Rapid Word Learning under Uncertainty via Cross-Situational Statistics

Chen Yu and Linda B. Smith

Department of Psychological and Brain Sciences and Program in Cognitive Science

Indiana University

Address correspondence to:

Chen Yu

Department of Psychological and Brain Sciences

1101 East 10th Street

Indiana University

Bloomington, IN, 47405

Email: chenyu@indiana.edu

Phone: 812 856-0838

Fax: 812 855-4691

Word count: 3923

References: 24

Abstract

There are an infinite number of possible word-to-world pairings in naturalistic learning environments. Previous proposals to solve this mapping problem focus on linguistic, social, and representational constraints at a single moment. This paper investigates a cross-situational learning strategy based on computing distributional statistics across words, across referents, and most importantly across the co-occurrences of these two at multiple moments. We briefly exposed adults to a set of trials containing multiple spoken words and multiple pictures of individual objects with no information about word-picture correspondences within a trial. Nonetheless, subjects learned over trials the word-picture mappings through cross-trial statistical relations. Different learning conditions compared the degree of within-trial reference uncertainty, the number of trials and the length of trials. Overall, the remarkable performances of learners in various learning conditions suggest that they calculate cross-trial statistics with sufficient fidelity and by doing so rapidly learn word-referent pairs even in highly ambiguous learning contexts.

Quine (1960) famously presented the core problem for learning word meanings from their co-occurrence with perceived events in the world. He imagined an anthropologist who observes a speaker saying "gavagai" while pointing in the general direction of a field. The intended referent (rabbit, grass, the field, or rabbit ears, etc.) is indeterminate from this experience. The solution to this indeterminacy problem requires that the learning system be somehow constrained.

Research on children's word learning has concentrated on how this learning might be constrained in a single trial such that the word is correctly mapped to the referent *on that trial*. This literature suggests attentional (Smith, 2000), social (Baldwin 1993, Tomasello, 2000), linguistic (Gleitman, 1990) and representational (Markman, 1990) constraints that enable learners to "fast map" words to referents in a single encounter. However, the indeterminacy problem may also be solved *cross-situationally*, not in a single encounter with a word and potential referent but across multiple encounters and learning trials. A learner who is unable to unambiguously decide the referent of a word on any single learning trial might nonetheless store possible word-referent pairings across trials, evaluate the statistical evidence, and ultimately map individual words to the right referents through this cross-trial evidence. There has been very little systematic investigation of whether human learners do this kind of learning, and if they do, of the nature of the underlying learning processes.

This constitutes a significant gap in current understanding of human learning in general, and in word learning in particular. Not all opportunities for word learning outside of the laboratory are as uncluttered and as constrained as the experimental settings in which fast-mapping has been demonstrated. Instead, in everyday scenarios, there are typically many words, many potential referents, limited cues as to which words go with which referents, and rapid attentional shifts among the many entities in the scene. Such highly ambiguous learning contexts could nonetheless play the dominant role in real-world word learning *if* learners calculate and use statistical information across multiple encounters with words and referents.

Several formal simulations suggest the plausibility of cross-situational word learning (Siskind, 1996; Vogt & Smith, 2005; Yu and Ballard, in press). In these simulations, learners keep track of many words and many referents, accruing evidence over many trials as to the word-referent pairings. Given the infinite number of potential meanings, these cross-situational learning mechanisms must also be constrained and there are a variety of potential constraints that work reasonably well in simulations studies. Further, Akhtar and colleagues (Akhtar, 2002; Akhtar and Montague, 1999) have shown that human learners (children) use information about the labels of two objects within a single learning trial (see also Namy and Gentner, 2002) and that learners will combine information across trials in which a single object is unambiguously labeled prior to an ambiguous trial in order to discover the relevant referent or property (see also Markman, 1990; Carey & Bartlett, 1978). These are both critical components of cross-situational learning. However, there is no evidence as to whether human learners are capable of learning from highly ambiguous contexts involving many words and referents, and whether they are able to compute statistics over many possible word-referent pairs and in doing so narrow in on the right word-referent mappings. The following experiments provide experimental evidence for this learning mechanism in adults.

