Linking Learning to Looking: Habituation and Association in Infant Statistical Language Learning

Daniel Yurovsky, Shohei Hidaka, Chen Yu, and Linda B. Smith
{dyurovsk, shhidaka, chenyu, smith4} @indiana.edu
Department of Psychological and Brain Science, and Cognitive Science Program
1101 East 10th Street Bloomington, IN 47405 USA

Abstract

Recent experiments have shown the importance of statistical learning in infant language acquisition. Computational models of such learning, however, often take the form of corpus analyses and are thus difficult to connect to empirical data. We report a cross-situational learning experiment which demonstrates robust individual differences in learning between infants. We then present a novel generative model of cross-situational learning combining two competing processes – *habituation* and *association*. The model's parameters are set to best reproduce each infant's individual looking behavior from trial-to-trial in training and testing. We then isolate each infant's word-referent learning function to explain the variance found in preferential looking tests.

Keywords: statistical learning; computational modeling; cognitive development; language acquisition

Introduction

Language acquisition should be hard but young children nonetheless move from babbling to complex sentences in a remarkably short time. One might thus expect the underlying language learning mechanism to also be complex, involving constraints and biases (Markman, 1990) and sophisticated inferences (Xu & Tenenbaum, 2007). However, even if the final mechanism is complex, it must begin with something simple – language learners develop. By understanding the tools available to very young learners, we may develop insight into how more complex mechanisms are created and how they might be understood as products of simpler mechanisms.

One candidate for a simple mechanism is the accumulation of associations between words and objects in a child's ambient environment (Hollich, Hirsh-Pasek, & Golinkoff, 2000, Smith, 2000). If the co-occurrence structure of the world is informative, such that words frequently occur with the objects they label, a child who can attend to this information could find a wedge into learning the more complicated structural aspects of her language (Landau, Smith, & Jones, 1988).

Recently, Smith and Yu (2008) have provided evidence of just such a sensitivity in 12 and 14-month-old infants. In the cross-situational learning paradigm, infants are exposed to a series of individually ambiguous learning trials containing multiple words and objects. While each trial contains several potential mappings, some of which are spurious, a child who can attend to the overall co-occurrence structure can unambiguously determine the correct mappings.

Attempts to understand the mechanism underlying this competence, however, have been aimed primarily at the abstract computational level. Computational models have taken the form of corpus analyses (Fazly, Alishahi, & Stevenson, in press, Frank, Goodman, & Tenenbaum, 2009, Yu, 2008) and thus have resisted direct comparison to empirical data. To understand the mechanisms available to budding language learners, however, models must account for and explain the *behavior* of young infants.

Because preferential looking is the primary measure of learning in studies of preverbal infants, it is this looking data that computational model must explain. Yu and Smith (in press) took a first step towards this goal. Using an associative model, they found that the *number* of words for which an individual infant showed preferential looking behavior was predictable from that infant's own eye movement in training. This might seem to fit an associative learning mechanism: one learns to associate words to the objects at which one is looking when one hears the words. However, Yu and Smith were unable to predict *which* word-referent mappings were learned. If associative learning is the relevant mechanism, something is still missing.

We propose to take two more steps towards understanding the mechanism supporting cross-situational learning. First, whereas Yu and Smith's model was descriptive – using patterns in training behavior to predict test behavior – we present a *generative* model of eye movements. That is, we construct a model which produces eye-movement behavior matching that of infants during training, and then show that the same model accounts for the test data. Second, we predict not only *how many* word-referent mappings each infant learned, but also *which ones*. This modeling is done at the *individual infant* level, allowing us to explain behavior as it unfolds trial-by-trial throughout training and testing.

To motivate our model, we first present results from a cross-situational learning experiment with 15-month-old infants. Analysis of preferential looking test results shows robust individual differences among infants, underscoring the importance of understanding cross-situational learning at a process level. We then construct a model that generates fixations through the competition of two well-known processes that organize infant behavior and learning – *association* (Smith, 2000) and *habituation* (Hunter & Ames, 1988. Model parameters are fit to best account for each individual infant's looking behavior over the course of the experiment, and then inferences about learning are drawn from these parameter fits.

Experiment

Method

Infants were exposed to a cross-situational word learning task (Smith & Yu, 2008; Yu & Smith, in press). Each child viewed a series of trials pairing two novel objects with one novel label. While the correspondence between words and objects on an individual trial was ambiguous, cross-trial co-occurrence statistics between words and objects indicated the correct pairings. After 60 training trials, preferential looking tests were used to determine whether infants had learned the correct pairings.

