

**From the holdings of the Department of Special Collections , Princeton University Library,
Princeton, NJ**

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyright law. We do not charge any permission or usage fees for the publication of images of material in our collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff. For more information regarding copyright, credit, and citation please review the information on the Copyright, Credit and Citation Guidelines page of our website.

<https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines>. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. <https://library.princeton.edu/special-collections/ask-us>

Doctoral Dissertations, 2011-2026

Psychology

2024

A multi-method investigation into learning words from social cues during discourse

Lee, Crystal Yujin

Princeton, NJ : Princeton University

<http://arks.princeton.edu/ark:/88435/dsp01gh93h287r>

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. <https://www.copyright.gov/title17/title17.pdf>

A MULTI-METHOD INVESTIGATION INTO LEARNING WORDS FROM SOCIAL CUES
DURING DISCOURSE

Volume I

Crystal Yujin Lee

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE BY

THE DEPARTMENT OF PSYCHOLOGY

Adviser: Casey Lew-Williams

May 2024

© Copyright by Crystal Yujin Lee, 2024. All rights reserved.

Abstract

Children learn words in a social environment. In my dissertation, I examine how caregivers' social cues facilitate young children's word learning in settings that mirror their typical, dynamic learning environments. In Chapter 1, I overview prior work examining how social cues may support word learning, focusing on possible mechanisms underlying their benefits. In Chapter 2, I investigate how social cues support children's word learning over the time course of discourse topics, as opposed to the focus on single labeling events in prior studies. In Chapter 3, I expand these findings to understand how caregivers use social cues during natural discourse at home. Finally, in Chapter 4, I explore whether neural synchrony (the alignment of neural activity) within caregiver-child interactions has the potential to explain relations between caregivers' social cues and children's word learning. Together, my findings suggest that, despite the information-density and messiness of natural learning environments, social cues reduce referential ambiguity during natural caregiver-child discourse in a way that ultimately supports children's word learning.

Acknowledgments

This dissertation would not have been possible without the support of so many kind and generous people. First and foremost, I would like to express my gratitude to my advisor, Casey Lew-Williams. Casey embodies what I believe a mentor should be like: a rigorous scholar with a deep sense of empathy for his students. He has always given me the space to think and speak freely, and provided guidance when I need it. Casey, I'm so grateful you accepted me into your lab, and I'm so glad that I chose to join. Thank you for working with me, and teaching me to be a better researcher.

I would also like to thank my secondary advisor, Adele Goldberg. She has an incredible ability to identify interesting research questions, and create a community of amazing language researchers. She and her lab filled the language gap that I was missing at Princeton. Adele, I'm so appreciative that I had the opportunity to work with you. Thank you so much for your guidance and mentorship.

There are a number of other research mentors to whom I would like to express my gratitude. My undergraduate mentors, Chigusa Kurumada and T. Florian Jaeger, introduced me to the world of psychology and language research, and I would not be here without them. Caroline Rowland allowed me to join her department and spend a year at the MPI while working on my first large research project. I really believe that this extra year was invaluable for preparing for graduate school.

I'd also like to thank my collaborators and undergraduate research assistants in the Baby Lab who contributed to this dissertation – Mira Nenchewa, Jessica Kosie, Richard Yang, Eleanor Myers, Annie Jain, and Imuetiyan Eweka – and the families who participated in this research. This would not have been possible without their help and participation.

And finally, I would like to thank my community at Princeton. To the former and current members of the Baby Lab and Language Group – thank you for being such a warm and welcoming group of people. Your feedback was always so helpful, and I’m so glad I got a chance to know you. And to my cohort, COVID-19 messed up a lot of stuff, but I’m still so happy we met, started grad school together, and became friends.

To my 엄마 and 언니, my mom and sister, my Lee ladies, thank you for sneaking into conferences to see my presentations, and being my biggest cheerleaders.

I owe my biggest thanks to my partner, David Weissman. You were the best part of my grad school experience. I’m so excited to see what our future holds.

Table of Contents

Abstract.....	3
Acknowledgments.....	4
Introduction.....	7
Chapter 1. Broadening the scope of the role of social cues during word learning.....	18
Chapter 2. Speech and social cues combine at discourse boundaries to promote word learning..	55
Chapter 3. How social cues are presented across a natural discourse by caregivers.....	81
Chapter 4. Relating social cues, neural synchrony, and word learning in natural discourse.....	117
General Discussion.....	158
References.....	169

Introduction

Children learn language within social contexts. And caregivers support children's emerging language abilities by providing a suite of behavioral social cues during interactions. For example, children learn new words by following others' eye-gaze (Houston-Price, Plunkett & Duffy, 2006; Paulus & Fikkert, 2014) and gestures (Booth, McGregor & Rohlfing, 2008; Rader & Zukow-Goldring, 2012) toward objects of joint attention. In addition, a child's willingness to learn new words from others is influenced by whether they consider their interlocutor to be reliable or unreliable (Krogh-Jespersen & Echols, 2012; Scofield & Behrend, 2008; Sobel, Sedivy, Buchanan, & Hennessy, 2012), and knowledgeable or ignorant (Brooker & Poulin-Dubois, 2013; Mangardich & Sabbagh, 2018; Sabbagh & Baldwin, 2001, 2009). In doing so, young children use cues from their interlocutors' social behaviors to facilitate their own language learning.

However, previous experiments do not directly capture the complexity of learning from social cues in dynamic environments. Prior research on the role of social cues in word learning have typically involved static experiments where participants are taught the names of novel objects either with or without supportive social cues. As such, these experiments typically assess in-the-moment word learning. While in-the-moment associations are an important component of word learning, these experimental methods overlook the complexities of language learning from social cues in naturalistic, dynamic environments (Lee & Lew-Williams, 2022, 2023; Wojcik, Zettersten, & Benitez, 2022). In particular, they do not speak to the timescale or complexity of natural discourse, which usually includes a series of utterances that focus on a particular topic over time (Frank, Tenenbaum, & Fernald, 2013; Rohde & Frank, 2014; Sperber & Wilson,

1986). To fully understand how word learning from social cues manifests in daily life, it is critical to expand the lens under which social cues and its effects on word learning are viewed.

To that end, my dissertation work investigates the role of social cues for learning new words within discourse, and uses a variety of methods to do so, in order to flesh out a clearer picture on how social cues facilitate word learning in dynamic contexts. This dissertation augments the field's understanding of this process by: 1) overviewing how previous research has considered the role of social cues during word learning, 2) providing experimental evidence about how young children use social cues while learning new words within a discourse topic, 3) carrying out a descriptive study examining when caregivers provide social cues during natural discourse, 4) conducting an fNIRS study on brain-to-behavior coupling between children and parents to study how the presentation of social cues enhances communication, and 5) determining whether neural synchrony during caregivers' presentation of social cues facilitates children's novel word learning. Together, my multi-method dissertation evaluates the informativity of social cues during naturalistic word learning environments, and further determines the role of social cues as children learn new words.

Learning words from social cues in experiments

Children demonstrate the ability to track and exploit social cues from a young age. Infants as young as three months of age can use the eye movements of others to find objects (Caron, Caron, Caldwell & Weiss, 1973; D'Entremont et al., 1997; Gredebäck et al., 2010; Maurer & Salapatek, 1976; Vecera & Johnson, 1995). Infants by 6 months of age have been shown to use eye movements to determine referential intent (Baldwin, 1991, 1995; Butler, Caron, & Brooks, 2000; Deligianni, Senju, Gergely, & Csibra, 2011; Senju & Csibra, 2008;

Striano & Reid, 2006). Similar results have been found in experiments using gestures, albeit at a slightly older age. Between the ages of 9-12 months, infants can reliably follow pointing gestures to nearby objects (Bates, 1976). At about 14-15 months of age, children can follow deictic gestures (e.g., pointing) to objects further away or outside of their visual field (Delgado et al., 2002; Desrochers, Morissette, & Ricard, 1995).

The information provided by social cues has also been shown to facilitate word learning. For example, experimental investigations into the functions of social cues have found that the presence (versus absence) of eye-gaze cues promotes novel word learning (Baldwin et al., 1996; Booth, McGregor, & Rohlfing, 2008; Briganti & Cohen, 2011; Houston-Price, Plunkett, & Duffy, 2006; Hirotani et al., 2009; Nurmsoo & Bloom, 2008; Paulus & Fikkert, 2014; Rader & Zukow-Goldring, 2012; Yu & Ballard, 2007; see also Scofield & Behrend, 2011). Baldwin (1991) taught 16-19-month-old infants novel names of objects when a speaker was either: focused on the same object as the infant or looked at an object different from the infant. Their experiment was designed to determine whether infants use nonverbal, social cues from another to infer word-object mappings or if they indiscriminately learned mappings regardless of where a speaker is looking. They found that children both learned the novel label when the participants' and speakers' attention coincided, as well as avoided mapping errors when the object of focus did not coincide with the speaker's view. This work suggests that children can use eye-gaze as a non-verbal cue that is relevant to the formation of novel label inference. Furthermore, the ability to follow eye-gaze has been linked to growth in vocabulary (Akhtar & Gernsbacher, 2007; Brooks & Meltzoff, 2005, 2008; Carpenter et al., 1998; Morales, Mundy & Rojas, 1998; Morales et al., 2000a; Morales et al., 2000b; Tomasello & Farrar, 1986). As one example, Brooks and

Meltzoff (2008) found that 10- and 11-month-old infants who better exploited adult eye-gaze cues to find nearby objects showed significantly greater vocabulary at 24 months of age.

Gestures, too, are a present source of information throughout development (Goldin-Meadow & Singer, 2003; Kosie & Lew-Williams, 2023; Vigliocco et al., 2019). Prior work has found that deictic gestures (e.g., pointing) are the most prevalent in a child's environment, but representational gestures, such as holding up two fingers in a V-shape to represent a rabbit's ears, are also present (Iverson et al., 1999; Özçalışkan & Goldin-Meadow, 2005; Rowe et al., 2008). The information conveyed by caregiver gestures is ultimately incorporated during word learning. For example, observing deictic gestures that disambiguate a referent in space are associated to accurate word learning (Booth, McGregor, & Rohlfing, 2008; Brooks & Meltzoff, 2008; Carpenter et al., 1998; Gogate et al., 2000; Kalagher & Yu, 2006; Harris et al., 1995; Vogt & Kauschke, 2017). A similar result has been found with representational gestures, where observing representational gestures are related to improvements in word learning (Boyce, 2021; Goodwyn et al., 2000; Lüke & Ritterfield, 2014; McGregor, 2008). Thus, experiments examining both eye-gaze and gestures have shown improvements in children's word learning.

Learning words from social cues in the wild

Although social cues have been shown to facilitate word learning, recent work has debated the generalizability of these findings to natural learning scenarios. The reason for this is primarily due to the differences in word learning settings between experimental sessions and typical learning environments. Many of these prior experiments expose young children to only labeling events that teach associations between novel objects and novel labels using social cues.

For example, a quintessential novel word learning experiment presents a naive learner with a novel label such as “*Dax*” or “It’s a *dax*” with another interlocutor’s eye-gaze directing their attention to a target novel object (Bannard & Tomasello, 2012; Booth, McGregor, & Rohlfing, 2008; Briganti & Cohen, 2011; Houston-Price, Plunkett, & Duffy, 2006; Yow et al., 2017). There are two issues with this that prevent certainty in generalizing this result to natural word learning. First, because these experiments track only children’s formation of in-the-moment object-label mappings, this approach is limited in that word learning in the wild not only occurs in the moment, but on longer timescales as well (McMurray, 2016; Piazza, Nenchewa, & Lew-Williams, 2021). Thus, we are limited in what we know of the influence of social cues on word learning across time, such as over a discourse. However, what we do know contrasts with prior experimental work. A particularly emblematic example is a study conducted by Frank, Tenenbaum, and Fernald (2013). Using 24 videos of play sessions between caregiver-child dyads, Frank, Tenenbaum, and Fernald (2013) found that social cues (in this case, eye-gaze and pointing) were not particularly informative nor effective in identifying a referent during naturalistic discourse. While these cues were often present during play, social cues alone were not sufficient in disambiguating the intended referent.

Second, prior experiments have typically paired social cues with labels. However, language input to a child encompasses much more complex linguistic input than just labeling information. For example, Tamis-LeMonda et al. (2017) found that structured play between caregivers and 13-month-olds elicited, on average, 64 word tokens per minute. For object labels specifically, prior studies have found that parents produce a range of 4-9 labels per minute (Abney, Suanda, Smith, & Yu, 2020; Chang, de Barbaro, & Deák, 2016; Pereira, Smith, & Yu, 2014; Yu, Suanda, & Smith, 2019). If we take these averages at face value, then there are upward

of 50 word tokens per minute that are not object labels. This suggests that social cues and label associations are not so straightforwardly presented in natural interactions compared to how they were presented in prior experiments. This raises questions as to whether social cues might facilitate word learning in the face of this linguistic complexity.

Learning words from discourse

Everyday speech to children tends to occur over time within the natural structure of discourse. Discourse refers to neighboring utterances that focus on a particular topic, one after the next (Frank, Tenenbaum, & Fernald, 2013; Rohde & Frank, 2014). Everyday discourse with children is naturally complex, including variations in discourse length (Rohde & Frank, 2014), lexical diversity (Lucariello & Nelson, 1986; Tare, Shatz, & Gilbertson, 2008), syntactic structure (Chang & Deák, 2019; Cameron-Faulkner, Lieven, & Tomasello, 2003), and reference displacement (i.e., “here-and-now” talk, Veneziano, 2001). This suggests that in order to learn new words, children must be able to contend with continuous input that varies in these, and likely other, dimensions.

Previous research has indeed found that young learners can indeed learn new words from input through discourse (Akhtar, Carpenter, & Tomasello, 1996; Horowitz & Frank, 2015; Sullivan & Barner, 2016; Sullivan et al., 2019), and that the structure of discourse shapes children’s word learning (Hoff, 2010; Horowitz & Frank, 2015). Thus, children seem to demonstrate an ability to learn even when learning occurs during more complex scenarios (i.e., through discourse). To adequately examine children’s word learning (and how the social context may support their learning), we must implement experiments that represent the linguistic complexity that they hear on a daily basis in the form of discourse.

Expanding how we examine social cues during word learning

Another important point of limitation in prior research, when examining the effect of social cues during word learning, is that this work has been mostly behavioral either in experiments or through observational research. The most common way to measure word learning from social cues is to present videos or static images to children in an experiment, and then measure word learning through looks or points to a target object (Booth, McGregor, & Rohlfing, 2008; Briganti & Cohen, 2011; Houston-Price, Plunkett, & Duffy, 2006). Such studies have found that presenting social cues during learning positively influences their performance during test phases. But like mentioned previously, there are inherent issues with examining word learning in social contexts in this way. Such experiments reduce the complexity of the stimuli presented by generally presenting only labeling utterances, and eliminating all conversational turn-taking and response-contingencies that are present in typical interactions.

Observational research improves upon these experiments by examining typical, dyadic interactions (e.g., Frank, Tenenbaum, & Fernald, 2013). They allow caregivers and their children to act as they normally would, allowing for a stronger sense of how social cues are presented to children. However, this methodology has suggested an opposite conclusion compared to typical experiments with experimenter manipulations: that social cues are not as informative as previously thought. Thus, implementing and observing only behavioral methods have resulted in contradictory conclusions in the field. Investigating how social cues facilitate word learning using multiple methods is necessary to better understand how word learning is situated within a social context.

One method that has been underutilized in social word learning research is neuroscience methods. Indeed, only one paper has been identified that studied the facilitation of word learning

through presenting social cues, such as joint attention, as measured by EEG (Hirotani et al., 2009). In this dissertation, I will investigate why these cues are important for children's word learning by using functional near-infrared spectroscopy (fNIRS). fNIRS is a brain imaging technique that non-invasively measures cortical hemodynamic responses. This method has become increasingly popular for infant and child research because it is safe, portable, relatively motion-robust, and increases task compliance compared to other neuroimaging techniques (Azhari et al., 2020; Lloyd-Fox et al., 2010).

Within the past decade, the fNIRS technique has been expanded to include “hyper-scanning”, or two-brain fNIRS. This method has been used to study dyadic interactions and shared neural activation across higher-order brain regions, known as “neural synchrony”. Recent work with adults using this method has demonstrated that shared understanding between two people is reflected by enhanced neural synchrony (i.e., Hasson, Ghazanfar, Galantucci, Garrod, & Keysers, 2012; Stephens, Silbert, & Hasson, 2010). This method has also been extended to examine how neural synchrony affects communication between adult-child dyads as well (Piazza, Hasenfratz, Hasson, & Lew-Williams, 2020). Alongside experimental and descriptive methods, fNIRS has the potential to adjudicate the tension between the two aforementioned methods by not only considering behavior, but also measuring shared activation across brains. That is, social cues may enhance learning by enhancing neural synchrony across brains, allowing children to understand the ebbs and flows of conversations and take advantage of the complex input provided by caregivers, resulting in enhanced word learning.

Harnessing the complexities of social cues in naturalistic contexts

What has not yet been investigated is how social cues may promote word learning in contexts that are representative of children's natural environment – specifically in relation to natural presentation of caregivers' social cues accompanied by complex discourse. Research on word learning must expand simplistic representations of a natural learning scenario in order to accurately examine the word learning problem faced by young children on a daily basis. Thus, one goal of this dissertation is to explore how social cues may facilitate word learning in naturalistic and semi-naturalistic contexts.

Furthermore, this research has generally only been conducted using behavioral tasks in an experimental setting, such as using eye-tracking. Conducting experiments that are more parallel to a child's typical learning environment, and using a different array of methods, would enable word learning theories to contend with the full scope of social cues as well as improve the field's understanding of how this nonverbal information promotes learning. The second, methodologically focused goal of this dissertation is to understand what kind of traction can be gained by employing a combination of experimental, observational, and neuroscience methods to investigate word learning in social contexts.

Overview of the dissertation

Chapter 1 reviews existing literature on the role of social cues during children's word learning, theoretical attempts to account for findings in this research area, and outlines the limitations of prior work. Critically, I note areas in which improvements can be made to fully investigate the social influence on word learning. A version of this chapter was published as an article in *Infant and Child Development* (Lee & Lew-Williams, 2023).

Chapter 2 moves beyond the simple experimental designs in prior research by investigating how social cues facilitate word learning at the level of discourse. I find that social cues in combination with labeling events often support children's word learning within discourse topics. A version of this chapter was published as an article in *Cognitive Development* (Lee & Lew-Williams, 2022).

Chapter 3 analyzes a video corpus of natural caregiver-child interactions to broaden the approach of the previous chapter by documenting how caregivers use social cues (such as eye-gaze) during interactions with their children. This is in an effort to determine if caregivers present social cues in a systematic pattern across discourse topics. If so, this would suggest that caregivers present disambiguating social cues across a conversation, and such presentation may facilitate learning.

Chapter 4 tests two hypotheses in two experiments. In Experiment 3, I hypothesize that caregivers present social cues in systematic patterns during natural discourse, resulting in enhanced neural synchrony with their children. I predict that the presence (vs. absence) of social cues from a caregiver, across discourse topics, promotes neural synchrony during interaction with their child.

In Experiment 4, I examine if links between social cues and neural synchrony during caregiver-child discourse predict children's novel word learning. I predict that social cue usage will be related to greater neural synchrony, and that greater neural synchrony between caregivers and children will be positively related to children's word learning.

Overall, this work seeks to broaden investigations into how social information influences the word learning process. By expanding both the types of learning scenarios considered (across time and discourse) and methodologies used (using fNIRS), this dissertation enhances the field's

understanding of how caregiver social cues are utilized by naive learners in learning environments that are more similar to their daily learning scenarios.

Chapter 1. Broadening the scope of the role of social cues during word learning

1.1 Abstract

Children learn words in a social environment, facilitated in part by social cues from caregivers, such as eye-gaze and gesture. A common assumption is that social cues convey either perceptual or social information, depending on the age of the child. In this review of research on word learning and social cues during early childhood, we propose that (1) the functions of social cues are not categorically perceptual or social, and (2) social cues support word learning in four interdependent ways, by helping children to orient attention, extract relevant information, disambiguate referents, and understand others' referential intent. We conclude with specific recommendations for theory-building and suggest that the dynamic and complex functions of social cues need to be accounted for in any complete theory of word learning.

1.2 Introduction

As children learn language, they interact with other people. Their social environment contains a suite of social cues – provided by caregivers, siblings, and other children and adults – that support their emerging language abilities. For example, to learn new words, children are known to follow another’s eye-gaze (Bannard & Tomasello, 2012; Booth, McGregor & Rohlfing, 2008; Gliga et al., 2012; Houston-Price, Plunkett & Duffy, 2006; Paulus & Fikkert, 2014; Tenenbaum et al., 2014) or body movements and gestures (Booth, McGregor & Rohlfing, 2008; Rader & Zukow-Goldring, 2012) toward an object of joint attention. In addition, a child’s willingness to learn new words from others is influenced by whether they consider their interlocutor to be reliable or unreliable (Krogh-Jespersen & Echols, 2012; Scofield & Behrend, 2008; Sobel, Sedivy, Buchanan, & Hennessy, 2011), and knowledgeable or ignorant (Brooker & Poulin-Dubois, 2013; Mangardich & Sabbagh, 2018; Sabbagh & Baldwin, 2001, 2009). In doing so, young children use cues from their interlocutors’ behaviors to facilitate language learning.

Prior studies have exposed children to a range of social cues – namely pointing and eye-gaze – to understand how they support the learning of different types of words, including nouns, verbs, and adjectives. Experimental studies (Akhtar, Carpenter, & Tomasello, 1996; Baldwin et al., 1996; Houston-Price, Plunkett, & Duffy, 2006; Striano, Chen, Cleveland, & Bradshaw, 2006; but see Scofield & Behrend, 2011) and longitudinal studies (Brooks & Meltzoff, 2008; Morales et al., 2000a, 2000b; Tenenbaum et al., 2014) have determined that the presence of social cues promotes children’s accuracy in identifying words during test phases that, for example, involve selecting or looking at a target referent. As an example, Morales et al. (2000a) found that individual differences in infants’ responses to initiations of joint attention by their caregivers between 6 and 18 months were significantly related to language outcomes at 24 months and 30

months of age. They also found that an aggregate measure of joint attention skills (between 6 and 18 months) predicted language outcomes at 30 months, even after accounting for language skills at 24 months. This suggested that children who possess better vocabulary skills at 24 months did not simply maintain their level of vocabulary in comparison to their peers at 30 months of age. Instead, the gap between higher- and lower- vocabulary individuals widened over time based on variation in social aspects of communication. Thus, individual differences in children's ability to engage with social cues seems to be predictive of their later language outcomes.

This leads us to ask: *why* do social cues facilitate word learning? What functions do social cues serve during the observation, processing, and learning of new words? There are three broad proposals that have aimed to answer these questions. First, in the socio-pragmatic theory of word learning, word learning requires children's recognition of language as a means to communicate, where social cues would support word learning by representing an interlocutor's communicative intent (Akhtar, Carpenter, Tomasello, 1996; Akhtar & Tomasello, 1998, 2000; Baldwin, 1993; Bruner, 1983; Nelson, 1988; Tomasello, 2000, 2001). In a seminal study, Baldwin et al. (1996) found that 18- to 20-month-old children were more likely to establish a word-object mapping when learning occurred when a speaker paid attention to the same object as the infant, compared to when the speaker was not physically present. They concluded that the speaker's referential intent was an important component of children's learning. Tomasello, Strosberg, and Akhtar (1996) came to a similar conclusion using affect-related social cues. In their study, 18-month-old children played a game with an experimenter. The experimenter announced her intention to find a novel object using a nonce word (e.g., "Let's find the *gazzer*"). She picked up one nonce object, rejected it with disappointment, and gleefully picked up the target object. Children immediately mapped the nonce word to the latter. These studies argue that

children are sensitive to the communicative nature of word learning and perceived these cues to be indicative of the referential intent of an interlocutor. Thus, the primary function of social cues is to provide *social information* – or communicative information about a speaker’s goal or intention (such as which object is being referred to in space).

A second proposal is based on associative mechanisms by arguing that social cues drive word learning via general cognitive processes, such as attention and memory (Colunga & Smith, 2005; Samuelson & Smith, 1998; Smith, 2000). This proposal prioritizes associative mechanisms over socio-pragmatic information, on the assumption that very young children would be unlikely to know how to interpret social information in a skillful way. What is likely happening, they argue, is that social cues provide information related to timing, spatial information, or simple associations rather than intent (Deák, Krasno, Triesch, Lewis & Sepeta, 2014; Yu & Smith, 2012). For example, Samuelson and Smith (1998) argue against the socio- pragmatic conclusions drawn by Akhtar et al. (1996), suggesting that if learners’ behaviors seem to indicate the recognition of referential intent, they may instead be behaviors that were “guided by the general effects of a contextual shift on memory and attention” (Samuelson & Smith, 1998, p. 94). This work emphasizes that word learning, particularly at younger ages, may be based mostly on general cognitive mechanisms rather than a recognition of the communicative intent of others.

A third approach has, in some sense, united these two proposals, suggesting that perceptual and social information work in tandem to support word learning. The Emergentist Coalition Model (ECM; Golinkoff & Hirsh-Pasek, 2006; Hollich et al., 2000) combines social-pragmatic and association-based theories and consists of three central tenets: 1) children are sensitive to perceptual, social, and linguistic cues, 2) the weights of these cues change over time, and 3) principles of word learning also change over time. This model suggests that children first

consider perceptual information when learning new words, such as object salience. For example, an infant may map a novel word to an object that grabs their attention, suggesting that perceptual information provides the initial momentum to help the word learning process. This hypothesis is corroborated by the results of Pruden et al. (2006), where 10-month-olds were found to rely on perceptual cues over social cues to learn words. However, the weights of these cues change over time. As children develop, they deemphasize the weight of perceptual cues in favor of social cues such as eye-gaze or pointing. ECM suggests that these cues transform in a similar way: social cues initially function to provide low-level perceptual information, but ultimately change to provide abstract social information such as a speaker's goals or intentions during word learning.

What remains underspecified in current frameworks are the diverse, complex functions that social cues play during word learning. Socio-pragmatic and association-based theories, as well as ECM, have generally assumed that social cues provide either social or perceptual information at different points in developmental time. However, these theories do not fully consider the dynamics of interactions between caregivers and children. It may not be realistic to neatly categorize the functions of social cues as perceptual or social. Doing so hides the richness of social cues for elucidating word learning in the moment and for explaining children's learning outcomes. Instead, their functions may fall along a gradient between perceptual and social information.

In what follows, we examine existing literature on young monolingual children's exposure to (mostly concrete) words and social cues, and we find that the supportive role of social cues cannot be neatly discretized into a perceptual vs. social division. Our review points to four distinct but interrelated functions of social cues in children's word learning, which vary

based on the type of social cue and the age of the child. How do different functions support early word learning, separately and in combination? We present this review as a way of informing frameworks of word learning to accommodate the richness and diversity of natural social cues.

We first provide an overview of a range of social cues and their suggested functions during word learning. We include articles that primarily focus on noun learning by monolingual children, although verb- and adjective-learning studies are described at times. Then, we discuss the two social cues that have been studied thoroughly in previous research on word learning – eye-gaze and gesture – and we detail their possible functions. Following this overview, we discuss key directions for future research that are needed to create a complete theory of how children learn words within socially-rich contexts.

1.3 What are the functions of social cues?

The first finding from our literature review is that there are four distinct but interrelated functions that social cues may serve in word learning: orienting attention, extracting information, disambiguating referents, and indicating referential intent (Figure 1). These four functions were generated based on wording from many empirical articles described below, and/or from our interpretations of the functional roles played by social cues in the experiments. This way of categorizing different functions, although imperfect, builds upon the general assumption in prior literature that social cues provide only perceptual or social information. Orienting attention refers to a caregiver's cues that serve to orient a child's attention to an area in space. This is mostly a perceptual function that may operate separately from a caregiver's social desire to orient attention. Extracting information refers to the influence of a caregiver's cues on the type of information that a child detects and/or encodes from a visual scene, such as an object's features

(e.g., Reid et al., 2004). Disambiguating referents refers to social cues from a caregiver that reduce ambiguity between a word and its possible referents or meanings. Lastly, indicating referential intent refers to a caregiver displaying social cues to convey their intent to communicate with the child.

The second idea emerging from our review is that these four functions can provide a range of perceptual and social information, and to differing extents depending on the function. One could argue that orienting attention relies predominantly on perceptual information, at least initially, but with increasing influence from social context. Similarly, one could argue that inferring referential intent requires an aptitude for inferring abstract social information, but importantly, perceptual information cannot be divorced from this process. In addition, disambiguation could be categorized as perceptually- and/or socially-based, depending on context and interpretation. Finding a target referent among distractor referents could be the result of a perceptual mechanism, or from an implicit or explicit understanding that the referent was highlighted based on a caregiver's desire to communicate. Finally, a child's ability to extract relevant information from a scene may be due to the perceptual nature of a social cue, or by inferring a speaker's communicative intent in using it.

These functions are also likely to be at least somewhat interdependent on one another. Categorizing the functions of social cues in a binary way as perceptual or social is probably not an accurate way to think about their influence on learning. Instead, functions are likely to depend on each other, or the weights of perceptual and social information are likely to vary for different social cues. For example, extracting relevant information about objects, events, and/or people requires the ability to follow cues from others in a way that shifts attention, but also the ability to use a caregiver's attention to infer referential intent. It is also likely that how children extract

information can be influenced by whether the referential intent of the caregiver is detected, such that information may be extracted better/faster when a child recognizes communicative intent. Similarly, a child's visual attention may be oriented faster when referential intent is detected. These examples suggest that the functions of social cues are inherently interactive, that the boundaries between each function are somewhat fuzzy, and that the functions may build upon each other to facilitate learning. Considering social cues as providing only perceptual or social information may mask the full richness of information offered by eye-gaze, gesture, and other social cues.

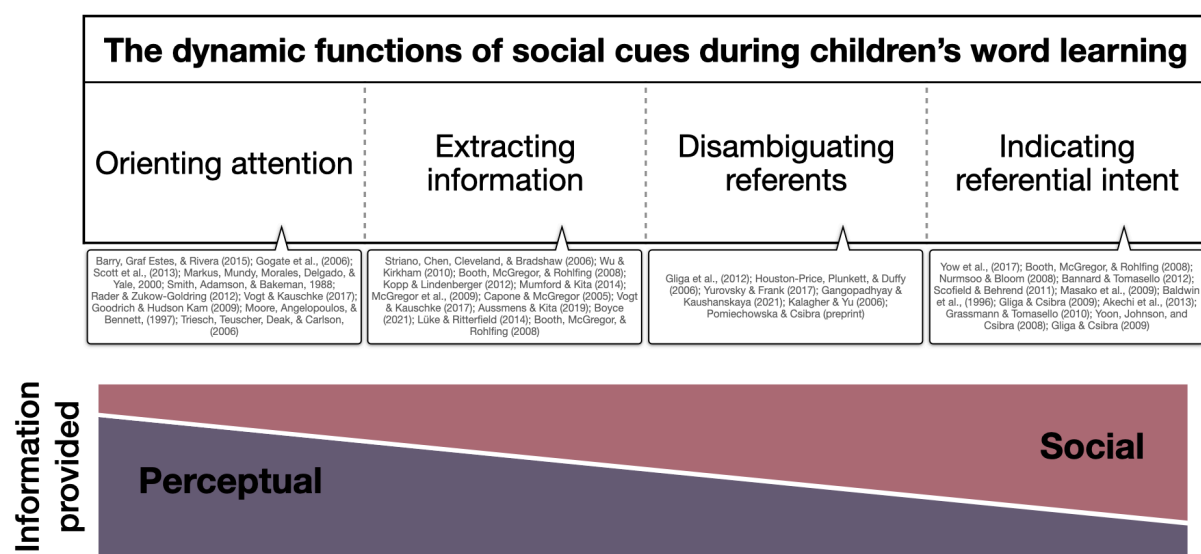


Figure 1. Key functions of eye-gaze and gestures reported in previous studies. The upper portion of the figure depicts the functions of social cues described in prior literature, with relevant studies cited underneath. The dashed lines between functions represent the fuzzy boundaries and interactiveness between functions. The lower portion of the figure depicts the type of information that social cues may provide across the four functions, demonstrating that each may provide

ranging “amounts” of perceptual and social information, but with notable overlap between the two.

We also do not find a specific point in time where children transition from using social cues as a source of perceptual information to the same cues being a source of social information. While many studies do show that children follow the developmental trajectory hypothesized by ECM, there is sizable variability within age points. Across studies within the same developmental time period, social cues seem to play different functions in different experiments. Our review indicates that there is more to be explained about the functions of social cues across children and across time. Therefore, in the sections that follow, we focus on eye-gaze (Section 1.4) and gestures (Section 1.5) and examine the four main ways in which they support children’s word learning. Within the descriptions of each function, we provide a brief overview of relevant studies. Where pertinent, we include the ages of participants, which range from 6 months to 6 years across different studies, to provide a window into developmental change.

1.4 The functions of eye-gaze as a social cue

Gaze following refers to the ability to follow the eye-gaze of another during an interaction (Brooks & Meltzoff, 2005; Moore, 2008), and infants demonstrate this ability from early on in life. They begin paying attention to the eyes of others as early as 2-3 months of age (Caron, Caron, Caldwell & Weiss, 1973; Maurer & Salapatek, 1976). This ability continues to develop in subsequent months (Vecera & Johnson, 1995) and stabilizes around 6 to 8 months of age (D’Entremont et al., 1997; Gredebäck et al., 2010). This early attentiveness to eyes is important as it is required for two important social skills: gaze following (in sighted children) and joint attention. Children use this skill to develop the ability to determine and share common

reference points between themselves and a caregiver, which is also known as joint attention (Çetinçelik, Rowland, & Snijders, 2021; Mundy & Newell, 2007). For the purposes of this review, we collapse “gaze-following” and “joint attention” because both involve similar behaviors (i.e., following the eye-movements of another), and are deeply intertwined as following eye-gaze is a crucial aspect of engaging in joint attention in sighted children (Çetinçelik, Rowland, & Snijders, 2021).

The ability to follow eye-gaze has been linked with vocabulary development (Akhtar & Gernsbacher, 2007; Brooks & Meltzoff, 2005, 2008; Carpenter et al., 1998; Morales, Mundy & Rojas, 1998; Morales et al., 2000a; Morales et al., 2000b; Tomasello & Farrar, 1986). Brooks and Meltzoff (2005) found in a longitudinal study that gaze-following abilities at 10-11-months were positively related to word and phrase comprehension, as well as to total gestures when measured again at 18 months, as measured by the MacArthur Child Development Inventory (CDI; Fenson, Dale, Reznick, Bates, Thal & Pethick, 1994). In a different study, positive relationships between gaze-following abilities and later productive vocabulary were also found (Brooks & Meltzoff, 2008). Given the link between gaze-following and language learning, what is the function of eye-gaze cues during moments when children have the potential to learn a new word? We turn to each of the four purported functions of eye-gaze cues in turn.

Orienting attention

One way in which word learning outcomes can be influenced by social cues is through orienting attention, which at first pass seems categorizable as a perceptual function. Infants can follow the movements of the eyes or pupils (Farroni, Johnson, Brockbank, & Simion, 2004) and the direction of head motion (Corkum & Moore, 1998; D’Entremont, 2000; Moore & Corkum,

1998) to shift their attention to a specific object in a visual scene, and this variation in attention predicts subsequent word learning. Gogate et al. (2006) found that the extent to which 6- and 8-month-old infants gazed between their caregiver and the gazed-at-object positively predicted subsequent novel word learning. Moreover, individual differences in the number of parent-infant joint attention events in a fixed time window has been found to predict variation in early vocabulary size (Carpenter, Nagell, & Tomasello, 1998; Farrant & Zubrick, 2012; Markus, Mundy, Morales, Delgado, & Yale, 2000; Scott et al., 2013; Saxon, 1997; Smith, Adamson, & Bakeman, 1988; Tomasello & Farrar, 1986). Joint attention can be further analyzed as cases that are caregiver-initiated or child-initiated. For instance, Scott et al. (2013) found that, compared to when mothers followed infants' attention to engage in joint attention, how often infants followed mothers' attention related to later vocabulary growth, suggesting that parent- infant joint attention events are indeed important in lexical development.

Extracting information

Beyond directing and orienting attention to a stimulus in a visual scene, taking advantage of eye-gaze may enhance the information extracted from visual scenes. This could also be categorized as providing perceptual information, perhaps in a way that follows naturally from a redirection of attention. In a typical joint attention scenario, the eye- gaze cues provided by another speaker could enhance the processing of a gazed-at object by improving encoding or by highlighting relevant information about the object. This indirectly facilitates the mapping between a given label and object in the world (Çetinçelik, Rowland, & Snijders, 2021).

Several studies have found that the presence of eye-gaze cues does facilitate extracting information from a visual scene (Hirotani et al., 2009; Hutman et al., 2016; Kopp & Lindenberger, 2011; 2012; Reid et al., 2004). For example, Reid et al. (2004) first established different ERP signals when an object was cued by another's eye-gaze or not; during a second viewing, 4-month-old infants showed an increased positive slow wave (PSW) for objects that were not cued by another's gaze, an indication for a memory representation of a partially encoded stimulus (Snyder, 2010). This result suggests that objects that were cued through eye-movements were more familiar than objects that were uncued. The subsequent viewing of the uncued object led to an updating of its mental representation. In comparison, objects that had been accompanied by eye-gaze toward them initially required less of an update, suggesting that information about the object had already been encoded by virtue of its initial presentation (i.e., gazing at the same object another has looked toward).

Eye-gaze cues can also facilitate information extraction even when situations are more cognitively challenging. For example, Striano et al. (2006) examined the effect of joint attention on object processing in 9- and 12-month-old infants. In a joint-attention condition, a speaker alternated gaze between a single toy and the infant. In an object-only condition, a speaker gazed at the toy and at the ceiling, but never the infant. At test, infants were presented with the toy they saw previously, and a novel toy. The researchers found that only infants in the joint-attention condition preferred to look at the novel toy at 9 months. By 12 months, infants looked longer toward the novel toy in both conditions. This suggests that while 12-month-olds were able to process enough information about the object in either condition, encoding the object was significantly improved for the 9-month-olds by the presence of the speaker's eye-gaze. The authors suggest that infants' dependence on eye-gaze as an informative cue depends on the

complexity of the learning task; in this experiment, encoding of the object may have been complex for 9-month-old infants but not for 12-month-old infants, such that they could rely on eye-gaze in the absence of bids for joint attention from the speaker.

Eye-gaze cues from another not only affects how much information is extracted from a visual scene, but also what type of information is extracted (Hoehl et al., 2012; Michel et al., 2019; Reid & Striano, 2005; Wahl et al., 2013, 2019; Wu & Kirkham, 2010; Wu et al., 2011). Wu and Kirkham (2010) examined how infants' learning of multi-modal audio-visual events was affected when attention was cued to a portion of a screen from either eye-gaze cues or perceptual cues. They found that while social eye-gaze cues helped 4-month-olds learn where to look on a screen, those same social cues additionally helped 8-month-olds learn the specific multimodal events they were exposed to. Learning effects were not found with perceptual cues such as flashing central squares, despite reliably orienting attention to sides of the screen. Some have speculated that the information abstracted from a situation can also be retained longer when eye-gaze from the speaker is present. Kopp and Lindenberger (2011, 2012) examined 4- and 9-month-olds' immediate recognition and long-term retention of novel objects depending on whether these objects were presented during high or low moments of joint attention (JA). Infants were exposed to a series of novel objects, and an experimenter either alternated her gaze between the object and speaker (high-JA) or only looked at the object (low-JA). The researchers found that presence of joint attention modulated 9-month-olds' recognition both immediately after exposure, and after a one-week delay. In contrast, they found that 4-month-olds' recognition was evident only after a one-week delay, although they attribute this difference to experimental issues of ERP detection in young infants. Nevertheless, they argue that the presence of joint attention

during learning scenarios may have facilitated elaborated encoding and storage, improving long-term retention.

Disambiguating referents

Another proposed function of eye-gaze is to disambiguate referents in a visual scene. Children can readily use the eye-movements of another to disambiguate objects in a sensitive way: Hembacher, deMayo, and Frank (2020) found that 3- and 4-year-old children used the gaze of another to identify a referent in different ways depending on the scenario. They looked to the experimenter to provide disambiguating information the most when two novel objects were presented, an intermediate amount when the object set included one novel and one familiar object, and the least when both objects were familiar. Likewise, in one of the experiments conducted by Graham et al., (2010), researchers sought to determine how 24-month-old children used eye-gaze to map novel labels to objects when presented with multiple sets of referents (both familiar and novel). They found that children were equally likely to map novel labels to the gazed-at object, regardless of whether the objects were familiar or novel. This suggests that young learners can indeed use eye-gaze to disambiguate a referent.

The ability to disambiguate visual referents through eye-gaze has also been shown to improve word learning (Houston-Price, Plunkett, & Duffy, 2006; Law, Houston-Price, & Loucas, 2012). Law, Houston-Price, and Loucas (2012) tested 18-month-olds in a word-learning task that required them to follow a speaker's gaze to one of two novel objects and learn its novel label. Vocabulary measures were collected at 18, 24, and 30 months of age. They found that 18-month-old children were able to follow eye-gaze to disambiguate a visual referent and learn a

novel label, and that performance on this disambiguation task was related to vocabulary both concurrently and longitudinally.

