Adaptive Constraints and Inference 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

Previous research shows that human learners can acquire word-referent pairs over a short series of individually ambiguous situations each containing multiple words and referents (Yu & Smith, 2007). In this kind of cross-situational statistical learning based on the repeated co-occurrence of words with their intended referents, the application of principles such as mutual exclusivity and contrast can leverage prior experience to reduce the complexity in situations with multiple words and multiple referents. However, these principles can also block the learning of one-to-many mappings. In a study analogous those done in traditional associative learning, we manipulate the early and late evidence for particular pairings in the cross-situational learning paradigm, and examine the effects on learning of both one-to-one and many-to-many mappings. Two major findings are: 1) participants use mutual exclusivity and contrast to facilitate learning; and 2) given sufficient evidence, learners can adaptively disregard these principles and learn many-to-many mappings.

Keywords: statistical learning; language acquisition; cross-situational associative learning; blocking; highlighting

Introduction

Human infants and adults can acquire word-object pairings after experiencing a small number of individually ambiguous situations, each of which consists of several words and referents. The abilities required in cross-situational learning are to remember at least some of the co-occurrence statistics of nouns and their objects and to integrate statistical information across multiple learning situations, if one assumes that these words often occur when their referents are present, and that words and their referents will appear together in different situations. This idea, cross-situational learning, has been proposed as an essential means by which infants acquire language (Pinker, 1989; Gleitman, 1990). Recently, Smith & Yu (2008) empirically demonstrated that young infants can learn nouns through cross-situational learning.

In the more complex adult cross-situational word learning paradigm (Yu & Smith, 2007), participants were instructed to learn which word goes with each object and were then shown a sequence of training trials. Each trial consisted of a display of a few novel objects and a few successively spoken pseudowords. Each word referred to a particular on-screen object, but the correct referent for each word was not indicated, thus making meanings ambiguous on individual trials. In one learning condition with four words and four objects on each trial, participants attempted to learn 18 word-object pairings from 27 12-second trials. Thus, each stimulus—and each correct pairing—appeared six times. Learning was assessed by a 4AFC test of each pseudoword after training, and showed that participants on average acquired slightly more than 9 of the 18 pairs.

How might participants learn so many pairings from such a short series of trials, each of which contains 16 possible word-referent pairings? Some reasonable principles that learners may apply during training can significantly restrict the space of possible pairings. Among others, the mutual exclusivity (ME) constraint, which holds that learners will try to map words to referents in a 1-to-1 way (Markman, 1990), has been demonstrated in various word learning tasks. In the context of cross-situational learning, for example, suppose a learner hears words *A B C D* and sees referents *b a c d* on some trial. However, she realizes that she has heard *A* and *B* on some prior trial, and seen objects *a* and *b*, but the other stimuli are novel. Even if she does not know that *A*-*a* and *B*-*b* are the correct pairings (they may not even be unambiguous at this point), employing ME she may assume that {*A*, *B*} map to {*a*, *b*}, and that {*C*, *D*} map to {*c*, *d*}. Thus, by applying the ME constraint with some minimal previous experience the learner can whittle the 16 possible pairings down to four. This sort of mechanism was used to model cross-situational learning of pairs that appeared in consecutive trials, as well as the pairs that were not temporally contiguous, both of which were learned better than in conditions with no temporal contiguity (Kachergis, Yu, & Shiffrin, 2009b).

The role of prior knowledge has also been investigated (Klein, Yu, & Shiffrin, 2008), and it was found that participants can use pre-studied pairs to facilitate subsequent learning. Their experimental conditions were the same with that of (Yu & Smith, 2007) described above (27 trials, four pairs per trial) except that three pairs were unambiguously pre-trained, learners acquired a mean of 13.7/18 pairs (10.8 of the un-pretrained pairs). Presumably, pre-training helped participants learn more un-pretrained pairs (compared to 9.5 pairs in Yu & Smith) by reducing the number of possible pairings in trials containing any pre-trained pairs. Further evidence of bootstrapping made possible by the assumption of ME was found in a study that varied pair frequency and contextual diversity (which other pairs a given appears with during training). Kachergis, Yu, and Shiffrin (2008a) found that pairs appearing only thrice during training were learned significantly better when they were allowed to co-occur with pairs appearing nine times than when all pairs appeared solely with pairs of the same frequency.

