Published in final edited form as:
Lang Learn Dev. 2013 ; 9(1): . doi:10.1080/15475441.2012.707104.

Visual attention is not enough: Individual differences in statistical word-referent learning in infants

Linda B. Smith and **Chen Yu**
Department of Psychological and Brain Sciences, Program in Cognitive Science, Indiana University Bloomington, IN

Abstract

Recent evidence shows that infants can learn words and referents by aggregating ambiguous information across situations to discern the underlying word-referent mappings. Here, we use an individual difference approach to understand the role of different kinds of attentional processes in this learning: 12-and 14-month-old infants participated in a cross-situational word-referent learning task in which the learning trials were ordered to create local novelty effects, effects that should not alter the statistical evidence for the underlying correspondences. The main dependent measures were derived from frame-by-frame analyses of eye gaze direction. The fine- grained dynamics of looking behavior implicates different attentional processes that may compete with or support statistical learning. The discussion considers the role of attention in binding heard words to seen objects, individual differences in attention and vocabulary development, and the relation between macro-level theories of word learning and the micro-level dynamic processes that underlie learning.

Keywords

Word learning; Statistical learning; Development; Infant learning; Attention; Cross-situational word-referent learning

---

The problem of how infants break into word learning is still not well understood. A baby who knows no (or very few) words must attach names to objects as a consequence of experiencing co-occurring words and their referents. Young learners might learn their first words primarily in very clear cases in which the intended referent is the unambiguous focus of the speaker's and the learner's attention (e.g., Baldwin, 1993; Brent & Siskind, 2001; Hollich, Hirsh-Pasek, & Golinkoff, 2000). Yet many potential learning contexts are less than ideal and present the young learner with more ambiguous and less certain information (e.g., Woodward & Markman, 1998). Recent evidence suggests that infants, as well as adults, do learn words and referents in less than ideal contexts, aggregating ambiguous information across situations to discern the underlying word-referent mappings (Yu, Ballard, & Aslin, 2005; Yu & Smith, 2007; L. Smith & Yu, 2008; Vouloumanos, 2008; Vouloumanos & Werker, 2009; Scott & Fisher, 2009). These previous studies were centered on demonstrating the existence of cross-situational learning and as yet very little is known about the underlying mechanisms. Here we consider how different processes of visual attention may support or not support cross-situational learning. The findings indicate that some forms of visual attention, including novelty-driven attention, do *not* support statistical name-referent learning whereas other forms of attention do.

Our focus on processes of visual attention and their relation to statistical learning was motivated by previous findings of individual differences in infant cross-situational word-referent learning (Yu & Smith, 2010) and by theoretical analyses that suggest that the nature of the underlying attentional processes is a critical factor for statistical word-referent

learning under all theoretical assumptions (Yu & Smith, 2012). The prior empirical study used a "looking while listening" paradigm in which infants were presented with a series of visual scenes and co-occurring words as illustrated in Figure 1. On one trial, the infant might hear the words "regli" and "toma" in the context of seeing object **a** and object **b**. Without other information, the hypotheses that "regli" refers to object **a** and that "toma" refers to object **b** versus the hypotheses that "regli" refers to **b** and "toma" refers to **a** cannot be decided. However, if the next trial presents the referents of **b** and **c** in the context of the words "regli" and "gasser" and if the learner can remember the co-occurrences trial-to-trial and can combine the conditional probabilities of co-occurrences across trials, the learner could be more certain that "regli" refers to object **b** because **b** is the only candidate referent that has co-occurred with "regli" on both trials. In the first experiment using this method (Smith & Yu, 2008), 12- and 14-month old infants were presented with a randomly ordered stream of 30 such trials with 6 objects and 6 words to be learned across the trials. At the end of this experience, infants were tested: two visual objects were presented in the context of one spoken word and looking time was measured. The results showed that 12-and 14-month old infants looked more to the correct referent than the foil. To do this, they must have attended to, stored and statistically evaluated the information across the individually ambiguous training trials.

Yu and Smith (2010) added eye-tracking methodology and in this way tracked learning as it occurred, examining the object to which the infant attended when each word was heard during the ambiguous training trials. This method revealed marked individual differences in looking behavior that were strongly related to whether or not individual infants learned the underlying correspondences. At the beginning of training, looking was similar for all infants, with many rapid shifts of attention from one object to the other within a trial and little systematicity. Diffuse looking is potentially relevant to statistical learning, since infants might benefit from an initially broad sampling of the data on the pairings. However, on later looking trials, the looking patterns of infants who actually learned the word-referent associations as measured at test became more focused and different from those of nonlearners. More specifically, by the middle of the training trials, the learners' looking patterns were systematic, selective, and sustained on individual objects and they were often -- though not always -- directed toward the correct referent for the just heard word. However, the learners' attention but the nonlearners –at least as the learning trials progressed –became more controlled by the heard words whereas nonlearners' looking behavior did not. Looking at an object in the context of a heard word is both the means through which infants pick up information about the word-object correspondences and also the behavior experimentalists use to measure that learning. Because the differences in looking behavior during the training emerged across those trials, these differences most likely reflect differences in what infants had learned from the early trials about the word-referent correspondences. However, because this early learning organizes visual attention within trials, it may be essential to learning during later trials, for example, to the correction of spurious correlations, and thus to the overall success of statistical learning.

Importantly, both the infants who ultimately learned the correspondences and those who did not looked at the objects on all trials, but the looking behavior was *different*. This fact suggests that looking and listening is not enough to ensure statistical learning and raises the possibility that different forms of visual attention are differentially supportive of statistical word-referent learning. Recent advances in both theory and research suggest fundamentally different forms of attention (see Talsma, Senkowski, Soto-Farao & Woldorff, 2010, for review) that operate over different time scales (see Smith, Colunga & Yoshida, 2010, for review) and that support different cognitive functions (see, Talsma et al, 2010; Wright & Ward, 2008). In particular, studies of both adults (e.g., Fiebelkorn, Foxe & Molholm, 2010) and infants (Wu & Kirkham, 2010; Benitez & Smith, 2012) suggest that association-based

(or "endogenous") attention is critical to the binding of multimodal information within and across trials. That is, attention that is directed to a spatial location by a learned cue (and thus by expectancy and top-down processes) supports deeper cognitive processing and specifically the binding of one stimulus event to another. Thus, one possibility is that some infants in the Yu and Smith study did not learn the statistical correspondences between words and referents because their attention within any trial was primarily organized not by word-object associations (spurious or correct) but by individual stimulus saliency or the local novelty effects across trials. In contrast, the learners' visual attention may have been organized by word-object associations, and even though those associations may have been initially spurious, the expectancy based nature of attentional cuing may have led to better statistical learning. The current experiment was designed to test this idea.

Our focus on visual attention and on word-object associations runs counter to some theoretical approaches to cross-situational learning and to word-referent learning more generally. These alternative approaches do not focus on visual attention but on hypothesis testing and conceptualize learning not in terms of associations but in terms of reference (see Yu & Smith, 2012, for a review, as well as the other papers in this special issue). In formal theoretical analyses of hypothesis testing versus associations, Yu and Smith (2012) argue that the distinctions are not as formally clear-cut as our everyday intuitions might suggest and moreover that implicit assumptions about how attention works is a critical determiner of the success of both hypothesis-testing and association models of statistical word-referent learning. The distinction between words "as associates" versus "as referring" also may not be as clear-cut as some have proposed. For example, Waxman and Gelman (2009) dismissed associations as relevant to word learning arguing that words do not merely co-occur with objects but *point* to (or are "about") the object as the intended focus of interest. The mechanistic implications of a distinction between co-occurrence and reference is not obvious (see Yoshida & Smith, 2003). However, one possible behavioral implication of "reference" is that words predict what will be seen and thus cue looking behavior. From this perspective, looking that is too strongly determined by visual events alone –for example, by trial-to-trial changes in the particular objects in view or their momentary location -–– may compete with the role of words as referring and thus as cues to attention. In brief, the present focus on different kinds of visual attention may also inform and be relevant to other theoretical views of cross-situational word-referent learning, a point to which we return in the Discussion.

The rationale behind our experimental approach was to manipulate the trial structure so as to potentially capture infant looking behavior via local novelty effects and then to determine if infants who were more susceptible to these local visual effects were also less likely to learn the word-referent co-occurrences. The trials were also structured to examine the interaction of attention at different times scales, including local novelty effects and the more temporally extended effects of word-object associations across trials. The experimental task was the same cross-situational learning task used by Yu and Smith (2010) but the trials were rearranged to create what we expected to be strong local novelty effects within the stream of visual objects. These local novelty effects did not alter the underlying statistics of word-referent co-occurrences across the whole training set. In this way, we pitted two kinds of attention against each other: (1) looking to an object because it is new relative to the just previous trial and (2) looking to an object because of its statistical relation to a heard word. Because the cross-trial statistics for word-referent correspondences are the same as in the Yu & Smith (2010) study, with only the order of the trials differing, infants should learn the word referent correspondences if they keep track of these statistics. Moreover, if these statistics *increasingly organize attention during learning*, then children's looking behavior in response to the presentation of the words should change over trials, as was found for the learners in the Yu and Smith study (2010). Critically, if infants attended *only* to the locally

novel object, the relevant co-occurrence statistics for the underlying words and referents *would still be strong and expected to yield learning*. Thus if *any* form of visual attention supports statistical learning, then even infants whose attention is strongly organized by novelty could learn the word-referent correspondences. If, however, local visual novelty effects compete with the binding of heard words to seen objects, then infants who show looking behavior strongly organized by the novelty effects should be less likely to learn the word-referent correspondences.