At the same time, using eye-gaze as a disambiguating cue is not always successful. Frank, Tenenbaum, and Fernald (2013) examined the efficacy of using mothers' eye-gaze to disambiguate a visual referent in mother-child dyads. They found that mothers' eye-gaze was noisy and did not fully disambiguate the referred-to object, suggesting that, in naturalistic situations, relying on an eye-gaze cue may not always lead to successful disambiguation. Furthermore, children do not always rely on eye-gaze cues for disambiguation when other sources of information are available. Graham et al. (2010) found that children are not likely to use eye-gaze as a disambiguating cue when they can rely on the "mutual exclusivity" assumption as another source of information. The mutual exclusivity assumption is the assumption that one object can have only one label (Markman, 1992; Merriman & Schuster, 1991). The question in the experiment by Graham et al. was whether children would link a novel label to a familiar object (as directed by eye-gaze) or a novel object (as directed by a mutual exclusivity assumption). Children were presented with sets of one novel object and two familiar objects, and on half of the trials, the experimenter asked for the novel object using a novel label. Critically, the experimenter's eye-gaze provided an inconsistent clue as she looked at a familiar object instead. They found that children mapped the novel label onto the novel object, suggesting the mutual exclusivity assumption overrode information provided by eye-gaze. Thus, eye-gaze as a disambiguating cue is not necessarily used when it contrasts with prior knowledge.

Indicating referential intent

As mentioned previously, some scientists propose that the ability to follow gaze is evidence for infants' inferring another's referential intent, indicating a direct uptake of social information (Deligianni, Senju, Gergely, & Csibra, 2011; Senju & Csibra, 2008; Striano & Reid, 2006; Tomasello, 2000; Tomasello & Barton, 1994; Tomasello, Strosberg, Akhtar, 1996). Senju and Csibra (2008) investigated whether 6-month-old infants would follow the eye-gaze of a speaker more often if the speaker presented direct eye contact to the participant, compared to an interesting perceptual cue (i.e., a moving flower) prior to the initiation of any eye movements. The authors found that infants who were presented with direct eye contact made more first eye movements to the gazed-at object, as well as made more saccades between the face of the speaker and the gazed-at object. In contrast, the perceptual attention-getter did not facilitate as much gaze-following behavior as the ostensive cue (direct eye-contact). They interpret these results to suggest infants can indeed follow the gaze and attention of others, but require strong social, ostensive cues to do so. Perceptual cues or attention-getters alone were not enough to initiate gaze-following.

Beier and Carey (2014) attempted to figure out whether 12-month-old infants inferred perceptual or social information from an interaction by presenting them with a novel entity (a colored blob) and examined whether infants would follow its movement (i.e., attentional focus). They found that infants followed the attentional focus of the blob when it responded contingently (e.g., when an experimenter looked at it and it lit up), which could be construed as infants inferring either perceptual or social information. However, crucially, infants demonstrated differentiated behaviors when the context did vs. did not suggest a social interaction. When an experimenter clapped two sticks together with a neutral facial expression and the blob lit up, infants did not follow the entity's focus. However, when an experimenter clapped her hands

together and smiled warmly at the blob, and when it then lit up, the infant shifted their gaze. This suggests that infants attributed agency to the non-human object and inferred referential intent when the context seemed to communicate a social interaction. This study provides further evidence for children's ability to detect social information.

Studies examining gaze-following in children older than 12 months routinely find that the function of following eye-movements is to infer the referential intent of a speaker (Baldwin, 1991, 1995; Butler, Caron, & Brooks, 2000). For example, Baldwin (1991) taught 16-19-month-old infants novel names of objects when a speaker was either: focused on the same object as the infant or looked at an object different from the infant. Their experiment was designed to determine whether infants use nonverbal, social cues from another to infer word-object mappings or if they indiscriminately learned mappings regardless of where a speaker is looking. They found that children both learned the novel label when the participants' and speakers' attention coincided, as well as avoided mapping errors when the object of focus did not coincide with the speaker's view. This work suggests that children can use eye-gaze as a non-verbal cue that is relevant to the formation of novel label inference.

Understanding the many functions of eye-gaze for word learning

The functions of eye-gaze as a cue to language learning vary widely. From very early ages of development, the eye movements of others serve to orient infants' attention. Beyond attention, eye-gaze has been shown to influence the information extracted from a visual scene, such as enhancing object processing or retaining different types of information. Moreover, eye-gaze can be used to indicate referential intent. The proposed transition from perceptual to social effects, as in the ECM model by Hollich and colleagues (2000), may depend on the age of

individual learners. For instance, as infants mature, the information abstracted from eye movements may change from perceptual to social (e.g., from relying on gaze to guide their own attention vs. later using it to understand a speaker's intention).

At the same time, it is reasonable to question if this proposed developmental transition is an accurate characterization of how children extract perceptual and social information. We argue that this is not the case: in our overview, we find that eye-gaze can act in perceptually- or socially-based ways at the same time point in development. For example, some studies conclude that, under the right conditions, the social intent of a speaker can be inferred by infants younger than 12 months of age (see Senju & Csibra, 2008; Senju, Csibra, & Johnson, 2008, for a review see Striano & Reid, 2006). In opposition, other studies with the same age of participants suggest that only perceptual information is conveyed (Brooks & Meltzoff, 2005; Corkum & Moore, 1998; Moore & Corkum, 1998). However, eye-gaze is likely to take on a richer social meaning as children age, based on the increasing diversity of social and communicative experiences that children engage in across time as well as the maturation of underlying neural and cognitive mechanisms. Furthermore, the different functions of eye-gaze can occur in some combination or along a gradient – and it is again likely that a single eye-gaze cue can serve an increasing number of functions as children develop. It does not seem to be the case that there is a developmental time point where children qualitatively change their usage of a cue from perceptual to social in service of word learning. Rather, children likely take advantage of social cues to support word learning in whatever way makes sense in the moment. For example, eye-gaze can be used to guide attention throughout development, but even from early in life, gaze may indicate referential intent only in certain contexts. This variation in functional roles is likely to be explained by other factors (see Section 1.6).

Relatedly, some studies cannot definitively determine the function of an eye-gaze presented in an experiment. Often, studies do not find evidence for eye-gaze functioning as conveying perceptual information over social information. Consider Striano et al. (2006). In this study, 12-month-olds looked toward a novel object in both a joint-attention and object-only condition, while 9-month-olds looked toward the novel object reliably more than chance only in the joint-attention condition. What function did joint attention serve for 12-month-olds in this experiment? The authors conclude that while 9-month-olds benefitted from joint attention to enhance object processing in a cognitively challenging scenario, 12-month-olds used domain-general mechanisms to encode object characteristics regardless of experimental condition. However, it remains possible that the heightened performance of the 12-month-olds is the result of inferring referential intent from following the gaze of the speaker in both conditions, as opposed to using eye-movements for simply orienting attention to a spatial location. The experimental setup may have facilitated object processing regardless of whether it was establishing joint attention or simply following the eye-movements of another. This alternative explanation remains plausible, as other research has argued that 12-month-olds can determine the referential intent of another (Csibra & Volein, 2008; Okumura et al., 2013; Senju & Csibra, 2008; Striano & Reid, 2006). For example, Okumura et al., (2013) found that 12-month-olds anticipated the appearance of objects on a screen when the objects were previously referred to by a human agent, compared to a robot's gaze.

In the opposite way, consider Butler, Caron, and Brooks (2000). Fourteen- and 18-month-old infants were engaged in joint attention events with an experimenter who looked toward two visual events (one on each side) in varying scenarios. These scenarios included: when an experimenter's view was unobstructed, when their view was obstructed by opaque screens, and

when their view was obstructed by opaque screens but included windows to the targets.

Eighteen-month-olds turned in accordance with the experimenter's gaze much more when the experimenter's view was unobstructed, suggesting that they understood the referential nature of eye-gaze. Fourteen-month-olds, however, were much more variable – they turned less in accordance with the experimenter in comparison to the 18-month-olds, regardless of whether their view was obstructed. This variation would seem to contradict the inferences concluded by research such as Okumura et al. (2013) in that 14-month-old infants may have been able to follow the gaze of another, but they did not understand the referential intent behind this movement. It could be the case that the elaborate set-up of the experiment might have presented too difficult a situation for the 14-month-olds, but nevertheless, we cannot infer that the 14-month-olds are detecting referential intent.

In summary, the functions of eye-gaze cues during word learning present a much less defined picture compared to the black-and-white dichotomy between perceptual and social information that has sometimes been assumed. While it is fairly clear that the functions of eye-gaze cues vary by the age of the child, there is not a clear transition from inferring perceptual to social information, as all age points, including infancy, may infer both perceptual and social information. Furthermore, it is difficult to determine whether a social cue is providing only perceptual or only social information. A more conservative conclusion is that eye-gaze cues can convey a combination of perceptual and social information at the same time. Therefore, other factors may be at play that influence the function of social cues during word learning.

1.5 The function of gesture as a social cue

Gestures – hand movements that co-occur with speech – are prevalent in a sighted child’s visual input throughout development (Goldin-Meadow & Singer, 2003; Vigliocco et al., 2019). While deictic gestures (e.g., pointing) are the most prevalent in a child’s environment, representational gestures provided by caregivers are also present (e.g., holding up two fingers in a V-shape to represent a rabbit’s ears) (Iverson et al. 1999; Özçaliskan & Goldin-Meadow 2005; Rowe et al., 2008). (Conventional gestures – i.e., gestures that are used by a particular culture or community, such as waving a hand to say “hi” – are certainly present in children’s communicative environments, but we are not aware of research that suggests they facilitate word learning). Children exploit information conveyed by gestures beginning at an early age. For example, between 9-12 months of age, infants can reliably follow pointing gestures to nearby objects (Bates, 1976). At about 14-15 months of age, children demonstrate the ability to follow pointing gestures to objects further away or outside of their visual field (Delgado et al., 2002; Desrochers, Morissette, & Ricard, 1995). These gestures are ultimately incorporated as important sources of information during word learning scenarios. For example, deictic gestures that indicate a referent in space are often related to more accurate word learning (Booth, McGregor, & Rohlfing, 2008; Brooks & Meltzoff, 2008; Carpenter et al., 1998; Gogate et al., 2000; Kalagher & Yu, 2006; Harris et al., 1995; Vogt & Kauschke, 2017). Representational gestures too are related to improvements in word learning (Boyce, 2021; Goodwyn et al., 2000; Lücke & Ritterfield, 2014; McGregor, 2008). But exactly how do gestures benefit word learning?

Orienting attention

One function of gestures is to orient attention to a specific location in space. Thoermer and Sodian (2001) examined whether infants at 10 and 12 months could follow the pointing gesture of an adult over multiple trials. They found that most infants at 10 months of age did not reliably follow gestures, while they did at 12 months of age. But in a visual dishabituation task, they also examined whether infants interpreted pointing as goal-directed or referential. To do so, an experimenter presented infants with two objects. They habituated infants to movement of the experimenter, where she looked and pointed to one of the objects. They then switched the locations of the objects and examined whether infants' looking behaviors differentiated depending on whether they saw the experimenter follow a new path (with points to the same object) or point to a new object (but following the same spatial path). If infants interpreted pointing behavior as indicating referential intent, they would exhibit longer looking times when dishabituated to the experimenter pointing to a new object compared to a new path. If they were simply following spatial motion, they would show longer looking times when they saw the experimenter point to a new path compared to a new object. Researchers found no significant difference between the conditions in 10-month-old infants but 12-month-olds looked longer at changes in path, suggesting that pointing behavior was not interpreted as indicating the referential intent of the experimenter, but rather allowed infants to follow the spatial location of the object. They thus conclude that at 12 months, infants were more likely to follow perceptual information conveyed by pointing. Other studies have also found that deictic gestures are used to orient attention (Vigliocco et al., 2019) and, as a result, enhance word learning (Deák et al., 2008; Deák, Flom, & Pick, 2000; Deák et al., 2018; Rader & Zukow-Goldring, 2012).

Only one study examined whether representational gestures facilitate word learning by orienting attention. Goodrich and Kam (2009) examined how 2- to 4-year-old children and adults

used co-speech representational gestures to infer novel intransitive verb meanings. They taught participants four novel intransitive verbs either with representational gestures matching the action described by the verb, with uninformative gestures (e.g., pointing at the child), or with speech alone. Both children and adults were better able to infer the novel meanings when speech was paired with representational gestures, compared to when learning occurred with uninformative gestures or speech alone. The researchers conclude that the function of representational gestures, at least in the realm of verb learning, is mainly to direct and sustain a child's attention to a given scene.

Extracting information

By comparison, representational gestures have more often been thought to support information extraction because they are usually schematic in nature, allowing gestures to convey what is critical in a visual scene (Aussems & Kita, 2019; de Ruiter, 2000; Kita, 2000; Kita, Alibali, & Chu, 2017; Vigliocco et al., 2019). Aussems and Kita (2019) explored how representational gestures impacted what information 3-year-olds encoded from observing various events. They found that participants who saw actors moving in unusual ways, along with gestures that represented that movement, recalled these events better than control conditions. This suggested that seeing representational gestures while encoding events schematically highlighted critical aspects of the scene, which improved later recall.

Exposure to representational gestures not only impacts what information is extracted, but also facilitates word learning (Boyce, 2021; Capone & McGregor, 2005; Lüke & Ritterfield, 2014; McGregor et al., 2009; Mumford & Kita, 2014; Vogt & Kauschke, 2017). Capone and McGregor (2005) taught six novel labels to 2-year-olds and varied whether the labels were

accompanied by representational gestures or not. They found that the toddlers retrieved the novel labels more accurately both immediately after learning and after a four-day delay when exposure occurred with the gestures. The inclusion of the representational gestures appeared to reinforce the semantic content of the novel label: the gestures may have drawn attention to the critical semantic information of the word learning problem, allowing for improved retrieval and recall.

This increased attention to relevant information also improves children's generalization of novel meanings. Mumford and Kita (2014) found that 3-year-olds used representational gestures seen during novel verb learning to guide their generalizations of those meanings. When generalizing verbs to novel scenes that could have change-of-state or manner meaning, children demonstrated a stronger manner generalization when they had been exposed to manner gestures in comparison to either end-of-state gestures or no gestures. In this way, the representational gestures appeared to help children identify which information in a visual scene was relevant or critical to the meaning of the novel verb, which facilitated learning and generalization of the verb's meaning.

Disambiguating referents

A third main function of gestures is to disambiguate speech in context. Studies have found that deictic gestures are relied on by children to disambiguate ambiguous scenarios (Grassmann & Tomasello, 2010; Kalagher & Yu, 2006; Pomiechowska & Csibra, preprint). Grassmann and Tomasello (2010) pit mutual exclusivity and socio-pragmatic conventions against each other by presenting 2- and 4-year-olds with a novel and familiar object. In their critical condition, an experimenter pointed to a known object (e.g., cup) and requested a novel label (e.g., the modi). They sought to determine whether participants would prefer to follow social cues (e.g., choosing

the object that was pointed to) or adhere to a mutual exclusivity bias and choose the novel object (since they already knew the word cup at this age). They found both age groups were more likely to provide the pointed-to object in response to the experimenter's request. This indicated that children were more likely to use gestures as a source of information, compared to lexical information like mutual exclusivity, to disambiguate the immediate environment.

This study mirrors observational research (O'Neill et al., 2005; Özçalışkan & Goldin-Meadow, 2005; Vigliocco et al., 2019). O'Neill et al. (2005) examined the gestures that English-speaking mothers used in two contexts – a counting task and a free-play session – when interacting with their 20-month-old infants. They found that mothers overwhelmingly used deictic gestures like pointing when interacting with their child (97% of gestures in the free play session, and 91% of gestures in the counting task). These gestures were then used most often to disambiguate referents in the environment. Indeed, all 12 mothers in this task used deictic gestures most often to convey disambiguating speech acts. In addition, this work found that representational gestures were used sparingly in free-play contexts and completely unused in the counting task. Similarly, a study conducted by Vigliocco et al. (2019) examined caregivers' use of gestures in a free-play session with their 2-year-olds. They found that deictic gestures were primarily used to point out target objects that were close in space to other objects.

Although representational gestures are common in natural interactions, our literature review found few studies suggesting that they assist in the disambiguation of referents. In the free-play study by Vigliocco et al. (2019), representational gestures were found to be helpful for disambiguation in contexts when target referents were absent. Other work has shown that children can use representational gestures to disambiguate referents and learn novel verbs

(Goodrich & Kam, 2009), but the argument in this work was that representational gestures function to heighten and direct attention to a certain aspect of a visual scene.

Indicating referential intent

Finally, gestures have been shown to inform children of speakers' referential intent. Deictic gestures have often been included to indicate referential intent in communicative scenarios (Akechi et al., 2013; Behne, Liszkowski, Carpenter, & Tomasello, 2012; Booth, McGregor, & Rohlfing, 2008; Csibra & Volein, 2008; Gliga and Csibra, 2009; Paulus & Fikkert, 2014; Sodian & Thoermer, 2004; Yoon, Johnson, and Csibra, 2008). Gliga and Csibra (2009) investigated whether 13-month-olds appreciate deictic gestures as communicative. In one condition, they presented infants with objects located on the left and right sides behind a barrier. An experimenter pointed to and named one of the hidden objects (e.g., duck), and measured infants' looking times in response to raising the barrier and finding the object either on the indicated or the unexpected location. They found that infants looked longer when the object was presented in the unexpected location, compared to the indicated location, suggesting that they believed the experimenter to have referred to a unique object at a specific location. Furthermore, the researchers argued that a simple associative view cannot explain their results. If infants were following only the pointing motion, different looking behaviors would not be predicted for an unexpected object. In addition, if infants were only using the verbal label to anticipate the hidden object, there should not have been different looking behavior because the expected object was always present. They argue that the only way to explain infants' looking behavior is that the infants believed the pointing gesture and labeling to refer to a combination of object and location.

Understanding the many functions of gesture for word learning

Taken together, prior studies point to a number of ways in which gestures facilitate word learning (see also Cook, 2018), including orienting attention, extracting information, disambiguating referents, and indicating referential intent.

As in our review of research on eye-gaze, we can ask the question: Do gestures serve a perceptually- or socially-based function, or some combination of the two? Is there a clear developmental trend that captures the many documented ways that children exploit gestural information? Similar to our conclusion about the function of eye-gaze, we do not find a clear answer. Instead, we find an overlap of functions that could be categorized as perceptually- or socially-based within the same time frame. Even at the age of 12 months and younger, researchers have concluded that infants use gestures to both orient attention and infer referential intent. This suggests that gestures do not simply serve one stable function, and there may be no inflection point in development where social cues change functions in a qualitative way.

Like eye-gaze, we also find that there are no clear, distinct boundaries between the different types of functions. Instead, there are four interrelated functions that seem to guide word learning. Booth, McGregor, and Rohlfing (2008) sought to determine how 2-year-olds use social cues to learn new words in a novel word-learning task. Words were taught to children in one of the following conditions: eye-gaze cues alone, gaze + point, gaze + touch, gaze + manipulation (movement). Booth et al. reasoned that if orienting attention was the primary function of social cues, accuracy should increase as cues increased in directness (e.g., touching a referent is more unambiguous than pointing). However, they found that the only difference in learning was found between the gaze alone and gaze + point conditions. No increases in performance were observed

in learning conditions that provided increasingly direct social cues. Instead, children simply seemed to attend longer during the condition that involved both gaze and touch. And notably, time spent looking at the target object accounted for small but significant amounts of test accuracy variance. What this study suggests is that information presented in multiple modalities or multiple cue types increases attention, which in turn explains some (though not all) of children's success in word learning.

Booth et al. put forth two (not mutually-exclusive) possibilities: one explanation is that that social cues vary in the degrees to which they indicate referential intent, and that a pointing gesture may sufficiently indicate intent for 2-year-olds. A second explanation is that social cues heighten attention to a learning scenario, and that gestures enable deeper encoding or more elaborated associations between labels and referents. What do we make of these explanations? If we are to consider these explanations as serving social versus perceptual functions, respectively, we would be unable to determine which explanation more accurately explains how children used the social cue presented to learn the novel words. The first explanation seems to support a socially-based perspective on social cues, the second a perceptually-based one. The interpretations of Goodrich and Kam (2009) demonstrate how difficult it is to parse the boundaries between each proposed function. They examined how representational gestures might inform verb learning for 2- to 4-year-olds and find that children can use representational gestures to disambiguate the meaning of a novel verb. Is the function of the gestures presented in this task to disambiguate meanings? Not necessarily, as the authors assume that the gestures served to “direct a listener’s attention to a given scene” (p. 86). This supports the idea that it is often unclear why a given social cue facilitates learning. To improve our understanding of why word

learning is enhanced by the presence of social cues, we need to grapple with the idea that there are (at least) four different functions of social cues, without distinct boundaries between them.

These studies indicate that the functions of gestures should not be categorized in a binary way as perceptually- vs. socially-based. Their functions instead both co-occur and vary depending on a number of contextual factors. Moreover, different types of gestures are likely to serve different functions: several prior studies have concluded that deictic gestures function primarily to direct attention as well as to disambiguate visual referents, representational gestures more often facilitate information extraction, and both are likely to convey referential intent. This interplay between gesture type and function has yet to be considered in a systematic examination of word learning in early childhood.

1.6 Implications for building a complete theory of word learning in context

Spanning our review of developmental research on eye-gaze and gesture, we can make two main conclusions. First, the functions of social cues during children's word learning do not involve a clear distinction between perceptual and social functions. That is, existing research does not support the idea of a purely associative or purely social mechanism to explain the role of social cues in word learning, and a number of studies have examined the overlap between these mechanisms (Houston-Price, Plunkett & Duffy, 2006; Moore, Angelopoulos & Bennett, 1999; Pruden et al., 2006; Yu & Ballard, 2007; Yurovsky & Frank, 2017). Second, we have seen that social cues can serve a variety of functions that include orienting attention, extracting information, disambiguating referents, and indicating referential intent. But importantly, it is not straightforward to pinpoint where a function exists along this perceptual-social dimension. For example, a given cue does not seem to just direct attention or disambiguate a visual referent

without also influencing what information is extracted – or vice-versa. Plus, in some cases, indicating referential intent may not even be possible without also directing attention. There are also interactions between different social cues (such as deictic or representational gestures) and their possible functions in particular learning contexts, and children’s abilities to infer social information becomes more sophisticated with age.

How do these conclusions relate to those of the most closely relevant word learning theory, the Emergentist-Coalition Model (Hollich et al., 2000)? ECM is currently the model of word learning that can best accommodate some of the conclusions from our literature review. To recap, ECM consists of three central tenets: 1) children are sensitive to perceptual, social, and linguistic cues, 2) the weights of these cues change over time, and 3) principles of word learning also change over time. Our review speaks to the first two tenets. Indeed, consistent with the first tenet, prior findings point to children’s robust ability to use social cues from a speaker. And like the second tenet, we can conclude that children increasingly infer social information over development. Overall, our findings are broadly consistent with ECM. However, in an effort to improve upon how the field conceives of social cues during word learning, we outline seven factors that we believe should be explicitly incorporated in theory-building.

Age

First, the age of the child influences how social cues are used when learning new words – an idea consistent with ECM. Our review of the literature is consistent with ideas articulated previously that very young infants use social cues in a perceptually-based way (e.g., orienting attention) and that older children are more successful in using social cues for socially- based functions (e.g., indicating referential intent). These age-related changes suggest that researchers’

predictions and conclusions are generally consistent with the predictions of ECM. However, ECM does not adequately contend with the dynamic functions of social cues during word learning across development. ECM primarily suggests a sequential change from perceptual functions to socially-based ones, but based upon our findings, these functions cannot be easily categorized. What is more likely, we argue, is that the functions of eye-gaze and gesture are intertwined from the beginning, spanning a range of functions that depend on a number of factors, including age. We suggest that theories of word learning should explicitly address the diversity of social cue functions both within and across ages.

Types of social cues

One key dimension needed to explain how social cues are integrated during word learning is the type of social cue used. We find overlapping functions between eye- gaze and gesture cues, but some types of these cues are used differently than others. For example, we determine that the type of gesture (representational or deictic) may determine its function during learning. Representational gestures may be more often used to influence the information extracted from a learning event, while deictic gestures may be used to orient attention or indicate referential intent. Eye-gaze and gestures are not the only social cues that should be considered. Other cues include affect, prosody, and speaker reliability, amongst others (Berman, Graham, Callaway, & Chambers, 2013; Doan, 2010; Herold, Nygaard, & Namy, 2011; Jaswal & Neely, 2006; Kelly, Morisset, Barnard, Hammond, & Booth, 1996; Krogh-Jespersen & Echols, 2012; Nicely et al., 1999; Nygaard & Queen, 2008; Sobel, Luchkina, & Tummeltshammer, 2020). It is important to consider how different types of social cues may improve word learning both

separately and simultaneously in order to obtain a comprehensive understanding of how social cues impact real-world word learning.

Functions of social cues

In our literature review, we also find that there are four main functions of social cues: orienting attention, extracting information, disambiguating referents, and indicating referential intent. As we stated previously, these functions may or may not be separable, as it is difficult to determine (for example) how indicating referential intent can occur without also orienting attention. That is, these functions do not cleanly fit into perceptual vs. social categories; a single function, especially in natural interactions, is likely to be supported by related functions along a perceptual-to-social continuum. Researchers should aim to understand how certain functions manifest in diverse communicative contexts, and how multiple functions occur together.

Word types

Our review primarily focused on how social cues intersect with children's learning of concrete nouns. This limited perspective precludes the potential for further variability as we consider various types of word learning goals. For example, there remains the possibility that some social cues are better suited for noun learning compared to verb learning. Research regarding the distribution of social cues and lexical types finds that caregivers often use gaze and touch regularly when labeling objects in the environment (Custode & Tamis-LeMonda, 2020). By comparison, they do not point to and label ongoing actions with the same frequency as referring to objects (Gleitman, 1990; Tomasello & Kruger, 1992). To our knowledge, there has not yet been a study examining the different functions of social cue for learning different word

classes (e.g., noun, verb, adjective). It is quite likely that social cues serve differentiable roles as children begin homing in on the meanings of different word types.

Times-scales of learning

Prior research has not focused on investigating another important aspect of word learning: how children learn from social cues across shorter and longer timescales, from seconds to years. Our examination of the literature has only reviewed language learning in the context of naming events, and considered the information provided by social cues in reference to these smaller timescales. These timescales are typically referred to as “fast-mapping” scenarios, where labels and referents are initially associated in memory in a first encounter (Carey, 1978).

However, researchers have also considered the importance of viewing word learning as a culmination of learning events rather than as a singular exposure (Wojcik, Zettersten, & Benitez, 2022). This extended period of learning enriches and strengthens the associations created, known as slow-mapping (Carey, 1978; Swingley, 2010). Studies examining slow-mapping find that different underlying mechanisms influence this learning, compared to fast-mapping (Horst, Parsons, & Bryon, 2011; McMurray, Horst, and Samuelson, 2012; Munro et al., 2012; Twomey, 2013). Thus, it is imperative to consider that if social cues have been shown to improve word learning in a fast-mapping context, do they also facilitate slow-mapping between labels and objects, attributes, and events? Some studies have begun to examine this. For example, Capone and McGregor (2005) found that representational gestures improved word retrieval for novel nouns at both fast-mapping and slow-mapping intervals (i.e., on average 12 days between testing). Deictic gestures, on the other-hand, may not: in a control condition, Axelsson, Churchley, and Horst (2012) found that pointing gestures facilitated fast-mapping, but not

slow-mapping, novel labels. A deeper examination of the effect of social cues across larger timescales of learning, using methods that include diverse retention intervals during exposure to novel labels, would help create a more complete view of the functions of such cues.

Multiple levels of linguistic input in naturalistic settings

Another consideration is that the facilitation of social cues has largely been investigated in conjunction with labeling events in controlled lab experiments (e.g., It's a *dax*). This generally diverges from natural language input to children from their caregivers, which encompasses a variety of utterance types beyond just labeling events, including descriptive comments, imperatives, questions, and routinized interactions (Chang & Deák, 2019; Peters & Yu, 2020; Tamis-LeMonda et al., 2017, 2019). In real, complex input, children need to sift through relevant and irrelevant linguistic input and use social cues in a way that facilitates comprehension. Thus, it is important to consider language input in naturalistic vs. experimental settings, as caregiver and child behaviors may differ between the two (see Tamis-LeMonda et al., 2017 for an example). Frank, Tenenbaum, and Fernald (2013) examined whether social cues assist in disambiguating visual referents in a child's real environment. In contrast to experimental research, they found that these cues did not cleanly delineate a given referent. Perhaps, then, social cues are not actually very helpful in real learning environments. In contrast, Cartmill et al. (2013) found individual differences in the informativity of social cues across play sessions of caregiver-child dyads, where clearer instantiations of social cues were related to more accurate identifications of referential intent. Thus, to truly understand whether and how children capitalize on social cues in their environment, we need more research in naturalistic settings to capture how social cues occur within and across different words, sentence types, and communicative acts.

Individual differences

Finally, theories will need to grapple with individual differences in how children use social cues in service of word learning. Individual differences in how reliably infants follow the gaze of another have been positively related to their vocabulary growth at later ages (Adamson et al., 2004; Delgado et al., 2002; Morales et al., 1998, 2000a; Mundy et al., 2007; Mundy & Gomes, 1998; Smith & Ulvund, 2003; Tomasello & Todd, 1983). Likewise, observing gestures during learning has been shown to positively enhance word learning (Capone & McGregor, 2005; Kapalková, Polišenská, & Süssová, 2016; Rowe et al., 2013). For example, Rowe et al. (2013) found that gestures were most beneficial among learners with less advanced language skills. Thus, investigations into variation in the use of social cues during word learning are necessary to create a clearer picture of early language development. Beyond just how well a given child exploits a social cue, there may be important variation in how different children interpret social cues in the first place. That is, they may differ in how much they interpret a social cue as perceptually- or socially-based, and there may be an advantage to learning sooner in development that social cues provide rich social information, such that children would better encode and retain word-object mappings. This form of variation could explain why children demonstrate varying degrees of learning in experimental tasks (e.g., Axelsson, Churchley, & Horst, 2012).

Furthermore, it will be important to consider variation in the use of social cues across different populations of children. For example, some studies have demonstrated that bilingual children seem to have a heightened sensitivity to social cues when learning new words, relative to monolinguals (Brojde, Ahmed, & Colunga, 2012; Groba et al., 2018; Yow & Markman, 2011; Yow et al., 2017). One idea behind this is that bilingual children may interpret these cues more

readily as conveying social information, yielding an advantage in learning novel words in experiments. That is, given the complexities of dual-language use and/or the prevalence of communication breakdowns, they may need to learn from a young age that social cues can be used to navigate their social environment. This may lead bilingual children to capitalize on learning opportunities afforded by social cues more often than monolingual children. Thus, a complete theory of word learning will need to consider the range of factors that contribute to individual differences in children's interpretation and use of social cues. Going forward, it is necessary to acknowledge that social cues are not monolithic; they offer different information to different children in different circumstances.

1.7 Conclusion

Social cues have been shown to support children's word learning, and the current review aimed to understand why this is the case. We propose that social cues such as eye-gaze and gestures serve at least four different functions: orienting attention, extracting information, disambiguating referents, and indicating referential intent. Similarly, social cues can serve different functional roles: while there may be broad age-related changes in how children use social cues for perceptual and social purposes, there is also evidence that infants younger than 12 months can use social cues both to guide attention and to read relevant social information. Contrary to implicit assumptions in prior studies, we argue that eye-gaze and gesture can, depending on the context, provide mostly perceptual information, mostly social information, or a combination of the two. However, there is considerable complexity left to be unexplained to gain a comprehensive view of children's word learning in a social world. Experimental and observational studies will need to scrutinize how diverse social cues may serve different

purposes for different children's word learning across learning contexts, sentence contexts, and developmental time.

Chapter 2. Speech and social cues combine at discourse boundaries to promote word learning

2.1 Abstract

Social cues, such as eye-gaze and referential gestures, have been shown to promote word learning in children, usually by exposing children to a label-object association with vs. without a social cue. However, social cues are not always so cleanly linked to referents in the world. Instead, label-object associations occur imperfectly across time in a discourse. We conducted an experiment to examine how labels and eye-gaze interactively support 4- and 5-year-olds' word learning during exposures to a series of brief discourse topics. We found that children learned novel words better when social cues and word labels jointly marked the beginning or end of a discourse topic, compared to when they were misaligned. The findings indicate that young learners can use social cues to inform their understanding of word meanings across successive utterances, not just within single labeling events. We discuss how social cues may support children's ability to find where discourse topics begin and end within their ongoing speech input.

2.2 Introduction

Caregivers exhibit a suite of social behaviors during interactions with children, such as shifts in eye-gaze and referential gestures. These cues have been shown to support children's word learning (e.g., Baldwin & Tomasello, 1998). Prior experiments on the role of social cues in word learning have typically involved teaching participants the names of novel objects either with or without supportive social cues. While in-the-moment associations are an important component of word learning, they do not capture the dimensionality of natural word learning contexts (Wojcik, Zettersten, & Benitez, 2022). In particular, they do not speak to the timescale or complexity of natural discourse, which usually includes a series of utterances that focus on a particular topic over time (Frank, Tenenbaum, & Fernald, 2013; Rohde & Frank, 2014; Schulze, Grassmann, & Tomasello, 2013; Sperber & Wilson, 1986). Here, in a simulation of discourse-like input from an adult, our aim was to examine how caregivers' presentation of social cues interacts with the production of object labels to support children's word learning across multiple utterances.

Many studies demonstrate that young children have the ability to track and exploit social cues. Infants as young as three months of age use the eye movements of others to find objects (Caron, Caron, Caldwell & Weiss, 1973; D'Entremont et al., 1997; Gredebäck et al., 2010; Maurer & Salapatek, 1976; Vecera and Johnson, 1995). Slightly older infants (older than 6 months) have been shown to use eye movements to determine referential intent (Baldwin, 1991, 1995; Butler, Caron, & Brooks, 2000; Deligianni, Senju, Gergely, & Csibra, 2011; Senju & Csibra, 2008; Striano & Reid, 2006). The information provided by eye-gaze has also been shown to facilitate language learning, and prior research has focused, in particular, on word learning. Experimental investigations into the functions of social cues have found that the presence (versus

absence) of eye-gaze cues promotes novel word learning (Baldwin et al., 1996; Booth, McGregor, & Rohlfing, 2008; Briganti & Cohen, 2011; Houston-Price, Plunkett, & Duffy, 2006; Hirotani et al., 2009; Nurmsoo & Bloom, 2008; Paulus & Fikkert, 2014; Rader & Zukow-Goldring, 2012; Yu & Ballard, 2007; see also Scofield & Behrend, 2011). Furthermore, the ability to follow eye movements has been linked to growth in vocabulary (Akhtar & Gernsbacher, 2007; Brooks & Meltzoff, 2005, 2008; Carpenter et al., 1998; Morales, Mundy & Rojas, 1998; Morales et al., 2000a; Tomasello & Farrar, 1986). As one example, Brooks and Meltzoff (2008) found that 10- and 11-month-old infants who better exploited adult eye-gaze cues to find nearby objects showed significantly greater vocabulary growth at 24 months of age.

Although social cues have been shown to facilitate word learning, many of these experiments primarily expose learners to simple labeling events that teach associations between objects, labels, and social cues, e.g., hearing “*Dax*” or “It’s a *dax*” with eye-gaze that directs attention to a target object (Bannard & Tomasello, 2012; Booth, McGregor, & Rohlfing, 2008; Briganti & Cohen, 2011; Houston-Price, Plunkett, & Duffy, 2006; Yow et al., 2017). These experiments generally track children’s formation of in-the-moment object-label mappings, but this approach is limited in that word learning in the wild not only occurs in the moment, but on longer timescales as well (McMurray, 2016; Piazza, Nenchewa, & Lew-Williams, 2021).

Virtually any segment of speech in child language corpora reveals that an object label does not occur in a vacuum, but as part of a longer timescale that encompasses more linguistic input. Tamis-LeMonda et al. (2017) found that structured play between caregivers and 13-month-olds elicited, on average, 64 word tokens per minute. For object labels specifically, prior studies have found that parents produce a range of 4-9 labels per minute (Abney, Suanda, Smith, & Yu, 2020; Chang, de Barbaro, & Deák, 2016; Pereira, Smith, & Yu, 2014; Yu, Suanda, & Smith, 2019). If

we take these averages at face value, then there are upward of 50 word tokens per minute that are not object labels.

If a majority of young children's language input is not object labels, then previous research has not directly grappled with how social cues might facilitate word learning in the face of naturalistic complexity. This leaves open questions: How do children track social cues, object labels, and other speech information across sequences of utterances and across time? Are social cues still effective in facilitating word learning beyond experiments with highly simplified presentation of nouns? Experiments that capture the complexities of children's environments are needed to evaluate the role of social cues in supporting word learning.

To that end, we must reconsider the speech that children hear daily, as well as the speech that scientists usually present in experiments. Everyday speech to children tends to occur over time within a natural structure of discourse. Discourse refers to neighboring utterances that focus on a particular topic, one after the next (Frank, Tenenbaum, & Fernald, 2013; Rohde & Frank, 2014), although the timing of switches between topics is variable (Rohde & Frank, 2014). Our investigation is premised on the idea that social cues not only support in-the-moment associations between objects and labels, as demonstrated by previous research, but also guide children's attention during complex, continuous input that supports the segmentation of discourse topics.

Previous research has shown that the structure of discourse shapes children's word learning (Hoff, 2010; Horowitz & Frank, 2015). In a study by Schwab and Lew-Williams (2020), four-year-olds were taught novel object labels either across several related utterances referring to the same object (e.g., "This is a *gazzer*. This [*gazzer*] is small and green."), or across randomly interleaved references to different objects (e.g., "This is a *gazzer*. This [*cheem*] is big and

purple.”). In both conditions, children demonstrated above-chance learning, but learning was less robust when the speaker switched frequently between labeling of different objects. These results suggest that discontinuity in discourse structure hinders the initial formation of object- label associations, while continuity in discourse supports this early step in word learning.

In another prior study, children were shown to use both social cues and discourse-level information when interpreting novel words (Horowitz & Frank, 2015). An experimenter presented two novel objects to children of ages 2-6 years, and then described the objects across three descriptive sentences. The sentences were structured such that the second sentence contained a novel label but the first and third did not; critically, the experimenter was never looking at the target object while producing its label during the second sentence. Three-, 4-, and 5-year-olds (but not 2-year-olds) were able to match the ambiguous label to the object that the experimenter had looked at during the first and third sentence. This suggests that children used the surrounding discourse in tandem with social cues to infer the meaning of an ambiguous label, and that there is developmental growth in discourse comprehension between the ages of 2 and 6 years.

But how do children discover discourse topics in the first place? One idea is that social cues may help carve up continuous linguistic input by guiding attention to each relevant discourse topic, and their occurrence at the beginning and/or end of each topic may support this process of segmentation. That is, children may use social cues to construct the boundaries of discourse structure, as shown in the study by Horowitz and Frank (2015), where both looks and touches to an object marked it as a discourse topic. By signaling the onset or offset of a topic, these cues may scaffold learning by narrowing children’s attention to the most relevant referents

or events. However, the timing of social cues may be important for identifying discourse topics. For example, in Rohde and Frank (2014), points to objects were used more frequently toward the beginning of a discourse topic, perhaps serving a key introductory function. This raises the possibility that children benefit more when social cues are used at the onset (vs. later) in a discourse topic.

Words can also mark discourse topics, but it is unclear how they might do so in tandem with the timing of caregivers' social cues. Must labels and social cues be aligned in time in order to create clear shifts between topics? In Horowitz and Frank (2015), social cues did not need to be aligned with labeling events in order to assist in children's inference of referential labels. However, this finding stands in contrast to recent findings from observational studies of natural caregiver-child interaction, where labeling events are often accompanied by multisensory social cues that scaffold the problem of identifying discourse structure (Cartmill et al., 2013; Pereira, Smith, & Yu, 2013; Suanda, Smith, & Yu, 2018). Experiment designs have perhaps exaggerated the prevalence of labeling events, but even so, they may be helpful in providing optimal moments for learning, particularly in combination with simultaneous multisensory cues. As an example of this, Pereira, Smith, and Yu (2014) recorded toddlers' visual experience during a play session with caregivers and found that toddlers experienced different visual perspectives depending on whether or not the caregiver was producing a labeling event. When a labeling event occurred, visual information was cleaner and contained less visual competition between objects. Such moments of clarity may be instrumental when children are learning words from complex input. Thus, it is currently unknown whether social cues and labeling events must be aligned in time or can be misaligned in time in order to facilitate word learning.

To this end, our goal was to directly examine how the alignment or misalignment of social cues and labeling events supports 4- and 5-year-olds' word learning during brief exposures to successive discourse topics. This age range was selected for three reasons: it matched previous studies on children's word learning during discourse (Horowitz & Frank, 2015; Schwab & Lew-Williams, 2020; Sullivan et al., 2019); it allowed for the exploration of possible age-related differences in learning from discourse; and young children (3-year-olds) were unable to learn any words in a pilot version of the study. Social cues were operationalized as shifts in the adult experimenter's eye-gaze with a simultaneous turn of the head. It is important to note that there are discrepancies in prior research regarding infants' abilities to follow eye movements alone versus the spatial cues provided by head-turns (e.g., Corkum & Moore, 1995; Hoehl, Wahl, & Pauen, 2014; Hoehl et al., 2009; Moore & Corkum, 1998). However, our 4- to 5-year-old participants presumably possessed the ability to follow both to find a referent in the world (e.g., Baron-Cohen & Cross, 1992; Doherty & Anderson, 1999; Doherty, Anderson, & Howieson, 2009). For example, Doherty and Anderson (1999) found that 4-year-olds can follow an adult's gaze direction with or without simultaneous head-turn cues. Hereafter, we refer to eye-gaze as the social cue available to children in the experiment, although they had information available from both eye-gaze and head-turns.

