RESEARCH ARTICLE

What are the building blocks of parent–infant coordinated attention in free-flowing interaction?

Drew H. Abney[1] | Sumarga H. Suanda[2] | Linda B. Smith[3,4] | Chen Yu[5]

[1]Department of Psychology, University of Georgia, Athens, Georgia

[2]Department of Psychological Sciences, University of Connecticut, Storrs, Connecticut

[3]Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana

[4]School of Psychology, University of East Anglia, Norwich, UK

[5]Department of Psychology, University of Texas at Austin, Austin, Texas

Correspondence
Drew H. Abney, Department of Psychology, University of Georgia, Athens, GA, USA.
Email: drew.abney@uga.edu

Funding information
National Institutes of Health, Grant/Award Number: K99/R00HD082358, R01HD074601 and T32HD007475-22; National Science Foundation, Grant/Award Number: BCS 0924248; Indiana University

Abstract
The present article investigated the composition of different joint gaze components used to operationalize various types of coordinated attention between parents and infants and which types of coordinated attention were associated with future vocabulary size. Twenty-five 9-month-old infants and their parents wore head-mounted eye trackers as they played with objects together. With high-density gaze data, a variety of coordinated attention bout types were quantitatively measured by combining different gaze components, such as mutual gaze, joint object looks, face looks, and triadic gaze patterns. The key components of coordinated attention that were associated with vocabulary size at 12 and 15 months included the simultaneous combination of parent triadic gaze and infant object looking. The results from this article are discussed in terms of the importance of parent attentional monitoring and infant sustained attention for language development.

1 | INTRODUCTION

Parents and infants during a social interaction look at each other and jointly at objects that are the topic of the interaction. A large literature on coordinated attention shows that looking patterns to one's social partner and to objects of joint interest are essential to successful human communication and for infants' language learning (Bakeman & Adamson, 1984; Brooks & Meltzoff, 2005, 2008; Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998; Morales, Mundy, & Rojas, 1998; Tomasello & Farrar, 1986;

Tomasello & Todd, 1983). However, despite considerable research and interest in these patterns of parent–infant joint looking, there are distinct emphases in different literature areas on the key components of parent–infant coordinated attention that support infant learning (Adamson, Bakeman, Suma, & Robins, 2019; Yu, Suanda, Smith, 2019). The current study examines a large variety of parent–infant coordinated attention bout types in face-to-face free play interactions and the association with future vocabulary size.

1.1 | The many views on coordinated attention

The traditional (and expansive) literature on coordinated attention focused on infants' social skills—how infants look to the parent face and then follow parent gaze to an object, as well as how infants initiate joint interest in an object by looking at the partner's face to check the partner's response (Baldwin, Moore, & Dunham, 1995; Brooks & Meltzoff, 2005, 2008; Bruner, 1983; Carpenter et al., 1998). Much (but not all) of the empirical evidence on these patterns of looking derive from experimental discrete trial studies (Brooks & Meltzoff, 2005, 2008). The focus on triadic attention—looks to the partner's face and to objects—derives, in part, from both mature communicative behaviors and theories of communication focused on the listener's goal of determining the referential intent of a speaker (Brooks & Meltzoff, 2005, 2008; Carpenter et al., 1998; Morales et al., 1998; Senju, Csibra, & Johnson, 2008). From this perspective, a sequential pattern of *infant* looking that includes both looks to the parent face and to the object is critical because the infant must monitor parent behavior to establish coordinated attention and to determine referential intent.

A second, and equally expansive, literature has focused on how well parents monitor an infant's behavior and interests, as well as how parental responsiveness supports coordinated attention and word learning (Tamis-LeMonda, Bornstein, & Baumwell, 2017). These findings suggest that word learning is enhanced when parents do not direct their infant's attention to an object when naming an object but instead uses their child's self-directed attention to an object as the signal as to when to name an object (Hirsh-Pasek et al., 2015; McCathren, Yoder, & Warren, 1995; Tamis-LeMonda, Bornstein, Kahana-Kalman, Baumwell, & Cyphers, 1998; Tomasello & Farrar, 1986; Harris, Jones, Brookes, & Grant, 1986; Gogate, Walker-Andrews, & Bahrick, 2001; Zukow-Goldring, 1990). From this perspective, a pattern of *parent* looking behavior that attends to and follows infant gaze, and then talks about the object, would be critical because the parent must monitor infant looking behavior so as to coordinate their attention with infant attention and to name objects at the right moment for infant learning.

A third and more recent literature derives from temporally precise measures of parent and infant gaze during social interactions and the *objective* definition of joint attention in terms of gaze directed to the same object at the same time (Suarez-Rivera, Smith, & Yu, 2019; Wass et al., 2018; Yu & Smith, 2016a, 2017). For example, Wass et al. (2018) compared gaze patterns during joint and solo play and observed that infants increased their attention to objects in joint play with a caregiver compared to solo play, suggesting the parent's behaviors help support their infant's sustained attention onto the objects. These studies showed that during these social interactions, infants rarely look to parent faces, although parents look to infant faces, and that key factors for infant word learning in these interactions are infant sustained attention to objects when parents name them (Pereira, Smith, & Yu, 2014; Yu, Suanda, & Smith, 2019). From this perspective, the pattern of *infant* looking to objects during naming events is the critical component because it is the infant who must link the name to the attended object.

Finally, there is a fourth pattern of joint looking behavior that has been of considerable interest to developmental researchers: mutual gaze (Blass, Lumeng, & Patil, 2017; Hoehl, Michel, Reid, Parise, & Striano, 2014; , Franchak, Kretch, & Adolph, 2018; Klinnert, Campos, Sorce, Emde, & Svejda,

1983; Leong et al., 2017; Niedźwiecka, Ramotowska, & Tomalski, 2018; Piazza, Hasenfratz, Hasson, & Lew-Williams, 2018; Striano, Reid, & Hoehl, 2006; Yamamoto, Sato, & Itakura, 2019). Although not directly linked to parent naming of objects, gaze patterns in which parent and infant jointly look to each other's faces are critical components of parent–infant interactions in early development (e.g., Lavelli & Fogel, 2005; Tronick & Cohn, 1989). Mutual gaze is central to shared emotional affect and affiliation (Feldman, 2007), plays a role in emotional regulation (Klinnert et al., 1983; MacLean et al., 2014) and attentional control (Niedźwiecka et al., 2018), and is often a marker of communicative intent (Senju & Csibra, 2008).

