Temporal Contiguity in Cross-Situational Statistical 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

Recent research has demonstrated that participants often learn a surprising number of word-referent pairings solely from their co-occurrence statistics across individually ambiguous trials. To isolate processes, past designs prevented the same pairing from appearing in two consecutive trials. Yet such temporal contiguity often appears in real world settings, and seems likely to improve learning. The present research examines and models the effects of such repetitions. Our results show that allowing word-referent pairs to appear in adjacent trials indeed increases overall learning. Not only are the repeated pairs improved, but other pairs are improved, as well. Repetition seems to allow segregation of pairs that are and are not repeated from the previous trial, thereby allowing differential attention between the subsets. However, attention also seems to shift away from pairs that are repeated many times—to their detriment, but to the benefit of concurrent unrepeated pairs. The findings are explored with an associative learning model to provide a formal account of the underlying learning mechanisms.

Keywords: statistical learning; cross-situational word learning; language acquisition; ambiguity; attention

Introduction

To learn a language, people must learn which words refer to which physical referents. A learner listening to a speaker needs to ascertain which objects in the shared environment are the referents intended by the speaker. Almost all occurrences of words occur in settings where there are multiple possible referents. Particularly early in learning, there will be high ambiguity concerning the correct referent. Learning nonetheless takes place because correct pairings of words and referents tend to reoccur over many situations, a phenomenon termed 'cross-situational learning' (Pinker, 1984; Gleitman, 1990). A recent proposal of interest in cognitive development implements this well-established idea using statistical learning, which has been shown to work in many distinct perceptual domains (e.g., Conway & Christiansen, 2005). In a cross-situational word learning experiment, a learner acquires word meanings by tracking the co-occurrences of multiple words and referents across situations, ultimately discovering the most likely word-referent mappings. Such statistical word learning has been observed in both infants (Smith & Yu, 2008) and adults (Yu & Smith, 2007).

In typical adult studies, participants are instructed to learn which object each (novel) word denotes. They are presented with a series of study trials, each consisting of an array of several novel objects (e.g., a photograph of a metal sculpture, another of a tool, etc.) and successively spoken pseudowords (e.g., "manu", "bosa", etc.). Each pseudoword refers to a single onscreen object, but the correct referent for each pseudoword is not indicated, making referents ambiguous on individual trials. In a typical learning scenario, participants attempt to learn 18 pseudoword-object pairings from 27 12-second trials, with four pseudowords and four objects presented per trial. This configuration allows each stimulus (and hence each correct word-referent pairing) to be presented six times. One way to learn the correct pairings would involve the accumulation of pseudoword-object co-occurrence statistics across the training trials. To assess the learning that has taken place during training, each pseudoword is individually presented, and participants are asked to choose the appropriate object from a subset of the 18 objects. Yu & Smith (2007) found that adults learned an average of 9.5 of the 18 pairings when choosing from four alternative referents at test (i.e., 4AFC). Remarkably, several participants manage to learn all 18 mappings. Thus, humans can use the co-occurrences of multiple words and objects across individually ambiguous trials to learn word-object mappings.

In order to isolate processes and better control the first studies of this sort (Yu & Smith, 2007), an artificial constraint was used in previous designs of cross-situational learning studies: Word-referent pairs were not allowed to appear in consecutive trials. However, such repetitions are common in real learning environments, partly because of temporal contiguity inherent in the physical environment. For instance, a visual object that a learner is attending to at one moment is quite unlikely to suddenly disappear at the next; rather, it will gradually move away from the central to the peripheral visual field, remaining in sight for some time. This temporal contiguity in the environment may aid learning in a number of ways if the cognitive system is able to make use of this regularity.

