DOI: 10.1111/infa.12500

RESEARCH ARTICLE

INFANCY THE OFFICIAL JOURNAL OF THE INTERNATIONAL CONGRESS OF INFANT STUDIES WILEY

# The real-time effects of parent speech on infants' multimodal attention and dyadic coordination

**Sara E. Schroer** | **Chen Yu**

University of Texas at Austin, Austin, Texas, USA

**Correspondence**
Sara E. Schroer, University of Texas at Austin, Austin, TX, USA.
Email: saraschroer@utexas.edu

**Funding information**
National Science Foundation, Grant/Award Number: DGE-1610403 (GRFP); National Institutes of Health, Grant/Award Numbers: R01HD074601, R01HD093792, T32HD007475

**Abstract**

Parental responsiveness to infant behaviors is a strong predictor of infants' language and cognitive outcomes. The mechanisms underlying this effect, however, are relatively unknown. We examined the effects of parent speech on infants' visual attention, manual actions, hand-eye coordination, and dyadic joint attention during parent-infant free play. We report on two studies that used head-mounted eye trackers in increasingly naturalistic laboratory environments. In Study 1, 12-to-24-month-old infants and their parents played on the floor of a seminaturalistic environment with 24 toys. In Study 2, a different sample of dyads played in a home-like laboratory with 10 toys and no restrictions on their movement. In both studies, we present evidence that responsive parent speech extends the duration of infants' multimodal attention. This social "boost" of parent speech impacts multiple behaviors that have been linked to later outcomes—visual attention, manual actions, hand-eye coordination, and joint attention. Further, the amount that parents talked during the interaction was negatively related to the effects of parent speech on infant attention. Together, these results provide evidence of a trade-off between quantity of speech and its effects, suggesting multiple pathways through which parents impact infants' multimodal attention to shape the moment-by-moment dynamics of an interaction.

# 1 | INTRODUCTION

Over the first few years of life, parental responsiveness plays a critical role in improving children's cognitive abilities, including memory and problem-solving, social skills, language outcomes, and attentional abilities like habituation rates and exploratory behavior (e.g., Belsky et al., 1980; Riksen-Walraven, 1978; Shannon et al., 2002; Tamis-LeMonda, Bornstein, & Baumwell, 2001). Parental responsiveness is often defined as moments when parents provide responses that are prompt and relevant to the infant's needs. In everyday contexts such as toy play, parent responses—and the child behaviors that elicit them—can be measured across many modalities including visual attention, manual action (holding and manipulating objects), and speech. Yet despite the well-documented link between parental responsiveness and early development (Figure 1), we lack a mechanistic understanding of how and why responsiveness matters. One potential pathway is through shaping infant multimodal attention. We know that infants' multimodal attention on objects predicts concurrent and future outcomes (e.g., Kannass & Oakes, 2008; Schroer & Yu, 2022; Slone et al., 2019; Yu & Smith, 2012), but we know little on how parental responsiveness influences infant attention in real time. The present paper tests for the presence of this specific pathway (dashed red arrow in Figure 1) by examining how infant multimodal attention varies when parents are (or are not) verbally responsive in the moment.

## 1.1 | Power of parental responsiveness

By definition in the literature, micro-level responsive behaviors from parents have two critical characteristics: (1) temporal contingency: parents need to promptly respond to an infant's behavior; and (2) semantic relevancy: parents need to appropriately respond to what their infant is doing (e.g., Mesman, 2010, for an overview). Parent responses include a diversity of other behaviors (e.g., gestures, looks to an object), but here we will focus on responsive parent speech, which is used by parents as a primary way to respond to their child's needs. By providing contingent feedback and linguistic input at relevant moments, parental verbal responsiveness supports the development of early language abilities and cognitive functioning. Because the long-lasting effects of parental responsiveness are far from trivial, it has been identified as a malleable factor and targeted by many longitudinal and intervention-based studies. Early work found that increasing maternal stimulation, including responsive and attention-focusing behaviors, was associated with faster habituation rates, increased

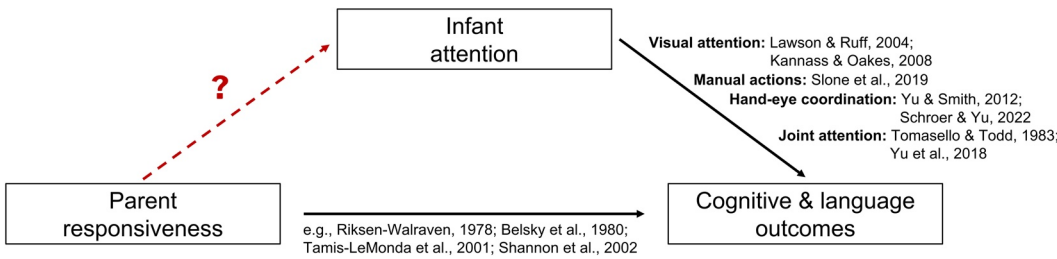



**FIGURE 1** What are the mechanisms through which parental responsiveness can predict developmental outcomes? To answer this question, we need to study how responsiveness is shaping interactions in the moment, as opposed to just correlating responsiveness to later outcomes. One mediating factor could be infant multimodal attention, as multiple modalities of infant attention have been linked to cognitive and language outcomes. The goal of this paper is to test the missing link in this mechanistic hypothesis (dashed red arrow), to show that parental responsiveness shapes multimodal infant attention in real time.

exploratory behavior, and improved learning of contingencies in an operant-conditioning task (Belsky et al., 1980; Riksen-Walraven, 1978). Parents that were consistently rated as "highly responsive" had children with faster cognitive growth across the first 4 years of life (Landry et al., 2001). And the effects of early responsive language extend to children's cognitive outcomes, vocabulary size, and language comprehension even 10 years later (Gilkerson et al., 2018).

## 1.2 | Going beyond speech

What are the pathways through which responsiveness impacts early development? A clear line can be drawn from parental responsiveness to language and cognitive outcomes (Figure 1), because responsive parent speech creates the linguistic environment children are exposed to. Indeed, most research on parental responsiveness focuses on its role in creating rich input for language development (e.g., Masur et al., 2005; Tamis-LeMonda et al., 2001). For example, a diverse vocabulary, the use of rituals to scaffold interactions, and promoting joint engagement and turn-taking are all aspects of parental responsiveness that have been shown to create input that predicts greater language outcomes (Hirsh-Pasek et al., 2015; Rowe, 2012). But parent talk is doing more than just providing the linguistic data for infants to learn. After all, if children are not hearing the right words at the right moment, and if they are not attending to relevant visual information when hearing words, then language learning would not be successful. Thus, one pathway linking parental responsiveness and infant learning is that responsiveness has real-time effects on infant multimodal attention, creating and extending infants' sustained attention on relevant information to support learning. Before considering infant attention as a mediating factor between parental responsiveness and language learning outcomes, we first need to establish that infants' own multimodal attention is also predictive of later outcomes (the second solid line in Figure 1).

## 1.3 | Multimodal attention as an underlying mechanism

While exploring their world, infants are creating learning opportunities through their visual attention and manual activities. The ability of an infant to sustain their attention to an object or activity during play-alone tasks is predictive of later cognitive and language outcomes (Kannass & Oakes, 2008; Lawson & Ruff, 2004). These moments of sustained attention have been shown to provide a rich opportunity for children to learn about the properties of objects (Ruff et al., 1990). Moreover, hands and manual actions can be viewed as a part of the infant's attention system because sustained visual attention is most often accompanied by manual actions (Yu et al., 2009). While they play with a toy, infants manipulate the object and move it around in space, creating a vast number of different views of the toy. The extent to which infants self-generate this variability in object views is positively related to their vocabulary growth (Slone et al., 2019). Informative object views are thought to be created through infant's hand-eye coordination. While holding and looking at an object, the infant aligns their body at the midline—bringing the object to their center as well as centering their eyes and head (Bambach et al., 2016). As a result, hand-eye coordination places attended objects big and centered in the infant's field of view. Sustained attention and big, centered, and varied object views not only support learning about visual objects, but also invite parents to provide needed linguistic input for word learning at the same time. Parents are more likely to label objects when their infant is in hand-eye coordination (West & Iverson, 2017) and labeling in these moments supports real-time word learning (Pereira et al., 2014; Schroer & Yu, 2022; Yu & Smith, 2012). Beyond infants' own multimodal

attention, the ability of parents and infants to coordinate their visual attention, often termed joint attention (JA), has long been known to predict infant developmental outcomes (e.g., Tomasello & Todd, 1983; Yu et al., 2018).

## 1.4 | Coordinating visual attention

Dyadic differences in JA, or parent and infant visually attending to the same object at the same time, is predictive of individual differences in child vocabulary size (Tomasello & Todd, 1983). Bouts of JA emerge from multimodal dyadic behaviors, with hands playing a guiding role (Deák et al., 2018; Yu & Smith, 2013, 2017). Critically, JA shapes infant visual attention in the moment. When parents jointly attend to the object their infant is gazing at, the infant's ability to sustain their attention to that object is extended (Yu & Smith, 2016). It is these moments of visual sustained attention within JA that are the most predictive of the infant's later vocabulary (Yu et al., 2018). Of course, parent visual attention is not the only modality that can influence infant visual attention. Redundancy of parent behaviors—such as talking, holding, and looking at the object—extends infant attention beyond the effect of visual JA alone (Suarez-Rivera et al., 2019). And multimodal redundancy across all three modalities has the strongest effect on infant attention. Further, when parents were instructed to drastically reduce their responsiveness, infants had fewer bouts of sustained manual actions and shifted between objects more frequently (McQuillan et al., 2019).

