# Actively Learning Object Names Across Ambiguous Situations

## George Kachergis, Chen Yu, Richard M. Shiffrin

*Department of Psychological & Brain Science / Cognitive Science Program, Indiana University*

## Abstract

Previous research shows that people can use the co-occurrence of words and objects in ambiguous situations (i.e., containing multiple words and objects) to learn word meanings during a brief passive training period (Yu & Smith, 2007). However, learners in the world are not completely passive but can affect how their environment is structured by moving their heads, eyes, and even objects. These actions can indicate attention to a language teacher, who may then be more likely to name the attended objects. Using a novel active learning paradigm in which learners choose which four objects they would like to see named on each successive trial, this study asks whether active learning is superior to passive learning in a cross-situational word learning context. Finding that learners perform better in active learning, we investigate the strategies and discover that most learners use immediate repetition to disambiguate pairings. Unexpectedly, we find that learners who repeat only one pair per trial—an easy way to infer this pair—perform worse than those who repeat multiple pairs per trial. Using a working memory extension to an associative model of word learning with uncertainty and familiarity biases, we investigate individual differences that correlate with these assorted strategies.

*Keywords:* Active learning; Language acquisition; Statistical learning; Cross-situational learning; Temporal contiguity; Individual differences

## 1. Introduction

Human infants learn words quite quickly despite many challenges facing them, including uncertainty and ambiguity in the language environment. Recent research has studied how learners may acquire word meanings from statistical regularities in the co-occurrence of words and referents (e.g., objects). Such cross-situational statistical word learning relies on two assumptions: (a) that spoken words are often relevant to the visible environment and

Correspondence should be sent to George Kachergis, Department of Psychological & Brain Science/Cognitive Science Program, Indiana University, Bloomington, IN 47405, USA. E-mail: gkacherg@indiana.edu

(b) that learners can to some extent remember the co-occurrence of multiple words and objects in a scene. Thus, as words and their intended referents are observed in different situations over time, learners can discover the correct word-object mappings. Relying only on the regularity of the linguistic environment and basic memory and attention processes, this may be an important method of learning nouns for infants, and even adult travelers.

In adult cross-situational learning studies (e.g., Yu & Smith, 2007), participants are asked to learn the referent of novel words by watching a series of training trials. On each trial, learners see an array of unfamiliar objects (e.g., four sculptures) and hear pseudowords (e.g., *stigson, bosa*). The referent of each pseudoword is ambiguous on a given trial, because although each word refers to a single onscreen object, the intended referent is not indicated. In a typical learning scenario, participants attempt to learn 18 word-object pairings from 27 trials, with four words and four objects given per trial. In this design, each word-referent pair is presented six times over the five-minute training period. Learning a correct word-object pairing requires accumulating word-object co-occurrences in some fashion. When tested on each word and given four trained objects to choose from, participants can choose the correct object for roughly half of the 18 words, on average (Yu & Smith, 2007).

However, in the real world even infant learners are not passive observers, merely watching the world go by. Rather, as learners shift their attention, their eyes, heads, and hands move, changing the objects in their view moment by moment. Indeed, recent evidence from studies with head cameras on infants suggests that infants' attention is selective: Even in a messy environment with lots of objects, only a few objects are in view at once (Smith, Yu, & Pereira, 2011; Yu & Smith, in press). If caregivers notice these attention shifts, they may be more likely to name objects that are currently being attended. Thus, learners may be able to increase the likelihood of hearing an object name by shifting their attention toward this object as a way to elicit the name from caregivers. This is a form of active learning, a concept studied extensively in machine learning (cf. Settles, 2009), in which a learner can query an information source for the labels of particular data points. That active learning plays an important role in statistical learning is also suggested by an otherwise similar experiment in which learning in the study phase was incidental (Kachergis, Yu, & Shiffrin, 2010). In this study, participants were either told to remember how many times each word and object co-occurred or were asked to do an oddball detection task during training. At test there was only slight evidence of incidental learning, suggesting that explicit attention may be critical to the success of statistical word learning.

In this study, we introduce active cross-situational word learning, in which learners choose which four objects they would like to see named on each successive trial. Thus, learners control when to repeat pairs, when to stop experiencing pairs they may feel they know, and when to attempt to learn more pairs. This gives us a glimpse of their preferred strategies. For example, participants may choose to repeat a single pair from the previous trial, and leverage working memory to quickly learn that the repeated word refers to the repeated object, perhaps while ignoring the other three word-object pairs on the trial. Equivalently, a learner may prefer to repeat three pairs from the previous trial and quickly learn the novel pairing that was not present. Kachergis, Yu, and Shiffrin (2009a)

manipulated this sort of temporal contiguity in a passive cross-situational learning study and found not only that repeated pairs are learned more easily but also are unrepeated pairs in conditions with some repeats. This suggests that simple inference supported by working memory is not the only learning mechanism at work.