All of these results indirectly imply that learners assume mutual exclusivity during training, and demonstrate the added power it can yield when pairings are 1-to-1. Yurovsky and Yu (2008) gave participants a cross-situational task with some non-mutually exclusive pairings: halfway through training, half of the referents ceased appearing (e.g., A-a_1), and each was replaced by a second referent (e.g., A-a_2) which henceforth always co-occurred with the original referent's word (e.g., *A*). Thus, by the end of training half the words had both a primacy referent (a_1) and a recency referent (a_2). In separate 4AFC tests of primacy and recency referents, participants learned more than 50% of both the primacy and the recency referents. By the law of the excluded middle, some participants must have learned both pairings (e.g., A-a_1 and A-a_2), and thus built lexicons that violated ME. In another condition, the trials with primacy and recency referents were randomly interleaved, and learners still acquired nearly as many non-ME pairings as ME pairings (and above 50%, on average).

Ichinco, Frank, and Saxe (2009) studied ME using a different modification of the cross-situational word learning paradigm: Halfway through training, instead of replacing primacy referents with recency referents, they added an extra referent that always co-occurred with a particular old word-referent pair. That is, halfway through training, trials began to contain one more referent (i.e., 4) than words (3). Examining conjoint probability of learning ME-violating associations for each word, Ichinco *et al*. found that most learned pairings respected ME. Similar behavior was found in a condition with an added word instead of an added referent. Thus, although some participants learned some ME-violating pairings, the majority of learning behavior seems to follow a mutual exclusivity bias. This is unsurprising if one views cross-situational learning as a more complex form of traditional associative learning.

In a typical associative learning task, participants are given some subset of cues on each trial, asked to predict an outcome, and then shown the actual outcome. The subject's learning of associations between particular cues and outcomes is tracked over time. In cross-situational learning, the words can be construed as cues, and the objects as outcomes (or vice-versa). No trial-to-trial feedback is given, but the learner may generate it on the basis of the preceding training trials. *Blocking* is an associative learning effect often observed in experiments with two training stages: in the first stage, cue A is repeatedly paired with outcome X, and in the next stage A and B are jointly paired with X. The association between B and X is found to be weaker than when only the second stage training occurs: thus, B has been blocked by A's pretraining. Ichinco *et al*.'s design closely matches a blocking design (see Table 1), and their results—weak learning of the old word (or referent) to new referent (or word) association—are indeed a blocking effect.

Training Stage	Yurovsky & Yu, 2008	Ichinco, *et al.*, 2009	present study
Early	w_1-o_1	w_1-o_1	w_1-o_1
Late	w_1-o_2	w_1-$\{o_1, o_2\}$	$\{w_1, w_2\}$-$\{o_1, o_2\}$

Table 1: Comparison of three cross-situational ME designs. *N.b.*: these examples suppress other concurrent trained pairs.

The goal of the present paper is to systematically investigate how statistical learners accumulate and use current statistical evidence in subsequent learning. In the present study, for the first time we set up a strong test of inference, akin to associative learning's highlighting: will participants use knowledge of pairs acquired early in training, in addition to the principle of ME, to quickly learn pairs introduced late in training? If so, will this mechanism block the learning of many-to-many mappings? With two word-referent pairs sharing the same referent, one mapping appears in the early part of training and the other appears in the late training, will subjects prefer one over the other if we vary the amount of evidence (i.e., co-occurrence statistics) given during the early and late stages of training? Will they learn both pairs eventually? The set of experiments in the present paper allows us to answer those questions and examine how learners may adaptively incorporate evidence to potentially overwhelm biases.

Experiment 1

Participants are tasked with learning many word-referent pairs from a series of individually ambiguous training trials according to the cross-situational word learning paradigm (Yu & Smith, 2007). In the present study, each training trial is composed of two objects and two spoken pseudowords. On any given trial, participants can only guess which word refers to which object, since the order of presentation of the words is randomized, and there is no indication of which word refers to which object. However, since words only occur on trials with their intended referents, the correct pairings are disambiguated over the series of trials.

