Cognitive Development 64 (2022) 101254

Contents lists available at ScienceDirect

Cognitive Development

journal homepage: www.elsevier.com/locate/cogdev

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

Crystal Lee [1,*], Casey Lew-Williams [2]

Department of Psychology, Princeton University, USA

ARTICLE INFO

Keywords:
Eye gaze
Discourse
Social cues
Word learning

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, but not 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.

1. 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, 2021). 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,

* Correspondence to: Princeton University, Department of Psychology, Princeton, NJ 08540, USA.
E-mail address: cl40@princeton.edu (C. Lee).
[1] https://orcid.org/0000-0001-5452-6997.
[2] https://orcid.org/0000-0002-8781-4458.

https://doi.org/10.1016/j.cogdev.2022.101254
Received 19 April 2022; Received in revised form 17 September 2022; Accepted 19 September 2022
Available online 26 September 2022

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, Markman, Bill, Desjardins, Irwin & Tidball, 1996; Booth, McGregor, & Rohlfing, 2008; Briganti & Cohen, 2011; Houston-Price, Plunkett, & Duffy, 2006; Hirotani, Stets, Striano & Friederici, 2009; Nurmsoo & Bloom, 2008; Paulus & Fikkert, 2014; Rader & Zukow-Goldring, 2012; Yu & Ballard, 2007; see also Scofield & Behrend, 2008). Furthermore, the ability to follow eye movements has been linked to growth in vocabulary (Akhtar & Gernsbacher, 2007; Brooks & Meltzoff, 2005, 2008; Carpenter, Nagell, Tomasello, Butterworth & Moore, 1998; Morales, Mundy & Rojas, 1998; Morales et al., 2000; 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 time-scales as well (Piazza, Nencheva, & Lew-Williams, 2021; McMurray, 2016). 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, Kuchirko, Luo, Escobar, and Bornstein (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, Armstrong, Gleitman,

Goldin-Meadow, Medina & Trueswell, 2013; Pereira, Smith, & Yu, 2014; Suanda, Smith, & Yu, 2016). 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, Boucher, Kiefer, Williams & Barner, 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. Method

2.1. 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 as being White, 3 as Asian, 2 as Black, and 1 as Hispanic (the remaining 19 participants did not report their ethnicity). The [name of university omitted for blind review] 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.

2.2. Experimental design

2.2.1. 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.

2.2.2. 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 Fig. 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 2 × 2 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 Fig. 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.

2.2.3. 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 of 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 s ($SD = 4.15$), and 15 trials (0.9 % of total trials) were greater than 20 s; 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.

2.2.4. 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 s 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, gazzer, 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 × 400 pixels in size and presented against a white background.

2.2.5. Procedure

Children participated in this study from their homes on either a smartphone or tablet, and an experimenter interacted with the child

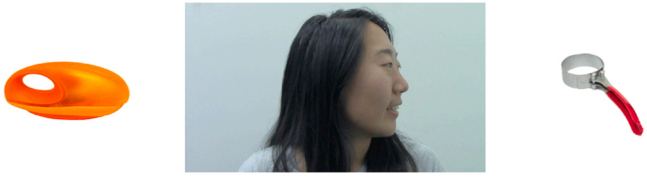

Fig. 1. Example of stimuli used in the learning phase. A speaker turned her head to label each object in sequence. These labeling utterances were embedded in brief, content-neutral discourses about each object (e.g., "*I like the shape*").

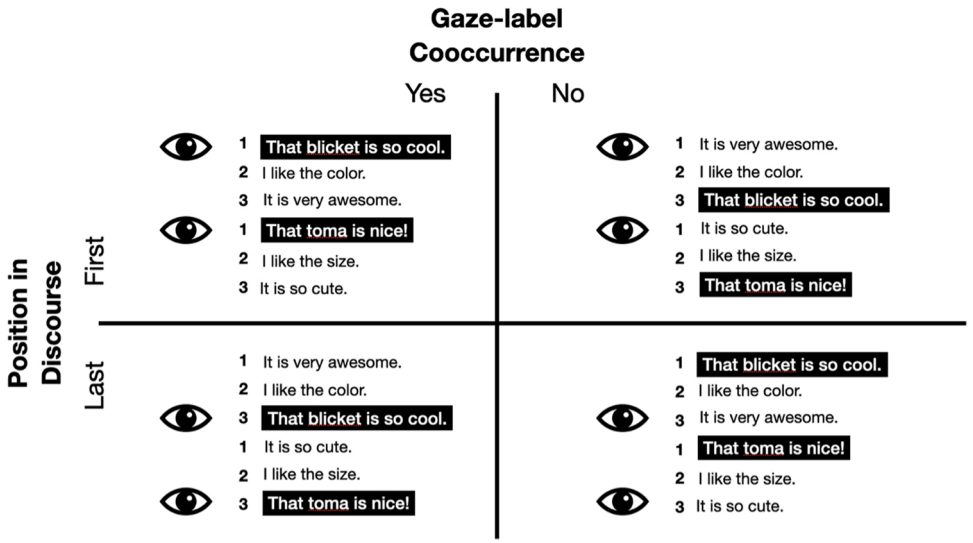

Fig. 2. The four within-subject learning conditions presented in the experiment. We manipulated whether a social cue was presented first or last in a discourse, and whether the cue co-occurred with a labeling sentence. The eye icon represents the presentation of a social cue, and the highlighted text 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.