In fact, investigating active learning can reveal what information and mechanisms a learner has at his or her disposal, and characterizing the observed strategies—and their performance—will motivate building more cognitively plausible learning models. For example, our recent associative model of cross-situational learning assumes that learners have access to both their familiarity and their uncertainty about the word-object pairings present on a given trial, and that attention competes for uncertain stimuli and for already-strong pairings (Kachergis, Yu, & Shiffrin, 2012). This model matches adult behavior in passive cross-situational experiments investigating mutual exclusivity, a bias to find 1-to-1 word-object mappings that is present even in 2.5-year olds (Markman & Wachtel, 1988). If active learners have access to their knowledge of pairing strength and stimulus uncertainty, then these cues can be combined to produce a few active learning strategies. One strategy is to choose one object you have never seen before (i.e., one with maximal uncertainty) and fill the remaining three slots on the trial with familiar objects. Alternatively, learners may choose novel combinations of familiar objects to disambiguate mappings; we have previously found such contextual diversity to aid learning (Kachergis, Yu, & Shiffrin, 2009b). Detailed analysis of active learning strategies can reveal what knowledge is available to learners and how they attempt to employ it to learn the correct mappings. It may even be that people are worse at actively structuring the learning environment than the randomly constructed passive trial sequences they normally experience in word-learning experiments.

In the experiment, participants did two blocks of passive cross-situational learning, as well as of two blocks of active cross-situational learning in which they chose the objects that they see named on each successive trial. Although there are many other possible formulations of active cross-situational learning, we chose this instantiation because it most closely matches the passive task, and it somewhat matches the real world, where learners can attend to objects and likely increase the chance of a teacher labeling those objects.

Active learning can change and improve performance by altering study time, the allocation of attention, and modes of processing. It is not clear how (and is probably not possible) to match passive and active conditions on all these factors. In this study, we allowed the participants in the active conditions to take the time they needed to make the choices for the next trial without trying to change the passive conditions to match. Both the changes in performance and the patterns of those changes are still useful in highlighting differences in active and passive statistical learning.

## 2. Experiment

Participants were asked to learn 18 word-referent pairs from a series of individually ambiguous training trials using the cross-situational word learning paradigm (Yu &

Smith, 2007). Each training trial was comprised of a display of four novel objects and four spoken pseudowords. With no indication of which word refers to which object, learners have little chance of guessing the four correct word-referent mappings from the 16 possible pairings. However, since words always appear on trials with their proper referents, the correct pairings may be learned over the series of trials.

The key manipulation of this study allows learners in active conditions to choose which four objects they want to see named on the next trial. In both conditions, 18 word-referent pairs were experienced over a series of 27 training trials. Importantly, the same pair was never allowed to appear in neighboring trials in passive conditions. In both conditions, each pair could only appear six times during the training session. Thus, both the number of exposures per pair and the ambiguity on each trial (i.e., number of pairs) were matched in active and passive learning conditions. To compare passive and active learning performance, each participant underwent two training and test blocks of each.

## 2.1. Subjects

Participants were 41 undergraduates at Indiana University who received course credit for participating. None had participated in other cross-situational experiments.

## 2.2. Stimuli

Each training trial consisted of an array of four uncommon objects (e.g., sculptures) and four spoken pseudowords. The 72 pseudowords generated by computer are phonotactically probable in English (e.g., "bosa") and were spoken by a monotone, synthetic female voice. These 72 objects and 72 words were randomly assigned to four sets of 18 word-object pairings, one set for each training condition.

Training for each condition consisted of 27 trials. Each training trial began with the appearance of four objects, which remained visible for the entire trial. After 2 s of initial silence, the four words were heard in a random order (1 s per word, with 2 s of silence after each) for a total duration of 14 s per trial.

## 2.3. Procedure

Participants were told that they would see a series of trials with four objects and four artificial words, but that the order of presentation of the words was random. They were also told that their knowledge of which words belong with which objects would be tested at the end. In the active learning conditions, participants were instructed that they would be able to choose four objects they wanted to see named next. In active learning training blocks, after each trial a display of all 18 objects in the to-be-learned set was shown, and participants chose four objects to be named on the next trial by clicking on them. Objects that had already been chosen six times were removed from the choice set.

