

Statistical Word Learning is a Continuous Process: Evidence from the Human Simulation Paradigm

Yayun Zhang*, Daniel Yurovsky⁺, Chen Yu*

yayzhang@indiana.edu, yurovsky@stanford.edu, chenyu@indiana.edu

*Department of Psychology & Brain Science, Indiana University, 1101 E. 10th Street, Bloomington, IN 47405 USA

⁺Department of Psychology, Stanford University, 450 Serra Mall, Stanford, CA 94305 USA

Abstract

In the word-learning domain, both adults and young children are able to find the correct referent of a word from highly ambiguous contexts that involve many words and objects by computing distributional statistics across the co-occurrences of words and referents at multiple naming moments (Yu & Smith, 2007; Smith & Yu, 2008). However, there is still debate regarding how learners accumulate distributional information to learn object labels in natural learning environments, and what underlying learning mechanism learners are most likely to adopt. Using the Human Simulation Paradigm (Gillette, Gleitman, Gleitman & Lederer, 1999), we found that participants' learning performance gradually improved and that their ability to remember and carry over partial knowledge from past learning instances facilitated subsequent learning. These results support the statistical learning model that word learning is a continuous process.

Keywords: statistical learning; word-referent mapping; learning mechanisms

Introduction

Many recent studies have shown that both adults and children acquire new vocabulary by using word-object co-occurrences to discover which linguistic labels map on to which objects (e.g. Yu & Smith, 2007). Despite the fact the natural learning environment is noisy and ambiguous, human learners are still able to keep track of multiple possible word-object pairings simultaneously (Yurovsky, Smith, & Yu, 2013). They continuously store and update the word-object co-occurrences across word learning moments and make statistically appropriate decisions based on aggregated statistics (Smith, Suanda, & Yu, 2014).

However, this aforementioned cross-situational word learning strategy and its supporting associative learning model (AL) have been challenged by another learning model called the hypothesis testing model (HT). Although both computational modeling results and behavioral data provide evidence showing that the two models do interact to some degree during word learning and the learning outcomes generated by these two models can be similar (Yu & Smith, 2012; Smith et al., 2014; Romberg & Yu, 2014), they do suggest fundamentally different learning pathways (Trueswell, Medina, Hafri, & Gleitman, 2013). One major difference between these two models is how learners process past information when learning object labels in

subsequent moments. The AL model suggests that learners can keep track of multiple co-occurrences of object-label mappings in one naming situation. Because a label and its correct referent are likely to co-occur more consistently than do other pairs, with enough exposure, the correct mapping can be accomplished by using cross-trial statistical relations (Yu & Smith, 2007). A more recent study supports the AL model by showing that word learning is not an "all-or-none" process. Instead, it is an incremental process that involves forming partial knowledge of word-object associations (Yurovsky, Fricker, Yu, & Smith, 2014). Therefore, labels are learned gradually by accumulating knowledge from past learning experience. However, the HT model suggests that learners only make one hypothesis on an object-label pairing in one context. If this hypothesis were confirmed in later contexts, it would be considered as learned knowledge. If the hypothesis were rejected, then learners would pick another hypothesis from scratch and repeat the process until getting the correct mapping (Medina, Snedeker, Trueswell, & Gleitman, 2011; Trueswell et al., 2013). Thus, the HT model supports a "fast mapping" process with fewer exposures whereas the AL model suggests gradual statistical learning with lots of data.

Not only do these two competing mechanisms define statistical learners very differently, the experimental paradigms used to support these two models are quite different as well. For a typical adult study that supports the AL model, participants are asked to learn object and word mappings through a series of learning trials wherein each trial contains multiple pseudowords and multiple novel objects without information about which words map on to which objects. While the cross-situational learning paradigm provides us useful data on how learners process information in ambiguous learning situations, one remaining question is whether learners employ the same learning strategy in the real world as real-life learning moments can be much more uncertain and noisy than laboratory tasks. One recent study done by Medina et al. (2011) used the Human Simulation Paradigm (Gillette et al., 1999) to study whether adults are able to learn gradually by accumulating evidence from multiple naturalistic learning instances from child-parent interactions (details described in the method section below). What they found was that incremental learning from multiple ambiguous learning moments did not occur. Instead, successful word learning depended on the presence of unambiguous learning moments. Participants learned the best when the unambiguous learning moments happened

early during training, suggesting that word learning requires an initial “one trial learning” step followed by a confirmatory process (Trueswell et al., 2013).