Table 1 provides the structure of the training trials. There were in total 6 word-referent pairs to be learned in a set of 30 training trials. Each training trial presented 2 words and 2 referents with no within-trial information about which word went with which referent. Across training trials, labels and their referents always co-occurred. Thus, if infants register and track the co-occurrence information across trials, they should, as in the original experiments, be able to determine the underlying word-referent pairs. Within this overarching structure, the design imposes blocks of 5 trials in which one object (unique to that block) is repeatedly presented at the same location and paired on successive trials with five different objects, a local sequence that might be expected to bias attention to the one new object on each trial, that is to the location that changes trial to trial within a block. Across the set of 30 trials, there were six blocks of 5-trials each with a different object selected to be the repeated object within each block. Thus, each word-referent was the repeated word and object across the 5 trials in one of the six blocks in the course of training. By our description of the task structure in terms of visual attention, there are at least three different factors, operating at different time scales, that may influence how much infants look at the repeated and changed objects within a trial: (1) the increasing local novelty of the changing object (relative to the repeated object) across the 5 trials *within a block*; (2) the familiarity of individual visual objects should increase *across blocks* and thus potentially diminish these local novelty effects; and (3) the number of correct word-referent co-occurrences, if registered, should increase the strength of correct associations *across all learning trials*.

There is, however, another description of the structure of this task that is not based on kinds of visual attention. If infants are trying to solve a reference problem by testing hypotheses about word-referent meanings, then the arrangement of trials may be ideal for learning. From a strictly statistical evidence point of view, the first 5 trials in Table 1 provide unambiguous information that word “A” refers to object a; and if learners adhere to some form of mutual exclusivity (see Markman, 1990; Halberda, 2006; Yu & Smith, 2011), there is also unambiguous evidence that “B” refers to b and word “C” refers to c and so forth. Thus, by a hypothesis testing and statistical learning account that assumes all information presented is considered by the cognitive system, the structure of the task should be highly supportive of learning.

Finally, our central question concerns the nature of individual differences in cross-situational learning. If the attentional processes that organize learners’ and nonlearners’ looking behavior are fundamentally different –with learners’ looking organized by learned associations (or by the goal of testing hypotheses about words and references) but nonlearners’ attention is driven by more local visual effects, then the present arrangement of learning trials should heighten these individual differences. Such a result would provide insight into the possible origins of the individual differences observed in the Yu and Smith study, to the mechanisms that underlie successful cross-situational learning, and to the limitations of this kind of learning mechanism. Such a result might also provide a link between studies of statistical word-referent learning and other evidence that indicates a predictive relation between measures of visual attention in early infancy and later

vocabulary development (Tamis-LeMonda & Bornstein, 1989; Dixon & Hull Smith, 2008), an issue we consider in the Discussion.

Method

Participants

The participants, drawn from a working and middle-class population of a midwestern college town, were 24 12-month-old infants (+/− 4 weeks) and 24 14-month-old infants (+/− 4 weeks). Within each age group, half the participants were male and half were female. Three additional infants began but did not finish the experiment.

Vocabulary measure

All parents were asked to complete the MacArthur Communicative Development Inventory: Words and Gestures (Fenson et al, 1994), a measure of children's vocabulary and vocabulary size. Using this checklist parents reported the words that their infant *comprehended* and the number that they *produced.*

Stimuli and design

The 6 "words" for the cross-situational learning task --*bosa, gasser, manu, colat, kaki and regli* – were the same as those used by Smith and Yu (2008) and Yu and Smith (2010). They were recorded by a female speaker in isolation and were presented to infants over loudspeakers located at both sides of the screen. The 6 "objects" were brightly colored drawings of novel shapes (the same as used in the previous studies). The names and objects were randomly paired as corresponding words and referents. On each trial, two objects (12 by 14 inches in projected size and separated on the screen by 30 inches) were simultaneously presented on a 47 by 60 inch white screen.

There were 30 training slides. Each presented two objects on the screen for 4 sec; the onset of the first word was presented at 368 msec after the onset of the slide and the second word at 1850 msec after the onset of the slide. The mean duration of the spoken words was 745 msec (range 570 to 960); thus, there was at least 500 msec between the offset of the first word and the onset of the second. Across trials, the temporal order of the words and spatial order of the objects were varied such that there was no relation between temporal order of the words and the spatial position of the referents. Over the series of 30 training trials, each correct word-object pair occurred 10 times and each word also co-occurred with each of the other five objects twice. Training trials were arranged in a blocked fashion with each of the 6 blocks defined by the one word-referent pair that repeated in that block and such that, within the block, the repeated object occurred on the same side of the slide. Thus, the 10 repetitions of each word-referent pair consisted of five times with the object in the role of the repeated object within its block and five times in the role of one of the varying objects (once in each of the other five blocks). The order of trials within a block and the order of blocks were randomly determined. These 30 training trials were followed by the test trials.

There were 12 test trials (2 per target word), each 8 seconds in duration. Each test trial presented one word, repeated 5 times, with 2 objects – the target and a foil – in view. The five word repetitions occurred at the 0 (onset of trial), 1.8, 3.5, 5.2 and 6.9 secs points in the 8 sec trial. Since only one word is presented, if the that word cues attention to the target object – the object has co-occurred most often with the word, then infants should look to that target object. Looks to the target at test are considered correct responses. The foil that was pitted against the target object was drawn from the training set. Each of the 6 words was tested twice. The foil for each trial was randomly determined such that each object occurred twice as a foil over the 12 test trials. The left-right locations of objects on the test slides and

the order of test trials was randomly generated with the target appearing on the left on half of the slides and on the right on the other half.

Procedure

Infants sat (on their mother's lap) 3.5 feet in front of the screen with the mother's chair set at the center of the screen. Infants' direction of eye gaze was recorded from a camera centered at the base of the screen and pointed directly at the child's eyes. Parents were instructed to keep their own eyes shut through the entire procedure so as to not influence their infant's behaviors. A camera directed on the parent throughout the procedure confirmed their adherence. As in Smith and Yu (2008), centering slides were presented at the beginning of the procedure and were interspersed periodically (every 3 to 5 slides but not coincident with the start of a new training block) during training. These centering slides consisted of a centered presentation of a Sesame street character (3 sec). The entire procedure took about 4 minutes.

Coding

The direction of eye-gaze of the infant during training and test was determined frame by frame (30 frames per sec) by a coder using the MacShapa coding system (Sanderson et al, 1994) who decided for each frame whether the infant's direction of eye gaze was to the left, right, or center of the screen or whether it was not toward the screen. A second coder independently coded a random sample of 25% of the frames. Agreement on the coding of these frames was 98%. Because the analyses also concern the fine-grained dynamics of looking, we also examined the reliability of coders' timing of shifts in looking behavior. The second coder scored a random sample of 25% of the frames in which the main coder had marked a shift in looking from the just previous coded frame. The two coders agreed on the same shift direction within 1 frame of each other on 92% of these trials.

Results

We present first the data on performance at test and how, from these data, we partitioned infants into learners and nonlearners. We also analyze the receptive and productive vocabulary sizes of learners and nonlearners using the parent report measure of vocabulary. Second, we consider the effects of the main manipulation, the local novelty effects, on the looking behavior of learners and nonlearners during the training trials across several temporal scales –within and across blocks. Third, we determine the experienced word-referent correspondences for individual infants and the relation of these experienced statistics to learning. Finally, we present finer-grained analyses of the role of words in cueing looking behavior during the training trials.

Learners and nonlearners

The task used in this experiment differs from those used in previous experiments in that the stimuli were arranged to create local novelty effects that were unrelated to the word-referent correspondences. In comparison to previous findings, the overall performance of the infants at test suggests that these local novelty effects made the learning of the underlying word-referent correspondences more difficult; in contrast to both Smith and Yu (2008) and Yu and Smith (2010), there is no overall difference between looks to target and foil on the test trials for either 12-or 14-month olds, mean proportion (of the 8 sec trial) looking to the target versus the foil was .43 versus .38 for the 12-month- olds and was .39 versus .33 for the 14-month-olds ($t(23) < 1.2$, $p > .30$ in both cases). This is potentially interesting in its own right as the statistical evidence for the word-referent correspondences is unaffected by this arrangement even if learners attend primarily to the locally novel object and indeed, this arrangement might be expected to lead to better learning based on some hypothesis testing

accounts because the repetition of word-object pair across two consecutive trials provides infants with the relevant information to immediately confirm or reject their particular pairing hypotheses one trial to the next. But overall infants did not learn the pairings as they did in previous studies with randomly ordered presentations of the very same information.

