# Frequency and Contextual Diversity Effects in Cross-Situational Word Learning

**George Kachergis, Chen Yu, and Richard M. Shiffrin**
**{gkacherg, chenyu, shiffrin}@indiana.edu**
Department of Psychological & Brain Science / Cognitive Science Program
Bloomington, IN 47405 USA

## Abstract

Prior research has shown that people can use the co-occurrence statistics of words and referents in ambiguous situations to learn word meanings during a brief training period. The present studies investigate the effects of allowing some words and referents to appear more often than others, as is true in real learning environments. More frequent word-referent pairs are often—but not always—learned better, and also boost learning of other pairs. Superior learning for training sets with varying pair frequency may be a result of learning frequent pairs first, and using this knowledge to reduce ambiguity in later trials to learn other items. However, contextual diversity – the number of other pairs a given pair appears with – is naturally confounded with frequency, and presents an alternative explanation. The experiments in the present study systematically manipulate three critical factors in cross-situational learning – frequency, contextual diversity, and within-trial ambiguity – and measure their individual and combined effects on statistical word learning.

**Keywords:** statistical learning; language acquisition; cross-situational learning; contextual diversity; word frequency

## Introduction

Human infants learn word meanings with astonishing speed (e.g., Bloom, 2000). As in many other cognitive learning tasks, a key challenge in word learning is to deal with uncertainty and ambiguity in everyday environments. Recent research has focused on how regularities in the co-occurrence of words, objects, and events in the world can significantly reduce ambiguity across situations. This approach, dubbed statistical learning, relies on two assumptions: 1) that spoken words are often relevant to the visible environment, and 2) that learners can remember to some degree the co-occurrence of multiple words and objects in a scene. Thus, as the same words and objects are observed in different situations across time, learners can apprehend the correct word-object mappings. Both infants (Smith & Yu, 2008) and adults (Yu & Smith, 2007) have demonstrated cross-situational statistical word learning.

In the adult version of cross-situational learning studies reported in Yu & Smith (2007), participants were instructed to learn which word goes with which object and then studied a series of training trials. Each trial consisted of a display of several novel objects and pseudowords spoken in succession. Each pseudoword referred to a particular onscreen object, but the correct referent for each pseudoword was not indicated, thus making meanings ambiguous on individual trials. In a typical learning scenario, participants attempted to learn 18 pseudoword-object pairings from 27 12-second trials, with four pseudowords and four objects concurrently presented in a trial. This design allowed each stimulus (and hence each correct word-referent pairing) to be presented six times. In one form or another, the learning of a pairing involved the accumulation of pseudoword-object co-occurrence statistics across the training trials. Learning was assessed by testing pseudowords after the brief training, showing that participants acquired on average nine of the 18 pairs.

Although previous statistical learning studies (Klein, *et al.*, 2008; Yurovsky & Yu, 2008; Smith & Yu, 2008; Yu & Smith, 2007) convincingly demonstrated that human learners can efficiently learn word-referent pairs through statistical information alone, it remains unknown exactly which statistical regularities help (or hinder) learning. One intuition is that statistical learning will succeed better for more frequently appearing word-referent pairs, including the greater number of opportunities to store and learn them. However, a closer examination of the learning task suggests that manipulating frequency alters other factors, some of which may benefit low frequency pairs. These other factors typically involve learning, memory and inference. Consider several examples: Suppose one has learned via co-occurrence statistics several word-referent pairings: *A-a*, *B-b*, and *C-c*. Encountering a trial with four words (*A B C D*) and referents (*a b c d*) then allows the participant to infer that *D-d* is the correct pairing, even if this is the first occurrence of *D-d*. As another example, suppose the first two trials are (*A B C D*; *a b c d*) and (*A B C E*; *a b c e*). Memory for both trials would allow the participant to infer that *D-d* and *E-e* are correct pairings, though the mappings from (*A B C*) to (*a b c*) would remain ambiguous. Many other and more complex examples of this sort can be generated. Note that a pair *A-a* will be learned better if it appears in a set of trials with sufficiently diverse contents (i.e., *contexts*). If words (*A B*) and referents (*a b*) always occur together, then the correct pairings for these stimuli would remain ambiguous, regardless of the number of occurrences of these trials. Thus, a stimulus pair that appears with only a few other specific stimuli (i.e., has low contextual diversity) will be difficult to learn. Conversely, the more diverse the contexts in which a pair appears, the more likely is the acquisition of that pair. Motivated by these examples and by modeling efforts (Yu…2005; 2007; Frank, *et al.*, 2007), the present study focuses on three potentially influential factors: 1) **frequency**: repetitions per word-referent pair, 2) **within-trial ambiguity**: the number of co-occurring words and referents per trial; and 3) **contextual diversity**: the diversity of other pairs each pair appears with over time. The role of each individual factor in the context of cross-situational learning has not been systematically studied. Moreover, the potential interactions