Participants. Twenty-five 15-month-old infants (14 females, *M* = 14 mos, 23 days, range: 13;22 to 16;4) composed the final sample. Twelve additional infants were excluded due to fussiness (*N*=11) or experimental error (*N*=1).

Stimuli. Six pseudoword labels were recorded by a female native English speaker in isolation and presented to infants over loudspeakers. Six novel two-dimensional objects, each a unique bright color, were presented to infants two at a time on a 47" by 60" white screen. All stimuli were constructed to be comparable to those used in previous cross-situational learning experiments (Smith & Yu, 2008, Yu & Smith, in press).

Procedure. Infants sat on their mother's laps 3.5 feet away from a large white projection screen. Direction of gaze was recorded by a Tobi X60 eye-tracker as well as a camera directed at the child's eyes. Parents were instructed to shut their eyes during the course of the experiment so as not to influence infant behavior.

Training consisted of 60 2-second long training slides. Each slide presented two objects, one on each side of the screen, and was accompanied by one of the recorded labels. A slide's label was presented 700ms after the objects' onsets. On each slide, one of the objects was the label's correct referent and one was a foil. This correspondence was uncorrelated with spatial location, but could be determined from cross-trial co-occurrence statistics: each label occurred 10 times with its correct referent and only 2 times with each of the other objects. Training trials were interspersed with presentations of Sesame Street characters intended to maintain infant attention. Total training lasted approximately 4 minutes.

Following training, infants were exposed to 6 testing trials, each 8 seconds long. Test trials began with approximately 1 second of silence, followed by six repetitions of a label – each separated by 1 second. Two objects were visible for the entire 8 seconds – the label's correct referent and a distractor object. Each of the 6 labels was tested once, and each object appeared equally often as a target and a distractor. Figure 1 illustrates the time course of training and testing with sample trials.

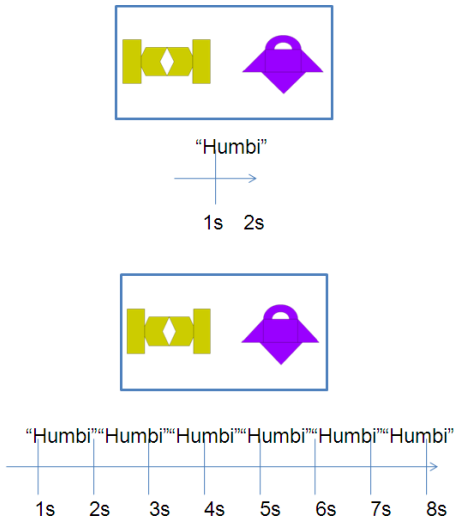

Figure 1: The time course of training (above) and testing (below) trials. Infants saw two objects and heard a label produced either once (training) or 6 times (test). The first 1 second window of each was silent; every subsequent window contained an auditory label.

Data. Gaze position was recorded via eye-tracker at a rate of 50Hz. Because of movement or looking away during the experiment, there were some discontinuities in automatic gaze recording. On average, 57.8% of each infant's gaze points were recorded. Naïve coders blind to the contents of each slide coded each of the remaining frames for direction of gaze (left, right, away/unknown). After hand-coding, 74.5% of all gaze points were mapped to a screen position where one of the objects appears.

Results and Discussion

Infants looking times to target and distractor objects on each of the 12 preferential looking test trials were submitted to a 2 (Target/Distractor) x 6 (Word) x 25 (Subject) mixed ANOVA. The analysis revealed no main effects, but showed a highly significant interaction between Target/Distractor and Subject (F = 3.66, p < .001, η^2 = .1). Individual infants thus showed reliably different looking patterns at test: some looked reliably longer at targets than distractors; others looked reliably longer at distractors than targets (Figure 2). This is consistent with previous work on slow vs. fast habituators (Cashon & Cohen, 2000, Schöner & Thelen, 2006).

Why should there be reliable individual differences? It is well known that the function that maps learning onto looking is *nonmonotonic* – it switches directions (Hunter & Ames, 1988 – Figure 3). This complicates the interpretation of looking behavior, with some investigators of word learning behavior suggesting that increased looking to the target indicates learning (e.g. Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987) whereas others interpret increased looking to the distractor as evidence of learning via violation of expectation (e.g. Stager & Werker, 1997).

