Online Processing of Speech and Social Information in Early Word Learning

Daniel Yurovsky
yurovsky@stanford.edu
Department of Psychology
Stanford University

Anna Wade
wadea@neurosurg.ucsf.edu
Department of Neurological Surgery
University of California, San Francisco

Michael C. Frank
mcfrank@stanford.edu
Department of Psychology
Stanford University

Abstract

Although word learning unfolds over days, weeks, and months, individual naming events are over in a matter of seconds. To benefit from a naming event, children must at least hear the label and see the referent. We tested 1-, 2- , 3-, and 4-year old children in a naturalistic word learning task with two conditions: one that taxed both speech processing and rapid gaze-following, and one in which a social cue-to-reference was available for an extended time. The development of word-learning in the *extended* condition paralleled the development of speech processing, but learning in the *brief* condition lagged behind. However, learning from both the *brief* and *extended* cues was predicted by individual differences in speech processing and cue-following together. Thus, even through the 4th year, real-time processing of social and linguistic information are a critical bottleneck for word learning.

Keywords: Language acquisition, word learning, attention, social cues, development

Introduction

Language learning is a fundamentally social endeavor – it relies critically on input from social partners. This is because many aspects of natural languages, like the mappings between words and their referents, are conventions that can vary from community to community (Chater & Christiansen, 2010). To learn their first words, children must track the relationship between the sounds that speakers produce and the things in their world (Pinker, 1984; Siskind, 1996).

Not everything about a word and its referents needs to be learned in a single shot. Instead, it is likely that this relationship is refined over multiple exposures, over a period of days, weeks or months (Carey & Bartlett, 1978; Smith & Yu, 2008; McMurray, Horst, & Samuelson, 2012). Nonetheless, each individual naming event occurs in real-space and real-time (Samuelson, Smith, Perry, & Spencer, 2011; Spencer, Perone, Smith, & Samuelson, 2011). If a child does not hear the label, or does not see the target referent, the information "available" in the naming event is effectively unavailable to the child (Yu & Smith, 2012; Yurovsky, Smith, & Yu, in press). Thus, a critical bottleneck in language acquisition is the ability to process the right information at the right time (Fernald & Marchman, 2012).

Young children are slow processors of both speech and visual input (Kail, 1991; Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998). They are also slow to re-deploy their attention in response to changing visual information (Colombo, 2001). Thus, it is perhaps unsurprising that the properties of child-directed speech and child-directed actions seem well-designed to scaffold slow processors. Child-directed utterances are slower, shorter, and have larger pitch contours than do utterances spoken to adults. Repetition is common in child-directed speech, and key words are made more salient through minor local variations (Onnis, Waterfall, & Edelman, 2008). In addition, new labels are often introduced through a small set of common naming phrases that facilitate attention tom and learning of, new words (Fernald & Hurtado, 2006; Yurovsky, Yu, & Smith, 2012). Exposure to these kind of structure makes a real difference, predicting individual differences in vocabulary development (Hoff, 2003). Child-directed actions are similarly exaggerated, marked by bigger, simpler, and more repetitive movements (Brand, Baldwin, & Ashburn, 2002). Indeed, multi-modal synchrony between these exaggerated visual and auditory inputs may be precisely the the information that young children use to to learn their first words (Gogate, Bahrick, & Watson, 2000).

However, while social partners may sometimes scaffold young language learners, simplifying their speech and timing their naming events so that labels coincide with the focus of children's visual attention, this kind of pedagogical naming likely accounts for a minority of the relevant input from which words and their referents could be learned. For instance, while isolated words facilitate speech processing and word learning, they make up less than 10% of all speech to children (Brent & Siskind, 2001). Similarly, referential expressions produced while children are already looking at the target object facilitate word learning, but make up only a portion of the naming events produced to young children (Tomasello & Farrar, 1986; Frank, Tenenbaum, & Fernald, 2013). In the remaining naming events, successful learning requires the child to check with the speaker to determine the target of her reference (Baldwin, 1991). Consequently, the ability to quickly process speech, and to quickly follow a speaker's social cues, should both give learners access to more and more useful information.