among these factors, as illustrated in the above examples, remain unexplored.

Until a pair has appeared with all other pairs in the vocabulary, increasing within-trial ambiguity can yield greater contextual diversity. Will the toll of increased ambiguity outweigh the advantages of increased contextual diversity? Similarly, greater pair frequency can yield greater contextual diversity until that pair has been seen with all other pairs. Are repetitions solely crucial as learning opportunities, or as a means to increase contextual diversity? The current studies systematically investigate these three factors – both individually and in combination – and measure their effects on word learning. More specifically, Experiment 1 will focus on frequency alone, while Experiment 2 will explore contextual diversity and within-trial ambiguity. Experiment 3 will explore the interaction of contextual diversity and frequency. By manipulating the learning input and measuring the learning results, we can not only discover factors that are predictive of successful learning, but also shed light on the principles and constraints of the underlying learning mechanisms that operate on such learning input. Although these experiments investigate the learning of novel word-referent pairs, these findings may generalize to the learning of other types of associations.

## Experiment 1

Participants were asked to simultaneously learn many word-referent pairs from a series of individually ambiguous training trials using the cross-situational word learning paradigm (Yu & Smith, 2007). Each training trial is comprised of a display of four novel objects with four spoken pseudowords. With no indication of which word refers to which object, learners have a small chance of guessing the four correct word-referent pairings from the 16 possible ones. 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 is to repeat some pairs more often than others within the same set of trials. As discussed above, the more often a stimulus pair is repeated, the more opportunities there are to deduce and rehearse that pairing. In addition, more frequent pairs appear with more other pairs, and thus have greater contextual diversity. In light of this, we created two training conditions with subsets of pairs that appear with different frequency. In both conditions, training consisted of 27 training trials containing 18 word-referent pairs, four of which were displayed on each trial. In the **two frequency subsets** condition (Fig. 1, left), 9 of the stimulus pairs appeared 9 times, and 9 of the pairs appeared only 3 times. In the **three frequency subsets** condition, 6 pairs appeared 3 times, 6 pairs appeared 6 times, and 6 pairs appeared 9 times. A dramatic frequency effect was predicted: the more frequent pairings would be learned more often, and pairs with a mere 3 repetitions may not be learned at all. Importantly, the same pair was never allowed to appear in neighboring trials.
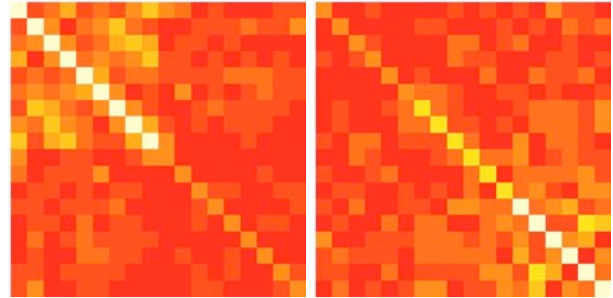


Figure 1: Word-referent co-occurrence matrices for the two learning conditions in Exp. 1. Each cell represents the co-occurring frequency of a specific word-referent pairs. The 18 correct pairs are on the diagonal. The other cells show spurious co-occurrences of incorrect word-referent pairs. Co-occurrences range from 0 (red) to 9 (white). **Left:** in the two frequency condition, 18 pairs form two frequency groups: 9 repetitions (the top 9 pairs) and 3 repetitions (the bottom 9). **Right:** in the three frequency condition, 18 pairs appear at three different frequencies: 3, 6, and 9 (the top, middle, and bottom 6 pairs, respectively).

