition and 222 tokens for the Switch-First condition.

Learning Phase

In the learning phase, participants showed hypothesis-testing behavior, as predicted by MBP. First, the number of target-referent selections are shown below in Table 4.

Recall that in the Same-First condition, MBP predicts most of the tokens will be selected 3 times. Human behavior is slightly more error-prone but largely in line with this prediction, with 70% of tokens having the target referent selected in all three exposures. In the Switch-First condition, MBP predicts that just under half of the tokens will have two target-referent selections (the maximum number in this condition), while global models predict most tokens will receive two target-referent selections. Human behavior reflects MBP's prediction, with 42% of tokens receiving two selections (see Table 4). This suggests MBP's memory-limited hypothesis testing approach provides a strong fit to the trial-by-trial learning patterns evident in adults' cross-situational word-learning.

Table 4: Target selections

	# Target selections:	0	1	2	3
Same-first	Pre-flush	–	17	45	98
	Post-flush	–	6	29	125
	MBP Pred.	–	*∼ 0*	*>0*	*<160*
	Global Pred.	–	*∼ 0*	*∼ 0*	*∼ 160*
Switch-first	Pre-flush	50	48	62	–
	Post-flush	48	40	72	–
	MBP Pred.	*∼60*	*>20*	*<80*	
	Global Pred.	*∼ 0*	*∼ 0*	*∼ 160*	–

Note. Human behavior supports MBP predictions of token counts of different number of target-referent selections.

Second, we compare differences in participants' behavior as a function of their previous selections. In the Switch-First condition, the first exposure selection was always incorrect, and in the Same-First condition, the first exposure selection was always correct. We examine behavior on the 3rd exposure based on whether the target object was previously selected. When the target object was not selected in exposures 1 and 2, participants are at chance (M = 24.6%), whereas when the target object was previously selected, participants selected the target object at rates far above chance (M = 70.53% for the Switch-First and 87.5% for the Same-First).

Test Phase

Figure 4 presents adults' accuracy in selecting the target referent at test. We constructed a logistic mixed-effects model predicting participants' selection of the target referent on each test trial, with fixed effects of Condition (Same-First vs. Switch-First) and Block (Pre-flush vs. Post-flush), as well as random effects of participant, item, and a Block-by-participant random slope. The model revealed a significant effect of Condition, $\beta = 3.26$, $SE = 0.71$, $p < 0.001$, with higher accuracy in the Same-First condition than the Switch-First condition. Thus, as predicted by MBP, participants' accuracy was significantly higher for items which they had confirmed three times, as opposed to only once or twice. We also observed a significant effect of Block, $\beta = 1.87$, $SE = 0.60$, $p = 0.002$, with higher accuracy for the post-flush block than the pre-flush block. Moreover, the interaction between block order and manipulation was significant, $\beta = 1.22$, $SE = .60$, $p = .041$, indicating that the effect of Block was greater in the Switch-First condition than in the Same-First condition. These results indicate that, as predicted, words which learners confirmed three times are relatively robust to interference from exposures to additional words, whereas words which learners confirmed only once or twice were more vulnerable to this interference. This pattern of performance, while not as pronounced as in the MBP simulation results (see Figure 3), is consistent with the model's key distinction between an interference-resistant lexicon storing highly confirmed word meanings and an interference-susceptible, size-limited memory buffer storing less certain word meanings.

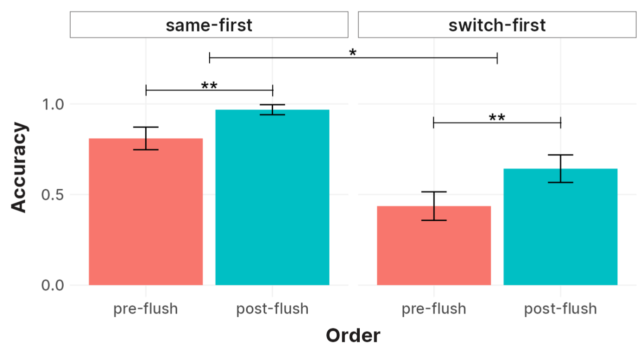

Figure 4: Adults' accuracy in selecting the target referent at test. As predicted by MBP, there are significant main effects and significant interactions. Error bars indicate standard error over tokens.

The 95% confidence intervals are shown in Figure 4. While the adults' exact means are not identical to the MBP predictions (Figure 3), the confidence intervals overlap for 3 of the 4 conditions, excluding the pre-flush portion of the Same-First condition in which MBP overestimates human performance.

Conclusion

A central challenge of cross-situational word learning is retaining word-referent mappings across exposures. Here, we evaluated Memory-Bound Pursuit (MBP), a hypothesis-testing model of cross-situational word-learning which aims to account for learners' memory constraints. First, we found that this model accounted well for both children's and adults' cross-situational word-learning success under varying levels of ambiguity in previous findings from Smith and Yu (2007) and Suanda et al. (2014). The model successfully captures both patterns of performance by changing the value of a single parameter: the average size of the memory buffer. The choice of a smaller memory buffer for children is, of course, in line with substantial work showing that children's memory capacity increases over development. This finding also suggests that child and adult word learning may share a similar underlying mechanism, despite quite disparate levels of absolute performance.