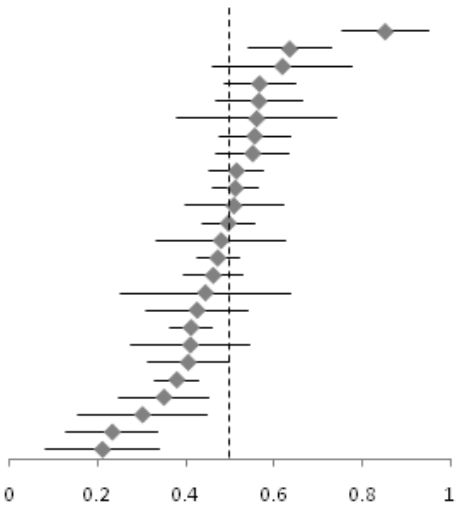

Figure 2: A plot of mean(std err) preferential looking to target for each infant. Values greater than .5 indicate an average preference for the target; those less than .5 indicate preference for the distractor.

The above analysis indicates that individual infants show reliable looking patterns when tested for their preference to look to or away from a label's referent. However, since individual infants show different patterns, it is unclear how to interpret their behavior. For which infants should we infer learning? In the following computational modeling effort, we propose to show that an unambiguous answer can be found through model selection. If we are explicit about the mechanisms which combine to generate looking behavior, we can ask if a learning mechanism is necessary to explain individual infants' looking behavior at test.

Computational Model

Throughout the experiment, infants were exposed to a series of slides presenting two objects along with an auditory label word. Infants responded to these stimuli - at any point in time - by fixating one of the two objects on the screen. Our goal was to derive a generative model for each infant that produced fixation patterns that best approximated his or her own generated fixations.

Because of the structure of the training and testing trials, we divided the time course of fixations into a series of 1 second bins (Figure 1). Proportion of looking to each of the two on-screen objects was calculated in each such window, and model was fit to this data.

Conceptually, the model is simple. Let us suppose that fixation patterns within a given window are generated by the combination of two processes: *habituation* to each of the objects on the screen, and *association* between each of the objects and the label being heard. Let us also suppose that each of these processes is a function of looking time to the input. However, because we do not know the true form of these functions (although see Schöner & Thelen, 2006), we *approximate* them with arbitrary degree polynomials. These polynomial approximations allow us to make inferences about the shape of the functions without making claims about their exact form.

We use each infant's individual training data to infer the *habituation* and *association* functions which best account for that infant's behavior. Because one cannot learn what one does not see, habituation and association are functions of gaze duration rather than occurrence frequency. We thus produce an explicit *linking function* from learning to looking at test, and this function is used to infer what each child learned from her looking behavior in training. Doing so allows us to move beyond preferential looking as a measure of learning, and to make deeper and more specific conclusions about the mechanisms supporting cross-situational learning in real time.

Data

In the experiment, infants were exposed to 60 training trials followed by six test trials. The label for each 2s trial was heard 700ms into the trial. Adding 367ms to the label's onset to account for processing time (Swingley & Aslin, 2000) results in two ~1s windows (Figure 1, top). In the first window, we assume that fixations are being driven by the objects (*habituation*) only, and in the second we assume that fixations are driven both by the objects (*habituation*) and the co-occurring word (*association*).

Test trials had a similar structure (Figure 1, bottom). Each began with a short period of silence, followed by the onset of a label which was then repeated 5 more times at 1 second intervals. We divide each testing trial into 7 1s windows: 1 in which fixations are driven only by the objects, and 6 in which fixations are driven by both objects and the label. The natural logarithm of the odds of looking at each of the on-screen objects was computed in each window, and these are the data to which the model was fit. Log odds is similar to proportion of looking, but has nicer mathematical properties for this particular analysis (see also, Barr, 2008). Any windows in which there was no fixation data for an infant were left out of that infant's dataset.

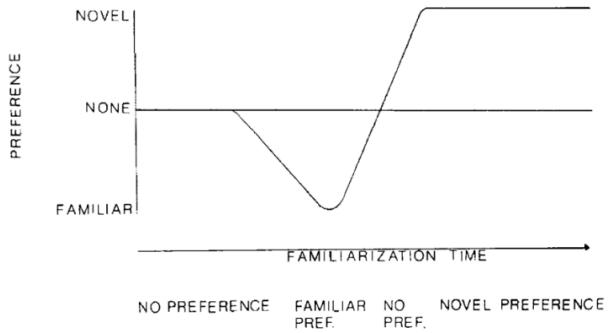

Figure 3: A schematic of the infant looking preference function reproduced from Hunter and Ames (1988). Because the function is *nonmonotonic* – direction of preference changes in opposite directions across time – looking time data resists straightforward interpretation.