The importance of socially scaffolded visual attention suggests a specific pathway linking parental responsiveness with cognitive and language outcomes—when parents are selective in when and how they choose to respond to their infants, parental responsiveness shapes infants' sensorimotor and social experiences. Very little work has been done, however, on the real-time effect of parental responsiveness on infants' multimodal attention (the dashed red arrow in Figure 1). In the present paper, we measure how parents' responsive speech can scaffold the coordination between parents and infants during social interactions.

## 1.5 | Present study

The present paper bridges studies of parental responsiveness, which has often been studied as a parenting characteristic at a macro level (e.g., Hirsh-Pasek et al., 2015; Masur et al., 2005), with more fine-grained analyses of multimodal behaviors within parent-infant social interactions (e.g., Chang, de Barbaro, & Deák, 2017; Suarez-Rivera et al., 2019; Yu & Smith, 2016). Our goal was to examine the function of responsiveness at a novel developmental timescale, looking at variations in fractions of seconds instead of milestones over months. By studying parent talk and its effects as a real-time event (i.e., a single utterance), as opposed to a global characteristic of parental responsiveness, we can measure how parent speech influences infant multimodal attention and test specific mechanistic hypotheses as to why responsiveness matters.

Based on the literature, our goal was to quantify how parent speech influences four types of multimodal attention and dyadic coordinated attention: (1) infants' visual attention and (2) manual actions, (3) the sensorimotor coordination of visual attention and manual action, and (4) dyadic coordination in the form of JA to objects. In line with recent work studying real-time attention in parent-infant interactions, we used head-mounted eye trackers to capture eye gaze (e.g., Franchak et al., 2018; Suarez-Rivera et al., 2019; Yu et al., 2018; Yu & Smith, 2016). Head-mounted eye tracking captures natural behaviors in real time that can be analyzed frame by frame, typically at a rate of 30 frames/sec,

allowing visual attention to be quantified at fractions of a second. Using head-mounted eye trackers, we asked how parent speech that responds to changes in infant attention (e.g., shifting gaze to an object) affects infant engagement with that object in real time.

We present findings from two laboratory settings in different play contexts. In Study 1, parents and infants were seated on the floor of a room decorated with child-appropriate artwork, a rug, and cushions to approximate a seminaturalistic home or daycare environment (Figure 2a). Parents and infants played with 24 toys and were given no instructions beyond "play as you would at home." To test the robustness of the observed effects of parent speech, we used data collected in a different lab environment (Study 2) that was designed to look like a studio apartment—with a play space, living room, and kitchenette (Figure 2b). Parents and infants were given 10 toys to play with while they wore wireless head-mounted eye trackers. Going wireless allowed the dyads to move around freely—with infants (and parents!) scooting, crawling, running, and climbing as they would at home.

We hypothesized that parent-spoken utterances would support infant's ability to sustain their multimodal attention. If parents responded to a change in infant attention, then that bout of attention would last longer than an attention bout that parents did not respond to. As a first step toward understanding the multimodal attention pathway of parental responsiveness, we chose to not consider the content in parent speech, and instead study all parent speech that was contingent to infant multimodal attention to capture overall real-time effects of parent speech. Further, based on previous work showing individual differences in parent speech during free play (e.g., Hirsh-Pasek et al., 2015), we conducted a second set of analyses to test for potential individual differences in the effects of parent speech. We divided participants into two groups of "more talkative" and "less talkative" parents and examined whether quantitative differences in the amount of parent speech could predict qualitative differences in the effects of parent speech on infants' multimodal attention. In both Study 1 and Study 2, we found that parents' speech scaffolds infant multimodal attention, suggesting a pathway through which parental responsiveness may shape the sensorimotor and social experiences of the infant.

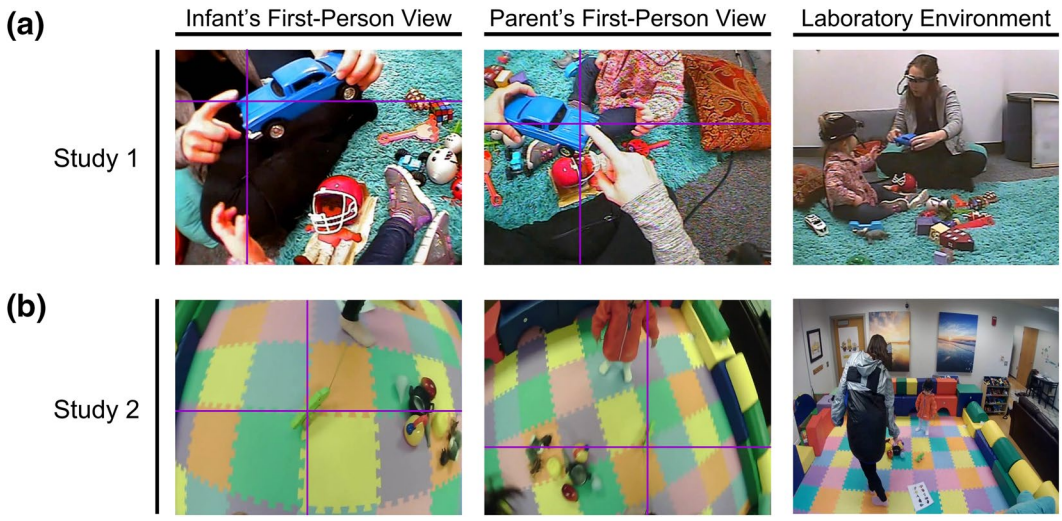



**FIGURE 2** (a) In Study 1, participants were seated on the floor while playing with 24 toys. (b) In Study 2, dyads played with 10 toys and wore wireless eye trackers that allowed participants to move freely. For one frame from each study, examples are shown of the first-person views captured from the eye trackers worn by infants (left) and their parents (middle) with the center of the purple crosshair representing the participant's gaze, as well as the laboratory environments (right).

# 2 | STUDY 1

## 2.1 | Methods

### 2.1.1 | Participants

Thirty-four infants (16 female) and their parents were recruited from Bloomington, Indiana, a large college town in midwestern United States, to participate in a study about parent-infant interactions during toy play. The community of Bloomington and the surrounding area is primarily white, non-Hispanic, working- and middle-class families. The average age of infant participants was 18.7 months (SD = 3.02), though the range was distributed across the entire second year of life (12.3–24.3 months). Despite a wide age range, age did not correlate with any dependent variable (as will be discussed in further detail). We chose to study 12- to 24-month-old infants as the importance of parental responsiveness in this age range has been studied extensively at the macro-level (e.g., Masur et al., 2005; Tamis-LeMonda et al., 2001) and micro-level patterns in parent-infant interactions have been described (e.g., Suarez-Rivera et al., 2019; Yu & Smith, 2016), but the hypothesized mechanistic link between the two (as in Figure 1) has not yet been studied. A sample size of 34 infants was deemed appropriate for our analyses, as previous micro-level studies with temporally dense data like our own used similar sample sizes (e.g., Franchak, et al., 2018; Yu & Smith, 2016). Further, since our analyses were on the event level, as opposed to subject level, the data we analyzed varied between 800 to over 2000 instances of behaviors. The present study was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent or guardian for each child before any assessment or data collection. All procedures in this study were approved by the Human Subjects and Institutional Review Boards at Indiana University.

### 2.1.2 | Data collection

Parents were asked to play on the floor with their infant while both participants wore head-mounted eye trackers (Figure 3). Dyads were asked to play for 10 minutes and an average of 7.2 min of useable data were collected per subject (range: 3.9–11.6 min). There was a weak positive correlation of age and experiment duration ($r = 0.356$, $p = 0.039$). At the beginning of the play session, 24 toys were randomly spread out across the floor and parents were asked to play as they would at home. Parents were told that they could sit in any orientation (behind, next to, in front of their infant), but were asked to keep their infant sitting on the floor due to the eye tacker's cable that was attached to a nearby computer.

The eye tracker system (Positive Science) used a scene camera on the participant's forehead to record images from the wearer's perspective with a visual field of 108°. An infrared camera pointed to the participant's right eye to record eye movement. The infant's eye tracker was affixed to a hat and the parent wore their eye tracker like a pair of glasses. Additional cameras and microphones were placed in the room to capture traditional third-person views of the dyad.

The experiment was run by two researchers. The session began by one researcher placing the eye tracker on the parent and adjusting the scene and eye cameras, while the other researcher engaged the infant with toys that were not used as experimental stimuli (e.g., a pop-up toy that played music). Afterward, both researchers worked together to place the eye tracker on the infant. One researcher and the parent continued to distract the infant as the other researcher set up the eye tracker on the infant.