After each training block, participants' knowledge of word-object mappings was assessed using 18-alternative forced choice (18AFC) testing: On each test trial, a single

word was played, and the participant was instructed to choose the appropriate object from a display of all 18 trained objects. Each of the 18 words was tested once in a random order.

Every participant did four blocks of training and testing: Half did two active learning blocks followed by two passive learning blocks, and the other half did the reverse.

## 2.4. Results and discussion

A repeated measures ANOVA on accuracy[1] by training type (active or passive) and training type repetition (1st or 2nd), nested by condition order (active-first or passive-first), revealed a significant main effect of training type ($F(1, 39) = 15.17$; $p < .001$). Test performance after active learning is far better than after passive learning (active $M = .59$; passive $M = .35$). Moreover, participants did not improve much on their second block of either training type: There was no significant effect of repetition ($F(1, 39) = 2.08$; $p = .15$). There was no significant interaction of condition order and repetition ($F(2, 38) = 1.62$; $p = .20$), nor of training type and repetition ($F < 1$), but there was a significant interaction of training type and condition order ($F(2, 38) = 4.53$; $p < .05$). As shown in Fig. 1, doing active learning first improves performance in the passive conditions (passive $M = .30$ if passive-first, $M = .39$ if active-first). To preview the discussion, active learning may allow learners to practice different information selection and rehearsal strategies, which in turn may help them selectively attend to a subset of word-referent pairs in the passive conditions. Individual performance after the different types of training was significantly correlated (Pearson's $r = .62$, $t(38) = 4.81$, $p < .001$). Fig. 2 shows that almost every participant performed at least as well and most often better after active training.
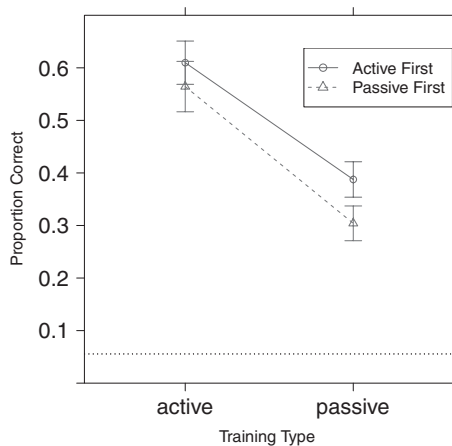


Fig. 1. Accuracy by type of first condition and training type in the experiment. Active learning resulted in far higher test performance than passive learning. Moreover, learners who did active learning first performed better in the passive learning conditions. Error bars show +/− SE, and the dotted line shows chance performance (.056 for 18AFC).
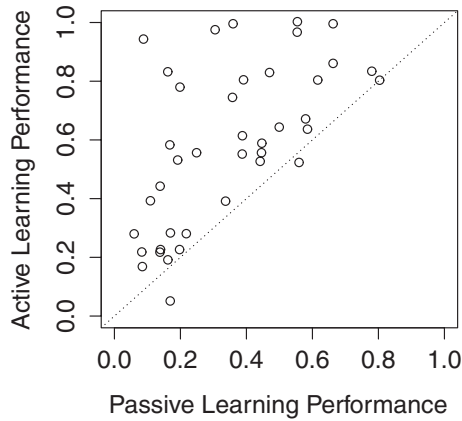
Fig. 2. Comparison of performance after passive versus active learning for each participant. Performance after the two types of training is correlated ($r = .62$), but learners are almost universally better after active training.
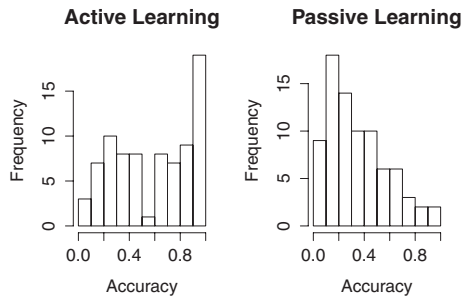


Fig. 3. Histograms of performance after active learning (left) and passive learning (right) training blocks (two per subject per condition). Accuracy after active learning is bimodal, indicating that some strategies are quite successful while others are mediocre.

