Integration and Inference: Cross-situational Word Learning Involves More than Simple Co-occurrences

Alexa R. Romberg and Chen Yu
{aromberg, chenyu}@indiana.edu
Department of Psychological and Brain Sciences, 1101 E. 10th Street
Bloomington, IN 47405 USA

Abstract

Statistical word learning involves forming and aggregating associations between words and objects that co-occur across contexts (e.g., Vouloumanos & Werker, 2009; Smith & Yu, 2008; Yu & Smith, 2007). However, the mechanisms that support such learning are currently under debate, including the extent to which learners carry forward multiple ambiguous associations (e.g., Trueswell et al., 2013). The current study presented adults with a set of statistical word learning tasks designed to measure the statistical computations learners employ to build label-object mappings and to probe what information from past contexts is available to further process and integrate with new information. Results reveal that learners use the co-occurrence of label-object pairings to make inferences both about objects and labels currently present and those presented on previous trials. Further, the strength of learners' memory for past contexts moderated their inferences, suggesting a role for a rich information structure in cross-situational word learning.

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

Introduction

Imagine an infant on a walk with his father. The father, like many parents, comments on the things they see together: "There's a doggie and a kitty in the window!" and a few moments later: "Look, the man is walking the doggie!". How might the father's comments help the infant learn the meanings of words like *doggie*, *kitty* and *man*? Recent research has demonstrated that learners readily form label-object mappings by gathering co-occurrence statistics. Human infants (Smith & Yu, 2008; Vouloumanos & Werker, 2009), children (Scott & Fisher, 2011) and adults (Kachergis, Yu & Shiffrin, 2012; Suanda & Namy, 2012; Yu & Smith, 2007) are all capable of converting multiple individually ambiguous learning instances into specific knowledge as demonstrated by above-chance performance on a post-learning test or by an improvement in selection of the correct referent in a combined training and test procedure (Trueswell, Medina, Hafri & Gleitman, 2013). However, the precise ways in which learners resolve the local ambiguities have been relatively unexplored.

Specifying how exactly learners use the information available is an important step to understanding the mechanisms contributing to success. When learners perform some computations but not others, this offers important constraints to any model of their learning and can inform discussion about the nature of the information stored. In the context of cross-situational word learning, two primary mechanisms, associative learning and hypothesis testing, have been proposed for how learners accrue information over time. These mechanisms differ largely in the amount of information stored and, consequently, in how prior information influences later learning (Yu & Smith, 2012). In particular, associative learning proposes that learners form multiple associations between the objects and labels present during each learning instance, storing a relatively rich information network. Hypothesis testing proposes that learners store only a single link between a label and possible referent, discarding other co-occurrence information. Distinguishing these possible mechanisms has been challenging thus far because of a lack of data regarding how learners process information on a trial-by-trial basis.

Details about what information learners store and how they use it during cross-situational word learning is vital for advancing theories of this process. The typical cross-situational word learning experiment uses a fairly large novel vocabulary (up to 18 to-be-learned label-object mappings) and consists of a series of trials that each present a subset of the labels and objects. Thus, the learner is faced with the difficult task of tracking these many labels and objects across trials (typically between 27 and 60 trials) and using what co-occurrences they can glean to generate as many correct mappings as possible. While this experimental design is daunting for the participant, it is also daunting for the experimenter, as there are inevitably many possible paths to success. One cannot know definitively how participants arrived at a particular mapping over the course of statistical learning or whether the same types of computations were used for all learned mappings.

The present study sought to alleviate these analytical ambiguities for the experimenter while maintaining the learning ambiguities for the participants. Rather than have participants view many trials across which to learn many mappings, learners were presented with a series of "miniature" cross-situational word learning tasks. These tasks consisted of only 2 or 3 trials and were constructed so that some, though not all, label-object mappings could (theoretically) be disambiguated, depending on which information learners stored and which inferences they made. The miniature tasks were constrained so that there was only one pathway to disambiguation, allowing us to infer the computations successful learners employed.