However, the evidence also suggests that individual infants did learn the underlying correspondences; t-tests (with the 12 test trials as the random factor, $t(11) > 2.20$, $p < .05$) were conducted on each individual's durations of looks to target and foil during test to determine whether individual infants reliably looked at the target more than the distractor. These analyses indicated that 19 (9 younger and 10 older) of the 48 subjects looked reliably longer at the target than the distracter during test, showing clear evidence of learning the word-referent pairings. These 19 infants were classified as learners and the remaining 29 infants were classified as nonlearners. Figure 2 shows the frame-by-frame analysis of looking at test: the proportions of infants during each frame who were looking at the target or the foil. Figure 2 specifically shows the mean proportion of infants looking to the target and foil for the temporal window of 1.7 seconds after the onset of the word on the test trial; this is averaged across the 5 repetitions of the test word within a test trial/ Thus, the effect of the word on infant looking is seen, for learners, from the start of period. The 1.7 second window spans the onset from the first repetition to the onset of the next. As is evident, the presentation of a word at test strongly cues visual attention to the associated target object for both younger and older learners, but not for the nonlearners.

The learners and nonlearners also differed in their receptive and productive vocabulary sizes as measured by the MCDI, a fact that in and of itself shows a relation between performance in laboratory cross-situational learning task and real-world word learning. For the 12-month-olds, the mean receptive vocabulary of the learners was 102 words (s =64.3) and their productive vocabulary was 19.5 words (s=19.5); for the 12-month-old non-learners, the mean receptive vocabulary was 53 words (s= 59.1) and the productive vocabulary was 9 words (s=13.8). For the 14-month-old learners, the mean number of words in receptive and productive vocabulary, respectively, were 180.6 (s=41.0) and 47.9 (s=47.9), but for the nonlearners they were 106.5 (s=56.7) and 20 (s=21.6). The number of words reported to be in the infants vocabulary was submitted to a 2 (Age) by 2 (Learner/NonLearner) X 2 (Receptive/Productive Vocabulary) yielded highly reliable main effects of Age, $F(1,44)=15.91$, $p<.001$, and Learn/Not learn, $F(1,44)=13.47$, $p<.001$. There was also, of course, a highly reliable main effect of receptive versus productive vocabulary, $F(1,44) = 130.1$, $p < .001$, which interacted with age, $F(1,44) = 9.39$, $p < .01$, as the difference between receptive and productive vocabulary by parent report increased with age. Within an age group, there were no reliable differences in the ages of the learners and the nonlearners, $t<1.00$, in both cases. In brief, the infants who learned the word-referent correspondences by aggregating co-occurrences across individually ambiguous learning trials were the infants with the most advanced vocabularies for their age. The underlying skills that support cross-situational word learning thus appear to be related to vocabulary growth. The main goal of the remainder of the analyses is to try to understand those skills by determining how learners and nonlearners responded to the challenge of local novelty, a challenge that may have competed with registering and evaluating the co-occurrence statistics of words and objects.

Looking during learning

For these main analyses of looking behavior during the training trials (as well as in the measure of looking at test in Figure 2), we used looking time to potential referents as a proportion of *total* looking time (rather than considering only the relative proportion of looks to the two objects). For learning, in our view, the total amount of time looking at the objects (versus looking off screen) and not just a preference for looking at one object versus

the other is the relevant measure. This is the proper measure (rather than proportions of total looking that leave out looks away from the screen) because it seems highly likely that learning depends on actually looking at the objects for some duration. Further, the present approach is transparently provides information about total looking time since the sum of the looks to the two objects is the measure of mean total looking time. Moreover, proportions of looking time are not stable when calculated over look durations that are very small. Finally, when we removed trials in which infants looked at the screen for the less than 1 sec (of the 4 sec training trial in which proportions of total looking are not meaningful), analyses based on proportions yielded the same conclusions.

Local novelty within a block—These analyses examine changes in looking behavior within block as function of the number of repetitions of the repeated object. The dependent measures are the proportions of time within a trial that the infant looked to the repeated and the novel object during each training trial. A 2 (Age) X 2 (Learners/Nonlearners) by 6 (Block) by 5 (Repetition) ANOVA revealed main effects of Block, $F(5, 225) = 9.43$, $p < 0.001$) and Repetition ($F(4,180) = 12.91$, $p<. 01$) and a marginal interaction between Learners/Nonlearners, Repetition and Block, $F(20,900) = 1.50$, $p= .073$. The effect of Age did not approach significance, $p > .80$. As shown in Figure 3, there were strong effects of the manipulation of local repetition and thus local visual novelty for both Learners and Nonlearners (collapsed across age because of the lack of age differences). The general pattern is this: Infants look equally often to both objects on Trial 1 within a block; on this trial both the to-be-repeated and to-be-varied objects have changed with respect to the *just previous* trial (the last trial in the previous block). With increasing repetitions of one object at the same location within that block, all infants look less to that repeated object and more to the new object on each trial, showing the attentional draw of the contextually novel object.

Changes in looking across blocks—These analyses consider how the preference to look at the contextually novel rather than the repeated object changes across blocks, which might be expected as infants become more familiar with each of the individual objects and if they are learning word-referent correspondences that might compete with these (potentially weakening) novelty effects. Again, the dependent measures are the proportions of times infants looked to the contextually novel and repeated object averaged across repetitions in a block. A 2 (Age) by 2 (Learning versus No Learning) by 6 (Block) by 2 (Repeated/Varying Object) analysis of variance yielded reliable main effects of Repeated/Varying, $F(1,45) =151.84$, $p< .001$) and Block ($F(5,225) = 7.51$, $p < .001$). The interaction between these two factors was also reliable, $F(5,225) = 5.15$, $p < .001$). The main effect of Block is evident in Figure 4 which shows the proportion of time infants looked to the repeated and novel object within a trial across the 6 training blocks (collapsed across trial within a block and collapsed across Age for Learners and Nonlearners). Again, neither the main effect of Age nor any interactions with this factor approached significance. Critically, however, there was a reliable interaction between Learners/Nonlearners and Repeated/Varying, $F(1,45) = 7.16$, $p < .01$. As is evident in Figure 4, the local novelty effect, the preference to look at the new object on each trial lessened over training blocks for learners but not nonlearners. In brief, a new object at one location relative to an unchanged object at the other appears to capture visual attention. However, these local novelty effects diminish across blocks of trials for learners, but not for nonlearners, perhaps reflecting their learning of word-object associations. However, this difference between learners and nonlearners is clearly one of degree. Still, it is consistent with the idea that nonlearners' attention, relative to that of learners, is more sensitive to local stimulus factors and less sensitive to long term regularities.

Accumulated statistics—Statistical learning from word-referent co-occurrences could emerge from the passive registration of words heard while looking at an object, with each perceived co-occurrence of the heard word and the seen object adding a tally to the co-occurrence matrix, regardless of the reason that visual attention might have been directed to the object. In this view, statistical learning would depend only on the data itself and thus the differences between learners and nonlearners should lie in the different statistics gathered by the two groups of infants, which is the main result observed by Yu & Smith (2010). In the present study, learners –because they became less influenced by local novelty over time – may have collected better statistics. Alternatively, the underlying processes directing attention may matter as to whether the associations between heard words and seen objects are registered as such.

Accordingly, following the approach used by Yu and Smith (2010), we created an "experienced" co-occurrence matrix for words and objects for each infant. An experienced co-occurrence was defined as look to an object *just after* the heard word. More precisely, and using the same temporal windows as in Yu and Smith study that was the impetus for the present study, we counted the total looking times to the two objects beginning from the 500 msec following the word's onset to the onset of the next word (for the first word) or the end of a trial (for the second word in the trial, see Yu and Smith, 2010, for details). The definition of the relevant window as beginning 500 msec after word onset is based on the assumption that it takes at least 300 msec to process and recognize a word and that it takes at least 200 msec for infants to plan and execute an eye movement (see Yu and Smith, 2010). So defined, looking times to both objects in this window were added across trials to create a 6-word by 6-object co-occurrence matrix. The mean of the individual experienced co-occurrence matrices derived in this way are shown in Figure 5. As is apparent, there is no difference between learners and nonlearners at either age group. This result is in direct contrast to the findings of Yu & Smith (2010) who showed that learners –through their looking behavior –collected better statistics than nonlearners such that the strength of correct associations in the co-occurrence matrix strongly predicted subsequent performance at test. The present results show that although experiencing the relevant statistics may be necessary to learning it is not sufficient.

Although the co-occurrence matrices for learners and nonlearners are the same and the evidence within the matrices looks strong for the underlying word-referent correspondences, there is no evidence that the words cued looking behavior in either learners or nonlearners during the training trials. More specifically, we scored each look during the presentation as correct (using the same temporal window used to create the co-occurrence matrices) if the look was to the referent of the just presented word and incorrect if it was to the other object on the screen. Figure 6 shows the results for younger and older learners and nonlearners as a function of training block. As is evident and as confirmed by a 2 (Age) by 2 (Learners/nonlearners) by 6 (Block) by 2 (Correct/Incorrect) analysis of variance, there were no reliable effects ($p > .30$). By this measure, there were no cueing effects of associated words on visual attention during the learning phase for either learners or nonlearners and there was no increase in cued looks to the referent of the word across blocks of training. Critically, however, at test, which occurs right after block 6 but within which there are no competing local novelty effects and just one repeated word, we know –as shown in Figure 2 – that the statistically correct word cues attention to the associated object for learners but not for nonlearners. We know from the analyses shown in Figure 4, that learners –at the end of training but before test – are beginning to be influenced by something other than novelty. The implication is that this learning was too weak to clearly show itself prior to the test trials.