   Given that adults can actively structure their environment to effectively learn the word referents, we next investigate the strategies effective learners use to disambiguate mappings. However, we must first consider that there are many strategies, and that not all of them result in swift learning. Performance in cross-situational word learning is typically highly variable, both within- and between-subjects. This is likely because what is learned on a given trial depends on what has been attended and learned on all previous trials (Yu & Smith, 2011; Yu, Zhong, & Fricker, 2012), and both the ambiguity on each trial and the fallibility of human memory means that people often learn different things. Giving learners an opportunity to structure statistical regularities in training may yield a more diverse set of learning states, and thus may increase variability in performance. Fig. 3 shows a histogram of learning performance after each block of active and passive learning. Although accuracy after passive training is unimodal and positively skewed, accuracy after active learning looks roughly bimodal, with peaks at .25 and at .95, which may

reflect strategies of differing utility. In the following analysis, we will examine strategy differences in two systematic (and complementary) ways: (a) performance-based, by doing a median split on the performance of active learners and analyzing the strategies used by each group; and (b) selection-based, by investigating what pairs selected trial-by-trial, clustering the active training trials into groups, and comparing the mean performance of each cluster.

One apparent active learning strategy involves the repetition of some pairs from one trial to the next. In our passive training conditions, no pairs were allowed to repeat in consecutive trials. Even if constructed randomly, a given passive trial would average only .22 pairs repeated from the previous trial. In comparison, active learners selected to repeat a mean of 1.5 pairs per trial, suggesting that learners use repetition to disambiguate pairs. To distinguish individual strategies, which may also vary over time, we carried out a clustering analysis on the structure of the training trials. A complementary way to measure trial-to-trial repetitions is to count how many unique pairs are seen within a two-trial window: A learner who repeats all four pairs from the previous trial sees only four unique pairs, whereas a learner who chooses four new pairs will experience eight pairs, as in the passive training sequence. More generally, we can measure how many unique pairs were contained in a window of $m$ trials, with a minimum of four and a maximum of $4m$ pairs. A window-size of five trials maximized the deviation between passive training and active training: The passive training sequence had a mean of 16.1 unique pairs (16.1/5 = 3.2 pairs per trial)—almost all 18—in the five-trial windows (range: 15–18), whereas active learners viewed a mean of only 11.5 (11.5/5 = 2.3 pairs per trial) unique pairs every five trials (range: 4–18). We found two groups of active training blocks by using partitioning around medoids to cluster the number of unique pairs in five-trial windows, estimated by the optimum average silhouette width (Kaufman & Rousseeuw, 1990). Cluster 1 contained 43 of the active training structures, and Cluster 2 contained the other 37. Fig. 4 shows the trial-by-trial mean number of unique pairs in five-trial windows for each cluster, and for the passive training sequence. Active learners in both clusters chose to view far fewer pairs in five-trial windows than they were shown in the random passive sequence. As training progressed, both clusters steadily viewed fewer unique pairs, but Cluster 1 overall chose fewer pairs than Cluster 2 (10.2 and 13.1 mean unique pairs, respectively).