We focused on three fundamental processes that could serve as building blocks for sophisticated statistical learning. The first was the tracking of co-occurrence information – noticing that some labels and some objects appear together across multiple trials. The simplest of statistical learning models, such as the "dumb associative model" outlined by Yu & Smith (2012) do only this co-occurrence tracking to fill an association matrix. The second process was "forward integration", by which learners use information that they carry forward across trials about some objects and labels to make an inference about another object-label mapping. Mutual exclusivity (Markman, 1990) would be a strong form of forward integration, when one rules out objects with known labels as possible referents for a novel label. Recent evidence indicates that learners do employ mutual exclusivity during cross-situational word learning and that this type of inference could arise through basic attentional processes (Kachergis, Yu & Shiffrin, 2012). The third process was "backward inference," by which learners use information on the current trial to infer something about an object/label experienced on a previous trial (but not on the present trial). This last process can be thought of as learning from negative evidence, as it entails noting the absence of particular objects and labels.

We compare performance on three different "miniature" cross-situational word learning tasks to assess learners' ability to use the available information in the three processes of co-occurrence tracking, forward integration and backward inference. The tasks were designed to look specifically at how trial-by-trial information is retained and processed. We also relate performance on the miniature tasks to a "full" cross-situational word learning task, to investigate whether these fundamental processes are also employed in larger-scale statistical word learning.

Method

Participants

Participants were 38 undergraduates (20 females) at Indiana University who earned course credit for their participation. The mean age was 20.9 years.

Materials

The auditory stimuli consisted of 108 nonce words synthesized with the Ivona voice Jennifer using the TextSpeaker program. Nonce words consisted of one or two syllables (264 ms to 795 ms in duration) and followed English phonotactics. The visual stimuli were 123 color photographs of real objects or 3D models either of novel objects or objects that were not readily nameable. Images were displayed in the 4 corners of a 17" monitor, on a white background at a size of approximately 3" square.

Experimental Design

There were 3 types of "mini-tasks" (see Figure 1), each made up of 2-3 training trials and then 3-4 test trials. Each mini-task was independent of the others and no stimuli were repeated across tasks. When objects and labels were repeated on multiple trials they were always presented in different spatial or temporal positions.

Across all parts of the experiment, each trial consisted of viewing 4 objects on a screen and listening to 4 nonce words. Each trial began with the objects displayed in silence for 3 seconds. The onsets of the words were 3 seconds apart and the total trial length was 15 seconds. Every time an object was on screen the corresponding label was provided.

The training trial structure of each of the 3 types of mini-tasks is given in Figure 1**.** For all tasks, the R items refer to the object-label pairs R_1, R_2 and R_3, which were presented on Trials 1 and 2. T1 refers to the object-label pair presented on Trial 1 but not Trial 2. T2 refers to the object-label pair presented on Trial 2 but not Trial 1. The Base and Familiar Context tasks each consisted of a total of 5 word-object pairings: R_{1-3}, T1 and T2. The Novel Context task consisted of a total of 8 word-object pairings: R_{1-3}, T1 and T2 and the 3 novel, label-object pairs presented only on Trial 3 (N_{1-3}).

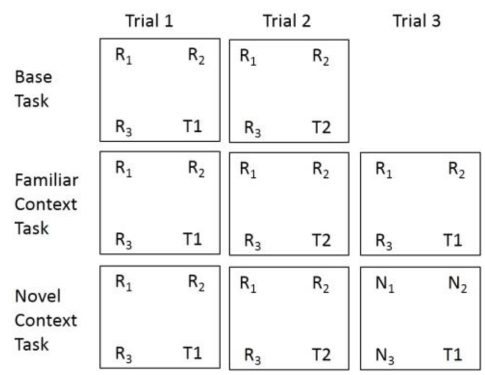

Figure 1. Schematic representing the training trial structure for the 3 mini-tasks, with letters representing objects filling different roles in the design.

For each of the mini-tasks, the training trials were followed immediately by a series of test trials. On each test trial one word was presented auditorally and participants were instructed to click on the object the word most likely referred to out of all objects presented on the task plus a novel distracter object. For the Base and Familiar Context tasks, participants selected from 6 objects and for the Novel Context task participants selected from 9 objects. The tested words came from the different categories of items in the task (R, T1, T2 and, for Novel Context only, N). While the tested items aligned structurally across tasks, the information available to participants differed, enabling us to test hypotheses about what information participants track and what inferences they make.