However, Yurovsky et al. (2013) followed Medina et al.’s method and found very different results. Instead of studying word learning from only the observer’s perspective, they recorded training videos from both the child’s view (captured by a head-mounted camera) and the observer’s view (captured by a tripod-mounted camera) in order to study whether adult participants can learn from ambiguous learning events after viewing multiple child’s view naming moments. Previous studies with head-mounted cameras have shown that children’s visual field is selective and only includes one or a very few dominant objects at a naming moment. Therefore, naming events seen from the child’s view videos may be less ambiguous than those seen from the observer’s view, which may facilitate cross-situational learning. The results demonstrated that participants’ learning performance improved significantly after watching multiple highly ambiguous child’s view videos but not after when the same naming events were seen from the observer’s perspective. Thus, statistical aggregation may indeed characterize learning from the kind of naming events children experience; learning may not necessarily require unambiguous learning moments (Yurovsky et al., 2013).

Given the conflicting findings from previous literatures, in the current study we aim at investigating the underlying mechanisms of how learners process information across multiple learning contexts with the following questions: 1) Do learners gradually accumulate knowledge from multiple naturalistic naming moments? 2) Are ambiguous learning events enough for successful cross-situational learning? Are unambiguous learning instances necessary? 3) How do unambiguous learning trials interact with ambiguous learning trials, and how do they influence learners’ performance? To answer these questions, our design followed the Human Simulation Paradigm to closely simulate learning moments in the real world. Meanwhile, we systematically selected and manipulated a set of videos that vary in their ambiguity, allowing us to measure and analyze participants’ learning patterns trial-by-trial in order to examine how statistical learning unfolds over time. Experiment 1 was designed to provide a baseline of learning performance for individual naming moment; Experiment 2 focused on statistical learning solely from ambiguous learning instances; and Experiment 3 focused on information integration through a set of interleaved ambiguous and unambiguous learning moments.

Experiment 1

In order to examine the detailed word-learning patterns, we first selected a set of naming instances from the video corpus collected by Yurovsky et al. (2013) for their original study. The videos included play sessions from eight parent-child dyads. Parent-child dyads were asked to play naturally with 25 toys for about 10 minutes while their interactions

were recorded by a tripod-mounted camera and a head-mounted camera in order to get both the observer’s view and the child’s view at each naming moment. The current study only used videos from the child’s view because ultimately only visual inputs perceived by the child enter into the learner’s cognitive system.

The goal of Experiment 1 was to provide a baseline measure of the ambiguity of naming events. Following Medina et al. (2011) and Yurovsky et al. (2013), we replaced each object names in the videos with an identical beep to measure the baseline information of each video seen in isolation. However, instead of asking participants to type back the names of the referents as in previous studies, we made the test trials more straightforward by giving participants a forced-choice test of learning performance. We believed that this testing paradigm would reduce the demand on vocabulary retrieval and avoid potential disagreements during response coding, thus provide us a cleaner and more reliable measure of learning. The purpose of Exp 1 was to get a baseline measure of the ambiguity of each naming video by using the new forced-choice test.

Participants. Seventeen Indiana University undergraduates (4 Male, $M_{age} = 19.82$ $SD_{age} = 1.47$) participated in exchange for course credits. None had participated in other cross-situational word learning experiments.

Materials. Ninety-six child’s view naming moment vignettes were selected from the video corpus. The correct referents were twelve different toys (e.g. elephant, mickey, tiger, etc), each of which had eight naming instances from at least four different parent-child dyads. Based on previous baseline data reported in Yurovsky et al. (2013), 3 of these 8 naming instances (Figure 1) were highly unambiguous ($M = .98$, $SD = .04$) and 5 of them were highly ambiguous ($M = .11$, $SD = .13$). These 96 vignettes were grouped into 8 blocks. Twelve vignettes in each block referred to 12 different toys. Vignettes were pseudorandomized within block and the ambiguity of vignettes did not follow any specific order.