It turns out that these strategy clusters—constructed solely from the active training sequences—result in different overall levels of performance: Cluster 1's mean of .71 is significantly higher than Cluster 2's mean of .45 (Welch $t(76.7) = 3.87$, $p < .001$). Corroborating this clustering result, a median ($Mdn = .61$) split on active learning performance identifies a similar grouping: Cluster 1 contained 28 better blocks, whereas Cluster 2 contained 26 of the worse blocks ($\chi^2 = 8.60$, $p < .01$). Overall, viewing fewer unique pairs is correlated with accuracy (Pearson $r = .33$, $t(78) = 3.11$, $p < .01$). Why does constraining the number of pairs one has viewed in the past several trials help learning? In addition to limiting working memory load, it is likely that active learners used trial-to-trial repetitions to aid learning. Indeed, on average, learners in Cluster 1 repeated 1.84 pairs per trial, significantly more than Cluster 2's mean of 1.07 repetitions (Welch $t(76.8) = 9.82$, $p < .001$). Fig. 5 shows a heatmap of how many pairs were repeated
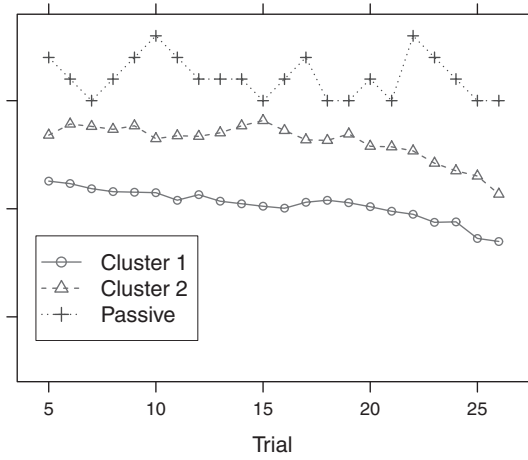
Fig. 4. The mean number of unique word-object pairs seen in five-trial windows across training by the two clusters of active learners and the number seen in passive training. Learners in Cluster 1 chose to see fewer unique pairs than learners in Cluster 2.
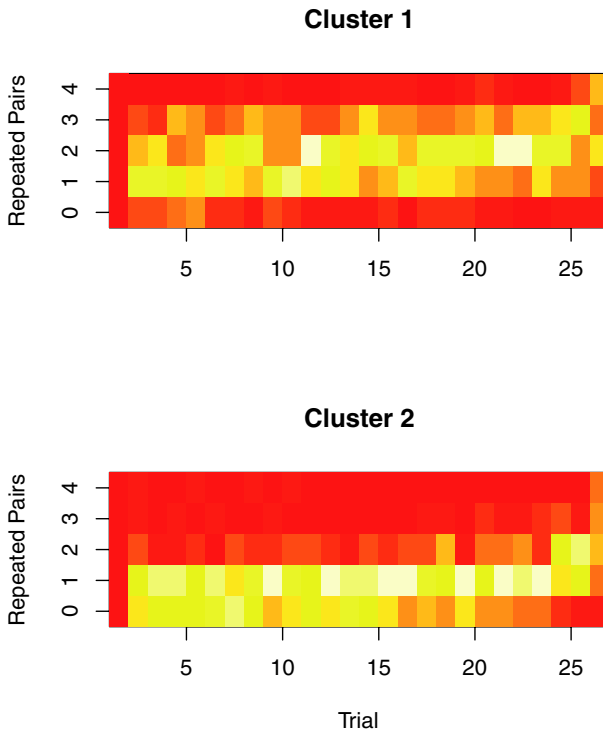


Fig. 5. The number of pairs active learners chose to repeat on each consecutive trial, accumulated for each of the two clusters (red = 0, white = 21). Learners in Cluster 2 most often repeated one—or even zero— pairs, while those in Cluster 1 chose from one to three repeats per trial.

trial-by-trial in each cluster. Learners in Cluster 2 often chose to repeat one or no pairs per trial until near the end.[2] Cluster 1 shows a much more varied approach, repeating anywhere from one to three pairs. Repeating more than one pair seems to be a good strategy—indeed, the mean number of pairs repeated per trial in active training is correlated with learning (Pearson $r = .28$, $t(78) = 2.61$, $p = .01$).

Why do repetitions of two or more pairs improve learning? Repeating three items highlights the new pair, and repeating one item highlights the old pair, when comparing successive trials. Yet repeating one pair introduces many more new items into memory and may thereby increase proactive interference, thereby explaining a benefit for more repetitions. We note that repeating two pairs also yields information—only 8 associations are reasonable using repetition information, instead of 16 on a normal trial—but two repetitions are particularly useful if a learner already knows one of the repeated pairs: This strategy allows learning of the unknown pair and practice of the known pair.

## 3. Model

To understand learners' apparent sensitivity to temporal contiguity in active training, we extend a recent associative model of cross-situational word learning with a working memory mechanism to see whether it explains the advantage of repeating one or more pairs per trial. The starting point is the associative model of cross-situational word learning proposed by Kachergis et al. (2012), which assumes that learners do not equally attend to all word-object pairings on a trial. The model incorporates a form of "mutual exclusivity": a bias to assume that a stimulus has only one referent and vice versa. It does so by assuming a competition between two effects of selective attention: strengthening associations between words and objects that have co-occurred previously, and strengthening associations between stimuli that have no strong associates (e.g., novel stimuli). These competing familiarity and uncertainty biases allow the model to exhibit fast mapping, since a novel word-novel object combination will demand more attention, and a novel word will only become weakly associated with an already-known referent (Kachergis et al., 2012). For example, suppose word $w_1$ and object $o_1$ have appeared together and are thus somewhat associated, while $w_7$ and $o_7$ are novel. Given a trial with both pairs: $\{w_1, o_1, w_7, o_7\}$, $w_1$–$o_1$ demands more attention than $w_7$–$o_1$, $w_1$–$o_7$, or $w_7$–$o_7$, since $w_1$–$o_1$ is stronger than baseline. However, attention is also pulled individually to $w_7$ and to $o_7$, since both of these novel stimuli have no strong associates. Uncertainty is measured by the entropy of each stimulus's association strengths. Because of the high joint uncertainty of $w_7$ and $o_7$, more attention is given to the association $w_7$–$o_7$. Thus, attention is mostly divided between $w_1$–$o_1$ and $w_7$–$o_7$, although the other pairings will be strengthened a bit.