Over the first two years, children make rapid gains in the rate at which they process auditory input, picking referential words out of continuous speech (Fernald et al., 1998). During this time, individual differences in rate of spoken language processing predict individual differences in vocabulary size (Fernald, Perfors, & Marchman, 2006). Over the same two year period, children also improve in their abilities to attend to and use social cues indicating the target of a speaker's reference (Scaife & Bruner, 1975; Hollich, Hirsh-Pasek, & Golinkoff, 2000). As with speech processing, individual differences in children's gaze- and point-following predict their language development over these two years (Brooks & Meltzoff, 2006).

But, while the second year is marked by an increase in the rate at which children learn new words, this acceleration in

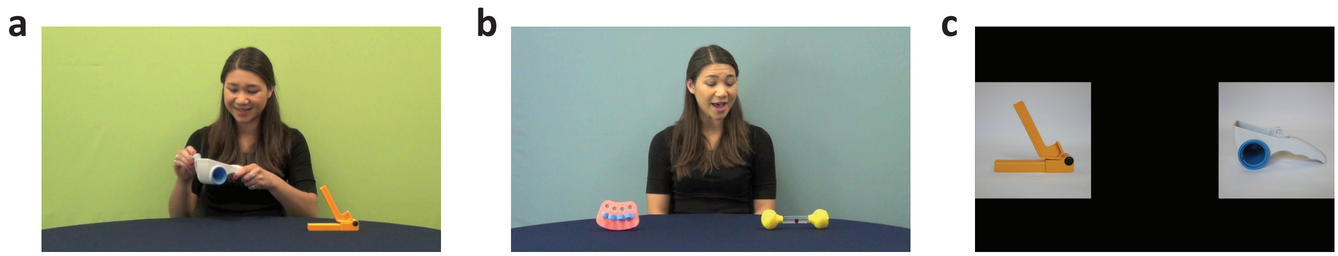

Figure 1: Example training and test trials from the Experiment. On *Extended* Cue trials (a), the speaker picked up and interacted with the target toy over the course of 10-20 seconds, providing consistent social information about her target reference. On *Brief* Cue trials (b), the speaker made a quick glance at the target toy, and then looked forward into the camera for the remainder of the trial. Determining her target of reference on these trials thus required rapid social-cue following. On test trials (c), children saw two of the toys from training and heard the speaker's voice ask them to find the target toy.

vocabulary growth continues into the third and fourth years and beyond (Bloom, 2000). Do these same skills in spoken language processing and rapid social-cue following continue to develop and continue to predict word learning over this extended period of vocabulary acceleration?

To determine how speech processing, social-cue following, and word learning co-develop over the first four years, we tested a large cross-sectional sample of children in age from 1-5 years in a short, naturalistic word-learning task. Over the course of approximately four minutes, we measured each child's ability to process speech containing a known referent, learn a new word when an *Extended* social-cue continuously provided disambiguating information about the target referent, and learn a new word the speaker gave only a *Brief* social-cue to indicate her target referent, requiring rapid gaze-following. We subsequently fit a linear mixed-effects model to children's looking times to determine how speech processing and social-cue following predicted word learning in both conditions over the course of development.

Method

Children's eye movements were tracked while they watched a series of naturalistic word-learning videos. In each, children saw a speaker seated at a table between two novel toys. She introduced them to one of the toys, providing a label and several interesting facts about it. Crucially, on some of the trials she provided an *Extended* Cue indicating the target of her reference – picking up the object and interacting with it over the course of the video. On other trials, she provided *Brief* Cue – only a quick glance to the target object when she first labeled it. After each learning trial, children were tested for their knowledge of the referent for the new word using the Looking While Listening procedure (Fernald et al., 1998). In addition, similar test trials were administered for known objects to measure children's processing of familiar words. Because eye movements were recorded during both learning and test, we were able to analyze the relationship between children's behavior during learning and test trials.