### Subjects

Participants were 33 undergraduates at Indiana University who received course credit for participating. None had participated in other cross-situational experiments.

### Stimuli

Each training trial consisted of four uncommon objects (e.g., strange tools) concurrently shown while four pseudowords were spoken sequentially. The 36 pseudowords generated by computer are phonotactically-probable in English (e.g., "bosa"), and were spoken by a monotone, synthetic female voice. These 36 arbitrary objects and 36 words were randomly assigned to two sets of 18 word-object pairings, one set for each learning 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 seconds of initial silence, each word was heard (randomly ordered, duration of one second) followed by two additional seconds of silence, for a total duration of 14 seconds per trial.

After each training phase was completed, participants were tested for knowledge of word meanings. A single word was played on each test trial, and all 18 referents were displayed. Participants were instructed to click on the correct referent for the word. Each of the 18 words was presented once, and the test trials were randomly ordered.

### Procedure

Participants were informed that they would see a series of trials with four objects and four alien words. They were also told that their knowledge of which words belong with which objects would be tested at the end. After training, their knowledge 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. Condition order was counterbalanced.

## Results & Discussion

Fig. 2 displays the learning performance[1] for the subsets of pairs in both training conditions. In the condition with two frequency subsets, participants were significantly more likely to learn 9-repetition pairs ($M = .46$) than 3-repetition pairs ($M = .34$, paired $t(30) = 3.070$, $p < .001$), which agrees with our hypothesis of the frequency effect on statistical learning. However, this frequency advantage disappeared in the condition with three subsets of differing frequency: all three subsets were learned approximately equally well. Overall, participants learned almost the same number of pairings in the two conditions (two freq. subsets: $M = .40$, three freq. subsets: $M = .38$, paired $t(30) = .373$, $p > .05$).
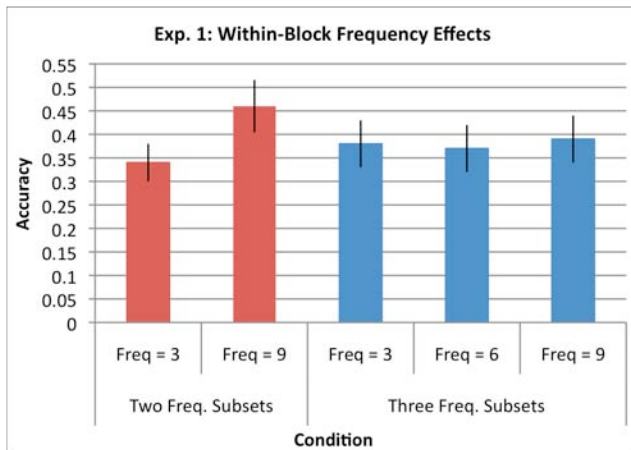


Figure 2: Accuracy for subsets of pairs with different frequency in two training conditions. Learning was well above chance (18AFC chance = .056) in every condition. Error bars show +/-SE.

Why did increased frequency aid learning in one condition, but not the other? How can it be explained that pairs of frequency 3, 6, and 9 are learned at equal levels? One plausible explanation is that once a pair is learned, future trials containing that pair effectively have reduced within-trial ambiguity. For example, if a learner sees (*A B*; *a b*) and has already learned *A-a*, then *B-b* may be inferred where it would not otherwise be certain. In this way, high frequency pairs may reduce the degree of ambiguity in later trials, and increase the learning of low frequency pairs. If this is true, the contexts in which each high frequency and low frequency pairs co-occur should play a critical role in effective statistical learning. In the next experiment, contextual diversity is varied in order to understand the counterintuitive finding in Experiment 1 and to directly measure the role of context diversity.