In the present cross-situational study, we divide each set of learning experiences into an early stage and a late stage, and systematically vary the number of times pairs appear in each stage. Half of the pairs appear in both the early training stage and the late stage, and the remaining half the pairs only appear in the late stage. As shown in Table 2, when a pair w1-o1 from the early stage appears in the late stage, another specific pair only appearing in the late stage (w7-o7) always co-occurs with w1-o1. Thus, in the late stage, {w1,o1,w7,o7} always co-occur, therefore, all of the four possible associations are reasonable (w1-o1, w1-o7, w7-o1, w7-o7). In fact, there is no additional information in the late stage that can be used to identify which ones are better than others. However, the key manipulation in this experiment is to vary the strength of w1-o1 in the early stage. More specifically, the early pairs co-occur 0 (no early training), 3, 6 and 9 times before they appear together with the late pairs. Given that we already know that people can effectively extract co-occurrence statistics in cross-situational learning, it is reasonable to assume that participants in this study would form some form of knowledge about w1-o1 when they enter the late stage trials. The research questions are: 1) whether they would prefer w7-o7 by applying the ME constraint; 2) whether they still learn w1-o7 and w7-o1 – a violation of ME; 3) how the amount of evidence about w1-

o1 may influence how they process the otherwise-ambiguous information in late trials with {w1,o1,w7,o7}.

Training Stage	Repetitions	Example Trials
Early [pairs 1-6]	0, 3, 6, or 9	$\{w_1, w_2, o_1, o_2\}$, …, $\{w_1, w_5, o_5, o_1\}$,
Late [pairs 1-12]	3 (Exp. 1) 6 (Exp. 2) 9 (Exp. 3)	$\{w_1, w_7, o_7, o_1\}$, …, $\{w_1, w_7, o_7, o_1\}$

Table 2: Experiment design, with example trials. Early pairs are indexed 1-6, and late-only pairs are 7-12. Pairs 1-6 also appear in the late stage, and thus occur more than pairs 7-12.

Subjects

33 undergraduates at Indiana University participated to receive course credit. None had participated in other cross-situational experiments.

Stimuli

On each training trial, two unusual objects (e.g., sculptures) are simultaneously shown while two pseudowords were sequentially heard. The 48 computer-generated words are phonotactically-probable in English (e.g., “bosa”), and were spoken by a monotone, synthetic female voice. These 48 objects and 48 words were randomly assigned to four sets of 12 word-object pairings, one set for each learning condition. Within each set, 6 pairings only appear in the late training and the other 6 appear in both the early and late trainings. Each 8-second training trial began with the appearance of two objects, which remained visible for the entire trial. After 2 s of initial silence, each word was heard (randomly ordered, duration of 1 s) followed by 2 s of silence.

There were four learning conditions in this study. The late training was the same in those four conditions, and was composed of 18 trials. Each pair appeared 3 times late in training. Therefore, the same trial {w1,o1,w7,o7} also appeared 3 times. Four conditions varied in the early training. There was no early training in condition 1. In condition 2, each of 6 early pairs appeared 3 times, forming 9 early trials before the late training. In conditions 3 and 4, each early pair appeared 6 or 9 times. Accordingly, there were 18 and 27 early training trials in those two conditions.

Procedure

Learners were instructed that they would see a series of trials with two objects and two alien words, and that they should try to figure out what each word means for a test at the end. Participants were not told there were training stages, and there was no perceptible break. After training, their knowledge was assessed using 11-alternative forced choice (11AFC) testing: on each test trial a single word was played, and the participant was instructed to choose the appropriate object from a display of 11 of the 12 trained referents. Participants were instructed to click on the best referent for the word. Each word was tested twice in 11AFC trials: once without its corresponding early referent as one of 11 choices to test its association with the late referent (‘early-late’ and ‘late-late’), and once without its corresponding late referent as one of 11 choices to test its association with the early referent (‘early-early’ and ‘late-early’). In this way, we tested their knowledge of all of the four possible associations showing in Figure 1 (two associations for each word), and we access their knowledge of each of four possible associations individually in this test.

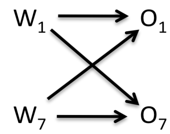

Figure 1: Example of associations tested by 11AFC.

Training condition order was counterbalanced, and each learner participated in all four conditions.