Model Description

In a given window, each of the two objects had an activation level as described below. Odds of looking to each of the objects were computed using the ratio form of the Luce choice axiom (Luce, 1959). We additionally adjust the odds ratio in two ways.

Because saccades are controlled by a vision system subject to physical constraints, and because we model looks in 1 second windows, the current window depends on the location of the eye in the last window. For this reason, as an approximation, we modify $OddsLook_t(O)$ by a parameter p times the odds of looking in the previous window to the object in O's current location. Further, infants are known to display preferences for one side of the screen over another. For this reason, we build in a constant term b which models each infant's potential preference for the left or right side of the screen.

Thus, on trial t, if objects O_1 and O_2 are present,

$$OddsLook_t(O_1) = \frac{e^{Act_t(O_1)}}{e^{Act_t(O_2)}} \times \Big(b \cdot \big((Loc(O_1) = left)\big)\Big) \times \Big(p \cdot OddsLook_{t-1}(Loc(O_1))\Big)$$

Silent Window. In a window in which no label is present, activation is driven by an infant's *habituation* to each of the objects present. Habituation to an object was approximated by an arbitrary degree polynomial function *habit* evaluated on the cumulative looking time to that object so far in the experiment. Estimation of the parameters of this function for each infant will be described below.

$$Act_t(O) = habit(O)$$

Label Window. For windows in which a label was heard, we assume that activation is also driven by the association between each object and the label W. For these windows,

$$Act_t(O) = habit(O) + assoc(O, W)$$

Association and Habituation. Each infant's individual *habituation* and *association* functions were approximated by arbitrary degree polynomial functions. For each infant, all possible orders 0 to 2 were tried for each function, with the optimal parameters fit as described below. The final order of each function was chosen using AIC to be the most parsimonious fits for the infants looking behavior.

Formally, if t_o is cumulative looking time to an object, and $t_{o|w}$ is cumulative looking time to an object in the presence of a word,

$$habit(t_o) = \sum_{n=1}^{N_h} h_n \cdot {t_o}^n \qquad assoc(t_{o|w}) = \sum_{n=1}^{N_a} a_n \cdot {t_{o|w}}^n$$

Thus, one infant might have a quadratic *habituation* function (N_h = 2) and a linear *association* function (N_a = 1), while for another infant the best model may have been a linear *habituation* (N_h = 1) function and no *association* function (0 degree) at all.

Model Fitting

In order to determine the best approximation to each infant's individual learning functions, we constructed all 9 possible combinations of orders 0 to 2 for both *association* and *habituation* functions. The optimal parameters for each function were selected to best account for the infant's fixation data. Subsequently, model selection using AIC was performed for each infant by selecting from these models the one which also gave the best account of the individual infant's testing eye fixations without overfitting.

Results and Discussion

On average, the best generative model for each infant predicts a significant (r = .307, p < .001) proportion of the variance of looking. In comparison, a null model, which includes only a side bias (b) and inertia (p) term for each infant picks up a significantly small proportion of the variance (r_g = .307, r_n = .203, t = 2.68, p = .01).