For a simple example in the cross-situational paradigm, consider two successive trials on which pseudoword *A* and object *a* occurred, but all other stimuli – three other words and objects on each of the trials – differed. Assuming memory for the previous trial, the participant could infer that *A-a* is a correct pairing ("a" was the only repeated pseudoword, and *A* was the only repeated object). As a second example, consider two successive trials on which three of the four word-referent pairs are repeated (e.g., *D E F* and *f d e* appear on both trials), but the first trial contained *B-b* and the latter *C-c*. Given memory for the first trial, the participant could infer that *D E F* must be associated with *f d e*, albeit the exact pairings would remain ambiguous. Regardless of this ambiguity, it would be possible to infer that *B-b* must be correct and *C-c* must be correct, since these are the only remaining possibilities. These are just two

examples that illustrate the kinds of learning gains that are possible when pairs are repeated over successive trials. The gains are of course dependent on memory for the items in the previous trial. Given perfect memory for all trials, such inferences would not be restricted to repetitions on successive trials. However, participants have imperfect memories, and tend to remember things best from the immediate past. Thus, learning enhancements due to repetitions are expected to be largest when repetitions occur on successive trials. In the following studies we assess the effects of different degrees of temporal contiguity and model the results.

Experiment 1

Participants were asked to learn word-referent pairs from a series of individually ambiguous training trials using the cross-situational word learning paradigm (Yu & Smith, 2007). On each training trial, four novel objects were simultaneously shown while four pseudowords were serially spoken. No information was given to denote which word refers to which object on a given trial. Without any learning from previous trials that could be used to reduce ambiguity, there would be four equally possible referents for each pseudoword or each object, and thus 16 equiprobable 1-to-1 mappings of the four pseudowords onto the four objects. However, since words always appeared on trials with their proper referents, and there was mixing of groups of pairs over trials, the correct pairings reoccur more often over trials, and hence can be learned.

The manipulation of interest in this study is the repetition of some pairs in consecutive trials. As discussed above, the degree of overlap between two trials affects the type of inferences that can be made. Possible effects range from making a single repeated pairing obvious, to making the only unrepeated pairing obvious, to merely reducing the number of possible associations. In general, it is expected that trial orderings with more trial-to-trial repetitions will yield higher overall learning.

Each trial consisted of four words and four referents, allowing construction of training sequences of 18 pseudoword-object pairs presented over 27 trials in which each 'correct' pairing occurred six times ('incorrect' pairing co-occurrences ranged from 0 to 4, M = 1.5). The control condition was a fixed temporal sequence containing no pairs that repeated on successive trials. In all of the other conditions, the individual training trials were the same set used in this control condition. However, the *order* of the trials was shuffled many times to create orderings with degrees of trial-to-trial repetitions. Because the orderings were all constructed from the same set of trials, all co-occurrence statistics remained identical across conditions.

The successive repetition (SR) score of a trial ordering is the mean number of word-referent pairings that overlap across all consecutive pairs of trials. The minimum SR score is 0, as no pairs ever overlap. (A single trial repeated 27 times would give an SR score of 4, but of course this is not a reasonable learning situation.) The maximum SR score we were able to obtain by reshuffling the control sequence was 2.04. That is, in this condition, on average, a little over two word-referent pairs repeated in every pair of successive trials. The sequences we constructed and used had SR scores of 0.00, 0.33, 0.67, 1.00, 1.41, and 2.04.

Assuming memory of the preceding trial, repeating some of the pseudoword-referent pairs from trial *n-1* on trial *n* allows segregation of possible pairings into two subgroups – repeated pairs and unrepeated pairs – perhaps as a result of attention being drawn to the repeated stimuli. Such segregation of a large set with many possible pairings (i.e., 4×4=16) into two smaller subsets with a fewer total number of pairings (i.e., 2×2+2×2=8) reduces ambiguity, so it was expected that conditions with higher SR scores would result in increased learning.

Subjects

Participants were undergraduates at Indiana University who received course credit for participating. There were 50 participants in condition SR=0, 36 in conditions SR=.33, .66, 1.0, and 1.41; and 31 in condition SR=2.04. None had participated in other cross-situational experiments.

Stimuli

On each training trial, pictures of four uncommon objects (e.g., a metal sculpture) were simultaneously shown while four spoken pseudowords were serially played. The 72 computer-generated pseudowords are phonotactically-probable in English (e.g., "bosa"), and were spoken by a synthetic, monotone female voice. The 72 words and 72 objects were randomly assigned to four sets of 18 word-object pairings.