Our novel word-learning task was conducted using tablets in families' homes. On each trial, a video was presented of a centrally-located female, who labeled and directed her gaze to two novel objects, one at a time. Brief descriptions of each object were also presented, each including one labeling utterance and two neutral filler sentences, none of which could be used to disambiguate the referent. This helped to isolate the effect of social cues in marking discourse boundaries. The timing of labeling and the presentation of eye-gaze cues were manipulated in

four within-subjects conditions in order to determine how different combinations would affect children's novel-label learning. This design further allowed us to evaluate whether children can use markers at the beginning or end of a discourse topic to help determine its boundary. In each test phase, children were tested on their learning of the two labels in a forced-choice task.

Our experiment was designed to examine how social cues support children's learning of object labels depending on 1) when social cues are presented in a discourse, and 2) whether or not social cues align with labels. We hypothesized that social cues support children's word learning in two situations: when social cues aligned with word labels, and when social cues preceded word labels. That is, we predicted that children would learn words successfully when social cues preceded labeling events or occurred at the same time as labeling events (similar to Horowitz & Frank, 2015). However, another possibility is that social cues and word labels must jointly mark a discourse topic in order for children to learn successfully, such that learning would not be detectable when social cues and labels are not aligned in time.

2.3 Method

Participants

We conducted a power analysis to determine the minimum sample size to detect an effect of social cues, label position, and an interaction between these two factors. Using simulations of pilot data ($n = 8$), it was determined that 50 participants were needed, and our goal was to recruit 25 4-year-olds and 25 5-year-olds. To meet this sample size, a total of 63 full-term 4- and 5-year-old participants were recruited to participate in the study. Children had no known vision or hearing impairments and were from monolingual English-speaking backgrounds ($> 85\%$ English exposure). Data from 13 participants were excluded: 7 participants were excluded because of

technical difficulties, and 6 participants were excluded for failing the practice phase. After exclusion, we retained a final sample of 50 participants (29 females, $M = 4;8$ months, $SD = 6.39$). Of these participants, 25 reported being White, 3 Asian, 2 Black, and 1 Hispanic (the remaining 19 participants did not report their ethnicity). Based on the information collected at the time of recruitment, all caregivers of this sample graduated high school, and most had completed college ($n = 15$; 37%) or graduate school ($n = 25$; 61%). The Princeton Institutional Review Board approved this research protocol (Language Learning: Sounds, Words, and Grammar; IRB record number 0000007117), and a legal guardian provided informed consent for each child. Participants each completed four within-subjects conditions.

Experimental design

Practice phase. The tablet-based experiment was created using jsPsych (de Leeuw, 2015) and implemented using Pavlovia (<https://pavlovia.org>). The experiment began with a practice phase that consisted of two trials. Each trial asked participants to select one of two familiar objects (“*Can you touch the apple?*”), with order counterbalanced. This portion was designed to acclimate the participants to the experiment, and ensure they were able to select objects on both sides of the screen. Participants were excluded if they did not respond correctly to both questions during the practice phase.

Learning phase. On each trial during the learning phase, participants were exposed to a centrally-located female speaker and two novel objects (one on either side of the speaker). Each trial began with a video of the female looking at the child and smiling. In the video, the speaker then turned to gaze at each object, one at a time (see Figure 2.1), and she said three sentences about each object (with a brief 2-second pause between each set of sentences). Each learning trial

pertained to one novel word exposure. For each, she produced one sentence containing a novel label (e.g., “*Look at the blicket*”) and two neutral filler sentences (e.g., “*I like the color*”). This experiment used a 2x2 design that manipulated the timing of the gaze cue and the labeling sentence. The gaze cue was presented either first or last in the discourse and this cue either did or did not co-occur with the labeling sentence, resulting in a total of four within-subjects conditions (see Figure 2.2). Whenever eye-gaze was not directed at an object, the speaker looked centrally toward the child. In the Gaze-and-Label First condition, the eye-gaze cue co-occurred with the labeling sentence, and this occurred as the first of three sentences in the discourse topic. In the Gaze-and-Label Last condition, the eye-gaze cue co-occurred with a labeling sentence, and this occurred as the last of three sentences in the discourse topic. In the Gaze First, Label Last condition, the eye-gaze cue was still seen at the beginning of the discourse topic, but did not co-occur with a labeling event – which appeared as the last of three sentences. Instead, the eye-gaze cue was presented simultaneously with the first filler sentence. In the Gaze Last, Label First condition, the eye-gaze cue was seen at the end of the discourse topic, and the labeling sentence occurred first. Children participated in all four conditions, and they were taught four novel words within each (two at a time), for a total of 16 novel word-object pairs across the whole experiment. Conditions were presented in a randomized order.



Figure 2.1. Example of stimuli used in the learning phase. A speaker turned her head to label each object in sequence. These labeling utterances were embedded in brief, content-neutral discourses about each object (e.g., “*I like the shape*”).

		Gaze-label Cooccurrence	
		Yes	No
Position in Discourse	First	 1 Look at the blicket. 2 I like the color. 3 It is very awesome.	 1 It is very awesome. 2 I like the color. 3 Look at the blicket.
		 1 Look at the toma. 2 I like the size. 3 It is so cute.	 1 It is so cute. 2 I like the size. 3 Look at the toma.
	Last	 1 It is very awesome. 2 I like the color. 3 Look at the blicket.	 1 Look at the blicket. 2 I like the color. 3 It is very awesome.
		 1 It is so cute. 2 I like the size. 3 Look at the toma.	 1 Look at the toma. 2 I like the size. 3 It is so cute.

Figure 2.2. The four within-subject learning conditions presented in the experiment. We manipulated whether a social cue was presented first or last in a discourse, and whether the cue co-occurred with a labeling sentence. The eye icon represents the presentation of a social cue, and the highlighted text represents a speaker producing an object label. Learning trials in the Gaze-and-Label First and Gaze-and-Label Last conditions presented social cues and labeling utterances simultaneously (first or last in a discourse, respectively). Learning trials in the Label-First Gaze-Last condition presented a labeling utterance first and a social cue last in a discourse, while the Gaze-First Label-Last condition presented these in the reverse order.

Test phase. Following the presentation of each set of six sentences (three in reference to each object), participants were immediately tested on their knowledge of the two novel objects. They were asked to “*Touch the [novel label]*” and select one of the two objects seen during the learning phase. Each object served as the target twice, making a grand total of 32 test trials. The side presentation of objects and test trial orders was randomly generated. Participants could only respond to the test sentence after the noun was completed, and they had an unlimited amount of time to respond. Average response time was 2.79 sec (SD = 4.15), and 15 trials (0.9% of total trials) were greater than 20 seconds; however, analyses of response times were not preregistered and no trials were excluded based on slow responses. After every four test trials, positive feedback was presented to the participant (an image of a gold star with a speaker saying “*Good job*”).

Participants were assigned to one of four counterbalanced lists designed to account for spurious associations between novel objects and labels. Two lists were generated that randomly matched novel labels and object images. As objects were presented in pairs during learning trials,

two additional lists were created based on the randomly generated lists, which reversed the object-label mappings within pairs of objects.

Stimuli. During each of the two practice trials, two images were presented (e.g., apple and puppy) along with a simple sentence that asked the child to select a target (e.g., “*Can you touch the apple?*”). All auditory stimuli were recorded by a native English-speaking female using child-directed speech.

During the learning phase, a series of images of novel objects were presented, two at a time, and positioned to the left and right of a video of a female actor. Each video was 14.0 seconds in length. A total of 16 images of novel objects and 16 novel labels were presented across the learning trials. All words were bisyllabic (*blicket, toma, gazzar, numo, fopin, nissle, chamu, dakup, vugim, paegu, klefu, jimo, reko, tentle, miffle, and zibo*). For each ‘discourse topic’ – three sentences about one object – the speaker uttered one labeling sentence (e.g., “*Look at the blicket*”) and two neutral filler sentences that did not provide disambiguating information about the referent (e.g., “*I like the shape*”). The ordering of labeling and filler sentences depended on experimental condition. In half of the conditions, labeling sentences were heard first, followed by filler sentences. In the rest, the order was reversed.

Test trial stimuli consisted of novel labels recorded within the sentence frame, “*Can you touch the [novel label]?*”. Each sentence was normalized for duration in Praat by splicing each file between the sentence frame (“*Can you touch the*”) and the novel label (Boersma, 2001). Each resulting audio file was of the same duration: the sentence frame was 1175 ms and the novel label was 687 ms. Following each set of test trials, positive feedback was given to the participant by showing them an image of a gold star in the center of the screen. All images seen were 400 x 400 pixels in size and presented against a white background.

Procedure. Children participated in this study from their homes on either a smartphone or tablet, and an experimenter interacted with the child and caregiver via Zoom. The experimenter greeted the caregiver and child and explained that the game was about learning new words. The caregiver was encouraged to observe the child's progress to ensure that they stayed on task but was also instructed to not interfere with the child's touches. The experimenter turned off her video and microphone for the duration of the experiment in order to reduce the possibility of distractions. Participants had to press a start button to begin the experiment. The rest of the experiment proceeded automatically, except for the start of the learning trials where the videos began only after a participant touched an image on the screen. Families received a \$10 Amazon gift card for participating. Materials are publicly available at https://osf.io/ke69t/?view_only=19a216425ac74884b26991da4d091b36.

2.4 Results

Our preregistered analysis plan is available at https://aspredicted.org/135_PZK. Data and analyses are publicly available at https://osf.io/ke69t/?view_only=19a216425ac74884b26991da4d091b36. All analyses were conducted using R version 4.0.2.

Per our preregistration, our first analysis examined whether there were differences in accuracy between participants who experienced different counterbalanced pairings of novel objects and novel words. Using a one-way ANOVA, we found that there were no statistically significant differences in accuracy between the groups, $F(3, 1596) = 1.55, p = 0.20$. Thus, we did not include this as a fixed effect in subsequent analyses.

Our main analysis was a logistic mixed-effects model predicting accuracy during the test phase, using the R package lme4 (version 1.1-23). Our model included fixed effects of co-occurrence (whether the social cue aligned with a labeling utterance, 1 or -1), position in a discourse (whether a social cue occurred first or last in a discourse, 1 or -1), and their interaction. A fixed effect of (continuous) age was also included. In addition, the model also had a random intercept of participant, and random slopes of social cue and position.

For our main effects, we found a main effect of age, indicating that older participants were more likely to produce a correct response, $\beta = 0.04$, 95% CI = [0.001, 0.07], OR = 1.04, $z = 2.07$, $p = 0.04$. Critically, we also found a main effect of co-occurrence, suggesting that participants demonstrated higher accuracy during the test phase when social cues did (vs. did not) co-occur with a labeling utterance, $\beta = 0.31$, 95% CI = [0.11, 0.52], OR = 1.36, $z = 3.07$, $p = 0.002$. We did not observe a main effect of position, $\beta = -0.09$, 95% CI = [-0.31, 0.12], OR = 0.91, $z = -0.98$, $p = 0.37$. This indicates that there were no statistical differences in learning the novel words when social cues were presented first or last in a discourse. We did not find a significant interaction between co-occurrence and position, $\beta = -0.09$, 95% CI = [-0.21, 0.02], OR = 0.91, $z = -1.64$, $p = 0.10$. This suggests that overall, participants learned the novel words better when social cues and labeling utterances were presented simultaneously, and there were no statistical differences in learning depending on whether they marked the beginning or end of a discourse topic. That is, when social cues and labeling events were presented together at the boundary of a discourse topic, children were more successful at learning the two novel words. When social cues did not co-occur with labeling events, children's word learning was not significantly different from chance, $t(799) = 0.64$, $p = 0.52$.¹

¹ In an exploratory analysis, we replicated this model structure but with exclusion of response times greater than 3 standard deviations above the mean (33 trials). This model replicated the results from our preregistered analysis.

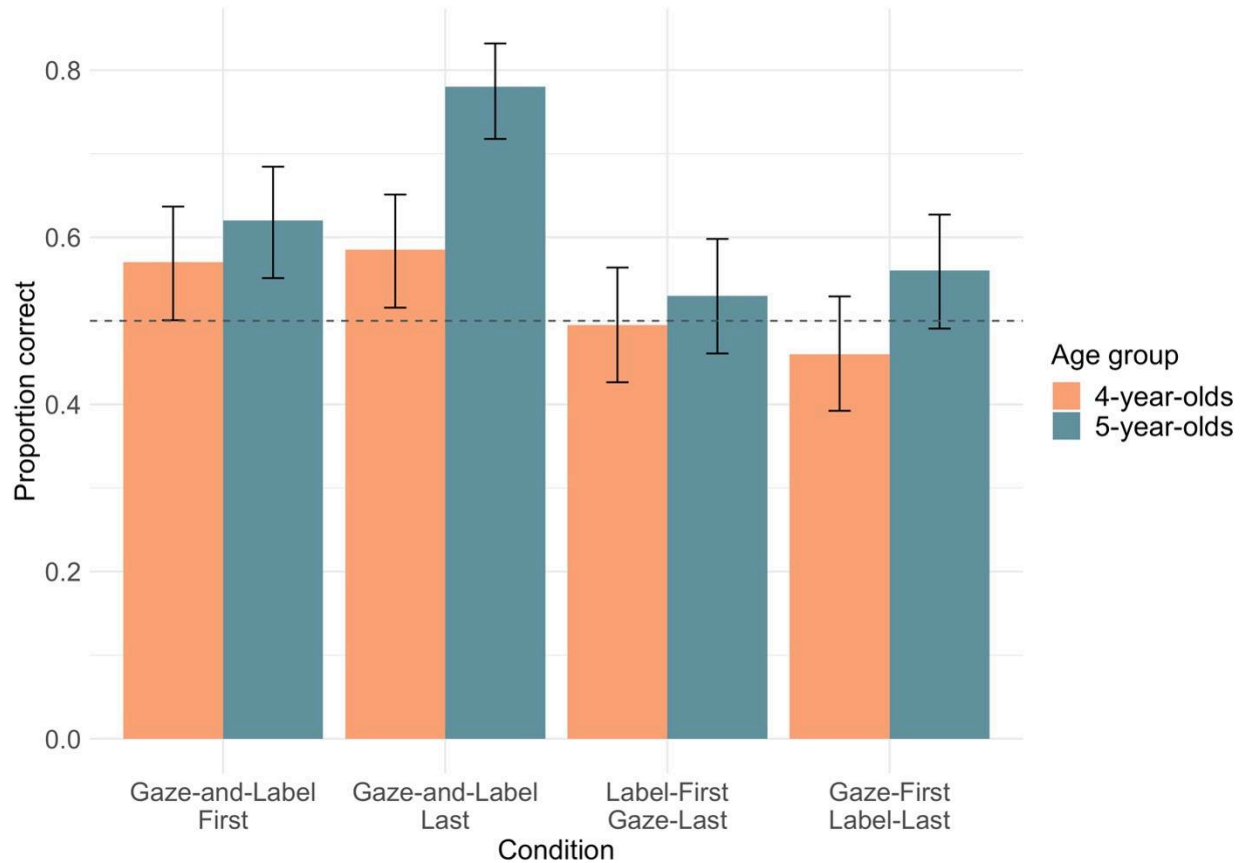


Figure 2.3. Proportion of correct responses by condition and by age group. Chance is shown by the dashed line (50%). Error bars represent 95% confidence intervals.

Due to 5-year-olds' overall greater accuracy on test trials, we conducted an exploratory analysis to further examine possible age-related effects. We used the same variables described previously – co-occurrence, position, and age – but examined them as a three-way interaction . Note that the sample size for this study was not determined with the intent of conducting this three-way interaction, and a larger sample size would be needed in future research to evaluate its robustness. In the model, we replicated all prior effects and found a significant three-way interaction between co-occurrence, position, and age, $\beta = -0.02$, 95% CI = $[-0.04, -0.002]$, OR = 0.98, $z = -2.15$, $p = 0.03$. To further understand why the three-way interaction was significant, we

conducted simple slope analyses and found a trending two-way interaction between co-occurrence and age, but only when social cues were presented last, $\beta = 0.07$, 95% CI = [-0.007, 1.37], OR = 1.07, $z = 1.78$, $p = 0.08$. When they were presented first, this interaction did not trend toward significance. To further understand this interaction, we focused only on trials where social cues occurred last during the learning phase, and examine whether the simple slope of age was significant when social cues did vs. did not co-occur with labels. In this subset of the data, we found the simple slope of age to be significant only when social cues co-occurred with labels, $\beta = 0.08$, 95% CI = [0.03, 0.13], OR = 1.08, $z = 3.00$, $p = 0.003$. This suggests that the original three-way interaction was driven mainly by the older children in our sample: 5-year-old children showed robust learning when social cues and labels co-occurred in the last sentence of a discourse topic. Together, the results of the experiment support the overall hypothesis that participants can use social cues to learn novel words in the context of multi-utterance discourse.

Given the within-subject design, participants experienced the four conditions in different orders, and the order of presentation could have contributed to performance during the test phase. To examine this possibility, we constructed the same logistic mixed-effects model, again predicting accuracy on each trial, but additionally included a predictor of the order of conditions (e.g., Gaze-and-Label First was seen in block 2, etc.). We found that, with this model, order did not significantly predict performance on test trials, $\beta = 0.14$, 95% CI = [-0.009, 0.29], OR = 1.15, $z = 0.07$, $p > 0.05$. Furthermore, the addition of this predictor did not significantly improve model fit in comparison to our preregistered analysis, $X^2 = 3.41$, $p > 0.05$. This suggests that there were no learning-related advantages to the order in which the four conditions occurred. That is, children did not necessarily benefit from experiencing particular cue combinations before others.

2.5 Discussion

Previous experiments have found that the presentation of social cues facilitates young children's word learning (Baldwin et al., 1996; Booth, McGregor, & Rohlfing, 2008; Briganti & Cohen, 2011; Houston-Price, Plunkett, & Duffy, 2006; Hirotani et al., 2009; Nurmsoo & Bloom, 2008; Paulus & Fikkert, 2014; Rader & Zukow-Goldring, 2012; Yu & Ballard, 2007). However, these experiments have mainly investigated in-the-moment associations between labels and objects, and have not focused on discourse-level input from adults. We addressed this by examining how the presentation of social cues and labeling events influences word learning across multi-utterance discourse that simulates natural learning environments. Social cues here were operationalized as simultaneous eye-gaze shifts and head-turns toward an object. We found that 4- to 5-year-old children learned better when social cues were presented concurrently with labeling events. There were no statistical differences in learning depending on whether their converging information was provided at the beginning or end of a discourse topic. We also found, in an exploratory analysis, that 5-year-old children, compared to 4-year-old children, learned the novel words better when social cues and labeling events were presented last in a discourse. Overall, this suggests that children can use information conveyed by social cues – in combination with object labels – to mark discourse topics and, in turn, to facilitate word learning.

Many studies have shown that children are adept at integrating social and linguistic information to learn new words (Bannard & Tomasello, 2012; Booth, McGregor, & Rohlfing, 2008; Briganti & Cohen, 2011; Houston-Price, Plunkett, & Duffy, 2006; Yow et al., 2017). We extend this prior research in two ways: by examining multi-sentence discourse and by testing how social cues and labeling events, both separately and in combination, support word learning. We found that social cues facilitated learning when they co-occurred with labeling events either

at the beginning and end of a discourse topic. This suggests that children may use social cues to segment incoming speech input into different discourse topics, but only when they also occur with the presentation of a novel word. Importantly, in this experimental context, social cues alone were not sufficient markers of discourse topics, regardless of whether they were presented at the beginning or end of a discourse topic, to support word learning. This was also the case for labels alone. This suggests that clear markers – social cues and labels together – are necessary for children to segment and track discourse topics. This converges with research on the impacts of visual (un)clarity on word learning. For example, Cartmill et al. (2013) examined the referential transparency of the environments of 14- to 18-month-olds in order to determine how easily a word meaning can be inferred from the surrounding visual environment. They found great variability in referential transparency during natural caregiver-child interactions, but higher adult ratings of referential transparency were linked to children having larger vocabularies at four years of age. Our findings extend this work by showing that word labels and social cues together create clear markers in continuous input and facilitate learning.

Why were simultaneous social cues and labeling events equally helpful to children at the beginning or end of a discourse topic? This result was unexpected, and it stands in contrast with prior research showing that converging cues at the beginning of a topic are the most helpful. Rohde and Frank (2014), in a study of natural cues that mark discourse boundaries in child-directed speech, found that caregiver pointing occurred more often toward the beginning of a discourse topic, and then gradually decreased across time. This finding, although focused on gesture as opposed to gaze, suggests that discourse-initial social cues in our study should more effectively support learning relative to discourse-final cues. We found that 5-year-olds (compared to 4-year-olds) demonstrated greater flexibility in learning from these simultaneous cues both at

both the beginning and end of a discourse topic. This may have resulted from their extensive experience following eye-gaze, which allowed them to benefit from eye-gaze cues (in combination with labels) no matter where they occurred. It is possible that younger participants, such as 2- and 3-year-olds, would display a different pattern of results, presumably with great reliance on very clear markers of the onset of discourse topics.

Contrary to the findings, we predicted learning to occur even when social cues were presented before labeling sentences. This prediction was based on the finding in Frank, Tenenbaum, and Fernald (2013) and Rohde and Frank (2015) that caregivers were more likely to produce social cues toward the beginning of discourse topics. We hypothesized that if social cues are used to index a discourse topic, as indicated by previous research, then referring to an object through a social cue would maintain the child's attention on that object throughout the discourse. Any mention of a label thereafter could be assumed to refer to it. However, we found that social cues and labels had to co-occur in order for children to learn words successfully. Does this mean that social cues do not independently serve to index discourse topics? We suggest that this may not be the case for several reasons. First, task difficulty was fairly high. Participants were exposed to 16 unique novel words in just 7 minutes, which is not the typical rate of new words in natural settings. Second, although social cues do support joint attention in subsequent moments (Baldwin et al., 1996; Houston-Price, Plunkett, & Duffy, 2006), participants in our experiment had to maintain information in their working memory and – in the Gaze First, Label Last condition – integrate misaligned cues across time to arrive at the correct word-label mapping. Finally, in our exploratory analyses, we observed some hints that 5-year-olds could use social cues to index discourse topics when these cues were presented prior to a label. In this condition, their accuracy during test trials was not significant, but they had an average accuracy of 56%

correct, with a 95% confidence interval of 49-63% (chance = 50%). Indeed, this trend aligns with Horowitz and Frank (2015). In this experiment, children were presented with misaligned social cues and labeling events. Specifically, eye-gaze and touch were presented along with non-labeling linguistic input, and then a labeling event was presented after. Three- to 5-year-old children showed evidence of word learning. One possible reason behind this difference from our experiment is that they provided two instances of social cues, while we provided only one. Taken together, it may still be the case that children can use eye-gaze to guide word learning even if labels are presented several sentences later.

But why did 5-year-olds demonstrate greater learning than 4-year-olds when social cues and labeling events co-occurred last in a discourse? We provide three possible reasons. The first concerns differences in their ability to both follow eye-gaze and sustain attention to the referred-to object. Children have been shown to follow eye-gaze from an early age and continue to develop this skill over time (Brooks & Meltzoff, 2005; Flom, Lee, & Muir, 2017; Moore, 2008), and this ability is linked to their vocabulary growth (Akhtar & Gernsbacher, 2007; Brooks & Meltzoff, 2005; 2008; Carpenter et al., 1998; Morales, Mundy & Rojas, 1998; Morales et al., 2000a; Tomasello & Farrar, 1986). Researchers have determined more recently that the link between following eye-gaze and learning new words is not necessarily just following eye-gaze, but the ability to follow and sustain attention to a target object (Suarez-Rivera, Smith, & Yu, 2019; Yu, Suanda, & Smith, 2019; Pereira, Smith, & Yu, 2014). Yu, Suanda, and Smith (2019) found that sustained attention was the key component of 9-month-olds' looking behavior that predicted vocabulary at 12 and 15 months. It is an open question of whether this extends to the much older children of the current study, but it is possible that the 5-year-olds in the present experiment were better at sustaining attention in this learning context compared to 4-year-olds.

This is especially possible considering the context of the tablet-based experiment, which was completed at home in an uncontrolled setting with various distractors. Eye-gaze measures will be important for follow-up experiments, so that we could track the dynamics of young children's looks to the two objects – and the experimenter – over time.

A second possible explanation concerns individual differences in working memory. Working memory is a mechanism that enables individuals to hold information in their mind for later processing (Baddeley, 2003). This ability undergoes significant development during early childhood years, with large gains occurring between the ages of 2 and 5 years (Garon, Bryson, & Smith, 2008; Hughes & Ensor, 2007). Differences in working memory capacity have been linked to growth in vocabulary development among both typically-developing children (Adams & Gathercole, 2000; Gathercole, 2006; Stokes & Klee, 2009) and children with language impairments (Jackson, Leita, & Claessen, 2016; Vugs et al., 2016). It is possible that the 5-year-olds in our sample had stronger working memory abilities compared to our 4-year-olds, enabling them to maintain the cued object more effectively in memory while perceiving later linguistic information. Unfortunately, we did not assess or control for working memory capacity or other cognitive abilities in the current sample, and future studies will be needed to understand if or how they guide children's word learning from natural discourse.

Finally, the observed three-way interaction may have resulted from age-related differences in discourse comprehension. Children have been shown to detect changes in discourse topics as early as 2 years of age (Allen, 2000; Clancy, 2004), but there is evidence to suggest that there are developmental changes in the ability to comprehend the content of discourse in early childhood. For example, Sullivan and Barner (2016) found significant improvements in children's ability to use surrounding discourse to learn new words between the

ages of 2 and 4. Similarly, Horowitz and Frank (2015) observed differences in the abilities of 4- vs. 5-year-old children to learn novel words from multi-sentence discourse. These findings suggest that the ability to comprehend and learn from discourse is still ongoing during early childhood, and could explain why 5-year-olds in our study were able to learn successfully when social cues and labels co-occurred at the end of a discourse topic. A key feature of the current study is that we examined how social cues are integrated into word learning in discourse-like learning scenarios, at least to a greater extent than in the designs of prior experiments (e.g., Horowitz & Frank, 2015). By doing so, we expand what is currently known about the functions of social cues in word learning. Prior studies have shown that social cues can support in-the-moment word learning and help disambiguate visual (Baldwin et al., 1996; Booth, McGregor, & Rohlfing, 2008; Briganti & Cohen, 2011; Hirotani et al., 2009; Houston-Price, Plunkett, & Duffy, 2006). Here, we specifically manipulated the presentation of social cues and labels. We show that children can learn novel words from social cues during longer timescales, which is crucial given that natural input to children from caregivers is continuous and contains many sentences and events. In natural discourse, like in the present experiment, caregivers in fact tend to focus on one discourse topic for some period of time and then move on to another topic (Rohde & Frank, 2014; Frank, Tenenbaum, & Fernald, 2013). During these longer discussions, social cues can likely be used to mark the beginning or end of a discourse topic in a way that supports efficient processing. That is, the presence of a brief social cue may provide enough information in continuous auditory and visual input to keep the child's attention on what matters, without a speaker having to constantly look at, point to, or touch objects that are under ongoing discussion. This boundary-marking may allow for the child to learn words even if moment-to-moment information in the discourse is noisy. This converges with previous research on how social

information effectively organizes continuous input in a child's learning environment (Ferguson & Lew-Williams, 2016; Marcus, Fernandes, & Johnson, 2007; Werchan & Amso, 2021).

This study also points to important limits in what social cues provide during word learning. We found that novel word learning was more successful when social cues and labeling events co-occurred during exposure, and that learning was limited (or sometimes absent) when social cues and labeling events were misaligned in time. This suggests that, although children do benefit from social cues when learning new words across three-sentence discourse, they cannot reliably learn under imperfect learning conditions, i.e., when social cues and labeling events occur slightly separated in time. We thus present an experimental validation of previous naturalistic research suggesting that brief, highly-informative events – events that provide concurrent labeling information and social cues to reference – likely help children learn object labels (Cartmill et al., 2013; Pereira, Smith, & Yu, 2014; Suanda, Smith, & Yu, 2016). When presented with discourse that approximates the complexity of naturalistic experience, children are better able to learn new words when exposed to converging social cues and labels.

Our design includes some limitations that will need to be addressed in future research. It is important to note that our discourse-like input was not real, natural discourse, such as what one might see in natural play between caregivers and children. Rather, this was a controlled experiment aimed at examining how eye-gaze and speech separately and jointly shape word learning across multiple sentences. A second limitation is that the experiment used neutral filler sentences, which were not readily interpretable as referring to either object, in contrast with prior studies that used sentences with relevant descriptive information. For example, Sullivan and Barner (2016) found that children can use knowledge gained from informative discourse, compared to neutral discourse, to infer novel word meanings. Because we used neutral sentences,

we are unable to determine with certainty that children actually inferred a discourse-like structure – one in which neutral sentences were used in combination with a labeling sentence and/or eye-gaze cue. Neutral sentences were necessary in order to isolate the influence of social cues outside of linguistically-informative sentences within a discourse, but future studies will still need to consider social cues within descriptive, informative discourse. This would complement our findings by providing a window into another form of natural content in caregiver discourse. A promising avenue for future research will be to analyze caregivers' use of social cues during discourse in large-scale natural video corpora of caregiver-child interactions. Annotation of cues such as eye-gaze, referential gestures, and touch – in addition to speech – will enable further understanding of the discourse-related functions of social cues in a way that is grounded in a child's real learning environment.

2.6 Conclusion

Researchers have examined the influence of social cues on word learning for some time and have generally come to the same conclusion: social cues facilitate word learning. However, these investigations have primarily examined in-the-moment associations between social cues, labels, and objects, usually by providing one or more exposures to specific labeling events. Here, we show that children can use social cues to make inferences about word meaning across successive utterances. This suggests that despite some naturalistic research concluding that social cues provide messy information, young learners can utilize these cues in complex learning scenarios by identifying discourse topics. Future research should expand our knowledge of how children integrate social information during learning by investigating how social cues affect

learning in naturally complex learning contexts, with children embedded in environments surrounded by continuous perceptual and social input. Our experiment suggests that children will be able to exploit converging cues to discourse topics to support their learning, even within the natural noisiness of their multimodal input.

Chapter 3. How social cues are presented across a natural discourse by caregivers

3.1 Abstract

A wide range of experimental work has found that the presence of social cues enhances word learning in young toddlers and children. However, these experiments have typically been limited in their generalizability to typical learning scenarios. Thus, we endeavor to determine if this finding may extend to more naturalistic contexts. To examine this, we document the temporal distribution of eye-gaze cues across discourse topics provided by caregivers during caregiver-child play interactions recorded in the family's home. We found that caregivers systematically provide referential eye-gaze cues across discourse topics, such that eye-gaze cues towards the referent of discussion are more likely to occur at the beginning of discourse topics, compared to later on. We also found that this effect varied depending on the utterance: caregivers are likely to pair referential gaze cues with labeling information at any point in a discourse, while such gaze cues paired with any other utterance decreases in likelihood across a discourse. These findings suggest that caregivers present social cues (and labels) in a systematic way to clearly introduce a new topic of discussion. This behavior, in turn, likely scaffolds the multimodal input, increasing the likelihood that they can follow and comprehend changes in discourse and increase the number of learning opportunities present to form word-meaning mappings.

3.2 Introduction

Social cues – social behaviors such as eye-gaze, gestures, and pointing – have been long-studied in the process of language learning. Many studies have found that the presence of such cues supports word learning (Baldwin, 1995; Bloom, 2002; Bruner, 1983; Houston-Price, Plunkett, & Duffy, 2006; Law, Houston-Price, & Loucas, 2012; Lee & Lew-Williams, 2022; Tomasello & Farrar, 1986). However, these experiments are limited in their generalizability to explaining learning in typical environments because they investigate the effect of social cues on word learning at the level of individual utterances. Such studies typically provide only one to a few exposure utterances, while natural linguistic input to a child contains inherent structure and regularities, where multiple utterances are locally relevant to one discourse topic (Frank, Tenenbaum, & Fernald, 2013; Rohde & Frank, 2014; Suanda, Smith, & Yu, 2018). Thus, it remains an open question as to whether the utility of social cues expands beyond the enhancements of learning at an utterance-level in the lab, to the natural environments of children learning language. Here, our aim was to investigate whether caregivers' natural presentation of social cues (in this case, eye-gaze) systematically aligned to the discourse structure of interactions with their children.

Caregiver eye-gaze

One social cue that has been predominantly studied in word learning is an interlocutor's referential eye-gaze, or gaze to an object in a scene or space. Infants demonstrate the ability to follow their caregivers' referential gaze from early on in life. They begin paying attention to the eyes of others as early as 2-3 months of age (Caron, Caron, Caldwell & Weiss, 1973; Maurer &

Salapatek, 1976), and continue to develop this ability in subsequent months (D'Entremont et al., 1997; Gredebäck et al., 2010; Vecera & Johnson, 1995).

This early attentiveness to eyes is important as it is required to follow the gaze of their caregivers and others in their environment. Researchers have argued the reason why following caregiver eye-gaze is important is due to resolving ambiguity. Infants can use their caregivers', or another interlocutors', gaze to resolve their intent to refer to objects or events (Baldwin, 1995; Bloom, 2002; Bruner, 1983; Houston-Price, Plunkett, & Duffy, 2006; Law, Houston-Price, & Loucas, 2012; Lee & Lew-Williams, 2022; Tomasello & Farrar, 1986). They can then infer that the incoming linguistic content (e.g., a label such as *ball*) refers to the object of attention (i.e., the object ball). For example, Law, Houston-Price, and Loucas (2012) tested 18-month-olds in a word-learning task that required them to follow a speaker's gaze to one of two novel objects and learn its novel label. Vocabulary measures were collected at 18, 24, and 30 months of age. They found that 18-month-old children were able to follow eye-gaze to disambiguate a visual referent and learn a novel label, and that performance on this disambiguation task was related to children's vocabulary size. There has also been a wealth of work that has found that following a caregiver's gaze to an object during infancy is largely predictive of later language outcomes in toddler years (Brooks & Meltzoff, 2005, 2008; Carpenter, Nagell & Tomasello, 1998; Çetinçelik et al., 2021; Morales et al., 2000b; Morales, Mundy & Rojas, 1998; Mundy & Gomes, 1998; Mundy, Kasari, Sigman & Ruskin, 1995; Tenenbaum et al., 2014; Young, Merin, Rogers & Ozonoff, 2009).

Another form of caregiver gaze is gaze towards the child. Gaze towards the child has been studied in the context of early learning broadly, particularly in regards to mutual gaze between the interlocutor and child (Csibra & Gergely, 2009, 2011; Fais et al., 2012; Lavelli &

Fogel, 2005; Marno et al., 2016). One notable finding is called the eye-contact effect, whereby eye-contact between two agents modulates the processing of co-occurring and imminently following stimuli (Senju & Johnson, 2009). At birth, newborns demonstrate a looking preference for faces with direct gaze compared to averted gaze (Farroni, Csibra, Simion, & Johnson, 2002). At a slightly older age, infants demonstrate differential processing of objects when presented with faces that demonstrate eye-contact (or mutual gaze), vs. not. For example, 4-month-old infants show enhanced processing of objects that are gazed at compared to objects at which the adult does not gaze (Hoehl, Wahl, & Pauen, 2014).

Mutual gaze also impacts how children control their attention. For example, Niedźwiecka et al., (2018) found that infants who engaged in more mutual gaze with their caregiver at 5 months demonstrated better attentional control at 11 months. Following this, they argue that mutual gaze influences children's attentional network, and may assist in the development of orienting attention. In this way, an infant learns to focus on a caregiver's face, follow their gaze, sustain their attention on a given scene, and then reorient back to the caregiver's face. This research suggests that children can use the eye-gaze of their interactional partner as a source of information to guide their attention to objects or scenes in their environment.

Not only for attentional control, mutual gaze has been argued to indicate communicative intent, even for young infants. Senju & Csibra (2008) presented 6-month-old infants with videos of an experimenter with two toys. The experimenter then presented an action: either an ostensive cue (i.e., gaze at the child) or not (control, head down). Senju & Csibra found that the infants only followed the gaze of the experimenter when mutual gaze was presented, and not during the control action. They argue that this preference to follow gaze only after an ostensive action allows infants to effectively respond to referential communication directed to them. Thus,

caregiver gaze to their children can serve to both facilitate their child's attention as well as indicate that communication is underway.

However, children are not always engaging in mutual gaze with their caregiver. In fact, a number of studies have pointed to the rarity with which children look towards their parents' faces (Abney et al., 2020; Carpenter et al., 1998; Deák, Krasno, Triesch, Lewis, & Sepeta, 2014; de Barbaro, Johnson, Forster, & Deak, 2016; Yu & Smith, 2013; Yu & Smith, 2016, 2017). Abney et al., (2020) found this result as well, but also found evidence that caregiver gaze to the child can still impact word learning. After recording play interactions between caregivers and their children, the researchers found that much of the coordinated joint attention between the dyads in their sample included parental gaze to their child, and that this gaze to their child predicted future vocabulary size. They argue that the prevalence of this cue points to the importance of parental monitoring in word learning, where gazing toward their child allows caregivers to determine highly-informative learning instances such as when to produce a label (Tamis-LeMonda & Bornstein, 1994; Tamis-LeMonda, Kuchirko, Song, 2014; Tomasello & Farrar, 1986). Thus, caregiver gaze to a child can take the form of many functions that may impact word learning (influencing a child's attention, indicating communicative intent, identifying high-impact learning scenarios).

Limitations in prior work examining social cues and word learning

Caregiver eye-gaze has been shown to facilitate novel word learning in a number of developmental psychology experiments (e.g., Baldwin & Tomasello, 1998). However, as noted previously, many of these experiments track only children's formation of *in-the-moment* object-label mappings through minimal exposure or learning trials where object-label

associations are strengthened by the presence or absence of co-occurring social cues. But such clean associations are not entirely ecologically valid when considering a child's typical learning environment (Cartmill et al., 2013; Frank, Tenenbaum, & Fernald, 2013). For example, Frank, Tenenbaum, and Fernald (2013) argue that cues such as referential eye-gaze are noisy in naturalistic settings and do not fully disambiguate potential referents for children. As such, their utility to scaffold the language learning process may be limited. However, as noted in Chapter 2, there are more reasons that warrant further investigation into the utility of social cues in actual learning environments that are bolstered by the artificial setups of experiments. Two of these reasons include: language input is more than labels, and language input occurs across time.

Isolated labels vs natural discourse. Many of the experiments that investigate word-learning enhancements from observing social cues present uncomplicated word learning scenarios. Specifically, many experiments present only labeling information (with or without social cues, Bannard & Tomasello, 2012; Booth, McGregor, & Rohlfing, 2008; Briganti & Cohen, 2011; Houston-Price, Plunkett, & Duffy, 2006; Yow et al., 2017). There are some studies that include additional linguistic content in their exposure trials, but they are still modest in complexity and include 1-3 additional utterances (e.g., Yurovsky, 2018). However, there have been even fewer studies that directly investigate whether the enhancement effect of social cue presentation is still present when linguistic complexity increases (see Chapter 2). It is imperative to determine the generalizability of this effect to more complicated, natural learning environments, because even young children receive complicated linguistic input. For example, Tare, Shatz, & Gilbertson (2008) examined non-object terms in mothers' child-directed speech to children between the ages of one and five. They found that those mothers' speech included a variety of different lexical terms, such as references to color, number and time. This inclusion

was consistent across different play contexts, which the researchers interpreted to mean that children can hear a variety of lexical terms in a variety of different contexts and scenarios. As such, it remains an important endeavor to determine whether there is an added benefit to learning new words from observing social cues. If they are still beneficial, it suggests that language learning must be considered in light of social interactions and information (Akhtar, Carpenter, Tomasello, 1996; Akhtar & Tomasello, 1998, 2000; Baldwin, 1993; Bruner, 1983; Nelson, 1988; Tomasello, 2000, 2001). However, if they are not informative, it suggests that social information is a noisy cue that does not dramatically alter the word learning process.

Language input across discourse. Furthermore, language input occurs dynamically across time. Contemporary research has begun to directly address time-related research questions as it affects language learning theories (McMurray, 2016; Piazza, Nenchewa, Lew-Williams, 2021; Piazza, Cohen, Trach, & Lew-Williams, 2021). In particular, the acknowledgement that communicative input can occur across time and representational hierarchies is an important consideration for children making sense of complicated input that ultimately results in language proficiency. One particular aspect of language that naturally fits into such considerations is the continuous nature of speech, which is set within a natural structure of discourse (Frank, Tenenbaum, & Fernald, 2013; Rohde & Frank, 2014; Hoff, 2010; Horowitz & Frank, 2015; Schawb & Lew-Williams, 2020). Linguistic discourse topics can be structured into utterances that meaningfully relate to one another (Sperber & Wilson, 1986), and can thus create a sequence of utterances over time that are relevant to a higher order discourse topic. There has been some work conducted on how discourse structure may be beneficial to word learning (see Frank, Tenenbaum, and Fernald, 2013; Horowitz & Frank, 2015; Sullivan & Barner, 2016; Sullivan et al., 2019). But can social cues still facilitate learning when considering this longer timescale?

Suanda, Smith, & Yu (2016) argue that to be the case: in their study, recorded interactions between caregivers and their one- to two-year-old children playing with novel objects, and coded for a variety of social cues (on the part of the caregiver and child) as well as discourse structure. They found that longer discourse episodes (average length of four seconds) were associated with greater rates of social cues (e.g., infant haptic touch, parent gestures/action) when comparing both to a baseline chance level and to brief discourse episodes (average length of one second). They suggest that it is with the lens of extended discourse topics that researchers may identify a different pathway of children's language learning, where extended discourse topics structure include a greater amount of social cue information. In this form of regularity, social cues may still be beneficial to word learning.