Results & Discussion

Fig. 2 displays the learning performance[1] in the training conditions with 3, 6, and 9 repetitions of early pairs. A 3x2x2 ANOVA with factors of early repetitions (3, 6, or 9), pair stage (early or late), and pairing type (within-stage or across-stage) showed only a significant main effect of pairing type ($F(1,30) = 8.62$, $p<.01$)[2]. As shown in Fig. 2, learning of within-stage (i.e., early-early and late-late word-object) pairs was much greater than learning of across-stage pairs (within-stage $M = .71$, across-stage $M = .15$). Even this small proportion of mean between-stage pair learning was significantly above chance, by subject (11AFC chance = .091; paired $t(30) = 2.29$, $p<.05$). Thus, although within-stage pairings—those consistent with ME—were clearly favored, participants also learned some ME-violating across-stage pairings. However, the indistinguishable, high level of performance on both early-early and late-late pairings is evidence of strong, ME-based inference: a given late pair could only be unambiguously learned by filtering out the consistently co-occurring early pair.

It is surprising that the number of early pair repetitions did not have a consistent effect on performance ($F(2,30) = .90$, $p>.05$). That is, even three repetitions of each early pair was enough prior experience to allow participants to infer the correct late pairs, thus achieving performance equal to the proportion of early pairs learned—no benefit was conferred by additional repetitions of early pairs (i.e., 6 or 9). In the condition with no early stage (0 early repetitions), participants learned 32% of the 2-to-2 pairings, on average—well above chance (paired $t(30) = 8.83$, $p<.001$).

Experiment 1 demonstrated that participants can efficiently leverage the ME constraint to learn late-appearing pairs that always co-occur with early pairs, and would thus be ambiguous, if not for prior experience. Indeed, performance on pairs learned using this filtering inference technique was no less than the performance on the

[1] Data from two subjects were excluded because their average performance in all four conditions was at chance (11AFC chance = .091). The outcomes of statistical tests were unaffected.

[2] We will report the results of the 0-early-training condition later as those results can be used to compare the data across experiments.

early pairs, which were learned by ambiguous cross-situational training, and which appeared more times, overall. Moreover, in tests of across-stage pairings, participants showed some learning of ME-violating associations.

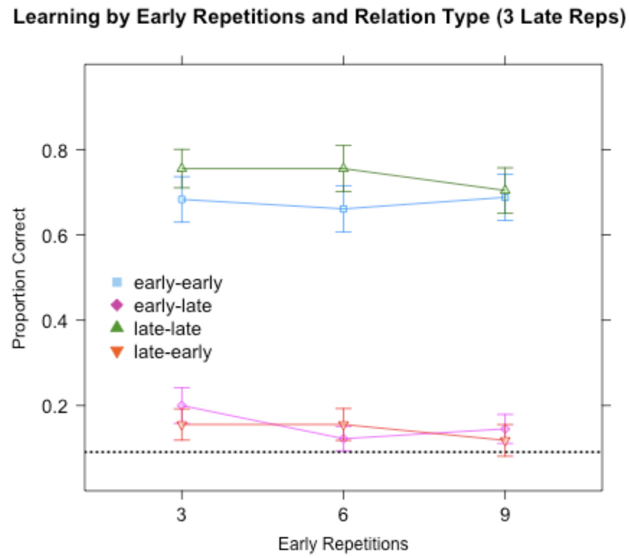

Figure 2: Mean accuracy by condition for the four types of associations (within and between early and late pairs). Error bars show +/-SE and the dotted line shows chance (.091).

In the next experiment, we increase the amount of late evidence, which we expect will cause participants to adaptively relax the ME constraint and learn more across-stage pairings (e.g. w7-o1). If they do relax ME, will they also learn fewer late-late pairings, or can they learn both?

Experiment 2

Experiment 1 showed that participants tend to utilize the principle of mutual exclusivity in combination with prior cross-situational training to quickly infer the referents of late-appearing words. In fact, performance on late and early pairs was undifferentiated, even when early pairings appeared more frequently (in both the early and late stages) than the late pairs (only the late stage). It appears that three more repetitions of early pairs grants enough knowledge to highlight the late-late pairs—which appear only three times—and bring performance up to the same level as the early-early pairs.

Thus, to some degree, Exp. 1 is analogous to blocking studies in associative learning literature, but with multiple co-occurring cues and outcomes on each trial. We show that just three repetitions of early pairs dramatically changed how they processed statistical information later and they apparently applied ME-based learning as evident by very few across-stage pairs learned later (i.e., early- late or late-early). In Experiment 2, we provide learners with more evidence for across-pair associations via additional repetitions of late stage pairs. Will participants adapt to this change, evaluate the statistical evidence in a different way and begin to count more on statistical information in the late part? Will they learn more ME-violating pairings?