In brief, the potentially aggregated statistics –*if* they depend *only* on hearing a name while looking at an object -- appear to be the same and therefore cannot explain the differences in individual performances at test. However, even when one's sensors come into contact with an auditorally presented word or a looked-at object, a perceiver might not actually bind the information together and store the heard word and seen object as an association. The first step to forming such an association is to perceive and register both the seen object and the heard word. The effect of local visual novelty shows that both learners and nonlearners were –at the very least –attending to the visual objects. The next set of analyses show that both learners and nonlearners were also listening and that the presented words had direct effects on visual attention for both learners and nonlearners.

Words do affect looking—Although there were at best minimal effects of the associated word on looking during training even for learners, there were unexpected and strong effects of the words on looking for all infants. These effects appear unrelated to the word-referent co-occurrence statistics. Figure 7 shows the frame-by-frame analyses of looking behavior during the 4 sec training trials: the proportion of infants looking during each frame that were looking at the varying, locally novel, object. The two vertical lines indicate the onset of the presentation of the first and second word presented during each training trial (collapsed across whether the first or second word referred to the changed, that is, novel object, on that trial). The darker line shows these average proportion of infants looking to the varying object for the first two blocks. The dashed lines shows the average proportion of infants looking to the varying object for the last two blocks. The middle two blocks are not shown (to make the pattern more easily discerned) but fall intermediate between the first two and last two blocks. At the start of the training trial, when the visual objects are presented and *before the first presentation of the first word*, there is no visual preference for the varying object. It is the auditory presentation of the first word that directs attention to an object. For all infants, looks to the locally novel object increases markedly *after the first word is presented*. This first auditory stimulus (half the time the name of the repeated object and half the time the name of the varying object) appears to force visual selection of the *more locally novel object* over the object that is repeated from the just previous trial. The effect is remarkably strong and uniform across all infants, seeming almost reflexive. At the point of 750 msec after the trial begins (and thus about 250 msec after the onset of the first word), .82, .90, .93, and .79 of infants in the 12-month-old learner and nonlearner groups and in the 14month-old learner and nonlearner groups, respectively, are looking at the local novel object. The strength and uniformity of this behavior suggests some near mandatory auditory (or word) cueing effect on visual attention to a contextually novel object.

The second word presented in a trial shows a similar, but much weaker, cueing effect, again to the contextually novel object (which by the second word, of course, is less novel). Between the presentations of the first and second word, looks to the contextually novel object diminished. The auditory cueing effect of the first word to the varying object diminishes more from the first to the third pair of blocks for learners than for nonlearners and on later blocks diminishes more within a trial for learners than nonlearners, again suggesting increasingly weaker local novelty effects for learners across the training trials.

This auditory (or word) cueing effect was quantified by comparing the overall fixation duration on the varying object between two temporal windows – the first one is defined as from the onset of the first word to the onset of the second word (the duration between the two lines in Figure 7) and the other window is defined from the onset of the second word to the end of trial (the duration from the second line to the end in Figure 7). Although the global pattern is similar for younger and older learners versus younger and older nonlearners, the fine-grained-dynamics –particularly in the period between the first and second word – appear in the figure to vary with age. Accordingly, we conducted four

separate ANOVAs, one for each combination of age and learning status, to examine these patterns. These analyses confirm the decrease over blocks in looking at the contextually novel object at the presentation of the first word for all four learner-by-age groups over the course of training, that is from the first two blocks to the last two blocks, minimum F ratio, $F(2,39)=8.37$; $p<0.001$. The presentation of the second word also appears to cue attention to the locally novel object and this effect also decreases reliably from the first two to the last two blocks for learners, $F_{12\text{-month-old}}(1,24) =5.94$, $p<0.005$; $F_{14\text{-month-old}}(1,27) = 7.02$, $p<0.005$, but not for nonlearners, $F_{12\text{-month-old}}(1,42) = 0.96$, $p=0.43$; $F_{14\text{-month-old}}(1,39) = 1.02$, $p=0.37$. Thus all infants strongly show the auditory (word) cueing effect but the dynamics of the effect, within a trial and across blocks of trials, are different for learners than nonlearners and for older and younger infants. Looks to the contextually novel object begin to decline *before* the second word is played and thus appear to reflect the dynamics of the auditory cueing effect itself, dynamics that differ for learners and nonlearners and for younger and older infants.

To summarize, nonlearners are more affected by local repetitions and changes than are learners suggesting that attention may be more driven by dynamically local effects for nonlearners than for learners. In contrast, learners' looking behavior changed more over the course of the training trials indicating a stronger role of longer-term dynamics on attention. These differences in looking during the learning trials *do not* affect the experienced statistics of the word-referent co-occurrences and do not appear to result from stronger effects of the associated names on looking for the learners than nonlearners. The word-onset cueing effect provides clear evidence that both learners and nonlearners are attending to the words in the sense that all infants show a strong effect of the first played word in a learning trial on looking behavior and a weaker effect of the second word. But these effects are not dependent on the co-occurrence statistics between the words and referents for either learners or nonlearners. Finally, this "auditory cueing effect" to the locally novel stimulus diminished within and across trials more for learners than nonlearners, again suggesting that aggregated experiences over the long term compete more effectively with dynamically local attentional processes for learners.

Overall, the findings provide at best partial support for the originating hypothesis. As predicted, the nonlearners were more susceptible to local novelty effects and this greater susceptibility. However, the learners were also susceptible to these effects and although their susceptibility diminished over the course of training, we had expected to also see, for learners, increasing cuing of attention to the target object *during training*. However, the word-referent correspondences learned by the learners was only evidence at test, when the challenge of local novelty was removed.

General Discussion

The findings show that attention to the words and the objects –even then the experienced co-occurrences yield clear statistical evidence for the word-referent correspondences --is insufficient for cross-situational word-referent learning. Some infants exposed to the training statistics showed clear behavioral signs of attending to both objects and words but did not learn the correspondences. Moreover, infants who did learn showed no evidence of that learning during training when looking was challenged by contextual novelty, but they did show learning when that challenge was removed at test. These findings place strong constraints on the possible mechanisms underlying statistical word-referent learning. The differences in looking behavior between nonlearners and learners in the learning phase also suggests that attention that is strongly influenced by more transient rather than longer-term regularities may be a marker of poor statistical learning and also of slower vocabulary development more generally. Finally, the frame-by-frame analyses of looking while

listening revealed what appears to be a near mandatory *auditory* (or word) cueing effect to the more contextually novel object in a display. We believe that this is the first evidence of such a phenomenon in infants. In the following, we consider, the implications of these findings for mechanisms of word-referent learning, for different kinds of attention and their relation to learning, and for individual differences in attentional processing as a predictor of individual differences in language learning. We conclude with some reflections on the possible disconnection between macro-level theories of word-referent learning and the more micro-level processes that may implement them.

Explaining cross-situational learning

There are two general classes of theories about cross-situational word-referent learning: associative learning (e.g., Smith, 2000; Yu & Smith, 2010; 2011; Fontanari, Tihkanoff, Cangelosi, Ilin & Perlovsky, 2009; Fazly, Alishahi & Stevenson, 2010) and hypothesis testing (e.g., Frank, Goodman & Tenenbaum, 2009; Snedecker, 2000; Blythe & Smith, 2010; Siskind, 1996). The present findings do not fit well with the usual construals of either of these mechanisms. First, they do not fit with the idea of word-referent learning as resulting from mere exposure to environmental statistics as might be proposed by some simple associative models (see Yu & Smith, 2012). The nonlearners were exposed to the same statistical regularities as the learners; the nonlearners' behavior suggests that they both heard the words and looked at the objects, and indeed the co-occurrence data based on each infants' own looking behavior –such that a co-occurrence is counted only when the infant is looked at an object after hearing its name -- suggests that the nonlearners experienced the same co-occurrence statistics that led to learning by other infants. But apparently these co-occurrences were not registered or not remembered by the nonlearners. In brief, mere exposure to the statistical regularities is not enough for word-referent learning from those regularities.

Second, the findings also do not fit well with usual notions about word-referent learning as hypothesis testing (e.g., Gillette, Gleitman, Gleitman & Lederer, 1999; Halberda, 2006; Xu & Tenenbaum, 2007). A growing number of researchers are using moment-to-moment eye-gaze in looking-while-listening paradigms (e.g, Halberda, 2006; Vouloumas & Werker, 2009; Marchman & Fernald, 2008; Yu & Smith, 2010) to make inferences about infant hypotheses in language processing and lexical learning tasks, and these looking behaviors have shown signs consistent with and revealing of possible internal decision processes directed at the disambiguation of word-referent correspondences (Fernald, Zangl, Portillo, & Marchman, 2008; Halberda, 2006). However, in the present study, the learners' looking behavior during the learning phase provided no evidence of attention to the word-referent correspondences and no evidence of attempts at disambiguation through such constraints as mutual exclusivity, which given the ordered structure of the training trials might have been expected to play a role. Moreover, the learners showed no evidence of confirming hypotheses during training in that they did not increasingly look at the target object for the heard referent as the statistical evidence for those correspondences increased. Instead, attention during the training trials for learners as well as nonlearners appeared to be primarily controlled by factors *unrelated* to the word-referent correspondences, that is, although these influences were greater for the nonlearners than the learners. The experimental design added local novelty with the goal of challenging attention to word-referent correspondences in order to determine whether all forms of attention support learning. This local novelty manipulation did make the task harder, as shown by the lack of success of many infants under this procedure compared to the earlier experiments (Smith & Yu, 2008).