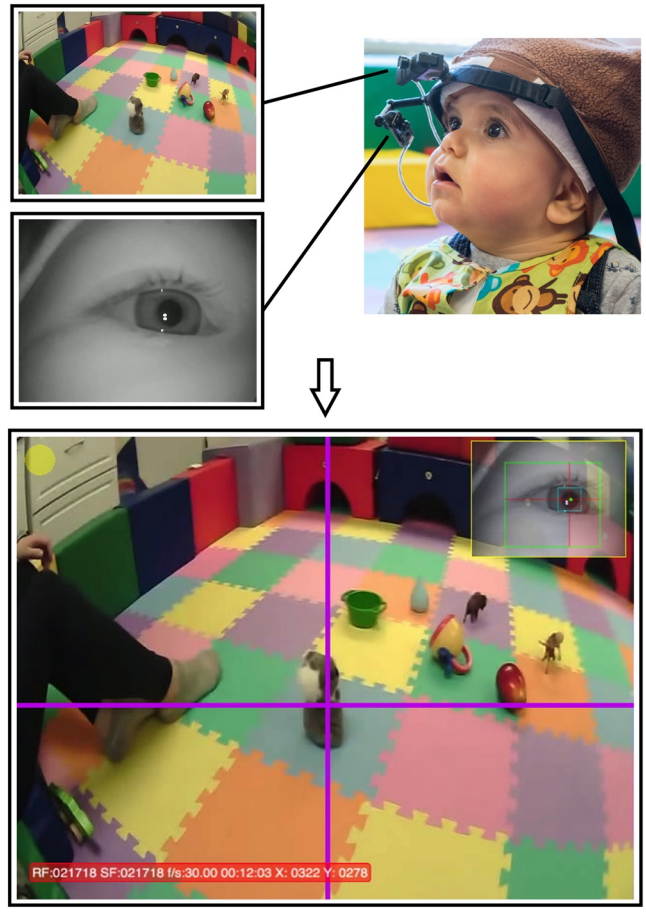



**FIGURE 3** Infant participants wore a modified head-mounted eye tracker attached to a small hat with strips of soft touch fastener. There were two cameras on the eye tracker—(1) a scene camera that captured the egocentric view of the participant and (2) a camera that captured the participant's eye. The videos were calibrated so that the x- and y-coordinates of the pupil could be mapped onto the egocentric view, generating the purple crosshair that indicates participant gaze. Parents wore the out-of-the-box models of eye trackers made by Positive Science (Study 1) and Pupil Labs (Study 2).

The researchers then monitored the experiment from an adjoining room. If the infant's (or parent's) eye camera was bumped or moved during play, the researchers reentered and quickly adjusted the camera.

### 2.1.3 | Data processing

Following the experiment, the eye tracking videos from the scene and eye cameras were synchronized and calibrated with Yarbus software (Positive Science) to generate a crosshair that indicated where the participant was looking during each frame of the video. Parent and infant visual gaze were then coded manually using the first-person view (from the scene camera) with the overlaid crosshair indicating where the participant gazed frame by frame (30 frames/second). Using an in-house program, the coder annotated which region of interest (ROI) the crosshair overlapped with. There were 25 ROIs—one for each toy and the social partner's face. For the presented work, however, only looks to objects were

analyzed. The scene cameras and third-person views were then used to annotate the objects being touched by a participant, frame by frame. Participants' left and right hands were coded separately.

Parent speech was transcribed using Audacity at the utterance level following Suarez-Rivera et al. (2019). There was no minimum length for an utterance, but separate utterances had to be 400 ms or more apart (otherwise they were collapsed together). All parent talk and vocal play (such as saying "vroom-vroom" or making a crashing sound) were considered speech. Due to the 400 ms criteria, chunks of speech that would be considered sentences, or communicative units, could be split apart and separate sentences could be counted as one utterance. Parents spoke an average of 17.106 times/minute (SD = 3.700) and each utterance lasted an average of 1.319 s (SD = 0.372). Neither infant age nor experiment duration correlated with frequency of parent speech or mean duration of utterances (*ps* > 0.149).

In the presented analyses, we were interested in four ways that parent speech shapes the sensorimotor and social experiences of the infant: infant visual attention, infant manual action, infant hand-eye coordination, and joint attention. Visual attention was defined as all infant looks to the toys. Manual action was similarly defined as all instances of the infant touching an object with either or both hands. Hand-eye coordination was defined as moments when the infant looked at and handled the same object, or the overlap of visual attention and manual action. Lastly, parent and infant attention were compared at the frame level to find moments of JA. For every frame, JA was objectively defined as when parent and infant were gazing at the same object—no other behaviors were needed to count as JA. For the analyses, a bout of JA had to last at least 500 ms, but could include short looks away from the attended object (Yu & Smith, 2017). For the three measures of infant multimodal attention, a minimum duration was also set at 500 ms. All transcribed parent utterances were counted as speech in the presented analyses.

To test the effects of parent speech, we categorized each bout of multimodal attention as "with speech" if the onset of a parent utterance began after the onset of the bout and before the offset of the bout (Figure 4). Thus, the responsive utterance followed the infant's attention by at least a single frame. Other bouts of multimodal attention, without any overlap with a responsive parent utterance, were categorized as "without speech". With this definition, the effects of parent speech were measured by comparing attentional bouts in the two categories.

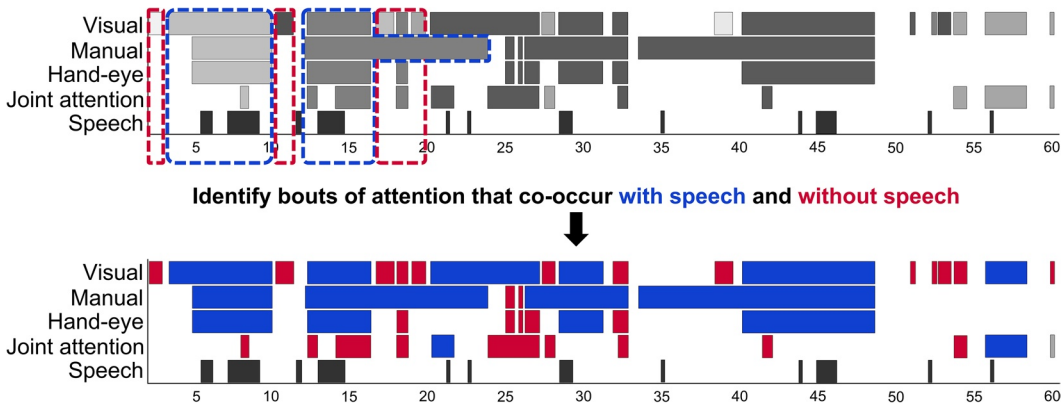



**FIGURE 4** A data visualization stream illustrating infant's visual attention, manual actions, and hand-eye coordination, as well as joint attention and parent speech, during 60s of an interaction (top). Each block represents an event (the onset, offset, and duration of a behavior) and the gray colors represent the objects being attended to by the infant. To examine the effects of parent speech, we determined whether each bout of multimodal attention co-occurred with parent speech (bottom). If a parent utterance began within a bout of multimodal attention, it was coded as "with speech" (blue) and all other bouts of multimodal attention were coded as "without speech" (red).

### 2.1.4 | Data analysis

The effects of parent speech on infants' visual attention, manual action, hand-eye coordination, and JA were analyzed separately. Across all subjects, there were 1998 instances of visual attention with speech and 2524 instances of visual attention without speech; 1408 instances of manual action with speech and 1092 instances of manual action without speech; 800 instances of hand-eye coordination with speech and 839 instances of hand-eye coordination without speech; and 894 instances of JA with speech and 901 instances of JA without speech. We used generalized mixed effects models to predict whether the duration of a bout of infant attention could predict whether it was accompanied by speech. Since the speech variable was dichotomous, binomial logistic regressions were used. We included two random effects for the object being interacted with and the subject contributing the data point (lme4 package for *R*; Bates et al., 2015). Each full model was then compared to a null model, with intercept and random effects only, using Chi-Square difference tests.

We then specifically tested whether parent speech co-occurs with sustained multimodal attention. To match the previously used definition of sustained visual attention from similar studies, bouts of sustained multimodal attention were defined as infant multimodal attention lasting 3 s or longer (Yu & Smith, 2016). The same binomial logistic regressions were used to separately analyze the subsets of visual attention, manual action, hand-eye coordination, and JA that exceeded the 3s threshold.

We also predicted that there would be considerable variability in parental responsiveness in terms of the frequency of parent speech, or the number of utterances per minute, and that this variability could impact the effects of parent speech on infant multimodal attention. Using a median split, dyads were placed into groups based on the frequency of parent speech. We then compared the durations of the infants' multimodal attention produced in the less talkative and more talkative groups. To directly measure the effects of parent speech, we only analyzed instances of attention that were accompanied by parent speech. For the less talkative group, there were 769 instances of visual attention with speech, 619 instances of manual action with speech, 348 instances of hand-eye coordination with speech, and 352 instances of JA with speech. As expected, there were more instances of multimodal attention with parent speech in the more talkative group: 1229 instances of visual attention with speech, 789 instances of manual action with speech, 452 instances of hand-eye coordination with speech, and 542 instances of JA with speech. Data was analyzed using linear mixed effects regressions (lme4 package for *R*; Bates et al., 2015), to see if group membership could predict the duration of attention. The object the infant was interacting with was included as a random effect. The model for each type of multimodal attention was then compared to a null model with intercept and random effect of object only, using Chi-Square difference tests. All reported *p*-values were not corrected for multiple comparisons.