Figure 1: Both highly unambiguous (A) and highly ambiguous (B) vignettes were used for all 3 experiments. The named object “mickey” can be easily identified in (A) as the dominant object in view, but not in (B) which contains multiple competing objects at the naming moment.

For each naming instance, the original sound was muted and the toy name was replaced by a beep at the onset of the

label. Most vignettes were 5 seconds long, with the name's onset occurring at exactly the third second. Two more seconds were added to the vignettes if mothers said the toy name again within 2 seconds after the first naming instance. Seven of the 96 vignettes included two naming instances and two included three naming instances. Four additional vignettes were included as examples. None of the correct referents in these examples were targets in actual training. The testing stimuli were 25 color photos of all toys given to the parent-child dyads during the free-play session. Images were displayed on a white background, in a 5x5 grid.

Procedure. Participants were instructed to watch the vignettes and identify the objects that correspond to the beeps. They were notified that for each test trial, they would see 25 pictures on the screen and they needed to choose the most likely referent by clicking on the picture. No feedback would be given. Participants then proceeded to see four sample vignettes, each followed by a testing trial. Once they were familiar with the study procedure, they were prompted to begin the actual experiment. Short breaks were given after the 2nd, 4th and 6th block.

Results and Discussion. The baseline results found using the forced-choice test was similar to the results of the original study (Yurovsky et al., 2013). Accuracy on the unambiguous trials was high ($M=.94$, $SD=.07$, $M_{min}=.71$; $M_{max}=1$) and accuracy on the ambiguous trials was low ($M=.14$, $SD=.16$, $M_{min}=0$; $M_{max}=.59$). This result suggests that the forced-choice test is a reliable measure to use for further learning tasks and these 96 vignettes are representative cases that resemble hard and easy learning instances in real life.

Experiment 2

To explore whether learners aggregate past knowledge across multiple ambiguous learning events, we asked participants to learn object names by observing a set of ambiguous vignettes and to make guesses on a trial-to-trial basis. If participants do carry over their previous knowledge, then we should see an incremental increase in guessing accuracy.

Participants. Twenty-six Indiana University undergraduates (7 Males, $M_{age} = 19.08$, $SD_{age} = 1.20$) participated and received course credits. None had participated in the previous baseline study or other cross-situational word learning experiments.

Materials. The same 96 vignettes used in Exp 1 were used in Exp 2. These 96 vignettes were divided into 12 blocks. Each block had 8 different vignettes all referring to the same toy. In any given block, the first 5 vignettes were ambiguous and the last 3 were unambiguous. The purpose of adding 3 unambiguous trials at the end was to measure how well people learn in unambiguous learning moments, and we expected participants to perform well with those easy cases.

Twelve one or two-syllable novel word labels (e.g. agen, gree, hage, etc) were recorded by a female native speaker of English. Instead of beeps, the novel word labels were now inserted to correspond to times in the original interactions that mothers used the object's English labels. Each toy had a unique label. The testing stimuli were the same as Exp 1.

Procedure. Participants were told that they would be trying to learn some words for some familiar objects in a new language. They would do this by watching mothers playing with their children and trying to guess which object the mothers were naming using the new label by choosing a most likely answer after each video. Testing instructions were the same as Experiment 1. After seeing the examples, participants were first presented with the first block of 8 vignettes. They were told that mothers in these 8 videos were naming the same object. Throughout the 8 testing trials, they were allowed to change their guess at any given trial. However, if they believed their previous answer was correct, they could choose the same answer again. They were not allowed to go back and change their previous answers and they were not aware of the ambiguity of each vignette. After each block, a prompt would appear to remind them to get ready for the next block of trials. Again, no feedback would be given.

Results and Discussion. Because we are interested in whether participants accumulate knowledge across ambiguous naming instances, we mainly focused on guess accuracy for the first 5 trials, which were all highly ambiguous trials. Figure 2 shows response accuracy for each trial, averaged across the 12 objects. Participants' responses on the first trial ($M_1 = .23$, $SD_1 = .16$) were low but still higher than baseline. Because of the block design of the current study, the mean first-trial accuracy was calculated by aggregating guesses across blocks. Participants tended to achieve better learning performance in later blocks, which is additional evidence on statistical cross-situational learning across multiple target words. The topic of cross-word statistical integration is worth future studies by itself. Nonetheless, the present study focuses on information aggregation from multiple learning instances of the same word.

