Mutual Exclusivity in Cross-Situational Statistical Learning

Daniel Yurovsky (dyurovsk@indiana.edu)
Department of Psychological and Brain Science, and Cognitive Science Program
1101 East 10th Street Bloomington, IN 47408 USA

Chen Yu (chenyu@indiana.edu)
Department of Psychological and Brain Science, and Cognitive Science Program
1101 East 10th Street, Bloomington, IN 47408 USA

Abstract

The Mutual Exclusivity (ME) constraint – a preference for mapping one word to one object – has been shown to be a powerful aid to children learning new words. We ask whether cross-situational language learning, in which word meanings are learned through computation of word-object co-occurrences across a series of highly ambiguous trials, is subject to the ME constraint. Our results show that participants can break the constraint to learn one-to-two word-referent mappings both when the referents are separated across time and when they are interleaved. This demonstrates the robustness of cross-situational statistical learning. We then use participants' ratings of their knowledge after individual trials to shed light on the underlying learning mechanism. Our results suggest that the ME constraint may be applied at multiple points along learning – within a single trial, across trials, and at test – which may explain one of its residual test effects found in the traditional language literature.

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

Introduction

The Mutual Exclusivity (ME) constraint – a preference for building one-to-one word-referent mappings – plays a critical role in human word learning. In Markman and Wachtel's (1988) classic experiment, a child is presented with a known object (*ball*) and an unknown object (*gyroscope*) and is asked by the experimenter to bring the "toma". Having never before heard "toma," the child will select the novel object *gyroscope* as its referent, but not the object *ball,* which already has a name. Since this original experiment, mutual exlcusivity has been reliably demonstrated in various studies. Subsequent research has also proposed several potential explanations for children's behavior, such as the Principle of Contrast (Clark, 1983), the Novel-Name-Nameless-Category Principle (Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992), and a Pragmatic Account (Diesendruck & Markson, 2001). Further, the ME constraint provides a powerful and reliable mechanism to deal with the reference uncertainty problem (Quine, 1960) in a single learning trial. Given a novel word and multiple potential referents, learners can apply mutual exclusivity to rule out those referents with known names and determine the correct referent on a single trial.

Recently, an alternative solution has been proposed to deal with the uncertainty in word learning: cross-situational statistics (Yu & Smith, 2007). A learner who is unable to determine the correct word-referent mappings in a single learning trial can nonetheless accumulate cross-trial statistics from a number of individually ambiguous training trials to ultimately discover correct word-referent associations. With respect to the above example, a learner may hear the words "toma" and "bosa" in the ambiguous context of seeing both a *loofah* and *gyroscope* without any information as to which word refers to which object. Although the mappings cannot be determined from this *single* situation, the learner could nonetheless solve the problem by keeping track of co-occurrences and non-occurrences *across situations*. Using this example, if the learner viewed a second scene with a *spatula* and *gyroscope* while hearing "diti" and "toma," *and* could combine co-occurrence probabilities across these two situations, the learner could correctly infer that "toma" maps to *gyroscope*, "bosa" to *loofah*, and "diti" to *spatula*. Recent empirical evidence has shown successful cross-trial learning in ambiguous situations in both adults (Yu & Smith, 2007) and children (Smith & Yu, 2008). Yu & Smith (2007) exposed adults to a series of 27 ambiguous trials, each containing 4 unknown words and 4 possible referents. Their participants learned 9.5 out of 18 mappings in less than 6 minutes. Thus, humans can use statistical information across multiple situations to learn word-object mappings by storing, computing, and continuously reducing a set of possible referents over time.

The present study intends to investigate the ME constraint in the cross-situational learning paradigm. Can adults learn multiple one-to-many word-referent pairings across a number of ambiguous learning trials? Since most previous ME experiments have focused on one-trial learning, and previous cross-situational language learning experiments used one-to-one stimuli, it is unclear what to expect. There are two potential outcomes. One is that human learners rely on mutual exclusivity in cross-situational learning in much the same way as in single-trial learning. As a result, they may still build one-to-one mappings while ignoring additional statistical regularities in the training data, or they may fail to learn at all due to the one-to-many confusion. This would suggest that ME is an additional constraint in cross-situational language learning. Alternatively, learners may break the mutual exclusivity constraint across multiple learning trials, demonstrating one-to-many mapping at test. This result would provide support for cross-situational learning as a fundamental mechanism supporting human learning in general, and word learning in particular.