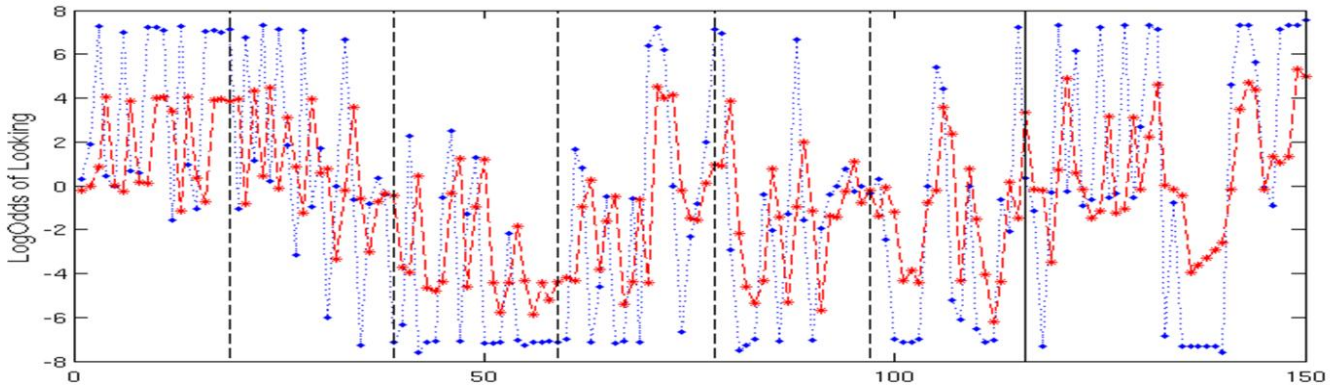

Figure 4: Log odds looking to the left side of the screen for one infant across both training and testing. Positive log odds indicate a preference for the left. Black dashed lines separate every 10 training trials and the black solid line indicates the start of testing. Infant behavior is the blue line with solid markers, model behavior is the red line with asterisk markers.

This indicates that *habituation* and *association* account for a significant proportion of each infant's looking behavior, both in training and testing. Further, functions which are appropriate for describing training can also describe test behavior. An example of the model's fit to one infant is shown in Figure 4 above.

Now that we have found the best model which accounts for each infant's looking behavior, we can determine which infants are likely to have learned word-referent mappings. Preferential looking behavior, while a good measure of learning at a group level, can be quite difficult to interpret at the individual level (Aslin, 2007; Houston-Price, Nakai, 2004; Hunter & Ames, 1988). There are several reasons for this. First, as mentioned above, the function which links learning to looking is *nonmonotonic*, and different infants learn at different rates. Hence whether preference for target or distractor should indicate learning in an individual infant is unclear. Second, as we have explicitly modeled, there are two principled reasons to move one's eyes in this task – in response to the objects on the screen (habituation), and in response to the relationship between objects and words (association). If we are interested in word-object mapping, then movement resulting from the first process adds noise to our measurement. Because both processes were modeled explicitly, however, we can probe association directly.

For each infant, model selection was used to determine which order polynomial best matched his or her association and habituation functions. If the optimal order of association for an infant was nonzero, then we can infer that the infant learned associations between words and objects. Thus, another way to measure whether an infant learned word-object associations is to ask about the order of that infant's association function. Of the 25 participants, 11 were best described as being driven by an associative process ($N_a > 0$). We can then look at what these association functions predict in the infant's test behavior.

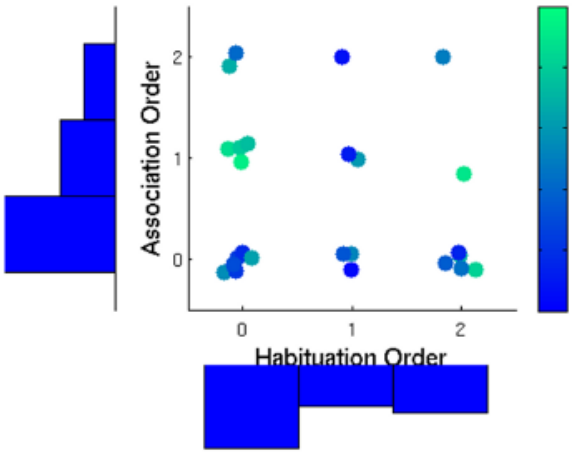

Figure 5: Distribution of association and habituation orders which best account for each infant. The scale on the right ranks infants from strongest preference for distractor (bottom) to strongest preference for target (top). Association order is correlated with strength of absolute preference.

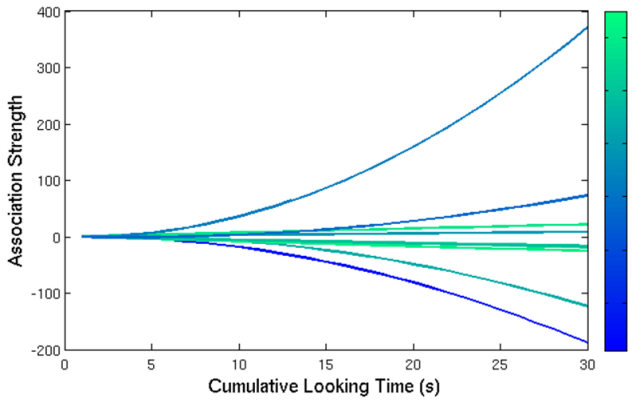

Figure 6: Theoretical association functions for each infant plotted over the course of 30 seconds of co-occurrence. The scale on the right ranks each infant by the strength of their preference for target or distractor. Throughout the entire 30 seconds, there is a significant correlation between the strength of the preference and the strength of the theoretical association function.