Formally, let $M$ be an $n$ word $\times$ $n$ object association matrix that is incrementally built during training. Cell $M_{w,o}$ will be the strength of association between word $w$ and object $o$. Strengths are subject to forgetting (i.e., general decay) but are augmented by viewing the particular stimuli. Before the first trial, $M$ is empty. On each training trial $t$, a subset $S$ of $m$ word-object pairings appears. If new words and objects are seen, new rows and

columns are first added. The initial values for these new rows and columns are $k$, a small constant (here, 0.01).

Association strengths are allowed to decay, and on each new trial a fixed amount of associative weight, $\chi$, is distributed among the associations between words and objects, and added to the strengths. The rule used to distribute $\chi$ (i.e., attention) balances a bias for attending to unknown stimuli with a bias for strengthening already-strong associations. When a word and referent are repeated, extra attention (i.e., $\chi$) is given to this pair—a bias for prior knowledge. Pairs of stimuli with no strong associates also attract attention, whereas pairings between uncertain objects and known words, or vice versa, draw little attention. To capture stimulus uncertainty, we allocate strength using entropy ($H$), a measure of uncertainty that is 0 when the outcome of a variable is certain (e.g., a word appears with one object and has never appeared with any other object), and maximal ($\log_2 n$) when all of the $n$ possible object (or word) associations are equally likely (e.g., when a stimulus has not been observed before, or if a stimulus were to appear with every other stimulus equally). In the model, on each trial the entropy of each word (and object) is calculated from the normalized row (column) vector of associations for that word (object), $p(M_{w,i})$, as follows:

$$H(M_{w,i}) = -\sum_{i=1}^{n} p(M_{w,i}) \cdot \log(p(M_{w,i}))$$

The update rule for allocating attention and adjusting strengths for the stimuli presented on a trial is:

$$M_{w,o} = \alpha M_{w,o} + \frac{\chi \cdot e^{\lambda \cdot (H(w)+H(o))} \cdot M_{w,o}}{\sum_{w \in S} \sum_{o \in S} e^{\lambda \cdot (H(w)+H(o))} \cdot M_{w,o}}$$

In this equation, $\alpha$ is a parameter governing forgetting, $\chi$ is the weight being distributed, and $\lambda$ is a scaling parameter governing differential weighting of uncertainty and prior knowledge (familiarity). As $\lambda$ increases, the weight of uncertainty (i.e., the exponentiated entropy term, which includes both the word's and object's association entropies) increases relative to familiarity. The denominator normalizes the numerator so that exactly $\chi$ associative weight is distributed among the potential associations on the trial. For stimuli not on a trial, only forgetting operates. After training, a learner is tested with each word and chooses an object from $n$ alternatives in proportion to the association strengths of each alternative to that word.

Using competing biases for familiar pairings and uncertain stimuli, this associative model learns on a trial-by-trial basis by distributing attention in a way that corresponds with both our intuitions about word learning and a number of empirical findings. However, although this model does exhibit training order effects, it has no working memory component that would confer additional benefit for successively repeated pairs. Thus, we augment the baseline model with a mechanism that segregates words and objects repeated from the last trial from unrepeated stimuli, and only strengthens associations within these

subsets. This working memory (WM) model will learn better than the baseline model whenever there are repetitions. Because of the 16 possible associations on the trial, it will not attend to the spurious ones between repeated stimuli and unrepeated stimuli: 6 in the case of one or three repeated pairs and 8 in the case of two repeated pairs. To estimate whether people are attending more to the repeated or unrepeated stimuli, we added an attention parameter $\beta$ to the WM model that apportions more weight to associations between repeated stimuli as $\beta$ approaches 1, and more weight to unrepeated pairs as $\beta$ approaches 0. When $\beta = .5$, the attention given to repeated versus unrepeated associations is proportional to the size of each subset.

Using maximum log-likelihood as a measure of goodness-of-fit, the 18 test trials for each of the 80 active training orders were used to fit three parameters ($\chi$, $\alpha$, and $\lambda$) for the baseline model and four parameters ($\chi$, $\alpha$, $\lambda$, and $\beta$) for the WM model.