The main difference found between learners and nonlearners is one of the degree of these dynamically local effects on attention. However, learners relative to nonlearners did show a greater sensitivity to the more global structure of the training sequence in the decreasing pull of local novelty over the training trials. Critically, though, during the training phase, learners did not look to the object labeled by a word and did not show evidence of learning until test, when the challenge of contextual novelty was no longer present. It is as if the learners were registering co-occurrences in the background while visual attention in the moment was organized by other more dynamically local factors, a result that might seem more like passive associative learning except for the critical fact that nonlearners exposed to the same statistics were unable to do this. Clearly, the findings do not contradict all possible variants of associative-learning nor hypothesis-testing accounts (see Yu & Smith, 2012) but they would seem to rule out passive associative learning from mere exposure or the necessity of active, explicit, hypothesis testing through the disambiguation for word-referent learning during training.

What we know from the results –and what will need to be explained by any theory is this: Attention strongly driven by contextual novelty competes with rather than supports statistical learning. Infants, the learners, can learn –showing evidence at test of learned correspondences –*without* showing any evidence during the training phase of strengthening word-referent associations. Thus a further lesson from the present results is that looking behavior –the behavior that instigates learning and that is also the standard measure of that learning in infants –is multiply determined –there multiple kinds of visual attention – and thus looking must be an imperfect measure of learning.

Attention and statistical word-referent learning

Any mechanistic account of cross-situational word learning –be it associative or hypothesis testing --has to assume that learners register word-referent *co-occurrences* not just words and referents as separate events unbound to each other. This is the first step to learning and prerequisite to the aggregation and evaluation of evidence for word-referent pairings. Successful aggregation and evaluation of the statistical evidence also requires long-term memory for bound words and referents. Thus there are two inter-related hypotheses as to what learners may have achieved during training that nonlearners did not: (1) the binding of heard words to seen objects, and (2) the formation of longer term rather than transient memories for those bound elements. Both of these hypotheses –and avenues for future empirical research –are informed by recent advances in the study of different kinds of attention and their different consequences for cognitive processing.

The attention literature in both adults and infants strongly suggests fundamental differences between attention that is captured by stimulus salience versus attention that is driven by longer-term associations (Colombo, 2001; Ristic & Kingstone, 2009; Wu and Kirkham, 2010; S. Smith & Chatterjee, 2008; Snyder & Munakata, 2008, 2010). The distinction in the adult literature is often characterized in terms endogenous, expectancy-driven, attention versus exogenous, stimulus-driven, attention. The difference between these two kinds of attention is readily seen in detection experiments. For example, an arrow that points to a location leads to faster reaction time to detect the target at that location in comparison to when the target's location is uncued. However, a salient flicker of light near the target location --a cue that presumably draws attention to the location through low level and involuntary processes --does not lead to more rapid detection (Posner, 1980; Jonides, 1981; Wright & Ward, 2008, for a review). The assumption is that the arrow has its pointing effect through the previous learning of its directional meaning, and is therefore a top-down cue for attention. Thus, the findings in this literature are also characterized in terms of the different cognitive consequences of top-down and bottom-up attention.

More recently, the distinction between attention cued by learned expectations versus stimulus salience has been proposed to also be critical to binding elements from different sensory modalities, such as binding a sound and a sight (e.g., Fiebelkorn, Foxe, & Molholm, 2010; Talsma, Senkowski, Soto-Faraco & Woldroff, 2010). The distinction between endogenous and exogenous cueing in multi-sensory phenomena (e.g. Fiebelkorn et al, 2010) is sometimes discussed not in terms of expectations versus stimulus effects but in terms of the role of longer-term representations versus transient working memory representations in linking events across modalities. Finally, in a recent study of infant multi-sensory binding, Wu & Kirkham (2010) showed that a salient visual cue next to a visual target blocked the learning of auditory-visual associations in 8-month-old infants whereas an expectancy-based cue that directed attention to the visual target supported learning. All of these findings are consistent with the idea that the *kind of attention,* the specific attentional mechanisms engaged by the task, may be critical to binding heard words to seen objects and thus to the statistical learning of word-referent correspondences.

The manipulations in the present study do not map directly onto the distinction between bottom-up exogenous attention and top-down endogenous attention as defined in the adult visual attention literature in that contextual novelty is not a *stimulus* property per se. Novelty effects depend on building working memory representations of the repeated stimulus (Turk-Browne, Scholl, & Chun, 2008; Schoner & Thelen, 2006). Nonetheless, as transient memory effects, attention driven by local stimulus changes may be akin to exogenous cueing and not support robust multi-sensory processing. Starting with this conjecture and in light of the related findings in the adult visual attention literature, we offer the following hypothesis: Nonlearners failed to bind the heard words to the seen objects and thus failed to gather evidence on word-referent correspondences. In other words, whereas the correspondence matrices for the learners in Figure 5 may be "psychologically accurate" in the sense of measuring the seen objects that co-occurred with the heard words for individual learners, these co-occurrences were not registered by the nonlearners. In this view, the nonlearners were not building co-occurrence matrices at all. For nonlearners relative to learners the more transient attentional processes may have remained stronger throughout the learning trials because the nonlearners did not learn the associated cues; in constrast, if learners were registering the co-occurrences, these may have effectively dampened the effects of contextual novelty. In brief, the advantage of learners relative to nonlearners may depend not on differences in these transient pulls on attention, contextual novelty grabs everyone's attention, rather the advantage of the learners may depend primarily on the better of the correspondences. Yu and Smith (2010) showed that cross-situational learning depended on looking behavior that yielded statistical evidence for the underlying word-referent pairs. Here, we propose that in addition, infants have to connect the words to the referents so as to register and store *co-occurrences* for those statistical evidence to matter.

Visual habituation and vocabulary development

The above interpretation suggests that learning associations dampens the strength of more local and transient pulls on attention. However, an explanation in the opposite direction is also possible: learners who habituated to the novelty effects fasters may have had a cognitive system more open to binding words and referents and to accumulating evidence over the longer term. Several longitudinal studies have shown that individual differences in early visual habituation predicts individual differences in later cognitive development (e.g., Fagan, Holland & Wheeler, 2007), including language learning (Tamis-LeMonda & Bornstein, 1989). The link between visual habituation and vocabulary found in previous research, unlike the present study, is *predictive*: rate of visual habituation at 5 months predicts larger vocabulary size at 13-months (Tamis-LeMonda & Bornstein, 1989; Dixon & Hull Smith, 2008). Rate of habituation to visual repetitions in 2 to 8 month old infants has

also been shown to predict performance in a variety of other cognitive tasks well into childhood and perhaps even in adults (Bornstein & Sigman, 1986; Gilmore & Hoben, 2002; Rose, Feldman & Jankowski, 2004; Fagan, Holland & Wheeler, 2007). Rapid habituation may be a signature of the ability to form more robust and longer lasting visual memories, the kinds of memories needed for learners to aggregate statistics across trials and also needed to build more expectancy and top-down forms of attention, the kinds of attention that support deeper processing and the formation of multi-sensory representations. Alternatively, or in addition, visual habituation may be crucial for other kinds of learning in that it frees the learner from the idiosyncracies of the specific context to discover the latent structure across contexts. These are critically important issues for determining the mechanistic bases of cross-situational word-referent learning, and language learning, and the observed individual differences.

A new auditory cueing effect

The most unexpected aspect of the present results, and perhaps also the singularly strongest result in the experiment is what appears to be a near-mandatory auditory cueing of visual attention, an effect that we do not believe has been reported in infants before. The start of any trial began with two visual objects on the screen. As evident in Figure 7, prior to the onset of the first word, infants *did not* look preferentially to the contextually novel object but looked equally often to both objects. However, just after the onset of the first word, virtually all infants looked to the contextually novel object. The magnitude and uniformity of this effect during the early training blocks, as shown in Figure 7, is remarkable. It is reflex-like and unrelated to the statistical association of the attended object and the specific word. Although we know of no report of this sort of cueing effect in the infant literature, there is a potentially related phenomenon in adult attention that is sometimes discussed as a multisensory alerting event: an auditory signal with a rapid rise time (but no prior relation to the presented visual information) leads to the rapid visual selection of the one different visual object in an array of like objects (e.g., Shinn-Cunningham, 2008; McDonald, Teder-Salejarvi & Hillyard, 2000; Vroomen and de Gelder, 2000). The dynamics of the adult effect are rapid and complicated and may not line up exactly with the effects observed here. But our finding from infants may be a developmentally early version of what has been observed in adults. With this new finding, there are a great many questions to be answered, but they are intriguing and potentially developmentally important: If words by their auditory properties alone bring visual attention to the more contextually unusual object in a scene, then words –perhaps very early and perhaps before word learning –may be organizing attention in ways that encourage social interaction and learning.