In addition, we hypothesize that social cues may be beneficial to word learning in a discourse not just because they are correlated with social information, but because they can assist children in carving up the continuous linguistic input into discourse *structure*. For example, social cues may be used by caregivers to signal topic changes across discourse topics. Rohde & Frank (2014), too, examined natural caregiver-child interactions, and found that caregiver pointing decreased over a discourse. This suggests that caregivers may have used this social cue to signal the new topic, and relied on it less as discourse proceeded. In turn, Rohde & Frank (2014) found that child gaze toward and touching the referenced object increased over a discourse. This together indicates that caregivers introduced a new topic of discussion via social cues (i.e., pointing) and children gradually recognize and uptake that new topic. This scaffolding may allow children to take advantage of subsequent learning opportunities presented to them later in a discourse. By expanding the window of consideration from in-the-moment label-gaze associations to associations across a discourse, children can learn from a larger number of

learning moments from in-the-moment label-gaze associations because the referent has already been determined. Follow-up information that is received as the discourse proceeds (the shape, size of an object, further linguistic information) can be used to strengthen a word-meaning relationship.

However, in Chapter 2, we found that both social cues *and* labeling information (e.g., “It’s a *dax*”) were necessary to promote learning of novel object labels in brief discourse structures, suggesting that both sources of information were crucial to carving up continuous utterances into discourse topics. Thus, our goal in this study is to examine the potential utility of social cues in a more naturalistic setting, directly addressing the systematicity of its presentation in light of discourse regularities and its possible co-occurrence with labeling information. In doing so, our work sought to expand upon the results of Rohde & Frank (2014) and Chapter 2. Namely, we sought to determine whether caregivers’ natural productions of social cues (in this case, eye-gaze to referents and the child) present some systematicity that may scaffold the language learning process in the face of more complex learning scenarios, i.e., across discourse topics in caregiver-child interactions, suggested by Rohde & Frank (2014). We also sought to determine whether our findings parallel the results in Chapter 2 (i.e., the alignment of social cues and labels).

We predict that caregivers are more likely to produce referential eye-gaze cues when introducing a new discourse topic, with the probability of using social cues decreasing over time for that discourse topic (Rohde & Frank, 2014). We also hypothesize that social cues are more likely to be produced alongside labeling information (e.g., *That is a dog*), as such moments would likely be highly-informative moments during longer, complex discourse (e.g., Cartmill et al., 2013). These predicted findings would suggest that there is ecological validity to the findings

presented in Chapter 2. This would lend credibility to the interpretation that social cues, such as caregiver eye-gaze, present some amount of utility in the language learning process, even when learning scenarios are more complex.

3.3 Method

The dataset coded and analyzed in this study was originally collected from Kosie & Lew-Williams (2023). The decision to use this dataset was two-fold: 1) it included high-quality video recordings of naturalistic interactions between caregivers and children at the age when social cues are likely to be of greater importance (Tomasello, 2001), and 2) the creation of this study began in March 2020, amidst the Covid-19 pandemic, when in-person data collection was not possible.

Participants

As described in Kosie & Lew-Williams (2023), the dataset includes 44 caregivers (41 mothers, 3 fathers) and their 18- to 24-month old infants ($M = 20.48$, $SD = 1.87$, 21 female). The participants were recruited from Central New Jersey and the surrounding area. All infants were monolingual (i.e., exposed to at least 80% English in their daily lives), full term (born after 37 weeks gestation), and had no reported developmental delays. The infants were predominantly white ($n = 34$ White; $n = 2$ Asian; and $n = 8$ more than one race) and came from higher-SES households. All mothers had graduated high school, and most had completed college ($n = 14$; 32%) or graduate school ($n = 27$; 61%). The Princeton Institutional Review Board approved this research protocol (Language Learning: Sounds, Words, and Grammar; IRB record number 0000007117), and a legal guardian provided informed consent for each child.

Procedure

As described in Kosie & Lew-Williams (2023), caregivers and infants were recorded during 10 minutes of free-play over Zoom. Before joining the session, caregivers were asked to identify a place in their home where they could play with their infant alone, gather some toys to play with, and turn off TV, music, or other distractions. Caregivers also completed the MCDI vocabulary questionnaire prior to their appointment (not analyzed here). At the start of the session, the experimenter joined the Zoom chat, greeted the caregiver, explained the procedure, and helped the caregiver adjust their camera so that both the caregiver and the infant were fully in view (see Figure 3.1). The experimenter informed the caregiver that she would turn off her camera and audio to minimize distractions and would not be watching or listening as the caregiver and infant played together. After answering any questions from the caregiver, the experimenter turned off her video and audio and minimized and muted the Zoom chat. Approximately 10 minutes later, the experimenter returned to the Zoom chat to let the caregiver know that the play session was finished and thanked the family for their time. After participating, families were emailed a \$10 Amazon gift card.



Figure 3.1. Example participants (caregiver and infant) during at-home, free-play session recorded on Zoom.

Coding caregiver eye-gaze and discourse topics

Two trained researchers coded all available videos for caregiver eye-gaze and discourse topics. Videos were coded continuously using ELAN, an open-source software designed for annotating video and audio files (Wittenburg et al., 2006). Researchers coded one dimension at a time, watching the video in its entirety and identifying the onset and offset of each occurrence of eye-gaze (first viewing) and discourse topics (second viewing) based within caregiver speech boundaries (transcriptions annotated from Kosie & Lew-Williams, 2023). The decision to code eye-gaze and discourse topics from within the boundaries of caregiver speech was made to examine the associations that exist between the presentation of eye-gaze and linguistic content.

Caregiver eye-gaze was determined by using the direction of gaze or caregiver head-turns. Caregiver eye-gaze was coded as looks to the Child, Toy, or Away. Looks to the Toy were characterized as looks to the toy of discussion (e.g., when discussing a ball, only looks to the ball would be annotated as “Toy”). However, due to the naturalistic nature of the study, it was sometimes difficult to singularly determine directions of gaze by the caregiver, i.e., if the caregiver was looking only to the Child or only to the Toy. Thus, if looks to the Child and Toy were confounded in direction (e.g., if a child held a toy directly in front of their face), coders were instructed to code both Child and Toy (following Frank, Tenenbaum, and Fernald, 2012). Looks Away were characterized as shifts of gaze during speech boundaries, or when gaze was occluded/off-screen. To evaluate the reliability of the data, a second independent coder produced annotations for three randomly selected videos. We then calculated Cohen’s κ for each direction of looks relevant to our hypotheses (Child or Toy; Cohen, 1960). Reliability ranged from fair to moderate (Child: $\kappa = 0.31$, Toy: $\kappa = 0.435$). This range of reliability is not uncommon when comparing coding of gaze to other naturalistic studies (e.g., Edmunds et al., 2017; Frank, Tenenbaum, and Fernald, 2013). We further detail how this may impact the interpretation of our results in the Discussion.

Caregiver discourse topics were determined based upon speech transcriptions as transcribed by Kosie & Lew-Williams (2023). The goal of coding was to determine the “topic” of conversation, in this case, what toy is being discussed, as well as how to group utterances together into discourse topics. In order to do so, coders used the following principles to establish the topics of utterances: if there was an explicit mention of a noun/toy, if there was reference to that toy by parts, adjectives, or colors, or if there was reference to previously mentioned toys through pronouns. Following Rohde & Frank (2014), coders were instructed to annotate

utterances to a particular discourse sequence to keep discourse topics continuous. Thus, if an irrelevant utterance (e.g., topic A) preceded and followed by an utterance that referred to the same topic (e.g., topic B), that irrelevant utterance could be annotated with the same topic (i.e., topic B). However, utterances could be annotated as “non-topical” when they occurred between utterances that referred to two different topics (e.g., between topics A and B). Reliability was determined by calculating a simple proportion equivalence of annotation assignments, as conducted by related literature on discourse analyses in computational linguistics (Bower & Hedberg, 2010; Rohde & Frank, 2014). The same independent coder produced annotations, and achieved high agreement (81.4% agreement).

3.4 Results

All analyses were performed in RStudio (RStudio Team, 2020) using R Statistical Software (R version 4.0.2 (2022-06-22), R Core Team, 2022). Mixed-effects analyses were performed using the lme4 package (Bates et al., 2015), and *p*-values were calculated using the lmerTest package (Kuznetsova et al., 2017).

3.4.1 *Describing the structure of caregiver discourse*

In a conversation, the structure of discourse varied between caregivers. Firstly, the total number of utterances varied widely across interactions ($M = 119.5$, $SD = 37.7$). The most talkative caregiver produced 201 utterances in a 10-minute interaction, compared to the least talkative caregiver who produced only 37 utterances. Caregiver discourse also varied in number of utterances per topic ($M = 3.7$, $SD = 8.0$, range = [1, 188]). Caregivers also switched frequently between discourse topics, and were fairly short in duration (Figure 3.2). On average,

caregivers switched discourse topics 44.2 times in this 10 minute interaction ($SD = 76.5$). They also lasted, on average, 8.6 seconds ($SD = 25.2$). However, there were significant variations in the structure of discourse across caregivers. For example, Caregiver 17 produced 90 discourse topics, where discourse topics lasted on average 2.2 seconds. And in comparison, Caregiver 26 produced 3 discourse topics, where discourse topics lasted on average 113 seconds.

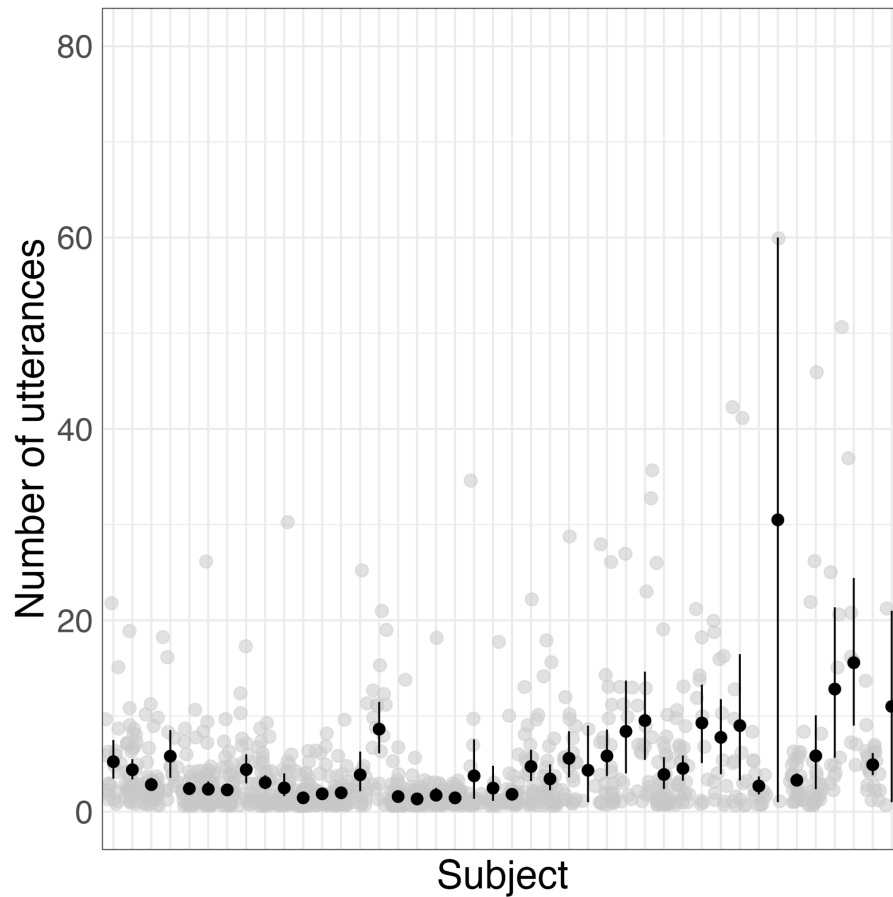


Figure 3.2. A scatterplot depicting the number of utterances produced by each caregiver for each discourse topic. The dark circle represents the average number of labeling utterances per caregiver, and the dark line represents a bootstrapped 95% confidence interval per caregiver.

Labels of referents (e.g., “It’s a dog” or “The dog is here”) were produced fairly minimally. On average, caregivers produced labels once per topic ($M = 1.1$, $SD = 1.6$). However, the number of referent labels varied depending on the discourse topic and by caregiver (see Figure 3.3). For example, Caregiver 6 produced a referent label 25 times for one discourse topic, and as few as one label for other topics. This minimal production of referent labels converges with prior naturalistic investigations into the content of language input from caregivers to their children (Abney, Suanda, Smith, & Yu, 2020; Chang, de Barbaro, & Deák, 2016; Pereira, Smith, & Yu, 2014; Yu, Suanda, & Smith, 2019; Yu et al., 2021).

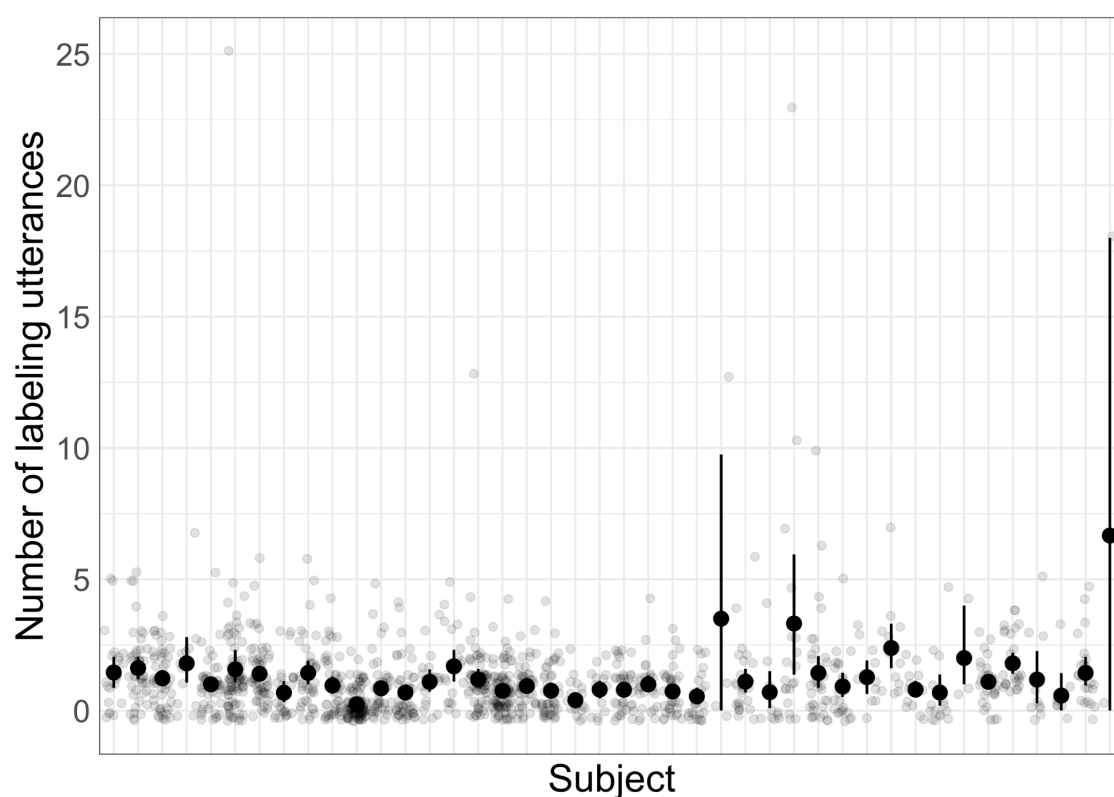


Figure 3.3. A scatterplot depicting the number of labeling utterances per caregiver and per discourse topic. The dark circle represents the average number of labeling utterances per caregiver, and the dark line represents a bootstrapped 95% confidence interval per caregiver.

In addition, there was a relationship between label production and the structure of caregiver discourse. We found that 49% of all first-utterances from caregivers contained labeling information, but for second utterances, this proportion drops to 27.8% (Figure 3.4). This patterning suggests that labeling information corresponds to changes in discourse topics: caregivers can use labels to introduce a new topic of conversation for the child.

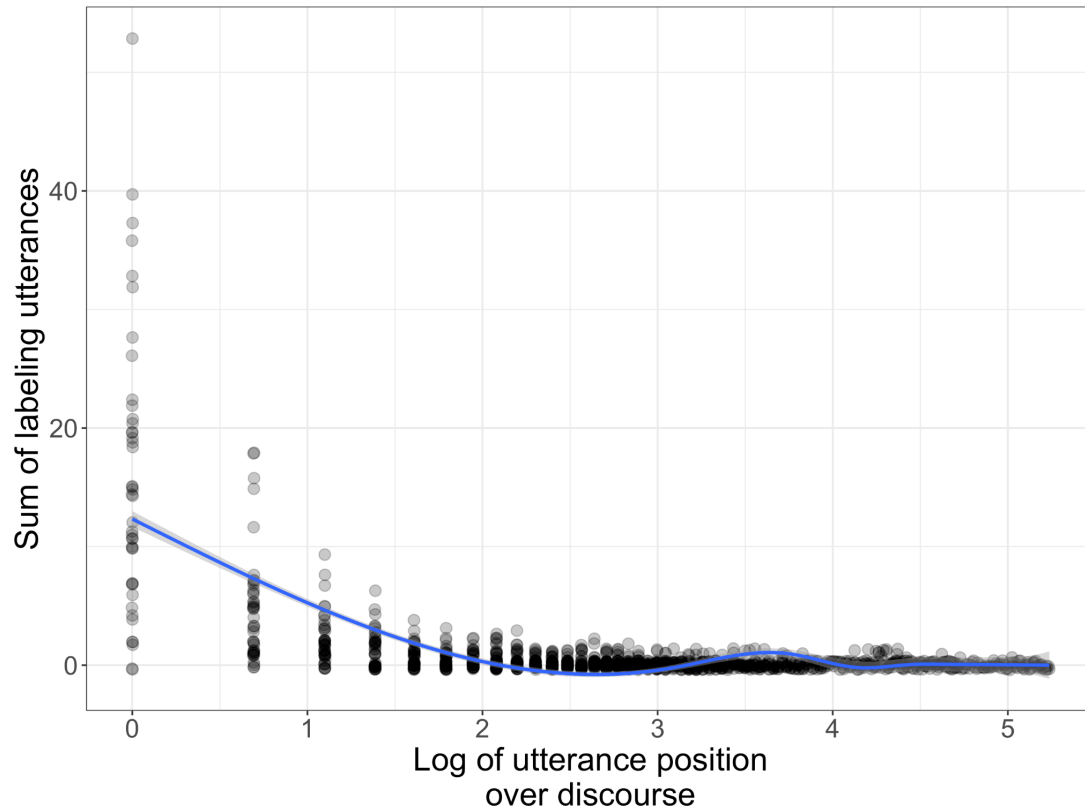


Figure 3.4. A scatterplot depicting the sum of labeling utterances per utterance position (log-scaled) across discourse topics, where each point represents a caregiver's utterance production at each discourse position. The line corresponds to the linear relationship between utterance position and the number of labeling utterances. The shaded intervals represent 95% confidence intervals.

3.4.2 How gaze toward referents and labels are distributed across discourse

One of our main predictions was that caregiver social cues would be systematically presented across a discourse topic, such that caregivers would be more likely to produce eye-gaze cues towards a referent of discussion (i.e., toy being played with) *earlier* in a discourse topic, compared to later. To analyze this, we first replicated the model structure of Rohde & Frank (2014), who found that other caregiver social cues (i.e., mothers' points) were less likely to occur as a discourse structure proceeds. We constructed a logistic mixed-effects model that predicted caregiver gaze to the referent of discussion (1 or 0, at the frame level) from fixed effects of the log of utterance Position in the discourse sequence (centered, numeric), Age of the child (centered, numeric), and an interaction between the two. We log-scaled the fixed effect of Position both to maintain similarity to Rohde & Frank (2014), and because prior works suggested that discourse features may be non-linear in their distribution over time (Dowman et al., 2008; Frank et al., 2013; Rohde & Frank, 2014). We also included a random intercept of discourse topic and subject (as conducted by Rohde & Frank (2014)).

We found converging results to Rohde & Frank (2014): caregiver gaze towards the referent of discussion was less likely to occur as the utterance position in a discourse increases ($\beta = -0.24$, OR = 0.79, 95% CI = [0.77, 0.80], $p < 0.001$). Similar to caregiver points, this finding suggests that caregivers use referential eye-gaze as a way to direct a child's attention when a new object becomes the topic of discourse. Like Rohde & Frank (2014), no other fixed effects were significant (Age or Age x Position). Broadly, this indicates that caregivers' tendency to direct gaze toward a new discourse referent seems to occur regardless of the age of the child.

However, because Rohde & Frank (2014)'s model does not consider how labeling information within discourse structures may affect how caregivers distribute social cue information, we created a new logistic mixed-effects model. We utilize the same model as

previously, but additionally include a fixed effect of Label (whether the co-occurring utterance includes labeling information or not, 1 or 0) and an interaction between Label and utterance Position. Finally, our model's random effects structure remained the same as previously (i.e., random intercept of discourse topic and subject). We attempted to include a random slope of discourse topic by subject, but the model failed to converge.

We replicated our initial findings: we found that the fixed effect of Position significantly predicted the proportion of looks to the toy of discussion ($\beta = -0.33$, OR = 0.72, 95% CI = [0.70, 0.74], $p < 0.001$, Figure 3.5). This suggests that caregivers were more likely to produce eye-gaze towards referents earlier in a discourse, compared to later.

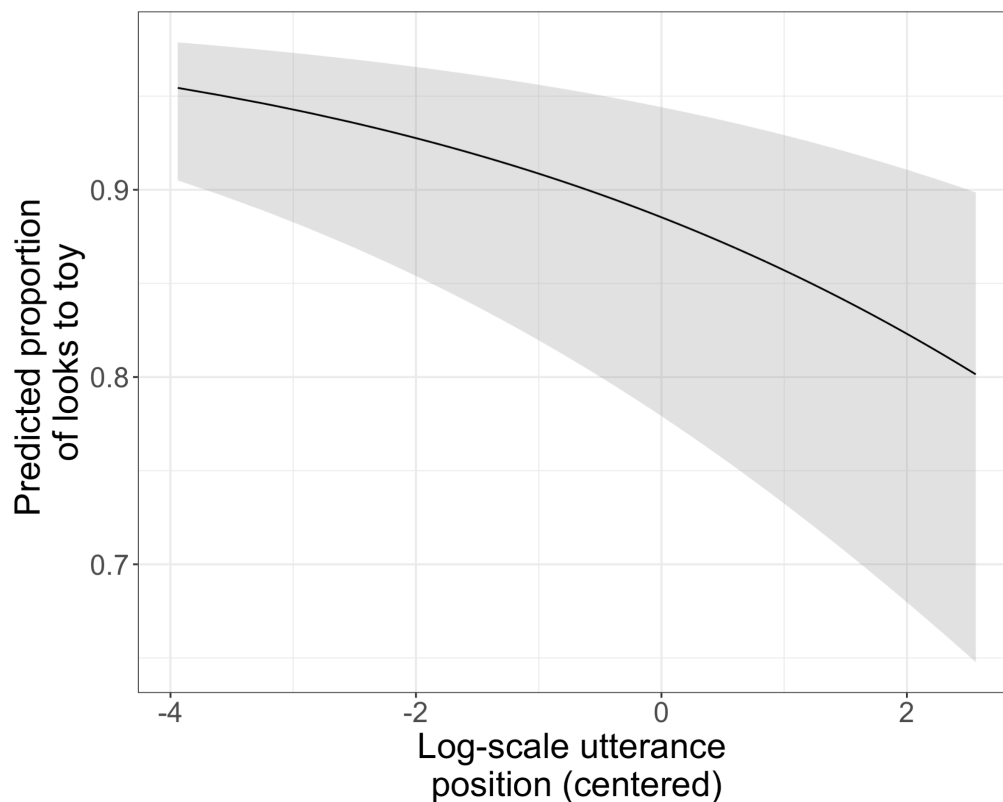


Figure 3.5. A line graph demonstrating the marginal effect of position of an utterance within a discourse (log-scaled) on the proportion of caregiver looks to an object. The shaded portion represents 95% confidence intervals.

We also found a main effect of Label ($\beta = -0.19$, OR = 0.83, 95% CI = [0.79, 0.87], $p < 0.001$), such that caregiver eye-gaze towards referents were less likely to occur with utterances that included labeling information, compared to utterances that did not. This suggests that, broadly, moments of labeling information are not always maximally informative in typical learning environments. However, to provide more clarity to this result, we examined whether gaze towards referents may differ depending on whether a label was produced (Label), as well as *when* it was produced in a discourse (utterance Position). Indeed, we found a significant interaction between Label and utterance Position ($\beta = 0.26$, OR = 1.29, 95% CI = [1.23, 1.35], $p < 0.001$, Figure 3.6). We conducted a follow-up simple slope analysis to determine the relationship between utterance Position and referential gaze for utterances that did and did not contain labeling information. Each of the simple slopes tested (for labels and non-labels) found a significant negative association between utterance Position and referential gaze, but was more strongly negative for non-labels ($\beta = -0.33$, OR = 0.72, 95% CI = [0.70, 0.74], $p < 0.001$) compared to labels ($\beta = -0.07$, OR = 0.93, 95% CI = [0.90, 0.97], $p = 0.001$). This suggests that caregivers' likelihood to present referential gaze cues across a discourse is higher for labels compared to non-labels. Caregivers produce referential gaze cues more readily for non-labels at the beginning of discourse topics, but decrease as discourse proceeds. In comparison, such cues are more likely to be paired with labeling information across a discourse.

This diverging relationship, combined with the prior results, indicates that caregivers combine (non-)labeling and gaze information differently depending on the point of discourse in an interaction. Taken in a broad view, caregivers do not seem to produce labeling and gaze information together, providing further evidence to the idea that a child's language learning

environment is not straightforward and social cues alone do not provide referential clarity (Cartmill et al., 2013; Frank, Tenenbaum & Fernald, 2013). However, they seem to do so when considering the structure of discourse, especially in comparison to non-labeling linguistic content. The caregivers in our sample were more likely to gaze towards referents at the end of discourse topics when producing labeling information, compared to non-labeling information. As such, children may take advantage of this labeling and social cue co-occurrence in service of their learning. Despite the relative infrequency of labeling information, the co-occurrence of gaze information may still strengthen word-object mappings, especially when discourse topics are extended in length and attentional resources may be limited.

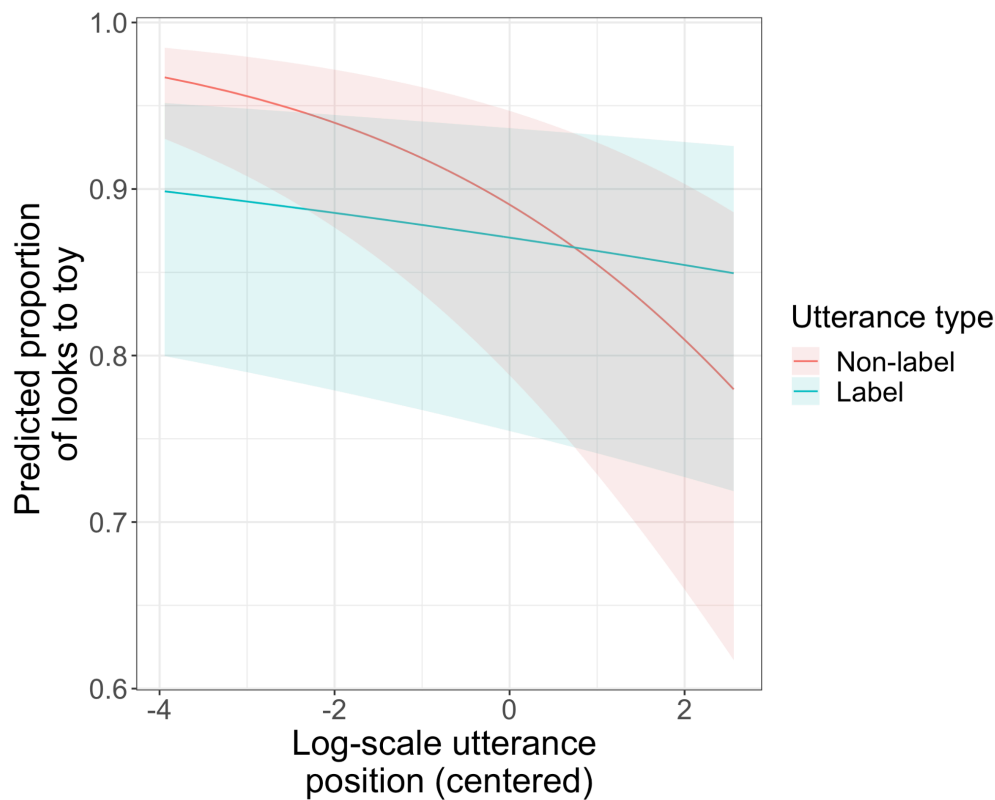


Figure 3.6. A line graph demonstrating the marginal effect of an interaction between position of an utterance within a discourse (log-scaled) and utterance type (label vs. non-label) on the

proportion of caregiver looks to an object. The shaded portion represents 95% confidence intervals.

Finally, we found a significant interaction between utterance Position and Age predicted gaze towards referents ($\beta = -0.03$, OR = 0.97, 95% CI = [0.95, 0.99], $p = 0.002$), but not Age alone. This suggests that the relationship between utterance Position and referential gaze differs depending on the age of the child. We conducted another simple slope analysis to determine the direction of this interaction. We found that caregivers of the oldest children in our sample (24 months, $\beta = -0.39$, OR = 0.68, 95% CI = [0.64, 0.71], $p < 0.001$) had a more negative relationship between utterance position and referential gaze compared to the youngest children (18 months, $\beta = -0.27$, OR = 0.76, 95% CI = [0.73, 0.79], $p < 0.001$). This means that, as discourse proceeds, caregivers of older children are less likely to present referential eye-gaze, compared to caregivers of younger children. This effect suggests that caregivers are sensitive to their children and present social cues to scaffold the changes in discourse topics when their children's knowledge or attentional capabilities may be more fragile due to their young age (Gogate, Maganti, & Laing, 2013; van Schaik et al., 2020). Older children may be able to grasp onto the referential gaze faster or more effectively, while younger children may need several presentations to identify the topic of discourse.

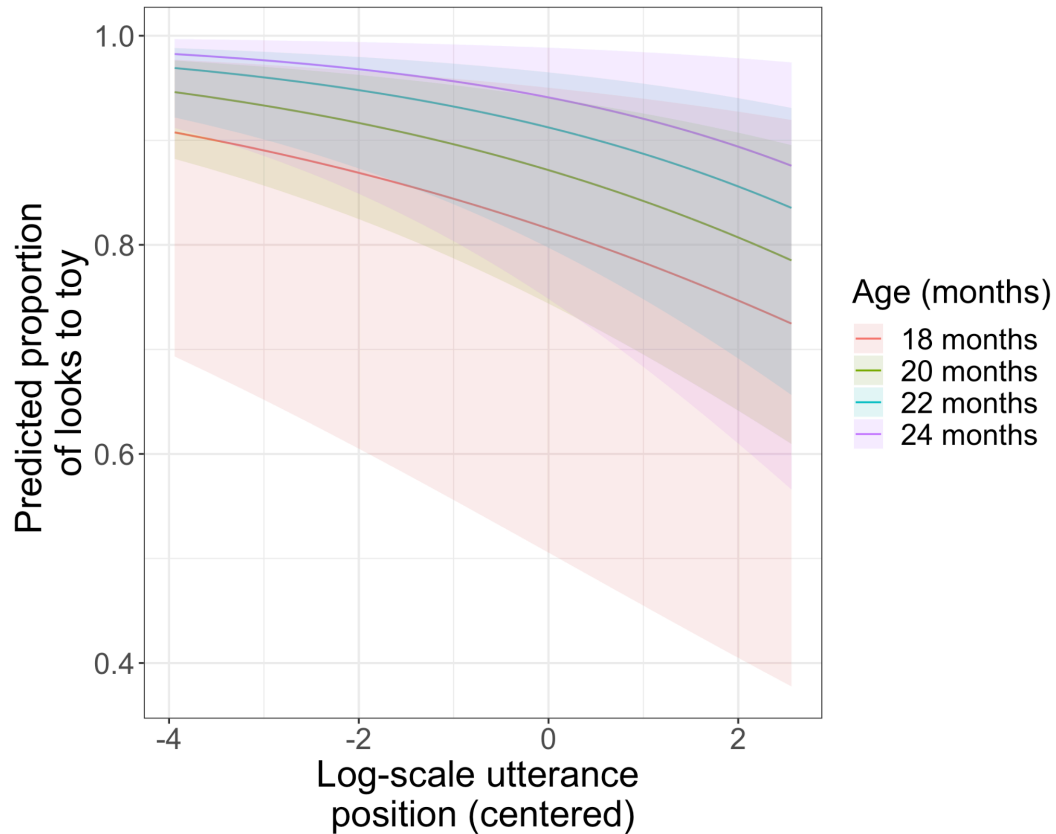


Figure 3.7. A line graph demonstrating the marginal effect of an interaction between position of an utterance within a discourse (log-scaled) and age of the child (18, 20, 22, and 24 months) on the proportion of caregiver looks to an object. The shaded portion represents 95% confidence intervals.

3.4.3 *How gaze toward the child and labels are distributed across discourse*

We also examined how the caregiver looks towards the child may differ across discourse structure. Our hypotheses were not predicted a priori. However, based on the literature, we predicted that caregiver gaze toward the child may increase as a discourse proceeded or remained uniform in an effort on the part of the caregiver to maintain a child's attention (e.g., Niedźwiecka et al., 2018).

To investigate this, we replicated our model structure used in 3.4.2, with the exception of predicting looks to the child (1 or 0, at the frame level). First, we found a significant effect of utterance Position ($\beta = 0.20$, OR = 1.22, 95% CI = [1.19, 1.25], $p < 0.001$), such that caregivers were more likely to look at their child later on in a discourse, compared to earlier. We interpret this as caregivers confirming and maintaining children's attention in a conversation. Second, we found a significant effect of Label ($\beta = 0.13$, OR = 1.14, 95% CI = [1.09, 1.20], $p < 0.001$). Caregivers were more likely to look at their child when producing labeling information, compared to non-labeling information. This behavior also seems to be in an effort to maintain a child's attention. Caregivers may produce pertinent labeling information while looking at their child to confirm they are paying attention, in order to maximize the number of opportunities for learning.

We also found a significant effect of Age ($\beta = -1.54$, OR = 0.21, 95% CI = [0.08, 0.58], $p = 0.003$), such that caregivers were less likely to look towards their child when their child was older. Caregivers may do so for two, related, reasons: (a) caregivers no longer need to look at their child to maintain their attention, and (b) caregivers can use linguistic content to confirm their attention. Taking these findings together, they suggest that caregiver looks to their child are systematically distributed based on three factors: utterance position in a discourse, caregiver label production, and the age of the child. Such looks seem to be in an attempt of caregivers to direct and maintain the attention of their children to the conversation at-hand (i.e., the play interaction).

We also found a significant interaction between utterance Position and Label ($\beta = -0.05$, OR = 0.95, 95% CI = [0.91, 1.00], $p = 0.03$, Figure 3.8), suggesting a different relationship between utterance position in a discourse and gaze to the child depending on the content of the

utterance (label vs. non-label). To further interpret this finding, we conducted a simple slope analysis. Each of the simple slopes tested (for labels and non-labels) found a significant positive association between utterance Position and referential gaze, but the strength of the relationship was greater for non-labels ($\beta = 0.20$, OR = 1.22, 95% CI = [1.19, 1.25], $p < 0.001$) compared to labels ($\beta = 0.15$, OR = 1.16, 95% CI = [1.12, 1.21], $p < 0.001$). Caregivers increase in likelihood of producing gaze to their child as discourse proceeds when utterances include non-labels, compared to labels. This result seems to converge with our prior interpretation: with non-labeling information, caregivers increase their looks to the children as discourse moves on to maintain their attention. With labeling information, caregivers also seem to increase their looks to their children, although to a lesser extent.

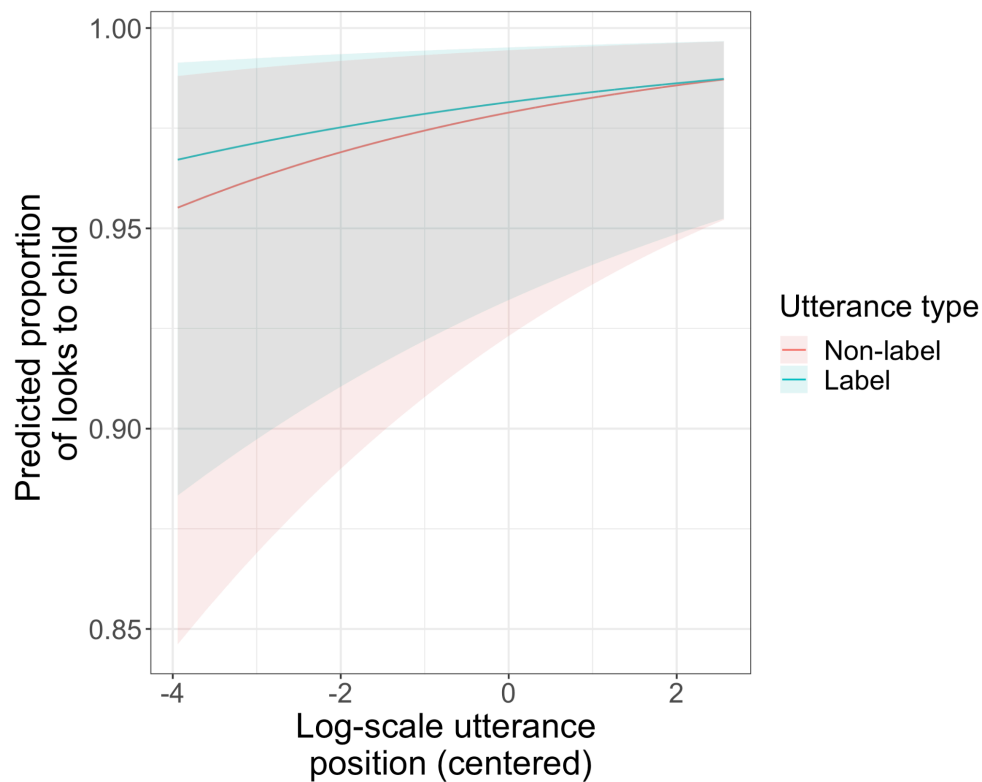


Figure 3.8. A line graph demonstrating the marginal effect of an interaction between position of an utterance within a discourse (log-scaled) and utterance type (label vs. non-label) on the

proportion of caregiver looks to their child. The shaded portion represents 95% confidence intervals.

Finally, we found a significant interaction between utterance Position and Age ($\beta = -0.07$, OR = 0.93, 95% CI = [0.91, 0.95], $p < 0.001$), such that the relationship between utterance position and caregiver gaze to their child differs depending on the age of the child. Conducting a simple slope analysis, we found that caregivers of the youngest ($\beta = 0.31$, OR = 1.36, 95% CI = [1.31, 1.41], $p < 0.001$) and oldest children ($\beta = 0.06$, OR = 1.06, 95% CI = [1.01, 1.11], $p = 0.02$) in our sample showed a significant positive relationship between utterance Position and gaze to their child. At 18 months of age, caregivers increase in the likelihood to look at their child as discourse proceeds. In comparison, at 24 months of age, caregivers increase in their looks to their child across a discourse as well, but to a lesser extent compared to the caregivers of younger children. What this suggests is that the main effect of utterance Position is driven primarily by caregivers of younger children. These caregivers seem to use their gaze towards their young children to confirm and maintain their attention, while caregivers of older children need not to confirm their child's attention as frequently through gaze as a conversation gets longer.

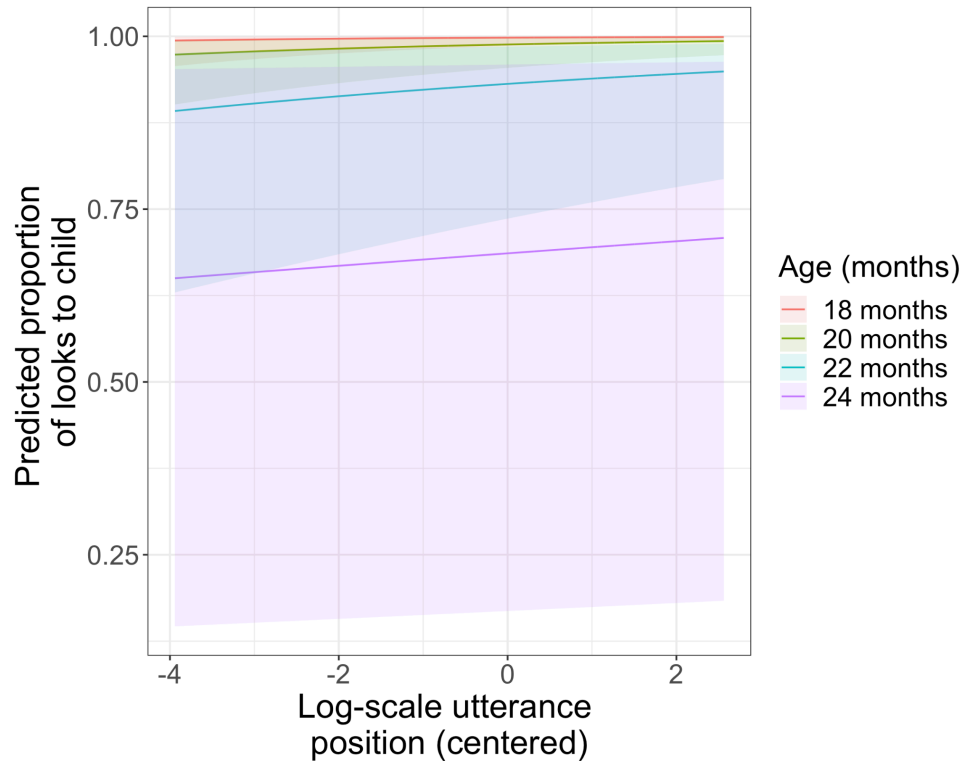


Figure 3.9. A line graph demonstrating the marginal effect of an interaction between position of an utterance within a discourse (log-scaled) and age of the child (18, 20, 22, and 24 months) on the proportion of caregiver looks to their child. The shaded portion represents 95% confidence intervals.