On each training trial, the four pictures appeared immediately. After two seconds of initial silence, each pseudoword was played for one second with two seconds of silence between pseudowords, for a total trial duration of 12 seconds. The pseudowords were presented in random order. Each training sequence consisted of 27 such trials, with each 'correct' pseudoword-object mapping occurring 6 times, and other mappings occurring from 0 to 4 times (M =1.5).

Upon completion of each training phase, participants were tested for their knowledge of the 'correct' (i.e. high frequency of occurrence) pairings. On each test trial, a single pseudoword was played and all 18 objects were displayed. Participants were asked to click on the correct object for that pseudoword. Each pseudoword was tested once, and the test order was randomized for each participant and condition. Participants completed four training/test conditions. Block order was counterbalanced.

Instructions

Participants were informed that they would see a series of trials with four pictures and four alien words played in random order. They were also notified that their knowledge of which words go with which pictures would be tested at the end.

Results

Figure 1 shows the overall performance achieved in each condition of Experiment 1. Increasing the degree of successive repetitions does produce increased performance, as predicted (r =.16, t(217)=2.40, p<.05). However, the increases are surprisingly modest relative to what might have been expected a priori. For detailed analysis of each condition, to-be-learned pairs were grouped according to the number of times they had successive repetitions in the sequence (see Table 1). Surprisingly, the relation between performance and SR was often non-monotonic. For example, consider the trial ordering with mean SR=1.0. In this condition, the non-repeated pairs were learned with 45% accuracy, the pairs that overlapped once were learned with only 30% accuracy, and the pairs that overlapped three times were learned with 66% accuracy. In some other conditions, performance was more or less equal for all degrees of repetition. Only in the two conditions of low SR (SR=.33 and SR=.67) did greater successive repetition confer a modest learning advantage. Nonetheless, overall the groups of pairs of different SR were correlated with performance across conditions (r =.11, t(343)=2.02, p<.05).

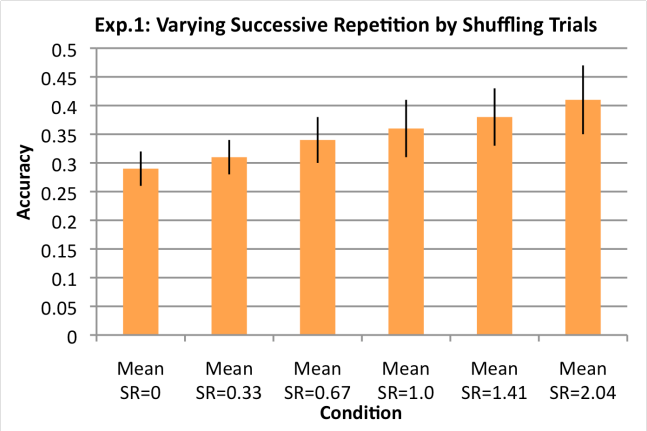

Figure 1: Accuracy (18AFC; chance=.056) for conditions with varying degrees of average successive repetitions (SR). Greater mean SR in a condition tended to improve learning performance, although not as drastically as expected. Error bars are +/-SE.

Table 1: Exp.1 accuracy for SR groups in each condition. Cells display: accuracy / number of pairs in SR group

SR	Mean SR=.33	Mean SR=.67	Mean SR=1	Mean SR=1.4	Mean SR=2
0	.32 /10	.29 / 4	.45/ 2	-	.46/ 4
1	.33 / 7	.37 /10	.30/ 7	.43/ 3	.39/ 1
2	.39 / 1	.43 / 4	.40/ 7	.45/10	.40/ 2
3	-	-	.66/ 2	.36/ 5	.42/ 7
4	-	-	-	-	.40/ 3
5	-	-	-	-	.39/ 1

Discussion

These results show that learning is increased by increasing the average degree of successive repetitions, even while leaving all within-trial co-occurrence statistics constant. However, based on the detailed analysis in Table 1, it is clear that this learning increase was not simply due to increased learning for successively repeated pairs within each condition. For instance, in the SR=1.0 condition, non-repeated pairs were learned better than 1- and 2-repetition pairs. The difficulty of learning a given pair is not solely due to the number of successive repetitions for that pair in a sequence. Perhaps the presence of other repeated pairs interfered with the learning of a given pair. Another relevant factor may be the interaction of spacing and sequential effects: for each case where a pair occurs in successive trials, that pair will appear one fewer time later in training (since each pair only appeared 6 times during training).