---

## Experiment 2

Experiment 1 showed that higher frequency can result in greater learning, but does not necessarily do so. In Experiment 2, we hold word-referent frequency constant and vary the contexts in which each pair appears to measure how the learning of a given pair can be affected by the other pairs it co-occurs with during training. The contextual regularities for each word-referent pair can be captured by two factors: 1) the number of co-occurring words and referents within a trial, namely, within-trial ambiguity; and 2) the number of *different* co-occurring words and referents over all the training trials, namely, contextual diversity (CD). The three conditions in this experiment manipulated both factors. In a **low/medium CD** condition, 18 pairs were divided into two groups. Six word-referent pairs in the low CD group were constrained to appear only with other pairs in this group during training. Likewise, the 12 pairs in the medium CD group only co-occurred with each other, and never with the 6 low CD pairs (Fig. 3, left). Thus, whenever a low CD pair appeared, the other stimuli on that trial had to be selected from the 5 remaining low CD pairs. In contrast, a given medium CD pair could appear with any of the 11 other medium CD pairs. Note that frequency was held constant – each of the 18 pairs was seen 6 times during training – and within-trial ambiguity was the same (3 words and 3 referents per trial). Only contextual diversity varied between these two groups. In each of the other two conditions in this experiment, all 18 pairs were randomly distributed to co-occur without constraint. To explicitly test the role of within-trial ambiguity, we implemented two versions of this design: the **uniform CD/3 pairs** condition with 3 words and 3 referents per trial, and the **uniform CD/4 pairs** condition with 4 words and 4 referents per trial (Fig. 3, middle and right, respectively).
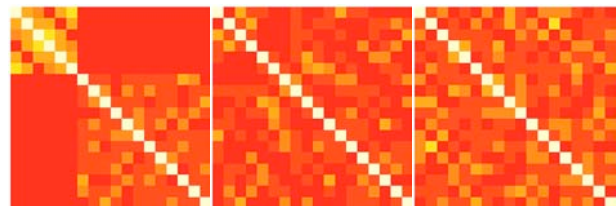


Figure 3: Word-referent co-occurrences for Exp. 2 (0=red, 6=white). **Left:** in the low/medium CD condition, each group's pairs co-occur only with other pairs within that group. **Middle and Right:** in the uniform CD/3 pairs and the uniform CD/4 pairs conditions, each pair randomly co-occurs with any of 17 other pairs.

Table 1 shows two metrics describing contextual diversity in this experiment: the mean number of other pairs that each pair co-occurs with during training, and the mean frequency of those co-occurring pairs. These two metrics are inversely related: if a given pair is made to co-occurs with more other pairs, it must occur with each of these other pairs fewer times, on average. For example, if pair *A-a* often appears with pair *B-b*, the incorrect associations *A-b* and *B-a* may be

learned. However, if *A-a* appears with many other pairs, it is unlikely to occur very often with any one of them (e.g., *B-b*). This is an example of how contextual diversity may be important for learning.

Table 1: Contextual Diversity in Experiment 2

| Condition \ CD | Low/Med | | Uniform/3 | Uniform/4 |
|---|---|---|---|---|
| Pairs per CD Group | 6 | 12 | 18 | 18 |
| Mean # of different co-occurring pairs | 4.0 | 9.2 | 8.8 | 12.2 |
| Mean frequency of co-occurring pair | 3.0 | 1.3 | 1.4 | 1.5 |

Greater within-trial ambiguity not only creates more possible associations on each trial, but also influences CD: In the 3 pairs/trial conditions, each pair appears on 6 trials, and thus appears with 12 other pairs during training (unique or not), In the 4 pairs/trial condition, each pair appears with 18 other pairs during training, as it occurs on 6 trials with 3 other pairs. Thus, pairs in the 4 pairs/trial condition appeared with more diverse pairs than pairs in the 3 pairs/trial conditions. Moreover, note in Table 1 that the 12 medium CD group pairs have very similar CD – by both metrics – to the uniform/3 pairs condition, since pairs in both these groups appeared with only 12 other pairs.