All these looking behaviors during parent–infant social interactions may be important components of parent–infant interactions and contribute to the quality of the interactions and infant learning from those interactions. However, these joint looking behaviors have been studied in isolation because different researchers focused on specific types of coordinated attention. As a result, we know very little about how these different components of joint looking naturally occur in free-flowing parent–infant interactions. Therefore, we created a free-flowing interaction context where infants and their caregivers sat at a table and played with a set of toys. In doing so, this particular interaction context afforded the infants and caregivers to have visual access to toys on the table and each other's faces, allowing for natural distributions of visual gaze to emerge. Note that recent research has also looked at gaze patterns of infants and their caregivers in other diverse interactional contexts (Franchak, Kretch, & Adolph, 2018; Yamamoto et al., 2019), and therefore, gaze patterns might differ as a function of interaction context. Finally, we also do not know whether individual differences in all or some of these gaze patterns are related to individual differences in early word learning. Accordingly, the main goals of the present study were to (a) determine the relative frequencies of different types of coordinated attention that are made up from different compositions of gaze patterns during a tabletop free-flowing interaction and (b) determine the predictive relation between those types of coordinated attention and infant vocabulary size.

To answer these questions, we used head-mounted eye trackers to simultaneously measure the gaze of parents and their 9-month-old infants as they jointly played with toys. Nine-month-olds were chosen because this is an age at which joint attention first emerges and in which large and well-documented individual differences that predict *later* vocabulary are reported (Adamson & Bakeman, 1991; Carpenter et al., 1998; Yu, Suanda, Smith, 2019). From the two streams of parent and infant gaze, we objectively defined seven mutually exclusive and comprehensive joint gaze types that encompass four perspectives discussed above and linked those patterns to later language learning outcomes.

Previous work has chosen specific terms when operationalizing particular types of coordinated attention. For example, *coordinated joint engagement* (Bakeman & Adamson, 1984) referred to when an infant coordinated with their attention with another person and an object for a particular duration (~3 s). In another example, *supported joint engagement* includes an asymmetrical interaction like when a parent produces overt behaviors like demonstrating an action that then supports coordinated attention (Adamson, Bakeman, & Deckner, 2004; Bakeman & Adamson, 1984; Kasari, Gulsrud, Wong, Kwon, & Locke, 2010). For the current paper, we chose the term, "coordinated attention," with the goal to be inclusive to the various types of coordinated attention that have been used in previous literature.

2 | METHODS

2.1 | Participants

Twenty-six parent–infant dyads participated (15 female and 11 male). The mean age of infants when participating in the parent–infant interaction component of the study was 9.21 months (*SD* = 0.23;

min = 9 months, 15 days; max = 10 months, 1 day). One dyad was excluded due to having an outlier 2.5 *SD*s above the mean on one of the variables; therefore, the final sample included 25 parent–infant dyads. Eight additional infants began the study but refused to wear the measuring equipment. The gender combinations of the parent–infant dyads included two male–male dyads, one male–female dyads, 13 female–female dyads, and nine female–male dyads. The entire sample of infants was broadly representative of Monroe County, Indiana (84% European American, 5% African American, 5% Asian American, 2% Latino, 4% Other), and consisted of predominantly working- and middle-class families. Infants were recruited through birth records and community organizations (e.g., museums, children's outreach events) that serve a diverse population. Parent reports of vocabulary were collected 3 and 6 months after the laboratory visit when the infants were 12 months old ($M = 12$ months, 15 days; $SD = 6$ days; min = 12 months, 3 days; max = 12 months, 27 days) and 15 months old ($M = 15$ months, 15 days; $SD = 8$ days; min = 15 months, 1 day; max = 16 months, 3 days). We measured vocabulary at these two intervals to ensure the reliability of our vocabulary measures and robustness of our findings. 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 involving human subjects in this study were approved by the University IRB (#0906000439) at the Indiana University.

2.2 | Stimuli

Two sets of three toys (car, cup, and train; and duck, plane, and boat) were used during the parent–infant play session. Each toy in each set was a unique uniform color (red, blue, green). The toys were chosen because they had been used in previous pilot tests and were observed to promote engaging infant–parent play for this particular age group.

2.3 | Experimental setup

Parents and their infants sat across from each other at a table (61 cm × 91 cm × 64 cm). The infants sat in a custom high chair, and the parents sat on the floor. Both infants and parents wore head-mounted eye trackers (positive science, LLC). The head-mounted eye-tracking system includes two cameras: (a) an infrared camera that is placed just below and is pointed to the right eye records eye images, and (b) a scene camera that is placed low on the forehead and is pointed outwards captures the user's first-person view (90° visual field). Each eye-tracking system recorded egocentric-view video and gaze direction (*x*-, *y*-coordinates) in that view, sampled at 30 Hz. Another camera (30 Hz) was mounted above the table and provided a bird's eye view of the dyadic interaction (see Yu and Smith, 2013 for additional technical details).

2.4 | Procedure

Parents and infants were fitted with the eye-tracking gear (see Figure 1). Once the eye-tracking gear was securely affixed to the participants, a calibration phase was completed. To collect calibration points for each eye tracker, an experimenter directed the infant's attention toward a toy that was only used for calibration, while another experimenter recorded the moment the child attended to the location of the toy. This procedure was repeated 15 times with the calibration toy placed in various

locations on the tabletop. A similar procedure was used to calibrate the parent's eye tracker. The calibration procedure took approximately 5 min.