Experiment 2

Experiment 1 showed that increasing the mean successive repetitions in a trial ordering facilitates learning for the entire set of pairings, but that greater numbers of successive repetitions does not always confer an advantage for the repeated pairs. Instead, learning of unrepeated pairs seems to improve with the presence of greater overall successive repetitions in the condition. To better understand the role that successive repetitions can play in statistical learning of both repeated and unrepeated pairs, we implemented three different types of temporal contiguity in three sequences of 27 training trials, exemplified in Table 2. In the **1 pair/2 trials** condition, 9 of the 18 pairs appeared in two consecutive trials at some point during the training. In the **1 pair/3 trials** condition, 9 of the 18 pairs appeared in three consecutive trials. Importantly, in both of these conditions, no other stimuli in the overlapping trials simultaneously overlapped. In the **2 pairs/2 trials** condition, however, each of the 18 pairs at some point appeared with another pair in two consecutive trials.

Importantly, these conditions offer different ways to perform inferences and consequently reduce the degree of ambiguity. For example, consider just two successive trials with one pair only repeated, as in the **1 pair/2 trials** condition. The repeated pair may be immediately inferred to be correct, but the remaining six pairs on the two trials remains ambiguous—there are still 9 possible pairings for each of the three remaining words and objects in each trial. For another example, when two pairs are repeated in two successive trials, as in the **2 pairs/2 trials** condition, no pairing can be unambiguously inferred. However, overall ambiguity is considerably reduced: If (*A B X Y; b a y x)* occurred on trial n, and (*A B E F*; *a b e f*) occurred on trial n+1, then memory for the preceding trial allows the following inferences: *(A, B)* and (*b, a*) must go together, in some pairing; (*E, F)* and (*f, e*) must go together, in some pairing; (*X, Y*) and (*y, x*) must go together, in some pairing. Thus no pairing can be unambiguously determined, yet the six items on the two trials have only eight possible

pairings—a significant reduction of ambiguity. Note that in all of these conditions, the conceivable inferences are reliant upon memory and attention.

Table 2: Summary of conditions in Experiment 2.

Trial	1 pair/2 trials	1 pair/3 trials	2 pairs/2 trials
n	**A**BCD, **a**bcd	ABCD, **a**bcd	**AB**CD, **ab**cd
n+1	**A**EFG, **a**efg	AEFG, **a**efg	**AB**EF, **ab**ef
n+2 ..	HIJK, hijk	**A**HIJ, **a**hij	GHIJ, ghij

Subjects

Participants were undergraduates at Indiana University who received course credit for participating. Twenty-three participants completed only the three SR conditions, and an additional 44 participants completed all conditions. None had participated in other cross-situational experiments.

Stimuli & Procedure

The sets of pseudowords and referents for Experiment 2, the number of trials, the number of stimuli per trial, and other details of the procedure were identical to those used in Experiment 1, except that individual trials and their orderings were constructed to be consistent with the three different types of successive repetitions described above.

Results

Figure 2 displays the overall learning performance for each training condition[1] in Experiment 2. Participants learned significantly more successively repeated pairs (M = .47) than non-repeated pairs (M = .27) in the 1 pair/2 trials condition (paired $t(66)=6.93$, $p < .001$), demonstrating that a single successive repetition boosts learning of that pair.