All participants also completed a "full" cross-situational word learning task, based on Yu & Smith, 2007, which consisted of 18 label-object pairings. These were presented 4 at a time across 27 training trials, so that each label co-occurred with its referent object 6 times. With the 4x4 design, objects co-occurred with other labels, but such "spurious" correlations were limited to no more than 3 times across the 27 trials. Training was followed immediately by 18 test trials. On each test trial all 18 objects were displayed, one auditory label was presented and participants selected the best referent by mouse click.

Procedure

Participants were given an overview of the experiment and informed consent was obtained. All participants first completed the Full CSL task. They were told that they there were 18 words and 18 objects, that they would see them 4 at a time and that the order of the labels on any trial did not correspond to the spatial location of the objects. They were instructed to learn as many label-object mappings as they could. Once participants completed the test for the Full task, they moved on immediately to the mini-tasks. There were a total of 15 mini-tasks, 5 of each task type. The tasks were grouped so that there was one of each type in each block of 3. The order of the 15 tasks was the same across all participants but the order of the test questions within each task were randomly determined for each participant. In the instructions for the mini-tasks, participants were told they would see a series of 15 tasks that were miniature versions of what they had just done and that they would be tested after just 2 or 3 training trials. They were told that no objects or words would be repeated across the mini-tasks. Participants were tested one at a time and listened to the auditory stimuli over headphones. The entire experiment took approximately 30 minutes.

Predictions

The Base task provides a baseline measure of each of the three processes we are examining: The R items represent the co-occurrence tracking process. For each of the tasks, the precise object-label mappings within this group remain ambiguous. However, successfully tracking the repetition of this group of objects and labels theoretically enables learners to perform two types of inference to disambiguate the T1 and T2 mappings. The T2 items represent forward integration: whether participants can use the familiarity of the R_1, R_2 and R_3 pairs on Trial 2 to make a mapping between the relatively novel T2 label and object. Finally, the T1 items represent backward integration: whether participants use the absence of T1 on Trial 2 to make a mapping between that label and object.

Backward integration relies on participants remembering the T1 pair across multiple ambiguous trials and was expected to be difficult. Thus, the other two mini-tasks were designed to test participants' memory for T1 by presenting it in either a novel or familiar context. This necessarily changes the interpretation of the T1 pair in the Novel and Familiar Context tasks, as participants no longer need rely solely on backward integration to learn the mapping.

In the Novel Context task, T1 is presented with 3 new objects and labels on the 3^{rd} trial. This task is the only task in which the association matrix distinguishes the T1 mapping, enabling a correct mapping if participants recognize T1 from the first trial. It is also possible that learners could employ forward integration, mapping the familiar-looking object to the familiar-sounding label without any memory specifically linking the two. If, however, participants do not retain any memory of T1, they should choose randomly from T1 and N_1, N_2 and N_3 on both the T1 and the N test trials.

Unlike the Novel Context task, the Familiar Context task does not provide any additional statistical certainty relative to the Base task. While participants get an additional T1 pairing, it occurs with the same items on both trials. However, participants could infer the T1 mapping by using forward integration in the same manner as the T2 item on Trial 2. Comparing performance between the Base and Familiar Context tasks provides further insight into how learners track information. In the most straightforward extrapolation from the Base task, accuracy on R and T1 should improve due to the extra trial and accuracy on T2 should decrease due to the extra trial between when T2 is presented and tested. Further, within the Familiar Context task, accuracy on T1 is expected to be higher than T2, as participants can use the same process to infer them and T2 is presented on the last trial of the experiment.

Results

All objects from the ambiguous groups ($R_{1\text{-}3}$ and $N_{1\text{-}3}$) were scored as correct. The baseline for chance performance varied between test items and between tasks. For the Base task and Familiar Context task participants selected from 6 objects, so chance performance was 50% for R test trials and 16.7% for T1 and T2 test trials. For the Novel Context task, participants selected from 9 objects, so chance performance was 33.3% and 11.1%, respectively. Statistical comparisons between trial types and tasks were performed with logistic mixed-effects models with random effects of subject (other random effect structures were tested but in no case improved model fit).