Once the calibration phase was complete, an experimenter placed one of the object sets on the table and the first play trial began. During object play, all three objects within a set were on the table and parents were instructed to engage with their infant as they naturally would. If an object was knocked off of the table, an experimenter seated in the corner would promptly place the object back onto the table. After approximately 90 s of play, an experimenter swapped out the objects with the second set of objects, and the second trial began. The ordering of object sets was counterbalanced across dyads. This procedure was repeated and dyads completed up to four trials for a total of 6 min of play. Not all dyads completed the full play session. Twenty-three dyads completed all four trials, and two dyads
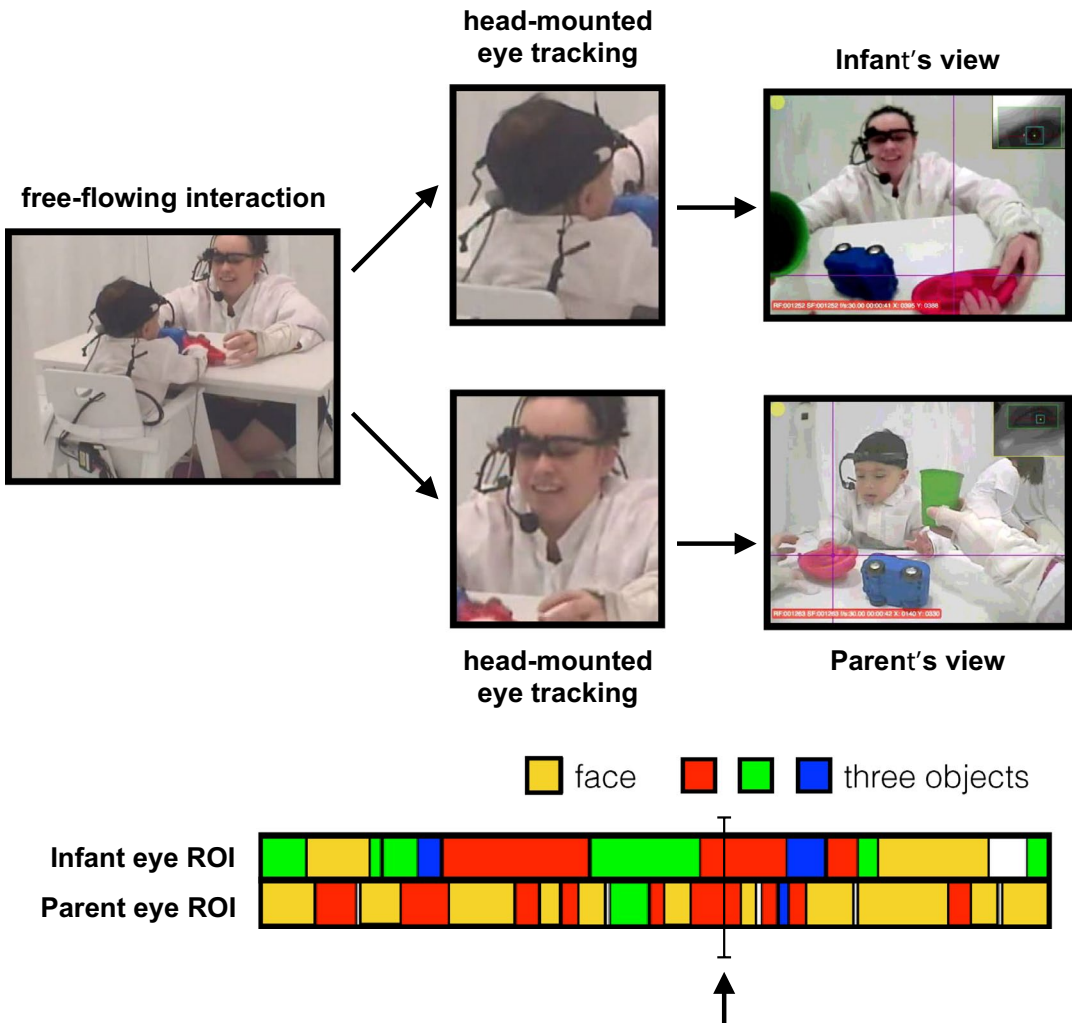

FIGURE 1 An infant and parent playing with a set of toys during a naturalistic free-flowing interaction. Both participants wore a head-mounted eye-tracking system that recorded eye gaze data from a first-person view. Crosshairs indicated frame-by-frame gaze direction. Two gaze streams from the parent and infant indicate region of interest fixations. The bracket corresponds to which ROIs the infant and parent were gazing toward: The infant's and parent's gaze were on the red object. White blocks indicate when an infant or parent is not looking at either an object or face

completed three trials, for a total average playtime of 5 min, 8 s. Parents were encouraged to play with their infants as they naturally would at home so behaviors were not restricted such as object naming, object touching, and object holding. Overall, natural toy play behaviors were observed such as the parents naming the objects and the infant and parent playing with objects together.

2.5 | Data processing

Eye-tracking software yielded scene camera footage with crosshairs superimposed, this footage was then sampled at a rate of 30 frames per second, using an in-house coding program, and trained coders blind to the hypotheses of the study annotated frame-by-frame the target of gaze. Calibration of gaze data was carefully processed and evaluated by trained experimenters. After processing of calibration was complete, each video was evaluated for accuracy and gaze data were recalibrated, if needed. The current study followed the best practices for calibration of head-mounted eye-tracking data communicated by Slone et al. (2018). For frames from the infant perspective, roughly 25% of frames were not codable either due to eye-tracking failure or the infant was off task and looking elsewhere than the regions of interest. Overall, over 10,000 frames were coded for each infant–parent play session.

Four regions of interest (ROIs) were defined: three toy objects and the partner's face. ROIs were manually coded frame-by-frame from a first-person view video. Coders independently coded either the infant's ROIs or the caregiver's ROIs. An ROI was annotated when a crosshair overlapped on any portion of an object or face (see Figure 1). If one ROI (e.g., blue object) went in front of another ROI (e.g., face), coders used the context before/after the instance to determine the proper ROI. To assess reliability, a second coder coded a randomly selected 10% of the frames with 95% agreement.