Experiment 1

Following the cross-situational learning paradigm presented by Yu and Smith (2007), we asked participants to simultaneously learn many word-referent pairs from a sequence of highly ambiguous learning trials. On each trial, learners were presented with 4 words and 4 objects without any information as to which referred to which. Thus, only 4 of the 16 possible pairings were correct on each trial. Although individual trials were ambiguous, words always co-occured with their correct referents.

The key manipulation in the present study was to introduce another kind of word-referent ambiguity on top of the within-trial ambiguity. Specifically, the whole training set could be viewed as two halves. In the first half, learners were exposed to one set of word-referent pairings (e.g. A-a1). Then, in the second half, the same words co-occurred with new referents (A-a2) in the absence of the previously presented referents (A-a1). Participants were not told that there would be a switch. Thus, word A was mapped to two referents a1 and a2 over the whole training session.

Which pairings would participants learn? Three outcomes were possible. First, participants could have treated the later trials as disconfirming evidence of the first word-referent pairs (A occuring without a1), acquired the new pairings (A-a2), and rejected those learned previously. Second, learning one referent for a word may have caused learners to reject further associations (A-a2) and to maintain acquired lexical knowledge (A-a1). Third, learners may have accepted both pairings. The first two results would support a strong ME constraint in cross-situational learning, while the third would clearly indicate that the constraint can be broken.

Two characteristics of this paradigm are worth noting. First, as shown in Figure 1, learners needed to solve the above problem in the context of many simultaneously co-occurring words and referents (A-a1, A-a2, B-b1, B-b2, C-c1, etc.) across few, highly ambiguous exposures to each individual pairing. Second, participants were not told in advance that some words mapped to two referents instead of one. Thus, we intended to measure their learning performance in a naturalistic context without providing additional cues to influence their learning processes.
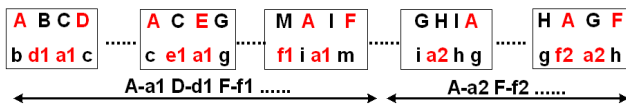

Figure 1: Participants learned one-to-two word-referent mappings (represented X-x1 and X-x2) in the context of 4 words and 4 referents per trial. For these *double* words, the first referent (x1) appeared exclusively in the first half of the trials while the second referent (x2) appeared only in the second half.

Subjects

Participants were forty-eight undergraduates who received course credit for volunteering. None had previously participated in any cross-situational learning experiments.

Stimuli

The stimuli were slides consisting of pictures of uncommon objects paired with auditorally presented pseudowords. These pseudowords were computer generated to broadly sample phonotactically-probable English forms and were produced by a synthetic female voice in monotone. 54 unique objects and 36 unique pseudowords were divided into three sets of 18 objects and 12 words. Each set of words was further divided into 6 *single* words, which were mapped to one unique object, and 6 *double* words, which were mapped to two unique objects. The first object for these *double* words will be referred to as the primacy referent, and the second as the recency referent. Each set thus contained 18 correct word-referent pairings to be discovered by the learner, and words and referents were presented 6 times each. However, because 16 potential associations could be made between the 4 words and 4 referents in each trial, learners experienced other "spurious" associations that made learning from individual trials difficult.

Referents of *single* words occurred randomly across all training trials. Each *double* word's first (primacy) referent occurred all 6 times in the first half of the trials, while the second (recency) referent occurred all 6 times only in the second half. As shown in Figure 1, the first half of the trials contained pairings of *single* words with their referents and *double* words with their primacy referents, while the second half contained the remaining occurrences of each *single* word with its referent along with all *double* word to recency referent pairings.

Each training slide consisted of 4 pictures, one in each corner of the screen, and 4 corresponding words, heard serially. There was no relationship between object and screen position, or screen position and word order. Each slide began with 2 seconds of silence, each word played for 1 second, and 2 seconds of silence were inserted between words for a total duration of 12 seconds per trial. Each training session consisted of 27 such slides, and each correct word-referent mapping was seen 6 times. Training slides were constructed the same way across all conditions.