## 2.2 | Results

### 2.2.1 | Effects of parent speech

We tested the relationship between parent speech and the four multimodal measures of infant and dyadic attention. Event-level analyses were used to compare the durations of multimodal attentional bouts with and without speech. The four types of multimodal attention were found to last longer when

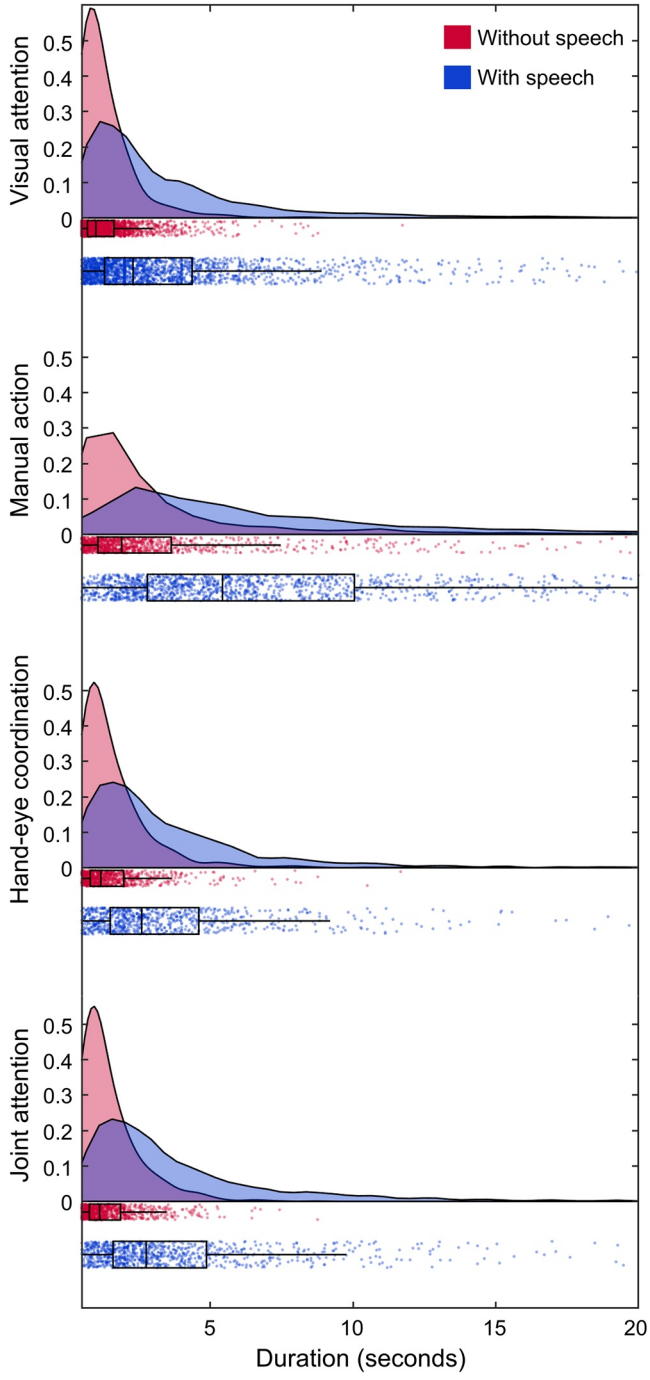



**FIGURE 5** Raincloud plots (Allen et al., 2019) showing the durations of visual attention, manual actions, hand-eye coordination, and joint attention that were accompanied by parent speech (blue) or unaccompanied by parent speech (red). Each type of multimodal attention is represented with a half-violin kernel density plot, a boxplot with mean and the first and third quartiles, as well as "raindrops" for each data point.

co-occurring with speech (Figure 5). Further, sustained bouts of each type of multimodal attention were also longer and more likely to occur with speech. Outputs from all regressions are reported in Table 1.

The mean duration of visual attention bouts with speech was nearly 3 times longer than the mean duration of bouts without speech ($M_{\text{with-speech}} = 3.604$ s, $M_{\text{w/o-speech}} = 1.333$ s, $p < 0.001$). When we examined the subset of visual attention bouts that were longer than 3s, the proportion of bouts with speech that met the sustained attention threshold was substantially higher than bouts without speech (With = 0.394, Without = 0.064). Further, the durations of sustained visual attention with speech were longer than instances without speech ($M_{\text{with-speech}} = 6.730$, $M_{\text{w/o-speech}} = 4.295$ s, $p < 0.001$).

The mean duration of manual action bouts with speech was more than double the mean duration of bouts without speech ($M_{\text{with-speech}} = 9.094$ s, $M_{\text{w/o-speech}} = 3.665$ s, $p < 0.001$). Most bouts of manual action with speech were sustained for 3s or more (0.733), while less than one-third of manual actions bouts without speech were 3s or more (0.308). As with visual attention, sustained manual action bouts that co-occurred with parent speech were longer than instances of sustained manual actions without parent speech ($M_{\text{with-speech}} = 11.734$ s, $M_{\text{w/o-speech}} = 8.617$ s, $p < 0.001$).

The mean duration of hand-eye coordination bouts with speech was double the duration of bouts without speech ($M_{\text{with-speech}} = 3.734$ s, $M_{\text{w/o-speech}} = 1.592$ s, $p < 0.001$). A higher proportion of hand-eye coordination bouts with parent speech were sustained for three or more seconds (With = 0.431, Without = 0.100). Sustained hand-eye coordination bouts increased in duration when accompanied by parent speech ($M_{\text{with-speech}} = 6.465$, $M_{\text{w/o-speech}} = 4.542$, $p < 0.001$).

Finally, the mean duration of JA bouts with speech was more than double the mean duration of JA bouts without speech ($M_{\text{with-speech}} = 3.979$ s, $M_{\text{w/o-speech}} = 1.479$ s, $p < 0.001$). Nearly half of all JA bouts with parent speech were sustained for three or more seconds (0.456), while few instances of JA without speech lasted longer than 3 s (0.092). Sustained JA bouts that co-occurred with speech were significantly longer than bouts that did not co-occur with speech ($M_{\text{with-speech}} = 6.660$, $M_{\text{w/o-speech}} = 4.045$, $p < 0.001$).

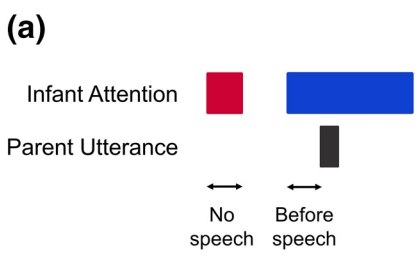
For all four measures of infant and dyadic multimodal attention, the duration of attention is longer when the bout is accompanied by parent speech. Moreover, when we specifically examined sustained

**TABLE 1** Descriptive statistics and modeling for corpus analyses in Study 1

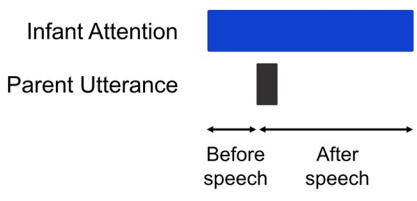
| | With speech | | Without speech | | Speech ~ dur + (1\|sub) + (1\|obj) | | |
|---|---|---|---|---|---|---|---|
| **Modality** | ***M*** | **SD** | ***M*** | **SD** | ***β*** | ***p*** | **Null model comparison** |
| Visual attention | | | | | | | |
| All instances | 3.604 | 3.892 | 1.332 | 1.007 | 0.786 | <0.001 | $\chi^2(1) = 1184$, $p < 0.001$ |
| Sustained attention (>3s) | 6.730 | 4.650 | 4.295 | 1.411 | 0.499 | <0.001 | $\chi^2(1) = 93.997$, $p < 0.001$ |
| Manual action | | | | | | | |
| All instances | 9.094 | 13.162 | 3.665 | 5.849 | 0.138 | <0.001 | $\chi^2(1) = 286.59$, $p < 0.001$ |
| Sustained attention (>3s) | 11.734 | 14.516 | 8.617 | 8.656 | 0.03 | <0.001 | $\chi^2(1) = 17.959$, $p < 0.001$ |
| Hand-eye coordination | | | | | | | |
| All instances | 3.734 | 3.876 | 1.592 | 1.269 | 0.604 | <0.001 | $\chi^2(1) = 353.67$, $p < 0.001$ |
| Sustained attention (>3s) | 6.465 | 4.595 | 4.542 | 1.718 | 0.327 | <0.001 | $\chi^2(1) = 28.828$, $p < 0.001$ |
| Joint attention | | | | | | | |
| All instances | 3.979 | 3.980 | 1.479 | 1.053 | 0.816 | <0.001 | $\chi^2(1) = 521.01$, $p < 0.001$ |
| Sustained attention (>3s) | 6.660 | 4.576 | 4.045 | 1.090 | 0.746 | <0.001 | $\chi^2(1) = 70.114$, $p < 0.001$ |

multimodal attention bouts, we saw that not only is sustained attention substantially more likely to occur with parent speech, but that bouts of sustained attention with parent speech are significantly longer. This finding, however, does not necessarily indicate that parent speech is supporting the infant's ability to sustain multimodal attention. Instead, multimodal attention that co-occurs with parent speech could be longer for one of two reasons, (1) longer bouts of attention afford more opportunities for coincidental overlap with parent speech, or (2) parent speech extends the duration of infant multimodal attention (Figure 6).

To examine these two hypotheses, we calculated "before speech"—or the interval between the onset of attention and the onset of speech—and the "after speech"—the interval of attention after the parent began speaking, for each instance of attention that co-occurred with parent speech. If hypothesis 1 is correct, then the onset of parent speech could occur at any time during a bout of attention and, as there are greater opportunities for coincidental overlap with increasing duration, the average duration of "before speech" would be greater than the duration of attention without speech. If hypothesis 2 is correct, then the duration of attention before speech should be shorter than the duration without speech. For each modality, we see that the duration of attention before speech is shorter than attention without speech—suggesting that parents are not just simply responding to longer bouts of infant attention. Further, if parent speech is *extending* infant behaviors, then the duration of attention after speech should be longer than the duration of attention before speech. Again, we see evidence that parent speech extends all four types of multimodal attention, as the "after speech" durations are significantly longer than the "before speech" durations. Our findings provide support that hypothesis 2 is correct—parent speech is associated with extended durations of infant multimodal attention.




| | No speech | Before speech | Comparison |
|---|---|---|---|
| Visual | 1.333s | 1.202s | β = -0.100, *p* <0.001; χ2(1)= 11.48, p < 0.001 |
| Manual | 3.667s | 1.846s | β = -0.200, *p* <0.001; χ2(1)= 162.80, p < 0.001 |
| Hand-eye | 1.593s | 1.201s | β = -0.288, *p* <0.001; χ2(1)= 41.47, p < 0.001 |
| Joint | 1.481s | 1.268s | β = -0.164, *p* <0.001; χ2(1)= 12.86, p < 0.001 |




| | Before speech | After speech | Comparison |
|---|---|---|---|
| Visual | 1.202s | 2.403s | β = 0.291, *p* <0.001; χ2(1)= 273.94, p < 0.001 |
| Manual | 1.846s | 7.248s | β = 0.323, *p* <0.001; χ2(1)= 622.09, p < 0.001 |
| Hand-eye | 1.201s | 2.533s | β = 0.374, *p* <0.001; χ2(1)= 149.29, p < 0.001 |
| Joint | 1.268s | 2.711s | β = 0.335, *p* <0.001; χ2(1)= 165.26, p < 0.001 |

**FIGURE 6** To test the hypothesis that parent speech extends infant multimodal attention we (a) compared the duration of attention without parent speech (red) to the lag between the onset of the bout of attention and the onset of speech ("before speech", blue) and found the duration of multimodal attention before speech was significantly shorter than bouts of multimodal attention without speech. (b) We also compared the durations of infant multimodal attention before and after the parent utterance began and found that attention after speech was significantly longer than attention before.