3.5 Discussion

Previous research has argued that social cues may be limited in their utility in service of word learning when in the context of natural learning environments (Cartmill et al., 2013; Frank, Tenenbaum, & Fernald, 2013). However, these experiments have not considered the inherent structure in these complex environments, namely, the hierarchical structure of language (Piazza, Nenchewa, & Lew-Williams, 2021). Utterances naturally fit within sequences of discourse topics within a conversation. Such hierarchy is present even within child-directed speech (Frank,

Tenenbaum, & Fernald, 2013; Rohde & Frank, 2014; Hoff, 2010; Horowitz & Frank, 2015). We address this by investigating whether social cues are differentially presented across discourse topics. We found that, indeed, caregivers present referential gaze cues to the topic of discussion in a systematic way: they are more likely to produce referential gaze when it co-occurs with utterances earlier on in a discourse, compared to later. We also found that the relationship between position in a discourse and referential gaze was different depending on the utterance content. Non-labeling utterances were much less likely to be paired with referential gaze, compared to labeling utterances. This converges with the findings from Chapter 2, where we found that only the co-occurrence between social cues and labels (positioned at the beginning and end of discourse sequences) predicted children's novel word learning. Our current results parallel this: in natural play interactions, caregivers seem to produce referential eye-gaze and labels together, specifically in the beginning of discourse topics, in an attempt for children to segment and track discourse topics. This caregiver preference suggests that children may benefit from this form of caregiver scaffolding, and may show improvements in learning words. Overall, these findings suggest an optimistic perspective of the utility of social cues. Caregivers seem to produce these cues in a sensitive fashion, where these cues are provided to denote broader changes in discourse topics, as well as provide maximally informative moments of learning (in the form of label and social cue co-occurrence).

Chapter 2 reported that consistent social cues are useful in word-learning whether they occur at the beginning or end of a short topic. In this chapter, we find that caregivers' referential gaze to the object of discussion were more likely to be present at the beginning of discourse sequences rather than at the end. This finding lends credence to the interpretation put forth in the Discussion of Chapter 2. Recall that the age range for the experiment in Chapter 2 was 4- and

5-year-olds, while the age sample of the current study was 18- to 24-month-olds. We argue that the older sample in Chapter 2 were better equipped to learn from the presented social cues at both the beginning and end of a discourse topic, stemming from their extensive experience following eye-gaze. In comparison, the caregivers of our younger participants may have been sensitive to their fragile language knowledge and gaze-following, and thus relied primarily on very clear markers of social cues and labels at the onset of discourse topics.

Why did caregivers produce referential gaze significantly more often for non-labeling utterances, compared to labeling utterances? This result was unexpected, as we predicted that caregivers would look towards referents more often for labeling utterances compared to non-labeling utterances. This prediction was based on the findings of Chapter 2, where 4- and 5-year-olds were better able to learn novel words in a discourse when social cues and labels aligned, compared to when they were misaligned. We hypothesized that if that sample was able to learn new words based on this alignment, the same relationship between social cues and labeling information may present itself in natural learning environments. However, that was not the case, and in fact, we found evidence that caregivers were likely to be looking at the child when producing labeling information. We interpret this finding in light of prior work on the richness of information in typical learning scenarios (Cartmill et al., 2013; Frank, Tenenbaum, & Fernald, 2013; Medina et al., 2011). Frank, Tenenbaum, and Fernald (2013) underscored the referential ambiguity still present even when social cues are presented by using signal detection analyses to predict object reference from multiple types of social cues. They found that, broadly, caregiver social cues were noisy and did not fully disambiguate the referential ambiguity present. In the current study, the caregivers in our sample may have produced noisy referential gaze cues, but produced clear gaze toward the child when labeling. This suggests that caregivers of children

between 18-24-months of age have a preference for producing labeling information alongside bids for mutual gaze. Their reliance on this form of gaze information may be in an effort to signal to the child that they are receiving a pedagogical and communicative instance (e.g., Senju & Csibra, 2008).

Interestingly, we found an interaction between labeling utterances and utterance position in predicting gaze. We take this to indicate the importance of considering the inherent structure of information presentation. Our results suggest caregivers utilize social cues systematically depending on both the type of utterance (labeling vs. non-labeling) *and* the position of an utterance in discourse. Not only is the likelihood of producing referential gaze alongside different utterance types (labels vs. non-labels) dependent on the position in a discourse topic, our study also found that caregivers' gaze to their children were also produced systematically over a discourse. Caregivers were more likely to look at their children later on in a discourse, compared to earlier, and this gaze to their children varied depending on what type of utterance was produced (label vs. non-label).

Considering these findings in tandem, this means that in the beginning of a discourse, caregivers can rely on both referential gaze and gaze to their children to provide a learning opportunity and present labeling information. However, as discourse proceeds, caregivers rely on producing referential gaze and labeling information to provide opportunities for their child to learn words. Additionally, they will look towards their child more often to engage their attention to the discourse, rather than to strengthen the word-meaning mapping (as bolstered by the evidence that co-occurrence between looks to child and non-labelling utterances increase as a discourse proceeds). This systematic presentation suggests that caregivers scaffold discourse structure to the ease of comprehension of their children. This can allow young learners to take

advantage of caregiver gaze to better understand both the content and structure of discourse, and inform their word-meaning mappings.

However, this finding seems to be driven primarily by the age of the child. We found that caregivers of older children are less likely to produce referential gaze as discourse proceeds, compared to caregivers of younger children. Contrastingly, caregivers of younger children are more likely to look at their child as discourse moves on, compared to older children. We interpret this finding as caregivers' efforts to scaffold the interaction to the comprehension of their child, particularly for younger children that possess limited working memory ability and less sustained attention. Caregivers of children younger than two may use referential gaze to introduce the discourse topic to their child (Rohde & Frank, 2014) and then look to their child to maintain their attention to receive further linguistic input (via the play interaction) (Csibra & Gergely, 2009; Farroni, Csibra, Simion, & Johnson, 2002; Hoehl et al., 2014; Senju & Csibra, 2008). For example, Csibra & Gergely (2008) demonstrated that even 6-month-old infants used mutual gaze as a cue for communicative input. Similarly, Hoehl et al., (2014) found that 9-month-old infants showed a desynchronization of alpha-band activity when engaging in joint attention with a live interlocutor vs. a video of an interlocutor. This difference in neural activity, particularly the desynchronization, is argued to reflect the engagement of general semantic knowledge (Klimesch, 2012). We interpret that caregivers' tendency to look towards their child at the upcoming ends of discourse topics is to re-engage their children in the discourse embedded in the play interaction. Older children may not have needed as much referential gaze to determine a discourse topic, nor needed as much mutual gaze to maintain sustained attention to the current conversation. Thus, caregivers are sensitive to the limits of their child's attention span, and seek to maintain their child's attention to the discourse at-hand.

What these findings broadly suggest is that caregiver eye-gaze may still hold utility in complex, dynamic learning environments to support children's word learning. While prior experiments have found positive relationships between the presence of social cues and word learning (Bannard & Tomasello, 2012; Booth, McGregor, & Rohlfing, 2008; Briganti & Cohen, 2011; Houston-Price, Plunkett, & Duffy, 2006; Yow et al., 2017), naturalistic investigations have raised doubt on the usefulness of such cues when learning new words in typical, information-dense environments (Cartmill et al., 2013; Frank, Tenenbaum, & Fernald, 2013). Our work suggests that caregivers work to scaffold continuous linguistic input in three ways. One, they present referential eye-gaze cues when discourse topics change. Two, they present referential gaze to the object of discussion when they are producing a label, to increase referential transparency and increase the likelihood that a word's meaning can be inferred. And three, they look toward their children to help maintain their attention to the current discourse, thereby increasing the number of learning opportunities that may be present.

We also found that discourse varied widely within and across caregivers; in the number of topics presented, the number of utterances per topic, and number of labels produced. For example, the grand mean of the number of utterances per discourse topic is 3.4 utterances ($SD = 6.3$). However, one caregiver produced a range of one utterance for a topic, to 22 utterances for a topic. We consider two reasons for this variation. One is the children's own utterances. Children are active conversational partners and can ultimately influence what shape a discourse takes. For example, research on turn-taking has found that children often initiate discourse topics, and their mothers then adapt the conversation to their initiation (Bloom et al., 1996; Chapman, Miller, MacKenzie & Bedrosian, 1981; Hoff-Ginsberg, 1987; Vandam et al., 2022). However, Hoff-Ginsberg (1987) found that children who are more competent speakers are more likely to

initiate conversations, and have their mothers follow, while for children who are less competent speakers, their mothers are more likely to initiate discourse topics. Thus, it is possible that the caregiver-child dyads varied in who initiated topics, and thus changed the length of the discourse. A second possibility is children's shifts in attention. Rather than utterances, discourse switches could have occurred as a result of caregivers tracking children's visual attention to different objects. If a child disengaged in sustained attention from one object and turned to another, caregivers may then switch their discourse to the second object. Our coding scheme allows for two irrelevant utterances to occur within a discourse topic, but children who are less able to control their attention may disengage more frequently. Thus, as caregivers monitor their children's behavior, they may switch discourse topics to focus on the object of the child's attention. There is a breadth of work that suggests that word learning is enhanced when parents name an object that a child has sustained attention for, rather than shifting their gaze to and naming a new object (Abney et al., 2020; Hirsh-Pasek et al., 2015; McCathren, Yoder, & Warren, 1995; Tomasello & Farrar, 1986; Harris, Jones, Brookes, & Grant, 1986; Gogate, Walker-Andrews, Bahrack, 2001; Zukow, 1990).

Limitations and future directions

This study sought to address experimental limitations conducted in previous research concerning the utility of social cues during word learning by implementing a more naturalistic methodology, recording caregiver-child interactions in the home. However, there are still some limitations that should be considered in future research. First, the data collection occurred during the Covid-19 pandemic, so the participant videos were recorded from a laptop via Zoom. This meant we possessed only one angle of recording to view the play interaction. This limited

perspective likely factored into why the coding reliability for caregiver eye-gaze was relatively low. Indeed, one study compared the discernment of gaze targets between a stationary camera, similar to the one implemented here, and a Point-of-View (POV) camera that was embedded into a pair of glasses (Edmunds et al., 2017). This enabled the experimenters to record one person's viewing perspective. They found that the coding reliability derived from PoV cameras were significantly higher than that of stationary cameras, revealing the difficulty to ascertain accurate coding of a live interaction from a 2D video. Other studies, as well, have reported low coding reliability for determining the targets of eye-gaze, with Cohen's Kappa ranging from 0.43 to 0.64 (Edmunds et al., 2017; Frank, Tenenbaum, & Fernald, 2013; Boundy, Cameron-Faulkner, & Theakston, 2016). Furthermore, it is likely that the experimental setting of data collection affected coding of eye-gaze as well. Recordings of caregivers and their children in their home provided the most optimal setting to record natural interactions, it also reduced the amount of experimental control. Kosie & Lew-Williams (2023) were unable to directly manipulate the placement of the camera, how caregivers and children faced the camera, whether the children wandered off-screen, how many toys were present, etc.

What, then, are we to make of the coding annotations for caregiver gaze? Is caregiver eye-gaze actually systematically combined with labeling information, and distributed in an intentional way across a discourse? Frank, Tenenbaum & Fernald (2013) put forth three possibilities in their interpretation of low coding reliability for their own samples' eye-gaze that apply here. One, it is possible that these cues are actually combined with other information and distributed systematically, and thus are likely to inform the word learning process, but they are difficult to code from video and hence there may be errors in the coding analysis. Another is that these cues have actually low utility because they are not reliably referential (in combination with

labeling utterances) and are not systematically distributed across a discourse. And a final explanation that Frank, Tenenbaum, & Fernald (2013) suggest is that the difficulty of coding eye-gaze is reflective of the difficulty in ascertaining one's eye-gaze in real life. They hypothesize that it is actually quite complicated to monitor another's gaze, and so the low reliability of the coding is reflective of determining the target of an eye-gaze in reality. In the context of the difficulty to code eye-gaze broadly, and the reduction of experimental control due to the nature of data collection, it is likely that both options 1 and 3 are applicable to our study. However, it is still likely that caregiver eye-gaze is intentionally paired with labeling information, and distributed across a discourse, given the strength of the predictors in our models. Future research is needed to validate these results, likely using more robust recording methodologies such as head-mounted eye-tracking (Bradshaw et al., 2023; Edmunds et al., 2017; Franchak et al., 2011; Franchak, Kretch, & Adolph, 2018).

Furthermore, the goal of the study here, broadly, was to examine how caregivers produce social cues and linguistic information (labels within a discourse) across time. Thus, we coded information only provided by the caregiver. However, there is a wealth of research suggesting the child's active role in interactions (Foushee, Srinivasan, & Xu, 2023). From infancy, children can selectively attend to input (Gerken et al., 2011), learn from overheard conversations (Floor & Akhtar, 2006), and seek out information to reduce referential ambiguity (Zettersten & Saffran, 2021). It is important to thus consider how children may lead conversations, by examining both their linguistic productions and what social cues they produce to indicate topics of discussion they want to speak about. For example, Chang, de Barbaro, & Deák (2016) found that caregivers' speech was contingent upon infants' gaze, manual actions and vocalizations. Caregivers were also more likely to name objects when infants looked to them or to the object of

discussion. These results suggest an infant-led, caregiver-follow approach to interactions, such that infants and children may come to learn that their exploration of objects, and the environment more generally, comes with information from their caregiver. In contrast, Sun & Yoshida (2022) found that referential input from caregivers co-occurred most often when caregivers and infants engaged in joint attention, suggesting a caregiver-led interaction. Future research on the utility of caregiver social cues during word learning should consider the active role of children.

Examinations into the instances when caregivers present social cues, either in response to a child-led interaction or to lead the discourse, could shed further light on the utility of social cues during word learning.

3.6 Conclusion

Experimental work investigating the enhancement of word learning from the presence of social cues has generally been limited in scope. The methodologies implemented in these studies have primarily focused on the associations between social cues and labels, and disregarded the continuous nature of linguistic input set across time, set within the structure of discourse. Here, we investigated whether social cues may still prove to be beneficial in natural learning scenarios by examining whether caregivers produced eye-gaze cues systematically across discourse in interactions with their children. We found that caregivers were more likely to produce referential eye-gaze cues early on when a discourse topic is initiated, as well as when labels are produced early on in a discourse. These findings suggest an optimistic view on the benefit of social cues in the word learning process. Despite the information-density and messiness of typical learning environments, social cues may still serve to reduce referential ambiguity, especially when considered in light of the natural regularities present in child-directed discourse.

Chapter 4. Relating social cues, neural synchrony, and word learning in natural discourse

4.1 Abstract

Language learning is embedded within multi-modal social situations. Previous work, as overviewed in Chapters 2 and 3, has found that caregivers' production of social cues can assist in children's word learning. However, the generalizability of these findings from labs to natural environments has been debated because these cues do not fully disambiguate potential referents. How might social cues support word learning in more naturalistic settings? We hypothesize that in natural caregiver-child interactions, social cues enhance neural synchrony between interlocutors, and that this enhancement in communication supports children's learning. We test this hypothesis across two experiments by recording the neural activity of both caregivers and their children during natural play interactions using functional near-infrared spectroscopy (fNIRS). Using both behavioral and fNIRS data in Experiment 3, we found that caregivers and children engaged in mutual gaze early in a discourse topic, caregivers and children showed greater neural synchrony. In Experiment 4, we found that the relationship between referential gaze and novel word learning accuracy was mediated by neural synchrony. These findings suggest that the type of social cues differentially contributes to neural synchrony, and that the utility of social cues depends on the goal of an interaction (play vs. learning goal). We discuss why and how social cues may impact children's comprehension of discourse topics, and their impact on word learning.

4.2 Introduction

Interactions between caregivers and their children are inherently multi-modal (Abu-Zhaya, Seidl, & Cristia, 2017; Frank et al., 2013; Kandemir et al., 2023; Kosie & Lew-Williams, 2023; Suarez-Rivera et al., 2022). Alongside speech, caregivers produce a number of behavioral cues to indicate communicative intent (i.e., social cues). Such cues include referential gaze (Çetinçelik, Rowland, & Snijders, 2021; Graham, Nilsen, & Nayer, 2017), joint attention (Morales et al., 1998; Mundy et al., 2007), mutual gaze (Carpenter et al., 1998), and gestures (de Villiers Rader & Zukow-Goldring, 2010; Kalagher & Yu, 2006; Paulus & Fikkert, 2014). Prior research suggested that the presence of these cues can enhance children's word learning (e.g., Baldwin & Tomasello, 1998) through a number of different mechanisms such as orienting attention, extracting information, disambiguating referents, and indicating referential intent (Lee & Lew-Williams, 2023).

As noted previously, a large number of studies have found a positive impact of social cues on word learning (Baldwin, 1995; Bloom, 2002; Bruner, 1983; Houston-Price, Plunkett, & Duffy, 2006; Law, Houston-Price, & Loucas, 2012; Lee & Lew-Williams, 2022; Tomasello & Farrar, 1986). However, such studies are somewhat limited in their application from lab experiments to natural environments. Lab sessions typically expose learners to simple labeling events that teach associations between novel objects and novel labels, by using social cues to direct attention to a novel object while hearing the label (Bannard & Tomasello, 2012; Booth, McGregor, & Rohlfing, 2008; Briganti & Cohen, 2011; Houston-Price, Plunkett, & Duffy, 2006; Yow et al., 2017). Such methodologies can track in-the-moment object-label mappings, but cannot scale to word learning that occurs at longer time-scales (Piazza, Nenchewa, & Lew-Williams, 2021; McMurray, 2016). It is critical to address this limitation when investigating

the utility of social cues during word learning, because a typical caregiver-child interaction exists on a much longer time-scale than lab experiments.

For example, the speech input directed to children does not solely consist of labeling information, but includes the complexities of discourse structure (Frank et al., 2013; Rohde & Frank, 2014; Slone et al., 2023; Suanda, Smith, & Yu, 2018). Discourse, as described in previous chapters, is defined as adjacent utterances referring to the same topic. If we directly consider the inherent structure of linguistic input to children, social cues serve a range of purposes in the moment, and not simply the identification of a referent. They are also used by caregivers to indicate changes in discourse topics, allowing children to ‘get on the same page’ as their caregivers, learn potentially important contextual information, thereby enhancing communication within these interactions by maximizing the amount of information transfer between conversation partners. As such, these processes may help children capitalize on the learning moments included in the discourse. Prior work suggests that this may be the case. For example, in Chapter 3, we found that caregivers spontaneously provide referential eye-gaze cues more frequently at the beginning of a discourse topic. This work parallels previous research. Horowitz and Frank (2015) found that children used social cues to construct the boundaries of discourse structure, where both looks and touches to an object marked it as a discourse topic. In addition, Rohde and Frank (2014) found that caregivers’ points to objects were used more frequently toward the beginning of a discourse topic, suggesting that such social cues served as a signal to introduce a new topic.

This identification of changes in discourse topics is likely to, in turn, promote word learning (Frank et al., 2013; Suanda, Smith, & Yu, 20168; Sullivan & Barner, 2015; Sullivan et al., 2019). One reason for this prediction is that, across a discourse topic, children have more

opportunities to take advantage of rich social information that helps them resolve referential ambiguity. For example, Suanda, Smith, and Yu (2018) found that within longer discourse topics, caregivers and children engage in longer bouts of sustained attention to objects and greater coordination of their behaviors, compared to shorter ones. They suggest that this difference likely affects word learning. During longer discourse topics, children may be more likely to resolve referential ambiguity in part thanks to the multifaceted social information that dynamically directs their attention.

A second reason is that, across a discourse, children can use the surrounding linguistic information to make hypotheses about novel word meanings. For example, Sullivan and Barner (2015) showed 2- to 4-year-old children sets of three pictures. Each set contained an image of an animal, a food item, and a drink. During training trials, the researchers said aloud “I’m very *x*. Look! There’s a *y* on the table”, alongside these pictures. They manipulated the adjective *x* to reflect the mental state of the narrator, whereas *y* was the novel referent label (e.g., “I’m very *thirsty*. Look! There’s a *blick* on the table.”). If participants were able to use discourse to make inferences of word meaning, they would assume that *blick* referred to the image of the drink. They then asked the participants to identify the referent of the novel word by pointing to it amongst the same three images. Sullivan and Barner (2015) found that their participants did indeed use the prior discourse to identify the referent of the novel word, and found that participants were able to generalize this meaning to new instances. They suggest that this finding indicates children as young as two-years-old can use the information presented in a discourse to bootstrap new word meanings.

In order to better understand how social cues influence engagement in a discourse topic and thus impact word learning, we recorded the neural activity of children and their caregivers

using functional near-infrared spectroscopy (fNIRS) to track the neural activity of both caregivers and children in real time. This allows for the measurement of each dyad's neural synchrony (also known as neural coupling). Neural synchrony is defined as the alignment of neural activity within a dyad (Liu et al., 2017). There are a number of reasons why fNIRS, and the measurement of neural synchrony, was chosen as this chapter's methodology. First, fNIRS is a non-invasive and toddler-friendly neuroimaging technique that does not require young participants to produce explicit responses and is tolerant to movement (Ahzari et al., 2020; Boas, Elwell, Ferrari, & Taga, 2014; Vanderwert & Nelson, 2014; Wilcox & Biondi, 2015). This allows young children to engage in an experiment while interacting freely in natural social scenarios. As such, it provides an experimental setting that not only strongly parallels children's natural environments. Second, fNIRS allows for the measurement of responses in real time. This allows for the investigation of the dynamics of neural activity alongside real time changes in discourse structure within child-directed speech, and in-the-moment presentations of caregiver social cues.

Most critically, prior work has found a relationship between neural synchrony and (multimodal) communication. First, increased neural synchrony, or shared neural responses, are associated with successful linguistic communication within a dyad (Liu et al., 2017; Stephens, Silbert, & Hasson, 2010). For example, Stephens, Silbert, & Hasson (2010) recorded the brain activity between speaker and listeners, and found neural synchrony between adult speakers and listeners only when communication was possible (i.e., neural synchrony disappeared when English listeners heard the same story, but in Russian). Furthermore, they found that stronger neural synchrony was associated with better comprehension of the story, suggesting that neural synchrony is a predictor of better understanding. This suggested that the neural synchrony detected cannot be due to the auditory input alone. Rather, the synchronized neural activity was

present, in part, by the shared understanding that came from successful communication. Second, recent work has found that observing social cues in an adult-infant dyad was related to increased neural synchronization (Leong et al., 2017; Piazza et al., 2020; Wass et al., 2018). Leong et al., (2017) used electroencephalography (EEG) to measure the neural activity during a live interaction between an experimenter and 15-month-old infants. They found that neural synchrony was significantly higher when the experimenter and participant engaged in mutual gaze, compared to moments of non-mutual gaze. Together, these findings suggest that the alignment of neural activity between two interlocutors, including caregiver-child interactions, can be related back to both environmental behavioral cues, such as social cues from a caregiver, and from successful communication through shared understanding from linguistic input (Wass et al., 2020). Utilizing this method thus allows for the measurement of in-the-moment shared understanding between caregivers and children as caregivers produced social cues during a natural play interaction.

However, to the best of our knowledge, research has not yet investigated how behavioral cues from a caregiver may influence the neural synchrony of an interaction at the level of discourse topics, nor its influence on subsequent word learning. Typically, investigations into neural synchrony between caregivers and children average synchrony values over a whole interaction, or between conditions (e.g., Miller et al., 2019; Nguyen, Hoehl, & Vrtička, 2021; Nguyen, Abney, et al., 2021; Nguyen, Schleihau, et al., 2021; Reindl et al., 2018). A few studies have examined synchronization in a free-flowing play interaction, and found that the temporal dynamics of neural synchrony were linked to behavioral or linguistic patterns from the interaction (Nguyen, Schleihau, et al., 2021) and have made such links back to word learning (Piazza, Cohen, Trach, & Lew-Williams, 2021). However, there are critical limitations. Nguyen,

Schleithauf, et al. (2021) considered how thematically related utterances may impact neural synchrony but did not find a relationship. However, as demonstrated in Chapter 3, caregiver social cues are distributed systematically across discourse topics, and Nguyen, Schleithauf, et al.'s (2021) study did not investigate how such cues may help delineate what content is relevant to the current discourse topic. In addition, Piazza, Cohen, Trach, & Lew-Williams (2021) was unable to measure neural synchrony due to unreliable measure of the caregivers' neural activity, so their analysis between neural activity and word learning was the consistency of neural activity across children, not synchrony within dyads. Thus, in this chapter, we strive to accomplish two goals. First, we build upon existing research to make novel connections between social cues and discourse topics to gain a fuller picture of the dynamic changes of neural synchrony between caregivers and their children. Second, we endeavor to identify a neural mechanism by which social cues from caregivers may impact children's word learning.

In this chapter, we hypothesize that, to benefit word learning, caregiver social cues serve to not only synchronize behavior presented in an interaction, but also synchronize neural activity. This interpersonal neural synchronization may serve as the mechanism that explains the relation between children's observation of social cues and enhancements in learning (in this case word learning). More specifically, we hypothesize that social cues can serve to enhance shared understanding in an interaction by directing children's attention to relevant information, thus allowing them to track changes in discourse topic as a conversation proceeds. This synchronization in shared understanding manifests as neural synchrony, where children and their caregivers can demonstrate temporal alignment in neural activity.

The social cues under consideration were referential gaze, gaze to the child, joint attention, and mutual gaze. We coded all the interactions for these caregiver social cues. These

specific social cues were examined because they have been identified in previous research as cues likely to enhance word learning (Carpenter et al., 1998; Çetinçelik, Rowland, & Snijders, 2021; Graham, Nilsen, & Nayer, 2017; Morales et al., 1998; Mundy et al., 2007). The play interaction included three sets of toy sets: toys with familiar labels, toys with unfamiliar labels, and novel objects with novel labels. Following the play interaction, the children were tested on their learning of words from the brief play session. We examined the data in various ways linked to our hypotheses. In Experiment 3, we included behavioral and neural activity data from the play interactions including the familiar and unfamiliar toy sets. In Experiment 4, we examined the behavioral and neural data from the play interaction involving the novel toy sets, as well as the behavioral data from the novel word task. For both experiments, we examined the relationship between each type of social cue in turn (i.e., referential gaze, joint attention, mutual gaze, gaze to the child) to determine whether each type of cue uniquely supports communication and word learning.

In Experiment 3, we assess whether the degree of neural synchrony within a dyad is related to the time course of caregiver social cues. We examined three critical hypotheses. First, we investigated whether the amount of social cues overall was related to neural synchrony. Second, we examined whether the distribution of social cues across discourse topics organize neural synchrony, and hypothesized that children and caregivers should have greater shared neural activity during discourse topics that present social cues earlier, compared to later. Finally, we examined whether social cues affect neural synchrony in-the-moment, such that the presentation of a social cue enhances neural synchrony.

In Experiment 4, we examine three hypotheses relating social cues, neural synchrony and word learning. First, we investigated whether overall caregiver social cues (both time and

frequency) related to children's word learning. Second, we determined whether caregiver-child neural synchrony in an interaction is related to children's learning of new words. This was defined as both neural synchrony across the interaction and at each moment a novel word is spoken by the caregiver.. And finally, we explored whether neural synchrony mediates the relationship between social cue presentation and performance in the word learning task.

Across these two experiments, we specifically examined the extent of neural synchrony in three regions of interest (ROIs): the prefrontal cortex (PFC), the temporoparietal junction, and parietal cortex. These areas were chosen because these are identified regions known to support attention, reasoning, language comprehension, and social cognition (Tsujimoto, 2008; Wilson et al., 2008; Iacoboni et al., 2004) and would likely be recruited to support comprehension during a live interaction (e.g., Redcay and Schilbach, 2019), and to support word learning (Piazza, Cohen, Trach, & Lew-Williams, 2021).

4.3 Experiment 3

In Experiment 3, we examined whether social cues promoted neural synchrony across discourse topics in a caregiver-child interaction. Namely, whether observing a caregiver's social cues enhances a child's understanding of a discourse topic, as measured by the synchrony of neural activity within the dyad. Children played with two sets of toys with their caregivers, while the neural activity of both the child and the caregiver were recorded using an fNIRS machine. We then coded a video recording of each interaction for caregiver social cues. The central question of this study was whether observing social cues was related to enhanced synchrony, both across discourse topics embedded within the interaction, and across the interaction as a whole.

4.4 Method

Participants

Sixty-seven two- to four-year-old children ($M = 31.96$ months, $SD = 5.70$, 41 female), with no history of vision or hearing problems and no known developmental delays, and their caregivers participated in the experiment. All children were monolingual (heard greater than 80% English speech at home) and full-term (> 37 weeks gestation). In a post-session optional survey, we asked caregivers further demographic information. Of the caregivers who opted into the survey, the children were predominantly from white households ($n = 37$ White; $n = 4$ Asian; and $n = 15$ more than one race) and came from higher-SES households. All mothers had graduated high school, and most had completed college ($n = 18$; 27%) or graduate school ($n = 47$; 69%). The Princeton Institutional Review Board approved this research protocol (Language Learning: Sounds, Words, and Grammar; IRB record number 0000007117), and a legal guardian provided informed consent for each child.

Efforts were made to preserve as much participant data as possible. Exclusion criteria of an entire participants' data were preregistered (https://aspredicted.org/FRD_4DY). Participants were included in the analysis if they provided any amount of neural or behavioral data during the study. Two children were excluded because of a technical issue with recording from the NIRS machine. Four children were excluded because they refused to wear and removed the fNIRS cap. In addition, nine participants were excluded from the behavioral analysis due to technical issues with the video recordings ($N = 5$) and because the caregivers' eyes were blocked from view (from hair or glasses, $N = 4$). A total of 50 participants provided a complete dataset (both behavioral and neural data). The Princeton Institutional Review Board approved this research

protocol (Language Learning: Sounds, Words, and Grammar; IRB record number 0000007117), and a legal guardian provided informed consent for each child.

Procedure

At the beginning of the experiment, caregivers and their children were welcomed into the lab space. They were asked to sit at a table, facing each other. After sitting, experimenters assisted placing fNIRS caps on both participants. Experimenters then introduced the task, where caregivers and their children would play with toys in three different sets of toys, in three blocks. The three sets of toys were toys with familiar labels (*dog, spoon, banana*), toys with unfamiliar labels (*ferret, whisk, radish*), and novel toys with novel labels (*gazzer, cheem, toma*). Each block of play lasted 3.5 minutes. After each block ended, an experimenter would interrupt the play interaction to introduce the new set of toys. The order of presentation of the familiar and unfamiliar blocks was randomized, such that each always appeared either first or second. The novel play block always appeared last. We focus on the play interactions during the familiar and unfamiliar blocks in this study (the novel block is analyzed in Experiment 4). Each session was recorded from three different angles, capturing a frontal view of the child, a frontal view of the caregiver, and a side view of the dyad. We used only the side view to code behavioral data (Figure 4.1). After participating, families were emailed a \$10 Amazon gift card.



Figure 4.1. An example of the experimental set-up.

Coding caregiver social cues

We focused on coding caregiver eye-gaze. Specifically, we coded the following behaviors: referential gaze to the object of discussion, gaze to the child, joint attention (between the child and play object), and mutual gaze between the caregiver and child. Two trained researchers coded all available videos for caregiver eye-gaze. Videos were coded continuously using ELAN, an open-source software designed for annotating video and audio files (Wittenburg et al., 2006). Researchers coded one dimension at a time, watching the video in its entirety and identifying the onset and offset of each occurrence. To evaluate the reliability of the data, one coder produced annotations for four randomly selected videos that they had not yet seen. We then calculated Cohen's κ for each dimension of caregiver eye-gaze relevant to our hypotheses

(Cohen, 1960). Reliability ranged from fair to strong (Toy: $\kappa = 0.66$, Child: $\kappa = 0.67$, Joint attention: $\kappa = 0.38$, Mutual gaze: $\kappa = 0.55$).

Coding discourse topics

To determine the “topic” of conversation during the play interaction, we used the same coding guidelines as discussed in Chapter 3. One trained researcher produced all annotations using ELAN. Reliability was determined by calculating a simple proportion equivalence of annotation assignments, as conducted by related literature on discourse analyses in computational linguistics (Bower & Hedberg, 2010; Rohde & Frank, 2014). An independent coder produced these annotations, and achieved high agreement (90% agreement).

fNIRS data acquisition and preprocessing

We simultaneously recorded brain activity from the adult and child of each dyad using a dual-brain LABNIRS system (Shimadzu Scientific Instruments, Columbia, MD). We recorded from 41 channels (3 cm in the adult, 2.5 cm in the infants) across the cortex of each participant. These channels covered ROIs: prefrontal cortex (7 channels), temporoparietal junction (17 channels), and parietal cortex (17 channels). The locations of these channels were homologously placed across the infant and the adult (Figure 4.1). The caps were positioned on the basis of known anatomical landmarks according to the international 10-20 system (e.g., the center point of the cap was approximately at Cz). Our primary dependent measure was intersubject correlations (ISC), focusing on the concentrations of oxyhemoglobin (HbO). We chose this measurement because prior research has found greater correlations between fMRI BOLD signals

and HbO for naturalistic stimuli, compared to BOLD and HbR (Liu et al., 2017; Piazza, Cohen, Trach, & Lew-Williams, 2021).

To preprocess the fNIRS data, we removed motion artifacts using both moving standard deviation and spline interpolation (Di Lorenzo et al., 2019). We also low-pass-filtered (1 Hz) and high-pass-filtered (0.01 Hz) the signal to remove physiological noise and drift, respectively. Before analyzing the fNIRS data, the raw data was visually inspected to ensure the quality of data from each channel, from each participant. This resulted in 18.9% of channels from the whole sample to be excluded from analysis. We then downsampled the fNIRS time series for each signal to 1HZ to match the behavioral data. The grand means of neural activity across each ROI for both the caregiver and child were then calculated. To calculate our measurements of neural coupling, or ISC, we conducted Pearson correlations for each caregiver ROI to the corresponding child ROI.

4.5 Results

All analyses were performed in RStudio (RStudio Team, 2023) using R Statistical Software (R version 4.0.2 (2022-06-22), R Core Team, 2022). Mixed-effects analyses were performed using the lme4 package (Bates et al., 2015), and *p*-values were calculated using the lmerTest package (Kuznetsova et al., 2017). These analyses were preregistered (https://aspredicted.org/FRD_4DY).

4.5.1 *Distribution of social cues across a discourse*

The first preregistered analysis sought to replicate the findings of Chapter 3, where caregivers' referential gaze was more likely to be produced earlier on in a discourse compared to

later. We replicated the same model structure from Chapter 3, but expanded our investigation of social cues to include not only referential gaze to the object of discussion and the child (as in Chapter 3) but also moments of joint attention and mutual gaze. Thus, our analysis included a logistic mixed-effects model predicting caregiver social cues (referential gaze, gaze to child, joint attention, mutual gaze) to the referent of discussion (1 or 0, at the frame level) from fixed effects of the log of utterance Position in the discourse sequence (log-scale, centered, numeric), Age of the child (centered, numeric), and an interaction between the two, as well as a fixed effect of Label (whether the co-occurring utterance includes labeling information or not, 1 or 0) and an interaction between Label and utterance Position. We also included a random intercept of discourse topic and subject.

We found that caregivers presented some social cues differently across time within a discourse topic. The models examining the temporal distribution of referential gaze and gaze to the child found a significant effect of position utterance Position, such that referential gaze was more likely to be produced earlier on in a discourse, compared to later ($\beta = -0.08$, OR = 0.92, 95% CI = [0.87, 0.98], $p = 0.009$), and gaze to the child was more likely to be produced later in a discourse, compared to earlier ($\beta = 0.07$, OR = 1.07, 95% CI = [1.01, 1.14], $p = 0.03$). Models looking at referential gaze and joint attention also found a significant effect of label (Referential gaze: $\beta = -0.28$, OR = 0.75, 95% CI = [0.68, 0.84], $p < 0.001$; Joint attention: $\beta = -0.25$, OR = 0.78, 95% CI = [0.70, 0.87], $p < 0.001$), meaning that these cues were less likely to co-occur with utterances containing labeling information, compared to utterances that did not contain labeling information.

We found a significant interaction only between utterance Position and Label when predicting how mutual gaze was distributed across a discourse topic ($\beta = 0.31$, OR = 1.36, 95%

CI = [1.12, 1.66], $p = 0.002$). Caregivers were significantly more likely to pair labeling information with moments of mutual gaze as discourse proceeded, compared to pairing such cue moments with non-labeling information. Overall, this suggests that caregivers use social cues in distinctive ways, depending on the type of cue. Caregivers seem to use referential gaze to indicate changes in discourse topics, and gaze to the child to maintain engagement across time. Engaging in mutual gaze seems to be used by caregivers to successfully transmit labeling information across a discourse topic. In comparison, our results regarding moments of joint attention seem to suggest that, while regarded in the literature as highly-informative teaching moments, such cues may not be as maximally informative during more natural play instances, given that they are less likely to be paired with co-occurring labeling information.

4.5.2 Relating social cue presentation to overall neural synchrony

The second preregistered analysis examined the relationship between caregiver social cue presentation and neural synchrony. We hypothesized that more social cues in the caregiver-child interaction would be positively associated with increased neural synchrony (ISC). To investigate this, we calculated the number of social cues in two ways; by the frequency of social cue production and the amount of time each social cue was presented. To calculate ISC, we computed a Pearson correlation between the averaged time series between child and caregiver participant for each ROI (e.g., frontal). Neural synchrony across dyads was only significantly different from 0 in the frontal ROI ($t(117) = 6.09$, $p < .001$), compared to the left or right ROI (Left: $t(117) = -0.37$, $p = 0.71$, Right: $t(120) = 0.14$, $p = 0.89$). Thus, we report on only synchrony values from that region. Furthermore, we found that frontal neural synchrony was not

significantly different between blocks of play (familiar vs. unfamiliar toys, $t(114) = 1.27$, $p = 0.21$), so the following analyses will collapse across toy sets.

Control analysis. We conducted a control analysis to determine whether measurements of neural synchrony were due to spurious effects, such as simply being in the same room or being involved in the same task (Leong et al., 2017; Nguyen et al., 2021; Piazza, Hasenfratz, Hasson, & Lew-Williams, 2021). To do so, we created a surrogate dataset of 1000 temporally-shuffled versions for each dyad. For each caregiver-child dyad, we randomly shuffled a child's frontal fNIRS time-course to a random time point within the caregiver's frontal fNIRS time-course. We then calculated the average ISC for each random pairing. Using a linear mixed-effects model, we predicted the average ISC from fixed effects of dyad (random or true, reference level = random) and a random intercept of dyad. We found that there was not a significant difference in the amount of neural synchrony depending on the toys played with (familiar vs. unfamiliar). Critically, we found a significant difference of dyad (random vs. true), such that true dyads were significantly more synchronized in neural activity compared to random dyads ($\beta = 0.06$, 95% CI = [0.04, 0.07], $p < 0.001$). Furthermore, the average correlation for random dyads was not significantly different from zero ($\beta = 0.00$, 95% CI = [-0.01, 0.02], $p = 0.72$). This suggests that the neural synchrony detected from our true dyads is not attributable to spurious factors, but more specifically related to what is occurring in the time-contingent, live interaction between caregivers and their children.

Overall social cues usage is not related to overall synchrony. We conducted a linear regression predicting the average ISC of each dyad from the frequency of each social cue produced (referential gaze, joint attention, mutual gaze, and gaze to the child). We found that for all social cues considered we found no significant relationship between social cue presentation

and ISC (Table 4.1). We found the same non-significant relationship when considering the amount of time each social cue was produced. This suggests that the total amount of social cues presented (by both frequency and time produced) does not affect overall neural synchrony between caregivers and their children.

Social cue type	Frequency – ISC		Time – ISC	
	β	p	β	p
Joint attention	0.001	0.35	0.00	0.78
Referential gaze	0.002	0.11	0.001	0.08
Mutual gaze	0.002	0.13	0.00	0.98
Gaze to child	0.001	0.32	0.00	0.47

Table 4.1. Beta estimates and p -values when predicting ISC from frequency and time of each social cue type.

4.5.3 Relating social cue presentation and neural synchrony within discourse topics

Our third preregistered analysis investigated how the timing of social cue presentations affected the neural synchrony of caregiver-child interactions within discourse topics. We predicted that if caregivers presented social cues earlier on in a discourse topic, compared to later, child-caregiver neural synchrony would increase. To assess this, we calculated the total amount of time a social cue (referential gaze, gaze to child, joint attention, and mutual gaze) was presented in the first vs. second half of a discourse topic. This calculation included social cue coding that occurred when caregiver utterances were not produced to provide a more comprehensive understanding of how social cues are presented within a discourse topic. This

means that the total calculation time producing social cues was the sum amount of time each social cue was presented across the interaction, regardless of whether caregivers were talking. Discourse topics where caregivers produced social cues for a greater amount of time in the first half of the discourse (vs. the second half) were termed “early” presentations of social cue, and coded as 1, while all other discourse topics were coded as 0.

We then calculated the average ISC for each dyad, and for each discourse topic. We calculated ISC the same as previously. However, to also account for the potential time lag in brain–behavior comparisons due to the hemodynamic response function (HRF), we exploratorily lagged the fNIRS signal by 4 seconds (see Piazza, Hasenfratz, Hasson, & Lew-Williams, 2021). We then recalculated this lagged average ISC for each dyad, for each discourse topic.