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.

3. 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, $t(799) = 0.64$, $p = 0.52$.[3] (Fig. 3).

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.

4. 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 and 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

[3] 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.

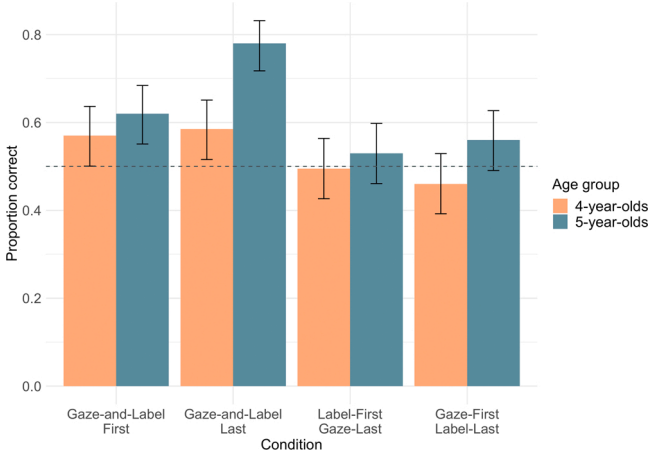

Fig. 3. Proportion of correct responses by condition and by age group. Chance is shown by the dashed line (50 %). Error bars represent 95 % confidence intervals.

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 (2014) 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 min, 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., 2000; 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 looking 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; Baddeley & Hitch, 1974). 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, Leitao, & Claessen, 2016; Vugs, Knoors, Cuperus, Hendriks & Verhoeven, 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 time-scales, 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, Yu, & Smith, 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.

5. 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.

Conflicts of interest

We have no conflicts of interest to disclose.

Data availability

Materials, data, and analyses are publicly available at https://osf.io/ke69t/?view_only=19a216425ac74884b26991da4d091b36.

Acknowledgements

This work was supported by the National Science Foundation Graduate Research Fellowship [DGE-2039656] awarded to CL, and the National Institutes of Child Health and Human Development [grant number 5R01HD095912] awarded to CLW.

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/infa.12365

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.

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

Allen, S. E. (2000). A discourse-pragmatic explanation for argument representation in child Inuktitut. *Linguistics, 38*, 483–521.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*(3), 189–208.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *In Psychology of learning and motivation, 8* pp. 47–89). Academic Press.

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. (1995). Understanding the link between joint attention and language. *Joint Attention: Its Origins and Role in Development, 131*, 158.

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.

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

Bannard, C., & Tomasello, M. (2012). Can we dissociate contingency learning from social learning in word acquisition by 24-month-olds? *PLoS One, 7*(11), Article e49881. https://doi.org/10.1371/journal.pone.0049881

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. https://doi.org/10.1111/j.1468-0017.1992.tb00203.x

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot Int., 5*, 341–345.

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

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

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

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

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), 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 USA, 110*(28), 11278–11283. https://doi.org/10.1073/pnas.1309518110

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

Clancy, P. M. (2004). The discourse basis of constructions: Some evidence from Korean acquisition. *In Proceedings of the thirt second Stanford child language research forum*, 20–29.

Corkum, V., & Moore, C. (1995). Development of joint visual attention in infants. In C. Moore, & P. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 61–83). Hillsdale, NJ: Erlbaum.

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

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods, 47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

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

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.

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., Lee, K., & Muir, D. (2017). *Gaze-following: Its development and significance*. Psychology Press.
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., & Smith, I. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*, 31–60.
Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*(4), 513–543.
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
Hirotani, 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
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
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. https://doi.org/10.1111/j.1467-8624.2009.01311.x
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
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
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). Positive and Protective: Effects of Early Theory of Mind on Preschool Problem Behaviors. *Journal of Child Psychology and Psychiatry, 48*, 1025–1032. https://doi.org/10.1111/j.1469-7610.2007.01806.x
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.
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
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
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
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., & 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., & 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
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
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
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
Piazza, E. A., Nencheva, M. L., & Lew-Williams, C. (2021). The development of communication across timescales, 09637214211037665 *Current Directions in Psychological Science*. https://doi.org/10.1177/09637214211037665.
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
Rohde, H., & Frank, M. C. (2014). Markers of topical discourse child-directed speech. *Cognitive Science, 38*(8), 1634–1661. https://doi.org/10.1111/cogs.12121
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
Scofield, J., & Behrend, D. A. (2008). Learning words from reliable and unreliable speakers. *Cognitive Development, 23*(2), 278–290.
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
Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition, 142*. Harvard University Press.
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.
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
Suanda, S. H., Smith, L. B., & Yu, C. (2016). 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., 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), Article e12702. https://doi.org/10.1111/cogs.12702
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), Article e12456. https://doi.org/10.1111/desc.12456
Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development, 57*(6), 1454–1463. https://doi.org/10.2307/1130423
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
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.
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), Article e13088. https://doi.org/10.1111/desc.13088
Wojcik, E., Zettersten, M., & Benitez, V. (2021). The map trap: Why and how word learning research should move beyond mapping. WIREs Cognitive Science.

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), Article e12482. https://doi.org/10.1111/desc.12482

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., 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), Article e12735. https://doi.org/10.1111/desc.12735