## Subjects

Undergraduates at Indiana University received course credit for participating. The varied contextual diversity condition had 63 participants, the 3 pairs/trial 18 condition had 38, and the 4 pairs/trial 18 condition had 77 participants. None had previously participated in cross-situational experiments.

## Stimuli & Procedure

The sets of pseudowords and referents for Experiment 2 were identical to those used in Experiment 1, but several new trial orderings were constructed to vary contextual diversity and within-trial ambiguity. The 27-trial, 4 pairs/trial conditions had the same timing as Experiment 1. The 36-trial, 3 pairs/trial conditions also had 3 seconds per stimulus pair, with 2 seconds of initial silence, making a total of 11 seconds. Knowledge was assessed after the completion of each condition using 18AFC testing, as in Experiment 1.

## Results & Discussion

Figure 4 displays the average levels of learning achieved in Experiment 2. In the low/medium CD condition, the 12 medium CD pairs were learned significantly better than the 6 low CD pairs (12 pairs $M = .47$, 6 pairs $M = .34$, paired $t(62) = 4.11$, $p < .001$), demonstrating a clear advantage for greater contextual diversity. Moreover, incorrect responses in the low/medium CD condition were largely chosen from the subset of pairs within the same group (thus co-occurring with the target pair): 56% percent of incorrect answers for low CD words were chosen from the 6 low CD referents (chance=33%, $t(55) = 5.48$, $p < .001$), and 76% of incorrect

answers for medium CD words were chosen from the 12 medium CD referents (chance=66%, $t(55) = 3.72$, $p < .001$). Thus, even incorrect answers reflected co-occurrences encountered during training, rather than arbitrary guesses.

Overall performance on the low/medium CD condition was no different than on the uniform CD/3 pairs (low/medium CD $M = .43$, uniform CD/3 pairs $M = .43$, Welch $t(72.7) = .08$, $p > .05$). As discussed, these two conditions have nearly the same degree of CD (see Table 1), which may explain their equal difficulty.
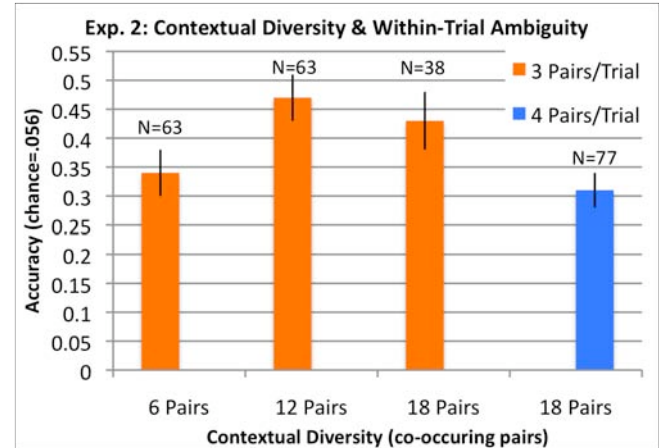


Figure 4: Accuracy for 4 pair groups that differ in CD and within-trial ambiguity. Error bars show +/-SE.

Finally, the uniform CD/4 pairs condition yielded less learning than in the uniform CD/3 pairs condition (4 pairs/trial $M = .31$, 3 pairs/trial $M = .43$, Welch $t(59.6) = 2.13$, $p < .05$), suggesting that the increased within-trial ambiguity of the 4 pairs/trial condition is more deleterious than any advantage conferred by the increased CD in this condition. In sum, this experiment demonstrated that greater CD alone improves learning, but that increased within-trial ambiguity is a powerful inhibitor. In Experiment 3, frequency and contextual diversity are manipulated within several conditions to clarify the interactions of these factors.