Forward integration and backward inference

We first address performance on individual mini-tasks before turning to relationships between the mini-task and full task and comparisons between mini-tasks. Mean accuracy for each type of test item is shown in Figure 2. The results from the Base task reveal that learners do engage the three processes it was designed to test: co-occurrence tracking, forward integration and backward inference. Each of the three trial types has accuracy significantly above chance performance (see confidence intervals on figure). While forward integration accuracy was quite high, our prediction that backward inference would be relatively challenging was confirmed, with participants performing significantly better on T2 items than T1 items on the Base task (b=1.898, z=7.88, $p < 0.001$).

Results from the Novel Context and Familiar Context tasks point to the robustness of co-occurrence tracking and forward integration. In the Novel Context task, neither R items nor T2 were presented in the final trial, so learners must maintain that information while concurrently learning about additional objects and labels. Despite this challenge, participants were significantly above chance on both R and T2 items for the Novel Context task (see Figure 2). In the Familiar Context task, T2 information must be maintained while familiar objects and labels from Trial 1 are repeated in Trial 3. Again, learners were quite successful, performing significantly above chance. Surprisingly, there was no decrement in performance for T2 from the Base to the Familiar Context task (see further discussion below).

While accuracy was significantly above chance for backward integration T1 items on the Base task, it was not very high. Backward integration relies on memory for the T1 pair, as the inference must be made in the absence of the object and label. Remembering the T1 label and object may pose a particular challenge since the mapping between them is ambiguous when they are first presented; it is possible that rapid decay of this information is responsible for the relatively low performance on backwards inference.

However, results from the Novel Context task demonstrate that participants recognized the T1 pair as familiar on Trial 3 and distinguished it from $N_{1\text{-}3}$. Indeed, accuracy on T1 is numerically much higher than the Base and Familiar Context tasks even though the chance baseline is lower. The error pattern also suggests that participants were not likely to confuse T1 and $N_{1\text{-}3}$. On N trials, participants selected T1 only 4.7% of the time, less than they selected T2 (10%), which did not co-occur with the N group. On T1 trials, participants selected one of $N_{1\text{-}3}$ 18.4% of the time, less than is expected for random guessing (33.3%) and much less than they selected T1 (65.3%).

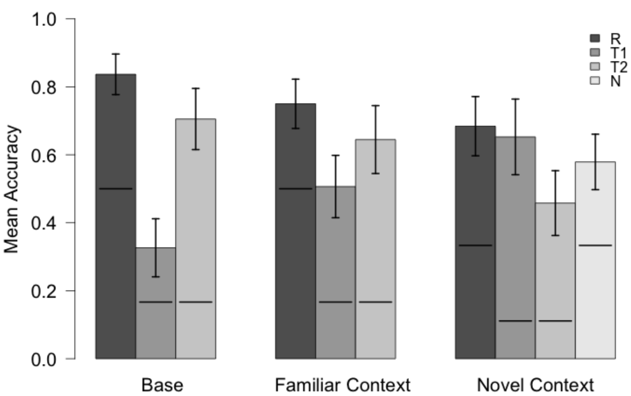

Figure 2. Mean percentage correct for each of the three mini-tasks. The error bars are 95% confidence intervals of the mean. The horizontal lines within each bar represent chance performance for that test item.

Are forward integration and backward inference relevant for statistical word learning beyond the mini-tasks? While the mini-tasks used in our experiment are structured similarly to the design of the larger cross-situational word learning paradigm employed in previous research (e.g., Yu & Smith, 2007; Yu, Zhong & Fricker, 2012) the mini-tasks had much more trial-to-trial overlap than other cross-situational learning paradigms. Thus, it is possible that learners don't rely on these inferential computations in the larger task, but simply accumulate co-occurrence statistics.

