Social contexts shape information gathering, attention, and memory during familiar language comprehension and novel word learning

Kyle MacDonald

Dissertation overview

Learning a first language should be hard. Consider that even concrete nouns are often used in complex contexts with multiple possible referents, which in turn have many conceptually natural properties that a speaker could talk about. This ambiguity creates the potential for an (in principle) unlimited amount of referential uncertainty in the learning task.[1]. Moreover, finding meaning in language requires rapdily establishing reference during real-time interaction where the incoming information is dynamic, multimodal, and transient. Remarkably, word learning proceeds despite this uncertainty, with estimates of adult vocabularies ranging between 50,000 to 100,000 distinct lexical concepts (P. Bloom 2002). How do learners infer and retain such a large variety of word meanings from data with this kind of ambiguity?

Social-pragmatic theories of language acquisition emphasize the importance of the social context for word learning (P. Bloom 2002; Clark 2009; Hollich et al. 2000). Experimental work has shown that even children as young as 16 months prefer to map novel words to objects that are the target of a speaker's gaze and not their own (Baldwin 1993). In an analysis of naturalistic parent-child labeling events, Yu and Smith (2012) found that young learners tended to retain labels that were accompanied by clear referential cues, which served to make a single object dominant in the visual field. And correlational studies have demonstrated strong links between early intention-reading skills (e.g., gaze following) and later vocabulary growth (Brooks and Meltzoff 2005; Carpenter et al. 1998). Moreover, studies outside the domain of language acquisition have shown that the presence of social cues: (a) produce better spatial learning of audiovisual events (Wu et al. 2011), (b) boost recognition of a cued object (Cleveland, Schug, and Striano 2007), and (c) lead to preferential encoding of an object's featural information (Yoon, Johnson, and Csibra 2008).

Together, the evidence suggests that social information shapes the basic information gathering processes that support word learning. My dissertation work directly tests this information-seeking proposal. To do this, I decompose the concrete word learning task into three parts (see M. Frank, Lewis, and MacDonald (2016)) and ask how information from a social partner affects each sub-component:

1. language comprehension
2. social cue following

---

[1]This problem is a simplified version of Quine's *indeterminacy of reference* (Quine 1960): That there are many possible meanings for a word ("Gavigai") that include the referent ("Rabbit") in their extension, e.g., "white," "rabbit," "dinner." Quine's broader philosophical point was that different meanings ("rabbit" and "undetached rabbit parts") could actually be extensionally identical and thus impossible to tease apart.

3. linking a word to a concept

In our previous work, we characterized how children and adults choose to allocate visual attention between a social partner and objects during American Sign Language (ASL) comprehension (Chapter 1). We then compared the dynamics of eye movements during ASL processing to those of spoken language learners, shwoing that ASL-learners gather more information about the linguistic signal before shifting away from a language source compared to spoken language learners. We proposed an information-seeking account to explain these modality differences and tested predictions of our account across a variety of language comprehension contexts. We found the same pattern of behavior for English-speaking adults processing displays of printed text and for children and adults processing degraded speech (Chapter 2). These results suggest that listeners flexibly adapt to the value of seeking higher value visual information with respect to the goal of rapid langauge comprehension.
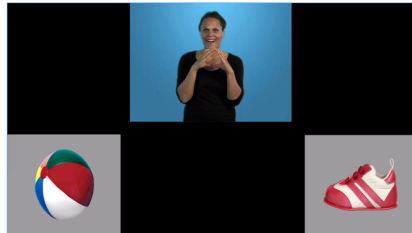
In a separate line of work, we asked whether the presence of a social cue – a speaker's gaze – could change the representations underlying novel word learning (Chapter 3). Our results suggest that word learneers stored representations with different levels of fidelity depending on the amount of ambiguity present during learning. In the absence of a referential cue to word meaning, learners tended to store more alternative word-object links. In contrast, when gaze was present learners stored less information, showing behavior consistent with tracking a single hypothesis. Thus, word learners flexibly respond to the amount of ambiguity in the input, and as referential uncertainty increases, they tend to store more information.