## Experiment 3

Experiment 2 showed that greater contextual diversity results in greater learning of those pairings, but that greater within-trial ambiguity can counteract this advantage. In Experiment 3, within-trial ambiguity was held constant, and frequency and contextual diversity were varied within 4 training conditions. Each condition had 18 pairs divided into 3 subsets of 6 pairs occurring at 3 frequencies: 3, 6, and 9. In the **low CD** condition, the pairs in each of the three frequency subsets appeared on trials only with pairs in the same group – never with pairs in other groups (Fig. 5a). That is, a 3-repetition pair would only be seen with other 3-repetition pairs, and similarly for 6- and 9- repetition pairs. In this way, learning a 3-repetition pair could help disambiguate only other 3-repetition pairs, etc. In the **high CD** condition, pairs of different frequencies co-occurred randomly throughout training (Fig. 5b). In this condition,

learning a given pair may help participants learn any pairs it co-occurred with in the future. In the final two conditions, the 12 pairs from two frequency subsets were allowed to co-occur, and the remaining 6 pairs co-occurred only with themselves (i.e., within-frequency). In the **3/6 mingled** condition, the 3- and 6-repetition pairs co-occurred during training, and the 9-repetition pairs only appeared with other 9-repetition pairs (Fig. 5c). In the **3/9 mingled condition**, 3- and 9-repitition pairs were mixed, and the 6-repetition pairs could only appear with other 6-repetition pairs (Fig. 5d). This final condition (Exp. 3-b) was run on a separate group of participants than the first three conditions (Exp. 3-a).
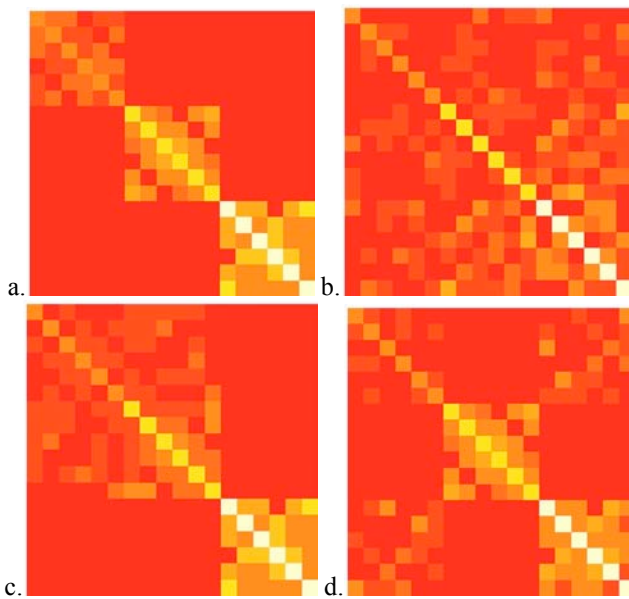


a.     b.
c.     d.

Figure 5: Exp 3. Co-occurrence matrices (0=red, 9=white). There were three frequency subsets in each condition (3, 6 and 9). To-be-learned pairs were manipulated in four ways to co-occur within and between each subset.

Table 3: Frequency & Contextual Diversity in Experiment 3 Cells display (mean number of different co-occurring pairs) / (mean frequency of those co-occurring pairs)

| Freq. \ CD | Low | High | 3&6 | 3&9 |
|---|---|---|---|---|
| 3 | 4 / 1.5 | 5 / 1.2 | 5.5 / 1.1 | 4.5 / 1.4 |
| 6 | 4 / 3 | 8.5 / 1.4 | 7.5 / 1.7 | 4 / 3 |
| 9 | 4.7 / 3.9 | 9.8 / 1.8 | 4.7 / 3.9 | 7.5 / 2.4 |

## Subjects

Participants were undergraduates at Indiana University who received course credit for participating. Exp. 3-a had 36 participants, and Exp. 3-b had 31. None had previously participated in cross-situational experiments.