Participants were tested after completion of each training phase. Test slides consisted of one trained word presented along with four pictures, one in each corner of the screen. One slide was made for each word, and they were presented in random order. Participants indicated the correct referent for a word by clicking on its picture. The pictures present varied by test condition. *Single* word tests were the same in each condition – one picture was the correct referent and the other three were foils (referents for other words). In the Primacy condition, each *double* word test consisted of the word's primacy referent (a1, etc.) and 3 foils. In the Recency condition, each *double* test consisted of the word's recency referent (a2, etc.) and 3 foils. In the Both condition, each *double* test consisted of both the primacy referent and recency referent and 2 foils. Thus, the primacy and the recency items were both present, and competed with each other, in the Both condition.

Procedure

Participants were told that they would be seeing pictures of objects and hearing words, and that they should try to determine which referred to which. They then engaged in three training sessions each followed by one of the three test conditions. Test conditions were counterbalanced for order.

Results

First, participants clearly learned both *single* and *double* words in all three test conditions. Even the lowest accuracy, for *double* referents in the Recency test condition, was significantly above chance ($M = .510$, $t(47) = 5.237$, $p < .001$). Thus, participants learned even under ME violations.

Second, accuracy for *single* words did not differ by test condition (MPrimary = .642; MRecency=.642; MBoth = .642, $F(2,47) = .016$, $p > .05$). Equivalence of these control words provides evidence that the three test conditions are directly comparable. A 2x2 ANOVA with factors of word-type (*single* or *double*) and test condition (Primacy or Recency) showed a significant effect of word type ($F(1,47) = 11.455$, $p < .002$) but not condition ($F(1,47) = .760$, p > .05) and no interaction ($F(1,47) = 1.580$, p > .05). Thus, *single* words were learned better.

Because *double* words were tested against two correct referents in the Both condition, they cannot be directly compared with *single* words. Instead, we compared choice frequency of primacy and recency referents in the Both condition. Figure 2 shows the significant preference effect, with primacy referents being chosen almost twice as frequently as recency referents (Mprimary = .524, Mrecency=.292; $t(1,47) = 3.403$, $p < .002$).
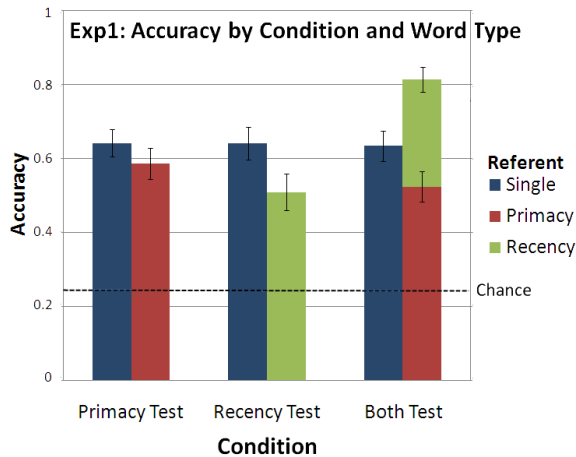

Figure 2: Accuracy of three referent types in three test conditions. Error bars represent +/-SE.

Discussion

While participants were more accurate on referents of *single* words, they successfully learned both *double* word referents, breaking the constraint of mutual exclusivity. Thus, learning the primacy referent did not prevent learning of the recency referent, and learning the recency referent did not force rejection of the learned primacy referent. Our results also reveal that learners can concurrently keep track of multiple word-referent mappings, which may explain their impressive performance in cross-situational learning.

Interestingly, participants showed a strong preference for the primacy referent when presented with both at test. This suggests that while mutual exclusivity did not prevent the acquisition of a second referent, it did produce a bias in favor of the first. This result is in line with a similar finding in the single-trial learning domain. Doherty (2004) trained children on second meanings for known words (for instance the word "cake" could also refer to a *tapir*). When read a story and asked to point to the object referred to by the homonymous word, children successfully chose the new meaning. However, when the referent corresponding to the known meaning (*cake*) was one of the test options, children often selected it even when the new referent was contextually appropriate. This first-referent preference also generalized to two consecutively learned meanings for novel words. Thus, mutual exclusivity may have driven primacy selection at test even when both referents were learned.