Participants

Parents and their 1–5 year-old children were invited to participate in a short language learning study while they visited the San Jose Children's Discovery museum. All-together, we collected demographic and experimental data from 114 children, 39 of whom were excluded for one or more of the following reasons: abnormal developmental issues ($N = 7$), failure to calibrate ($N = 26$), less than 75% exposure to English ($N = 13$), and fussiness or inattention ($N = 22$). The final sample consisted of 18 1-2 year olds (M_{age} = 1 yr.; 7 mo., 9 girls), 25 2-3 year olds (M_{age} = 2 yr.; 6 mo., 9 girls), 21 3-4 year olds (M_{age} = 3 yr.; 6 mo., 8 girls), and 11 children over the age of 4 (M_{age} = 4 yr.; 8 mo., 5 girls).

Stimuli

The experiment consisted of two kinds of trials: learning and test. Learning trials were 10-20 second video clips in which a speaker first introduced herself to the child, and then produced a short monologue about one of the two toys on the screen, labeling it three times. The script for the first learning trial, for example, was "Hey there, can I show you my friend's toys? This is a *manu*. I really like the *manu*. The *manu* is fun to play with." The exact script varied from trial, but always followed this general format. On *Extended* Cue trials, the speaker picked up the target toy and engaged with it over the course of labeling (Figure 1a). In contrast, on the *Brief* Cue trials, the speaker indicated her target of reference only with a quick glance to the toy when she first produced it's label. She looked straight into the camera for the rest of the trial (Figure 1b). Thus, learning from *Brief* Cue trials required children to rapidly follow her gaze.

Test trials followed the standard Looking While Listening protocol (Fernald et al., 1998). On each test trial, children saw two objects – one on each side of the screen – heard a short audio clip of the speaker from the learning trials asking them to find a target object (Figure 1c). Each test trials was 7.5 seconds long. On *Familiar* test trials, both the target and distractor were common objects familiar to young children (e.g. book vs. dog). On *Novel* test trials, both the target and distractor were novel objects from the previous learning trial.

Finally, the experiment ended with a calibration check: a short video in which small dancing stars appeared in four places on the screen. Because eye-tracker calibration can be imprecise, especially with younger children (Morgante, Zolfaghari, & Johnson, 2011), this check allowed us to adjust initial calibration settings to minimize the discrepancy between the behavior children produced and the behavior we analyzed (for details, see Frank, Vul, & Saxe, 2012).

Design and Procedure

The experiment began with a 4-point calibration and then proceeded into a series of learning/test blocks. In each block, children first watched a learning trial in which a speaker labeled one of two on-screen toys. Following this learning trial, children were given a Looking While Listening test trial in which they saw both of these toys and were asked to find the toy labeled on the previous learning trial (e.g. "Can you find the *manu*?"). Each block consisted to two such learning/test combinations: one for a toy indicated by an *Extended* Cue, and one for a toy indicated by a *Brief* Cue (Figure 1a and b). The same toys and the same label (*manu*) were used for all *Extended* Cue trials, and a different set of toys and a different label (*bosa*) were used for all *Brief* Cue trials. The entire experiment consisted of three such blocks, and two *Familiar* test trials were inserted between each block. Thus, in total, each child participated in three learning and test trials in each Cue condition, and four *Familiar* test trials.

Data Analysis

Children's eye movements during both learning and testing were analyzed using a Regions of Interest (ROI) approach. On learning trials, bounding-box ROIs were drawn by a human coder frame-by-frame for the speaker's face and for the two objects. On test trials, a bound-box ROI was drawn for each of the two static images. To ensure that recorded eye movements were mapped to the correct ROIs, children's calibrations were first adjusted by fitting a robust linear regression for their fixations during the calibration check video and using this model to transform eye movements during the rest of the experiment (Frank et al., 2012).

Children's learning and test behaviors were quantified by measuring their proportion of looking to each ROI on each trial. To ensure that proportions were representative, individual test trials were excluded from analysis if eye gaze data was missing for more than half of their duration. To compute age-group looking proportions, proportions were computed first for each individual trial, averaged at the individual-child level, and then averaged across children.