Experiment 1

Our goal was to capture in a laboratory task some of the complexity and ambiguity perhaps of real-world word learning and to examine adult learners' ability to make word-referent mappings under those conditions. To these ends, we ask adult learners to simultaneously learn relatively many word-referent pairs (18 at a time) from individual learning trials that are highly ambiguous. On each trial, the learner is presented with multiple labels and multiple referents with no information as to which label goes with which referent. The experiment includes three conditions that differ in their degree of within-trial ambiguity: 2 words and 2 possible referents, 3 words and 3 possible referents, and 4 words and 4 possible referents on each trial. The 2×2 condition yields 4 possible word-referent associations per trial. The 3×3 condition yields 9 potential word-referent associations per trial. The 4×4 condition yields the seemingly overwhelming number of 16 word-referent associations per trial.

Although there is no information on any individual trial as to which label goes with which word, the underlying word-referent mappings are certain in that individual labels are present in a training trial if and only if the referent is also present. Can learners keep track of the simultaneous co-occurrences of many labels and referents across trials such that they can learn these mappings? Is this easily accomplished in relatively few learning trials, from (relatively) few highly ambiguous exposures to each individual word?

The key ingredient to learning from highly ambiguous individual trials would seem to be the keeping track and comparing of information across trials. This is illustrated in Table 1 for the case in which the learner hears two words while viewing two objects (with neither spatial nor temporal cues linking the words to particular referents). On trial 1, a learner could mistakenly link word A to referent b (and possibly also link word A to referent a). Notice, however, that on trial 4, this mistake can be corrected, the cognitive system can *rule out* A-b as a possible word-referent pair *if* the system registers that word A occurred on trial 4 *without* possible referent b. If the cognitive system remembers prior word-referent pairings, if it registers both co-occurrences and non co-occurences, and if the system calculates the right statistics, it should be able to learn as many as 18 word-referent pairs from relatively few and highly ambiguous individual learning trials.

Insert Table 1 about here

Method

Participants. 38 students at Indiana University received course credit or $7 for their participation.

Stimuli and design. The stimuli were slides containing pictures of uncommon objects (e.g. canister, facial sauna, and hitch haul) paired with auditorally presented pseudowords. These artificial words were generated by a computer program to sample broadly phonotactically-probable English forms and were produced by a synthetic female voice in monotone. There were 54 unique objects and 54 unique pseudowords partitioned into three sets of 18 words and referents for use in the three conditions. The training trials were generated by randomly pairing each word with one picture; these are the word-referent pairs to be discovered by the learner. Three learning conditions differed in the number of words and referents presented on each training trial. In the 2×2

condition, each trial presented 2 words and 2 pictures. There was no indication of which picture went with which word. In the 3 × 3 condition, each trial presented 3 words and 3 pictures and in the 4 × 4 condition, each trial presented 4 words and 4 pictures. In all conditions, on each trial the referents were visually presented simultaneously on the screen to start the trial. The names were then presented auditorally over the loud speakers. The temporal order of the spoken names was not related in any systematic way to the spatial location of the referents.

To form each trial, we randomly selected several (2, or 3, or 4, depending on conditions) word-referent pairs from the 18 word-referent pairs for that condition such that across trials in a condition, each word and referent were presented 6 times. That is, over training trials, the learner would experience 6 repetitions of each word-referent pair. However, given that multiple words and references are presented on each trial, the learner will experience other "spurious" associations that might be expected to make learning from these ambiguous individual trials difficult. Specifically, on average each word co-occurred with 5 other wrong referents in the 2 × 2 condition, 8.78 wrong referents in the 3 × 3 condition, and 12.22 wrong referents in the 4 × 4 condition, which is proportional to within trial ambiguity. The probability, during training, of the correct referent given its name $p(a|A)$ was 1.0 in all conditions whereas the average probability of irrelevant but co-occurring referents was 0.205, 0.231, and 0.247 in the 2 ×2, 3 × 3, and 4 × 4 conditions respectively. Notice that despite the considerable differences in within-trial uncertainty across conditions, the strength of the spurious correlations varies only moderately across these conditions.

Because the same number of word-referent pairs (18) is taught in each condition, and because we sought, across conditions, to keep the number of exposures to each word-referent pair constant, other presentation factors necessarily varied across conditions. These are summarized in Table 2. Across conditions, the number of repetitions of each unique word and referent, and the total time of the training session were kept constant and thus the number of total trials differs across the three conditions as does the duration of each trial. Order of trials within a condition was randomly determined. Order of the three conditions (a within-subject manipulation) was counterbalanced across subjects.