Our experimental results also support MBP's predictions. During learning, adults showed the expected hypothesis-testing behavior, with little indication they preferred previously co-present but un-selected meanings over meanings that had not previously co-occurred with the word. At test, learners were also more accurate in their selections when they had confirmed the meaning three times, sufficient for the meaning to enter MBP's lexicon, than when they confirmed it only once or twice, leaving it in MBP's memory buffer. More simplistic hypothesis-testing models (e.g., Trueswell et al., 2013) cannot account for this pattern, as they do not posit a role for additional confirmations of a hypothesis. Moreover, the word-referent mappings which MBP identified as stored in the lexicon were also less susceptible to interference from exposures to additional words (during the intervening flush) than mappings which MBP identified as stored in the memory buffer. This is consistent with MBP's distinction between the more permanent lexicon and the size-limited, interference-susceptible memory buffer. This pattern cannot be accounted for by models of cross-situational word-learning which do not incorporate memory limitations, including both local (Stevens et al., 2017) and global (e.g,. Fazly et al., 2010) models.

This result echoes recent calls to build memory processes into our models of word learning (Bhat et al., 2021; Kachergis et al., 2012; Vlach, 2019). Several of these memory-based models, such as those proposed by Bhat et al. (2021) and Kachergis et al. (2012), likely do not fully account for both the hypothesis-testing behavior during learning and the greater decline in learners' retention of Switch-First items at test, though future work is needed to directly test this question. Indeed, MBP also does not provide a perfect fit to the data: MBP overestimates participants' ability to retain word meanings stored in the lexicon across subsequent exposures to other words. While a modest reduction in performance after interference is compatible with prior memory research (e.g., Sosic-Vasic et al., 2018), it suggests that either some of these mappings did not truly enter the lexicon, despite 3 confirmations, or that more general interference factors affect accuracy for words stored in the lexicon. Future work might ask how factors like inter-trial interval, degree of ambiguity, and individual differences in memory affect this relationship between meaning confirmation and retention.

More broadly, we believe that MBP's success in accounting for both child and adult performance across a range of studies is especially notable given that it relies on only a single parameter, which varies predictably with age. By increasing the size of the memory buffer across age groups, MBP provides a straightforward, memory-driven account of developmental changes in cross-situational word-learning without postulating additional word-learning mechanisms or tuning parameters. In short, by integrating a hypothesis-testing approach with insights from the memory literature, MBP provides a strong and straightforward baseline model of cross-situational word-learning across development.

References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of learning and motivation* (Vol. 2, pp. 89–195). Elsevier.

Bhat, A. A., Spencer, J. P., & Samuelson, L. K. (2022). Word-object learning via visual exploration in space (wolves): A neural process model of cross-situational word learning. *Psychological Review*, *129*(4), 640.

Dautriche, I., & Chemla, E. (2014). Cross-situational word learning in the right situations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 892.

Ebbinghaus, H. (1913). Memory (H. R. . C. Bussenius, Trans.). *New York: Teachers College.(Original work published 1885)*, *39*.

Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science*, *34*(6), 1017–1063.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological science*, *20*(5), 578–585.

Healy, A. F., & McNamara, D. S. (1996). Verbal learning and memory: Does the modal model still work? *Annual review of psychology*, *47*(1), 143–172.

Holehouse, J., & Blythe, R. A. (2018). Cross-situational learning of large lexicons with finite memory. *arXiv preprint arXiv:1809.11047*.

Ibbotson, P., López, D. G., & McKane, A. J. (2018). Goldilocks forgetting in cross-situational learning. *Frontiers in psychology*, *9*, 1301.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2012). An associative model of adaptive inference for learning word–referent mappings. *Psychonomic bulletin & review*, *19*(2), 317–324.

Koehne, J., Trueswell, J. C., & Gleitman, L. R. (2013). Multiple proposal memory in observational word learning. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 35).

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive psychology*, *20*(2), 121–157.

Quine, W. V. O. (1960). *Word and object*. MIT press.

Sadeghi, S., Scheutz, M., & Krause, E. (2017). An embodied incremental bayesian model of cross-situational word learning. In *2017 joint ieee international conference on development and learning and epigenetic robotics (icdl-epirob)* (pp. 172–177).

Smith, K., Smith, A. D., & Blythe, R. A. (2011). Cross-situational learning: An experimental study of word-learning mechanisms. *Cognitive Science*, *35*(3), 480–498.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.

Soh, C., & Yang, C. (2021). Memory constraints on cross situational word learning. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 43).

Sosic-Vasic, Z., Hille, K., Kröner, J., Spitzer, M., & Kornmeier, J. (2018). When learning disturbs memory–temporal profile of retroactive interference of learning on memory formation. *Frontiers in psychology*, *9*, 82.

Stevens, J. S., Gleitman, L. R., Trueswell, J. C., & Yang, C. (2017). The pursuit of word meanings. *Cognitive science*, *41*, 638–676.

Suanda, S. H., Mugwanya, N., & Namy, L. L. (2014). Cross-situational statistical word learning in young children. *Journal of experimental child psychology*, *126*, 395–411.

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive psychology*, *66*(1), 126–156.

Vlach, H. A. (2019). Learning to remember words: Memory constraints as double-edged sword mechanisms of language development. *Child Development Perspectives*, *13*(3), 159–165.

Vlach, H. A., & DeBrock, C. A. (2017). Remember dax? relations between children's cross-situational word learning, memory, and language abilities. *Journal of memory and language*, *93*, 217–230.

Yang, C. (2020). How to make the most out of very little. *Topics in cognitive science*, *12*(1), 136–152.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological science*, *18*(5), 414–420.

Yurovsky, D., & Yu, C. (2008). Mutual exclusivity in cross-situational statistical learning. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 30).

Zehr, J., & Schwarz, F. (2018). PennController for internet based experiments (IBEX). *DOI: https://doi.org/10.17605/OSF. IO/MD832*.