Here, I propose a study that will connect our prior work on information seeking eye movements during familiar language comprehension with our work on the effects of social cues on cross-situational word learning. This goal of this study is to test predictions of our information seeking account in the novel word learning context. Testing these preditions will increase our understanding of how listeners flexibly adapt the dynamics of their eye movements to seek higher value information to support word learning. Overall, these results will synthesize ideas from research on active learning, social-pragmatic theories of language acqusition, and work on goal-based vision to increase our knowledge of how language processing and acquisition unfolds across a wider variety of populations and learning contexts.
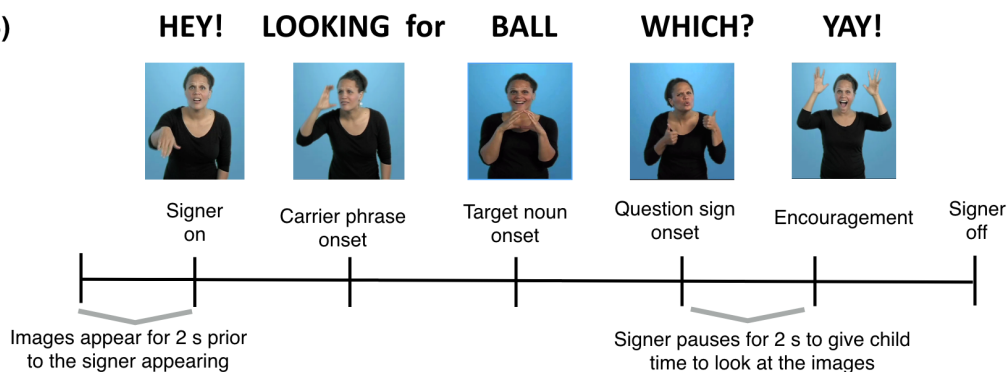
Completed work

Chapter 1: Eye movements during real-time American Sign Language comprehension

A)

B)

HEY!	LOOKING for	BALL	WHICH?	YAY!

| Signer on | Carrier phrase onset | Target noun onset | Question sign onset | Encouragement | Signer off |