Window-of-analysis selection began by coding the point of disambiguation for each trial. This was the onset of the target label for test trials, and the rotation of the speaker's head for learning trials (marked '0' in the graphs in the Results section). The window for each trial began 500ms after this point of disambiguation to allow children of all ages enough time to process. The window ended at the end of test trials, and the point at which the label was heard for a second time on learning trials: 2.5 seconds after the point of disambiguation.

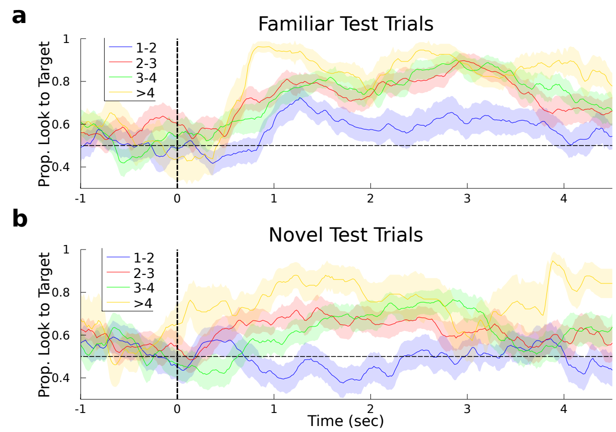

Figure 2: Children's probabilities of fixating the correct target of each label over the course of each test trial. The point labeled 0 indicates the onset of the label, and different colors indicate different age groups. Each line indicates the mean proportion of looking for one age group, and shaded areas repent ±1SE. A proportion of .5 indicates chance performance.

Results and Discussion

Children's patterns of fixation provide a continuous record of their moment-by-moment visual attention over the course of both learning and test trials. We first present an analysis of word learning and familiar word recognition over development. We then connect test behavior to children's patterns of looking during learning. Figure 2 shows gaze trajectories over the course of both Familiar and Novel test trials for each age group. To quantify children's learning with standard analyses, we aggregated these patterns of looking over time to compute the aggregate proportion of looking at the target object on each test trial.

Test Trials

Overall, children in each age range showed evidence of recognizing familiar words – looking at the correct target on Familiar trials for a greater proportion of time than expected by chance ($M_{1-2} = .60$, $t(16) = 2.27$, $p < .05$; $M_{2-3} = .76$, $t(22) = 10.31$, $p < .001$; $M_{3-4} = .78$, $t(20) = 7.95$, $p < .001$; $M_{4+} = .82$, $t(9) = 8.46$, $p < .001$). A linear model showed that familiar word recognition improved significantly across development ($\beta_{age} = .07$, $t(67) = 4.12$, $p < .001$; $r = .46$).

When tested for their knowledge of the word from *Extended* Cue trials, children in the youngest age group did not show evidence of learning ($M_{1-2} = .48$, $t(17) = -.41$, $p = .68$), but children in the older age groups did ($M_{2-3} = .64$, $t(20) = 3.14$, $p < .01$; $M_{3-4} = .71$, $t(20) = 5.22$, $p < .001$; $M_{4+} = .75$, $t(8) = 8.46$, $p < .001$). Learning from *Extended* Cue trials also improved significantly across development ($\beta_{age} = .09$, $t(69) = 4.28$, $p < .001$; $r = .45$).

Finally, when tested for knowledge of the word from *Brief* Cue trials, 1-2 year olds did not show evidence of learning ($M_{1-2} = .48$, $t(17) = -.44$, $p = .67$), 2-3 year olds showed marginal evidence of learning ($M_{2-3} = .59$, $t(21) = 1.96$, $p = .06$), and the older two age groups showed significant evidence of learning; $M_{3-4} = .60$, $t(20) = 2.41$, $p < .05$; $M_{4+} = .74$, $t(9) = 6.57$, $p < .001$). As with the other conditions, learning from *Brief* Cue trials improved significantly across development ($\beta_{age} = .08$, $t(69) = 3.64$, $p < .01$; $r = .40$).