Subjects

29 undergraduates at Indiana University received course credit for participating. None had participated in previous cross-situational experiments.

Stimuli & Procedure

The sets of pseudowords and referents used in Experiment 2 are identical to those used in Experiment 1. In each condition, the late stage of training was simply doubled from 18 trials to 36 trials wherein each {w1,o1,w7,o7} appears six times (instead of three times in Experiment 1), yielding three more repetitions of late and early pairs.

Results & Discussion

Figure 3 displays the average[3] levels of learning achieved in Exp. 2. Once again, a 3x2x2 ANOVA with factors of early repetitions (3, 6, or 9), pair stage (early or late), and pairing type (within-stage or across-stage) showed only a significant main effect of pairing type ($F(1,24) = 5.45$, $p<.05$). As in Exp. 1, learning of within-stage (i.e., early-early and late-late) pairs was much greater than learning of across-stage pairs (within-stage $M = .74$, across-stage $M = .28$). Thus, the increase of statistical evidence in the late stage with three more repetitions of both early and late pairs, didn't improve the learning of within-stage pairs (Welch $t(51.8) = .55$, $p>.05$). Nonetheless, learning of across-stage pairings increased—as predicted—due to increased late stage pairings (Welch $t(37.5) = 2.35$, $p<.05$). That is, having six rather than three repetitions of each late pair with its matched early pair in the late stage increased the learning of early word to late object (and vice-versa) pairings; the pairings that violate ME. Thus, although people are initially inclined to assume mutual exclusivity, and are able to use it to quickly infer the meaning of novel words, people will also adaptively relax ME in the face of greater evidence that words are being mapped to additional objects. As in Exp. 1, there was no significant effect of the number of early repetitions on learning ($F(2,24) = .06$, $p>.05$).

In summary, this experiment demonstrates that learners react to increased evidence that late and early pairs go together by learning more pairings. By doing so, they exhibit the ability to violate ME in order to learn 1-to-many mappings without reduced learning of ME-compliant pairings. In Experiment 3, we increase the late stage evidence once more to determine whether learners will continue to adaptively relax the ME constraint and learn more across-stage pairings.

[3] Data from four subjects were excluded because their average performance in all four conditions was at chance (11AFC chance = .091). The outcomes of statistical tests were unaffected.

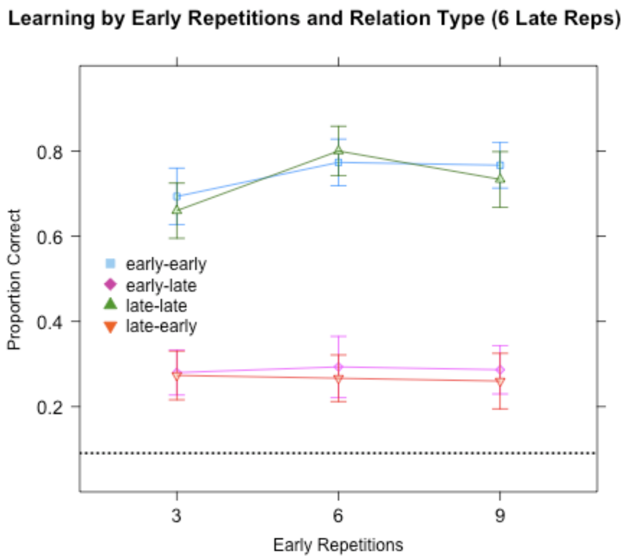

Figure 3: Accuracy by number of early pair repetitions and association type. Error bars show +/-SE.