Images appear for 2 s prior to the signer appearing

Signer pauses for 2 s to give child time to look at the images

When children interpret spoken language in real time, linguistic information drives rapid shifts in visual attention to objects in the visual world, which can provide insights into the development of efficiency in lexical access. But how does language influence visual attention when the linguistic signal and the visual world are both processed via the visual channel? We developed precise measures of eye movements during real-time comprehension of a visual-manual language, American Sign Language (ASL), by 29 native, monolingual ASL-learning children (16-53 mos, 16 deaf, 13 hearing) and 16 fluent deaf adult signers. All signers showed evidence of rapid, incremental language comprehension, initiating eye movements prior to sign offset. Deaf and hearing ASL-learners showed remarkably similar gaze patterns, suggesting that the in-the-moment dynamics of eye movements during ASL processing are shaped by the constraints of processing a visual language in real time and not by differential access to auditory information in day-to-day life. Finally, variation in children's ASL processing was positively correlated with age and vocabulary size. Thus, despite channel competition, allocation of visual attention during ASL comprehension reflects information processing skills that are fundamental for language acquisition regardless of language modality.
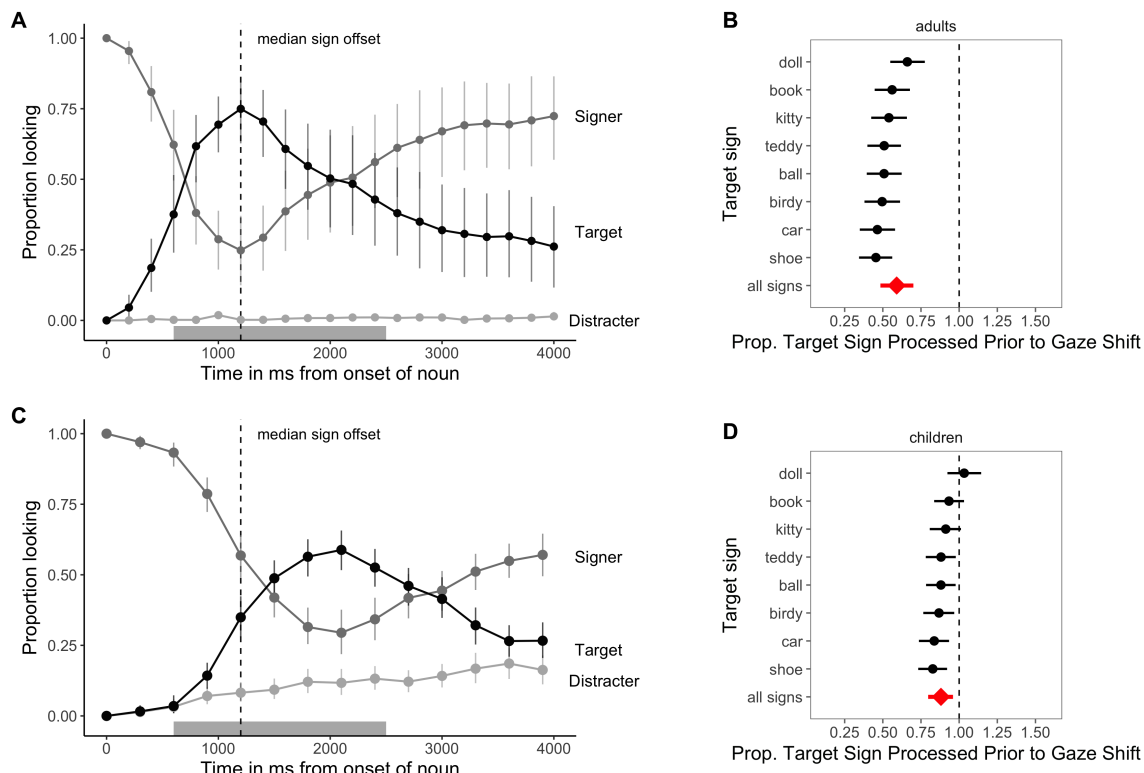
## Chapter 2: An information seeking account of eye movements in spoken and signed language comprehension

The study of eye movements during language comprehension has provided fundamental insights into the interaction between conceptual representations of the world and the incoming linguistic signal. For example, research shows that adults and children will rapidly shift visual attention upon hearing the name of an object in the visual scene, with a high proportion of shifts occurring prior to the offset of the word (Allopenna, Magnuson, and Tanenhaus 1998; Tanenhaus et al. 1995). Moreover, researchers have found that conceptual representations activated by fixations to the visual world can modulate subsequent eye movements during language processing (Altmann and Kamide 2007).

The majority of this work has used eye movements as a measure of the output of the underlying language comprehension process, often using linguistic stimuli that come from a disembodied voice. But in real world contexts, people also gather information about the linguistic signal by fixating on the language source. Consider a speaker who asks you to "Pass the salt" but you are in a noisy room, making it difficult to understand the request. Here, comprehension can be facilitated by gathering information via (a) fixations to the nonlinguistic visual world (i.e., encoding the objects that are present in the scene) or (b) fixations to the speaker (i.e., reading lips or perhaps the direction of gaze). But, this situation creates a tradeoff where the listener must decide what kind of information to gather and at what time. How do we decide where to look? We propose that people modulate their eye movements during language comprehension in response to tradeoffs in the value of gathering different kinds of information.