### 3.1. Results and discussion

Overall, both models achieved quite good fits to the data, with $R^2 = .98$ for both models on predicted and actual block accuracy, shown in Fig. 6. The WM model's BIC was 4404.6 and the baseline model's BIC was 4481.9, so the WM model is preferred despite the additional parameter.

Given the large number of repetitions used by active learners, it is surprising that the baseline model can approach the fit of the WM model without explicit awareness of repetitions. This may indicate that individual differences contribute much of the variability. Using the strategy clusters recovered from the active training sequences, we investigate whether there are systematic differences in the estimated parameters for these groups. Shown in Table 1, only the learning rate $\chi$ was significantly higher for Cluster 1—the high-accuracy, greater repetition group—than for Cluster 2 for both the baseline model (Welch $t(70.0) = 4.12$, $p < .001$) and the WM model (Welch $t(76.9) = 3.87$, $p < .001$). That is, the model suggests that the learners in Cluster 1, who use more concurrent
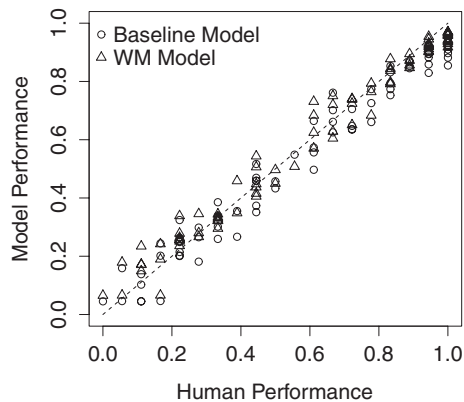


Fig. 6. Human versus model performance on active training blocks. Both the baseline model and the WM model were able to match human performance quite well, but the WM model is preferred by BIC.

Table 1
Mean best-fitting model parameters by cluster

| Param | Baseline Model | | | WM Model | | | |
|---|---|---|---|---|---|---|---|
| | $\chi^*$ | $\alpha$ | $\lambda$ | $\chi^*$ | $\alpha$ | $\lambda$ | $\beta$ |
| Cluster 1 | 18.9 | .87 | 4.4 | 21.0 | .91 | 6.8 | .59 |
| Cluster 2 | 5.0 | .90 | 3.4 | 7.5 | .90 | 5.9 | .59 |

*Notes. * indicates a significant parameter difference between clusters*

repetitions and generally learn more, may be faster learners than those in Cluster 2, but that they do not systematically differ in memory fidelity nor in attentional biases toward uncertain versus familiar, or repeated versus unrepeated pairs. Overall, the parameter estimates for the WM model and the baseline models are not strikingly different: The addition of the WM mechanism does not seem to have significantly influenced any of the parameters.

Finally, since the WM model is preferred by BIC, we examine its attention parameter, $\beta$, that might help us understand the range of strategies. Do learners focus more on repeated ($\beta \approx 1$) pairs, unrepeated pairs ($\beta \approx 0$), or do they split attention ($\beta \approx .5$)? Fig. 7 shows the distribution of the estimated $\beta$ values: Many people focused more on learning the repeated pairs—even exclusively, but several attended only to unrepeated pairs, and a number split attention roughly equally. Once again, we see individual differences spanning the range of possibilities, although the peaks are of interest. Estimated $\beta$ values were significantly, but weakly correlated with accuracy (Pearson $r = .22$, $t(78) = 2.07$, $p < .05$). Thus, the WM model observed a wide range of attention strategies for repeated pairs, and more attention to repeated pairs roughly corresponds to better performance.

## 4. General discussion

Active learning can speed language acquisition if the learner can implement an appropriate strategy that is based on the information available to him or her within the confines of the human memory system. In the context of cross-situational word learning, we have shown that many adults can generate strategies that improve their overall learning. Indeed, people who did active learning first were better at passive learning, suggesting that some part of their active information selection strategy carried over—though not trial-to-trial repetitions, since there were none during passive training. In active training,
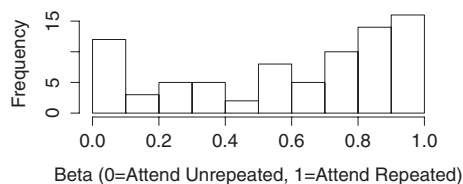


Fig. 7. Histogram of best-fitting $\beta$ values, showing a multimodal distribution peaked at 1—attend repeated pairs, and 0—attend unrepeated pairs, but with many other values.

learners preferred to use many repetitions, which kept the rate that new pairs were introduced far below the rate in passive training. This likely allowed participants to slowly learn new pairs, even as they practiced some already-known pairs. Active learning strategies were clustered into two groups of nearly equal size: Cluster 2 strategies typically repeated only one pair per trial—easily disambiguating that one, whereas Cluster 1 often repeated multiple pairs, and outperformed Cluster 2, on average.