Our model was constructed to determine how the timing of caregiver social cues influenced neural synchrony. We created a linear mixed-effects model predicting the average ISC for each dyad and discourse topic from fixed effects of discourse-cue presentation (i.e., “early” presentations or not, 1 or 0), and included random intercepts of participants and discourse topics. We repeated this model for each type of social cue.

Early moments of mutual gaze promote neural synchrony. We first implemented our model with reference to mutual gaze. We found that, when predicting our unlagged measurements of neural synchrony, the temporal distribution of mutual gaze significantly predicted neural synchrony within discourse topics ($\beta = 0.02$, 95% CI = [0.01, 0.03], $p = < 0.001$). This suggests that when dyads engage in mutual gaze earlier in a discourse topic, they were, on average, more synchronized. When predicting our lagged measurements of neural synchrony, we found that none of our fixed effects were significant predictors. In addition, all

other models using the temporal distribution of other social cues (referential gaze, joint attention, and gaze to the child) as predictors of neural synchrony were not significant.

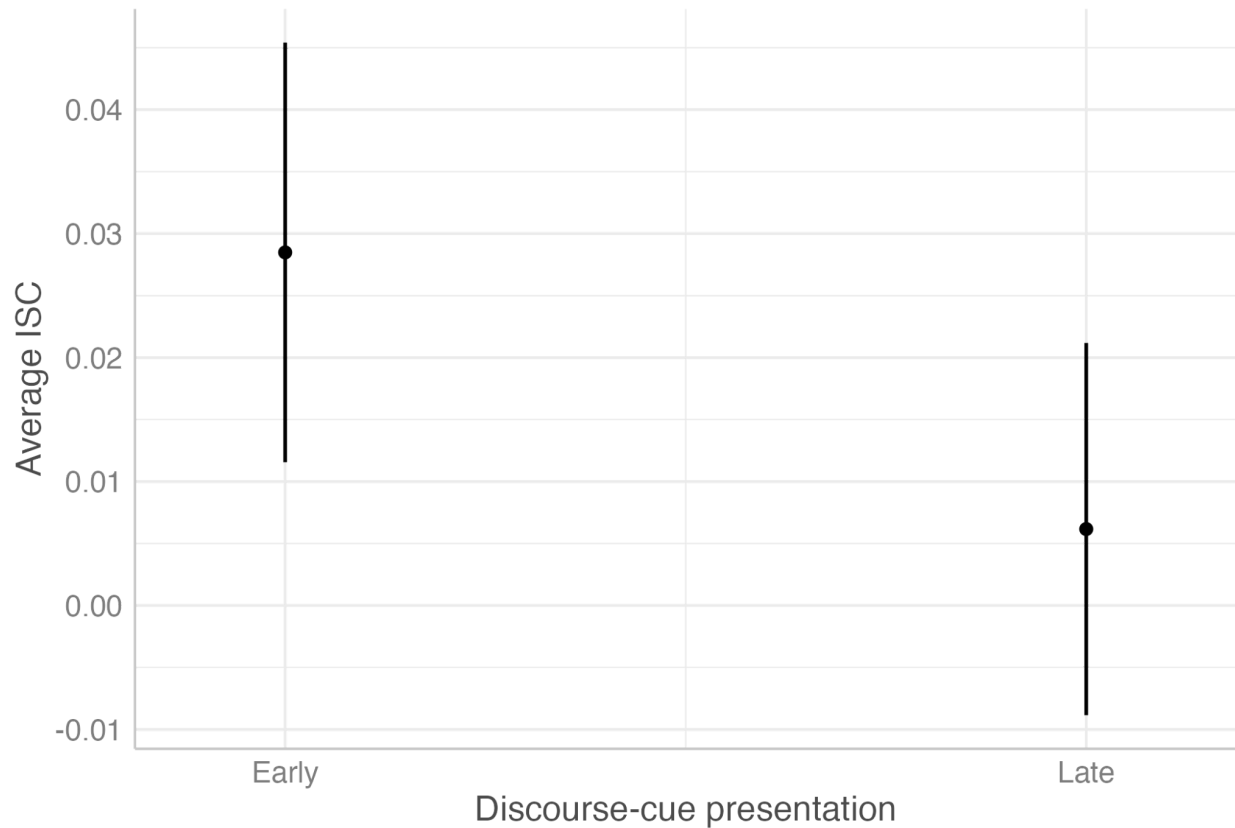


Figure 4.2. A graph demonstrating the marginal effect of discourse-cue presentation – presenting mutual gaze early or later within a discourse topic – on average ISC within that discourse topic. The point represents the average ISC for each type of discourse-cue presentation (early vs. late), and the line portion represents 95% confidence intervals.

4.5.4 Relating in-the-moment social cue presentation and neural synchrony

To further investigate how social cues may influence neural synchrony, we analyzed how in-the-moment presentation of social cues impact dyads' neural synchrony. We defined in-the-moment presentation as 10 seconds before the onset of a social cue. We predicted that

caregiver social cues would significantly enhance the neural synchrony of a dyad. To calculate neural synchrony, we found the average ISC of the caregiver-child dyads at the onset of a social cue (time onset = 0) and 10 seconds prior (in 1 second steps). We again exploratorily lagged the fNIRS time signal by 4 seconds to account for HRF.

To examine this analysis, we constructed a linear mixed effects model predicting average neural synchrony within a dyad, from a fixed effect of time window (numeric, [-10, 0]), and a random intercept of dyad. We conducted this same analysis for each social cue type (joint attention, referential gaze, mutual gaze, gaze to the child). However, we failed to find a significant relationship between synchrony and in-the-moment presentation of each type of social cue during the prespecified window. These results replicated when including lagged values of ISC.

4.6 Discussion

We examined how the distribution of social cues (joint attention, referential gaze, mutual gaze, gaze toward the child, and deictic gestures) across discourse topics influences the synchrony of neural activity between caregivers and children during a naturalistic play session. First, we replicated some of the results from Chapter 3: referential gaze from caregivers was more likely to be produced earlier on in a discourse topic, compared to later, and this cue was less likely to be paired with labeling information, compared to non-labeling information. In addition, like in Chapter 3, caregivers were more likely to look toward their child later in a discourse topic. We also expanded upon our results from Chapter 3 in two ways. One, we found that joint attention was also less likely to be paired with labeling information, compared to

non-labeling information. Two, mutual gaze cues were more likely to be paired with labeling information later on in a discourse topic, compared to earlier.

Second, we did not find a relationship between the overall amount of social cues produced (frequency of production or amount of observed) and overall amount of neural synchrony within dyads across the whole play session. However, we did find that the *temporal* distribution of social cues across a discourse topic predicted neural synchrony. Specifically, dyads were more synchronized within a discourse topic when caregivers produced mutual gaze earlier on in a discourse topic, compared to later. In comparison, we did not find that the temporal distribution of referential cues towards objects (referential gaze, joint attention, gaze to the child) were related to neural synchrony in the first or second half of discourse topics. Finally, we examined how social cues may influence a dyad's neural synchrony at shorter time-scales, 10 seconds prior to the onset of a social cue. We failed to find any significant increase in neural synchrony during the time window when each of the social cues were presented.

Our critical finding presents evidence that the timing of engagement in mutual gaze enhances neural synchrony shared understanding between caregivers and their children during natural interactions, specifically within discourse topics. This behavioral synchrony at the beginning of a conversation was reflected in shared neural activity, which led dyads to show increased neural synchrony as the conversation proceeded. This finding converges with prior research (Leong et al., 2017; Luft et al., 2022; Piazza, Hasenfratz, Hasson, & Lew-Williams, 2021; Vanoncini et al., 2024), where mutual gaze was found to promote neural synchrony within the PFC between caregiver-child dyads. Our work suggests that when caregivers and children attend to each other, the dyad aligns in neural activity, suggesting that children were better able to understand discourse content as the conversation proceeded.

Why did engaging in mutual gaze track neural synchrony within discourse topics?

Ostensive gaze cues have been thought to be a salient cue to indicate communicative intent (Csibra & Gergely, 2009; Leong et al., 2017; Senju & Csibra, 2008). According to the natural pedagogy hypothesis, a speaker can use ostensive cues, such as mutual gaze or general gaze to the child, to signal an impending attempt at communication (Csibra & Gergely, 2006). Children may use such cues to anticipate referential communication directed towards them. For example, Senju & Csibra (2008) found that, in one experiment, 6-month-old infants only followed the gaze of an actor when the actor engaged in mutual gaze with the infant, suggesting that the participants used this ostensive signal to identify possible referential communication, and behave accordingly (i.e., following the actor's gaze). Furthermore, this synchronization of behavior within a dyad has also been found to have aligned neural mechanisms and has been suggested to promote communication (Leong et al., 2017; Piazza, Hasenfratz, Hasson, & Lew-Williams, 2021; Vanoncini et al., 2024). Our results converge with this prior research. When dyads engaged in mutual gaze at the beginning of discourse topics, their neural activity aligned as they directed their gaze to process similar types of information (i.e., each other) and anticipate communication.

In Chapters 2 and 3, we previously hypothesized that referential gaze may help children track changes in discourse content and enhance their understanding of an interaction, as indexed by neural synchrony within a dyad. Thus, we predicted that early presentations of this cue would be associated with increased neural synchrony. However, we failed to find this relationship. Why might this have been the case? Prior research has found that referential gaze is a pervasive cue in caregiver-child interactions (Abney et al., 2020; Goupil et al., 2024; Phillips et al., 2023; Sun & Yoshida, 2022; Wildt & Rohlfing, 2024). Very early in life, infants as young as nine months can readily follow the gaze of their parents (Mundy et al., 2007; Flom et al., 2004). This dynamic

switches to the *caregivers* following children's gaze at around two years of age (Deák et al., 2014; Suarez-Rivera et al., 2019; Wass et al., 2018). Indeed, Sun and Yoshida (2022) found that it was significantly more common for caregivers to follow their child's attention, compared to the children following their caregiver's attention. It is likely, then, that the reason the referential cues investigated in the current study did not track neural synchrony is because of this temporal dynamic. Incorporating whether caregivers follow their child's attention, or vice versa, may relate to synchrony better than the current measurements of referential gaze, and serve to illustrate the role of this cue embedded within leader-follower dynamics.

4.7 Experiment 4

In Experiment 4, we investigated whether levels of neural synchrony mediated the relationship between observation of social cues and accuracy of word learning within caregiver-child dyads. From the same experimental session as Experiment 3, children played with a set of novel toys with novel labels with their caregivers, while the neural activity of both the child and the caregiver were recorded using the dual-brain fNIRS paradigm described in Experiment 3. Following the play session, an experimenter administered a behavioral test of children's novel word learning. The central question of this study was whether observing caregivers' social cues (joint attention, mutual gaze, gaze to the child, and referential gaze) and neural synchrony of a dyad are related to children's accuracy of word learning exhibited in the behavioral task.

4.8 Method

Participants

The same 67 dyads from Experiment 3 participated in Experiment 4. Again, efforts were made to preserve as much participant data as possible. Exclusion criteria of an entire participants' data were preregistered (https://aspredicted.org/S3F_GZD). Participants were included in the analysis if they provided any amount of neural or behavioral data during the study. From these, 52 participants provided behavioral data for the novel play session. For the novel word learning task, 12 participants' data were excluded because of technical difficulties (e.g., video camera failing to record), and an additional 4 participants were excluded for failing to meet our pre-registered criterion of success on practice trials in this task. This left 51 participants who provided word learning data. Finally, 55 dyads provide usable fNIRS data (i.e., providing at least one channel for at least one patch). A total of 38 participants provided a complete dataset (both behavioral and neural data). The Princeton Institutional Review Board approved this research protocol (Language Learning: Sounds, Words, and Grammar; IRB record number 0000007117), and a legal guardian provided informed consent for each child.

Procedure

Immediately following play with the two sets of toys used in Experiment 3, an experimenter brought out a set of novel toys with novel labels (*gazzer*, *cheem*, *toma*, Figure 4.3). Children and caregivers were instructed to play with the toys, and the labels corresponding to each toy were posted on a wall behind the child so that the caregiver could be reminded of their names whenever needed. Play lasted 3.5 minutes. As in Experiment 3, each session was recorded from three different angles, capturing frontal images of the child and the caregiver, as well as each dyad from the side. We used only the side view to code behavioral data.



Figure 4.3. The novel objects presented in Experiment 4, and were matched with a label (*cheem*, *gazzler*, *tobu*). Object-label mappings varied depending on the counterbalancing list.

After the play session finished, child participants were then administered a behavioral word learning task to assess their learning of the novel labels. On each trial of this task, they asked, “Which is the *X*? Can you give me the *X*?” and slid a tray towards the child that contained two of the novel objects. Participants’ choices were recorded as whichever object they eventually handed to the experimenter, regardless of which object was touched first. At the end of each trial, experimenters praised their performance, regardless of accuracy, by saying “Good job!”.

Two practice trials were presented, with each object presented possessing a familiar label (*dog*, *banana*). We preregistered that participants had to pass both practice trials in order for their data to be included; this helped to ensure that their touches and reaches were indicative of their underlying knowledge, rather than random behavior. Following the practice trials, the test trials

were presented with the novel objects seen during the play session. Each novel object was queried three times, for a total of 9 test trials. The order of trials and side presentation was the same across participants. The specific order of trials was initially randomized, but modified such that each novel label was not queried more than two times in a row. The side of presentation of each object (left or right) was specified in order to counterbalance how often a target and foil label appeared on either side. Namely, each novel label appeared as a target either 1 or 2 times on each side, and 1 or 2 times on each side as the foil. Finally, participants were assigned to one of three counterbalanced lists that varied the novel object–label association, designed to account for spurious relationships between novel objects and labels.

Coding of caregiver social cues

Coding of caregiver social cues was the same as in Experiment 3, except we included data from the novel play session.

fNIRS data acquisition and preprocessing

This experiment used the same data acquisition and preprocessing pipeline as in Experiment 3, except we included data from the novel play session.

4.9 Results

All analyses were performed in RStudio (RStudio Team, 2023) using R Statistical Software (R version 4.0.2 (2022-06-22), R Core Team, 2022). Mixed-effects analyses were performed using the lme4 package (Bates et al., 2015), and *p*-values were calculated using the

lmerTest package (Kuznetsova et al., 2017). These analyses were preregistered (https://aspredicted.org/S3F_GZD).

4.9.1 Word learning performance

Children's performance in the behavioral word learning task was on average 66% ($SD = 30\%$), and significantly different from chance (chance = 50%; $t(46) = 3.64, p < 0.001$). Their performance suggests that, as a group, participants readily learned the novel object-label mappings within a short period of time (3.5 minutes).

Using a one-way ANOVA, we also found that there were no statistically significant differences in accuracy between these counterbalancing groups ($F(2, 44) = 0.80, p = 0.46$). Thus, we did not include this as a fixed effect in subsequent analyses.

4.9.2 Observing social cues are not related to learning of novel object labels

Our first preregistered analysis examined how the production of caregiver social cues (joint attention, referential gaze, mutual gaze, gaze to child) may be related to children's novel word learning. We constructed two linear models predicting the proportion of correct responses in the behavioral word learning test from: 1) the frequency of caregivers' production of each social cue, and 2) the time spent producing each cue. We found that none of our measurements of social cues (frequency or time, for each social cue) were related to word learning accuracy. In an exploratory analysis, we collapsed across all social cues to predict accuracy in word learning, and also did not find a significant relationship.

4.9.3 Relating neural synchrony to novel word learning

We replicated our control analysis from Experiment 3 to first determine whether measurements of neural synchrony were actually due to engagement in dyadic interactions, or due to spurious effects such as the experimental task (Leong et al., 2017; Nguyen, Schleithauf, et al., 2021; Piazza, Hasenfratz, Hasson, & Lew-Williams, 2020). If neural synchrony could be achieved by simply engaging in the same experimental task (i.e., by playing with the same set of toys), children of one dyad should present similar neural activation with their actual caregiver as well as a random caregiver of another dyad. Using the same analysis as previously, we found a significant difference of dyad (random vs. true), such that true dyads were significantly more synchronized in neural activity compared to random dyads ($\beta = 0.03$, 95% CI = [0.00, 0.05], $p = 0.05$). Furthermore, the average correlation for random dyads was not significantly different from zero ($\beta = 0.00$, 95% CI = [-0.02, 0.02], $p = 0.98$). This suggests that the neural synchrony detected from our true dyads was not attributable to simply engaging with the experimental task, but was more specifically related to what occurred in the time-contingent, live interaction between caregivers and their children.

Our preregistered analysis examined whether neural synchrony within caregiver-child dyads was related to children's accuracy on the novel word learning task. Neural synchrony was defined in two ways. First, we defined average ISC across the entire 3.5-minute play session. To calculate ISC, we computed a Pearson correlation between the averaged time series between child and caregiver participant for each ROI (e.g., frontal). Neural synchrony across dyads was only significantly different from 0 in the frontal ROI ($t(51) = 2.01$, $p < .04$), compared to the left or right ROI (Left: $t(50) = 1.41$, $p = 0.16$, Right: $t(52) = -0.83$, $p = 0.41$). Thus, we report on only synchrony values from that region. For our first preregistered analysis, we correlated average

neural synchrony and accuracy on the word learning task, but we did not find a significant relationship ($r = 0.06$, $p = 0.70$).

The second way we defined neural synchrony was average ISC within a 20-second time window surrounding each caregiver's production of one of the three novel words (10 seconds before and after). In our second preregistered analysis, we constructed a linear mixed-effects model to predict predict accuracy in selecting the novel words (*gazzer*, *toma*, *cheem*) from each dyad's average ISC when each novel label was spoken. To account for HRF, we also exploratorily lagged the fNIRS signals by 4 seconds and recalculated ISC. We included each variation of lagged-ISC as separate models. For all models, we included a random intercept of subject. However, for all our analyses, we again did not find a significant effect of (both lagged and unlagged) ISC when predicting word learning.

4.9.4 Does neural synchrony mediate the relationship between social cues and novel word learning?

Our last preregistered analysis examined our critical prediction: that neural synchrony mediates the relationship between observing social cues and enhanced word learning. We first examined how our overall measures of these three indices (social cue usage, neural synchrony, and word learning) were related to each other. In this analysis, we implemented a mediation model using the 'mediation' package in R (Tingley, Yamamoto, Hirose, Keele, & Imai, 2014). Our independent variable was social cue usage, our dependent variable was word learning performance, and our mediating variable was neural synchrony within a dyad.

We first operationalized social cue usage as both the frequency of production and total amount of time, neural synchrony as the average ISC across the session, and word learning

performance as the proportion of correct trials. Although we failed to find a relationship for nearly all of our measurements of social cue usage in the session and word learning performance in Section 4.9.2, we still conducted the mediation model as described because prior research suggests a relationship should be present (Shrout & Bolger, 2002). However, we failed to find significant mediation for all models tested (for each social cue, for each measurement). This result replicated when lagging ISC.

Next, we explored whether neural synchrony when novel words were spoken mediated the relationship between social cues and word learning. In this exploratory analysis, we predicted the degree of synchrony when novel words are spoken may mediate the relationship between when social cues are presented across discourse topics (discourse-cue presentation: earlier vs. later) and word learning accuracy (i.e., for each novel word).

To determine this, we first calculated the frequency and average amount of time each social was presented across the length of each discourse topic (i.e., novel word). If the average usage of a social cue was greater in the first-half of a discourse topic compared to later, we coded this discourse-cue presentation as “early” and all others as “later” (“early” = 1, “later” = 0). We calculated neural synchrony as the average ISC when each novel word was spoken (*cheem*, *gazzar*, *tobu*) within each discourse topic, for each dyad. We also exploratorily lagged the average ISC by 4 seconds to account for potential hemodynamic lags.

We failed to find significant mediation for all models concerning discourse-cue presentation measured by frequency of presenting each social cue. This result replicated at a 4-second lag. We also failed to find significant mediation for models concerning the amount of time joint attention, mutual gaze, and gaze to the child were presented across discourse topics.

However, we did find significant partial mediation for one model: the amount of time referential gaze was presented across discourse topics (Figure 4.4). Our results showed a significant total effect between when referential gaze was presented (early vs. later) and word learning accuracy ($\beta = -0.17$, 95% CI = $[-0.31, -0.03]$, $p = 0.02$), such that children learned words better when referential gaze was presented later in a discourse topic. For path A in the mediation model (see Figure 4.4), we found a significant effect of when referential gaze was presented (early vs. later) on neural synchrony when a novel word was spoken ($\beta = 0.08$, 95% CI = $[0.011, 0.014]$, $p = 0.02$), such that when caregivers provided more referential gaze at the beginning of discourse topics, dyads were more synchronized when a novel word was spoken. For path B, the relationship between neural synchrony when a novel word was spoken and word learning was trending in significance ($\beta = 0.48$, 95% CI = $[-0.04, 1.01]$, $p = 0.07$), where greater synchrony was associated with better word learning.

Finally, we found that the relationship between when referential gaze was presented (early vs. later) and word learning accuracy was still significant when controlling for neural synchrony ($\beta = -0.21$, 95% CI = $[-0.35, -0.07]$, $p = 0.006$). We used a bootstrapping approach to assess the significance of the indirect effects (simulated 5000 times, Preacher & Hayes, 2004), and found that the indirect effect was also significant ($\beta = 0.05$, 95% CI = $[0.002, 0.12]$, $p = 0.04$). This result suggests that neural synchrony, time-locked to the production of novel words, partially explains the relationship between when caregivers' social cues are used across discourse topics and their impact on word learning.

To further clarify this relationship, we investigated the timing of label production. We found that the majority of labels are produced earlier in a discourse topic compared to later ($t(760) = 7.40$, $p < 0.001$). This finding, in combination with the aforementioned mediational

model, points to a particular scenario that may maximize learning. When dyads become synchronized in neural activity from producing the novel labels, children can use subsequent disambiguating social cues later in the conversation to learn words better. This mediation was not present when we accounted for a 4-second hemodynamic lag. When repeating this model and controlling for age (by adding a main effect of age, numeric and centered), the effect of partial mediation remains.

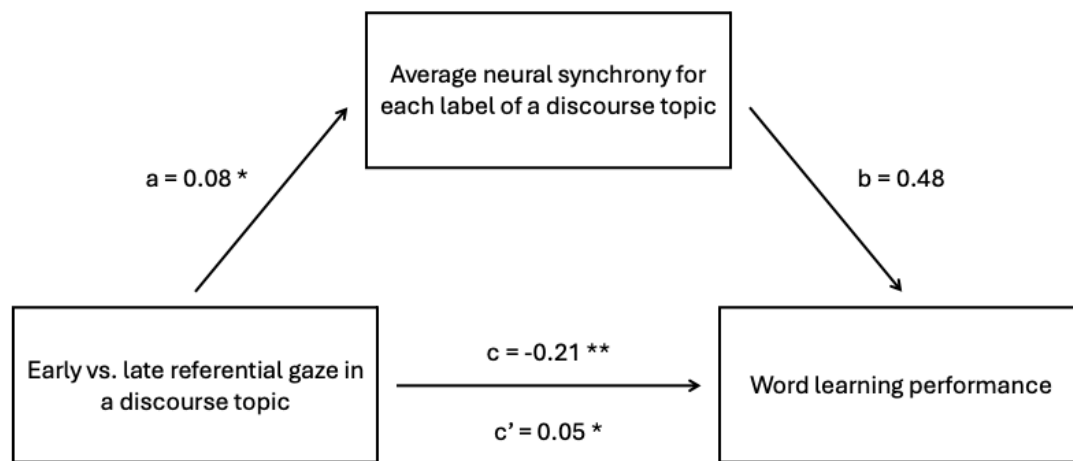


Figure 4.4. The mediation model shows the link between temporal distribution of referential gaze in a discourse topic (early vs. late; early = 1, late = 0) and word learning performance (numeric, proportion), as mediated by average neural synchrony when labels are presented in a discourse topic. Along the lower path, (c) and (c') show results when the mediator was not included and was included in the model, respectively. Asterisks indicate significant paths (* $p < 0.05$, ** $p < 0.01$).

4.10 Discussion

Our goal for this study was to explore the relationship between caregiver social cues, neural synchrony, and children's word learning. Our central hypothesis was that neural synchrony during natural caregiver-child interactions would mediate the relationship between children's observation of social cues and enhanced word learning. While we did not find a significant relationship between our preregistered measures of these constructs, we further examined discourse structure and labeling events in exploratory analyses. We found that neural synchrony when novel labels were produced partially mediated the relationship between when the temporal distribution of caregivers' referential gaze across a discourse topic and children's word learning accuracy.

This result characterizes a particular pathway of learning novel words in a dynamic, naturalistic setting. Specifically, when caregivers use novel labels to introduce a discourse topic, this serves to synchronize neural activity between them and their child. This suggests that the dyad shared increased common ground in these moments, perhaps reflecting an increase in shared attention or a similar understanding of linguistic information. Receiving referential gaze information after understanding the topic of discussion allows children to disambiguate potential referents and thus strengthen the mapping between labels and objects.

Prior research has cast doubt on the benefit of learning words from social cues in natural settings (e.g., Frank, Tenenbaum, & Fernald, 2013). Specifically, this work has suggested that cues such as referential gaze do not have as much utility in the learning process as suspected because they do not cleanly refer to potential referents in typical learning environments. Our results that examined overall use of social cues converge with this idea. Our measure of the production of social cues across the whole play session was unrelated to children's word

learning. However, our exploratory findings suggest the social cues carry value in supporting word learning in more specific ways. When learning a novel word from a play interaction, it seems that observing earlier referential gaze is less beneficial compared to observing later referential gaze, which converges to some extent with prior experimental work (Lee & Lew-Williams, 2022). This benefit is mediated by the degree of neural synchrony when a caregiver produces a novel word. As labels were presented primarily in the first half of a discourse topic, children could presumably use this linguistic information to help identify a discourse topic. When children were able to ‘get on the same page’ as their caregivers when novel labels are spoken, they took advantage of later-produced social cues to determine potential referents of new, unfamiliar labels. Thus, the subsequent social cues may have strengthened the mapping between the new labels and novel objects. This asymmetry in the presentation of labels and social cues converges with our prior work. In Chapter 3, we found that referential gaze cues were less likely to co-occur with labeling information, compared to non-labeling linguistic input, during a natural play interaction, and this result was replicated in Experiment 3. We build upon this work in Experiment 4 by demonstrating that this (mis)alignment in timing has potentially helpful impacts on word learning.

Our findings further our understanding of the role of social information from others during the word learning process. To the best of our knowledge, there has not yet been a study examining if the link between social cues and word learning is mediated by neural synchrony. The most relevant study by Piazza, Cohen, Trach, and Lew-Williams (2021) recorded the neural activity of 3.5- to 4.5-year-old children while reading a book with an adult experimenter, and related this neural activity to subsequent word learning. However, their primary measure of synchrony was synchrony *between children* and not between children and adults. Their findings

suggest that there is a particular neural signature that tracks word learning within learners, while our results present a possible dyadic characterization of the learning contexts between caregivers and their children that may maximize word learning.

In Experiment 3, we found that dyads were more synchronous when moments of mutual gaze occurred earlier in a discourse topic compared to later. We interpreted this finding to mean that early opportunities for children to “get on the same page” with their caregiver impact their understanding of a discourse topic. In Experiment 4, which focused on word learning, the efficacy of social cues was tied to neural synchrony between dyads when labels were presented. Why, in Experiment 4, was mutual gaze unrelated to neural synchrony and word learning, relative to referential gaze? One potential explanation is that engaging in behavioral, and thus neural, synchrony is easier when play interactions involve familiar toys compared to novel toys. In Experiment 3, the goal of an interaction involved only play, and both partners (caregivers and children) had some familiarity with the objects presented. In Experiment 4, the children were completely naive to the toys presented. Only the caregivers were told the labels of the novel objects. Thus, it is possible that when the toys are more familiar for both conversation partners (caregivers and children), the partners are better able to synchronize their behavior, such as engage in mutual gaze, and align their neural activity to attend to each other.

To explore this, we conducted an exploratory linear mixed-effects model with a fixed effect of play block (familiar, unfamiliar, novel, reference level = novel) and a random intercept of subject (Figure 4.5). We found that dyads demonstrated significantly higher neural synchrony during play with familiar objects, compared to the novel objects ($\beta = 0.04$, 95% CI = [0.01, 0.07], $p = 0.02$), and that there was not a significant difference in synchrony during play with unfamiliar vs. novel objects. This suggests that alignment in neural activity, and in behavior, may

be easier when play includes familiar objects, where previous shared experiences allow dyads to synchronize neural activity and behavioral engagement.

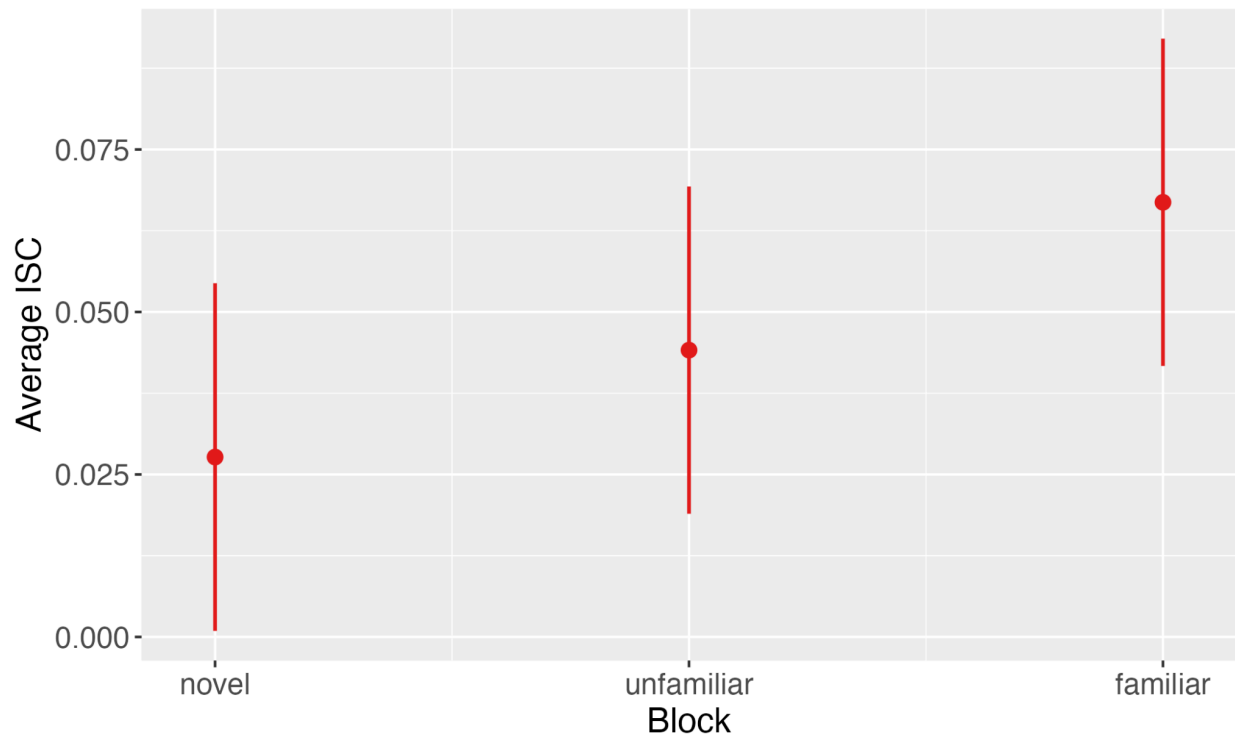


Figure 4.5. A graph demonstrating the marginal effect of play block (novel, unfamiliar, and familiar) on the average neural synchrony of a dyad within that block. The line represents 95% confidence intervals.

Minimal research has examined the consistency of play routines in childhood, but routines as a whole are well known to be a part of children's daily experiences, such as feeding and changing. These routines help structure meaningful interactions. Furthermore, these routines have proved to help infants to coordinate with the adult's actions (Trevarthen, 1979) and to become skilled cooperative agents as they participate in them (Lerner et al., 2011; Rączaszek-Leonardi et al., 2013). Routines, as they are familiar and predictable, also orient the

infant's capacity to anticipate the other's action and create expectations on the other's behavior (Ambrosini et al., 2013; Reddy, Markova, & Wallot, 2013). Thus, engaging in familiar games or playing with familiar toys can become predictable activities for both the child and caregiver, allowing for the synchronization of behavior and neural activity. Indeed, recent research has found that the language directed towards children across routines and play are consistent and context-dependent (e.g., Tamis-LeMonda et al., 2019). This points to the possibility that play with familiar (vs. relatively unfamiliar) toys encompasses shared prior experiences that allow dyads to become more closely synchronized in both brain and behavior.

If unfamiliar play leads to less synchronized behavior and neural activity, how can children learn words from their caregivers? The answer to this question may be rooted in the idea that children and caregivers adapt and react to each other in real time. As previously mentioned, natural interactions between children and caregivers include leader-follower dynamics (Abney et al., 2020; Goupil et al., 2024; Phillips et al., 2023; Sun & Yoshida, 2022; Wildt & Rohlfing, 2024). And prior work has found that word learning is promoted when adults follow their child's attention (Tomasello & Todd, 1983; Tomasello & Farrar, 1986; Tomasello, Mannle & Kruger, 1986; Carpenter et al., 1998; Smith, Adamson & Bakeman, 1988). However, that is not the only pathway to learning new words. Shimpi and Huttenlocher (2007) differentiated between successful and unsuccessful "follow-in" scenarios, where mothers label the object that is currently in the child's focus, and "lead-in" scenarios, where mothers attempt to redirect their child's attention to an object out of focus. They found that caregivers' use of successful lead-in labels was positively correlated with children's vocabulary, suggesting that "follow-in" scenarios are not the only efficient scenario to learn new words. Indeed, when caregivers are able to redirect their child's attention, word learning can still proceed successfully. Thus, referential gaze

may be a more pertinent or informative social cue in this scenario, compared to mutual gaze. Caregivers who were able to establish the topic of discourse using novel labels increased the synchronization of neural activity and assisted in children identifying the topic of conversation. After determining the discourse topic, caregivers could then be successful in redirecting their child's attention to the aforementioned object. Such misalignment likely assisted in the children's learning.

Limitations and future directions

In Experiments 3 and 4, we sought to examine how the temporal distributions of social cues across discourse topics may impact neural synchrony of an interaction, as well as whether neural synchrony within a dyad explained the relationship between children's observation of social cues and their subsequent word learning. However, there are still some limitations that should be considered in future research. First, we considered neural synchrony only within the PFC because that was the only area of measurement from our fNIRS system that demonstrated significant synchronization between caregivers and children. However, it is likely that other areas of the brain are critical to the word learning process. For example, the temporoparietal junction (TPJ) is also recruited during interpersonal interactions as it is associated with language processing, self-reference and processing of one's own and others' mental states (Monticelli et al., 2021). In order to fully understand the neural mechanisms of discourse comprehension and whether it is facilitated by observing social information, as well as the neural mechanisms of word learning, other regions that support processes central to language learning (e.g., memory, retrieval) should be considered in finer-grained detail (e.g., Takashima et al., 2019).

Second, we considered each social cue's relationship to word learning and neural synchrony in separate models. However, live interactions proceed with dynamic, multidimensional information. Thus, there remains a large landscape to explore in exactly how multimodal social information impacts discourse comprehension, word learning, and neural synchrony. For example, Kosie and Lew-Williams (2023) demonstrated the wealth of multimodal information present in caregiver-child play interactions, such as facial expression, different categories of gestures, and social touch. Indeed, touch and facial expressions (operationalized as emotions) have been shown to influence the neural synchrony within dyadic interactions (Alonso, McDorman, & Romeo, 2024; Morgan et al., 2023; Nguyen, Abney et al., 2021; Santamaria et al., 2020; Schwartz et al., 2022). In order to fully evaluate how social information influences the learning process, it will be critical to examine how each social cue, isolated and in tandem with other types of cues.

4.11 Conclusion

Experimental work investigating the enhancement of word learning from the presence of social cues has generally been somewhat limited in scope. The methodologies implemented in these studies have primarily focused on the associations between social cues and labels, and have generally not focused on the continuous nature of linguistic input across time or within the structure of discourse topics. Here, we investigated whether social cues are beneficial in natural learning scenarios by examining whether caregivers produced eye-gaze cues systematically across discourse topics in interactions with their children, and whether some of these cues modulated the synchronicity of neural activity within an interaction. Specifically, engaging in joint attention earlier in a conversation enhanced the average neural synchrony of a dyad across

time. We also found that the timing of referential gaze toward an object of discussion impacted word learning, and that this relationship was mediated by the level of synchronicity when the referent label was spoken. These findings suggest that, despite the information-density and messiness of natural learning environments, social cues still serve to reduce referential ambiguity across child-directed discourse topics.

General Discussion

The goal of this dissertation was to explore how social information provided by caregivers influences word learning in naturalistic learning environments. Prior research has been primarily limited to experimental sessions that simplify the learning problem, but here, I endeavored to investigate the mechanisms that explain the relationship between social cues and word learning in more natural, complex settings. I presented novel evidence showing that social cues are presented in a systematic fashion in natural settings, particularly in the structure of discourse topics. Furthermore, I identified mutual gaze and referential gaze specifically as informative cues that enhance children's understanding of discourse topics and support their word learning.

How social cues help children learn new words

A wealth of research has investigated how children learn novel words from observing another's social cues. Social cues that have been most prominently considered include referential eye-gaze, deictic gestures such as pointing, and representational gestures (i.e., using gestures to represent meaning). But what is the functional relationship that ties social information to enhanced word learning? In Chapter 1, I reviewed prior research to determine why social cues lead to better word learning for young children. I concluded that social cues provided both perceptual *and* social information, and served five distinct, but related, functions. This suggests that social cues are not a uniform source of information for children's learning. Depending on the context of a learning situation, the age of the child, the type of cue observed (amongst other factors), and the learning outcomes may shift and change. In subsequent chapters, I endeavored to characterize these changes in learning outcomes by examining how social cues may (or may

not) facilitate learning in settings more akin to typical learning environments. More specifically, I focused on how social cues may facilitate learning across more complex linguistic contexts, such as discourse topics.

Learning words from social cues across discourse topics

The generalizability of experimental studies demonstrating an improvement in children's word learning from observing caregiver social cues has been under debate. Seminal studies such as Frank, Tenenbaum, and Fernald (2013) suggested that the utility of these cues were low in typical learning environments, as they found these cues were likely unable to fully resolve referential ambiguity in their data. However, this study did not fully consider that real, dynamic environments contain statistical regularities that a young child can take advantage of to learn. In this dissertation, I examined whether accounting for the structure of linguistic information within discourse topics allows children to utilize social cues more readily in service of novel word learning. In Chapter 2, I presented 4- to 5-year-old children with novel objects, and provided a surrounding discourse that varied when labeling information was presented, and whether the novel labels were paired with social cue information (referential gaze). I found that children were more likely to learn the novel words when labels and gaze information were presented together. This suggests that children are able to learn new words across a somewhat protracted timescale, such as a discourse topic, lending credence to the idea that social cues may still hold utility in the word learning process when linguistic information is spread across time.

To further test this idea, I examined whether caregivers systematically produce social cues over discourse topics in Chapter 3. I coded caregiver-child play interactions for discourse topics and caregiver eye-gaze (referential gaze to the object of discussion and to the child).

Indeed, I found that caregivers provided gaze cues systematically across time. Caregivers produced more referential gaze cues at the beginning of a given discourse topic, compared to later. By contrast, caregivers were more likely to look towards their child at the end of discourse topics. This patterning suggests that caregivers produce referential gaze cues to introduce a new object of discussion and scaffold the content of the interaction to the ease of comprehension for the child. At the end of discourse topics, they tended to look at their child presumably to maintain their child's attention or to confirm their understanding. These descriptive results suggest that social cues may be helpful for children's learning of new words within discourse topics. If children are able to integrate their caregiver's cues to identify the content of an interaction, they may capitalize on subsequent learning moments embedded within the discourse topic.

The neural basis of learning words from social cues across discourse topics

As described in Chapter 1, social cues can serve different purposes in service of word learning: orienting attention, extracting information, disambiguating referents, and indicating referential intent. Prior neuroimaging work has demonstrated an additional function of observing another's behavior in an interaction: to synchronize neural activity, i.e., neural synchrony. Studies with adults have determined that this alignment in neural activity is associated with shared understanding of a conversation or story (Sievers et al., 2024; Stephens, Silbert, & Hasson, 2010). Even further, studies examining the interactions of adult-adult dyads and adult-child dyads have found that observing the behavior of your partner enhances neural synchrony (Nguyen, Abney, et al., 2021; Piazza, Hasenfratz, Hasson, & Lew-Williams, 2020). In Chapter 4, I extended these findings to examine social cues and word learning across natural

discourse topics, and recorded children's and caregivers' neural activity using a dual-brain fNIRS system. In Experiment 3, I found that engaging in mutual gaze at the beginning of a discourse topic enhanced overall neural synchrony within that conversation. In Experiment 4, I tested whether neural synchrony explains the effect of social cues on children's word learning. I found evidence for a particular pathway that may lead to enhanced word learning from observing referential gaze (though not for other social cues). Specifically, the relationship between when referential gaze is seen in a discourse topic and children's word learning performance was mediated by the level of neural synchrony at the moments when labels were presented. This means that child and caregiver brains became synchronized when novel labels were produced, and that subsequent referential gaze from the caregiver facilitated better learning of novel words.