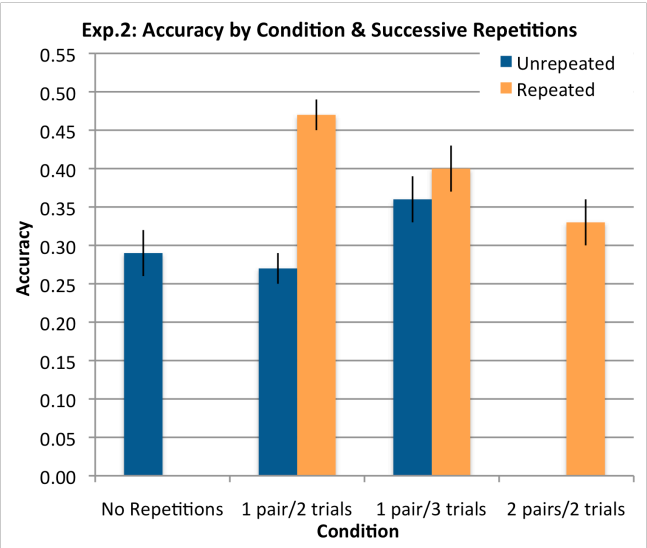

Figure 2: Accuracy (18AFC; chance=.056) for the three conditions in Exp. 2, and the TC=0 condition (no overlaps) from Exp. 1 for comparison. Error bars are +/-SE.

However, performance for repeated pairs (M = .40) in the 1 pair/3 trials condition was not significantly greater than for the non-repeated pairs (M = .36, paired $t(62)=1.31$, $p > .05$). Instead, a higher proportion of non-repeated pairs were learned in the 1 pair/3 trials condition than in the 1 pair/2 trials condition (paired $t(62)=3.24$, $p < .01$). Thus, although there was no SR advantage within the 1 pair/3 trials condition, more non-repeated pairs were learned instead, and overall pair learning in this condition (M = .37) was not significantly different than overall learning in the 1 pair/2 trials condition (M = .33, paired $t(62)=0.98$, $p > .05$). Although each of the conditions with some variety of successive repetition trended toward greater overall performance than the condition with no repetitions, none were significantly greater.

Discussion

It is rather striking that performance dropped for successively repeated pairs from condition 1 pair/2 trials to 1 pair/3 trials to 2 pairs/2 trials, even though the opportunities for unambiguous inference rose from 1 pair/2 trials to 1 pair/3 trials, and then dropped very low for 2 pairs/2 trials. It is equally intriguing that the non-repeated pairs benefited more when one pair repeated over three rather than two successive trials. Our working hypothesis is that this has to do with how statistical learners allocate their real-time attention during statistical associative learning, how the repetitions of certain pairs create a local attentional salience—either for the repeated pairs or for the unrepeated pairs, and how learners dynamically adjust their attentional weights while associating pairings trial by trial (see Kruschke 2003, e.g.). A formal account of this hypothesis requires a computational model to allow us to further investigate these processes that may comprise statistical word learning.

Modeling

Several computational models were constructed and evaluated. We present one model that captures the intuitions we have discussed, and that fits the observed data reasonably well. We fit the model to Experiment 2, and then predicted the results of Experiment 1 with those best-fit parameters. There are several critical principles encoded in the model:

1) There are learning limitations such that the total amount stored in long-term memory per trial is a constant.[2]
2) Observers are assumed to remember the previous trial, and also any item repeated over three successive trials.

[1] Data for four participants in the 1 pair/3 trials condition was lost due to computer failure.

[2] The final choice rule is stochastic, so some of our storage assumptions do not assume variable rules.

3) Based on this memory, association strengths are stored only for possible pairings that respect logical inferences.
4) Observers have access to the current state of their long-term memory and add association strength to a pair that is probabilistically chosen in proportion to the current strengths of present referents.
5) When some pairs repeat over successive trials and others do not, a bias that stores more or less association strength for the repeated subset of pairs than the non-repeated subset is allowed.
   a) This bias is allowed to vary for cases where the repeated pair(s) are just a one trial repeat, and the cases where an item repeats three or more trials in succession. The change in bias allows observers who believe they already 'know' a pairing (because it repeats three times in a row) to divert attention/storage to other pairings.

The model is instantiated as follows (see Figure 3 for pseudocode). An 18 (word) by 18 (referent) associative matrix is filled initially with constant values (1/18 in every cell), reflecting initial uncertainty about the pairings. Every trial, a fixed total amount, X, of additional association strength is added to this matrix. When two successive trials have no repeats, then when a pseudoword is presented an object is chosen in proportion to the strengths currently in the association matrix between that pseudoword and the four objects. For the chosen pairing, X/4 is added to its cell in the matrix. The three succeeding pseudowords are treated similarly, except for the constraint that objects are chosen without replacement, for the current trial. (Sampling without replacement is utilized for all following cases, as well.)