## Stimuli & Procedure

The 36 pseudowords and referents from Experiments 1 and 2 were used in Experiment 3, in addition to 36 additional word-referent unique pairs that were constructed in a similar fashion. Four conditions were constructed to vary contextual

diversity and frequency within-block. Each training block consisted of 36 trials, each of which displayed 3 stimulus pairs over the course of 11 seconds (equivalent to the 3 pair/trial conditions in Experiment 2). Word learning was measured using an 18AFC test after each set of training trials, as in Experiments 1 and 2.

## Results & Discussion

Figure 6 displays the average levels of learning achieved in Experiment 3. In the low CD condition, increased frequency resulted in significant increases in learning (freq=3 $M = .25$, freq=6 $M = .42$, freq=9 $M = .71$; freq 6>3 paired $t(35) = 3.74$, $p < .001$; freq 9>6 paired $t(35) = 5.7$, $p < .001$). However, in the high CD condition, in which all pairs were allowed to co-occur, significantly more 3- and 6- repetition pairs were learned than in the low CD condition (freq=3 $M = .45$, paired $t(35) = 4.29$, $p < .001$; freq=6 $M = .64$, paired $t(35) = 4.47$, $p < .001$), although slightly fewer 9-repetition pairs were learned ($M = .6$, paired $t(35) = 2.09$, $p < .05$). In fact, there was no significant difference between the 6- and 9-repetition groups. Overall, learning was greater in the high CD condition than in the low CD condition (high CD $M = .56$, low CD $M = .46$, paired $t(35) = 2.45$, $p < .05$). Thus, mixing pairs of different frequency increases learning of the lower frequency pairs, and allows more total pairs to be learned. This is further demonstrated in the two mingled conditions which mixed two of the three frequency subsets (Fig. 3 c & d). When the 3- and 6-repetition subsets were allowed to co-occur, more 3-repetition pairs were learned than in the low CD condition ($M = .38$, paired $t(35) = 2.45$, $p < .05$), and performance on the 6- and 9- repetition pairs remained about the same.
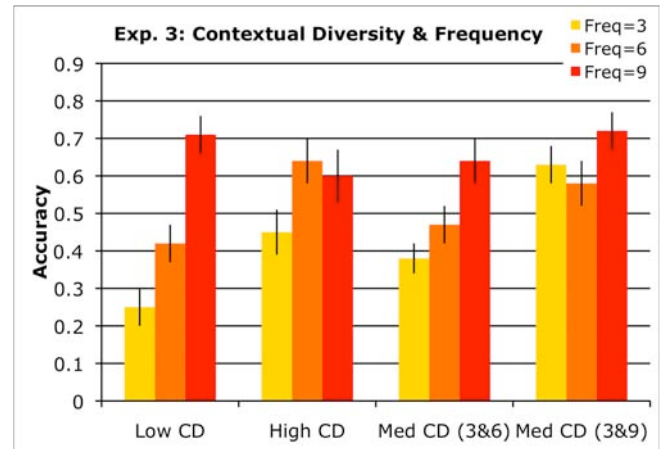


Figure 6: Accuracy for subsets of pairs with different frequency and contextual diversity. Error bars show +/-SE.

In the 3/9 mingled condition, 3-repetition pairs were learned even better than in the 3/6 mingled condition ($M = .63$, Welch $t(61.7) = 3.7$, $p < .001$). Finally, in the two mingled conditions, learning of each 9-repetition subset remained at the same level as in the low CD condition (3/6 mingled: paired $t(35) = 1.37$, $p > .05$; 3/9 mingled: Welch $t(63.2) = .08$, $p > .05$). Thus, increasing CD helped learning, on average, by boosting acquisition of low frequency pairs.