Given that active learners used many repetitions, but with apparently diverse strategies and outcomes, we extended the associative word-learning model from Kachergis et al. (2012) with a working memory mechanism to attempt to see how people were leveraging repetitions. Overall, the model accounted for active learning accuracy very well, but parameters told of a plurality of strategies: Many people ignore unrepeated pairs while several only attend to these pairs, but the majority span the spectrum, attending to both repeated and unrepeated pairs. It may be that this focus shifts during a block, as knowledge develops. Overall, greater focus on repeated pairs was somewhat correlated with accuracy. Future work should also focus on predicting which pairs people will choose next, perhaps based on their current state of uncertainty or familiarity. Moreover, it may not always be obvious to the learner what active learning strategies will be beneficial, and thus strategic learning is likely an important factor. What will lead participants to good strategies is a domain needing more exploration, but we note one factor: Previous work has shown that for context repetition to benefit learning, the contexts must vary somewhat (Jones, Johns, & Recchia, 2012).

In summary, although active learners likely benefit from extra time and thought while choosing stimuli for the next trial, the model parameters suggest a more nuanced interpretation: The learning rates were higher for active learning only for Cluster 1, suggesting that the more critical factor helping active learning is the strategy for choosing and the number of items repeated. Infants may also benefit from such contiguity in the real world. We note that there is much autocorrelation in scenes and conversations: As the head is turned or the eyes shifted, many objects remain in view, and conversations drift over minutes. Moreover, we suggest that infants likely influence their learning environment in a way that is analogous to the active learning paradigm we present here. By choosing to look longer at some objects, they may increase the likelihood that a caregiver will label one of those objects. The structure of experiences that active learners choose, combined with their performance, promises to reveal the arsenal of cues and mechanisms that human learners have at their command. When coupled with a model, these data may also be able to identify individual differences. Active learning is clearly a powerful learning aid, and with better understanding it can likely be harnessed in education to speed the learning of language and other knowledge domains.

## Acknowledgments

## Notes

1. Data from one subject were excluded after it was found that their average performance in all four blocks was below chance (chance in an 18AFC test is .056).
2. Due to the constraint of each pair appearing only six times—as in passive training—there are only a few objects that remain to choose from, with the final trial being completely determined.

## References

Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology*, *66*, 121–132.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2009a). Temporal contiguity in cross-situational statistical learning. In N. Taatgen, H. Van Rijn, J. Nerbonne, & L. Schomaker (Eds.), *Proceedings of the 31st annual meeting of the cognitive science society* (pp. 1704–1709). Austin, TX: Cognitive Science Society.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2009b). Frequency and contextual diversity effects in cross-situational word learning. In N. Taatgen, H. Van Rijn, J. Nerbonne, & L. Schomaker (Eds.), *Proceedings of the 31st annual meeting of the cognitive science society* (pp. 755–760). Austin, TX: Cognitive Science Society.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2010). Cross-situational statistical learning: Implicit or intentional? In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd annual conference of the cognitive science society* (pp. 1189–1194). Austin, TX: Cognitive Science Society.

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.

Kaufman, L., & Rousseeuw, P.J. (1990). *Finding groups in data: An introduction to cluster analysis*. New York: Wiley.

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

Settles, B. (2009). Active learning literature survey. *Computer Sciences Technical Report 1648.*, Madison: University of Wisconsin-Madison.

Smith, L. B., Yu, C., & Pereira, A. F. (2011). Not your mother's view: The dynamics of toddler visual experience. *Developmental Science*, *14* (1), 9–17.

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

Yu, C., & Smith, L. B. (2011). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental Science*, *14* (2), 165–180.

Yu, C., & Smith, L.B. (in press). Embodied attention and word learning by toddlers. *Cognition*.

Yu, C., Zhong, Y., & Fricker, D. (2012). Selective attention in cross-situational statistical learning: Evidence from eye tracking. *Frontiers in Developmental Psychology*, *3*, 1–16, doi: 10.3389/fpsyg.2012.00148.