2.6 | Vocabulary size

Parents completed the infant version of the MacArthur-Bates Communicative Development Inventory (Fenson et al., 1994) when the infants were 12 and 15 months old. We used total receptive vocabulary as our measure of vocabulary scores at each age.

2.7 | Parent naming events

Speech from the parent was transcribed into spoken utterances. Among utterances that included, the names of the toys were designated at naming events.

2.8 | Gaze measures

The analyzed data begin as gaze streams to the partner's face or to one of the three objects in play. We operationally defined gaze to an object as coordinated attention behavior by requiring, first, at least 500 ms of coordinated attention to the same object at the beginning of the bout (see Yu & Smith, 2013) while also allowing the inclusion of looks to the partner's face for up to five additional seconds. Moreover, we also allowed a maximum gap of 1 s between two identical ROIs when no other ROIs
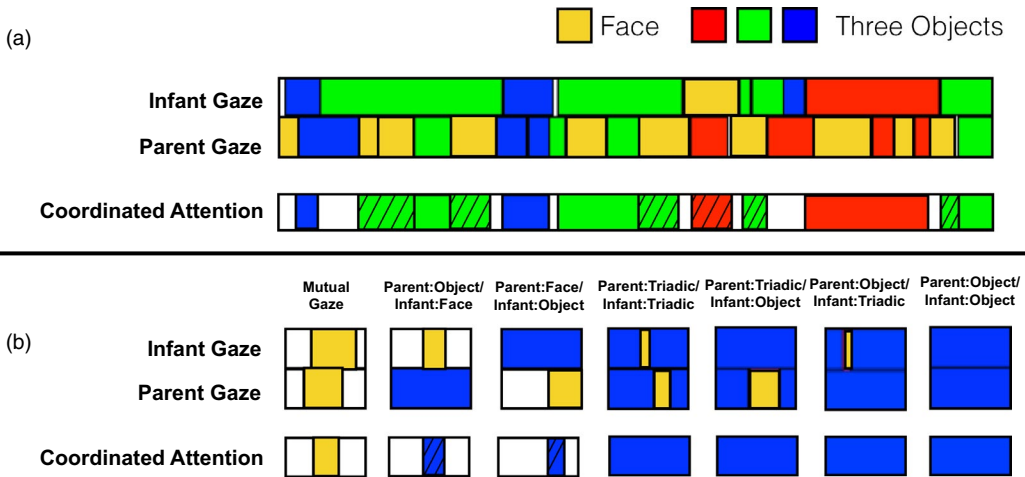
FIGURE 2 Overview of coordinated attention bout types. (a) An example of coupled ROI streams from one infant (first row) and parent (second row), and bouts of coordinated attention (third row). Solid coordinated attention bouts depict "object-oriented coordinated attention," and coordinated bouts with diagonal lines depict "face-oriented attention." (b) Examples of seven types of coordinated attention (from left to right): Mutual Gaze, Parent:Object/Infant:Face, Parent:Face/Infant:Object, Parent:Triadic/Infant:Triadic, Parent:Triadic/Infant:Object, Parent:Object/Infant:Triadic, and Parent:Object/Infant:Object. White space reflects gaze to object ROIs other than the partner's face or target ROI

were coded. These restrictions were to ensure temporal continuity between two looks to the same object while allowing for brief looks to the face of the partner and/or brief instances when no ROI was coded. To be inclusive of all types of coordinated attention, the definition required that both participants looked at the same object but could also include short or long looks by either participant to the face of the partner as shown in Figure 2. Table 1 provides an operational definition of the seven coordinated attention types:

TABLE 1 Operational definitions of the seven coordinated attention bout types

Coordinated attention types	Definition
Mutual Gaze	Simultaneous fixations of infant and parent face ROIs.
Parent:Face/Infant:Object	Simultaneous fixations of parent looking at infant's face and infant looking at any object ROI.
Parent:Object/Infant:Face	Simultaneous fixations of infant looking at parent's face and parent looking at any object ROI.
Parent:Triadic/Infant:Triadic	*Only* continuous alignment toward the same object ROI that also included looks to the face ROI from both parent and infant for less than 5 s.
Parent:Triadic/Infant:Object	*Only* continuous alignment toward the same object ROI that also included looks to the face ROI from parent for less than 5 s.
Parent:Object/Infant:Triadic	*Only* continuous alignment toward the same object ROI that also included looks to the face ROI from infant for less than 5 s.
Parent:Object/Infant:Object	*Only* continuous alignment toward the same object ROI that did not include any face looks from either parent or infant.

3 | RESULTS

3.1 | How frequently do different types of coordinated attention occur?

The seven coordinated attention bout types accounted for almost 50% of total interaction time ($M = 48\%$, $SD = 11\%$), clearly showing a strong role for looks to objects and to faces in parent–infant interactions. Table 2 provides the amount of total play that the dyads were in different types of coordinated attention as well as the relative proportions of coordinated attention types. We focus primarily on the amount of time spent in each coordinated attention bout type, and therefore, the following analyses incorporate the proportion and relative proportion properties. More specifically, proportions were computed as the total amount of time in each type of coordinated attention bout compared to the total trial time for each dyad and relative proportions were computed as the total amount of time in each type of coordinated attention bout compared to the overall amount of time in any type of coordinated attention bout. We did not have specific hypotheses for average duration and rate of each bout type. Two components of coordinated attention account for a majority of the total amount of time spent in coordinated attention: Parent:Face/Infant:Object (relative proportion: $M = 31\%$, $SD = 16\%$) and Parent:Triadic/Infant:Object (relative proportion: $M = 31\%$, $SD = 14\%$). The Parent:Face/Infant:Object type was principally made up of cases in which the infant looked at the object and the parent looked at the infant's face (without looking at the object). The Parent:Triadic/Infant:Object type was principally made up by cases in which the infant looked only to the object