For the case of one repeated stimulus pair, when the word is encountered during the trial, the pair's matrix cell is augmented by $\alpha \cdot X$ (α representing an attentional bias for repeated vs. non-repeated items; $\alpha = .5$ would represent no bias). When any of the three non-repeated pseudowords is encountered, an object is selected in proportion to the current values in the three relevant cells, and that cell augmented by a value $(1-\alpha) \cdot X/3$. Successive non-repeated pseudowords are treated similarly, albeit respecting the sampling without replacement constraint. In the case when the one repeated pair occurs three trials in succession, the same rules apply, but a different value for α is allowed: α'. The intuition is that attention may shift for pairs that are repeated many times: once learned, the regularity of a repeated pair may be used to reduce ambiguity for the remaining pairs.

For the case of two repeated items, the same rules apply, with the same α, save a repeated choice is made from the two repeated objects, with storage $\alpha \cdot X/2$, and a non-repeated choice is made from the two non-repeated objects, with storage $(1-\alpha) \cdot X/2$ (again respecting sampling without replacement).

At test, an object response is probabilistically selected in proportion to the association strengths stored in the matrix, in the row for the tested pseudoword.

N = number of to-be-learned word-referent pairs
M = N×N association matrix, initially filled with 1/N
X = total sum of associative weight distributed per trial
α = attention parameter [0,1] for 1 repetition pairs
α' = attention parameter [0,1] for >1 repetition pairs

Training:
 For each trial T:
  Separate repeated & unrepeated referents:
   {repeated}, {unrepeated}
   nRep = |{repeated}|, nUnrep = |{unrepeated}|
  If any referent was present on the previous two trials,
   Alpha = α'
  Else, Alpha = α
  For each word *w* on T:
   If *w* is repeated, choose* a referent *r* from {repeated}
    M[*w*,*r*] = M[*w*,*r*] + Alpha·X/nRep
   Else, choose* a referent *r* from {unrepeated}
    M[*w*,*r*] = M[*w*,*r*] + (1-Alpha)·X/nUnrep

Testing:
 For each word *w*:
  Choose* a response referent from row M[*w*,]

*Without replacement, and probabilistically according to the associations: M[w, each referent in that subset]

Figure 3: Pseudocode of statistical learning algorithm.

Results

Figure 4 displays the human and model performance on the SR and non-SR pairs in Experiment 2. The model achieves an excellent qualitative fit to the data. The best-fitting parameters were X=1.118, α=0.468, and α'=0.028 (i.e., attention shifts to unrepeated pairs when some pair is oft-repeated). Figure 5 shows the predictions when the model uses these parameters to fit the results for Experiment 1. The presented model does an excellent job, but is tailored to fit the observed data and hence should best be considered descriptive and representative. Future experimentation will be needed to test the assumptions and refine the model. The general lesson for statistical learning is the importance of strategies that are based on reasonable inferences drawn on the basis of memory for the immediately preceding trial (and perhaps earlier), particularly repetitions that provide logical constraints on possible pairings.

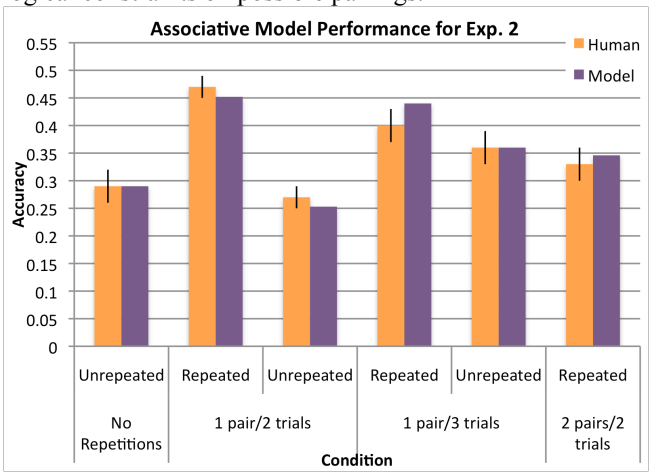

Figure 3: Comparison of the best-fit model performance and human learning in Exp. 2. Error bars are +/-SE.