Experiment 2

Experiment 1 revealed that learners are able to break the mutual exclusivity constraint to acquire one-to-two word-referent mappings in cross-situational learning. However, successful learning in Experiment 1 may have been due to the temporal separation of the primacy and recency referents. That is, *double*-primacy pairings occurred only in the first half of training, while *double*-recency pairings occurred only in the second half. In Experiment 2, we removed this separation. As shown in Figure 3, primacy and recency referents occurred in random order across trials. Thus, if one trial contained word A and referent a1, the very next trial could contain A and a2 (but not a1).

Subjects

Participants were forty-eight undergraduates who received course credit. None had participated in Experiment 1 or any other cross-situational word learning experiments.

Stimuli and Procedure

Stimuli for Experiment 2 were identical to those used in Experiment 1, except that the two halves of training were shuffled together such that primacy and recency occurrences were interleaved across time.
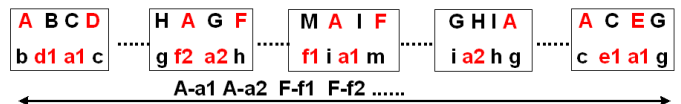

Figure 3: Participants learned one-to-two word-referent mappings (represented X-x1 and X-x2) in the context of 4 words and 4 referents per trial. This time, primacy(x1) and recency(x2) referent occurrences were interleaved across the entire training set.

Results

Figure 4 shows the results of Experiment 2. In all three test conditions, and for both *single* and *double* words,

participant accuracy was significantly above chance (smallest $M_{recency}$ = .517, $t(47)$ = 6.238, $p < .001$). A comparison of *single* word accuracy across test conditions showed no significant differences ($F(2,47)$ = .822, $p > .05$).

We again performed a 2x2 repeated measures ANOVA comparing *single* and *double* words in the Primacy and Recency test conditions. This time we found no effect of either condition ($F(1,47)$ = .426, $p > .05$) or word-type ($F(1,47)$ = 1.419, $p > .05$). Essentially, *single* word accuracy dropped to levels comparable to that of *double* words. We also compared primacy and recency selection frequency in the Both condition. Unlike in Experiment 1, we found no significant difference ($t(1,47)$ = .724, $p > .05$).

We then compared accuracy across the two experiments. Here we found a significant main effect of experiment on *single* word accuracy ($t(1,286)$ = 2.78, $p < .01$) but no effect of experiment on *double* accuracy in the Primacy and Recency conditions ($t(1,190)$ = .580, $p > .05$) or total *double* accuracy in the Both condition ($t(1,46)$ = 1.595, $p > .05$).
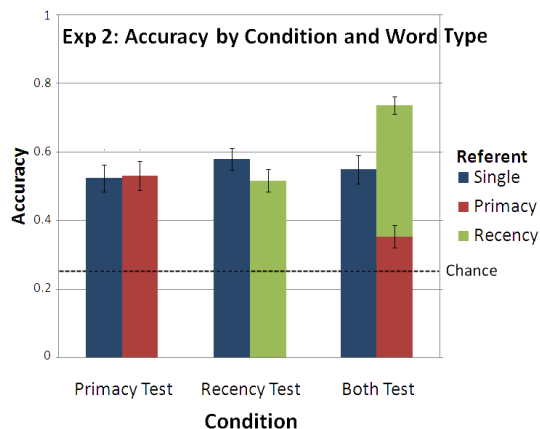

Figure 4: Accuracy for the three referent types in the three test conditions. Errors bars represent +/-SE.

Discussion

In Experiment 2, participants were able to learn one-to-two word-object mappings even when the two referents were interleaved during training. Whereas participants in Experiment 1 could have learned first one of the pairings and then the other, participants in Experiment 2 had to learn both simultaneously. Interestingly, *double* words were learned equally well in Experiments 1 and 2, suggesting that the temporal ordering had little effect on their difficulty. Thus the learning mechanism responsible for one-to-many mappings in cross-situational learning appears to operate in the same fashion in both cases.

Surprisingly, *single* and *double* words were learned equally well in Experiment 2. A cross-experiment comparison shows this to be the result of poorer *single* word learning in Experiment 2. Two explanations are possible. First, interleaved *double* referent ordering may have prompted learners in Experiment 2 to build many-to-one mappings for *single* words. As a result, they would not have been able to take advantage of mutual exclusivity for these words. An alternative explanation concerns the number of

spurious correlations possible in the two learning conditions. In Experiment 1, *single* words co-occur only with primacy referents in the first half of the trials and only with recency referents in the second half. This reduces the degree of uncertainty for *single* words on a given trial relative to the equivalent trial in Experiment 2.