We test this adaptive tradeoff account using two case studies that manipulate the value of different fixation locations for language understanding: a) a comparison of processing sign vs. spoken language in children (E1), and b) a comparison of processing printed text vs. spoken language in adults (E2). Our key prediction is that competition for visual attention will make gaze shifts away from the language source less valuable than fixating the source of the linguistic signal, leading people to generate fewer exploratory, nonlanguage-driven eye movements.
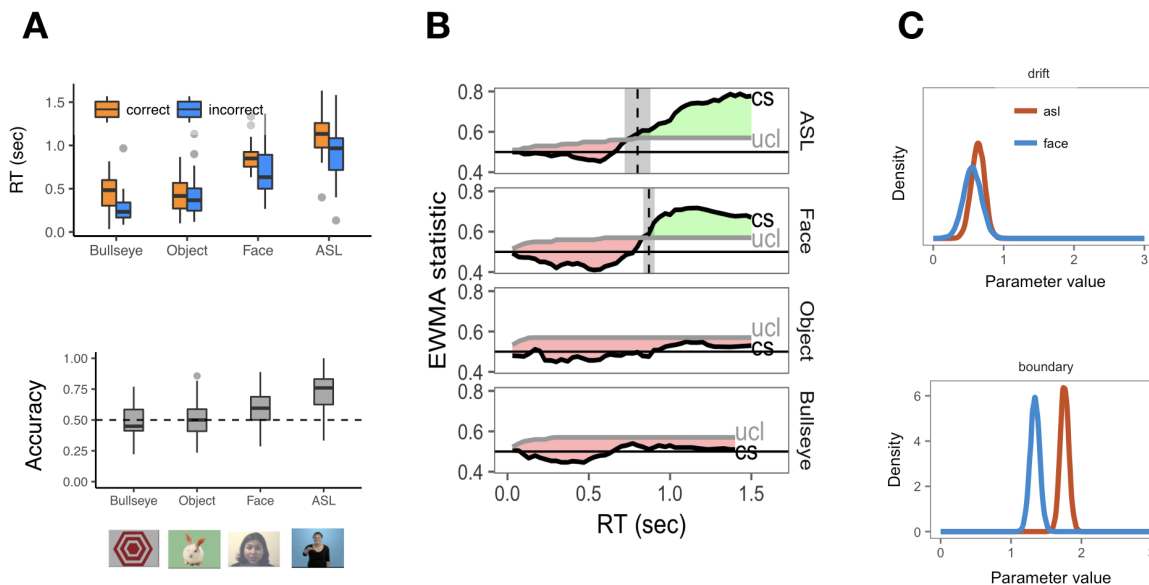

Some of our work has explored how the presence of another person changes the set of information seeking behaviors available (MacDonald et al. 2017). Inspired by theories of natural vision that characterize eye movements as an information seeking mechanism, we asked whether children and adults would allocate more visual attention to a speaker when the linguistic signal was noisy to support the goal of rapid language understanding. We used an eye-tracking task to measure participants' gaze patterns while they processed clear or degraded speech (speech with brown noise added). Both children and adults spent more time fixating on the speaker in the degraded speech context. Interestingly, children and adults were also more accurate in word recognition even though the speech was noisy and difficult to process. This result suggests that listeners were compensating for the uncertainty in the auditory channel by gathering visual information from the speaker. Critically, listeners would not have been able to gather this information if the speaker was not present (e.g., listening to a noisy recording) and in clear view.
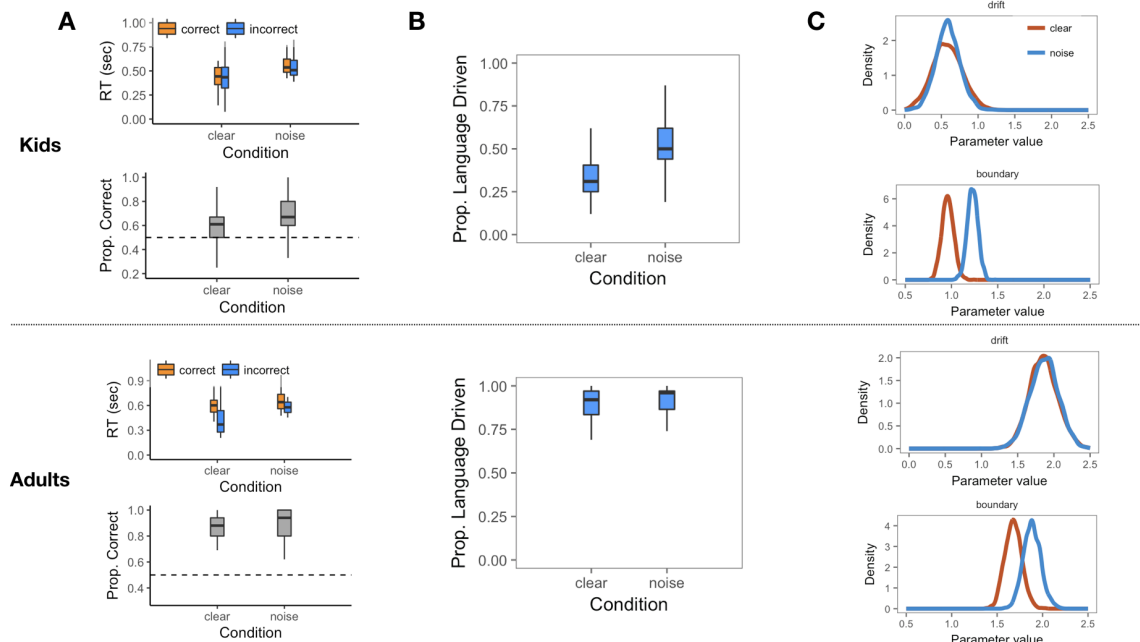
Chapter 3: Social information modulates attention and memory during word learning