In order to verify that forward integration and backward inference were relevant for cross-situational word learning in a larger set, we correlated participants' scores on the Base task with their scores on the Full CSL task (see Figure 3). The proportion of correct object-label mappings was positively correlated for the Base mini-task and Full CSL task (r=0.485, p=0.002), suggesting that these tasks tapped similar skills. We also tested correlations between participants' accuracy on the Full CSL and on each of the individual trial types in the Base task to investigate the role of the individual computations. Positive correlations were found for both backward inference (T1) items (r=0.407, p=0.011) and forward integration (T2) items (r=0.387, p=0.016). The relationship between accuracy on the co-occurrence tracking (R) items and Full CSL accuracy was marginally significant (r=0.303, p=0.065). Accuracy on R items was in general quite high and this measure of co-occurrence tracking may not have been sensitive enough to detect a significant relationship. However, as described above, the tracking of co-occurrence information is necessary for the other two computations. Together, these results strongly suggest that forward integration and backward inference are processes integral to cross-situational learning.

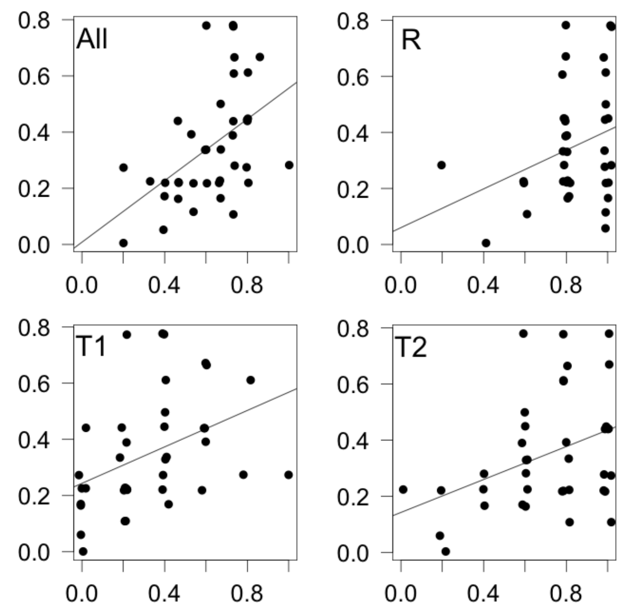

Figure 3. Scatterplot depicting the correlation between percentage correct on the Base Mini-Task (horizontal axis) and the Full Task (vertical axis). Values have been jittered so that all data may be seen. Overall performance on the Base Task is shown in the top left panel, R test trials in top right, T1 test trials in bottom left and T2 test trials in bottom right. Lines represent linear best fit.

The role of tracking multiple co-occurrences

We now turn to comparisons between the mini-tasks to further explore the computations learners employed to infer label-object mappings and participants' memory for ambiguous prior information. The Familiar Context task repeated the information from Trial 1 on Trial 3. We predicted that this repetition would lead to higher accuracy relative to the Base task for R and T1 items and lower accuracy for T2 items (because of the addition of Trial 3 between training on T2 in Trial 2 and test). Of these predictions, only the improvement on T1 is confirmed by the data. Accuracy on R items was not significantly different between the two tasks ($p > 0.29$), suggesting that the additional co-occurrence information did not lead to better mappings.

Changes in performance on T1 and T2 items were tested with a logistic mixed-effect model with Task (Base or Familiar Context) and Trial Type (T1 or T2) as fixed effects and Subject as a random effect. This revealed a significant Task by Trial Type interaction (b=-1.00, z=3.13, p=0.002). However, follow-up analyses confirmed that the pattern of effects was not as predicted. While accuracy on T1 was higher for the Familiar Context task than the Base task (b=0.631, z=2.94, p=0.003), accuracy on T2 was not different across the two tasks (b=-0.347, z=1.48, p=0.14). Accuracy was significantly higher for T2 than T1 for both tasks (Base: b=1.898, z=7.88, p<0.001; Familiar Context: b=0.785, z=3.64, p<0.001). Surprisingly, even though participants could use the same forward integration process to infer the T1 and T2 mappings and the T1 object and label were presented on the last training trial while T2 had to be maintained across this trial, participants were more accurate on T2 than T1.