Experiments 3a-b

Experiments 1 and 2 documented behavioral evidence of the robustness of cross-situational statistical learning – learners were able to effectively acquire statistical regularities of word-referent mappings by breaking the ME constraint. However, the underlying learning mechanism remained unclear. In Experiments 3a and 3b, we asked participants to report their level of knowledge of each referent trial by trial. This allowed us to assess their moment-by-moment learning and infer their internal knowledge states.

Subjects

Ninety-six undergraduates volunteered in exchange for course credit. None had participated in Experiments 1 or 2 or any other cross-situational word learning experiments. Half participated in Experiment 3a and half in 3b.

Stimuli and Procedure

Stimuli were identical to those used in Experiments 1 and 2. After each trial, participants were asked to enter a confidence rating for each picture by clicking a horizontal bar labelled 1 to 10 that appeared below it. After a rating was given for all 4 pictures, a button appeared in the center of the screen allowing the participant to move to the next trial. Input bars responded to multiple clicks, allowing participants to revise individual rating. Participants were given the same instructions as before, indicating that they should try to learn the mappings between words and objects. They were also told that a 1-10 scale would appear below each picture, and that they should indicate how confident they were that they knew the correct word for the corresponding picture. Trial ordering for Experiment 3a was equivalent to that of Experiment 1 and ordering for Experiment 3b was equivalent to that of Experiment 2.

Results

Experiment 3a Tests Forced-choice test results showed similar patterns to those found in Experiment 1. Participant accuracy was significantly above chance for both *single* and *double* words in all three conditions (smallest $M_{Recency}$ = .52, $t(47)$ = 5.336, $p < .001$). A comparison of *single* word accuracy across test conditions again showed no significant differences ($F(2,47)$ = .753, $p > .05$). A 2x2 repeated measures ANOVA for word type and condition showed a marginal main effect of word type ($F(1,47)$ = 10.642, p = .071), but no effect of condition ($F(1,47)$ = 1.992, $p > .05$), and no interaction ($F(1,47)$ = .060, $p > .05$). Comparison of primacy and recency selection in the Both condition showed a significant primacy bias ($t(1,47)$ = 2.312, $p < .03$).

Experiment 3b Tests Analyses of forced-choice test results showed similar patterns to Experiment 2. Accuracy was significantly above chance for both word types in all three test conditions (smallest $M_{Primacy}$ = .49, $t(47)$ = 5.754, $p <$ .001). A comparison of *single* word accuracy across test conditions showed no significant differences ($F(2,47)$ = 1.429, $p >$ .05). A 2x2 repeated measures ANOVA for word type and condition showed an effect of word type ($F(1,47)$ = 4.407, p = .05) but no effect of condition ($F(1,47)$ = .22, $p >$ .05) and no interaction ($F(1,47)$ = 3.685, $p >$ .05). No difference was found between primacy and recency selection in the Both condition ($t(1,47)$ = 1.047, $p >$ .05).

Confidence Data Validity In Experiment 3, we were primarily interested in using participants' confidence scores as a moment-to-moment measure of their learning states. To do so, we first asked whether participants' ratings accurately reflected their levels of knowledge. Since each referent occurred 6 times per training session, participants provided 6 ratings per referent. We averaged ratings by occurrence number across all referents to produce 6 values per participant – one for each occurrence number. Table 1 shows correlations between confidence scores and accuracy for each occurrence. There is a strong increasing trend across occurrences, with a significant positive correlation on all occurrences after the first.

Table 1: Correlation between Confidence and Accuracy.

Occurrence	1	2	3	4	5	6
Correlation	-.136	.253*	.375**	.454**	.507**	.592**

Figure 5 shows confidence data from one subject. Each row corresponds to an individual referent, and each column to a trial of the training session. On each trial, 4 confidence scores (four column cells) were given, one for each of the presented referents. The numerical scale used by our participants has been mapped to a gray scale, with brighter squares indicating higher confidence ratings.