TABLE 2 Properties of attention and coordinated attention types

Attention types	Proportion	Relative proportion	Duration (s)	Rate (per minute)
Object				
Parent	0.33 (0.12)		0.62 (0.17)	31.81 (9.06)
Infant	0.86 (0.06)		1.87 (0.58)	29.61 (7.29)
Face				
Parent	0.54 (0.14)		0.82 (0.28)	41.50 (11.01)
Infant	0.08 (0.05)		1.25 (0.59)	4.35 (3.22)
Coordinated attention				
Overall	0.48 (0.11)	1.00	1.23 (0.31)	30.53 (6.75)
Mutual Gaze	0.06 (0.05)	0.13 (0.11)	0.92 (0.29)	3.80 (2.72)
Parent:Object/Infant:Face	0.03 (0.02)	0.07 (0.06)	0.45 (0.19)	3.52 (2.85)
Parent:Face/Infant:Object	0.15 (0.07)	0.31 (0.16)	0.60 (0.22)	15.37 (5.42)
Parent:Triadic/ Infant:Triadic	0.05 (0.04)	0.10 (0.08)	2.47 (0.81)	2.60 (1.53)
Parent:Triadic/ Infant:Object	0.15 (0.08)	0.31 (0.14)	2.45 (0.87)	2.50 (1.53)
Parent:Object/ Infant:Triadic	0.005 (0.01)	0.01 (0.03)	0.56 (1.04)	0.16 (0.29)
Parent:Object/ Infant:Object	0.03 (0.03)	0.07 (0.08)	1.15 (0.28)	2.58 (1.69)

Note: Means and standard deviations (in parentheses). Proportion measures the overall proportion of time in bout. Duration is the average duration of bouts. Rate measures the number of bouts per minute.

and the parent looked at the infant's face and the object. Also noteworthy is that infants rarely looked to their parent's face either when their parent was looking at an object (Parent:Object/Infant:Face: $M = 3\%$, $SD = 2\%$) or when they were jointly looking at the same object as their parent (Parent:Object/Infant:Triadic: $M = 0.005\%$, $SD = 0.01\%$). Moreover, only 12 dyads displayed any amount of Parent:Object/Infant:Triadic coordinated attention.

3.2 | For the frequently occurring types of coordinated attention, which types were associated with future vocabulary size?

Vocabulary size at 12 months of age (min = 10, max = 185, mean = 83, median = 82, $SD = 50$) was positively correlated with vocabulary size at 15 months of age (min = 18, max = 206, mean = 134, median = 142, $SD = 54$), $r(24) = .83$, $p < .001$. Median vocabulary size at 12 months of age (median = 82) was between the 50% and 60% percentiles of the full American English Words and Gestures Form (Understanding) dataset from Wordbank (Frank, Braginsky, Yurovsky, & Marchman, 2016). Median vocabulary size at 15 months of age (median = 142) was between the 40% and 50% percentiles. Data were downloaded on 4/21/20.

For looking bouts (independent of coordinated attention), infants' proportion of looking at objects positively correlated with vocabulary size at 12 months ($r = .59$, $p = .002$) and 15 months ($r = .61$, $p = .001$), whereas proportion of infant looking at faces was negatively correlated with vocabulary size at 12 months ($r = -.56$, $p = .003$) and 15 months ($r = -.56$, $p = .003$).

The proportion of overall coordinated attention correlated with vocabulary size at 12 months ($r = .46$, $p = .019$) but not at 15 months ($r = .29$, $p = .153$). The two most frequent types of coordinated attention, accounting for 62% of total time in overall coordinated attention, were (a) Parent:Face/Infant:Object and (b) Parent:Triadic/Infant:Object. Table 3 reports the zero-order correlations between coordinated attention proportions and future vocabulary size. Despite the high proportion of total coordinated attention across the two frequent types, only the proportion of time in Parent:Face/Infant:Object correlated with vocabulary size at 12 months ($r = .53$, $p = .006$) and 15 months ($r = .44$, $p = .049$) (see Figure 3). All data and code are available here: https://osf.io/6kc5d/?view_only=a7aa63ce44704d458fda70e37d7a597f.

TABLE 3 Zero-order correlations between coordinated attention proportions and vocabulary size at 12 and 15 months

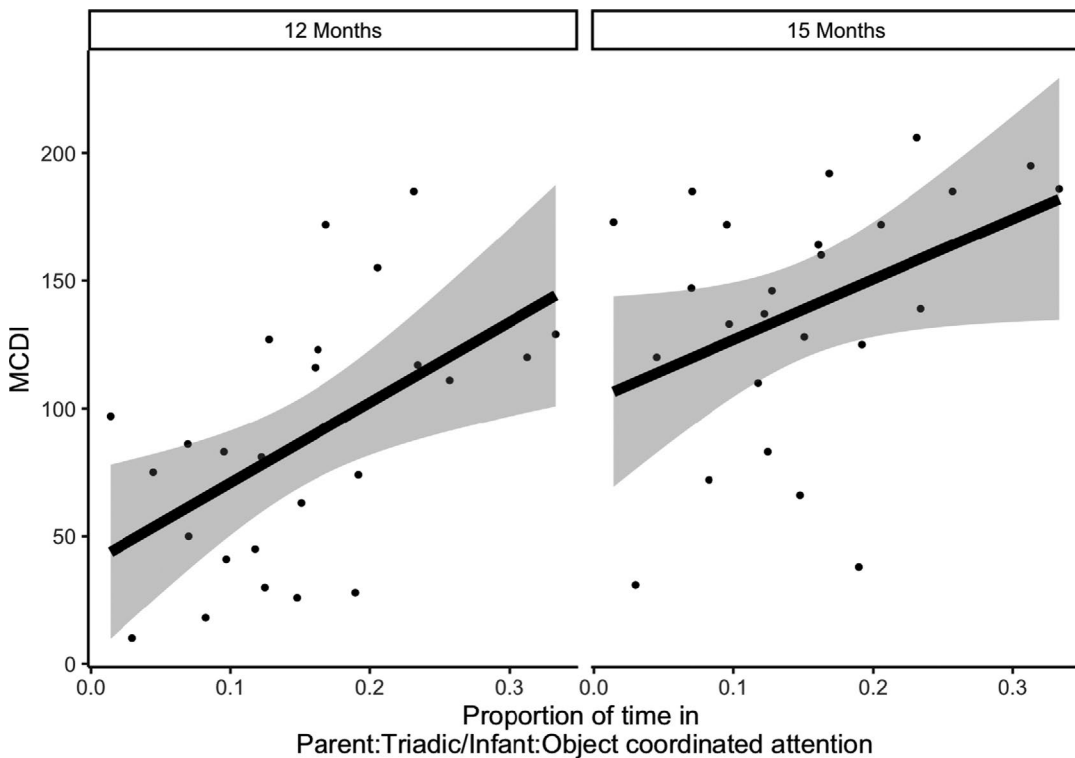| Type	12 months	15 months
Overall	$r = .46$, $p = .019$	$r = .29$, $p = .153$
Mutual Gaze	$r = -.16$, $p = .437$	$r = -.21$, $p = .314$
Parent:Object/Infant:Face	$r = -.33$, $p = .104$	$r = -.24$, $p = .243$
Parent:Face/Infant:Object	$r = .32$, $p = .118$	$r = .22$, $p = .281$
Parent:Triadic/Infant:Triadic	$r = -.006$, $p = .974$	$r = -.09$, $p = .664$
Parent:Triadic/Infant:Object	$r = .53$, $p = .006$	$r = .44$, $p = .049$
Parent:Object/Infant:Triadic	$r = -.02$, $p = .915$	$r = .15$, $p = .458$
Parent:Object/Infant:Object	$r = .07$, $p = .744$	$r = .08$, $p = .707$