Why might the results differ so much from our initial prediction? One possibility is that, for some participants, rather than increasing certainty, the repetition of the Trial 1 information on Trial 3 actually increased spurious correlations, and therefore the confusability, between $R_{1\text{-}3}$ and T1. The pattern of errors across tasks supports this interpretation. On R trials participants were significantly more likely to incorrectly select T1 for the Familiar Context task than the Base task (b=1.153, z=2.92, p=0.004). Selection of an R item on T1 test trials was equivalent across the two tasks, despite the improved performance on T1 for the Familiar Context task (b=-0.282, z=1.06, p=0.29).

The participants with the best memory of the first trial should be most likely to confuse $R_{1\text{-}3}$ and T1. The nature of the backward inference requires memory for T1, as well as memory for the context in which T1 occurred (i.e., the other objects and labels), since it is the absence of T1 from this context on Trial 2 that allows the inference. If memory of Trial 1 increases confusability between $R_{1\text{-}3}$ and T1, participants who were successful on backward inference in the Base task should improve *less* on T1 items for the Familiar Context task than participants who were not successful on backward inference.

Participants were split into two groups at the median for backward inference on the Base mini-task (N=19 in each group). Participants with 20% or less correct on T1 items for the Base task were labeled low-backward inference (low-BI, M=0.116, SD=0.322) and those with more than 20% correct were labeled high-backward inference (high-BI, M=0.537, SD=0.501). A logistic mixed-effect model predicting accuracy on T1 test items with Task (Base or Familiar Context) and BI (high or low) as fixed factors and Subject as a random factor revealed a significant Task by BI interaction (b=-2.346, z=4.86, p<.001). The low-BI group had significantly higher accuracy on the Familiar Context task (M=0.484, SD=0.502) than the Base task (b=1.973, z=5.18, p<0.001), while the high-BI group did not (M=0.453, SD=0.500; b=-0.349, z=1.19, p=0.23). Thus, participants with a weak memory of T1 from the first trial benefitted from the repetition of that information on Trial 3, while participants with a strong memory did not. This is further confirmed by the pattern of accuracy for R items across the two tasks, which was also subject to a Task by BI-group interaction (b=-1.122, z=2.01, p=0.044). The low-BI group improved slightly from the Base task to the Familiar Context (Base M=0.768, SD=0.424; FC M=0.80, SD=0.402) while the high-BI group actually declined (Base M=0.905, SD=0.294; Familiar Context M=0.789, SD=0.410), suggesting that the Trial 1 repetition disrupted their memory for the R items.

In contrast to the Familiar Context task, performance on the Novel Context task should only be aided by improved memory for T1 (and its first-trial context). Both high-BI and low-BI participants benefitted from the presentation of T1 in a novel context with high accuracy on T1 items (high-BI M=0.726, SD=0.448; low-BI M=0.579, SD=0.496).

While a better memory for context impeded performance on the Familiar Context mini-task, such memory should generally improve statistical word learning, as learners would have a more complete association matrix on which to build. We tested the role of contextual memory in cross-situational word learning by comparing performance in the Full CSL task for the high- and low-BI groups. As predicted, high-BI participants were significantly more accurate in the Full CSL task than low-BI participants (high-BI M=0.444, SD=0.192; low=BI M=0.254, SD=0.173, $t(36)$=3.2, p=0.003).

Discussion

The present study investigated three fundamental processes that may contribute to cross-situational word learning. We found that learners readily tracked co-occurrence information trial by trial and used those co-occurrence statistics to infer label-object mappings in new learning situations, a process we termed forward integration. We also found that learners inferred label-object mappings when the disambiguating evidence was the *absence* of the label and object on trials on which they would otherwise be expected, a process we termed backward inference. Further, we found that participants retained multiple co-occurrences between objects and labels presented on previous trials. Importantly, participants who best remembered multiple object-label co-occurrences within a learning trial were most successful at cross-situational word learning.