These studies converge with prior findings and expand our understanding of children's word learning in several ways. First, I found differential utility of the social cues considered in current experiments, suggesting that not all social cues are equally beneficial in understanding a conversation or in helping word learning. This generally converges with the conclusions of Frank et al. (2013). However, my results suggest that a more nuanced perspective of these cues are necessary in order to gain a comprehensive understanding of how observing the behavior of others influences young children's understanding of communicative scenarios and impacts what they learn. While examining the temporal structure of information to gain insights into child development is not new (e.g., McMurray, 2016; Piazza, Nenchewa, & Lew-Williams, 2021), the current results are novel in that I specifically accounted for and centralized discourse structures into this area of research. Consideration of discourse topics is imperative in order to expand our knowledge of how natural conversation impacts the word learning process. Thus, in future work,

it remains critical to consider the inherent structure and regularities of child-directed input in order to acquire an accurate, holistic understanding of how children learn words.

Limitations and future research

The novel results presented in this dissertation further our understanding of how behavioral cues from caregivers influence the word learning process in dynamic, naturalistic environments. However, further research is needed to uncover how exactly these cues are situated in the learning process. For example, in Chapters 2 and 3, I explored a number of age effects. I found in Chapter 2 that older children learn co-occurring social cues and labels differently from younger children, namely that older children were able to learn from co-occurring cues at both the beginning *and* end of discourse topics. I concluded that child age, and their associated cognitive abilities (e.g., working memory, discourse comprehension) may influence what information they glean from different learning situations, such as when social cues and labels were presented. In Chapter 3, I saw that caregivers scaffold interactions to the age of the child by changing the presentation of their social cues. I found that caregivers of older children presented social cues in different ways from caregivers of younger children. For example, caregivers of older children were less likely to present referential gaze cues within a discourse topic compared to caregivers of younger children, suggesting that caregivers of younger children are scaffolding interactions more often. These results converge with prior work on caregiver scaffolding (Gogate, Maganti, & Laing, 2013; Masek et al., 2021; Puccini et al., 2010). How scaffolding unfolds, and whether it is necessary, depends on the development of the child. As children develop their cognitive systems and social competencies, the type of scaffolding, and whether it is needed, is likely to change. Thus, longitudinal studies examining

how caregivers change their production of social cues in response to their child's development will clarify how these cues scaffold interactions at different age points, and their utility in the learning process.

Another limitation that needs to be addressed in future work is how word learning from social cues is conceptualized. In this dissertation, I assessed word learning from social cues in forms typically referred to as “fast-mapping” scenarios, where labels and referents are initially associated in memory in a first encounter, and learning was assessed in a test phase via selection of a target referent amongst two options (Carey, 1978). However, this fast-mapping scenario is likely not reflective of the actual word learning process. Contemporary work has considered the importance of viewing word learning as a culmination of learning events rather than as a singular exposure, referred to as “slow-mapping” (Carey, 1978; Swingley, 2010; Wojcik, Zettersten, & Benitez, 2022). Studies examining slow-mapping find that different underlying mechanisms are likely to influence this learning (Horst, Parsons, & Bryon, 2011; McMurray, Horst, and Samuelson, 2012; Munro et al., 2012; Twomey, 2013). Thus, it is critical to consider what actually is “learning” and how best to measure this in order to isolate the actual effect of observing social cues during the learning process. For example, Wojcik, Zettersten, & Benitez (2022) suggest one potential way to expand our view of word learning is to consider what aspects of the learning context learners absorb. They note that when children encounter new words, they encode not only labeling information, but also knowledge about semantic, perceptual, and collocational similarity (Fourtassi et al., 2020; Peters & Borovsky, 2019; Unger & Fisher, 2021; Wojcik & Saffran, 2015) and incorporate speaker characteristics (Brooker & Poulin-Dubois, 2013). It is possible that social cues not only assist learning in the form of “fast-mapping”, but also serve to enhance children's learning of contextual information

surrounding labels, such as increasing their perceptual knowledge of a novel object. Indeed, Hoehl et al. (2012) found that eye-gaze cues influenced 4-month-old infants' processing of novel objects, where caregiver eye-gaze made object-processing easier for the infants. It could be the case that this facilitation of processing positively impacts word learning. Future work should investigate a broader perspective of word learning to examine how social cues may influence not only the acquisition of labels, but how language-related information is processed and retained.

It is also important to consider that children incorporate behavioral information from their caregiver differently depending on the (a-)typicality of their development, as well as their learning environments. For example, research comparing autistic children (AD) and typically-developing children (TD) has uncovered differences in how AD and TD children utilize social cues, particularly for eye-gaze and joint attention cues. This line of research has found that AD children rely on social cues less often compared to TD children (Baron-Cohen et al., 1997; Liu et al., 2021; Pierce, Glad, & Schreibman, 1997; Senju, Tojo, Yaguchi & Hasegawa, 2005; Senju, Yaguchi, Tojo & Hasegawa, 2003; Stallworthy et al., 2021). However, depending on the severity of their diagnosis, AD children still possess language skills on par with mental age-matched TD children (Chan et al., 2005; Charman et al., 2003; McDuffie, Yoder, & Stone, 2005; Preissler & Carey, 2005; Tager-Flusberg, 2006). Parish-Morris et al. (2007) examined whether AD and TD children differ in their reliance on social or attentional information to learn new words. They found that variation in the AD group's reliance on social information was the most powerful predictor of their word learning, but not for the TD group, and concluded that word learning cannot be fully explained by children's reliance on social information. Thus, to further elucidate how social cues may be used by young learners, consideration should be given as to whether and/or how children actually incorporate social information when forming

hypotheses about word meaning. This, for example, could enrich what is currently known about how the timing of social cues in a discourse topic impacts comprehension and word learning.

Additionally, it will be important to consider how different learning environments impact the utility of social cues in service of word learning. For example, in this dissertation, I recruited only monolingual participants, but some studies have reported that bilingual children seem to have a heightened sensitivity to social cues when learning new words, relative to monolinguals (Brojde, Ahmed, & Colunga, 2012; Groba et al., 2018; Yow & Markman, 2011; Yow et al., 2017). For example, Brojde, Ahmed, and Colunga (2012) examined what types of information sources monolingual and bilingual children rely on to learn new words. Participants were introduced to a novel object (e.g., *zuly*). During the experimental phase, participants were asked to generalize this novel object label to other novel objects. In four conditions, the experimenters varied whether participants could use eye-gaze as a social cue and similarity along featural dimensions as an object cue to identify the other *zuly*. In the crucial condition where social cues and object cues conflicted, monolingual children attended to object cues more often, while bilingual children attended to social cues more often. They suggest that this result is indicative of certain demands in learning two languages; specifically, social cues may be more beneficial to bilingual children (vs. monolingual children) for accurately identifying the referent of a label across two languages.

Similarly, there may be variation across households in the frequency and timing of social cues, such that some children have the opportunity to exploit caregiver social cues consistently across developmental time, while others do not. Cross-cultural investigations have already noted significant differences in the information presented during caregiver-child interactions across non-WEIRD communities. For example, caregivers in Papuan, Testal Mayan, and Tsimane

communities generally produce less child-directed speech compared to Western communities (e.g., Casillas et al., 2020, 2021; Cristia, Dupoux Gurven, & Stieglitz, 2019). Prior ethnographic research has found that these communities, in comparison to Western countries, are centered primarily around adults and their work, with potentially less frequent child-directed speech input from caregivers and minimal play interactions (Casillas et al., 2021; Gaskins, 2000). Thus, it is possible that differences in linguistic communication carry over to differences in the frequency or amount of caregiver social cues as well.

Moreover, cultural norms across communities could dictate the usefulness of some social cues over others in service of word learning. For example, research has documented cultural differences in face scanning among young children (and adults), such that Japanese children are more likely to look towards others' eyes, while British children are more likely to look towards the mouth (Morin-Lessard et al., 2019; Senju, Vermetti, Kikuchi, Akechi, & Hasegawa, 2013; see also Haensel, Ishikawa, Itakura, Smith, & Senju, 2020). Sekiyama, Hisanaga, and Mugitani (2021) related this difference in looks to language development. They found that 2- and 3-year-old Japanese children's tendencies to gaze toward the eyes of another were positively related to larger vocabularies. In comparison, Tenenbaum et al. (2015) found that Western children's attention to the mouth was related with later vocabulary outcomes. These differences suggest that the cultural norms present in communities may influence which social cues children find useful in determining word meaning.

Thus, variations of the learning environment – target language(s), cultural norms, frequency and timing of social cues – may set children on different language learning trajectories, which calls for investigations of the early foundations of communication between caregivers and infants. The current work suggests that social cues are a potentially lucrative

source of information in light of the information-dense and continuous nature of discourse and interactions. However, determining the generalizability of these results is limited, as the majority of participants recruited here were from White, highly educated, mid-high SES households (Simons, Shoda, & Lindsay, 2017). It is possible that for children growing up in particular environments (e.g., bilingual vs. monolingual homes, or homes with many vs. fewer caregiver interactions), social cues may play a more or less important role in their language learning. For example, children growing up in households that present limited interactions with their caregiver may mean that social cues are not a determining factor of their language development. As such, they may rely on other sources of information to find meaning (e.g., timing, physical proximity to objects, Deák, Krasno, Triesch, Lewis & Sepeta, 2014; Yu & Smith, 2012). Determining how the utility of social cues changes depending on the learning environment will not only help elucidate the role of these cues during the word learning process, but will also clarify how social information constrains and supports children's language acquisition.

Conclusion

Using experimental, descriptive, and dual-brain neuroscience approaches, the studies presented in this dissertation indicate that social cues provide rich information that supports children's word learning in dynamic, natural interactions. I found that: 1) children can use social cues to learn new words across discourse topics, 2) caregivers systematically use social cues across discourse topics during natural play interactions, 3) the presence of mutual gaze early on in a discourse topic supports caregiver-child neural synchrony, and 4) neural synchrony partially explains the relationship between the timing of referential gaze cues and word learning during natural discourse. By examining the temporal distribution of social cues across the inherent

structure of child-directed input, I found novel evidence that social cues support children's learning in diverse ways.

References

- Abney, D. H., Suanda, S. H., Smith, L. B., & Yu, C. (2020). What are the building blocks of parent–infant coordinated attention in free-flowing interaction? *Infancy*, 25(6), 871–887.
<https://doi.org/10.1111/inf.12365>
- Abu-Zhaya, R., Seidl, A., & Cristia, A. (2017). Multimodal infant-directed communication: How caregivers combine tactile and linguistic cues. *Journal of Child Language*, 44(5), 1088–1116. <https://doi.org/10.1017/S0305000916000416>
- Adams, A. M., & Gathercole, S. E. (2000). Limitations in working memory: Implications for language development. *International Journal of Language & Communication Disorders*, 35(1), 95–116.
- Adamson, L. B., Bakeman, R., & Deckner, D. F. (2004). The development of symbol-infused joint engagement. *Child Development*, 75(4), 1171–1187.
<https://doi.org/10.1111/j.1467-8624.2004.00732.x>
- Akechi, H., Kikuchi, Y., Tojo, Y., Osanai, H., & Hasegawa, T. (2013). Brief report: Pointing cues facilitate word learning in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 43(1), 230–235.
<https://doi.org/10.1007/s10803-012-1555-3>
- Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development*, 67(2), 635–645.
<https://doi.org/10.1111/j.1467-8624.1996.tb01756.x>
- Akhtar, N., & Gernsbacher, M. A. (2007). Joint attention and vocabulary development: A critical look. *Language and Linguistics Compass*, 1(3), 195–207.
<https://doi.org/10.1111/j.1749-818X.2007.00014.x>

- Akhtar, N., & Tomasello, M. (1998). Intersubjectivity in early language learning and use. In *Intersubjective communication and emotion in early ontogeny* (pp. 316–335). Cambridge University Press.
- Akhtar, N., & Tomasello, M. (2000). The social nature of words and word learning. In R. Golinkoff & K. Hirsh-Pasek (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 115–135). Oxford University Press.
- Allen, S. (2000). A discourse-pragmatic explanation for argument representation in child Inuktitut. *Linguistics*, 38(3), 483-521. <https://doi.org/10.1515/ling.38.3.483>
- Alonso, A., McDorman, S. A., & Romeo, R. R. (2024). How parent–child brain-to-brain synchrony can inform the study of child development. *Child Development Perspectives*, 18(1), 26–35. <https://doi.org/10.1111/cdep.12494>
- Ambrosini, E., Reddy, V., Looper, A. de, Costantini, M., Lopez, B., & Sinigaglia, C. (2013). Looking ahead: Anticipatory gaze and motor ability in infancy. *PLOS ONE*, 8(7), e67916. <https://doi.org/10.1371/journal.pone.0067916>
- Aussems, S., & Kita, S. (2019). Seeing iconic gestures while encoding events facilitates children’s memory of these events. *Child Development*, 90(4), 1123–1137. <https://doi.org/10.1111/cdev.12988>
- Axelsson, E., Churchley, K., & Horst, J. (2012). The right thing at the right time: Why ostensive naming facilitates word learning. *Frontiers in Psychology*, 3, 88. <https://doi.org/10.3389/fpsyg.2012.00088>
- Azhari, A., Truzzi, A., Neoh, M. J.-Y., Balagtas, J. P. M., Tan, H. H., Goh, P. P., Ang, X. A.,

- Setoh, P., Rigo, P., Bornstein, M. H., & Esposito, G. (2020). A decade of infant neuroimaging research: What have we learned and where are we going? *Infant Behavior and Development*, 58, 101389. <https://doi.org/10.1016/j.infbeh.2019.101389>
- Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience*, 4(10), 829-839.
- Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, 62(5), 875–890. <https://doi.org/10.1111/j.1467-8624.1991.tb01577.x>
- Baldwin, D. A. (1993). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology*, 29(5), 832–843. <https://doi.org/10.1037/0012-1649.29.5.832>
- Baldwin, D. A. (1995). Understanding the link between joint attention and language. In *Joint attention: Its origins and role in development* (pp. 131–158). Lawrence Erlbaum Associates, Inc.
- Baldwin, D. A., Markman, E. M., Bill, B., Desjardins, R. N., Irwin, J. M., & Tidball, G. (1996). Infants' reliance on a social criterion for establishing word-object relations. *Child Development*, 67(6), 3135–3153. <https://doi.org/10.1111/j.1467-8624.1996.tb01906.x>
- Baldwin, D. A., & Tomasello, M. (1998). Word learning: A window on early pragmatic understanding. In *The proceedings of the twenty-ninth annual child language research forum* (pp. 3–23). Center for the Study of Language and Information.
- Bannard, C., & Tomasello, M. (2012). Can we dissociate contingency learning from social learning in word acquisition by 24-month-olds? *PLOS ONE*, 7(11), e49881. <https://doi.org/10.1371/journal.pone.0049881>
- Baron-Cohen, S., Baldwin, D. A., & Crowson, M. (1997). Do children with autism use the

- speaker's direction of gaze strategy to crack the code of language? *Child Development*, 68(1), 48–57. <https://doi.org/10.2307/1131924>
- Baron-Cohen, S., & Cross, P. (1992). Reading the eyes: Evidence for the role of perception in the development of a theory of mind. *Mind & Language*, 7(1-2), 172-186.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bates, J. E. (1976). Effects of children's nonverbal behavior upon adults. *Child Development*, 47(4), 1079–1088. <https://doi.org/10.2307/1128445>
- Behne, T., Liszkowski, U., Carpenter, M., & Tomasello, M. (2012). Twelve-month-olds' comprehension and production of pointing. *British Journal of Developmental Psychology*, 30(3), 359–375. <https://doi.org/10.1111/j.2044-835X.2011.02043.x>
- Beier, J. S., & Carey, S. (2014). Contingency is not enough: Social context guides third-party attributions of intentional agency. *Developmental Psychology*, 50(3), 889–902. <https://doi.org/10.1037/a0034171>
- Berman, J. M. J., Graham, S. A., Callaway, D., & Chambers, C. G. (2013). Preschoolers use emotion in speech to learn new words. *Child Development*, 84(5), 1791–1805. <https://doi.org/10.1111/cdev.12074>
- Bloom, L., Margulis, C., Tinker, E., & Fujita, N. (1996). Early conversations and word learning: contributions from child and adult. *Child Development*, 67(6), 3154–3175. <https://doi.org/10.1111/j.1467-8624.1996.tb01907.x>
- Bloom, P. (2002). *How children learn the meanings of words*. MIT Press.
- Boas, D. A., Elwell, C. E., Ferrari, M., & Taga, G. (2014). Twenty years of functional

- near-infrared spectroscopy: Introduction for the special issue. *NeuroImage*, 85, 1–5.
<https://doi.org/10.1016/j.neuroimage.2013.11.033>
- Booth, A. E., McGregor, K. K., & Rohlfing, K. J. (2008). Socio-Pragmatics and attention: Contributions to gesturally guided word learning in toddlers. *Language Learning and Development*, 4(3), 179–202. <https://doi.org/10.1080/15475440802143091>
- Boundy, L., Cameron-Faulkner, T., & Theakston, A. (2016). Exploring early communicative behaviours: A fine-grained analysis of infant shows and gives. *Infant Behavior and Development*, 44, 86–97. <https://doi.org/10.1016/j.infbeh.2016.06.005>
- Bower, M., & Hedberg, J. G. (2010). A quantitative multimodal discourse analysis of teaching and learning in a web-conferencing environment – The efficacy of student-centred learning designs. *Computers & Education*, 54(2), 462–478.
<https://doi.org/10.1016/j.compedu.2009.08.030>
- Boyce, K. C. M. (2021). *Can you see what I mean? The effects of iconic gestures on early word learning*. [Masters thesis, Dalhousie University]. Dalhousie University Repository.
<http://hdl.handle.net/10222/80330>
- Bradshaw, J., Fu, X., Yurkovic-Harding, J., & Abney, D. (2023). Infant embodied attention in context: Feasibility of home-based head-mounted eye tracking in early infancy. *Developmental Cognitive Neuroscience*, 64, 101299.
<https://doi.org/10.1016/j.dcn.2023.101299>
- Briganti, A. M., & Cohen, L. B. (2011). Examining the role of social cues in early word learning. *Infant Behavior and Development*, 34(1), 211–214.
<https://doi.org/10.1016/j.infbeh.2010.12.012>
- Brojde, C. L., Ahmed, S., & Colunga, E. (2012). Bilingual and monolingual children attend to

- different cues when learning new words. *Frontiers in Psychology*, 3, 155.
- Brooker, I., & Poulin-Dubois, D. (2013). Is a bird an apple? the effect of speaker labeling accuracy on infants' word learning, imitation, and helping behaviors. *Infancy*, 18(s1), E46–E68. <https://doi.org/10.1111/infa.12027>
- Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, 8(6), 535–543.
<https://doi.org/10.1111/j.1467-7687.2005.00445.x>
- Brooks, R., & Meltzoff, A. N. (2008). 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(1), 207–220.
<https://doi.org/10.1017/S030500090700829X>
- Bruner, J. (1983). Play, Thought, and Language. *Peabody Journal of Education*, 60(3), 60–69.
- Butler, S. C., Caron, A. J., & Brooks, R. (2000). Infant understanding of the referential nature of looking. *Journal of Cognition and Development*, 1(4), 359–377.
https://doi.org/10.1207/S15327647JCD0104_01
- Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, 27(6), 843–873.
- Capone, N. C., & McGregor, K. K. (2005). The effect of semantic representation on toddlers' word retrieval. *Journal of Speech, Language, and Hearing Research*, 48(6), 1468–1480.
[https://doi.org/10.1044/1092-4388\(2005/102\)](https://doi.org/10.1044/1092-4388(2005/102))
- Carey, S. (1978). The child as word learner. *Linguistic Theory and Psychological Reality*.
- Caron, A. J., Caron, R. F., Caldwell, R. C., & Weiss, S. J. (1973). Infant perception of the

structural properties of the face. *Developmental Psychology*, 9(3), 385–399.

<https://doi.org/10.1037/h0034910>

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, 63(4), i–174.

<https://doi.org/10.2307/1166214>

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, 110(28), 11278–11283.

<https://doi.org/10.1073/pnas.1309518110>

Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a Tseltal Mayan village. *Child Development*, 91(5), 1819–1835.

Casillas, M., Brown, P., & Levinson, S. C. (2021). Early language experience in a Papuan community. *Journal of Child Language*, 48(4), 792–814.

Çetinçelik, M., Rowland, C. F., & Snijders, T. M. (2021). Do the eyes have it? A systematic review on the role of eye gaze in infant language development. *Frontiers in Psychology*, 11, 3627. <https://doi.org/10.3389/fpsyg.2020.589096>

Chan, A. S., Cheung, J., Leung, W. W. M., Cheung, R., & Cheung, M. (2005). Verbal expression and comprehension deficits in young children with autism. *Focus on Autism and Other Developmental Disabilities*, 20(2), 117–124.

<https://doi.org/10.1177/10883576050200020201>

Chang, L., de Barbaro, K., & Deák, G. (2016). Contingencies between infants' gaze, vocal, and

manual actions and mothers' object-naming: longitudinal changes from 4 to 9 months. *Developmental Neuropsychology*, 41(5–8), 342–361.

<https://doi.org/10.1080/87565641.2016.1274313>

Chang, L. M., & Deák, G. O. (2019). Maternal discourse continuity and infants' actions organize 12-month-olds' language exposure during object play. *Developmental Science*, 22(3), e12770. <https://doi.org/10.1111/desc.12770>

Chapman, R., Miller, J., MacKenzie, H., & Bedrosian, J. (1981). *The development of discourse skills in the second year of life*. University of British Columbia, Vancouver, Canada.

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Drew, A., & Cox, A. (2003). Predicting language outcome in infants with autism and pervasive developmental disorder. *International Journal of Language & Communication Disorders*, 38(3), 265–285. <https://doi.org/10.1080/136820310000104830>

Clancy, P. M. (2004). The discourse basis of constructions: Some evidence from Korean acquisition. In *Proceedings of the 32nd Stanford Child Language Research Forum* (pp. 20-29).

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>

Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review*, 112(2), 347–382. <https://doi.org/10.1037/0033-295X.112.2.347>

Cook, S. W. (2018). Chapter Four - Enhancing learning with hand gestures: Potential mechanisms. In K. D. Federmeier (Ed.), *Psychology of Learning and Motivation* (Vol. 69, pp. 107–133). Academic Press. <https://doi.org/10.1016/bs.plm.2018.10.001>

- Corkum, V., & Moore, C. (1998). The origins of joint visual attention in infants. *Developmental Psychology*, 34, 28–38. <https://doi.org/10.1037//0012-1649.34.1.28>
- Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2019). Child-directed speech is infrequent in a forager-farmer population: A time allocation study. *Child Development*, 90(3), 759–773.
- Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. *Processes of change in brain and cognitive development. Attention and performance XXI*, 21, 249–274.
- Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1567), 1149–1157.
- Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, 13(4), 148–153. <https://doi.org/10.1016/j.tics.2009.01.005>
- Csibra, G., & Volein, Á. (2008). Infants can infer the presence of hidden objects from referential gaze information. *British Journal of Developmental Psychology*, 26(1), 1–11. <https://doi.org/10.1348/026151007X185987>
- Custode, S. A., & Tamis-LeMonda, C. (2020). Cracking the code: Social and contextual cues to language input in the home environment. *Infancy*, 25(6), 809–826.
- de Barbaro, K., Johnson, C. M., Forster, D., & Deák, G. O. (2016). Sensorimotor decoupling contributes to triadic attention: A longitudinal investigation of mother–infant–object interactions. *Child Development*, 87(2), 494–512. <https://doi.org/10.1111/cdev.12464>
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47, 1–12.
- de Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and*

Gesture (pp. 284–311).

de Villiers Rader, N., & Zukow-Goldring, P. (2010). How the hands control attention during early word learning. *Gesture*, 10(2–3), 202–221.

<https://doi.org/10.1075/gest.10.2-3.05rad>

Deák, G. O., Flom, R. A., & Pick, A. D. (2000). Effects of gesture and target on 12- and 18-month-olds' joint visual attention to objects in front of or behind them. *Developmental Psychology*, 36(4), 511–523. <https://doi.org/10.1037/0012-1649.36.4.511>

Deák, G. O., Krasno, A. M., Jasso, H., & Triesch, J. (2018). What leads to shared attention? Maternal cues and infant responses during object play. *Infancy*, 23(1), 4–28.

<https://doi.org/10.1111/infa.12204>

Deák, G. O., Krasno, A. M., Triesch, J., Lewis, J., & Sepeta, L. (2014). Watch the hands: Infants can learn to follow gaze by seeing adults manipulate objects. *Developmental Science*, 17(2), 270–281. <https://doi.org/10.1111/desc.12122>

Deák, G. O., Walden, T. A., Yale Kaiser, M., & Lewis, A. (2008). Driven from distraction: How infants respond to parents' attempts to elicit and re-direct their attention. *Infant Behavior and Development*, 31(1), 34–50. <https://doi.org/10.1016/j.infbeh.2007.06.004>

Delgado, C. E. F., Mundy, P., Crowson, M., Markus, J., Yale, M., & Schwartz, H. (2002). Responding to joint attention and language development. *Journal of Speech, Language, and Hearing Research*, 45(4), 715–719. [https://doi.org/10.1044/1092-4388\(2002/057\)](https://doi.org/10.1044/1092-4388(2002/057))

Deligianni, F., Senju, A., Gergely, G., & Csibra, G. (2011). Automated gaze-contingent objects elicit orientation following in 8-months-old infants. *Developmental Psychology*, 47(6), 1499–1503. <https://doi.org/10.1037/a0025659>

D'Entremont, B. (2000). A perceptual–attentional explanation of gaze following in 3- and

- 6-month-olds. *Developmental Science*, 3(3), 302–311.
<https://doi.org/10.1111/1467-7687.00124>
- D'Entremont, B., Hains, S. M. J., & Muir, D. W. (1997). A demonstration of gaze following in 3- to 6-month-olds. *Infant Behavior and Development*, 20(4), 569–572.
[https://doi.org/10.1016/S0163-6383\(97\)90048-5](https://doi.org/10.1016/S0163-6383(97)90048-5)
- Desrochers, S., Morissette, P., & Ricard, M. (1995). Two perspectives on pointing in infancy. In *Joint attention: Its origins and role in development* (pp. 85–101). Lawrence Erlbaum Associates, Inc.
- Di Lorenzo, R., Pirazzoli, L., Blasi, A., Bulgarelli, C., Hakuno, Y., Minagawa, Y., & Brigadoi, S. (2019). Recommendations for motion correction of infant fNIRS data applicable to multiple data sets and acquisition systems. *NeuroImage*, 200, 511–527.
<https://doi.org/10.1016/j.neuroimage.2019.06.056>
- Doan, S. N. (2010). The role of emotion in word learning. *Early Child Development and Care*, 180(8), 1065–1078. <https://doi.org/10.1080/03004430902726479>
- Doherty, M. J., & Anderson, J. R. (1999). A new look at gaze: Preschool children's understanding of eye-direction. *Cognitive Development*, 14(4), 549–571.
- Doherty, M. J., Anderson, J. R., & Howieson, L. (2009). The rapid development of explicit gaze judgment ability at 3 years. *Journal of Experimental Child Psychology*, 104(3), 296–312.
- Dowman, M., Savova, V., Griffiths, T. L., Kording, K. P., Tenenbaum, J. B., & Purver, M. (2008). A probabilistic model of meetings that combines words and discourse features. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(7), 1238–1248.
<https://doi.org/10.1109/TASL.2008.925867>
- Edmunds, S. R., Rozga, A., Li, Y., Karp, E. A., Ibanez, L. V., Rehg, J. M., & Stone, W. L.

- (2017). Brief report: Using a point-of-view camera to measure eye gaze in young children with autism spectrum disorder during naturalistic social interactions: A pilot Study. *Journal of Autism and Developmental Disorders*, 47(3), 898–904.
<https://doi.org/10.1007/s10803-016-3002-3>
- Fais, L., Werker, J. F., Cass, B., Leibowich, J., Barbosa, A. V., & Vatikiotis-Bateson, E. (2012). Here's looking at you, baby: What gaze and movement reveal about minimal pair word-object association at 14 months. *Laboratory Phonology*, 3(1), 91–124.
<https://doi.org/10.1515/lp-2012-0007>
- Farrant, B. M., & Zubrick, S. R. (2012). Early vocabulary development: The importance of joint attention and parent-child book reading. *First Language*, 32(3), 343–364.
<https://doi.org/10.1177/0142723711422626>
- Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences*, 99(14), 9602–9605.
<https://doi.org/10.1073/pnas.152159999>
- Farroni, T., Johnson, M. H., & Csibra, G. (2004). Mechanisms of eye gaze perception during infancy. *Journal of Cognitive Neuroscience*, 16(8), 1320–1326.
<https://doi.org/10.1162/0898929042304787>
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59(5), i–185.
<https://doi.org/10.2307/1166093>
- Ferguson, B., & Lew-Williams, C. (2016). Communicative signals support abstract rule learning

- by 7-month-old infants. *Scientific Reports*, 6(1), 25434.
<https://doi.org/10.1038/srep25434>
- Flom, R., Deák, G. O., Phill, C. G., & Pick, A. D. (2004). Nine-month-olds' shared visual attention as a function of gesture and object location. *Infant Behavior and Development*, 27(2), 181–194. <https://doi.org/10.1016/j.infbeh.2003.09.007>
- Flom, R., Lee, K., & Muir, D. (2017). *Gaze-Following: Its Development and Significance*. Psychology Press.
- Floor, P., & Akhtar, N. (2006). Can 18-Month-Old infants learn words by listening in on conversations? *Infancy*, 9(3), 327–339. https://doi.org/10.1207/s15327078in0903_4
- Fourtassi, A., Bian, Y., & Frank, M. C. (2020). The growth of children's semantic and phonological networks: Insight from 10 languages. *Cognitive Science*, 44(7), e12847.
- Foushee, R., Srinivasan, M., & Xu, F. (2023). Active Learning in Language Development. *Current Directions in Psychological Science*, 32(3), 250–257.
<https://doi.org/10.1177/09637214221123920>
- Franchak, J. M., Kretch, K. S., & Adolph, K. E. (2018). See and be seen: Infant–caregiver social looking during locomotor free play. *Developmental Science*, 21(4), e12626.
<https://doi.org/10.1111/desc.12626>
- Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-mounted eye tracking: a new method to describe infant looking. *Child Development*, 82(6), 1738–1750. <https://doi.org/10.1111/j.1467-8624.2011.01670.x>
- 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), 1–24. <https://doi.org/10.1080/15475441.2012.707101>

- Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: a review using an integrative framework. *Psychological Bulletin*, 134(1), 31.
- Gaskins, S. (2000). Children's daily activities in a Mayan village: A culturally grounded description. *Cross-Cultural Research*, 34(4), 375-389.
- Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, 27(4), 513-543.
- Gerken, L., Balcomb, F. K., & Minton, J. L. (2011). Infants avoid 'labouring in vain' by attending more to learnable than unlearnable linguistic patterns. *Developmental Science*, 14(5), 972-979. <https://doi.org/10.1111/j.1467-7687.2011.01046.x>
- Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1(1), 3-55. https://doi.org/10.1207/s15327817la0101_2
- Gliga, T., & Csibra, G. (2009). One-year-old infants appreciate the referential nature of deictic gestures and words. *Psychological Science*, 20(3), 347-353. <https://doi.org/10.1111/j.1467-9280.2009.02295.x>
- Gliga, T., Elsabbagh, M., Hudry, K., Charman, T., & Johnson, M. H. (2012). Gaze following, gaze reading, and word learning in children at risk for autism. *Child Development*, 83(3), 926-938. <https://doi.org/10.1111/j.1467-8624.2012.01750.x>
- 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(4), 878-894.
- Gogate, L. J., Bolzani, L. H., & Betancourt, E. A. (2006). Attention to maternal multimodal naming by 6- to 8-month-old infants and learning of word-object relations. *Infancy*, 9(3), 259-288. https://doi.org/10.1207/s15327078in0903_1

- Gogate, L. J., Maganti, M., & Laing, K. B. (2013). Maternal naming of object wholes versus parts to preverbal infants: A fine-grained analysis of scaffolding at 6–8 months. *Infant Behavior and Development*, 36(3), 470–479. <https://doi.org/10.1016/j.infbeh.2013.03.012>
- Gogate, L. J., Walker-Andrews, A. S., & Bahrick, L. E. (2001). The intersensory origins of word-comprehension: An ecological–dynamic systems view. *Developmental Science*, 4(1), 1–18. <https://doi.org/10.1111/1467-7687.00143>
- Goldin-Meadow, S., & Singer, M. (2003). From children’s hands to adults’ ears: gesture’s role in the learning process. *Developmental Psychology*, 39, 509–520. <https://doi.org/10.1037/0012-1649.39.3.509>
- Golinkoff, R. M., & Hirsh-Pasek, K. (2006). Baby wordsmith: From associationist to social sophisticate. *Current Directions in Psychological Science*, 15(1), 30–33. <https://doi.org/10.1111/j.0963-7214.2006.00401.x>
- Goodrich, W., & Kam, C. L. H. (2009). Co-speech gesture as input in verb learning. *Developmental Science*, 12(1), 81–87. <https://doi.org/10.1111/j.1467-7687.2008.00735.x>
- Goodwyn, S. W., Acredolo, L. P., & Brown, C. A. (2000). Impact of symbolic gesturing on early language development. *Journal of Nonverbal behavior*, 24, 81–103.
- Goupil, L., Dautriche, I., Denman, K., Henry, Z., Haresign, I. M., & Wass, S. (2024). *Leader-follower dynamics during early social interactions matter for infant word learning*. PsyArXiv. <https://doi.org/10.31234/osf.io/86pxt>
- Graham, S. A., Nilsen, E. S., Collins, S., & Olineck, K. (2010). The role of gaze direction and mutual exclusivity in guiding 24-month-olds’ word mappings. *British Journal of Developmental Psychology*, 28(2), 449–465. <https://doi.org/10.1348/026151009X424565>
- Graham, S. A., Nilsen, E. S., & Nayer, S. (2017). Following the intentional eye: The role of

- gaze cues in early word learning. In *Gaze-Following* (pp. 193–216). Psychology Press.
- Grassmann, S., & Tomasello, M. (2010). Young children follow pointing over words in interpreting acts of reference. *Developmental Science*, 13(1), 252–263.
<https://doi.org/10.1111/j.1467-7687.2009.00871.x>
- Gredebäck, G., Fikke, L., & Melinder, A. (2010). The development of joint visual attention: A longitudinal study of gaze following during interactions with mothers and strangers. *Developmental Science*, 13(6), 839–848.
<https://doi.org/10.1111/j.1467-7687.2009.00945.x>
- Groba, A., Houwer, A. D., Mehnert, J., Rossi, S., & Obrig, H. (2018). Bilingual and monolingual children process pragmatic cues differently when learning novel adjectives. *Bilingualism: Language and Cognition*, 21(2), 384–402. <https://doi.org/10.1017/S1366728917000232>
- Haensel, J. X., Ishikawa, M., Itakura, S., Smith, T. J., & Senju, A. (2020). Cultural influences on face scanning are consistent across infancy and adulthood. *Infant Behavior and Development*, 61, 101503.
- Harris, M., Barlow-Brown, F., & Chasin, J. (1995). The emergence of referential understanding: pointing and the comprehension of object names. *First Language*, 15(1), 19–34. <https://doi.org/10.1177/014272379501500101>
- Harris, M., Jones, D., Brookes, S., & Grant, J. (1986). Relations between the non-verbal context of maternal speech and rate of language development. *British Journal of Developmental Psychology*, 4(3), 261–268. <https://doi.org/10.1111/j.2044-835X.1986.tb01017.x>
- Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends in Cognitive Sciences*, 16(2), 114–121.

- Hembacher, E., deMayo, B., & Frank, M. C. (2020). Children's social information seeking is sensitive to referential ambiguity. *Child Development, 91*(6), e1178–e1193.
<https://doi.org/10.1111/cdev.13427>
- Herold, D. S., Nygaard, L. C., Chicos, K. A., & Namy, L. L. (2011). The developing role of prosody in novel word interpretation. *Journal of Experimental Child Psychology, 108*(2), 229–241. <https://doi.org/10.1016/j.jecp.2010.09.005>
- Hirotsu, M., Stets, M., Striano, T., & Friederici, A. D. (2009). Joint attention helps infants learn new words: Event-related potential evidence. *NeuroReport, 20*(6), 600–605.
<https://doi.org/10.1097/WNR.0b013e32832a0a7c>
- Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., Yust, P. K. S., & Suma, K. (2015). The contribution of early communication quality to low-income children's language success. *Psychological Science, 26*(7), 1071–1083.
<https://doi.org/10.1177/0956797615581493>
- Hoehl, S., Reid, V. M., Parise, E., Handl, A., Palumbo, L., & Striano, T. (2009). Looking at eye gaze processing and its neural correlates in infancy—Implications for social development and autism spectrum disorder. *Child Development, 80*(4), 968–985.
- Hoehl, S., Wahl, S., Michel, C., & Striano, T. (2012). Effects of eye gaze cues provided by the caregiver compared to a stranger on infants' object processing. *Developmental Cognitive Neuroscience, 2*(1), 81–89. <https://doi.org/10.1016/j.dcn.2011.07.015>
- Hoehl, S., Wahl, S., & Pauen, S. (2014). Disentangling the effects of an adult model's eye gaze and head orientation on young infants' processing of a previously attended object. *Infancy, 19*(1), 53–64. <https://doi.org/10.1111/infa.12035>
- Hoff, E. (2010). Context effects on young children's language use: The influence of

- conversational setting and partner. *First Language*, 30(3–4), 461–472.
<https://doi.org/10.1177/0142723710370525>
- Hoff-Ginsberg, E. (1987). Topic relations in mother-child conversation. *First Language*, 7(20), 145–158. <https://doi.org/10.1177/014272378700702006>
- Hollich, G., Hirsh-Pasek, K., & Golinkoff, R. M. (2000). I. What does it take to learn a word? *Monographs of the Society for Research in Child Development*, 65(3), 1–16.
<https://doi.org/10.1111/1540-5834.00091>
- Horowitz, A. C., & Frank, M. C. (2015). Young children's developing sensitivity to discourse continuity as a cue for inferring reference. *Journal of Experimental Child Psychology*, 129, 84–97. <https://doi.org/10.1016/j.jecp.2014.08.003>
- Horst, J., Parsons, K., & Bryan, N. (2011). Get the story straight: Contextual repetition promotes word learning from storybooks. *Frontiers in Psychology*, 2, 17.
<https://doi.org/10.3389/fpsyg.2011.00017>
- Houston-Price, C., Plunkett, K., & Duffy, H. (2006). The use of social and salience cues in early word learning. *Journal of Experimental Child Psychology*, 95(1), 27–55.
<https://doi.org/10.1016/j.jecp.2006.03.006>
- Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology*, 43(6), 1447.
- Hutman, T., Harrop, C., Baker, E., Elder, L., Abood, K., Soares, A., & Jeste, S. S. (2016). Joint engagement modulates object discrimination in toddlers: A pilot electrophysiological investigation. *Social Neuroscience*, 11(5), 525–530.
<https://doi.org/10.1080/17470919.2015.1114966>
- Iacoboni, M., Lieberman, M. D., Knowlton, B. J., Molnar-Szakacs, I., Moritz, M., Throop, C. J.,

- & Fiske, A. P. (2004). Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. *NeuroImage*, 21(3), 1167–1173. <https://doi.org/10.1016/j.neuroimage.2003.11.013>
- Iverson, J. M., Capirci, O., Longobardi, E., & Cristina Caselli, M. (1999). Gesturing in mother-child interactions. *Cognitive Development*, 14(1), 57–75. [https://doi.org/10.1016/S0885-2014\(99\)80018-5](https://doi.org/10.1016/S0885-2014(99)80018-5)
- Jackson, E., Leitao, S., & Claessen, M. (2016). The relationship between phonological short-term memory, receptive vocabulary, and fast mapping in children with specific language impairment. *International Journal of Language & Communication Disorders*, 51(1), 61-73.
- Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best: Preschoolers use past reliability over age when learning new words. *Psychological Science*, 17(9), 757–758. <https://doi.org/10.1111/j.1467-9280.2006.01778.x>
- Kalagher, H., & Yu, C. (2006). *The Effects of Deictic Pointing in Word Learning*. International Conference of Development and Learning.
- Kandemir, S., Ozer, D., & Aktan Erciyes, A. (2023). Multimodal language in child-directed vs. Adult-directed speech. *Quarterly Journal of Experimental Psychology* (2006), 17470218231188832. <https://doi.org/10.1177/17470218231188832>
- Kapalková, S., Polišíenská, K., & Süssová, M. (2016). The role of pictures and gestures as a support mechanism for novel word learning: A training study with 2-year-old children. *Child Language Teaching and Therapy*, 32(1), 53–64. <https://doi.org/10.1177/0265659015585373>
- Kelly, J. F., Morisset, C. E., Barnard, K. E., Hammond, M. A., & Booth, C. L. (1996). The

influence of early mother—Child interaction on preschool cognitive/linguistic outcomes in a high-social-risk group. *Infant Mental Health Journal*, 17(4), 310–321.

[https://doi.org/10.1002/\(SICI\)1097-0355\(199624\)17:4<310::AID-IMHJ3>3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1097-0355(199624)17:4<310::AID-IMHJ3>3.0.CO;2-Q)

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review*, 124(3), 245–266.

<https://doi.org/10.1037/rev0000059>

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and Gesture* (pp. 162-185).

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, 16(12), 606–617.

<https://doi.org/10.1016/j.tics.2012.10.007>

Kopp, F., & Lindenberger, U. (2011). Effects of joint attention on long-term memory in 9-month-old infants: An event-related potentials study. *Developmental Science*, 14(4), 660–672. <https://doi.org/10.1111/j.1467-7687.2010.01010.x>

Kopp, F., & Lindenberger, U. (2012). Social cues at encoding affect memory in 4-month-old infants. *Social Neuroscience*, 7(5), 458–472.