Confidence Data Results We first examined the primacy test bias found in Experiment 1. We computed a 2x3x6 Mixed ANOVA with a between-subjects factor of Experiment (a or b), and within-subjects factors of referent type (single, primacy, or recency), and occurrence (1-6). We found significant main effects of referent type ($F(2,95)$ = 4.808, $p <$ .01) and occurrence ($F(5, 95)$ = 158.491, $p <$ .001) and a marginal effect of experiment ($F(1, 47)$ = 3.491, p = .065). Two-way interactions were significant between referent-type and occurrence ($F(10,95)$ = 5.413, $p <$ .001), referent-type and experiment ($F(2,95)$ = 7.155, p = .001), and occurrence and experiment ($F(5,95)$ = 6.522, $p <$ .01). The three-way interaction was also significant ($F(10,47)$ = 13.556, $p <$ .001). As shown in Figure 6, ratings increase with occurrence, were higher for single and primacy referents in Experiment 3a relative to 3b, and *recency* was significantly different from *primacy* in 3a, but not 3b.
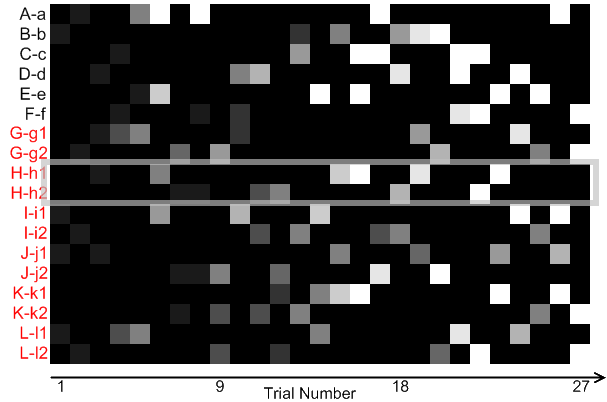

Figure 5: Confidence ratings provided by a subject in Experiment 3b. Rows correspond to individual referents and columns to individual trials. Two-to-one word-referent mappings are coded as X-x1 and X-x2. Just like the highlighted pairings H-h1 and H-h2, most *double* words showed increasing confidence scores across time for both primary and recency referents.

Next we examined the factors that influence the confidence rating of a specific referent trial by trial. We included rating on its most recent occurrence, experiment, word type (*single* or *double*), referent type (single, primacy, or recency), and also the average of the ratings of the three co-occurring referents on their most recent occurrences. A stepwise linear regression found significant effects of previous rating (β = .301, $t(92)$ = 50.221, $p <$ .001), *double* (β = -.019, $t(92)$ = -5.148, $p <$ .001), and the average of the co-occurring referents' previous ratings (β = .491, $t(92)$ = 81.909, $p <$ .001). Participant confidence for a referent was thus influenced not only by word type and knowledge of the same referent from previous experiences, but also by previous knowledge of the other co-occurring referents in the same trial.
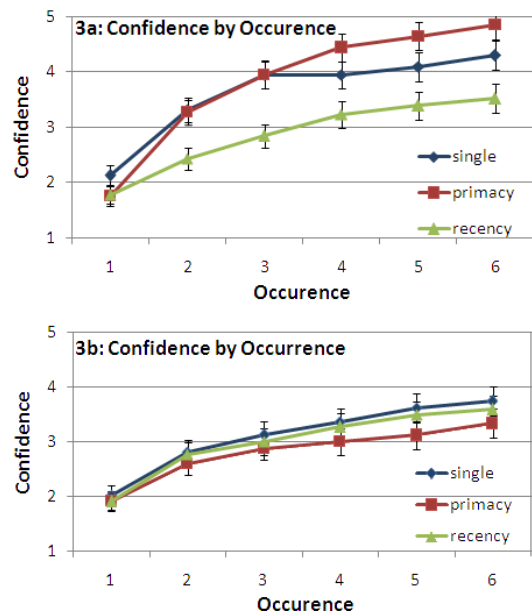

Figure 6: Confidence level by occurrence in Experiment 3

Discussion

In Experiment 3 we used confidence ratings provided by participants on each trial to assess the role of mutual exclusivity as learning progresses. These ratings allowed us to answer two main questions: first, what drove the primacy test bias in Experiment 1? Second, how was mutual exclusivity used within each trial?