Our results support the argument that cross-situational word learning involves learning a *system* of label-object mappings, in which learning about one set of items influences knowledge about other items. From a straightforward co-occurrence information point of view, the T1 and T2 objects are not more strongly associated with the T1 or T2 labels than the R_1, R_2 and R_3 labels for either the Base or Familiar Context task. In order to disambiguate these mappings participants must use the information available not just about the T1 and T2 pairs but also about the R_1, R_2 and R_3 pairs. In this way, participants draw on the entire association matrix to make inferences that are reasonable given their experience. Our results provide empirical evidence of these inferences, but do not tell us whether inferences were made by explicit reasoning or emerged from the dynamics of attention within and across trials (Kachergis, Yu & Shiffrin, 2012; Yu & Smith, 2012; Yu, Zhong & Fricker, 2012). If replicated in young word learners, these results suggest an important role for the contexts in which word learning occurs.

There has been debate about the nature of information selection and information processing by cross-situational word learners. The presence of multiple objects and multiple labels on an individual learning instance means that learners could potentially associate all labels with all objects – the multiple association account (e.g., Yurovsky, Smith & Yu, 2012). While equal attention may not be given to all possible mappings, this account predicts that learners will have a rich store of statistical information to draw on, so that if evidence for one particular mapping is contradicted (e.g., the label is given but the object is not present) there are other associations already in place that can inform the learners' inferences about the label's likely referent.

Alternatively, in the single-association account learners retain a single hypothesis for each object, discarding all other associations from a particular learning instance (Medina, Snedecker, Trueswell & Gleitman, 2011; Trueswell, Medina, Hafri & Gleitman, 2013). This account predicts that when a particular hypothesis is contradicted the learner must start from scratch, forming a new hypothesis at random based on the current learning instance.

These two accounts make disparate predictions for the present study, specifically within the Familiar Context task. The single-association account proposes that learners may form a hypothesis linking the T1 object and label during Trial 1 and that this hypothesis would be confirmed on Trial 3. However, because choice of hypotheses is random, there should not be systematic differences between which learners benefit from this extra information from one mini-task to the next. In direct contrast to this, our results suggest that for some learners, the repetition of information in Trial 3 was beneficial, improving accuracy on R and T1 items, and for some learners it was not. Crucially, what defined whether Trial 3 was beneficial was whether the participant had formed a strong memory for the first trial, both the potential T1 mapping AND the other objects present, as measured by their ability to perform backward inference. These findings raise important questions about how memory development may influence word learning in toddlers, as we found that better in the mini-tasks with high overlap, better memory led to potential interference, while in the larger task with little trial-by-trial overlap, better memory (i.e., backward inference) led to better performance. Our data suggest that those learners who are successful in cross-situational learning tasks carry multiple possible associations forward. These associations are integrated in both the forward and backward directions to discover likely object-label pairs. Thus, statistical associative learning is a powerful mechanism that is within the repertoire of human cognitive systems.

Acknowledgments

This work was supported by NIH R01 HD056029 to C. Yu and an NIH Ruth Kirschstein NRSA training grant to Indiana University: T32 HD007475-17.

References

Kachergis, G., Yu, C., & Shiffrin, R. M. (2012). An associative model of adaptive inference for learning word–referent mappings. *Psychonomic Bulletin & Review*, 1-8.

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*, 57-77.

Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *PNAS*, *108*(22), 9014-9019.

Scott, R. M., & Fisher, C. (2011). 2.5-Year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, *122*, 163-180.

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

Suanda, S. H., & Namy, L. L. (2012). Detailed Behavioral Analysis as a Window Into Cross‐Situational Word Learning. *Cognitive Science, 36, 545-559.*

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*, 126-156.

Vouloumanos, A., & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Developmental psychology*, *45*, 1611.

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

Yu, C., & Smith, L. B. (2012). Modeling cross-situational word–referent learning: Prior questions. *Psychological Review*, *119*, 21.

Yu, C., Zhong, Y., & Fricker, D. (2012). Selective attention in cross-situational statistical learning: evidence from eye tracking. *Frontiers in Psychology*, 3:148 doi: 10.3389/fpsyg.2012.00148

Yurovsky, D., Smith, L. B., & Yu, C. (2012). Does Statistical Word Learning Scale? It's a Matter of Perspective. In N. Miyake, D. Peebles, & R. Cooper (Eds.), *Proceedings of the 34nd Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.