Because Exp 2 gave participants the opportunity to make their guesses based on what they have learned before, we asked whether their accuracy improved significantly across trials. We fit a mixed-effects logistic regression predicting accuracy from trial number and baseline accuracy from Experiment 1 with a random effect of subject. This model had a highly-significant main effect of trial number ($\beta=.29$, $p<.001$) over and above the effect of baseline accuracy ($\beta=2.42$, $p<.001$), indicating significant learning across trials. Figure 2 shows this improvement, ranging from 23% accuracy on trial 1 to almost 50% on trial 5. This dramatic improvement suggests that word learning is a continuous process that learners make progress gradually by integrating what they have learned before. This result contradicts the

HT model arguing that highly ambiguous learning moments are not very useful as they might not be remembered over time to improve learning (Medina et al., 2011).

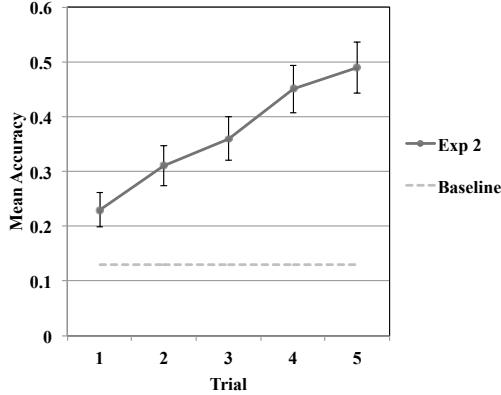


Figure 2: Mean accuracy (± 1 SE) across 5 ambiguous naming instances and baseline accuracy from Exp 1.

As expected, the mean accuracies for the 3 unambiguous trials (6th to 8th) at the end of each block were very high ($M_6 = .83$, $SD_6 = .38$; $M_7 = .90$, $SD_7 = .30$; $M_8 = .95$, $SD_8 = .23$).

Knowing that prior knowledge plays an active role in word learning, we further investigate to what degree learners' guesses depend on their previous experience. We calculated their learning performance conditioned on whether or not their previous guess was correct. When participants guessed the previous trial correctly, they were more likely to guess the current trial right ($M = .74$, $SD = .29$) compared to when they got the previous trial wrong ($M = .20$, $SD = .12$). To determine whether this difference was significant, we fit a mixed effect model as before, but this time added an additional main effect of Previous Trial, which was coded as -1 if the previous trial was incorrect, 1 if it was correct, and 0 for the participant's first trial. All previous factors remained significant, but additionally previous trial accuracy was a highly significant predictor ($\beta = 1.49$, $p < .001$).

Were participants learning even on trials for which they gave an incorrect answer? We subset the trials from Experiment 2 to just those following incorrect responses and asked whether accuracies on these differed from baseline accuracies on the comparable videos (mixed effects model: accuracy ~ experiment + (1|video) + (1|subj)). This model found a significant effect of experiment, indicating even when participants failed to get the correct answer for the previous trial, their current trial accuracy was still significantly above baseline ($\beta = .46$, $p < .01$). This finding suggests that without getting any feedback, learners used their prior knowledge to guide their current decision. They tended to use the previous accurate information in a more efficient way by choosing the same correct answer again. Even if their previous answer was wrong, they were still able to carry over partial knowledge that would allow them to improve their learning performance. This finding again contradicts Medina et al. (2011)'s finding showing that when participants guess incorrectly on a learning trial, their

guessing accuracy is at chance at the very next learning situation indicating no knowledge of previous contexts.