### 2.2.2 | Qualitative differences in the effects of parent speech

Studying behaviors at the event level allows us to gain a better understanding of proximate mechanisms—parental responsiveness matters *because* it is providing scaffolding to extend infant multimodal attention. Nonetheless, differences in parental responsiveness are robust and predictive of infant's language outcomes (e.g., Tamis-LeMonda et al., 2001). To begin understanding the influences of quantitative differences in responsive parent speech on infant multimodal attention, we examined whether more or less parent talk (*quantity*, by grouping at the dyadic level) has different effects on the infant's ability to sustain multimodal attention (*quality*, still analyzed at the event level).

Parents varied in how much they spoke to their infants. While the average parent produced 17.106 utterances/minute, the range of parent talk was quite wide (9.597 to 25.144 spoken utterances/minute). To test the relationship between parent speech and infant sustained multimodal attention, we divided the subjects into two groups based on a median split (median = 17.072 utterances/min). Parents in the more talkative group produced 20.091 utterances/minute on average while parents in the less talkative speech group produced on average 14.121 utterances/minute. The two groups did not different in infant age or duration of experiment ($ps > 0.553$). The less talkative and more talkative groups did not differ in the mean duration of parent utterances ($p = 0.947$, overall mean = 1.319 s), suggesting less talkative parents were truly producing less speech, not just fewer, longer utterances. Therefore, the duration of spoken utterances in the two groups was not a factor that would differentially influence infant's attention. Defining quantity of parent talk at the group level allowed us to look for potential trade-offs in quality and quantity in our moderately sized dataset, accounting in particular for potential variability in what parents are actually saying when they do speak to their infants.

Using the event-level datasets compiled for the more and less talkative groups, we compared the durations of infant multimodal attention that co-occurred with parent speech. There were significant differences in the effects parent speech had on infant behaviors in the less and more talkative groups. Although parents in the less talkative group spoke less to their infants, the durations of infants' multimodal attention in the less talkative group were longer than those of infants in the more talkative group (Table 2, Figure 7).

The mean duration of the less talkative group's visual attention bouts was significantly longer than mean duration of bouts of infants in the more talkative group ($M_{less} = 4.004$ s, $M_{more} = 3.354$ s, $p = 0.001$). A greater proportion of infants' visual attention crossed the 3s sustained attention threshold in the less talkative group (Less = 0.428, More = 0.373). Similarly, the duration of infant's manual actions was longer in the less talkative groups than the more talkative group ($M_{less} = 10.138$ s, $M_{more} = 8.275$ s, $p = 0.003$), and more actions crossed the 3s threshold in the less talkative group (Less = 0.764, More = 0.706). The duration of hand-eye coordination was longer in the less talkative group ($M_{less} = 4.270$ s, $M_{more} = 3.322$ s, $p = 0.004$) and more bouts were classified as sustained

**TABLE 2** Descriptive statistics and modeling to compare the speech groups in Study 1

| | Less talkative group | | More talkative group | | Dur ~ group + (1\|obj) | | |
|---|---|---|---|---|---|---|---|
| **Modality** | ***M*** | **SD** | ***M*** | **SD** | ***β*** | ***p*** | **Null model comparison** |
| Visual attention | 4.004 | 4.504 | 3.354 | 3.432 | −0.583 | 0.001 | $\chi^2(1) = 10.699$, $p = 0.001$ |
| Manual action | 10.138 | 13.682 | 8.275 | 12.687 | −2.124 | 0.003 | $\chi^2(1) = 9.08$, $p = 0.003$ |
| Hand-eye coordination | 4.270 | 4.516 | 3.322 | 3.246 | −0.791 | 0.004 | $\chi^2(1) = 8.238$, $p = 0.004$ |
| Joint attention | 4.437 | 4.840 | 3.681 | 3.275 | −0.593 | 0.028 | $\chi^2(1) = 4.855$, $p = 0.028$ |

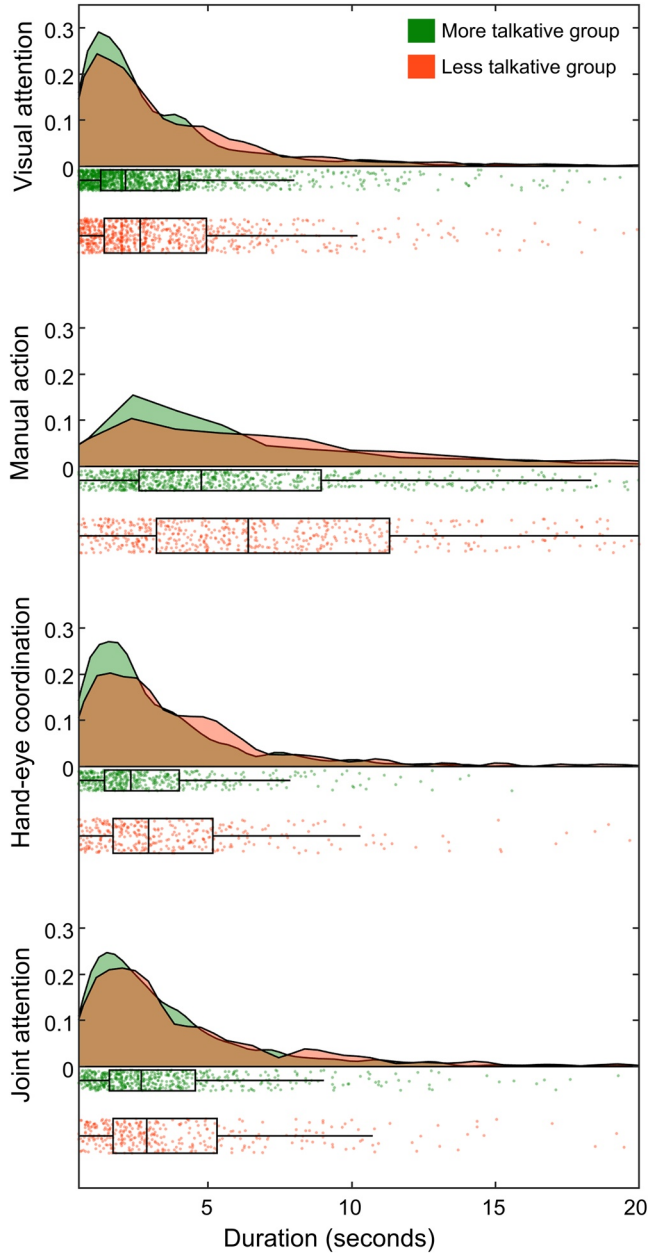



**FIGURE 7** Raincloud plots showing the durations of visual attention, manual action, hand-eye coordination, and joint attention that were accompanied by parent speech for the more talkative group (green) or the less talkative group (orange).

(Less = 0.494, More = 0.383). Finally, the duration of JA was longer in the less talkative group ($M_{less}$ = 4.437 s, $M_{more}$ = 3.681 s, $p$ = 0.028) and more bouts were classified as sustained in the less talkative group (Less = 0.480, More = 0.441).

Lastly, we found that there is equal opportunity for parents to respond to infant multimodal attention in the two groups. For each multimodal attention type, we compared the mean duration and frequency of all bouts (i.e., both with and without responsive speech) for infants in the less and more

talkative groups and found that the groups did not differ in duration ($ps > 0.263$) nor frequency (bouts per minute, $ps > 0.3415$). Similarly, infant age was not correlated with the mean duration ($ps > 0.111$) or frequency ($ps > 0.399$) of any of the four measured attention types.

### 2.2.3 | Discussion

In Study 1, we observed that parent speech scaffolds infant's multimodal attention. We present evidence of two groups of dyads, classified by how much speech parents produced in an interaction. These two groups coordinate their attention in different ways—in the less talkative group, there are less occurrences of speech-attention overlap. But, when parent speech co-occurs with the infant's attention, infants in the less talkative group had significantly longer durations of attention than infants in the more talkative group. Taken together, these results provide a new window into why parental responsiveness matters. Responsiveness is often used as a construct to assess the quality of language input. Here, we defined parental responsiveness as parent speech that co-occurs with infant multimodal attention, grounding it in the dynamics of an interaction. In doing so, we were able to demonstrate that responsiveness has real-time effects on infant multimodal attention—and that differences in responsiveness correspond to different effects on attention. We posit that the social extension of infant's sustained multimodal attention is one mechanism through which parental responsiveness leads to better language outcomes.