FIGURE 3 Scatter plots (with the best-fitting regression lines) show correlations between proportion of time in Parent:Triadic/Infant:Object coordinated attention and MCDI at 12 months (left) and 15 months (right). Ribbons reflect ± 95% CIs

3.3 | For the frequently occurring types of coordinated attention, which types were associated with more in-the-moment object naming by the parent?

To aid in understanding what other potentially important behaviors might occur during different types of coordinated attention, we conducted a post hoc, exploratory analysis to determine the amount of object naming a parent produces during coordinated attention. Previous research suggested that a key factor of infant word learning is that when an infant sustains attention on an object, a parent names the object (Pereira et al., 2014; Yu, Suanda, & Smith, 2019). Of the two most frequently occurring coordinated attention types, only Parent:Triadic/Infant:Object coordinated attention was associated with future vocabulary size. To determine why this type of coordinated attention, but not the other frequently occurring coordinated attention type (Parent:Face/Infant:Object), might be more associated with future vocabulary size, we calculated the proportion of coordinated attention bouts that included parent naming of the object jointly attended to by the infant and parent (target). We also calculated the proportion of coordinated attention bouts that included parent naming of an object that was not jointly attended to by the infant and parent (nontarget). We chose to use the proportion of coordinated attention bouts that included naming instances because such a measure controls for differences in other properties of coordinated attention that vary across the two frequently occurring coordinated attention bout types, namely average bout duration and bout frequency.

Overall, parents named objects approximately eight times per minute ($M = 8.23$ per/minute, $SD = 2.83$). The proportion of coordinated attention bouts that included instances of parent naming

TABLE 4 Proportion of object attention and coordinated attention bouts that include an instance of parent naming for target and nontarget naming

Type	Target	Nontarget
Mutual Gaze	0.41 (0.15)	0.07 (0.09)
Parent:Object/Infant:Face	0.48 (0.14)	0.09 (0.08)
Parent:Face/Infant:Object	0.38 (0.11)	0.15 (0.10)
Parent:Triadic/Infant:Triadic	0.06 (0.22)	0.0 (0.0)
Parent:Triadic/Infant:Object	0.50 (0.18)	0.07 (0.08)
Parent:Object/Infant:Triadic	0.16 (0.37)	0.0 (0.0)
Parent:Object/Infant:Object	0.30 (0.27)	0.07 (0.08)

was higher for Parent:Triadic/Infant:Object coordinated attention bouts ($M = 50\%$, $SD = 18\%$) compared to Parent:Face/Infant:Object coordinated attention bouts ($M = 38\%$, $SD = 15\%$), $t(24) = 3.26$, $p = .003$, $d = 0.80$. Overall, Parent:Triadic/Infant:Object coordinated attention bouts had higher instances of parent naming of the target object, further leading to the suggestion that the gaze behaviors that make up Parent:Triadic/Infant:Object coordinated attention are important pathways for the relationship between coordinated attention and future vocabulary size (see Table 4).

4 | GENERAL DISCUSSION

Coordinated attention represents an important but complex behavior for an infant developing in a social world. By definition, coordinated attention necessarily includes gaze patterns from an infant and a parent. The present study provides three new contributions for understanding the nature of early coordinated attention and its role in language development. First, from all of the potential and theorized combinations of components that make up various types of coordinated attention, only two account for a majority of the time infants and their parents spend in coordinated attention: when parents are looking to their infant's face while their infant was looking at an object (Parent:Face/Infant:Object) and when parents are looking to their infant's face *during* object-oriented coordinated attention (Parent:Triadic/Infant:Object). Second, of the coordinated attention bout types that frequently occurred, only the variability of time spent in coordinated attention that includes when parents look to their infant's face *during* object-oriented coordinated attention (Parent:Triadic/Infant:Object) is correlated with future vocabulary size. Third, parents name the object their infant is looking at more often during bouts of coordinated attention that includes when parents look to their infant's face *during* object-oriented coordinated attention (Parent:Triadic/Infant:Object).