Together, these results provide clear evidence of a developmental trajectory in both word learning and word recognition (Figure 3). Word recognition and learning from the *Extended* Cue improved particularly rapidly over early development, consonant with previous work examining the link between speech recognition and early word learning (Fernald et al., 2006; Fernald & Marchman, 2012). Because the speaker in the *Extended* Cue condition continued to provide a social cue-to-reference over the course of labeling, the primary hurdle to learning in this condition was speech processing rather than referential ambiguity. The *Brief* Cue condition, however, presented an additional challenge: children needed to rapidly follow the speaker's social gaze to determine her target of reference. In this condition, the biggest jump in performance came much later in development. While 2-3 year olds showed marginal evidence of learning from the *Brief* Cue, and learning in 3-4 year olds was statistically significant, only the oldest children showed robust evidence of learning. Because children in the middle age groups are well into the stage of development at which they attend to and learn from social cues, it is likely that *Brief* Cue condition was difficult for them precisely because it required rapid processing of referential information (Baldwin, 1991; Hollich et al., 2000).

These results suggest that a critical bottleneck in early word learning may be attention in-the-moment: children need to process speech and social information quickly enough to determine the label and target of reference.

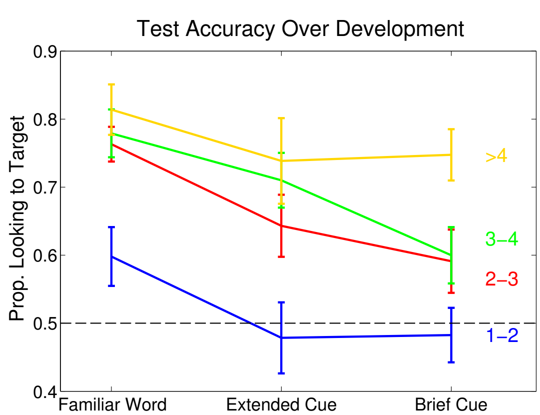

Figure 3: Children improved in their abilities to recognize familiar words, and to learn from both the *Extended* and *Brief* Cues over the course of development. Individual lines indicate different age groups and error bars indicate ±1SE.

Table 1: Predicting Learning of Novel Words.

Predictor	Estimate (SE)	*t* Value	Significance
Intercept	.34 (.09)	4.04	$p < .001$
Age (yrs)	.06 (.02)	3.18	$p < .001$
Brief Cue	.04 (.04)	1.00	$p = .34$
Familiar Test	.25 (.12)	2.09	$p < .05$
Face Prop.	-.23 (.08)	-2.91	$p < .01$

Connecting Learning and Test

In addition to recording children's patterns of looking on test trials, we also captured their looking behavior during learning. This allowed us to chart the developmental trajectories of looking at the caregiver's face and at the target of reference. Figure 4 shows the time course of looking for each age group in both Cue conditions around the point of disambiguation.

In the *Extended* cue condition, looking patterns were qualitatively similar across development. At all ages, children oriented to the speaker's face as she began speaking, and then switched their attention to her target of reference between 500ms and 1.5s after she produced the label. They continued to look predominantly at this target object for the next several seconds. There were apparent quantitative differences – for instance the youngest children were slowest to disengage form the face, but the *Extended* cue scaffolded children at all ages into finding the target of reference and sustained their attention on it.

In contrast, looking patterns in the *Brief* cue condition changed qualitatively across development. Children in the youngest two age groups generally maintained fixation on the speaker's face long after the point of disambiguation, and were relatively unlikely to attend to the target referent. Thus, they were not able to process the speaker's social gaze quickly enough to use it for disambiguation. In contrast, the 3-4 year olds, and especially children over the age of four, showed evidence of disengaging from the face and following the speaker's gaze to find her intended referent. These data provide evidence of children's developing abilities to track and use social information in real-time at a rapid rate.