Insert Table 2 about here

Procedure. The pictures were presented on a 17 inch screen and the sound was played by the speakers connected to the same PC. Subjects were instructed that their task was to learn the words and referents but they were not told that there was one referent per word. They were told that multiple words and pictures would co-occur on each trial and their task was to figure out across trials which word went with which picture. After training in each condition, subjects received a four-alternative forced-choice test of learning. They were presented with one word and four pictures and asked to indicate the picture named by that word. The target picture and the three foils were all drawn from the set of 18 training pictures.

Insert Figure 1 about here

Results and Discussion

Figure 1 shows that participants learned more word-referent pairs in each condition than expected by chance ($t(37)=8.785$, $p<0.001$, $p_{rep} > 0.99$, $d = 1.425$, one-tailed, for 4×4). They discovered on average more than 16 of the 18 pairs in the 2×2

condition and more than 13 of the 18 pairs in the 3×3 condition, all this in less than 6 minutes of training per condition. Even in the 4×4 condition with 16 potential associations per trial, subjects discovered almost 10 of the 18 word-referent pairs. Indeed, 9 subjects discovered over 75% of the pairs in this condition. The level of performance in the three conditions is remarkable; in a very short time, over relatively few trials, each highly ambiguous, subjects nonetheless find the underlying word-referent pairs. The degree of within-trial uncertainty clearly matters ($F(2,74)=76.069$, $p<0.001$, $p_{rep}< 0.99$, $\eta^2 = 0.631$). But just as clearly, subjects calculate cross-trial statistics with sufficient fidelity that they are able to acquire a significant number of word-referent associations and demonstrate this knowledge at test, despite the ambiguity of the individual learning trials.

The present experiment was designed as a first demonstration of the general viability of cross-situational learning given highly ambiguous individual learning trials, and as such can not specify the precise mechanisms that underlie this learning. Relevant to these mechanisms -- and to what participants actually learned from the experience – are the spurious correlations. At test subjects were presented with four alternatives – the correct referent for the tested word and three alternatives. It seems highly likely that at test participants simply chose the most strongly associated item among the presented alternatives. If subjects were able to track *all* word-referent co-occurrences in the training trials, they should be able to respond perfectly in all conditions, since the association between the correct word and referent (the p(referent|word)) is 1.0 and much greater than even the strongest spurious correlation in the training sets. However, if learners are keeping track of all co-occurrences during learning and if they are choosing the alternative most associated to the tested word, then if errors do occur, they should be related to the spurious correlations that arise given the presentation of multiple words and referents on specific learning trials. Specifically, errors should be related to what we call foil probability, the probability that the foil at test had co-occurred with the word during training. For example, a foil that had occurred with the tested word on 3 of the 6 repetitions of that word during training should be wrongly selected more often than a foil that occurred with the tested word only 1 test trial. The probability that the foils at test were spuriously associated with the tested word was, on average, not high:.056, .115, and .155 in the 2×2, 3×3, and 4×4 conditions respectively.

Because the probability that a tested foil had been associated with a target word was both greatest and most variable across foils in the 4 x 4 condition and because participants in this condition made the most errors, we more closely examined the relation between choices of foils and their strength of association with the tested word. Table 3 shows the probability that subjects chose the tested foils as a function of their association to the tested word (accumulated across multiple test trials and subjects). There appears to be little systematic relation. A strong conclusion that foil probability does not matter, however, is not warranted as the strength of spurious associations (and thus associations of foils to test words) was overall quite low. The low level of spurious correlations is the natural result of a large training set (and the random selection of co-occurring pairs during training) which may well also describe real-world word learning. We pursue the issues of spurious correlations, foil probability, and size of the data set in Experiment 2.