Macro and micro conclusions

Theories of word learning are often formed at the macro-level in terms of theoretical constructs about the nature of knowledge and operations on that knowledge, for example, in theoretical claims about hypotheses, concepts, constraints, referring and inferences. These constructs have worked well in capturing the evidence from experiments that measure macro-level behaviors such as the object chosen by a child in a name comprehension task or total looking time in a preferential looking task. However, with advances in more micro-level measures and analyses of behavior, using techniques such as the tracking of momentary eye gaze, experiments are beginning to reveal the micro-level structure underneath these macro-level behaviors. New research using these methods both in the study of on-line word comprehension and word-learning by infants (see, e.g., Fernald, Perfors & Marchman, 2006; Fernald et al, 2008) has led to new insights about the role of priming and lexical competition in word comprehension and word learning. This tension between macro- and micro- level in explanations is also seen in the adult literature particularly with respect to distinctions between cognition and perception as growing evidence shoes that conceptual

knowledge affects even the earliest stages of visual processing (e.g., Lupyan, Thompson-Schill & Swingley, 2010; Lupyan & Spivey, 2010). These newer findings, though not directly at odds with macro level concepts, also do not map in simple one-to-one ways onto those macro-level accounts, in part because the micro level is less about the knowledge that children or adults might have than about the dynamic processes that activate and create knowledge.

From this more micro-level perspective, the present results and our interpretation of them suggest that the binding of a co-occurring word and referent to form a unified multisensory memory – a cell in a co-occurrence matrix -- is critical to cross-situational learning. At the micro-level this may be implemented by the formation of multi-sensory associations, and these may depend critically on top-down expectations about where to look and what might be seen there rather than bottom-up or more transient influences on looking behavior. This mechanistic conjecture is not so much at odds with the notion of words as “referring” as perhaps providing hypotheses about the mechanisms through which “referring” is implemented.

Micro-level analyses that capture behavior in tasks at time scales not examined before also seem likely to reveal new phenomena. That is the case in the present study: the auditory (or word) cueing effect to the locally novel object was not expected and its potential importance to infant attention or word learning is not yet known. But it seems likely given the nearly mandatory and reflexive nature of this behavior on the part of the infants in this study that at least some portion of the total looking-time measure used in standard preferential-looking studies of word learning included looks driven –not by knowledge about words and their referents – but by such an auditory cueing effect. Thus, we are immersed in an exciting time: as researchers detail the micro-structure behind our macro-level constructs, they will give those constructs a richer bases in the temporal dynamics of the underlying processes and perhaps provide keys to linking them to neural mechanisms. But it is also possible that our methodological advances will lead to insights about micro level processes that just do not correspond to macro-level theoretical constructs at all (see Fodor, 1975).

Acknowledgments

We thank Char Wozniak for collection of the data, Justin Halberda for very helpful comments on an earlier version of this paper, and the reviewers for cogent criticisms and comments. This research was supported by National Institutes of Health R01 HD056029.

References

Baldwin DA. Early referential understanding: Infants’ ability to recognize referential acts for what they are. Developmental Psychology. 1993; 29(5):832–843.

Benitez, VL.; Smith, LB. Predictable locations aid early object name learning. 2012. (Under review)

Blythe RA, Smith K, Smith ADM. Learning times for large lexicons through cross- situational learning. Cognitive Science. 2010; 34:620–642. [PubMed: 21564227]

Bornstein MH, Sigman MD. Continuity in mental development from infancy. Child Development. 1986; 57(2):251–274. [PubMed: 3956312]

Brent MR, Siskind JM. The role of exposure to isolated words in early vocabulary development. Cognition. 2001; 81:B33–B44. [PubMed: 11376642]

Chater N, Tenenbaum J, Yuille A. Probabilistic models of cognition: Conceptual foundations. Trends in Cognitive Sciences. 2006; 10(7):287–291. [PubMed: 16807064]

Colombo J. The development of visual attention in infancy. Annual Review of Psychology. 2001; 52:337–367.

Dixon WE, Hull Smith P. Attentional focus moderates habituation-language relationships: Slow habituation may be a good thing. Infant and child development. 2008; 17:95–108.

Fagan JF, Holland CR, Wheeler K. The prediction, from infancy, of adult IQ and achievement. Intelligence. 2007; 35(3):225–231.

Fazly A, Alishahi A, Stevenson S. A probabilistic computational model of cross-situational word learning. Cognitive Science. 2010; 34:1017–1063. [PubMed: 21564243]

Fenson L, Dale PS, Reznick JS, Bates E, Thal DJ, Pethick SJ. Variability in early communicative development. Monographs of the Society for Research in Child Development. 1994; 59(5) Serial no 242.

Fernald A, Perfors A, Marchman V. Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the second year. Developmental Psychology. 2006; 42:98–116. [PubMed: 16420121]

Fernald, A.; Zangl, R.; Portillo, A.; Marchman, V. Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In: Sekerina, IA.; Fernández, EM.; Clahsen, H., editors. Developmental Psycholonguistics: On-line methods in children's language processing. John Benjamins; Amsterdam: 2008. p. 97-135.(2008)

Fiebelkorn IC, Foxe JJ, Molholm S. Dual mechanisms for the cross-sensory spread of attention: how much do learned associations matter? Cerebral Cortex. 2010; 20(1):109–120. [PubMed: 19395527]

Fontanari J, Tikhanoff V, Cangelosi A, Ilin R, Perlovsky L. Cross-situational learning of object-word mapping using Neural Modeling Fields. Neural Networks. 2009; 22:579–585. [PubMed: 19596549]

Fodor, JA. The language of thought. Harvard University Press; Cambridge: 1975.

Frank M, Goodman N, Tenenbaum J. Using Speakers' Referential Intentions to Model Early Cross-Situational Word Learning. Psychological Science. 2009; 20(5):578–585. [PubMed: 19389131]

Gillette J, Gleitman H, Gleitman L, Lederer A. Human simulations of vocabulary learning. Cognition. 1999; 73:135–176. [PubMed: 10580161]

Gilmore RO, Thomas H. Examining individual differences in infants' habituation patterns using objectives quantitative techniques. Infant Behavior & Development Special Issue: Variability in Infancy. 2002; 25(4):399–412.

Halberda J. Is this a dax which I see before me? use of the logical argument disjunctive syllogism supports word-learning in children and adults. Cognitive psychology. 2006; 53(4):310–344. [PubMed: 16875685]

Hollich GJ, Hirsh-Pasek K, Golinkoff RM. Breaking the language barrier: An emergentist coalition model for the origins of word learning. Monographs of the Society for Research in Child Development. 2000; 65(3 Serial No 262)

Jonides, J. Voluntary versus automatic control over the mind's eye's movement. In: Long, JB.; Baddeley, AD., editors. Attention and performance IX. Hillsdale, NJ: Erlbaum; 1981. p. 187-203.

McDonald JJ, Teder-Salejarvi WA, Hillyard SA. Involuntary orienting to sound improves visual perception. Nature. 2000; 407:906–908. [PubMed: 11057669]

Marchman V, Fernald A. Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. Developmental Science. 2008; 11:F9–16. [PubMed: 18466367]

Markman EM. Constraints children place on word learning. Cognitive Science. 1990; 14:57–77.

Posner MI. Orienting of attention. Quarterly Journal of Experimental Psychology. 1980; 32:3–25. [PubMed: 7367577]

Regier T. The emergence of words: Attentional learning in form and meaning. Cognitive Science. 2005; 29:819–865. [PubMed: 21702796]

Ristic J, Kingstone A. Rethinking attentional development: Reflexive and volitional orienting in children and adults. Developmental Science. 2009; 12(2):289–296. [PubMed: 19143801]

Rose SA, Feldman JF, Jankowski JJ. Dimensions of cognition in infancy. Intelligence. 2004; 32(3): 245–262.

Shinn-Cunningham BG. Object-based auditory and visual attention. Trends in Cognitive Sciences. 2008; 12:182–186. [PubMed: 18396091]

Sanderson PM, Scott JJP, Johnston T, Mainzer J, Wantanbe LM, James JM. MacSHAPA and the enterprise of Exploratory Sequential Data Analysis (ESDA). International Journal of Human, Computer Studies. 1994; 41:633– 681.

Schöner G, Thelen E. Using dynamic field theory to rethink infant habituation. Psychological Review. 2006; 113:273– 299. [PubMed: 16637762]

Scott R, Fisher C. 2-year-olds use distributional cues to intepret transitivity-alternating verbs. Language and Cognitive Processes. 2009; 24:777–803. [PubMed: 20046985]

Siskind JM. A computational study of cross-situational techniques for learning word-to-meaning mappings. Cognition. 1996; 61(1–2):1–38. [PubMed: 8990967]

Smith SE, Chatterjee A. Visuospatial attention in children. Archives of Neurology. 2008; 65(10): 1284–1288. [PubMed: 18852341]

Smith, K.; Smith, A.; Blythe, R.; Vogt, P. Lecture Notes in Computer. 2006. Cross-situational learning: a mathematical approach.

Smith L, Yu C. Infants rapidly learn word-referent mappings via cross-situational statistics. Cognition. 2008; 106(3):1558–1568. [PubMed: 17692305]

Smith, LB. How to learn words: An associative crane. In: Golinkoff, R.; Hirsh-Pasek, K., editors. Breaking the word learning barrier. Oxford: Oxford University Press; 2000. p. 51-80.