To test the replicability of the effect parent speech has in parent-infant social interactions, a second study was conducted in a more naturalistic laboratory environment (Figure 2b, Figure 8). The HOME Lab (Home-Like Observational Multisensory Environment) was a large, open room that was designed to look like an apartment. Furniture created three distinct spaces: a kitchenette with a sink, counter-top, refrigerator, and table and chairs; a living room with a couch, television, rug, and side table; and a play area with a colorful floor, surrounded by soft blocks, and with other toys and stuffed animals nearby. The entire space had art hanging on the walls, as well as seven third-person view cameras and microphones on the walls and ceilings. The home-like lab is, of course, still a novel environment for dyads that will elicit behaviors that are different than when they are comfortably at home. Nonetheless, the space is more naturalistically cluttered than a typical lab set-up, including the one used in Study 1. While more closely approximating a naturalistic home, we were still able to tightly control the environment dyads interacted in. There were additional methodological differences

**FIGURE 8** Photographs of the HOME Lab used in Study 2.

between Study 1 and Study 2—participants played with 10 toys (instead of 24) and the average age of participants was lower. Perhaps most importantly, a new wireless eye-tracker system was used in Study 2 that allowed participants to move around freely, rather than be seated on the floor during the entire experiment. Because human behavior is context-dependent, we did not expect to find the same patterns and frequencies of dyadic behavior across these two studies. However, by conducting the same analyses on data from a different toy play experiment, our goal was to examine whether the effect of parent speech would be observed in another context, and therefore be generalizable to many social interaction contexts in early development.

# 3 | STUDY 2

## 3.1 | Methods

### 3.1.1 | Participants

Fifty parents and infants (23 female) were recruited from the same population as in Study 1, but no dyads participated in both Study 1 and Study 2. Infant participants were an average of 16.6 months old (SD = 3.83, range: 12.4–25.8). Overall, participants in Study 2 were significantly younger than in Study 1 ($t(80.386) = 2.700$, $p = 0.008$).

### 3.1.2 | Data collection

Although data collection procedures in Study 1 and Study 2 were similar, there are a few differences worth noting. Study 2 was conducted in the HOME Lab environment. As in Study 1, dyads were asked to play for 10 min, which resulted in an average of 6.68 min of useable data (range: 1.95—11.26). Infant age did not correlate with duration of experiment ($p = 0.072$). Dyads were given 10 toys that were different than the toys used in Study 1 and wore wireless head-mounted eye trackers (Pupil Labs). Parents wore the eye tracker in the conventional set-up, like a pair of glasses. Parents did not need to remove corrective lens to wear the eye tracker. Infants wore a modified eye tracker that was affixed to a hat. The eye trackers were each connected with a short USB-C cord to an Android phone. Participants wore jackets while they played, and the mobile phones were placed in a secure pocket on the participant's back. The wireless eye trackers allowed participants to move freely around the play space. Parents were instructed to play as they would at home with their infant and were asked to remain in the play space (i.e., not move into the other parts of the HOME Lab). Participants varied in the amount of time they spent moving around the play space, as well as in the amount of time they spent engaged with the 10 toys. Participants also wore motion sensors while they played (see Schroer & Yu, 2021).

### 3.1.3 | Data processing and analysis

Data was processed and analyzed in a manner identical to Study 1. The only difference in data processing was that 10 regions of interest (ROIs) were used for coding participant's attention to the different toys available for play. The parents in Study 2 spoke an average of 17.352 times/minute (SD = 4.885) and each utterance lasted an average of 1.037 s (SD = 0.217). As in Study 1, neither infant age nor

experiment duration correlated with frequency of parent speech or mean duration of utterances ($ps > 0.109$).

## 3.2 | Results

### 3.2.1 | Effects of parent speech

As in Study 1, infant multimodal attention was extended by parent speech (Table 3). The duration of infant visual attention with speech was more than double the duration of visual attention without speech ($M_{\text{with-speech}} = 3.253$ s, $M_{\text{w/o-speech}} = 1.448$ s, $p < 0.001$). Visual attention was also more likely to be sustained when it co-occurred with parent speech. More than a third of visual attention bouts accompanied by parent speech passed the sustained attention threshold, while only 8% of visual attention bouts without speech were longer than 3s (With = 0.383, Without = 0.079). These bouts of sustained visual attention were also longer when they co-occurred with parent speech ($M_{\text{with-speech}} = 5.951$ s, $M_{\text{w/o-speech}} = 4.658$ s, $p < 0.001$).

Infant manual actions that their parents responded to were more than twice as long as bouts of manual action that did not co-occur with parent speech ($M_{\text{with-speech}} = 8.675$, $M_{\text{w/o-speech}} = 3.604$ s, $p < 0.001$). Manual actions that occurred with speech were also far more likely to be over the 3s sustained attention threshold (With = 0.706, Without = 0.295) and the sustained bouts of manual action were significantly longer when accompanied by parent speech ($M_{\text{with-speech}} = 11.511$ s, $M_{\text{w/o-speech}} = 8.764$, $p = 0.001$).

Similarly, durations of hand-eye coordination were more than twice as long when accompanied by parent speech ($M_{\text{with-speech}} = 3.216$ s, $M_{\text{w/o-speech}} = 1.522$ s, $p < 0.001$) and substantially more instances of hand-eye coordination with parent speech were considered sustained attention than bouts without speech (With = 0.391, Without = 0.089). Sustained hand-eye coordination lasted longer if it co-occurred with speech ($M_{\text{with-speech}} = 5.660$s, $M_{\text{w/o-speech}} = 4.659$, $p = 0.001$).

Finally, durations of JA were twice as long when accompanied by parent speech ($M_{\text{with-speech}} = 2.977$ s, $M_{\text{w/o-speech}} = 1.455$ s, $p < 0.001$). A third of JA bouts that co-occurred with speech were sustained past

**TABLE 3** Descriptive statistics and modeling for corpus analyses in Study 2

| | With speech | | Without speech | | Speech ~ dur + (1\|sub) + (1\|obj) | | |
|---|---|---|---|---|---|---|---|
| **Modality** | ***M*** | **SD** | ***M*** | **SD** | ***β*** | ***p*** | **Null model comparison** |
| Visual attention | | | | | | | |
| All instances | 3.253 | 2.958 | 1.448 | 1.231 | 0.628 | <0.001 | $\chi^2(1) = 933.74, p < 0.001$ |
| Sustained attention (>3s) | 5.951 | 3.214 | 4.658 | 1.998 | 0.241 | <0.001 | $\chi^2(1) = 38.715, p < 0.001$ |
| Manual action | | | | | | | |
| All instances | 8.675 | 12.354 | 3.604 | 8.050 | 0.102 | <0.001 | $\chi^2(1) = 239.12, p < 0.001$ |
| Sustained attention (>3s) | 11.511 | 13.740 | 8.764 | 13.458 | 0.019 | 0.001 | $\chi^2(1) = 12.559, p < 0.001$ |
| Hand-eye coordination | | | | | | | |
| All instances | 3.216 | 2.685 | 1.522 | 1.273 | 0.621 | <0.001 | $\chi^2(1) = 378.29, p < 0.001$ |
| Sustained attention (>3s) | 5.660 | 2.808 | 4.659 | 1.945 | 0.215 | 0.001 | $\chi^2(1) = 13.121, p < 0.001$ |
| Joint attention | | | | | | | |
| All instances | 2.977 | 2.622 | 1.455 | 1.143 | 0.633 | <0.001 | $\chi^2(1) = 323.65, p < 0.001$ |
| Sustained attention (>3s) | 5.610 | 2.954 | 4.411 | 1.732 | 0.287 | <0.001 | $\chi^2(1) = 16.143, p < 0.001$ |

3s (0.341), though few bouts of JA without speech were sustained (0.077). Sustained JA lasted longer if it co-occurred with speech ($M_{\text{with-speech}} = 5.610$ s, $M_{\text{w/o-speech}} = 4.411$ s, $p < 0.001$).

Lastly, we compared the time "before speech", time "after speech", and the duration of attention bouts without parent speech. For all four measures of multimodal attention, "before speech" was shorter than attention without speech. Additionally, the durations of infant attention "after speech" were significantly longer than attention before speech began. As in Study 1, these analyses suggest that parent speech is extending the duration of infant multimodal attention—as opposed to speech simply co-occurring with longer infant attention (Table 4).

### 3.2.2 | Qualitative differences in the effects of parent speech

We then split the participants in Study 2 in two different groups, based on the talkativeness of parents. Frequency of parent speech ranged from 7.475 utterances/min to 27.151 utterances/min, with a median of 16.340 utterances/min. The less talkative group spoke an average of 13.312 utterances/minute and the more talkative group spoke an average of 21.392 utterances/minute. The groups did not differ in the mean duration of utterances ($M = 1.037$ s, $p = 0.100$), infant age ($p = 0.684$), or duration of experiment ($p = 0.617$).

We compared the durations of infant multimodal attention that co-occurred with parent speech across the less talkative and more talkative speech groups. Our findings from Study 1 were partially replicated in this new laboratory environment (Table 5). Infants in the less talkative group had longer durations of visual attention than their peers in the more talkative group ($M_{\text{less}} = 3.595$ s, $M_{\text{more}} = 2.982$ s, $p < 0.001$). Infants in the less talkative group also had longer durations of hand-eye coordination with speech than infants in the more talkative group ($M_{\text{less}} = 3.464$ s, $M_{\text{more}} = 3.002$ s, $p = 0.014$). Additionally, dyads in the less talkative group had longer durations of JA with speech than dyads in the more talkative group ($M_{\text{less}} = 3.241$ s, $M_{\text{more}} = 2.772$ s, $p = 0.005$). We did not find differences in the durations of manual action between the less and more talkative speech groups ($M_{\text{less}} = 9.128$, $M_{\text{more}} = 8.292$ s, $p = 0.219$).

As in Study 1, we found that there was equal opportunity for parents to respond to infant multimodal attention in the two groups. For each multimodal attention type, we compared the mean duration and frequency of all bouts (i.e., both with and without responsive speech) for infants in the less and more talkative groups and found that the groups did not differ in duration ($ps > 0.333$) and frequency ($ps > 0.060$). Similarly, infant age was not correlated with the mean duration ($ps > 0.273$) or frequency ($ps > 0.195$) of any of the four measured attention types.