Insert Table 2 about here

What have participants learned from these brief experiences? Word-referent pairings are uncertain within a trial but (if participants can track all the information) certain across trials. However, given the real-time processing demands of attending to and remembering many words and referents and the relatively brief training regimen, participants' knowledge of most word-referent pairs may not have been certain. Indeed, many participants volunteered (quite wrongly) prior to test that they were sure they knew none of the pairings. Thus, participants may well have NOT learned, for example, that word A maps only to referent a. Rather, participants' knowledge may have been more of the form: word A is associated with referent a and b but not with anything else. Such partial knowledge can explain the present results. Such partial knowledge could also play a powerful role in real-world word learning. Our main point is this: The acquisition of this kind of knowledge (albeit potentially imperfect) requires calculations on cross-trial co-occurrences. Experiment 1 shows that adult human learners do this for relatively many word-referent pairs and despite the within-trial uncertainty of the pairings. Cross-situational statistical learning is within the repertoire of human learners.

Experiment 2

Experiment 2 was designed to replicate the findings in Experiment 1 and in particular to further explore learning in conditions of high within-trial ambiguity as a function of the number of word-referent pairs to be learned. Accordingly, in this experiment, each condition is a version of the original 4 x 4 condition of Experiment 1. We manipulate: (1) the total number of word-referent pairs to be learned and (2) the number of repetitions of each word-referent pair. In the 9 words/8 repetitions condition, subjects attempt to discover a total of 9 word-referent pairs each repeated 8 times over the course of training. In the 9 words/12 repetitions condition, subjects attempt to discover 9 word-referent pairs but are given 4 additional repetitions of each word-referent pair. Finally, the third condition is a replication of the 4×4 condition of Experiment 1; there are 18 word-referent pairs to be learned and 6 repetitions of each.

Intuitively, the 9 words/12 repetitions condition should improve the learning performance because, compared with the 18 words/6 repetitions condition, the number of words needed to be learned are reduced while their occurrence frequencies are doubled. However, for statistical learners smaller data sets may not be as good as large ones because spurious correlations are more likely to occur.

Method

Participants. 28 students at Indiana University received course credits for their participation. None participated in Experiment 1.

Stimuli and procedure. All aspects of the experiment are identical to Experiment 1 except for the composition of the three training conditions. All three conditions use the 4×4 presentation of 4 words and 4 pictures on each trial but differed in the number of word-referent pairs to be learned (9, 9 and 18) and in the number of repetitions of each word-referent pairs (12, 8 and 6, respectively). These conditions are summarized in Table 4. The random selection of co-occurring word-referent pairs during training and the random selection of foils at tests lead naturally to differences in foil probabilities, that is, in the associations of alternatives at test with the tested word. The foil probabilities are higher when 9 word-referent pairs are to be learned (0.375 in both 9 pair conditions) and lower (0.247) when 18 word-referent pairs are to be learned.

Insert Table 4 about here

Insert Figure 2 about here

Results and Discussion

The three conditions present equivalent within-trial uncertainty but differ in the number of word-referent pairs. In terms of the proportion of word-referent pairs to be discovered, participants performed comparably in the three conditions, ($F(2,54) = 0.52$; $p > 0.5$; $p_{rep} = 0.42$, $\eta^2 = 0.03$), discovering more pairs than expected by chance as shown in Figure 2 ($t(27) > 6.4$ in all three conditions, $p < 0.001$, $p_{rep} > 0.99$, $d = 1.249$). Again, adult learners acquire lexical knowledge from highly ambiguous exposures to word and potential referents. Considering the results from both Experiments 1 and 2 together, this suggests that within-trial uncertainty is a more critical factor in learning than the number of pairs in the learning set.

In terms of total number of pairs learned, subjects actually learned more pairs in the 18 word-referent condition (M=9.461, SD=2.907) than in the two 9 word-referent conditions (8 repetitions: M=5.111, SD=1.706; 12 repetitions: M=5.481, SD=2.089). The 18 word condition presents the same within-trial ambiguity, more word-referent pairs to be learned, and fewer repetitions of the individual word-referent pairs than the other two learning conditions. If numbers of co-occurrences were all that mattered, this condition should lead to the poorest overall performance. The advantage lies in the fewer spurious correlations (and thus also in the lower foil probabilities at test.) Herein lies the power of cross-situational statistical learning: Even when the referent of a word cannot be unambiguously determined on any single learning trial, across multiple trials involving many different words and many different potential referents, the word and referent will occur most systematically than any other. The more words and referents there are to learn and that may co-occur together on any learning trial, the more discernible the systematicity –across trials – of the underlying correct mappings. Could bigger lexicons (more pairs) really be easier to learn than smaller ones? Learning requires multiple processes, some of which (for example, memory for particular items and attention) will be negatively impacted by the size of the learning set. However, within these constraints, statistical learning of a *system* of word-referent pairs may well benefit from larger as opposed to smaller data sets.