Many theoretical discussions of coordinated attention (Baldwin et al., 1995; Bruner, 1983; Carpenter et al., 1998) have centered on the infant's ability to follow parent eye gaze and to infer parent momentary interests and thus referential intent. Several empirical investigations further demonstrate that the ability of parents and infants to coordinate attention in social interactions has also been shown to be highly predictive of later vocabulary development (Bakeman & Adamson, 1994; Brooks & Meltzoff, 2005, 2008; Carpenter et al., 1998; Morales et al., 1998; Tomasello & Farrar, 1986; Tomasello & Todd, 1983). Putting these two together would seem to suggest that one route to fostering vocabulary growth in children with small vocabularies would be to focus on the infant's deficiencies of reading parents' interests. The present results suggest an alternative route. The present results show that two pieces of coordinated attention critical for language development are as follows:

the prevalence of simultaneous parent monitoring of infant gaze and infant attention to objects during bouts of object-oriented coordinated attention. The significance of this finding derives in part from the well-documented link between individual differences in coordinated attention and individual differences in concurrent and future vocabulary size (e.g., Carpenter et al., 1998; Mundy, Sigman, & Karsari, 1990; Tomasello & Todd, 1983) and from the well-documented links between individual differences in vocabulary and cognitive and linguistic skills (Marchman & Fernald, 2008) and later school achievement (Murphy, Rowe, Ramani, & Silverman, 2014). It is therefore critical to understand the environments and situations where vocabulary learning does and does not occur.

If the reason for why coordinated attention correlates with future vocabulary size is because coordinated attention taps into an infant's emerging ability to infer referential intent, we should expect to observe a higher frequency of infants looking to their parent's face (e.g., Parent:Object/Infant:Face, Parent:Triadic/Infant:Triadic, or Parent:Object/Infant:Face). A stronger test of this conclusion would be in the form of adding additional variants of coordinated attention that include when infants and parents lead and/or follow into a bout of coordinated attention (Mundy & Newell, 2007). Nevertheless, the results observed in the current study, including that infants primarily fixate their gaze on objects and not their parent's face, point to a rethinking of the infant's role in coordinated attention for language development at different phases of development.

Consistent with other previous research that spans across different parent–infant interaction contexts (e.g., Carpenter et al., 1998; de Barbaro, Johnson, Forster, & Deak, 2015; Deák, Krasno, Triesch, Lewis, & Sepeta, 2014; Yu & Smith, 2013, 2016a, 2016b), in the current study, we observed that infants rarely looked to their parent's face. This is an important observation because previous research has observed that attentional skills like an infant's ability to follow the gaze of an adult during a discrete trial are predictive of later vocabulary (Brooks & Melzoff, 2005, 2008; Morales et al., 1998). The reason for this disparity is an open question, and future research should focus on better understanding the similarities and differences between experimental contexts such as free play and discrete trials during coordinated attention (e.g., Tamis-LeMonda, Kuchirko, Luo, Escobar, & Bornstein, 2017).

The finding that a specific type of coordinated attention made of two component behaviors—infant attention to objects while parent is attending to the same object and also monitoring the attention of the infant—correlates with later vocabulary size is consistent with the literature showing that parent responsiveness and infant sustained attention are important for concurrent and future language development. Recent research found that infant sustained attention during bouts of joint attention strongly predicted future vocabulary size (Yu, Suanda, & Smith, 2019). Parental monitoring and subsequent following behaviors have been shown to lead to enhanced word learning (Tamis-LeMonda & Bornstein, 1994; Tamis-LeMonda, Kuchirko, & Song, 2014; Tomasello & Farrar, 1986). One explanation for this is that responsive parents seize opportunistic moments wherein infants attend to a target object by providing verbal referents and accompanying deictic gestures to objects in their infant's visual attention. For infants to link visual objects and their referents, what is important is that their visual attention on the target object is temporally aligned with parent naming (Gogate et al., 2001; Yu & Smith, 2012; Yu, Suanda, & Smith, 2019; Zukow-Goldring, 1990) and results from our exploratory analyses add further elaboration suggesting that parents name the object their infant is looking at more often during bouts of coordinated attention that includes when parents look to their infant's face *during* object-oriented coordinated attention (Parent:Triadic/Infant:Object).

The observation that parental monitoring and infant object attention during coordinated attention are two critical components that drive the relationship between coordinated attention and future vocabulary size at this particular developmental period provides a starting point for a unifying explanation of the roles of the parent and the infant for the creation of important word learning moments. It was entirely possible that the variability of other types of coordinated attention bouts would

be correlated with increases in future vocabulary size, in addition to the coordinated attention bout where infant attention to objects while parent is attending to the same object and also monitoring the attention of the infant. However, a critical observation is that parental monitoring of infant attention *outside* of object-oriented coordinated attention bouts occurred frequently (Parent:Face/Infant:Object: 31% of all coordinated attention) but did not correlate with future vocabulary size. The key feature that bridges the parent's role (attentional monitoring) and the infant's role (attention to object) is the emergent object-oriented coordinated attention of the dyad.

A few limitations are noteworthy. Our decision to include face looks instead of looks to the partner's eyes constrains our interpretation of coordinated looking. Specifically, an interpretation of mutual gaze should include looks to any part of the partner's face. Moreover, the interactive context structured in the laboratory was not the natural setting in which infants and their parents are typically interacting within. Ongoing efforts in our laboratory are moving toward more naturalistic, laboratory-based interactions that simulate home-like contexts and minimize possible impacts of data collection methods such as the use of eye-tracking equipment. Future research should determine whether the patterns of results documented in the current study generalize to more naturalistic interactions. Moreover, although our analytic protocol took advantage of high-density temporal streams of behavior, the sample size of the current study was limited and future research in this domain should work to achieve larger samples when possible. Finally, given the multimodal nature of the interactive context, future work should investigate the role of other behaviors that might play a role in coordinated attention above and beyond the gaze patterns of objects and partners' faces, such as holding behaviors, object passing, and gaze behaviors independent of coordinated attention. Given the experimental design of the current study, namely that it consisted of joint play, it is difficult to disentangle gaze patterns that would be completely independent of a social partner's gaze patterns even during bouts of infant gaze that did not include parent looking at any of the ROIs. Future research could experimentally control for effects of coordinated attention on independent gaze patterns like infant sustained attention by implementing an experimental design similar to Wass et al. (2018) which included joint play and solo play sessions for each dyad.