One distinction between associative learning and hypothesis testing is that associative learners store and use lots of data – all prior experiences through the course of learning, while hypothesis testing learners only update their current hypotheses trial-by-trial. To measure how much learning depends on prior experiences, we examined participants' learning performance in the current trial conditioned on the proportion of correct answers from all of the prior trials for the same word. The analysis revealed a clear pattern showing that as participants' total number of previous correct trials increases, their performance on the current trial also improves (Figure 3). For example, at the second trial, participants who guessed correctly on their previous trial ($M = .67$, $SD = .39$) were more accurate on the current trial than those who guessed incorrectly on the previous trial ($M = .20$, $SD = .16$, $t(22) = 6.34$, $p < .001$).

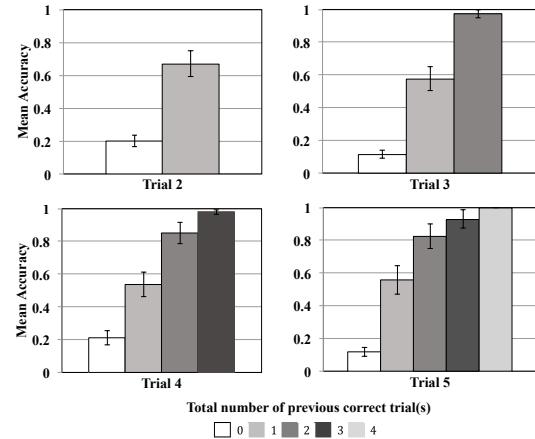


Figure 3: Mean accuracy across vignettes at each trial. Bars represent total number of previous correct trials.

To quantify this effect of accumulated learning as a continuous variable, we added another factor to the mixed effects model—the proportion of previous trials on which the participant was successful. An ANOVA showed that this addition significantly improved the model's fit ($\chi^2 = 1139$, $p < .001$). Proportion of previously correct trials was a significant predictor of accuracy over and above the contribution of previous trial accuracy ($\beta = 23.38$, $p < .001$). This finding reveals that learners not only carry over knowledge learned from the immediate previous learning trial, they also encode and use all their past learning experiences in highly ambiguous learning contexts, which again suggests that cross-situational learning is a cumulative and continuous process that involves tracking and integrating past contexts.

Experiment 3

Because real life situations often involve both ambiguous and unambiguous learning moments, we next investigate how these two types of learning instances interact with and influence each other.

Participants. Twenty-six Indiana University undergraduates (11 Males, $M_{\text{age}} = 20.27$, $SD_{\text{age}} = 2.32$) participated in exchange for course credits. None had participated in Experiment 1 or 2 or other similar experiments.

Materials. The same materials used in Experiment 2 were used again in Experiment 3, but the trials within each block were re-arranged. The overall design was to have 1 unambiguous trial followed by 2 ambiguous ones, so we put the preselected 3 Unambiguous trials (U) at the 1st, 4th and 7th positions and the other 5 trials were Ambiguous (A). The whole sequence is composed as U-A-A-U-A-A-U-A. In this way, ambiguous and unambiguous trials are interwoven.

Procedure. The procedure was the same as Experiment 2.

Results and Discussion. We calculated mean guessing accuracy for the first 6 trials across items, which consisted of two sets of “U-A-A” sequences (Figure 4). There are three distinctive patterns: (1) As expected, participants’ responses were highly accurate at unambiguous trial 1 ($M_1 = .93$, $SD_1 = .16$) and trial 4 ($M_4 = .94$, $SD_4 = .14$). In addition, we did not find any significant learning difference between the two unambiguous trials (trial 1 and 4, $t(25) = .36$, ns). This is contrary to Medina et al. (2011)’s finding that ambiguous learning moments hurt learners’ performance on later unambiguous learning moments. (2) Accuracy on trial 2 ($M_2 = .64$, $SD_2 = .34$) and trial 5 ($M_5 = .73$, $SD_5 = .35$) is much higher than baseline. Thus, there is a significant improvement of learning on an equally ambiguous naming situation after an unambiguous one. (3) There is a significant improvement from trial 3 ($M_3 = .63$, $SD_3 = .07$) to trial 5 ($t(25) = 4.2$, $p < .001$), suggesting that participants gradually improved their learning performance with more learning trials.