General Discussion

There is no doubt that human learners (including young children) "fast map" names to things by solving the indeterminacy problem in a single trial – linking a novel word correctly to the intended referent through the use of social (Baldwin, 1993; Tomasello,2000; Bloom, 2000), linguistic (Gleitman, 1990), attentional (Smith, 2000), and conceptual constraints (Gentner, 1982). The present results suggest the importance of an additional kind of learning that does not require such in-the-moment certainty but instead allows for substantial learning from far more ambiguous learning environments in which the correct mapping of a word to an intended referent cannot be guaranteed.

The robustness of the learning given brief training suggests a possible role for cross-situational learning in vocabulary development. Studies of that learning environment indicate that parents, on average, direct between 300 to 400 words *an hour* to their children (Hart & Risley, 1995). So many words in so little time seem likely to generate considerable ambiguity about intended referents. Yet this kind of learning environment with much in-the-moment ambiguity may, precisely because of the sheer amount of statistical data provided, yield considerable word learning. The present experiments constitute a first step in understanding the role of cross-situational statistical learning by showing robust learning of relatively many words from the co-occurrence data available in brief exposures (less than 6 minutes). The findings are reminiscent of recent evidence on adults' and infants' ability to discover segmental units in the sequential probabilities of sounds or visual events (Saffran, Aslin, & Newport, 1996; Gomez & Gerken 1999; Kirkham, Slemmer, & Johnson, 2002, Conway & Christiansen, 2005). Like the present results, the findings on learning sequential probabilities and segmentation suggest that the solution to fundamental learning problems central to language may be found by studying the statistical patterns in the learning environment and the statistical learning mechanisms in the learner (Saffran, Newport, & Aslin, 1996; Newport & Aslin, 2004).

What is the mechanism that gives rise to the observed effects? One possibility is a simple associative process that counts the number of co-occurrences and on test trials chooses the object most strongly associated with the test word. Alternatively, more complicated associative models that include competition and inhibition of competing associations might be required. Finally, statistical learning that explicitly compares alternative hypotheses and rules out wrong hypotheses might be needed to generate the fast learning of so many word-referent pairs from such minimal training data. These questions cannot be answered without formal modeling, the next step in our research agenda.

Relevant to all of these mechanisms are the kinds of constraints that must be imposed on the system to account for human learning. In the present experiments, the subjects were not explicitly told there was one word for each picture. Nonetheless, their post-experiment comments indicate that they almost adopted a one-word, one-object strategy, what is known as the mutual-exclusivity assumption in the child word-learning (Markman, 1990; Clark 1987). Does the learning mechanism *require* this constraint to succeed? Is it an *explicit* hypothesis testing strategy? Subjects' post-experiment comments indicate that many were quite sure that they had learned nothing from the training and were amazed at their own success. This suggests that cross-situational learning may go forward nonstrategically and automatically, steadily building a reliable lexicon.

A further critical question is the availability of these cross-situational learning mechanisms to infants and young children. This seems highly plausible in that considerable research suggests strong continuity in general learning mechanisms in infants, children and adults (Gillette, Gleitman, Gleitman, & Lederer, 1999). At the very least, the present results point to the value of the systematic study of the cross-situational learning and its mechanisms. In conclusion, the human learning environment is data rich. Past analyses questioned the quality of that data for language learning (Quine, 1960) because each datum is highly ambiguous in and of itself. But the data set as a whole –if human learners possess the right learning mechanisms – may readily solve this indeterminacy problem. The present results suggest that human learners may well possess these needed mechanisms.

Table 1: an example of cross-situational learning

Trial	Words	Potential referents in scene
1	A B	b a
2	C D	d c
3	E F	e f
4	G A	g a

Table 2: the statistics of the stimuli in three learning conditions in Experiment 1

condition	# of total words	# of occurrences per words	# of trial	time per trial (sec)	total time
2 × 2	18	6	54	6	324
3 × 3	18	6	36	9	324
4 × 4	18	6	27	12	324

Table 3: the correlation between incorrect answers and foil probabilities in the 4 $\times$ 4 condition of Experiment 1.

foil probability	Probability to select incorrect answers
0	0.162
1/6	0.216
2/6	0.222
3/6	0.077

Table 4: the statistics of the stimuli in three learning conditions in Experiment 2.