### 3.2.3 | Discussion

In Study 2, we were able to replicate most of our findings in Study 1 in a different laboratory environment with a significantly younger sample. Parent speech extended the durations of infants' visual attention, manual action, and hand-eye coordination, as well as dyadic JA—and this effect differed in dyads with more and less talkative parents. It is important to note that for the presented results in Study 1 and Study 2, we report uncorrected *p*-values. Although this may impact the interpretation of our findings, we did replicate our results across two samples in different laboratory environments. Taken together, these two studies suggest that the scaffolding of infant multimodal attention through parent speech is a robust phenomenon.

**TABLE 4** Comparisons to test the hypothesis that speech extends infant attention in Study 2

| Modality | No speech | Before speech | After speech | Before speech versus no speech | | | | Before versus after speech | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $\beta$ | $p$ | $\chi^2(1)$ | $p$ | $\beta$ | $p$ | $\chi^2(1)$ | $p$ |
| Visual attention | 1.448 | 1.142 | 2.111 | −0.219 | <0.001 | 58.402 | <0.001 | 0.323 | <0.001 | 264.770 | <0.001 |
| Manual action | 3.606 | 1.823 | 6.851 | −0.176 | <0.001 | 144.670 | <0.001 | 0.280 | <0.001 | 572.320 | <0.001 |
| Hand-eye coordination | 1.524 | 1.147 | 2.069 | −0.308 | <0.001 | 47.921 | <0.001 | 0.350 | <0.001 | 119.450 | <0.001 |
| Joint attention | 1.456 | 1.058 | 1.920 | −0.368 | <0.001 | 53.340 | <0.001 | 0.377 | <0.001 | 123.270 | <0.001 |

**TABLE 5** Descriptive statistics and modeling to compare the speech groups in Study 2

| | Less talkative group | | More talkative group | | Dur ~ group + (1\|obj) | | |
|---|---|---|---|---|---|---|---|
| **Modality** | ***M*** | **SD** | ***M*** | **SD** | ***β*** | ***p*** | **Null model comparison** |
| Visual attention | 3.595 | 3.215 | 2.982 | 2.708 | −0.612 | <0.001 | $\chi^2(1) = 21.706, p < 0.001$ |
| Manual action | 9.128 | 12.035 | 8.292 | 12.612 | −0.768 | 0.219 | $\chi^2(1) = 1.524, p = 0.217$ |
| Hand-eye coordination | 3.464 | 2.838 | 3.002 | 2.530 | −0.441 | 0.014 | $\chi^2(1) = 6.127, p = 0.013$ |
| Joint attention | 3.241 | 2.802 | 2.772 | 2.455 | −0.497 | 0.005 | $\chi^2(1) = 8.013, p = 0.005$ |

## 4 | GENERAL DISCUSSION

To study the mechanisms through which parental responsiveness shapes the sensorimotor and social experiences of the infant, we measured the association between parent speech and extended infant multimodal attention and dyadic JA. Expanding on prior work that has primarily focused on parents' visual attention (e.g., Yu & Smith, 2016), we found that parent speech scaffolds infants' ability to sustain visual attention. Further, we found evidence that parental responsiveness shapes infant's sensorimotor and social experiences across multiple levels, including visual attention, manual actions, hand-eye coordination, and dyadic JA. With these results, we suggest that one pathway through which parental responsiveness influences infant language and cognitive development is by this real-time shaping of infant multimodal attention.

### 4.1 | The role of parental responsiveness

We can begin to unpack this mechanistic mystery by studying responsiveness at the micro-level of an interaction and asking what happens when a parent talks to their infant. Responsive speech from parents contributes to language learning in two ways—providing linguistic input and shaping infants' sensorimotor and social experiences. While past research has often focused on the linguistic input (e.g., Rowe, 2012; Schwab et al., 2018; Tamis-LeMonda et al., 2001), the present study examined the effects of parent speech on infants' multimodal attention. Parent speech does far more than "just" provide the linguistic information for their child's language learning, also shaping the moment-by-moment dynamics of interactions. Through this pathway, parental responsiveness may impact early word learning at multiple levels.

One open question, however, is whether the two pathways created by responsive speech—the linguistic input and the effects on multimodal attention—contribute to language and cognitive development independently. One possibility is that different types of parent speech may play different roles. Some types of parent speech may contain rich linguistic information, and other types of parent speech may serve the role of socially boosting infant's multimodal attention. It is also plausible that a single utterance in parent speech may serve both roles—providing linguistic input and impacting infants' multimodal attention. Parent speech during high-quality attention (extended bouts and/or informative views) is directly linked to real-time word learning (Pereira et al., 2014; Schroer & Yu, 2022; Yu & Smith, 2012), suggesting that the influences of the linguistic and multimodal pathways may be coupled in time. Future work should endeavor to understand the extent to which these two pathways can be disentangled. A related question is whether infants' endogenous attention and the attention socially scaffolded by parents serve the same role in learning (see Wass et al., 2018). Infants who are

able to show sustained attention endogenously may not need much of a social boost from parents, and instead parents may just use speech to provide linguistic input at the right moments created by infants. For other infants who may not be able to establish sustained attention by themselves, they would still need responsive speech from parents to achieve this goal. Thus, the role of responsive speech from parents may evolve through time, depending on the development of individual infants' self-regulation and language knowledge.

## 4.2 | Individual differences in responsive parent speech

The group differences in the effects of parent speech offer insight into how differences in the quality and quantity of responsiveness might lead to differences in developmental outcomes. Parents that spoke infrequently during the interaction seemed to have more potent effects on their infants' multimodal attention than their more talkative parent peers, as infants in the less responsive group sustained their attention for longer periods of time. This relationship between the amount of parent speech and the scaffolding of infant attention could be explained by two accounts, which are not necessarily mutually exclusive. First, less talkative parents may be more selective in when they choose to talk. These parents may be finding more optimal moments to scaffold infant attention to objects and selectively talk during those instances. Alternatively, parents that talk less may be more responsive when they do choose to talk. Less talkative parents' utterances may be more semantically relevant to their infant's behavior by following their child's focus of attention, rather than redirecting infant attention to different objects (see McQuillan et al., 2019). Thus, less talkative parents may even have a different signal-to-noise ratio than more talkative parents, so that any given utterance is a more salient and rewarding signal. Recent work with macaques suggests that rare rewards had increased dopamine responses and improved learning (Rothenhoefer et al., 2021). By responding less frequently to their infant's attention, the rewarding response from a less talkative parent is rarer—and may be more potent as a result.

Regardless, both patterns of dyadic behavior create many opportunities for high-quality infant attention and dyadic coordination suggesting (at least) two pathways for parental responsiveness to support language learning and cognitive development. More parent speech provides more instances of scaffolding, whereas less parent speech leads to more effective scaffolding. Further research is needed to better understand the quantitative and qualitative differences in the parent speech that forms these two pathways, as well as how individual dyads adjust and adapt their own pathways and interaction patterns based on the history of their interactive experiences. Defining talkativeness as a dichotomous group variable prevents us from forming any conclusions about important individual differences in this effect. It is also worth noting that we do not have any baseline measures of infant sustained multimodal attention (such as in a play-alone session), so we cannot rule out that infants in the less talkative group sustain their attention longer for reasons independent of their parent's responsiveness. Additionally, the present study only focuses on one sequence of dyadic behavior: infant acts and the parent responds. More research needs to be done to understand how infants shape their own learning experiences, especially by eliciting changes in parent behavior (e.g., Elmlinger et al., 2019).

## 4.3 | Replicable findings from natural behavior

Our results also demonstrate that high-density data at the sensorimotor level produces reliable and replicable results. Not only did we replicate the effects of parent speech in two unique lab

environments, but we observed behaviors previously only described in a more sterile, "white room" lab (e.g., Suarez-Rivera et al., 2019; Yu & Smith, 2016). Even though these studies differ in several ways (e.g., the number of toys used, the lab environment), they share one critical property—parents and infants interacted in free-flowing play. Their behaviors are naturalistic as they are not elicited or constrained by experimenters and experimental stimuli. One open question is whether our findings would be generalized to other activities beyond toy play. The HOME lab environment is designed to answer this question as we can collect data from parent-infant interactions across a diversity of tasks, including book reading, play with more interactive toys (like a ball maze), meal preparation and feeding (Peters et al., 2020), and "grooming" the infant (after particularly messy snack times). Expanding our studies of dyadic behavior to different routines will improve our understanding of parental responsiveness in different contexts and the mechanisms underlying language learning and cognitive development (Schroer et al., 2022; Tamis-LeMonda et al., 2019).

## 4.4 | Limitations and future work

The present study contributes to the growing evidence that the causal link between parental responsiveness and real-time infant attention does exist. To test the whole pathway—from responsiveness to attention to learning outcomes—future work should endeavor to include real-time word learning tasks as well as longitudinal measures of language and cognitive development. Evidence of this mechanistic pathway could also be provided from studies that directly manipulate parental responsiveness to see how it effects infant's attention (see McQuillan et al., 2019). To better understand the roles of temporal contingency and semantic relevancy, we also need to analyze the content of responsive parent speech. Further, we focus on parent responses to infant multimodal attention, but we know there are many other behaviors that shape the dynamics of parent-infant interactions, such as infant vocalizations (e.g., Chang et al., 2017). Like the modalities we studied, parent responses to infant vocalizations can predict language outcomes (Tamis-LeMonda et al., 2001; Wu & Gros-Louis, 2014), shift patterns of behavior in real-time (Miller & Gros-Louis, 2013), and even lead to the real-time learning of novel vocal forms (Goldstein & Schwade, 2008). As there is dramatic communicative development in the second year of life, which likely shifts how and whether parents respond to infant vocalizations, studying the effects parental responsiveness has on vocalizations is fodder for many future studies.