To determine whether these developing abilities to process speech and social cues contribute to word learning, we fit a linear mixed-effects model to the data (Baayen, Davidson, & Bates, 2008). This model used children's age, their accuracy on Familiar test trials, and their looking during learning trials to predict their test accuracy for both *Extended* and *Brief* cue trials. Table 1 shows coefficient estimates and their significance for each of these predictors. While Cue type was not a significant predictor, age and Familiar test accuracy were both significant positive predictors of test accuracy, and looking to the speaker's face was a significant negative predictor. No interaction terms approached significance. [1]

[1] When looking to the target was included instead of looking to the face, this term was a significant positive predictor ($\beta = .18$, $t =$

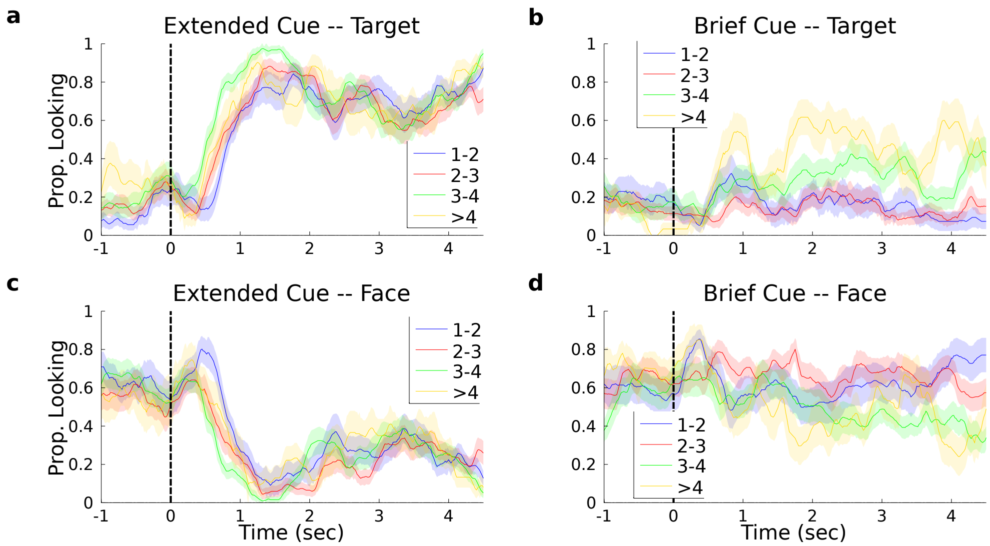

Figure 4: Children's looking patterns during learning for both *Extended* Cue and *Brief* Cue trials. The top row shows looking to the target referent and the bottom row shows looking to the speaker's face. Dotted lines at '0' indicate the point of disambiguation. Looking patterns were qualitatively similar in the *Extended* Cue condition across development, but diverged significantly in the *Brief* Cue condition. Only the oldest two groups of children were able to rapidly follow the speaker's social cue.

Thus, children who were fast at picking the label out of the speaker's utterance, and fast to follow her social cue in both Cue conditions were the most likely to learn the mapping between the word and its target referent. Because age was also a significant predictor, even after accounting for speech processing and cue-following, there must be additional changes in cognitive processing across development that moderate to the connection between real-time attention and ultimate word learning (e.g. working memory). Nonetheless, these data provide strong evidence that children's abilities to process both speech and social signals change over the course of the first four years, and that changes in these skills are important contributors to word learning.

Conclusion

Although children may learn words by aggregating information across a number of naming events (Pinker, 1984; Smith & Yu, 2008), their success must ultimately be constructed from the information they acquire in each individual event. Because both speech and social cues to reference are rapid, serial channels, getting the most out of each naming event requires processing words and identifying social referents quickly and accurately. Our data suggest that the ability to do both of these things develops significantly over the course of childhood, and that both of these abilities are related to the ability to learn novel labels for novel objects.

While a large body of work has established the relationship between children's language processing speed and their later language outcomes (Fernald et al., 1998, 2006), our study adds to this literature by suggesting that processing speed is important in social understanding as well. Much of the early social input that children receive from their caregivers is the social equivalent of child-directed speech: slow, clear, and focused on accessible referents. But as children develop and begin to interact with others, they may encounter an increasing proportion of situations in which they need to track a fleeting glance or a subtle reference. Being able to apprehend these brief social signals may play an important role in allowing children to learn across a range of environments.