Condition	# of total words	# of occurrences per words	# of trial	time per trial (sec)	total time
9 words/8 repetitions	9	8	18	12	216
9 words/12 repetitions	9	12	27	12	324
18 words/6 repetitions	18	6	27	12	324

References

Akhtar, N. (2002). Relevance and early word learning. *Journal of Child Language, 29*, 677-686.

Akhtar, N., & Montague, L. (1999). Early lexical acquisition: The role of cross-situational learning, *First Language, 19*, 347-358.

Baldwin, D. (1993). Early referential understanding: Infant's ability to recognize referential acts for what they are. *Developmental psychology* (29), 832-843.

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

Carey, S. & Bartlett, E. (1978). Acquiring a single new word. *Proceedings of the Stanford Child Language Conference, 15*, 17-29. (Republished in Papers and Reports on Child Language Development 15, 17-29.)

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.

Clark, E.V. (1987). The Principle of Contrast: a constraint on language acquisition. In B. MacWinney (Ed.), *Mechanisms of language acquisition* (pp. 1-33): Hillsdale, NJ: Lawrence Erlbaum Associates.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj II (Ed.), *Language development* (Vol. 2). Hillsdale, NJ: Erlbaum.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition, 73*, 135-176.

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

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*(2), 109-135.

Hart, B., & Risley, T. R. (1995). *Meaningful Differences in the Everyday Experience of Young American Children.* Baltimore, MD: Brookes Publishing.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*, B35-B42.

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

Namy, L. L,. & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology: General, 131*, 5-15.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*(2), 127-162.

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

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word Segmentation: The role of distributional cues. *Journal of memory and language, 35*, 606-621.

Siskind, J.M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition, 61*, 39-61.

Smith, L. B. (2000). How to learn words: An Associative Crane. In R. Golinkoff & K. Hirsh-Pasek (Eds.), *Breaking the word learning barrier* (pp. 51-80): Oxford: Oxford University Press.

Tomasello, M. (2000). Perceiving intentions and learning words in the second year of life. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 111-128): Cambridge University Press.

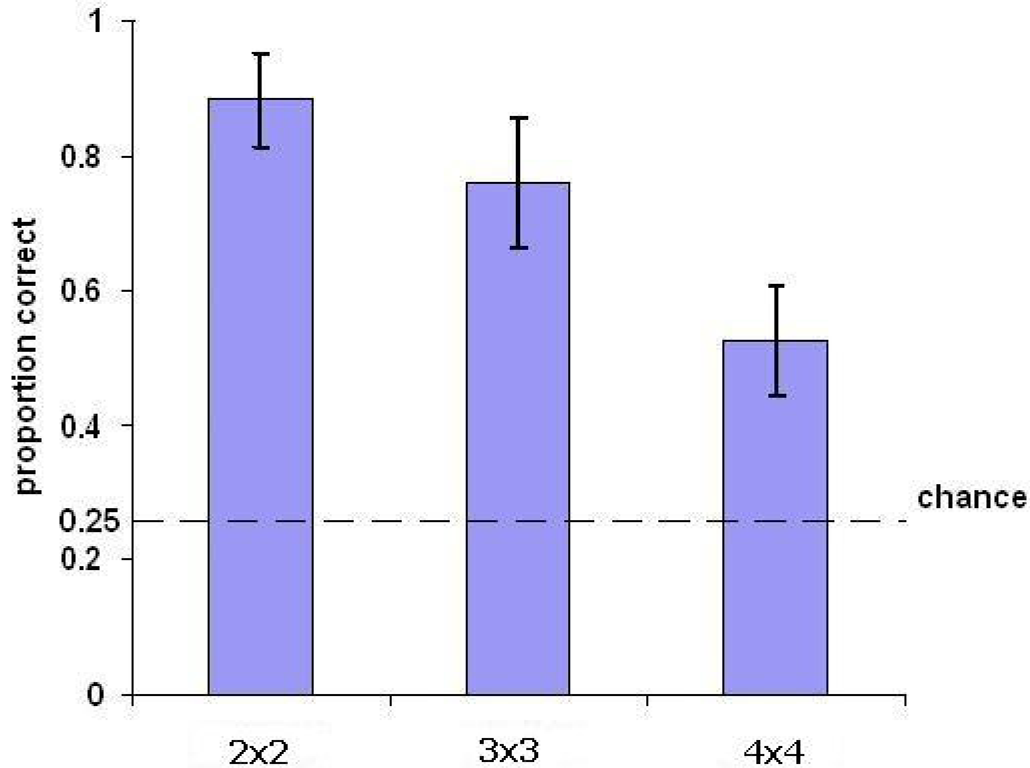
Vogt P. & Smith A.D.M. (2005). Learning colour words is slow: a cross-situational learning account. *Behavioral and Brain Sciences 28(4)*: 509-510.