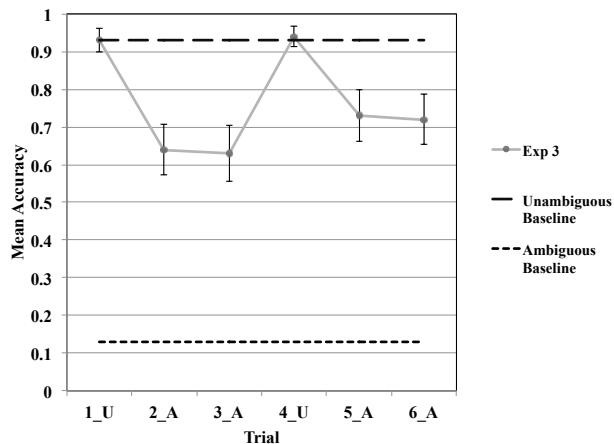


Figure 4: Mean guessing accuracy (± 1 SE) across the first 6 naming instances and baseline accuracy for both ambiguous and unambiguous trials from Exp 1.

To understand how ambiguous and unambiguous information is integrated trial-by-trial, we examined accuracy on the ambiguous learning trials (trial 2 and 5) that immediately followed the unambiguous instances (trial 1

and 4, see Figure 5A) and the ambiguous learning trials (trial 3 and 6) that immediately followed other ambiguous instances (trial 2 and 5, see Figure 5B). Data were further split by whether learners got the previous trial right or not. Guessing responses on trial 2 and 5 were collapsed because they were at the same position in the “U-A-A” sequence and trial 3 and 6 were also combined for the same reason.

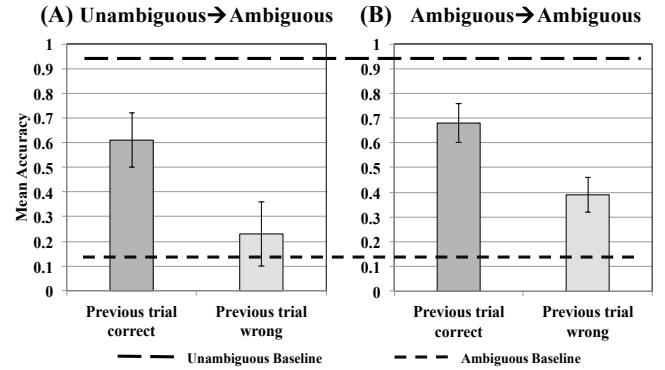


Figure 5: Mean accuracy of current trial as a function of whether participants answered the previous unambiguous trial (A)/ambiguous trial (B) correctly or not. Ten participants contributed to (A) and 22 participants contributed to (B).

When the previous trial was unambiguous, participants’ response accuracy on the following ambiguous trial was higher ($M = .61$, $SD = .35$) when they got the unambiguous trial right than when they got it wrong ($M = .23$, $SD = .42$, $t(9) = 3.16$, $p = .01$). To test whether each was significantly above baseline, we fit a mixed-effects model as in Exp 2 to determine if responses in Exp 3 were different from those on comparable trials in Exp 1. This effect was significant for trials following correct responses ($\beta = 4.64$, $p < .001$), but not for trials following incorrect responses ($\beta = -.50$, $p = .51$). This suggests that if participants missed the “obvious” cues from easy learning moments, they were not able to carry over any useful information that could potentially benefit subsequent learning.

However, when both the previous and current trials were ambiguous, the pattern of responses was similar to the finding of Experiment 2. Participants’ learning performance was significantly better when they made a right guess ($M = .68$, $SD = .36$) on the previous trial than a wrong one ($M = .39$, $SD = .33$, $t(21) = 4.00$, $p = .001$) and both scores were above baseline (by mixed-effects model as above, post-correct $\beta = 5.48$, $p < .001$, post-incorrect $\beta = .58$, $p < .001$). This finding again supports the statistical learning model that learning involves continuous interactions of knowledge on a moment-to-moment basis. From the current design, it is clear that remembering and carrying over partial knowledge, despite the uncertainty of the information, could facilitate learning and partial knowledge can be especially helpful when the learning situations are ambiguous. This finding is also consistent with previous work showing that partial knowledge learned from previous experience may be leveraged incrementally to bootstrap learning (Yurovsky, et al., 2014).