## General Discussion

Three factors that significantly determine the success of cross-situational statistical learning are word-referent frequency, contextual diversity, and the degree of within-trial ambiguity. These three factors are related: picking values for two of the factors somewhat constrains the value of the third. For example, consider a pair that is to appear 6 times during a 3 pairs/trial training set. On each of the 6 trials it occurs on, 2 other pairs must also appear. These 12 pairs may each be distinct, or some particular pairs may appear more often than once. If two pairs always co-occur, the "correct" word-referent pairs cannot be disambiguated. However, if one of the two pairs is learned prior to the appearance of the other (as may be the case for a high frequency pair), then the other may be learned more easily, since the prior knowledge of the frequent pair reduces the within-trial ambiguity.

Experiment 1 demonstrated that although varied frequency can result in increased performance for more frequent pairs, it is also possible for varied frequency to perplexingly yield equal performance, perhaps as a result of contextual diversity and effectively reduced within-trial ambiguity. Experiment 2 showed that increased contextual diversity improves learning for equal-frequency pairs. Moreover, pairs with greater within-trial ambiguity were learned less well – despite greater contextual diversity. Experiment 3 confirmed that more frequent pairs are learned more often, even when contextual diversity is controlled. In addition, increasing the contextual diversity of two groups of different frequency by allowing these groups to co-occur augmented learning of the less frequent of these groups. Indeed, the highest learning performance observed was in conditions with varied frequency and high contextual diversity. Intriguingly, these two characteristics that yield high performance are also embedded in real-world learning environments: words in any natural language have a skewed frequency distribution (Zipf, 1949), and naturalistic learning situations are highly complex, with many co-occurring words, events and objects (Hart & Risley, 1995). Varied frequency and contextual diversity seem to make situations more complex, but our results suggest that they facilitate statistical learning. Much structure is present in our world: some words and objects occur more than others, some appear in many situations and others in few, and situations vary greatly in complexity. The above experiments demonstrate that human learners are sensitive to these different kinds of regularities.

Frequency may be the most important factor yet investigated in cross-situational word learning: the more times a word-referent pair appears, the more opportunities there are to properly associate those stimuli. However, if that pair always appears with only a few other pairs, or simultaneously appears with many other pairs, each learning opportunity is worth very little. Thus, given enough contextual diversity to disambiguate proper pairings, and a reasonably small degree of within-trial ambiguity, learning can proceed with considerable ease. Because the presence of known high frequency pairs reduces within-trial ambiguity, highly ambiguous situations containing some familiar referents become feasible learning opportunities. This process of bootstrapping may account for the rapid acquisition of vocabulary in infants, who are known to learn frequent nouns earlier than less common nouns (Goodman, *et al.*, 2008). Once known, the ubiquitous nouns make possible the rapid acquisition of infrequent nouns. Thus, the results suggest a learning system that does not learn independent associations between individual words and referents, but one that rather learns a system of associations (see Yu, 2008). In such a system, a single word-referent pairing is correlated with all the other pairings that share the same word and all the other pairings that share the same referent, which are in turn correlated with more word-referent pairs—the whole system of them. We contend that the improvement in statistical word learning is in part due to the recruitment of accumulated latent lexical knowledge, used to learn subsequently appearing pairs. In future empirical and modeling studies, we plan to further investigate the mechanistic nature of statistical learning.

## References

Bloom, P. (2000). *How children learn the meaning of words*. Cambridge, MA: MIT Press.

Frank, M. C., Goodman, N. D., Tenenbaum, J. B. (2008). A Bayesian framework for cross-situational word-learning. *Advances in Neural Information Processing Systems 20* (pp. 457-464). Cambridge, MA: MIT Press.

Goodman, J. C., Dale, P. S., Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, *35*(03), 515-531.

Hart, B. & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Brookes.

Klein, K. A., Yu, C., & Shiffrin, R. M. (2008). Prior knowledge bootstraps cross-situational learning. *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1930-5). Austin, TX: Cognitive Science Society.

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

Yu, C. (2008). A statistical associative account of vocabulary growth in early word learning. *Language Learning and Acquisition*, *4*(1), 32-62.

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 exclusivity in cross-situational statistical learning. *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 715-720). Austin, TX: Cognitive Science Society.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley.