Cross-situational Word Learning Respects Mutual Exclusivity

Denise Ichinco, Michael C. Frank, & Rebecca Saxe
{**ichinco, mcfrank, saxe**}**@mit.edu**
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

Abstract

Learners are able to infer the meanings of words by observing the consistent statistical association between words and their referents, but the nature of the learning mechanisms underlying this process are unknown. We conducted an artificial cross-situational word learning experiment in which either words consistently appeared with multiple objects (extra object condition) or objects consistently appeared with multiple words (extra word condition). In both conditions, participants learned one-to-one ("mutually exclusive") word-object mappings. We tested whether a number of computational models of word learning learned mutually exclusive lexicons. Simple associative models learned mutually exclusive lexicons in at most one of the two conditions. In contrast, a more complex Bayesian model—which assumed that only some objects were being talked about and only some words referred—learned mutually exclusive lexicons in both conditions, consistent with the performance of human learners.

Keywords: mutual exclusivity; statistical learning; word learning; language acquisition

Introduction

Learning the meanings of words is hard. To begin learning a simple object noun like "dog" or "ball," a language learner must be able to identify what the word refers to. The word-world mapping problem presents a difficult learning challenge because there are an infinite number of possible referents available to be mapped in any given situation.[1] Each conversational situation, considered alone, is highly ambiguous.

In order to learn the meanings of words, learners often must aggregate information across multiple situations, and recent experiments suggest that human learners can succeed when this kind of computation is required. Yu et al. (2007) tested adult learners' ability to use cross-situational information in learning the meanings of object nouns. A single conversational "situation" was modeled by a single trial in which a number of different words and possible referents (represented by pictures of objects) were present. In three different conditions, Yu and colleagues presented sets of 2, 3, or 4 objects together with the corresponding words. Each situation gave only ambiguous evidence about which words and objects were paired. Despite this ambiguity, adult learners were able to learn the mappings between words and objects, and young children also showed evidence of success in a simplified version of the paradigm (Smith & Yu, 2008).

These experiments suggest that human learners are able to keep track of information about the co-occurrence of words and objects across situations, but the mechanisms underlying this ability are still unknown. Proposals about the computations carried out by human learners can be tested by instantiating these proposals as computational models, which can then be run on the stimuli from human experiments and evaluated on their fit to human performance. Unfortunately, although the pattern of results described in Yu & Smith (2007) is highly informative relative to the abilities of human learners, the task is simple enough that all existing computational models of word-world mapping are able to succeed. Thus, the existing data do not allow us to distinguish simple associative models from more sophisticated models of word learning (Yu & Ballard, 2007; Frank et al., in press).

The phenomenon of "mutual exclusivity" may provide a method for distinguishing models of word learning. Children prefer to map novel names to novel objects, rather than to familiar ones for which a name is already known (Markman, 1990). Although a variety of experiments have shown that even young infants are able to use mutual exclusivity to learn new words, there is no consensus on what mechanisms underlie mutual exclusivity inferences. Prior developmental explanations include both pragmatic accounts (Clark, 1987) and lexical constraints models (Markman, 1990). However, our previous computational work has shown that several models of cross-situational word learning—including the intentional word learning model proposed by Frank et al. (in press) and several simpler associative models—can succeed in conventional mutual exclusivity tasks without reference to either pragmatic principles or lexical constraints.

The success of these models in mutual exclusivity tasks suggests that the mechanisms of statistical learning may account for human learners' inferences in traditional mutual exclusivity experiments. This hypothesis can be tested by exposing participants to ambiguous, cross-situational word learning tasks which contain possible one-to-many mappings. If participants still acquire only one-to-one mappings, this would provide support for the view that mutual exclusivity inferences are compatible with (and perhaps even driven by) mechanisms of statistical learning. In addition, to the extent that models of cross-situational word learning make different predictions in a more probabilistic mutual exclusivity task, this may allow us to differentiate models on their fit to human performance.

Recent evidence suggests that word learning can occur in cross-situationally ambiguous paradigms even when not all mappings are one-to-one. Yurovsky & Yu (2008) trained participants using materials in which a single word was consistently associated with multiple objects. They found that

[1] Here we consider only the problem of mapping a word to a single referent and leave aside the problem of how to infer the full set of referents for a word from a limited set of examples (Xu & Tenenbaum, 2007).

Figure 1: An example trial in the 3×3 condition. Objects appeared on the screen simultaneously and the participant heard the words one after another in random order.

participants sometimes chose the first object and sometimes the second, indicating that they were able to succeed even in non-one-to-one learning situations. However, because their experiment averaged across words and participants and may have only tested the associations between any given word and a single object, it was not possible to determine whether the lexicons participants learned were consistent with mutual exclusivity.

Here we report the results of experiments designed to address this issue. Experiment 1 replicated Yu & Smith (2007)'s experiments with a new stimulus set. Experiment 2 then used this stimulus set to create a cross-situational mutual exclusivity situation in which we were able to test whether individual participants learned multiple (non-mutually exclusive) mappings for individual items, both words and objects. We found that participants learned one-to-one mappings for both words and objects. We then conducted simulations to test whether existing models of word learning could account for our results in Experiment 2. We conclude that while a Bayesian word learning model may account for our results, the bi-directional mutual exclusivity results that we observed pose serious problems for simpler associative models of statistical word learning.

Experiment 1

The goal of our first experiment was to replicate the finding of cross-situational word learning by adults (Yu & Smith, 2007). To mimic the various degrees of ambiguity possible in real speech, each participant received exposure to conditions in which 2, 3, and 4 words and pictures were presented during each trial.

Methods

Participants Twenty-four MIT students and members of the surrounding community participated and received $5 in compensation.

Stimuli Possible referents were represented by line drawings of unreal but geometrically possible objects, as shown in Figure 1 and first used in Kanwisher et al. (1997). Pictures were distinct from one another but not easily nameable. The images were presented in a horizontal row and displayed only as long as the words were being played. Novel words obeyed the phonetic rules of English and varied in length from 1-3 syllables. They were generated using AT&T's

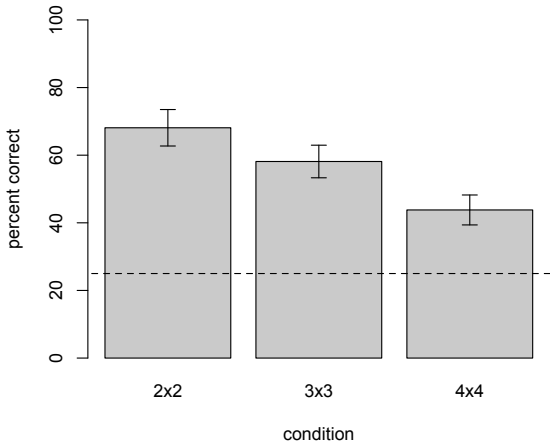

Figure 2: Percentage correct at test, plotted by training condition. Error bars indicate standard error of the mean. Participants learned more pairings between words and pictures when fewer words and pictures are presented in each trial. Because learning was tested with a four-alternative forced-choice test, chance was 25% correct, as indicated by the dotted line.

freely-available, web-based text-to-speech generator (voice: Christine). There were 54 words and pictures in total.

Procedure Each participant took part in three learning conditions, 2×2, 3×3 and 4×4, counterbalanced for order across subjects. Each condition consisted of a training and a testing phase.

The training phase was divided into trials in which a set of pictures was shown simultaneously and the corresponding words were played sequentially. No indications were given as to which word belonged with which picture, and the order of words with respect to pictures was randomized. Participants were told that each word they heard belonged with one of the pictures, but that they could not be sure which one. An example trial is shown in Figure 1.

A testing phase occurred immediately after each training phase. During each test trial, participants heard one word and saw four pictures and were asked to indicate the picture named by the word they heard. There was one trial in the testing phase for each word that was presented in the training phase. The target picture and the 3 foils were all drawn from the set of stimuli that were used in that condition.

Eighteen word/picture pairings were randomly selected for each condition from the total stimulus set, and no pairings were seen in two different conditions by the same subject. Within the training phase of each condition, each word was seen six times, so the number of trials and the duration of individual trials differed between conditions.

Results and Discussion

We were primarily interested in whether participants were able to learn correct word-object pairings from cross-

situationally ambiguous evidence. Condition means are shown in Figure 2. Participants who learned more than 9/18 pairings for any condition were significantly above chance for that condition, as defined by a chi-square test: $\chi^2(1, N = 18) = 6$, $p < .05$. Participants who learned 5/18 pairings were above chance; nearly all participants (24, 22, and 20 for the 2×2, 3×3, and 4×4 conditions) performed above chance for all conditions. Overall, the performance of the group was significantly above chance for each condition, as determined by a one-sample t-test comparing participants' correct answers to chance. For the 2×2 condition, $t(23) = 7.4$, $p < .001$; for the 3×3 condition, $t(23) = 6.4$, $p < .001$; for the 4×4 condition, $t(23) = 3.97$, $p < .001$.

A larger number of words and objects presented in a single trial caused poorer learning performance ($F(2, 23) = 13.07$, $p < .001$) due to the greater degree of ambiguity in which object mapped to which word. Our results directly replicate Yu & Smith (2007), who also found above chance performance in all conditions with a main effect of condition.[2]

Experiment 2

In Experiment 2, we build upon Experiment 1 by creating a situation in which two pairings are equally supported by the data but prior knowledge about an item in one of the pairings (a familiar word or picture) could allow a smart learner to disambiguate. Because several associative models of cross-situational word learning operate by estimating a uni-directional conditional probability measure, we included two conditions in our experiment: an "extra word" condition in which there was an opportunity to learn that a word mapped to two objects (shown in Figure 3), and an "extra picture" condition in which two words might map to a single picture.

We broke the experiment into two blocks. In the first block, we gave participants cross-situational experience with a set of words which were paired with pictures (henceforth called *old words* and *old pictures*). This block was identical to the 3×3 condition of Experiment 1. Then in the second block we added an extra item to each trial such that there were either an extra word with 3 pictures or 3 words with an extra picture. This extra item was chosen to co-occur perfectly with an *new picture* or *new word* (respectively) so that in order to learn the new pairing, participants would have to make a mutual exclusivity inference on the basis of their experience in the first block.

Previous experiments have attempted to test whether participants are able to learn in the presence of non-mutually exclusive stimuli (Yurovsky & Yu, 2008). In order to test whether the mappings participants learned were mutually exclusive, we included in our design a large number of test trial types to determine for each item and participant whether multiple mappings were learned.

[2]The means reported for the three conditions in the Yu & Smith (2007) study were significantly higher than the means we found. Although students and non-students show the same pattern, the student group in our study performed significantly better than the non-students, so we believe that differences in the sample population may account for the differences in means.

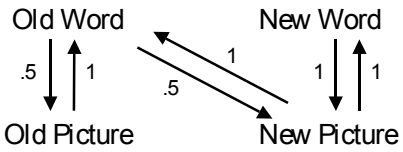

Figure 3: Co-occurrence frequencies in the "extra word" condition. An arrow from a to b should be read "if a has occurred, the probability that b has occurred is x."

Methods

Participants Thirty-two MIT students and members of the surrounding community participated in exchange for payment ($6). Fifteen participants were in the "extra picture" condition and 17 were in the "extra word" condition. None had also participated in Experiment 1.

Stimuli The stimulus set used in Experiment 2 was identical to that used in Experiment 1.

Procedure Participants were told that they had been kidnapped by aliens who were trying to teach them their language through two episodes of alien television. The experiment was divided into two blocks: an *initial exposure* block and a *mutual exclusivity* block. During the training phase of each block, a "television" was shown around the images to reinforce the idea of learning an alien language via a training video. Participants were not given any information about the relationship between the words they would hear and the pictures they would see—in particular, they were not told that each word belonged with only one picture.

The initial exposure block was identical to the 3×3 condition of Experiment 1. Participants were randomly assigned to either the "extra word" condition or the "extra picture" condition. The training phase of the mutual exclusivity block differed only in that an extra item (either an *old word* or an *old picture*, respectively) was shown during each trial. The extra item was chosen from the set of items in the initial exposure block for which the participant had given the most correct answers. The extra item always appeared in the same trials as a particular pairing of a *new word* and *new picture* and thus was perfectly correlated with that pairing within the second block. The co-occurrence statistics for the extra word condition are shown in Figure 3.

Each test trial was a forced choice in which participants were given a word and asked to choose between four pictures. Three trial types were used to ascertain, for each set of items, whether participants had learned a non-mutually exclusive set of pairings:

1. *Old word / old picture* trials tested whether the pairing learned in the initial exposure block was remembered.
2. *New word / new picture* trials tested whether the new

pairing was learned.

3. *Old word / new picture* (extra word condition) or *New word / old picture* (extra picture condition) tested whether a non-mutually exclusive pairing was made.

In addition to the three critical trial types, there were also two "preference" trial types. These trials tested each word that had multiple pairings, and both of the word's meanings were presented as options. These trials allowed participants to show (for those trials on which they learned both possible mappings) whether they preferred the old picture or new picture mapping. We also used these trials to confirm answers a subjects' answer on the corresponding critical trials: for example, we excluded a correct response for an old word / old picture trial when the same participant chose an *incorrect* response (neither the old picture nor the new picture) on that old word's preference trial.

Finally, control trials assessed subjects' ability to learn words that were paired with only one picture. These trials assessed mappings from both the initial exposure block and the mutual exclusivity block.

Each block had exactly 24 word/picture pairings and each pairing was shown 6 times. In the mutual exclusivity block, 8 words or pictures (depending on condition) that had been learned during the initial exposure block were chosen to be *old words* or *old pictures* and were paired with 8 *new words* and these old words were also shown exactly 6 times. There were 72 trials in the training phase of each block. The initial exposure block had 48 test trials, while the mutual exclusivity block contained 72 test trials: 24 critical trials (3 types $\times$ 8 words/objects), 16 preference trials (2 types $\times$ 8 words/objects), 16 control trials from the initial exposure block, and 16 control trials from the mutual exclusivity block. Test trials were presented in a random order. The experiment lasted approximately 30-40 minutes.

Results and Discussion

In our first analysis, we examined whether we replicated the results of Yu & Smith (2007) and Experiment 1. Results from the initial exposure block of Experiment 2 were comparable to the results of Experiment 1 in the 3×3 condition. On average, participants answered 56% of trials correctly. On control trials in the second (mutual exclusivity) block, participants gave correct answers on 51% of trials testing items from the first block and on 65% of trials testing items from the second block. Thus, the additional items in the mutual exclusivity block did not hinder learning, though performance could have improved in the second block due to practice and familiarity with the task, so we cannot conclude that the additional items were helpful.

In our second analysis, we looked at the critical trials for the mutual exclusivity block: the trials that tested which pairings each participants learned for each item. To perform this analysis, we tabulated for each participant and word which objects were paired with the word at test; we considered that a participant had learned a pairing only if they answered both the critical trial correctly and did not choose an incorrect

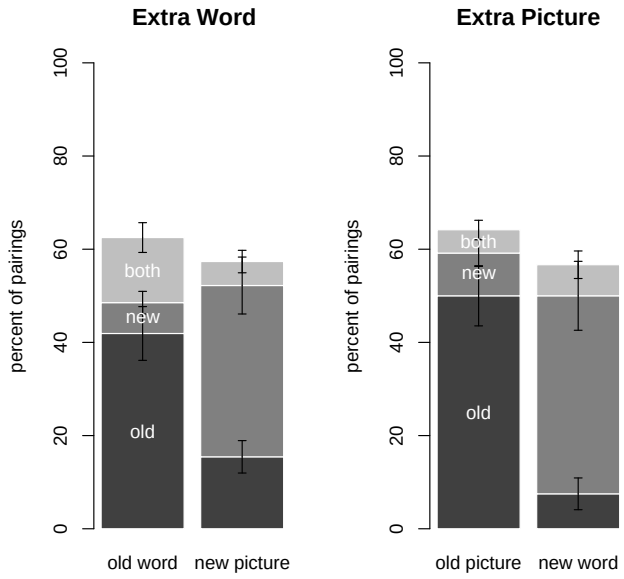

Figure 4: Results from the second, "mutual exclusivity" block of Experiment 2: the extra word condition (left) and the extra picture condition (right). For each condition, we show what percentage of mappings (with standard error of the mean) were made to each possible mapping target for both "new" items (introduced in the second block) and "old" items (introduced in the first block) that participated in multiple pairings. In the "extra word" condition, the old word and new picture participated in multiple pairings, and vice versa in the "extra picture" condition—see Figure 3.

answer on the corresponding preference trial. For example, for a participant in the extra word condition, for a particular *old word* learned in the initial exposure block, we tabulated whether the participant selected the corresponding *old picture* in the *old word / old picture* trial and whether they selected the corresponding *new picture* in the *old word / new picture* trial. The participant also needed to choose the old picture on the old word preference trial to be counted as having learned the old word-old object pairing; on the new word preference trial, the participant would be counted as having learned the correct pairing if they chose either the old picture or the new picture.

The results of this analysis are shown in Figure 4. Though participants did occasionally learn multiple pairings, they did this far less often than they mapped old words to old pictures and new words to new pictures. In general, we found that they were far more likely to learn one pairing for each word or picture than two pairings (all values of $t > 4.6$, all values of $p < .001$). Put another way, out of the set of 8 items in each condition, participants learned multiple pairings for an average of less than one item. For example, in the extra word condition, participants were more likely to remember

Model		Condition	
Name	Version	Extra Word	Extra Picture
Co-occurrence			
Cond. prob.	$p(w\|o)$	✓	
Cond. prob.	$p(o\|w)$		✓
Translation	$p(w\|o)$	✓	
Translation	$p(o\|w)$		✓
Mutual info.			
Bayesian model		✓	✓

Table 1: For each model in our comparison (and for each version of the model, where appropriate), whether the model learned a lexicon consistent with mutual exclusivity in the extra word and the extra picture conditions.

the pairing between the old word and the old picture than to learn the a pairing between the new word and the old picture; they were also more likely to learn the pairing between the new word and the new picture than the pairing between the new word and the old picture.

The number of exclusive pairings that were made for the old and new items did not differ reliably across conditions (for old word / old picture mappings in the extra word and extra picture conditions, $t(30) = -.93$, $p = .36$, and for the new word / new picture mappings, $t(30) = -1.63$, $p = .11$). More generally, the basic distribution of responses across conditions appeared highly similar.

In our final analysis, we asked whether, in the cases when participants selected multiple pairings, they had a consistent preference for one pairing over the other. For example, in the extra word condition, for an old word, a participant might answer correctly in both old word / old picture trials and old word / new picture trials, but have a consistent preference for the old picture when both pictures were presented side-by-side. To test this, we analyzed participants' responses in the preference test trials of the mutual exclusivity block. Even in the small number of cases that they chose two pairings for a single item, participants still had a consistent preference for the mutual exclusivity-consistent pairing, choosing the new word with new picture in the extra picture condition 73% of the time and choosing the old word with the old picture 68% of the time in the extra word condition.

To summarize our findings: participants learned equally well in Experiment 2 (which had data which might support non-mutually exclusive mappings) as they did in Experiment 1. The pairings that participants did learn were on the whole consistent with mutual exclusivity, and the pattern of performance was quite similar across the two conditions, suggesting that participants did not learn either many-to-one or one-to-many lexicons.

Computational Models

We next tested whether a range of computational models of word learning would capture the basic pattern of experimental results reported in the two conditions of Experiment 2. The set of comparison models for this section is the same as those reported in Frank et al. (in press): basic models based on co-occurrence, conditional probability models, and mutual information; the translation model of Yu & Ballard (2007); and the Bayesian intentional model proposed in Frank et al. (in press).

The models which relied on direct measures of association—including co-occurrence frequency, conditional probability, mutual information, and the translation model of Yu & Ballard (2007)—did not capture the basic result of Experiment 2: namely that participants learned mappings for ambiguous words which still respected mutual exclusivity (Table 1). The simple co-occurrence model learned non-mutually exclusive pairings because the co-occurrence of old words and new pictures was exactly equal to the co-occurrence of old words and old pictures in the extra word condition, and vice versa in the extra picture condition.

Models relying on unidirectional conditional probability (e.g., that compute either $p(w|o)$ or $p(o|w)$, where w is a word and o is an object picture)—including the translation model—succeeded in learning a mutually exclusive lexicon for one condition but failed for the other. For example, models that computed $p(w|o)$ learned a mutually-exclusive lexicon in the extra word but not extra picture conditions. Finally, the model based on the mutual information between words and objects ($MI(w,o) = \frac{p(w,o)}{p(w)p(o)}$) failed as well. To understand this finding, consider that, in the extra word condition, the mutual information of the old word and the old picture was the same as the mutual information of the old word and the new picture. No associative model that we evaluated was able to match human performance across both conditions.

In contrast, the Bayesian model of intentional word learning gave a higher posterior probability to the correct mutually exclusive lexicon than to the comparable non-mutually exclusive lexicon in both conditions. This model posits a *referential intention*—a particular object or set of objects that the speaker intends to talk about (implemented as a subset of the set of objects that are present in the context). The introduction of this hidden variable mediating between objects and words has two consequences. First, the model does not have to assume that all objects in the context are talked about. Second, the model assumes that some words refer to the objects in the referential intention, but that others are non-referential (e.g., function words or property terms); thus, not every word that is uttered must be mapped to an object.

We evaluated the Bayesian model by computing the posterior score (the non-normalized posterior probability) of possible lexicons on sample exposure sets from both the extra word and extra picture conditions. We evaluated three possible lexicons: (1) a lexicon which respected mutual exclusivity, mapping old words to old pictures and new words to new pictures (as participants in our experiment largely did), (2) a non-mutually exclusive lexicon corresponding to the full set of mappings that would be learned by a co-occurrence model

in the extra word condition, and (3) a corresponding non-mutually exclusive lexicon for the extra picture condition. We set the parameters of the model to the settings which resulted in learning the best lexicon in previous work and found that in both conditions, the model assigned higher posterior probability to the mutually exclusive lexicon.

To summarize: the associative models capture the basic human pattern of results in at most one condition of Experiment 2. In contrast, the Bayesian intentional model preferred the mutually exclusive lexicon in both conditions, consistent with human performance.

General Discussion

We investigated the mechanisms underlying cross-situational word learning via two experiments. The first experiment replicated the results of Yu & Smith (2007) and suggested that cross-situational learning in adults is a robust and general phenomenon. The second experiment set out to test whether mutual exclusivity inferences were made even in cross-situationally ambiguous contexts. In the first block, we taught participants one set of words; then in the second block we presented new words in situations in which they were confounded with words from the first block. We found that participants learned pairings that largely respected mutual exclusivity, de-confounding co-occurrence for the later words via the knowledge they learned in the earlier block. In addition, performance was similar when multiple words co-occurred with a single picture and when multiple pictures co-occurred with a single word.

What does this result tell us about the mechanisms of mutual exclusivity and cross-situational learning? To the extent that the performance of human participants in our experiment is due to the operation of learning mechanisms, our results can be used to test models of cross-situational learning. We evaluated a range of models of associative word learning on the stimuli from Experiment 2 and found that the pattern of performance from simple associative models based on co-occurrence and conditional probability failed to match human performance. However, the results from a more sophisticated intentional model were consistent with the performance of our participants. In addition, in a previous test of the model (Frank et al., in press), it was able to learn one-to-many or many-to-one maps when they were supported by the data, much as humans eventually come to learn polysemous, homonymous, or synonymous words.

Were the mutual exclusivity inferences that participants made in our experiment conscious or meta-cognitive inferences? As with the experiments reported by Yu & Smith (2007), some participants in our experiment did not believe that they had learned anything at all. On the other hand, it is possible that some participants did pursue a conscious strategy, either by making task-specific assumptions or by inferring that the mismatch between the number of words and the number of objects indicated that some words did not refer to objects or some objects did not refer to words. The latter kind of conscious inference is the same kind of inference that is licensed by the structure of the intentional word learning model, and it may well be explicit (at least in some cases), since it is not automatic or unavoidable (Halberda, 2006).

Human learners have the ability to acquire word-object mappings from ambiguous, cross-situational evidence. But the learning mechanisms underlying this ability should not be assumed to be simple on the basis of the simplicity of the experimental contexts in which this ability has been demonstrated. On the contrary, we hope that our current work provides some evidence that simple associative models of human learning are not sufficient to account for the rich variety of phenomena in children's early word learning.

Acknowledgments

We gratefully thank Nancy Kanwisher for providing the novel object stimulus set and the saxelab for helpful comments. MCF was supported by a Jacob Javits Graduate Fellowship.

References

Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition.* Hillsdale, NJ: Erlbaum.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (in press). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*.

Halberda, J. (2006). Is this a dax which I see before me? use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology*, *53*, 310–344.

Kanwisher, N., Woods, R., Iacoboni, M., & Mazziotta, J. (1997). A locus in human extrastriate cortex for visual shape analysis. *Journal of Cognitive Neuroscience*, *9*, 133–142.

Markman, E. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*, 57–77.

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

Xu, F., & Tenenbaum, J. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*, 245.

Yu, C., & Ballard, D. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, *70*, 2149–2165.

Yu, C., Smith, L., Klein, K., & Shiffrin, R. (2007). Hypothesis testing and associative learning in cross-situational word learning: Are they one and the same? *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*.

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

Yurovsky, D., & Yu, C. (2008). Mutual exclusiviy in cross-situational statisical learning. *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*, 715–720.