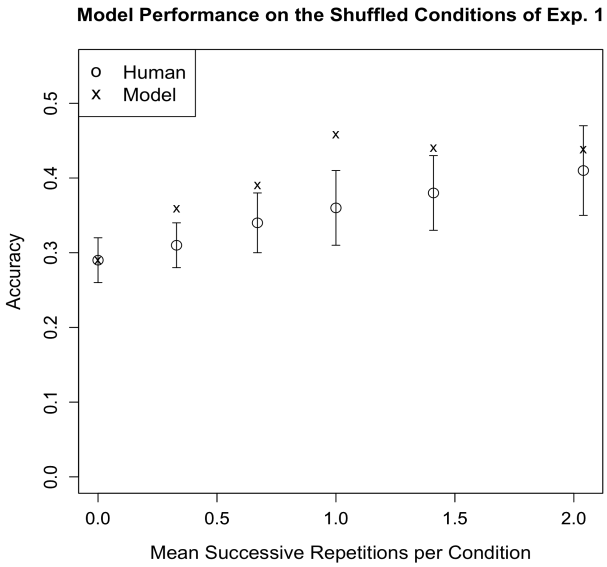

Figure 5: The model's predictions for Exp. 1 using the best-fit parameters from Exp. 2. Although the model slightly outperforms humans, especially for SR=1, the fit is not bad.

General Discussion

Learning from the successive presentation of instances requires that learning in one moment be connected to learning in previous moments. That is, in order for a learner to use a past trial containing the information that *A* is linked to *a* to rule out a link between *A* and *b*, the learning must connect the noticed association on this trial to previously learning. Given all that we know about human memory, the temporal separation of the two learning trials should matter. That is, unlike batch statistical processors, for human learners, the order of learning trials and the temporal contiguity of certain trials should matter. Two experiments in the present study confirmed our hypothesis. Adult learners performed significantly better in the learning conditions with repeated pairings. A closer look of the results in Experiment 1 showed that both repeated and non-repeated words were learned better. This observation suggests two plausible ways that the additional information in temporal continuity may be utilized to facilitate learning. In general, a key mechanism in a statistical associative learner is to decide in real time which word-referent pairs to attend to among all possible ones available in an unambiguous environment. From this perspective, repeated pairs may be temporally highlighted and therefore attract the learner's attention. On the other hand, the novelty of non-repeated pairs may demand attention, causing oft-repeated pairs to fade into the background as attention to them is attenuated (as evidenced by the small α' estimated for the model). These two mechanisms may operate in parallel and dynamically interweave. Indeed, the results in Experiment 2 provide direct evidence to support this proposal – human learners in that study seem to be able to fully take advantage of different types of temporal continuity by developing different computational inferences, suggesting that the learning system seems to be highly adaptive by discovering and adjusting to the most effective way to process the learning input.

Cross-situational statistical learning mechanisms may be criticized on the basis that these processes may not be efficient because they require the accumulation of statistical evidence trial by trial until it is strong enough to disambiguate learning situations. Nonetheless, statistical regularities and physical constraints in the real world may provide more information than the stimuli in our training. In the real world of real physics, there is likely to be considerable overlap between the objects present in a scene from one moment to the next and in the topics of discourse from one moment to next. Natural discourse seems likely—as in our overlapping conditions—to shift incrementally from one trial to the next. If language learners in the real world are sensitive to overlapping regularities as we suspect, cross-situational learning mechanisms which allocate more attention to novel stimuli may be well fit for the task of word learning.

Acknowledgments

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

References

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

Conway, C. & Christiansen, M. H. (2005). Modality constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *31*, 24-39.

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

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.

Kruschke, J. K. (2003). Attention in learning. *Current Directions in Psychological Science*, *12*, 171-175.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University 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. *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.