<https://doi.org/10.1080/17470919.2011.631289>

Kosie, J. E., & Lew-Williams, C. (2023). *Infant-Directed Communication: Examining the many dimensions of everyday caregiver-infant interactions*. OSF.

<https://osf.io/preprints/osf/hyqp2>

Krogh-Jespersen, S., & Echols, C. H. (2012). The influence of speaker reliability on first versus second label learning. *Child Development*, 83(2), 581–590.

<https://doi.org/10.1111/j.1467-8624.2011.01713.x>

- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13).
<https://doi.org/10.18637/jss.v082.i13>
- Lavelli, M., & Fogel, A. (2005). Developmental changes in the relationship between the infant's attention and emotion during early face-to-face communication: The 2-month transition. *Developmental Psychology*, 41(1), 265–280.
<https://doi.org/10.1037/0012-1649.41.1.265>
- Law, B., Houston-Price, C., & Loucas, T. (2012). Using gaze direction to learn words at 18 months: relationships with later vocabulary. *Language Studies Working Papers*, 4, 3–14.
- Lee, C., & Lew-Williams, C. (2022). Speech and social cues combine at discourse boundaries to promote word learning. *Cognitive Development*, 64, 101254.
<https://doi.org/10.1016/j.cogdev.2022.101254>
- Lee, C., & Lew-Williams, C. (2023). The dynamic functions of social cues during children's word learning. *Infant and Child Development*, 32(1), e2372.
<https://doi.org/10.1002/icd.2372>
- Leong, V., Byrne, E., Clackson, K., Georgieva, S., Lam, S., & Wass, S. (2017). Speaker gaze increases information coupling between infant and adult brains. *Proceedings of the National Academy of Sciences*, 114(50), 13290–13295.
<https://doi.org/10.1073/pnas.1702493114>
- Lerner, G. H., Zimmerman, D. H., & Kidwell, M. (2011). Formal structure of practical tasks: A resource for action in the social life of very young children. In J. Streeck, C. Goodwin, & C. D. LeBaron (Eds.), *Embodied Interaction: Language and Body in the Material World* (pp. 44–58). Cambridge University Press.

- Liu, Q., Wang, Q., Li, X., Gong, X., Luo, X., Yin, T., ... & Yi, L. (2021). Social synchronization during joint attention in children with autism spectrum disorder. *Autism Research, 14*(10), 2120-2130.
- Liu, Y., Piazza, E. A., Simony, E., Shewokis, P. A., Onaral, B., Hasson, U., & Ayaz, H. (2017). Measuring speaker–listener neural coupling with functional near infrared spectroscopy. *Scientific Reports, 7*(1), Article 1. <https://doi.org/10.1038/srep43293>
- Lloyd-Fox, S., Blasi, A., & Elwell, C. E. (2010). Illuminating the developing brain: the past, present and future of functional near infrared spectroscopy. *Neuroscience & Biobehavioral Reviews, 34*(3), 269-284.
- Lucariello, J., & Nelson, K. (1986). Context effects on lexical specificity in maternal and child discourse. *Journal of Child Language, 13*(3), 507-522.
- Luft, C. D. B., Zioga, I., Giannopoulos, A., Di Bona, G., Binetti, N., Civilini, A., ... & Mareschal, I. (2022). Social synchronization of brain activity increases during eye-contact. *Communications Biology, 5*(1), 412.
- Lüke, C., & Ritterfeld, U. (2014). The influence of iconic and arbitrary gestures on novel word learning in children with and without SLI. *Gesture, 14*.
<https://doi.org/10.1075/gest.14.2.04luk>
- Mangardich, H., & Sabbagh, M. A. (2018). Children remember words from ignorant speakers but do not attach meaning: Evidence from event-related potentials. *Developmental Science, 21*(2), e12544. <https://doi.org/10.1111/desc.12544>
- Marcus, G. F., Fernandes, K. J., & Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychological Science, 18*(5), 387–391.
<https://doi.org/10.1111/j.1467-9280.2007.01910.x>

- Markman, E. M. (1992). Constraints on word learning: Speculations about their nature, origins, and domain specificity. In *Modularity and constraints in language and cognition* (pp. 59–101). Lawrence Erlbaum Associates, Inc.
- Markus, J., Mundy, P., Morales, M., Delgado, C. E., & Yale, M. (2000). Individual differences in infant skills as predictors of child-caregiver joint attention and language. *Social Development, 9*(3), 302-315.
- Marno, H., Guellai, B., Vidal, Y., Franzoi, J., Nespor, M., & Mehler, J. (2016). Infants' selectively pay attention to the information they receive from a native speaker of their language. *Frontiers in Psychology, 7*.
<https://www.frontiersin.org/articles/10.3389/fpsyg.2016.01150>
- Masek, L. R., McMillan, B. T. M., Paterson, S. J., Tamis-LeMonda, C. S., Golinkoff, R. M., & Hirsh-Pasek, K. (2021). Where language meets attention: How contingent interactions promote learning. *Developmental Review, 60*, 100961.
<https://doi.org/10.1016/j.dr.2021.100961>
- Maurer, D., & Salapatek, P. (1976). Developmental changes in the scanning of faces by young infants. *Child Development, 47*(2), 523–527. <https://doi.org/10.2307/1128813>
- McCathren, R. B., Yoder, P. J., & Warren, S. F. (1995). The role of directives in early language intervention. *Journal of Early Intervention, 19*(2), 91–101.
<https://doi.org/10.1177/105381519501900201>
- McDuffie, A., Yoder, P., & Stone, W. (2005). Prelinguistic predictors of vocabulary in young children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research, 48*(5), 1080–1097. [https://doi.org/10.1044/1092-4388\(2005/075\)](https://doi.org/10.1044/1092-4388(2005/075))
- McGregor, K. K. (2008). Gesture supports children's word learning. *International Journal of*

Speech-Language Pathology, 10(3), 112–117.

<https://doi.org/10.1080/17549500801905622>

McGregor, K. K., Rohlfing, K. J., Bean, A., & Marschner, E. (2009). Gesture as a support for word learning: The case of under. *Journal of Child Language*, 36(4), 807–828.

McMurray, B. (2016). Language at three timescales: The role of real-time processes in language development and evolution. *Topics in Cognitive Science*, 8(2), 393–407.

<https://doi.org/10.1111/tops.12201>

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, 119(4), 831–877. <https://doi.org/10.1037/a0029872>

Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences*, 108(22), 9014–9019. <https://doi.org/10.1073/pnas.1105040108>

Merriman, W. E., & Schuster, J. M. (1991). Young children's disambiguation of object name reference. *Child Development*, 62(6), 1288–1301. <https://doi.org/10.2307/1130807>

Michel, C., Wronski, C., Pauen, S., Daum, M. M., & Hoehl, S. (2019). Infants' object processing is guided specifically by social cues. *Neuropsychologia*, 126, 54–61.

<https://doi.org/10.1016/j.neuropsychologia.2017.05.022>

Miller, J. G., Vrtička, P., Cui, X., Shrestha, S., Hosseini, S. M. H., Baker, J. M., & Reiss, A. L. (2019). Inter-brain synchrony in mother-child dyads during cooperation: An fNIRS hyperscanning study. *Neuropsychologia*, 124, 117–124.

<https://doi.org/10.1016/j.neuropsychologia.2018.12.021>

Monticelli, M., Zeppa, P., Mammi, M., Penner, F., Melcarne, A., Zenga, F., & Garbossa, D.

- (2021). Where we mentalize: Main cortical areas involved in mentalization. *Frontiers in Neurology*, 12.
<https://www.frontiersin.org/journals/neurology/articles/10.3389/fneur.2021.712532>
- Moore, C. (2008). The development of gaze following. *Child Development Perspectives*, 2(2), 66–70. <https://doi.org/10.1111/j.1750-8606.2008.00052.x>
- Moore, C., Angelopoulos, M., & Bennett, P. (1999). Word learning in the context of referential and salience cues. *Developmental Psychology*, 35(1), 60–68.
<https://doi.org/10.1037/0012-1649.35.1.60>
- Moore, C., & Corkum, V. (1998). Infant gaze following based on eye direction. *British Journal of Developmental Psychology*, 16(4), 495–503.
<https://doi.org/10.1111/j.2044-835X.1998.tb00767.x>
- Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Messinger, D., Neal, R., & Schwartz, H. K. (2000). Responding to joint attention across the 6- through 24-month age period and early language acquisition. *Journal of Applied Developmental Psychology*, 21(3), 283–298. [https://doi.org/10.1016/S0193-3973\(99\)00040-4](https://doi.org/10.1016/S0193-3973(99)00040-4)
- Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Neal, R., & Schwartz, H. K. (2000). Gaze following, temperament, and language development in 6-month-olds: A replication and extension. *Infant Behavior and Development*, 23(2), 231–236.
[https://doi.org/10.1016/S0163-6383\(01\)00038-8](https://doi.org/10.1016/S0163-6383(01)00038-8)
- Morales, M., Mundy, P., & Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behavior and Development*, 21(2), 373–377.
[https://doi.org/10.1016/S0163-6383\(98\)90014-5](https://doi.org/10.1016/S0163-6383(98)90014-5)
- Morin-Lessard, E., Poulin-Dubois, D., Segalowitz, N., & Byers-Heinlein, K. (2019). Selective

- attention to the mouth of talking faces in monolinguals and bilinguals aged 5 months to 5 years. *Developmental Psychology*, 55(8), 1640.
- Morgan, J. K., Santosa, H., Conner, K. K., Fridley, R. M., Forbes, E. E., Iyengar, S., Joseph, H. M., & Huppert, T. J. (2023). Mother–child neural synchronization is time linked to mother–child positive affective state matching. *Social Cognitive and Affective Neuroscience*, 18(1), nsad001. <https://doi.org/10.1093/scan/nsad001>
- Mumford, K. H., & Kita, S. (2014). Children use gesture to interpret novel verb meanings. *Child Development*, 85(3), 1181–1189. <https://doi.org/10.1111/cdev.12188>
- Mundy, P., Block, J., Delgado, C., Pomares, Y., Hecke, A. V. V., & Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy. *Child Development*, 78(3), 938–954. <https://doi.org/10.1111/j.1467-8624.2007.01042.x>
- Mundy, P., & Gomes, A. (1998). Individual differences in joint attention skill development in the second year. *Infant Behavior and Development*, 21(3), 469–482. [https://doi.org/10.1016/S0163-6383\(98\)90020-0](https://doi.org/10.1016/S0163-6383(98)90020-0)
- Mundy, P., Kasari, C., Sigman, M., & Ruskin, E. (1995). Nonverbal communication and early language acquisition in children with down syndrome and in normally developing children. *Journal of Speech and Hearing Research*, 38, 157–167. <https://doi.org/10.1044/jshr.3801.157>
- Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current Directions in Psychological Science*, 16(5), 269–274. <https://doi.org/10.1111/j.1467-8721.2007.00518.x>
- Munro, N., Baker, E., McGregor, K., Docking, K., & Arciuli, J. (2012). Why word learning is not fast. *Frontiers in Psychology*, 3, 41. <https://doi.org/10.3389/fpsyg.2012.00041>

Nelson, K. (1988). Constraints on word learning? *Cognitive Development*, 3(3), 221–246.

[https://doi.org/10.1016/0885-2014\(88\)90010-X](https://doi.org/10.1016/0885-2014(88)90010-X)

Nguyen, T., Abney, D. H., Salamander, D., Bertenthal, B. I., & Hoehl, S. (2021). Proximity and touch are associated with neural but not physiological synchrony in naturalistic mother-infant interactions. *NeuroImage*, 244, 118599.

<https://doi.org/10.1016/j.neuroimage.2021.118599>

Nguyen, T., Hoehl, S., & Vrtička, P. (2021). A guide to parent-child fnirs hyperscanning data processing and analysis. *Sensors*, 21(12), Article 12. <https://doi.org/10.3390/s21124075>

Nguyen, T., Schleihau, H., Kayhan, E., Matthes, D., Vrtička, P., & Hoehl, S. (2021). Neural synchrony in mother–child conversation: Exploring the role of conversation patterns. *Social Cognitive and Affective Neuroscience*, 16(1–2), 93–102.

<https://doi.org/10.1093/scan/nsaa079>

Nicely, P., Tamis-LeMonda, C. S., & Bornstein, M. H. (1999). Mothers' attuned responses to infant affect expressivity promote earlier achievement of language milestones. *Infant Behavior and Development*, 22(4), 557–568.

[https://doi.org/10.1016/S0163-6383\(00\)00023-0](https://doi.org/10.1016/S0163-6383(00)00023-0)

Niedźwiecka, A., Ramotowska, S., & Tomalski, P. (2018). Mutual gaze during early mother–infant interactions promotes attention control development. *Child Development*, 89(6), 2230–2244. <https://doi.org/10.1111/cdev.12830>

Nurmsoo, E., & Bloom, P. (2008). Preschoolers' perspective taking in word learning: Do they blindly follow eye gaze? *Psychological Science*, 19(3), 211–215.

<https://doi.org/10.1111/j.1467-9280.2008.02069.x>

Nygaard, L. C., & Queen, J. S. (2008). Communicating emotion: Linking affective prosody and

- word meaning. *Journal of Experimental Psychology: Human Perception and Performance*, 34(4), 1017–1030. <https://doi.org/10.1037/0096-1523.34.4.1017>
- Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., & Itakura, S. (2013). The power of human gaze on infant learning. *Cognition*, 128(2), 127–133. <https://doi.org/10.1016/j.cognition.2013.03.011>
- O'Neill, M., Bard, K. A., Linnell, M., & Fluck, M. (2005). Maternal gestures with 20-month-old infants in two contexts. *Developmental Science*, 8(4), 352–359. <https://doi.org/10.1111/j.1467-7687.2005.00423.x>
- Özçalışkan, Ş., & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition*, 96(3), B101–B113. <https://doi.org/10.1016/j.cognition.2005.01.001>
- Parish-Morris, J., Hennon, E. A., Hirsh-Pasek, K., Golinkoff, R. M., & Tager-Flusberg, H. (2007). Children with autism illuminate the role of social intention in word learning. *Child Development*, 78(4), 1265–1287. <https://doi.org/10.1111/j.1467-8624.2007.01065.x>
- Paulus, M., & Fikkert, P. (2014). Conflicting social cues: Fourteen- and 24-month-old infants' reliance on gaze and pointing cues in word learning. *Journal of Cognition and Development*, 15(1), 43–59. <https://doi.org/10.1080/15248372.2012.698435>
- Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review*, 21(1), 178–185. <https://doi.org/10.3758/s13423-013-0466-4>
- Peters, R., & Borovsky, A. (2019). Modeling early lexico-semantic network development: Perceptual features matter most. *Journal of Experimental Psychology: General*, 148(4), 763.

- Peters, R. E., & Yu, C. (2020). Examining a developmental pathway of early word learning: From qualitative characteristics of parent speech, to sustained attention, to vocabulary size. *42nd Annual Meeting of the Cognitive Science Society*, 7.
- Phillips, E. A. M., Goupil, L., Whitehorn, M., Bruce-Gardyne, E., Csolsim, F. A., Marriott-Haresign, I., & Wass, S. V. (2023). Proactive or reactive? Neural oscillatory insight into the leader–follower dynamics of early infant–caregiver interaction. *Proceedings of the National Academy of Sciences*, 120(15), e2122481120.
<https://doi.org/10.1073/pnas.2122481120>
- Piazza, E. A., Cohen, A., Trach, J., & Lew-Williams, C. (2021). Neural synchrony predicts children’s learning of novel words. *Cognition*, 214, 104752.
<https://doi.org/10.1016/j.cognition.2021.104752>
- Piazza, E. A., Hasenfratz, L., Hasson, U., & Lew-Williams, C. (2020). Infant and adult brains are coupled to the dynamics of natural communication. *Psychological Science*, 31(1), 6–17. <https://doi.org/10.1177/0956797619878698>
- Piazza, E. A., Nenchewa, M. L., & Lew-Williams, C. (2021). The development of communication across timescales. *Current Directions in Psychological Science*, 30(6), 459–467. <https://doi.org/10.1177/09637214211037665>
- Pierce, K., Glad, K. S., & Schreibman, L. (1997). Social perception in children with autism: An attentional deficit? *Journal of Autism and Developmental Disorders*, 27(3), 265–282.
<https://doi.org/10.1023/A:1025898314332>
- Pomiechowska, B., & Csibra, G. (2020). *Nonverbal action interpretation guides novel word disambiguation in 12-month-olds*. PsyArXiv. <https://doi.org/10.31234/osf.io/tkhsp>
- Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects

- in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, 36, 717-731.
- Preissler, M. A., & Carey, S. (2005). The role of inferences about referential intent in word learning: Evidence from autism. *Cognition*, 97(1), B13–B23.
<https://doi.org/10.1016/j.cognition.2005.01.008>
- Pruden, S. M., Hirsh-Pasek, K., Golinkoff, R. M., & Hennon, E. A. (2006). The birth of words: Ten-month-olds learn words through perceptual salience. *Child Development*, 77(2), 266–280. <https://doi.org/10.1111/j.1467-8624.2006.00869.x>
- Puccini, D., Hassemer, M., Salomo, D., & Liszkowski, U. (2010). The type of shared activity shapes caregiver and infant communication. *Gesture*, 10(2–3), 279–296.
<https://doi.org/10.1075/gest.10.2-3.08puc>
- R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rączaszek-Leonardi, Joanna, Nomikou, I., & Rohlfing, K. (2013). Young children’s dialogical actions: The beginnings of purposeful intersubjectivity. *Autonomous Mental Development, IEEE Transactions On*, 5, 210–221.
<https://doi.org/10.1109/TAMD.2013.2273258>
- Rader, N., & Zukow-Goldring, P. (2012). Caregivers’ gestures direct infant attention during early word learning: The importance of dynamic synchrony. *Language Sciences*, 34, 559–568.
<https://doi.org/10.1016/j.langsci.2012.03.011>
- Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews. Neuroscience*, 20(8), 495–505.
<https://doi.org/10.1038/s41583-019-0179-4>

- Reddy, V., Markova, G., & Wallot, S. (2013). Anticipatory adjustments to being picked up in infancy. *PLOS ONE*, 8(6), e65289. <https://doi.org/10.1371/journal.pone.0065289>
- Reid, V. M., & Striano, T. (2005). Adult gaze influences infant attention and object processing: Implications for cognitive neuroscience. *European Journal of Neuroscience*, 21(6), 1763–1766. <https://doi.org/10.1111/j.1460-9568.2005.03986.x>
- Reid, V. M., Striano, T., Kaufman, J., & Johnson, M. H. (2004). Eye gaze cueing facilitates neural processing of objects in 4-month-old infants. *NeuroReport*, 15(16), 2553–2555.
- Reindl, V., Gerloff, C., Scharke, W., & Konrad, K. (2018). Brain-to-brain synchrony in parent-child dyads and the relationship with emotion regulation revealed by fNIRS-based hyperscanning. *NeuroImage*, 178, 493–502. <https://doi.org/10.1016/j.neuroimage.2018.05.060>
- Rohde, H., & Frank, M. C. (2014). Markers of topical discourse in child-directed speech. *Cognitive Science*, 38(8), 1634–1661. <https://doi.org/10.1111/cogs.12121>
- Rowe, M. L., Özçalışkan, Ş., & Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting vocabulary development. *First Language*, 28(2), 182–199. <https://doi.org/10.1177/0142723707088310>
- Rowe, M. L., Silverman, R. D., & Mullan, B. E. (2013). The role of pictures and gestures as nonverbal aids in preschoolers' word learning in a novel language. *Contemporary Educational Psychology*, 38(2), 109–117.
- Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: links between preschoolers' theory of mind and semantic development. *Child Development*, 72(4), 1054–1070. <https://doi.org/10.1111/1467-8624.00334>
- Sabbagh, M. A., & Shafman, D. (2009). How children block learning from ignorant speakers.

- Cognition*, 112(3), 415–422. <https://doi.org/10.1016/j.cognition.2009.06.005>
- Samuelson, L. K., & Smith, L. B. (1998). Memory and attention make smart word learning: An alternative account of Akhtar, Carpenter, and Tomasello. *Child Development*, 69(1), 94–104. <https://doi.org/10.1111/j.1467-8624.1998.tb06136.x>
- Santamaria, L., Noreika, V., Georgieva, S., Clackson, K., Wass, S., & Leong, V. (2020). Emotional valence modulates the topology of the parent-infant inter-brain network. *NeuroImage*, 207, 116341. <https://doi.org/10.1016/j.neuroimage.2019.116341>
- Saxon, T. F. (1997). A longitudinal study of early mother-infant interaction and later language competence. *First Language*, 17(51), 271–281. <https://doi.org/10.1177/014272379701705112>
- Schulze, C., Grassmann, S., & Tomasello, M. (2013). 3-year-old children make relevance inferences in indirect verbal communication. *Child Development*, 84(6), 2079–2093. <https://doi.org/10.1111/cdev.12093>
- Schwab, J. F., & Lew-Williams, C. (2020). Discontinuity of reference hinders children's learning of new words. *Child Development*, 91(1), e29–e41. <https://doi.org/10.1111/cdev.13189>
- Schwartz, L., Levy, J., Endevelt-Shapira, Y., Djalovski, A., Hayut, O., Dumas, G., & Feldman, R. (2022). Technologically-assisted communication attenuates inter-brain synchrony. *NeuroImage*, 264, 119677. <https://doi.org/10.1016/j.neuroimage.2022.119677>
- Scofield, J., & Behrend, D. A. (2008). Learning words from reliable and unreliable speakers. *Cognitive Development*, 23(2), 278–290. <https://doi.org/10.1016/j.cogdev.2008.01.003>
- Scofield, J., & Behrend, D. A. (2011). Clarifying the role of joint attention in early word learning. *First Language*, 31(3), 326–341. <https://doi.org/10.1177/0142723710395423>

- Scott, K., Sakkalou, E., Ellis-Davies, K., Hilbrink, E., Hahn, U., & Gattis, M. (2013). Infant contributions to joint attention predict vocabulary development. In *CogSci 2013: the 35th Annual Conference of the Cognitive Science Society* (pp. 3384-3389). Cognitive Science Society.
- Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, 18(9), 668–671. <https://doi.org/10.1016/j.cub.2008.03.059>
- Senju, A., Csibra, G., & Johnson, M. H. (2008). Understanding the referential nature of looking: Infants' preference for object-directed gaze. *Cognition*, 108(2), 303–319. <https://doi.org/10.1016/j.cognition.2008.02.009>
- Senju, A., & Johnson, M. H. (2009). The eye contact effect: Mechanisms and development. *Trends in Cognitive Sciences*, 13(3), 127–134. <https://doi.org/10.1016/j.tics.2008.11.009>
- Senju, A., Tojo, Y., Yaguchi, K., & Hasegawa, T. (2005). Deviant gaze processing in children with autism: An ERP study. *Neuropsychologia*, 43(9), 1297–1306. <https://doi.org/10.1016/j.neuropsychologia.2004.12.002>
- Senju, A., Verneti, A., Kikuchi, Y., Akechi, H., & Hasegawa, T. (2013). Cultural modulation of face and gaze scanning in young children. *PLoS One*, 8(8), e74017.
- Senju, A., Yaguchi, K., Tojo, Y., & Hasegawa, T. (2003). Eye contact does not facilitate detection in children with autism. *Cognition*, 89(1), B43–B51. [https://doi.org/10.1016/S0010-0277\(03\)00081-7](https://doi.org/10.1016/S0010-0277(03)00081-7)
- Shimpi, P., & Huttenlocher, J. (2007). Redirective labels and early vocabulary development. *Journal of Child Language*, 34, 845–859. <https://doi.org/10.1017/S0305000907008112>
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: new procedures and recommendations. *Psychological Methods*, 7(4), 422.

Sievers, B., Welker, C., Hasson, U., Kleinbaum, A. M., & Wheatley, T. (2024).

Consensus-building conversation leads to neural alignment. PsyArXiv.

<https://doi.org/10.31234/osf.io/562z7>

Slone, L. K., Abney, D. H., Smith, L. B., & Yu, C. (2023). The temporal structure of parent talk to toddlers about objects. *Cognition*, 230, 105266.

<https://doi.org/10.1016/j.cognition.2022.105266>

Smith, C. B., Adamson, L. B., & Bakeman, R. (1988). Interactional predictors of early language.

First Language, 8(23), 143–156. <https://doi.org/10.1177/014272378800802304>

Smith, L. B. (2000). An associative crane. In R. Golinkoff & K. Hirsh-Pasek (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 51–80). Oxford University Press.

Smith, L., & Ulvund, S. E. (2003). The role of joint attention in later development among preterm children: Linkages between early and middle childhood. *Social Development*, 12(2), 222–234. <https://doi.org/10.1111/1467-9507.00230>

Snyder, K. A. (2010). Neural correlates of encoding predict infants' memory in the paired-comparison procedure. *Infancy*, 15(3), 270–299.

<https://doi.org/10.1111/j.1532-7078.2009.00015.x>

Sobel, D. M., Luchkina, E., & Tummeltshammer, K. (2020). Mechanisms for evaluating others' reliability when learning novel words. In J. B. Childers (Ed.), *Language and Concept Acquisition from Infancy Through Childhood: Learning from Multiple Exemplars* (pp. 179–196). Springer International Publishing.

https://doi.org/10.1007/978-3-030-35594-4_9

Sobel, D., Sedivy, J., Buchanan, D., & Hennessy, R. (2012). Speaker reliability in preschoolers'

- inferences about the meanings of novel words. *Journal of Child Language*, 39, 90–104.
<https://doi.org/10.1017/S0305000911000018>
- Sodian, B., & Thoermer, C. (2004). Infants' understanding of looking, pointing, and reaching as cues to goal-directed action. *Journal of Cognition and Development*, 5(3), 289–316.
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (Vol. 142). Harvard University Press.
- Stallworthy, I. C., Lasch, C., Berry, D., Wolff, J. J., Pruett, J. R., Marrus, N., Swanson, M. R., Botteron, K. N., Dager, S. R., Estes, A. M., Hazlett, H. C., Schultz, R. T., Zwaigenbaum, L., Piven, J., & Elison, J. T. (2022). Variability in responding to joint attention cues in the first year is associated with autism outcome. *Journal of the American Academy of Child & Adolescent Psychiatry*, 61(3), 413–422. <https://doi.org/10.1016/j.jaac.2021.03.023>
- Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, 107(32), 14425–14430. <https://doi.org/10.1073/pnas.1008662107>
- Striano, T., Chen, X., Cleveland, A., & Bradshaw, S. (2006). Joint attention social cues influence infant learning. *European Journal of Developmental Psychology*, 3(3), 289–299.
<https://doi.org/10.1080/17405620600879779>
- Striano, T., & Reid, V. M. (2006). Social cognition in the first year. *Trends in Cognitive Sciences*, 10(10), 471–476. <https://doi.org/10.1016/j.tics.2006.08.006>
- Stokes, S. F., & Klee, T. (2009). Factors that influence vocabulary development in two-year-old children. *Journal of Child Psychology and Psychiatry*, 50(4), 498–505.
- Suanda, S. H., Smith, L. B., & Yu, C. (2018). The multisensory nature of verbal discourse in

- parent–toddler interactions. *Developmental Neuropsychology*, 41(5–8), 324–341.
<https://doi.org/10.1080/87565641.2016.1256403>
- Suarez-Rivera, C., Schatz, J. L., Herzberg, O., & Tamis-LeMonda, C. S. (2022). Joint engagement in the home environment is frequent, multimodal, timely, and structured. *Infancy*, 27(2), 232–254. <https://doi.org/10.1111/infa.12446>
- Suarez-Rivera, C., Smith, L. B., & Yu, C. (2019). Multimodal parent behaviors within joint attention support sustained attention in infants. *Developmental Psychology*, 55(1), 96–109. <https://doi.org/10.1037/dev0000628>
- Sullivan, J., & Barner, D. (2016). Discourse bootstrapping: Preschoolers use linguistic discourse to learn new words. *Developmental Science*, 19(1), 63–75.
<https://doi.org/10.1111/desc.12289>
- Sullivan, J., Boucher, J., Kiefer, R. J., Williams, K., & Barner, D. (2019). Discourse coherence as a cue to reference in word learning: Evidence for discourse bootstrapping. *Cognitive Science*, 43(1), e12702. <https://doi.org/10.1111/cogs.12702>
- Sun, L., & Yoshida, H. (2022). Why the parent’s gaze is so powerful in organizing the infant’s gaze: The relationship between parental referential cues and infant object looking. *Infancy*, 27(4), 780–808. <https://doi.org/10.1111/infa.12475>
- Swingle, D. (2010). Fast mapping and slow mapping in children’s word learning. *Language Learning and Development*, 6(3), 179–183.
<https://doi.org/10.1080/15475441.2010.484412>
- Tager-Flusberg, H. (2006). Defining language phenotypes in autism. *Clinical Neuroscience Research*, 6(3), 219–224. <https://doi.org/10.1016/j.cnr.2006.06.007>
- Takashima, A., Bakker-Marshall, I., van Hell, J. G., McQueen, J. M., & Janzen, G. (2019).

- Neural correlates of word learning in children. *Developmental Cognitive Neuroscience*, 37, 100649. <https://doi.org/10.1016/j.dcn.2019.100649>
- Tamis-LeMonda, C. S., & Bornstein, M. H. (1994). Specificity in mother-toddler language-play relations across the second year. *Developmental Psychology*, 30(2), 283–292. <https://doi.org/10.1037/0012-1649.30.2.283>
- Tamis-LeMonda, C. S., Custode, S., Kuchirko, Y., Escobar, K., & Lo, T. (2019). Routine language: Speech directed to infants during home activities. *Child Development*, 90(6), 2135–2152. <https://doi.org/10.1111/cdev.13089>
- Tamis-LeMonda, C. S., Kuchirko, Y., Luo, R., Escobar, K., & Bornstein, M. H. (2017). Power in methods: Language to infants in structured and naturalistic contexts. *Developmental Science*, 20(6), e12456. <https://doi.org/10.1111/desc.12456>
- Tamis-LeMonda, C. S., Kuchirko, Y., & Song, L. (2014). Why is infant language learning facilitated by parental responsiveness? *Current Directions in Psychological Science*, 23(2), 121–126. <https://doi.org/10.1177/0963721414522813>
- Tare, M., Shatz, M., & Gilbertson, L. (2008). Maternal uses of non-object terms in child-directed speech: Color, number and time. *First Language*, 28(1), 87–100. <https://doi.org/10.1177/0142723707085316>
- Tenenbaum, E., Amso, D., Abar, B., & Sheinkopf, S. (2014). Attention and word learning in autistic, language delayed and typically developing children. *Frontiers in Psychology*, 5, 490. <https://doi.org/10.3389/fpsyg.2014.00490>
- Thoermer, C., & Sodian, B. (2001). Preverbal infants' understanding of referential gestures. *First Language*, 21(63), 245–264. <https://doi.org/10.1177/014272370102106303>
- Tingley D, Yamamoto T, Hirose K, Keele L, Imai K (2014). “mediation: R Package for Causal

- Mediation Analysis.” *Journal of Statistical Software*, 59(5), 1–38.
<http://www.jstatsoft.org/v59/i05/>.
- Tomasello, M. (2000). The social-pragmatic theory of word learning. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, 10(4), 401–413.
<https://doi.org/10.1075/prag.10.4.01tom>
- Tomasello, M. (2001). Perceiving intentions and learning words in the second year of life. In *Language development: The essential readings* (pp. 111–128). Blackwell Publishing.
- Tomasello, M., & Barton, M. E. (1994). Learning words in nonostensive contexts. *Developmental Psychology*, 30(5), 639–650. <https://doi.org/10.1037/0012-1649.30.5.639>
- Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, 57(6), 1454–1463. <https://doi.org/10.2307/1130423>
- Tomasello, M., & Kruger, A. C. (1992). Joint attention on actions: Acquiring verbs in ostensive and non-ostensive contexts. *Journal of Child Language*, 19(2), 311–333.
<https://doi.org/10.1017/S0305000900011430>
- Tomasello, M., Mannle, S., & Kruger, A. (1986). Linguistic environment of 1- to 2-year-old twins. *Developmental Psychology*, 22, 169–176.
<https://doi.org/10.1037/0012-1649.22.2.169>
- Tomasello, M., Strosberg, R., & Akhtar, N. (1996). Eighteen-month-old children learn words in non-ostensive contexts. *Journal of Child Language*, 23(1), 157–176.
<https://doi.org/10.1017/S0305000900010138>
- Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First Language*, 4(12), 197–211. <https://doi.org/10.1177/014272378300401202>
- Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of

- primary intersubjectivity. *Before Speech: The Beginning of Interpersonal Communication*, 1, 530-571.
- Tsujimoto, S. (2008). The prefrontal cortex: Functional neural development during early childhood. *The Neuroscientist*, 14(4), 345–358.
<https://doi.org/10.1177/1073858408316002>
- Twomey, K. E. (2013). *An Investigation of Fast and Slow Mapping*. [Doctoral dissertation, University of Sussex]. University of Sussex Repository.
<https://hdl.handle.net/10779/uos.23397443.v1>
- Unger, L., & Fisher, A. V. (2021). The emergence of richly organized semantic knowledge from simple statistics: A synthetic review. *Developmental Review*, 60, 100949.
- van Schaik, J. E., Meyer, M., van Ham, C. R., & Hunnius, S. (2020). Motion tracking of parents' infant- versus adult-directed actions reveals general and action-specific modulations. *Developmental Science*, 23(1), e12869. <https://doi.org/10.1111/desc.12869>
- VanDam, M., Thompson, L., Wilson-Fowler, E., Campanella, S., Wolfenstein, K., & De Palma, P. (2022). Conversation initiation of mothers, fathers, and toddlers in their natural home environment. *Computer Speech & Language*, 73, 101338.
<https://doi.org/10.1016/j.csl.2021.101338>
- Vanderwert, R. E., & Nelson, C. A. (2014). The use of near-infrared spectroscopy in the study of typical and atypical development. *NeuroImage*, 85, 264–271.
<https://doi.org/10.1016/j.neuroimage.2013.10.009>
- Vanoncini, M., Hoehl, S., Elsner, B., Wallot, S., Boll-Avetisyan, N., & Kayhan, E. (2024). Mother-infant social gaze dynamics relate to infant brain activity and word segmentation. *Developmental Cognitive Neuroscience*, 65, 101331.

- Vecera, S. P., & Johnson, M. H. (1995). Gaze detection and the cortical processing of faces: Evidence from infants and adults. *Visual Cognition*, 2(1), 59–87.
<https://doi.org/10.1080/13506289508401722>
- Veneziano, E. (2001). Displacement and informativeness in child-directed talk. *First Language*, 21(63), 323–356.
- Vigliocco, G., Motamedi, Y., Murgiano, M., Wonnacott, E., Marshall, C., Milán-Maillo, I., & Perniss, P. (2019). *Onomatopoeia, gestures, actions and words: How do caregivers use multimodal cues in their communication to children?* PsyArXiv.
<https://doi.org/10.31234/osf.io/v263k>
- Vogt, S., & Kauschke, C. (2017). Observing iconic gestures enhances word learning in typically developing children and children with specific language impairment. *Journal of Child Language*, 44(6), 1458–1484. <http://dx.doi.org/10.1017/S0305000916000647>
- Vugs, B., Knoors, H., Cuperus, J., Hendriks, M., & Verhoeven, L. (2016). Interactions between working memory and language in young children with specific language impairment (SLI). *Child Neuropsychology*, 22(8), 955–978.
- Wahl, S., Marinović, V., & Träuble, B. (2019). Gaze cues of isolated eyes facilitate the encoding and further processing of objects in 4-month-old infants. *Developmental Cognitive Neuroscience*, 36, 100621. <https://doi.org/10.1016/j.dcn.2019.100621>
- Wahl, S., Michel, C., Pauen, S., & Hoehl, S. (2013). Head and eye movements affect object processing in 4-month-old infants more than an artificial orientation cue. *British Journal of Developmental Psychology*, 31(2), 212–230. <https://doi.org/10.1111/bjdp.12001>
- Wass, S. V., Noreika, V., Georgieva, S., Clackson, K., Brightman, L., Nutbrown, R., Covarrubias,

- L. S., & Leong, V. (2018). Parental neural responsivity to infants' visual attention: How mature brains influence immature brains during social interaction. *PLOS Biology*, 16(12), e2006328. <https://doi.org/10.1371/journal.pbio.2006328>
- Wass, S. V., Whitehorn, M., Haresign, I. M., Phillips, E., & Leong, V. (2020). Interpersonal neural entrainment during early social interaction. *Trends in Cognitive Sciences*, 24(4), 329–342. <https://doi.org/10.1016/j.tics.2020.01.006>
- Werchan, D. M., & Amso, D. (2021). All contexts are not created equal: Social stimuli win the competition for organizing reinforcement learning in 9-month-old infants. *Developmental Science*, 24(5), e13088. <https://doi.org/10.1111/desc.13088>
- Wilcox, T., & Biondi, M. (2015). fNIRS in the developmental sciences. *WIREs Cognitive Science*, 6(3), 263–283. <https://doi.org/10.1002/wcs.1343>
- Wildt, E., & Rohlfing, K. J. (2024). The impact of maternal gaze responsiveness on infants' gaze following and later vocabulary development. *Infant Behavior and Development*, 74, 101917. <https://doi.org/10.1016/j.infbeh.2023.101917>
- Wilson, S. M., Molnar-Szakacs, I., & Iacoboni, M. (2008). Beyond superior temporal cortex: intersubject correlations in narrative speech comprehension. *Cerebral Cortex*, 18(1), 230–242. <https://doi.org/10.1093/cercor/bhm049>
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). *ELAN: a Professional Framework for Multimodality Research*.
- Wojcik, E. H., & Saffran, J. R. (2015). Toddlers encode similarities among novel words from meaningful sentences. *Cognition*, 138, 10–20.
- Wojcik, E. H., Zettersten, M., & Benitez, V. L. (2022). The map trap: Why and how word

- learning research should move beyond mapping. *WIREs Cognitive Science*, 13(4), e1596.
<https://doi.org/10.1002/wcs.1596>
- Wu, R., Gopnik, A., Richardson, D. C., & Kirkham, N. Z. (2011). Infants learn about objects from statistics and people. *Developmental Psychology*, 47(5), 1220.
- Wu, R., & Kirkham, N. Z. (2010). No two cues are alike: Depth of learning during infancy is dependent on what orients attention. *Journal of Experimental Child Psychology*, 107(2), 118–136. <https://doi.org/10.1016/j.jecp.2010.04.014>
- Yoon, J. M. D., Johnson, M. H., & Csibra, G. (2008). Communication-induced memory biases in preverbal infants. *Proceedings of the National Academy of Sciences*, 105(36), 13690–13695. <https://doi.org/10.1073/pnas.0804388105>
- Young, G. S., Merin, N., Rogers, S. J., & Ozonoff, S. (2009). Gaze behavior and affect at 6-months: predicting clinical outcomes and language development in typically developing infants and infants at-risk for autism. *Developmental Science*, 12(5), 798–814.
<https://doi.org/10.1111/j.1467-7687.2009.00833.x>
- Yow, W. Q., Li, X., Lam, S., Gliga, T., Chong, Y. S., Kwek, K., & Broekman, B. F. P. (2017). A bilingual advantage in 54-month-olds' use of referential cues in fast mapping. *Developmental Science*, 20(1), e12482. <https://doi.org/10.1111/desc.12482>
- Yow, W. Q., & Markman, E. M. (2011). Young bilingual children's heightened sensitivity to referential cues. *Journal of Cognition and Development*, 12(1), 12–31.
<https://doi.org/10.1080/15248372.2011.539524>
- Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, 70(13–15), 2149–2165.
<https://doi.org/10.1016/j.neucom.2006.01.034>

- Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, 18(5), 414–420.
<https://doi.org/10.1111/j.1467-9280.2007.01915.x>
- Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, 125(2), 244–262. <https://doi.org/10.1016/j.cognition.2012.06.016>
- Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLOS ONE*, 8(11), e79659. <https://doi.org/10.1371/journal.pone.0079659>
- Yu, C., & Smith, L. B. (2016). The social origins of sustained attention in one-year-old human infants. *Current Biology*, 26(9), 1235–1240.
<https://doi.org/10.1016/j.cub.2016.03.026>
- Yu, C., & Smith, L. B. (2017). Multiple sensory-motor pathways lead to coordinated visual attention. *Cognitive Science*, 41(S1), 5–31. <https://doi.org/10.1111/cogs.12366>
- Yu, C., Suanda, S. H., & Smith, L. B. (2019). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental Science*, 22(1), e12735. <https://doi.org/10.1111/desc.12735>
- Yu, C., Zhang, Y., Slone, L. K., & Smith, L. B. (2021). The infant's view redefines the problem of referential uncertainty in early word learning. *Proceedings of the National Academy of Sciences*, 118(52), e2107019118. <https://doi.org/10.1073/pnas.2107019118>
- Yurovsky, D. (2018). A communicative approach to early word learning. *New Ideas in Psychology*, 50, 73–79. <https://doi.org/10.1016/j.newideapsych.2017.09.001>
- Yurovsky, D., & Frank, M. C. (2017). Beyond naïve cue combination: Salience and social cues in

early word learning. *Developmental Science*, 20(2), e12349.

<https://doi.org/10.1111/desc.12349>

Zettersten, M., & Saffran, J. R. (2021). Sampling to learn words: Adults and children sample words that reduce referential ambiguity. *Developmental Science*, 24(3), e13064.

<https://doi.org/10.1111/desc.13064>

Zukow, P. G. (1990). Socio-perceptual bases for the emergence of language: An alternative to innatist approaches. *Developmental Psychobiology*, 23(7), 705–726.

<https://doi.org/10.1002/dev.420230711>