Some of our own work provides evidence that the social context can modulate the content of the learner's hypothesis space (MacDonald, Yurovsky, and Frank 2017). Inspired by ideas from Social-pragmatic theories of language acquisition that emphasize the importance of social cues for word learning (Clark 2009; Hollich et al. 2000; P. Bloom 2002), we showed adults a series of word learning contexts that varied in ambiguity depending on whether there was a useful social cue to reference available (a speaker's gaze). We then measured learners' memory for alternative word-object links. People flexibly responded to the amount of ambiguity in the input, and as uncertainty increased, they tended to store more word-object hypotheses. Moreover, we found that learners stored representations with different levels of fidelity as a function of the reliability of the social cue. When the speaker was a less reliable source of information, learners distributed attention and memory broadly, storing more hypotheses.

These results provide evidence that the content of learners' hypothesis spaces changed as a function of social information. Further suppport for this idea comes from experimental work showing that even children as young as 16 months prefer to map novel words to objects that are the target of a speaker's gaze and not their own (Baldwin 1993), and analyses of naturalistic parent-child labeling events shows that young learners tended to retain labels accompanied by clear referential cues, which served to make a single object dominant in the visual field (Yu and Smith 2012). One important direction for future research is to measure the full causal pathway from variation in social information through children's hypothesis spaces to their information seeking behaviors. For example, it would be interesting to know whether learners' subsequent questions or decisions about where to allocate attention would be affected by the social context in which they were first exposed to a new word.
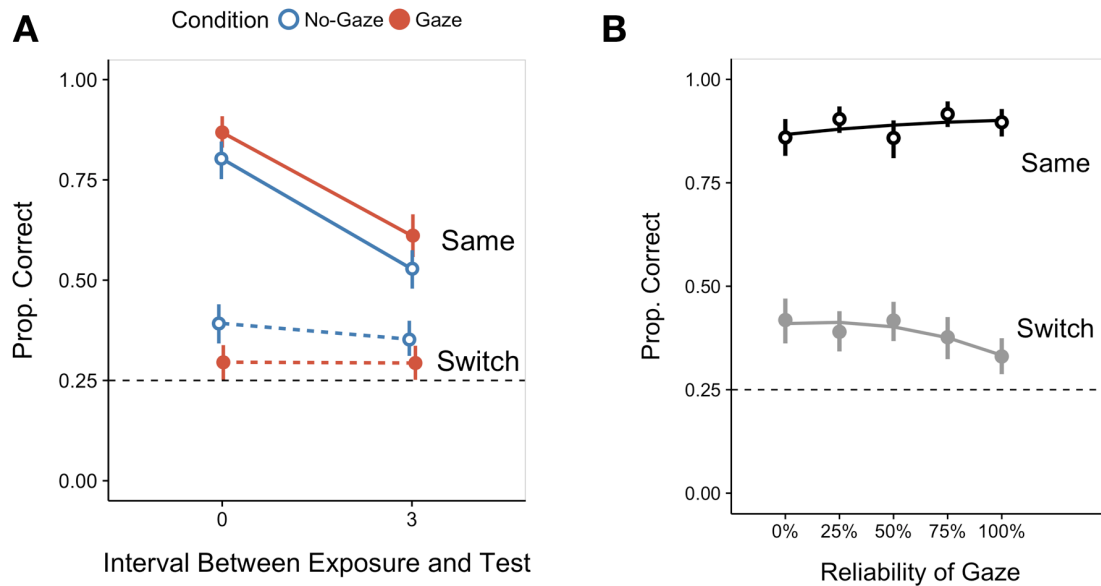
Proposed work