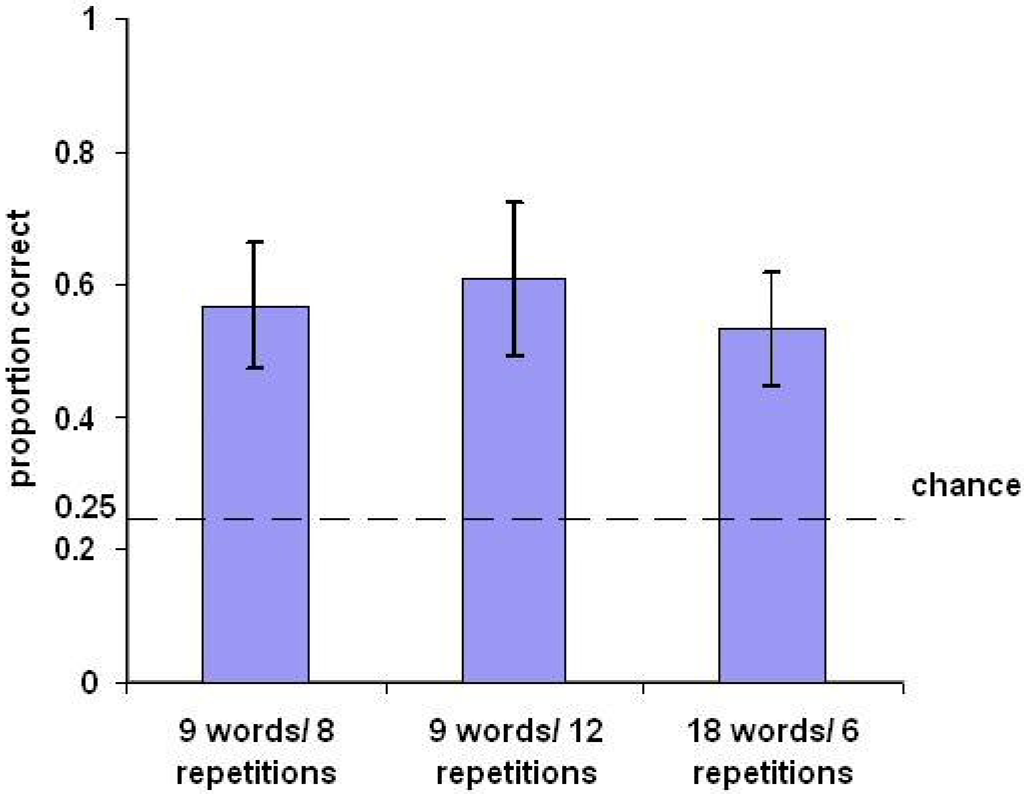
Yu C. & Ballard D. H. (in press). A unified model of early word learning: integrating statistical and social cues. *Neurocomputing*.

Figure captions

Figure 1: The results of three learning conditions in Experiment 1. Error bars reflect standard errors.

Figure 2: The results of three learning conditions in Experiment 2. Error bars reflect standard errors.


1
0.8
0.6
0.4
0.25
0.2
0
proportion correct
chance
2x2
3x3
4x4


1
0.8
0.6
0.4
0.25
0.2
0
proportion correct
chance
9 words/ 8 repetitions
9 words/ 12 repetitions
18 words/ 6 repetitions