More generally, becoming a better word learner is about getting more information out of less input. Many developments that are linked to better word learning – the emergence of mutual exclusivity, the shape bias, and increased speed in language processing (Yurovsky, Bion, Smith, & Fernald, 2012; Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002; Fernald & Hurtado, 2006) – have their effects because they allow children to glean information about word meanings from their environment more effectively. The work in this paper suggests that children's developing understanding of the social environment may have a similar role in early word learning.

Acknowledgments

The authors are grateful to Middy Casillas and to all of the members of the Language and Cognition Lab for feedback on the project and manuscript. This work was supported by a John Merck Scholars Fellowship to MCF.

1.996, $p < .05$). However, this model gave a slightly poor fit to the data, so we report the version including looking to the face.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, *62*, 875–890.

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

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for 'motionese': Modifications in mothers' infant-directed action. *Developmental Science*, *5*, 72–82.

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*, B33–44.

Brooks, R., & Meltzoff, A. N. (2006). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, *35*, 207–220.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, *15*, 17–29.

Chater, N., & Christiansen, M. H. (2010). Language acquisition meets language evolution. *Cognitive Science*, *34*, 1131–1157.

Colombo, J. (2001). The development of visual attention in infancy. *Annual Review of Psychology*, *52*, 337–367.

Fernald, A., & Hurtado, N. (2006). Names in frames: Infants interpret words in sentence frames faster than words in isolation. *Developmental Science*, *9*, F33–F40.

Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development*, *83*, 203–22.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*, 98–116.

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, *9*, 228-231.

Frank, M. C., Tenenbaum, J. B., & Fernald, A. (2013). Social and discourse contributions to the determination of reference in cross-situational word learning. *Language, Learning, and Development*, *9*, 1–24.

Frank, M. C., Vul, E., & Saxe, R. (2012). Measuring the development of social attention using free-viewing. *Infancy*, *17*, 355–375.

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, *71*, 878–94.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*, 1368–1378.

Hollich, G. J., Hirsh-Pasek, K., & Golinkoff, R. M. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society of Research in Child Development*, *65*.

Kail, R. (1991). Processing time declines exponentially during childhood and adolescence. *Developmental Psychology*, *27*, 259–266.

McMurray, B. A., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*, 831–877.

Morgante, J. D., Zolfaghari, R., & Johnson, S. P. (2011). A critical test of temporal and spatial accuracy of the Tobii T60XL eye tracker. *Infancy*, *17*, 9–32.

Onnis, L., Waterfall, H. R., & Edelman, S. (2008). Learn locally, act globally: Learning language from variation set cues. *Cognition*, *109*, 423–430.

Pinker, S. (1984). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA.: MIT Press.

Samuelson, L. K., Smith, L. B., Perry, L. K., & Spencer, J. P. (2011). Grounding Word Learning in Space. *PLoS ONE*, *6*, e28095.

Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature*, *253*, 265–266.

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*, 39–91.

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science*, *13*, 13–9.

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

Spencer, J. P., Perone, S., Smith, L. B., & Samuelson, L. K. (2011). Learning words in space and time: Probing the mechanisms behind the suspicious-coincidence effect. *Psychological Science*, *22*, 1049–1057.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *57*, 1454–63.

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*, 244–262.

Yurovsky, D., Bion, R. A. H., Smith, L. B., & Fernald, A. (2012). Mutual exclusivity and vocabulary structure. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), (pp. 1197–1202). Austin, TX: Cognitive Science Society.

Yurovsky, D., Smith, L. B., & Yu, C. (in press). Statistical word learning at scale: The baby's view is better. *Developmental Science*.

Yurovsky, D., Yu, C., & Smith, L. B. (2012). Statistical speech segmentation and word learning in parallel: Scaffolding from child-directed speech. *Frontiers in Psychology*, *3*, 374.