Figure 5 shows the distribution of association and habituation orders of the polynomial functions which best accounted for each individual infant's looking behavior. Points representing individual infants are color-coded by the strength of their preference for target(green) or distractor(blue). Analysis shows that the order of an infant's association function is strongly correlated with the strength of that infant's absolute mean preference in the 6 preferential looking trials (r = .55, p < .01). That is, the stronger an infant's preference at test (either for target or distractor), the higher the order of the association function that best described his or her data.

Second, because we have explicitly determined the polynomial function which best describes each infant's association learning, we can examine these functions in isolation. Figure 6 shows the association function for each infant plotted over 30 1 second exposures to a hypothetical word and object. We can compare the ordering of these functions – rank them in the order of their value after each window – and compare this to the mean preference for the target exhibited by each infant over the 6 test trials. The correlation between ordering and mean preference is significant at the .05 level over the entire course of the comparison, and peaks at four seconds (r_4 = .771, p < .001). This finding indicates that these theoretical learning functions are in deeply linked to preferential looking performance at test. The functions thus allow us to predict which infants will show *familiarity* preferences at test, and which infants will show *novelty preferences*. The two figures also reinforce the fundamental importance of understanding individual differences if we are to understand statistical learning. The 25 individual infants displayed the entire gamut of possible learning functions, and these functions fit sensibly to their looking performance at test.

General Discussion

Learning word-referent associations in cross-situational experiments, and in the world, must depend on moment-to-moment *behavior* of *individual* infants – what is looked at and when – and the co-occurrence of objects seen and words heard. Looking behavior, in turn, depends on previous experience in multiple ways and through multiple mechanisms. Two of these fundamental mechanisms are *habituation* and *association.* Repeated experience with an object increases the tendency to look away, but repeated experience with the object in a word's context increases the tendency to look towards the object in its presence.

The present analyses show what can be gained by attempting to understand the dynamic processes that underlie the *behaviors* used as indices of learning. Constructing trial-by-trial models of individual infants looking behavior in word-referent learning yields two major benefits. First, since looking behaviors themselves are commonly used as indices of learning, it allows us greater certainty in inferring learning in infants. Second, because we can track individual infants across the course of learning, it gives us a deeper theoretical understanding of how the mechanisms underlying this learning.

This work thus makes both specific and general contributions. First, the generative model of eye movements in cross-situational learning explains individual infant behavior in both training and testing. Second, we have delineated the interacting effects of two competing processes which produce infant eye fixations – habituation and association – and showed how they can be analyzed independently. Third, our experiment and model demonstrate the possibility of understanding cross-situational learning at the individual infant level, of making sense of the different ways in which different infants learn. Finally, we have demonstrated a novel methodology for analyzing infant learning tasks. In addition to the insight gained from preferential looking analysis, this work shows a promising role for model selection and the construction of explicit functions linking learning to looking.

Acknowledgments

This work was supported by a National Science Foundation Graduate Research Fellowship to the first author and National Institute of Health Grant R01HD056029. The authors would also like to thank Amara Stuehling and Melissa Elston for collecting much of the data. Finally, the authors would like to thank Mike Frank for thoughtful feedback on an earlier draft.

References

Aslin, R. N. (2007). What's in a look? *Developmental Science*, *10*, 48-53.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59*, 457-474.

Cashon, C. H., & Cohen, L. B. (2000). Eight-month-old infants' perception of possible and impossible events. *Infancy*, *1*, 429–446.

Fazly, A., Alishahi, A., & Stevenson, S. (in press). A probabilistic computational model of cross-situational word learning. To appear in *Cognitive Science*.

Golinkoff, R., Hirsh-Pasek, K., Cauley, K., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, *14*, 23–45.

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

Houston-Price, C. & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development*, *13*, 341-348.

Hunter, M.A. & Ames, E.W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In L.P. Lipsitt (Ed.), *Advances in child development and behavior* (pp. 69–95). New York: Academic Press.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*. *59*, 299-321.

Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.

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

Schöner, G. & Thelen, E. (2006). Using dynamic field theory to rethink infant habituation. *Psychological Review*, *113*, 273-299.

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

Smith, L. B. & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*, 333-338.

Stager, C. L., Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word learning tasks. *Nature*, *388*, 381-382.

Swingley, D. & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*, 147–166.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*, 245–272.

Yu, C. & Smith, L.B. (in press). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental Science*.