Eye movements for information seeking during novel word learning

The goal is to test how our information seeking account generalizes to contexts where there is uncertainty over word-object links. The word learning context is interesting because people are trying to make progress towards multiple goals: comprehending speech in the moment (a function of listening and looking to the speaker), mapping the new word to an object, and learning that mapping over time. Eye movements during concept formation can gather information about the speaker or about the nonlinguistic visual world. How do we explain where children look as they acquire more information in-the-moment of language comprehension and as they build a learning history about the correct word-object mapping? This question can be formalized as a sequential decision making problem where children make fixation choice based on (1) their knowledge of the target concept, (2) the value of fixating a speaker for linguistic processing, and (3) the cost of each eye movement.

The intuitive idea is that people balance fixating a speaker and fixating objects to support concept learning. The question is whether models of Bayesian concept learning and Optimal Experiment Design (Neslon & Cottrell, 2007) provide a good explanation of children's eye movements. How far can we get using a purely computational information seeking decision model?

Pilot

When gaze cued adults' visual attention, they showed stronger memory for the word-object link compared to when a gaze cue was absent (Figure 2). social information does more than modulate how people allocate their visual attention; instead, social contexts change strength of inferences
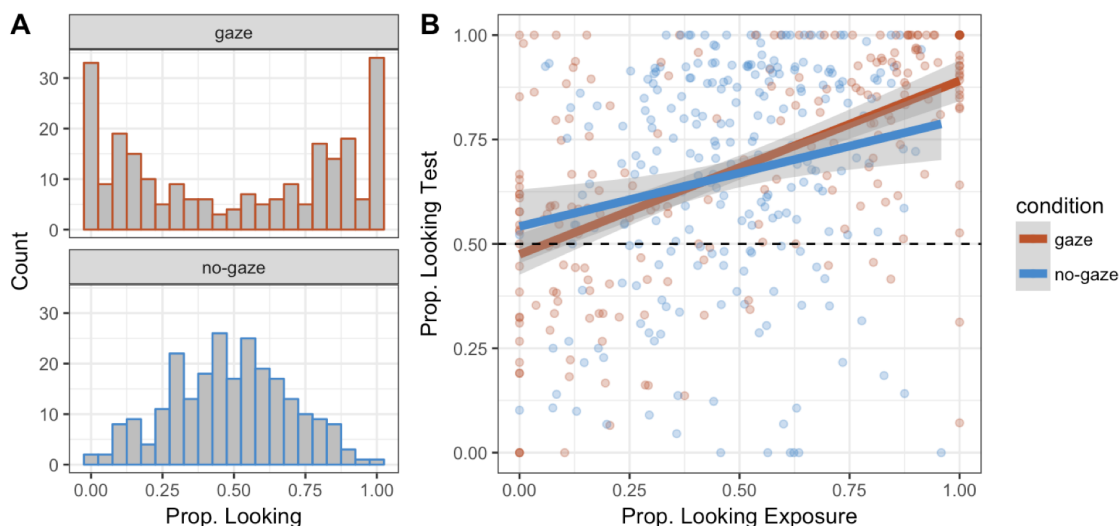
Design

Predictions

The prediction is that the dynamics of eye movement will shift over the course of learning. In the beginning of the task, learners will distribute fixations to prioritize gathering information about the objects or about disambiguating reference (e.g., gathering a gaze cue). After learning the word-object links, people will shift and start to distribute more fixations to the speaker to gather visual information that supports comprehension of the speech, showing the behavioral signatures measured in the familiar language comprehension task.