To answer the first question, we examined confidence rating growth for each referent type over time, finding two patterns. First, participants were more confident in general in Experiment 3a than 3b, indicating impaired learning of even *single* words due to the interspersed ordering of *double* words. Second, we found that confidence for recency referents lagged behind confidence for primacy and single referents in Experiment 3a, but not 3b. This may explain the primacy bias in Experiment 1. While participants formed equally strong associations with both of a *double* word's referents, they had stronger conscious feelings of knowledge for the primacy referent, and thus preferred it at test.

We addressed the second question via regression analysis. As expected, the confidence rating assigned to a referent on a given occurrence was influenced by how well it was previously known. However, it was even more heavily influenced by how well the co-occuring referents were known, suggesting the application of mutual exclusivity. Participants who knew labels for the co-occuring referents could rule them out within-trial as labels for the referent in question. Thus, even though participants broke the mutual exclusivity constraint *across* trials, they used it to bootstrap their learning *within* trials.

General Discussion

We began by offering cross-situational statistics as an alternative to the constraint-based approach to language learning. However, all previous cross-situational language learning experiments used stimuli compliant with the mutual exclusivity constraint. Our results demonstrate the robustness of cross-situational learning even in the face of stimuli that violate ME both in serial (Experiments 1 and 3a) and in parallel (Experiments 2 and 3b). Although participants were impaired by these manipulations, their enduring performance is strong evidence for cross-situational statistics as a fundamental mechanism supporting word learning. This point is further reinforced by the emerging relationship between findings in this paradigm and those in the classical literature.

The primacy bias found in Experiment 1 and replicated in Experiment 3a maps nicely to results found of research on homonym learning. Although most work on mutual exclusivity has focused on two-to-one word-object violations (Woodward & Markman, 1998), a growing body of research is beginning to examine the reverse direction via homonyms (for instance 'steer' is both the act of driving a car and a large hoofed mammal). Children demonstrate impaired learning of homonymous words as late as 10 years of age (Mazzoco, 1997). Test preference for a word's first learned referent, demonstrated by Doherty (2004), is reproduced in our results. Experiment 3a allows us to offer a possible explanation. While participants demonstrated equal objective knowledge of each referent, they showed significantly higher confidence in their knowledge of the first referent beginning with its second training trial. This subjective feeling of superior knowledge may drive test preference for the first learned referent.

Finally, the use of the cross-situational paradigm and trial-by-trial confidence ratings allowed us to examine mutual exclusivity from a new angle. Standard language learning experiments have shown mutual exclusivity to have a strong effect on a single trial. Our cross-situational investigation suggests that the ME constraint can be applied at multiple points: within a single trial, across trials, and at test. We found that application at these levels is not completely dependent – participants who broke mutual exclusivity on a cross-trial basis still apply it within-trial, and in some conditions during testing. The obvious complexity of applying the ME principle in cross-situational learning suggest a promising direction for future research.

References

Clark, E. V. (1983). Convention and contrast in acquiring the lexicon. In T. B. Seiler & W. Wannenmacher (eds.), *Concept development and the development of word meaning* (pp. 67-89). New York: Springer.

Diesendruck, G. & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology*, *37*, 630-641.

Doherty M. J. (2004). Children's difficulty in learning homonyms. *Journal of Child Language*, *31*, 203–214.

Golinkoff, R.M., Hirsh-Pasek, K., Bailey, L.M., & Wenger, N.R. (1992). Young children and adults use lexical principles to learn new nouns. *Journal of Developmental Psychology*, *28*, 99-108.

Markman, E. M. & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*, 121–57.

Mazzocco, M.M.M. (1997). Children's interpretations of homonyms: a developmental study. *Journal of Child Language*, *24*, 441–67.

Quine, W. V. O. (1960). Word and Object. Cambridge, MA: MIT Press.

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

Woodward, A. L. & Markman, E.M. (1998). Early word learning. In W. Damon (series ed.) and D. Kuhn & R. Siegler (vol. eds), *Handbook of child psychology: vol. 2. Cognition, perception, and language*. New York: Wiley.

Yu, C. & Smith, L. B. (2007). Rapid Word Learning under Uncertainty via Cross-Situational Statistics. *Psychological Science*, *18*, 414-420.