We found similar effects of parent responses in two different play contexts, but one open question is if these findings could be generalized to other contexts and activities, demographics, and age groups. The participants from each study were recruited from the same community. To further demonstrate the robustness of this finding, replications by other researchers working with demographically diverse populations are needed. Nonetheless, recent work found no differences in real-time patterns of visual attention, manual actions, and hand-eye coordination in toddlers with and without an autism-spectrum disorder diagnosis (Yurkovic et al., 2021), suggesting that micro-level behaviors maybe more consistent across groups than their macro-level differences might imply. Additionally, although there are many developmental milestones in the second year of life that could significantly change the dynamics of an interaction, we found no correlations between subject age and any of our dependent measures. In this first study, we deliberately chose to recruit a wide age to explore the effects of parental responsiveness across this developmental period. Nonetheless, in future work we will study more narrow age ranges to see how developmental changes in the second year of life may alter parental responsiveness.

## 5 | CONCLUSION

The present study provides new evidence on the real-time effects parent speech has on infants' multimodal attention and dyadic coordination. The scaffolding of infant attention and actions through parent speech creates rich opportunities for infants to learn language and about the world around them, yielding a developmental cascade that links parental responsiveness to developmental outcomes.

## ACKNOWLEDGMENTS

This work was supported by NIH R01HD074601 and R01HD093792 to CY. SES was supported by the NSF GRFP (DGE-1610403) and NIH T32HD007475. The authors declare no conflicts of interest with regard to the funding source for this study. We thank Linda B. Smith, the Computational Cognition and Learning Lab at Indiana University, and the Developmental Intelligence Lab at the University of Texas at Austin—especially Daniel Pearcy, Hannah Burrell, Dian Zhi, Tian (Linger) Xu, Julia Yurkovic-Harding, Catalina Suarez-Rivera, Anting Chen, Grace Lisandrelli, Lauren Slone, Drew Abney, Andrei Amatuni, and Jeremy Borjon—for their support in data collection, coding, and many fruitful discussions.

## ORCID

*Sara E. Schroer* https://orcid.org/0000-0002-6139-060X

## REFERENCES

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2019). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, *4*, 63. https://doi.org/10.12688/wellcomeopenres.15191.1

Bambach, S., Smith, L. B., Crandall, D. J., & Yu, C. (2016). Objects in the center: How the infant's body constrains infant scenes. *Proceedings of the sixth joint IEEE international conference on developmental learning and epigenetic robotics*.

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

Belsky, J., Goode, M. K., & Most, R. K. (1980). Maternal stimulation and infant exploratory competence: Cross-sectional, correlational, and experimental analyses. *Child Development*, *51*(4), 1168–1178. https://doi.org/10.2307/1129558

Chang, L., de Barbaro, K., & Deák, G. (2017). 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

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

Elmlinger, S. L., Schwade, J. A., & Goldstein, M. H. (2019). The ecology of prelinguistic vocal learning: Parents simplify the structure of their speech in response to babbling. *Journal of Child Language*, *46*(5), 998–1011. https://doi.org/10.1017/s0305000919000291

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

Gilkerson, J., Richards, J. A., Warren, S. F., Oller, D. K., Russo, R., & Vohr, B. (2018). Language experience in the second year of life and language outcomes in late childhood. *Pediatrics*, *142*(4), e20174276. https://doi.org/10.1542/peds.2017-4276

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*(5), 515–523. https://doi.org/10.1111/j.1467-9280.2008.02117.x

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

Kannass, K. N., & Oakes, L. M. (2008). The development of attention and its relations to language in infancy and toddlerhood. *Journal of Cognition and Development*, *9*(2), 222–246. https://doi.org/10.1080/15248370802022696

Landry, S. H., Smith, K. E., Swank, P. R., Assel, M. A., & Vellet, S. (2001). Does early responsive parenting have a special importance for children's development or is consistency across early childhood necessary? *Developmental Psychology*, *37*(3), 387–403. https://doi.org/10.1037/0012-1649.37.3.387

Lawson, K. R., & Ruff, H. A. (2004). Early focused attention predicts outcome for children born prematurely. *Journal of Developmental and Behavioral Pediatrics*, *25*(6), 399–406. https://doi.org/10.1097/00004703-200412000-00003

Masur, E. F., Flynn, V., & Eichorst, D. L. (2005). Maternal responsive and directive behaviours and utterances as predictors of children's lexical development. *Journal of Child Language*, *32*(1), 63–91. https://doi.org/10.1017/s0305000904006634

McQuillan, M. E., Smith, L. B., Yu, C., & Bates, J. E. (2019). Parents influence the visual learning environment through children's manual actions. *Child Development*, *91*(3), e701–e720. https://doi.org/10.1111/cdev.13274

Mesman, J. (2010). Maternal responsiveness to infants: Comparing micro-and macro-level measures. *Attachment & Human Development*, *12*(1–2), 143–149. https://doi.org/10.1080/14616730903484763

Miller, J. L., & Gros-Louis, J. (2013). Socially guided attention influences infants' communicative behavior. *Infant Behavior and Development*, *36*(4), 627–634. https://doi.org/10.1016/j.infbeh.2013.06.010

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., Zhi, D., Petersen, M., & Yu, C. (2020). Active vision in the perception of actions: An eye tracking study in naturalistic contexts. *Proceedings of the 42nd annual meeting of the cognitive science society*.

Riksen-Walraven, J. M. (1978). Effects of caregiver behavior on habituation rate and self-efficacy in infants. *International Journal of Behavioral Development*, *1*(2), 105–130. https://doi.org/10.1177/016502547800100202

Rothenhoefer, K. M., Hong, T., Alikaya, A., & Stauffer, W. R. (2021). Rare rewards amplify dopamine responses. *Nature Neuroscience*, *24*(4), 465–469. https://doi.org/10.1038/s41593-021-00807-7

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, *83*(5), 1762–1774. https://doi.org/10.1111/j.1467-8624.2012.01805.x

Ruff, H. A., Lawson, K. R., Parrinello, R., & Weissberg, R. (1990). Long-term stability of individual differences in sustained attention in the early years. *Child Development*, *61*(1), 60–75. https://doi.org/10.2307/1131047

Schroer, S. E., Peters, R. E., Yarbrough, A., & Yu, C. (2022). Visual attention and language exposure during everyday activities: An at-home study of early word learning using wearable eye trackers. In *To appear in the proceedings of the 44th annual meeting of the cognitive science society*.

Schroer, S. E., & Yu, C. (2021). The sensorimotor dynamics of joint attention. *Proceedings of the 43rd annual meeting of the cognitive science society*.

Schroer, S. E., & Yu, C. (2022). Looking is not enough: Multimodal attention supports the real-time learning of new words. *Developmental Science*, e13290.

Schwab, J. F., Rowe, M. L., Cabrera, N., & Lew-Williams, C. (2018). Fathers' repetition of words is coupled with children's vocabularies. *Journal of Experimental Child Psychology*, *166*, 437–450. https://doi.org/10.1016/j.jecp.2017.09.012

Shannon, J. D., Tamis-LeMonda, C. S., London, K., & Cabrera, N. (2002). Beyond rough and tumble: Low-income fathers' interactions and children's cognitive development at 24 months. *Parenting: Science and Practice*, *2*(2), 77–104. https://doi.org/10.1207/s15327922par0202_01

Slone, L. K., Smith, L. B., & Yu, C. (2019). Self-generated variability in object images predicts vocabulary growth. *Developmental Science*, *22*(6), e12816. https://doi.org/10.1111/desc.12816

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

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development*, *72*(3), 748–767. https://doi.org/10.1111/1467-8624.00313

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

Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First Language*, *4*(12), 197–211. https://doi.org/10.1177/014272378300401202

Wass, S. V., Clackson, K., Georgieva, S. D., Brightman, L., Nutbrown, R., & Leong, V. (2018). Infants' visual sustained attention is higher during joint play than solo play: Is this due to increased endogenous attention control or exogenous stimulus capture? *Developmental Science*, *21*(6), e12667. https://doi.org/10.1111/desc.12667

West, K. L., & Iverson, J. M. (2017). Language learning is hands-on: Exploring links between infants' object manipulation and verbal input. *Cognitive Development*, *43*, 190–200. https://doi.org/10.1016/j.cogdev.2017.05.004

Wu, Z., & Gros-Louis, J. (2014). Infants' prelinguistic communicative acts and maternal responses: Relations to linguistic development. *First Language*, *34*(1), 72–90. https://doi.org/10.1177/0142723714521925

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., Smith, L. B., Shen, H., Pereira, A. F., & Smith, T. (2009). Active information selection: Visual attention through the hands. *IEEE Transactions on Autonomous Mental Development*, *1*(2), 141–151. https://doi.org/10.1109/tamd.2009.2031513

Yu, C., Suanda, S. H., & Smith, L. B. (2018). 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

Yurkovic, J. R., Lisandrelli, G., Shaffer, R. C., Dominick, K. C., Pedapati, E. V., Erickson, C. A., Kennedy, D. P., & Yu, C. (2021). Using head-mounted eye tracking to examine visual and manual exploration during naturalistic toy play in children with and without autism spectrum disorder. *Scientific Reports*, *11*(1), 1–14. https://doi.org/10.1038/s41598-021-81102-0

**How to cite this article:** Schroer, S. E., & Yu, C. (2022). The real-time effects of parent speech on infants' multimodal attention and dyadic coordination. *Infancy*, *27*(6), 1154–1178. https://doi.org/10.1111/infa.12500