References

Allopenna, Paul D, James S Magnuson, and Michael K Tanenhaus. 1998. "Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models." *Journal of Memory and Language* 38 (4). Elsevier: 419–39.

Altmann, Gerry, and Yuki Kamide. 2007. "The Real-Time Mediation of Visual Attention by Language and World Knowledge: Linking Anticipatory (and Other) Eye Movements to Linguistic Processing." *Journal of Memory and Language* 57 (4). Elsevier: 502–18.

Baldwin, Dare A. 1993. "Infants' Ability to Consult the Speaker for Clues to Word Reference." *Journal of Child Language* 20 (02). Cambridge Univ Press: 395–418.

Bloom, Paul. 2002. *How Children Learn the Meaning of Words*. The MIT Press.

Brooks, Rechele, and Andrew N Meltzoff. 2005. "The Development of Gaze Following and Its Relation to Language." *Developmental Science* 8 (6). Wiley Online Library: 535–43.

Carpenter, Malinda, Katherine Nagell, Michael Tomasello, George Butterworth, and Chris Moore. 1998. "Social Cognition, Joint Attention, and Communicative Competence from 9 to 15 Months of Age." *Monographs of the Society for Research in Child Development*. JSTOR, i–174.

Clark, Eve V. 2009. *First Language Acquisition*. Cambridge University Press.

Cleveland, Allison, Mariah Schug, and Tricia Striano. 2007. "Joint Attention and Object Learning in 5-and 7-Month-Old Infants." *Infant and Child Development* 16 (3). Wiley Online Library: 295–306.

Frank, M, M Lewis, and Kyle MacDonald. 2016. "A Performance Model for Early Word Learning." In *Proceedings of the 38th Annual Conference of the Cognitive Science Society*.

Hollich, George J, Kathy Hirsh-Pasek, Roberta Michnick Golinkoff, Rebecca J Brand, Ellie Brown, He Len Chung, Elizabeth Hennon, Camille Rocroi, and Lois Bloom. 2000. "Breaking the Language Barrier: An Emergentist Coalition Model for the Origins of Word Learning." *Monographs of the Society for Research in Child Development*. JSTOR, i–135.

MacDonald, Kyle, Aviva Blonder, Virgnia and Marchman, Anne Fernald, and Michael C Frank. 2017. "An Information-Seeking Account of Eye Movements During Spoken and Signed Language Comprehension." In *Proceedings of the 39th Annual Conference of the Cognitive Science Society*.

MacDonald, Kyle, Daniel Yurovsky, and Michael C Frank. 2017. "Social Cues Modulate the Representations Underlying Cross-Situational Learning." *Cognitive Psychology* 94. Elsevier: 67–84.

Quine, Willard V. 1960. "0. Word and Object." *111e MIT Press*.

Tanenhaus, Michael K, Michael J Spivey-Knowlton, Kathleen M Eberhard, and Julie C Sedivy. 1995. "Integration of Visual and Linguistic Information in Spoken Language Comprehension." *Science* 268

(5217). The American Association for the Advancement of Science: 1632.

Wu, Rachel, Alison Gopnik, Daniel C Richardson, and Natasha Z Kirkham. 2011. "Infants Learn About Objects from Statistics and People." *Developmental Psychology* 47 (5). American Psychological Association: 1220.

Yoon, Jennifer MD, Mark H Johnson, and Gergely Csibra. 2008. "Communication-Induced Memory Biases in Preverbal Infants." *Proceedings of the National Academy of Sciences* 105 (36). National Acad Sciences: 13690–5.

Yu, Chen, and Linda B Smith. 2012. "Embodied Attention and Word Learning by Toddlers." *Cognition.* Elsevier.