General Discussion

To answer the study questions of how learners acquire correct word-object mappings through multiple naturalistic naming situations and whether unambiguous instances facilitate learning, we found that words are learned gradually by accumulating information across multiple naturalistic learning situations, and do not change suddenly from “unknown” to “known.” Although there is no doubt that learners achieve the highest learning performance when the naming moments are unambiguous, this does not mean that adults and children have to rely heavily on these “perfect” moments to learn words. Instead of focusing on the one-trial learning procedure that depends on locking in to a word’s correct label upon first encounter, word learning is more likely to be a continuous process that not only benefits from fast mapping, but also from aggregating statistics from past learning experiences.

Although successful fast-mapping of a word to its correct referent emerges quite early in development, successful retention of this mapping appears significantly later: 24-month-old infants show no evidence of learning after only 5 minutes delay (Horst & Samuelson, 2008; Bion, Borovsky, & Fernald, 2013). Results from studies with 3-year-old children and adults also suggest that despite participants’ ability to quickly form a new word-object mapping and perform well on an immediate test, they forget words over time in a curvilinear fashion (Vlach & Sandhofer, 2012). These results raise questions of whether the fast mapping strategy would be sufficient to help learners turn novel names into known ones for later retrieval or it is just an early disambiguation skill that does not directly relate to word learning (McMurray, Horst, & Samuelson, 2012). This is consistent with the learning pattern seen in Experiment 3. Even though learners achieved high accuracy in unambiguous learning trials (one-shot learning, etc.), their learning performance dropped significantly in subsequent ambiguous contexts in which they had to retrieve their previous mapping knowledge. Therefore, retention of word-object mappings might not be as consistently high as previously believed (Carey & Bartlett, 1978). Instead, it is very likely that word learning is a context dependent process that involves accumulating partial knowledge over a long time scale (Bion et al., 2013).

Despite the debate on whether word learning is a “fast mapping” procedure or a gradual statistical one, recent computational modeling results of these two models reveal that hypothesis testing model can actually be viewed as a special case of the associative learning model, suggesting that representations of these two models are exchangeable (Yu & Smith, 2012). Therefore, real world word learning is very likely to involve both learning mechanisms and individuals’ learning pattern is sensitive to the structure of information provided (Romberg & Yu, 2014).

Our view of word-referent learning as a continuous statistical learning process is supported by the current findings. By investigating both highly ambiguous and unambiguous learning events and the interaction between

the two, we believe that both types of learning instances contribute to a continuous process of word learning.

Acknowledgments

This research was supported by NIH R01 HD074601. Special thanks to Stella Huang for data collection.

References

- Bion, R. A., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word-object mappings in relation to vocabulary learning at 18, 24, and 30months. *Cognition*, *126*, 39-53.
- Carey, S. & Bartlett, E. (1978). Acquiring a single new word. *Proceedings of the Stanford Child Language Conference*, *15*, 17-29.
- Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*, 135-176.
- Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, *13*, 128-157.
- Romberg, A.R. & Yu, C. (2014). Interactions between statistical aggregation and hypothesis testing during word learning. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 1311-1316). Austin, TX: Cognitive Science Society.
- Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 9014-9019.
- McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological review*, *119*, 831-877.
- Smith, L. B., Suanda, S. H., & Yu, C. (2014). The unrealized promise of infant statistical word-referent learning. *Trends in Cognitive Sciences*, *18*, 251-258.
- Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*, 1558-1568.
- 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.
- Vlach, H. A., & Sandhofer, C. M. (2012). Fast mapping across time: Memory processes support children's retention of learned words. *Frontiers in Psychology*, *3*, 46.
- Yu, C., & Smith, L. B. (2012). Modeling cross-situational word-referent learning: Prior questions. *Psychological Review*, *119*, 21-39.
- Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*, 414-420.
- Yurovsky, D., Fricker, D., Yu, C., & Smith, L. B. (2014). The role of partial knowledge in statistical word learning. *Psychonomic Bulletin & Review*, *21*, 1-22.
- Yurovsky, D., Smith, L. B., & Yu, C. (2013). Statistical word learning at scale: The baby's view is better. *Developmental Science*, *16*, 959-966.