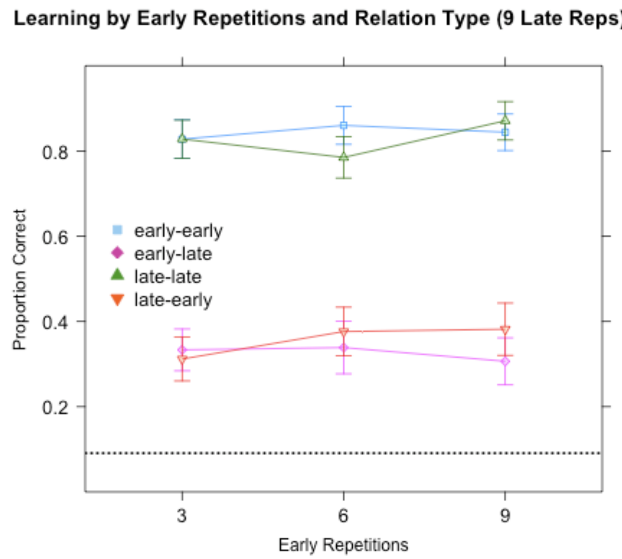

Figure 4: Accuracy by number of early pair repetitions and association type. Error bars show +/-SE.

Experiment 3

Experiment 2 provided six repetitions of late pairs with matched early pairs in comparison to Experiment 1's three repetitions, and learners indeed began to learn more late-early/early-late (i.e., across-stage, ME-violating) pairings. In Experiment 3, we once again increase the late stage by three repetitions (to nine), providing further evidence for cross-association of the matched early and late pairs.

Subjects

34 undergraduates at Indiana University received course credit for participating. None had participated in earlier cross-situational experiments.

Stimuli & Procedure

The same sets of pseudowords and referents were used in Experiment 3 as were used in Experiments 1 and 2. In each condition, there were 54 late-stage training trials, yielding three more repetitions of each late and early pair.

Results & Discussion

Figure 4 shows the mean[4] learning level by condition and pair type in Experiment 3. A mixed ANOVA (3, 6, or 9 late repetitions [between-subjects] x 3, 6, or 9 early repetitions x early or late pair stage x across- or within-stage pairing type) showed a main effect of pairing type ($F(1,101) = 161.76$, $p<.001$) and an interaction between pairing type and the number of late repetitions ($F(2,101) = 10.89$, $p<.01$). In brief, the patterns in Exp. 3 were consistent with what we observed in Exp. 2: participants learned within-stage pairs (e.g. w1-o1, w_7-o_7) quite well and also learned across-state pairs (e.g. w7-o1, w1-o7) with the additional evidence provided in late training.

[4] Data from three subjects were excluded because their average performance in all four conditions was at chance. The outcomes of statistical tests were unaffected.

General Discussion

When attempting to learn word meanings from a series of individually ambiguous trials, applying the mutual exclusivity constraint on each trial can significantly reduce the number of pairings a learner must consider, and even allow learner to quickly infer the meanings of novel words and referents. In the cross-situational experiments presented here, participants used such filtering—akin to highlighting in the associative learning literature—to quickly learn the same proportion of late-late pairings as early-early pairings. Thus, learners can use ME to infer the late-late pairings, even though each late-late pairing always co-occurred with one early pairing.

However, if pairings are not in fact mutually exclusive, or if word-referent mappings may change over time, assuming ME would be maladaptive. Across experiments, we increased the number of repetitions of late pairs, each of which always appeared with a particular early pair (e.g., w_1 and o_1 always appeared with w_7 and o_7, which had never appeared before the late stage). As an early stage pair appears more often with a particular late pair, a flexible learner would relax the assumption of ME.

Exp. 1, with three late pair repetitions, demonstrated that participants learn early-early pairs very well with only three early repetitions, and use this knowledge, in combination with ME, to learn the late-late pairs at a similar rate. However, even in this experiment, participants showed evidence of learning some ME-violating pairings (i.e., late-early and early-late pairings). Results from Experiment 2, with six late pair repetitions, looked very similar, but with slightly higher learning of ME-violating pairings. With nine late pair repetitions, Experiment 3 provided further evidence that participants should disregard ME, and indeed they learned more cross-stage pairings.

Viewing cross-situational statistical word learning as a complex form of associative learning in which there are multiple cues and outcomes on each training trial, it is unsurprising to find evidence for ME in cross-situational experiments, for ME may be responsible for well-known associative learning phenomena such as highlighting and blocking. However, our results demonstrate that participants do not merely use the ME constraint for logical inference. Instead, they utilize an adaptable learning strategy: as statistical evidence for violation of ME increases, they learn more ME-violating pairs. A comparison of the results from the 0-early-stage condition in three experiments in which each early pair (e.g., w_1-o_1) always co-occurred with a particular late pair (e.g., w_7-o_7). Figure 5 shows the mean proportion of learned ME-respecting pairings (i.e., only pairings involving each stimulus once) and ME-violating pairings (i.e., pairings that involve a stimulus twice) that participants learned by the number of late repetitions (i.e., experiment), which increases as more late repetitions give evidence that the ME constraint should be relaxed.