Smith LB, Colunga E, Yoshida H. Knowledge as process: Contextually cued attention and early word learning. Cognitive Science. 2010; 34:1287–1314. [PubMed: 21116438]

Snedecker, J. In: Clark, E., editor. Cross-Situational Observation and the Semantic Bootstrapping Hypothesis; Proceedings of the Thirtieth Annual Child Language Research Forum.; Stanford, CA: Center for the Study of Language and Information; 2000.

Snyder KA, Blank MP, Marsolek CJ. What form of memory underlies novelty preferences? Psychonomic Buletin & Review. 2008; 15:315–321.

Snyder HR, Munakata Y. Becoming self-directed: Abstract representations support endogenous flexibility in children. Cognition. 2010; 116(2):155–167. [PubMed: 20472227]

Snyder HR, Munakata Y. So many options, so little time: The roles of association and competition in underdetermined responding. Psychonomic Bulletin & Review. 2008; 15(6):1083–1088. [PubMed: 19001571]

Talsma D, Senkowski D, Soto-Faraco S, Woldorff M. The multifaceted interplay between attention and multisensory integration. Trends in Cognitive Science. 2010; 14:400–410.

Tamis-LeMonda CS, Bornstein MH. Habituation and maternal encouragement of attention in infancy as predictors of toddler language, play, and representational competence. Child Development. 1989; 60:738–751. [PubMed: 2737021]

Turk-Browne N, Scholl B, Chun M. Babies and brains: Habituation in infant cognition and functional neuroimaging. Frontiers in human neuroscience. 2008; 2:1–8. [PubMed: 18958202]

Tomasello, M. Perceiving intentions and learning words in the second year of life. In: Bowerman, M.; Levinson, S., editors. Language acquisition and conceptual development. Cambridge University; 2000. p. 111-128.

Vouloumanos A. Fine-grained sensitivity to statistical information in adult word learning. Cognition. 2008; 107:729–742. [PubMed: 17950721]

Vouloumanos A, Werker J. Infants' learning of novel words in a stochastic environment. Developmental Psychology. 2009; 45:1611–1617. [PubMed: 19899918]

Vroomen J, de Gelder B. Sound enhances visual perception: cross-modal effects of auditory organization on vision. Journal of Experimental Psychology: Human Perception & Performance. 2000; 26:1583–1590. [PubMed: 11039486]

Waxman S, Gelman S. Early word learning entails reference not merely associations. Trends in Cognitive Science. 2009; 13:258–263.

Woodward, A.; Markman, E. Early word learning. In: Damon, William, editor. Handbook of child psychology: Volume 2: Cognition, perception, and language. Wiley; Hoboken, NJ: 1998. p. 371-420.(1998)

Wright, RD.; Ward, LM. Orienting of attention. Oxford: Oxford University Press; 2008.

Wu R, Kirkham NZ. No two cues are the same: Depth of learning in infancy is dependent on what orients attention. Journal of Experimental Child Psychology. 2010; 107(2):118–136. [PubMed: 20627258]

Yoshida H, Smith LB. Correlations, concepts, and cross −linguistic differences. Developmental Science. 2003; 6(1):30–34.

Yu C, Ballard D, Aslin R. The role of embodied intention in early lexical acquisition. Cognitive Science: A Multidisciplinary Journal. 2005; 29(6):961–1005.

Yu C, Smith L. Rapid word learning under uncertainty via cross-situational statistics. Psychological Science. 2007; 18(5):414–420. [PubMed: 17576281]

Yu C, Smith L. What you learn is what you see: using eye movements to understand infant cross-situational statistical learning. Developmental Science. 2010

Yu C, Smith L. Hypothesis testing versus associative learning in cross-situational word-referent learning: Prior questions. Psychological Review. 2012; 119:21–39. [PubMed: 22229490]

Yurovsky, D.; Yu, C. In: Love, BC.; McRae, K.; Sloutsky, VM., editors. Mutual Exclusivity in Cross-Situational Statistical Learning; Proceedings of the 30th Annual Conference of the Cognitive Science Society; Austin, TX: Cognitive Science Society; 2008. p. 715-720.

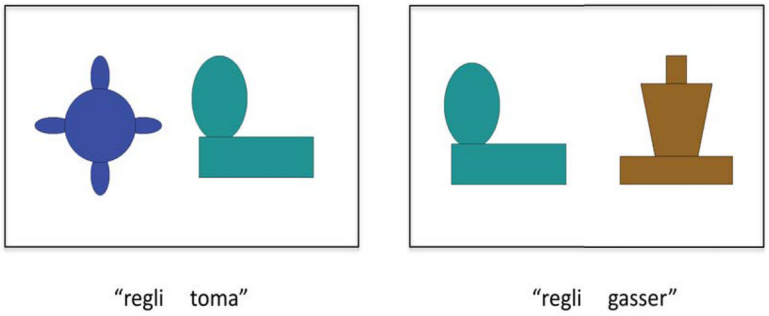

Figure 1.
Examples of two trials in the cross-situational learning task.

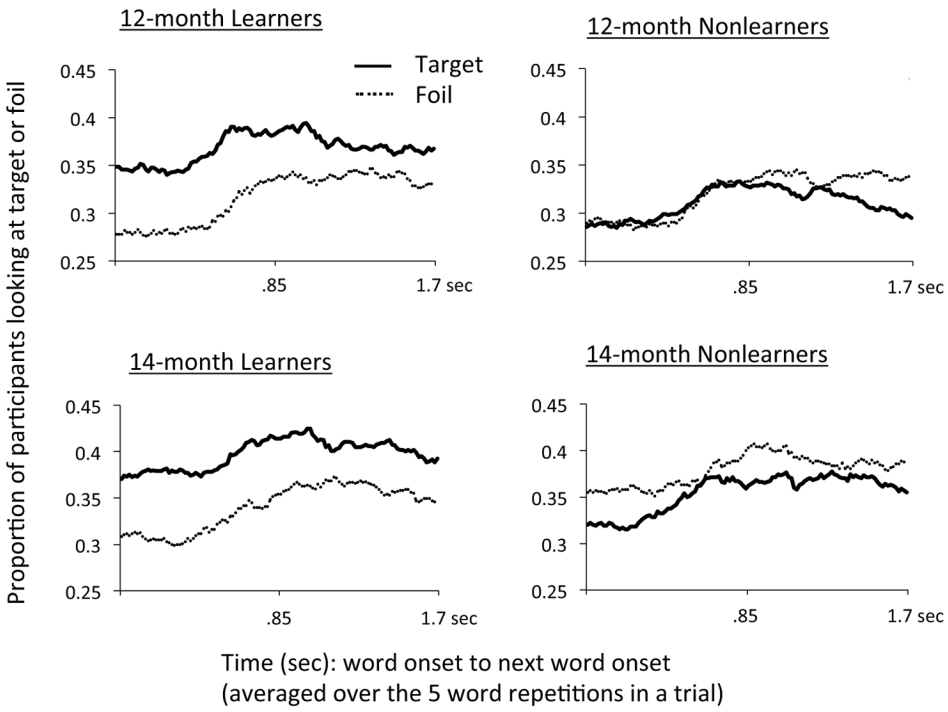

Figure 2.

Proportion of participants coded as looking at the target object and the foil object on each video frame (30 frames per sec) during the testing phase. The proportion of participants looking to each object is shown from the onset of the tested word to the onset of its repetition in the trial. The proportion of participants looking to the two objects is averaged over the five repetitions of the words in each test trial and across the 12 test trials.

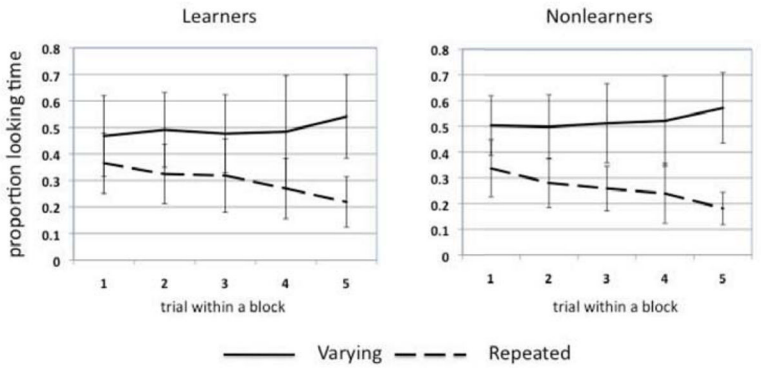

Figure 3.
Mean proportion total looking (and standard deviations) within a trial to the varying and to the repeated object as a function of the number of repetitions of the repeated object within a block (averaged across all 6 blocks) for Learners and Nonlearners.

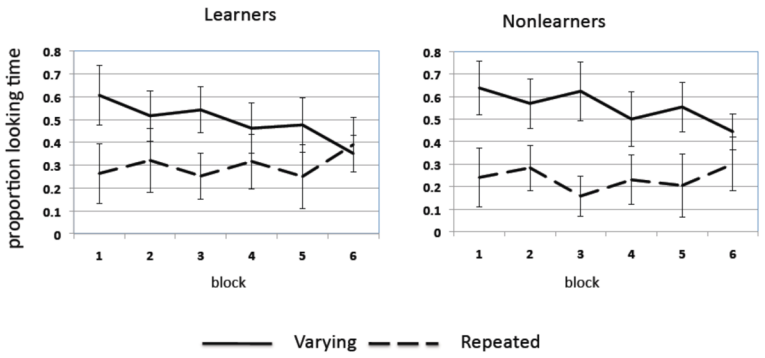

Figure 4.
Mean proportion of looking (and standard deviations) to the varying and repeated object as a function of block for the Learners and Nonlearners (averaged across trials in a block).