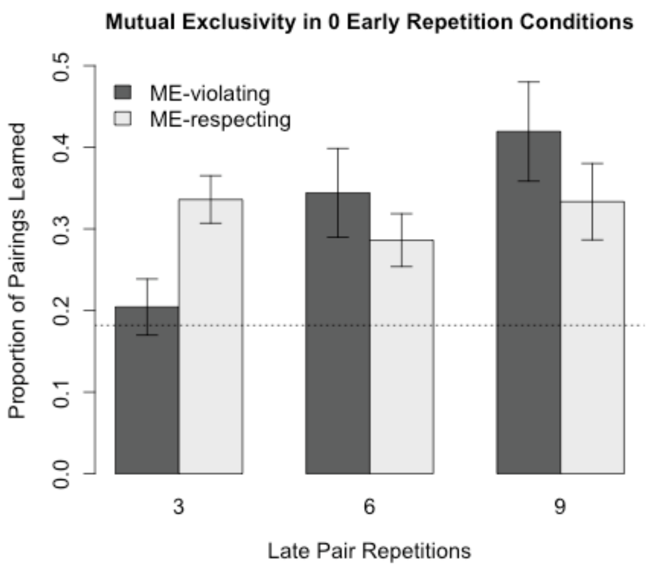

Figure 5: The mean proportion of pairings that each participant learned that violate ME and that comply with ME (across all experiments). Chance for ME-respecting pairings is the dotted line (2/11). Chance for ME-violating pairings is $(1/11)^2 = \sim.01$.

In summary, the results from the present study provide a complete picture of how participants use the ME constraint and how they accumulate statistical evidence over the course of learning. By varying the strengths of word-object associations in both the early and late training stages, we were able to demonstrate various learning behaviors that have been shown in previous studies. In fact, some of those studies have produced incompatible results. Therefore, the contribution of our work is to unify different views inferred from previous results. We argue that human statistical learning is adaptable: learners are able to adjust their learning strategy over the course of learning in response to changing amounts of evidence for particular types of pairings. One possible reason that previous results are not compatible is because each study took a snapshot of a continuous learning process. For example, many ME studies are conducted with one or two simple trials. We observe that learners' cross-situational learning strategies are dynamic and adaptable, and thus cannot be adequately portrayed by one snapshot. Rather, we need to examine learning trajectories by varying the amounts (repetitions) and types (ME-violating or –respecting) of evidence. We believe that the data presented here will quite useful in constraining models of cross-situational language acquisition. Future studies use real-time behavioral data (e.g., eye movements) to provide access online learning strategies.

Acknowledgments

This research was supported by National Institute of Health Grant R01HD056029 and National Science Foundation Grant BCS 0544995. Special thanks to Tarun Gangwani for data collection.

References

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

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 1-55.

Ichinco, D., Frank, M. C., & Saxe, R. (2009). Cross-situational word learning respects mutual exclusivity. In N. Taatgen, H. van Rijn, J. Nerbonne, & L. Schomaker (Eds.) *Proceedings of CogSci 31*.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2009a). Frequency and contextual diversity effects in cross-situational word learning. In N. Taatgen, H. van Rijn, J. Nerbonne, & L. Schomaker (Eds.) *Proceedings of CogSci 31*.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2009b). Temporal contiguity in cross-situational statistical learning. In N. Taatgen, H. van Rijn, J. Nerbonne, & L. Schomaker (Eds.) *Proceedings of CogSci 31*. Austin, TX: Cognitive Science Society.

Klein, K. A., Yu, C., & Shiffrin, R. M. (2008). Prior knowledge bootstraps cross-situational learning. *Proceedings CogSci 30*, Austin, TX: Cognitive Science Society.

Markman, E. M. (1990). Constraints Children Place on Word Learning. *Cognitive Science*, 14, 57-77.

Pinker, S. (1989). *Learnability and cognition*. Cambridge, MA: The MIT Press.

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

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

Yurovsky, D. & Yu, C. (2008). Mutual exclusivity in cross-situational statistical learning. *Proceedings of CogSci 30*. Austin, TX: Cognitive Science Society.