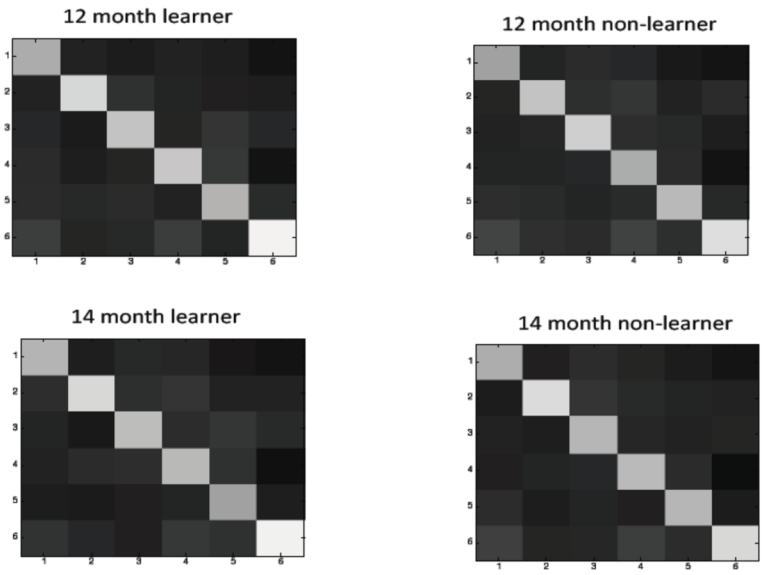

Figure 5.

Accumulated statistics for the four groups of participants calculated as in Yu and Smith (in press). Each cell represents the association probability of a word-object pair determined from the synchrony between a subject's looking to an object at the presentation of a word. The diagonal items are correct associations and other non-diagonal items are spurious correlations. Dark means low probabilities and white means high probabilities.

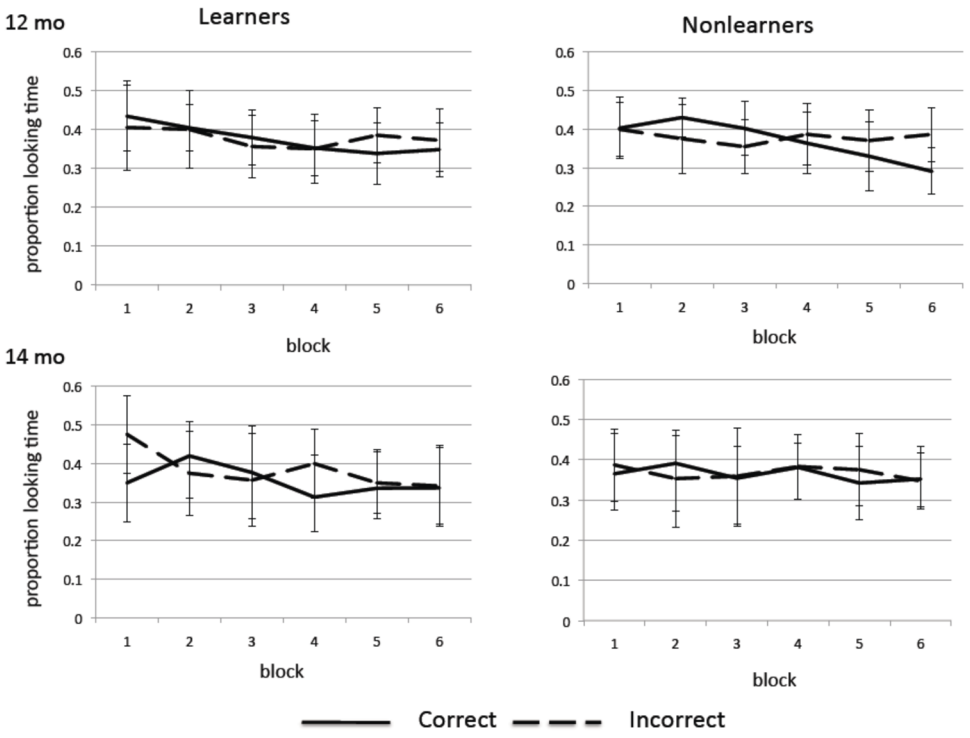

Figure 6.

Mean proportion looks (and standard deviations) to the two objects *just after* (see text for definition) hearing a word *during training* as a function of block (averaged across the 5 trials within a block and for both words presented within a training trial). “Correct” indicates looks to the target object that across trials is statistically associated with the word; and “incorrect” indicates looks to the other non-associated object during that same time period.

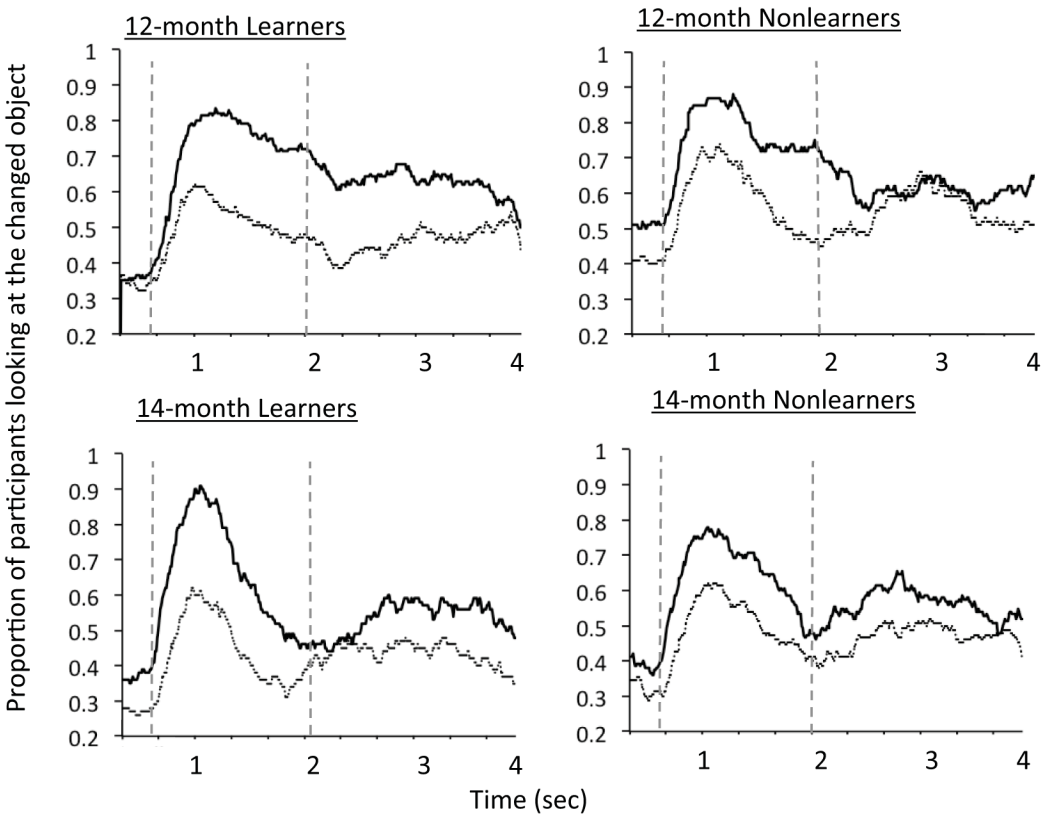

Figure 7.

Mean proportion of infants looking to the varying (changed) object on each video frame (30 frames per sec) within a 4 sec training trial. The solid line indicates the averages across the first two blocks (10 trials) and the dotted line indicates the averages on the last two blocks (10 trials). Thus the vertical lines indicate the onsets of the two words during the training trial.

Table 1

The structure of the training trials. Words are indicated by uppercase letters and objects by lowercase letters. The 30 trials are divided into 6 blocks, with each block defined by the word that is repeating. Potential influences on visual attention at three time scales are illustrated by the lines: (1) the increasing local novelty of the non-repeated object (relative to the repeated object) increases within each block; (2) the familiarity of individual visual objects increases across blocks; and (3) the number of correct word-referent co-occurrences increases across the 30 training trials. *Spatial and temporal order variation of words and objects is not indicated in this table, nor is the order of trials within a block (which is randomly determined; see text for clarification).*

Block/trial	Words	Objects	Time Scales
			Local Novelty Item Familiarity Associations
Block 1 Trial 1	A-B	a-b	
Trial 2	A-C	a-c	
Trial 3	A-D	a-d	
Trial 4	A-E	a-e	
Trial 5	A-F	a-f	
Block 2 Trial 6	B-A	b-a	
Trial 7	B-C	b-c	
Trial 8	B-D	b-d	
Trial 9	B-E	b-e	
Trial 10	B-F	b-f	
-	-	-	
-	-	-	
-	-	-	
Block 6 Trial 26	F-A	f-a	
Trial 27	F-B	f-b	
Trial 28	F-C	f-c	
Trial 29	F-D	f-d	
Trial 30	F-E	f-e	