Research spanning decades has provided important insights into the patterns of coordinated attention and what those patterns can predict in later development. Previous work has carefully chosen specific terminology when operationalizing particular types of coordinated attention. For example, Bakeman and Adamson (1984) coded behavior as *coordinated joint engagement* if an infant coordinated their attention with another person and an object for particular duration (~3 s) with brief attentional shifts elsewhere for less than 3 s. In another example, the coding of *supported joint engagement* includes an asymmetrical interaction such that the parent is producing overt behaviors that support the coordinated attention such as demonstrating an action with the object (Adamson et al., 2004; Bakeman & Adamson, 1984; Kasari et al., 2010). Our decision to choose the term "coordinated attention" was motivated by the goal to be inclusive to various types of coordinated attention that have been used in previous literature. The observation that one particular type of coordinated attention, Parent:Triadic/Infant:Object, correlated with later vocabulary development, is consistent with Bakeman and Adamson (1984)'s description of supported joint engagement.

Past research suggests that it is very likely that the role of the infant changes such that, at 9 months of age, there are very few fixations to the parent's face, but going into the second year of life, infants increase the frequency of alternating their gaze from their parents' face and objects in the environment (Bakeman & Adamson, 1984; Carpenter et al., 1998): a behavior that has been found to be an important predictor of language outcomes later in development (Carpenter et al., 1998; Mundy et al., 2007). Moreover, infant's sustained attention to objects increases in frequency and duration throughout the first and second year of life (Ruff & Lawson, 1990), which adds even more nuance to the developing

role of the infant during complex interaction with parents. It is possible that increases in both infants' alternating gaze between their parents' faces and objects in addition to increased and prolonged bouts of sustained attention provide the context where more complex social interactions lead to diverse language learning moments. Therefore, it is important to consider the results from the current study in the context of one particular age range during development and to motivate future investigations into the developmental trajectories of the patterns of gaze across infants and parents that are the building blocks of coordinated attention.

In summary, previous research studying social interactions between parents and infants has suggested various patterns of gaze behaviors. The current study highlighted the distribution of components that make up different types of coordinated attention that actually occur during parent–infant interaction and which types of coordinated attention are associated with future vocabulary size. Critically, the current study provided results consistent with a new explanation for this period of infancy: Coordinated attention is comprised of multiple component behaviors and the key components that are associated with future vocabulary size are momentary parental monitoring of infant visual attention and infant attention to the target object during object-oriented coordinated attention. These results point to the critical roles of parent and infant that are afforded by emergent temporary bouts of coordinated attention and unify two important roles of the parent–infant dyad known to be associated with later vocabulary size: parent responsiveness and infant sustained attention.

ORCID

Drew H. Abney https://orcid.org/0000-0003-3040-2927
Sumarga H. Suanda https://orcid.org/0000-0001-8773-5218

REFERENCES

Adamson, L. B., & Bakeman, R. (1991). The development of shared attention during infancy. In R. Vasta (Ed.), *Annals of child development* (Vol. *8*, pp. 1–41). London, UK: Jessica Kingsley Press.

Adamson, L. B., Bakeman, R., & Deckner, D. F. (2004). The development of symbol-infused joint engagement. *Child Development*, *75*(4), 1171–1187.

Adamson, L. B., Bakeman, R., Suma, K., & Robins, D. L. (2019). An expanded view of joint attention: Skill, engagement, and language in typical development and autism. *Child Development*, *90*(1), e1–e18.

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development*, 1278–1289.

Baldwin, D. A., Moore, C., & Dunham, P. J. (1995). Understanding the link between joint attention and language. In C. Moore, & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 131–158). Hillsdale, NJ: Lawrence Erlbaum.

Blass, E. M., Lumeng, J., & Patil, N. (2017). Influence of mutual gaze on human infant affect. In R. Flom, K. Lee, & D. Muir (Eds.), *Gaze-following: Its development and significance* (pp. 113–141). Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, *8*(6), 535–543.

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*(01), 207–220.

Bruner, J. (1983). *Child's talk: Learning to use language*. New York, NY: Norton.

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*, i–174.

de Barbaro, K., Johnson, C. M., Forster, D., & Deák, G. O. (2015). Sensorimotor decoupling contributes to triadic attention: A longitudinal investigation of mother– infant–object interactions. *Child Development*, *87*, 494–512.

Deak, G. O., Krasno, A. M., Triesch, J., Lewis, J., & Sepeta, L. (2014). Watch the hands: Infants can learn to follow gaze by seeing adults manipulate objects. *Developmental Science*, *17*(2), 270–281.

Feldman, R. (2007). On the origins of background emotions: From affect synchrony to symbolic expression. *Emotion*, *7*(3), 601.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*, 1–173.

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.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of child language*, *44*(3), 677.

Gogate, L. J., Walker-Andrews, A. S., & Bahrick, L. E. (2001). The intersensory origins of word-comprehension: An ecological–dynamic systems view. *Developmental Science*, *4*(1), 1–18.

Harris, M., Jones, D., Brookes, S., & Grant, J. (1986). Relations between the non-verbal context of maternal speech and rate of language development. *British journal of developmental psychology*, *4*(3), 261–268.

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., … Suma, K. (2015). The contribution of early communication quality to low-income children's language success. *Psychological Science*, *26*(7), 1071–1083.

Hoehl, S., Michel, C., Reid, V. M., Parise, E., & Striano, T. (2014). Eye contact during live social interaction modulates infants' oscillatory brain activity. *Social Neuroscience*, *9*(3), 300–308.

Kasari, C., Gulsrud, A. C., Wong, C., Kwon, S., & Locke, J. (2010). Randomized controlled caregiver mediated joint engagement intervention for toddlers with autism. *Journal of Autism and Developmental Disorders*, *40*(9), 1045–1056.

Klinnert, M. D., Campos, J. J., Sorce, J. F., Emde, R. N., & Svejda, M. A. R. I. L. Y. N. (1983). Emotions as behavior regulators: Social referencing in infancy. In R. Plutchik, & H. Kellerman (Eds.), *Emotions in early development (Vol. 2): The emotions* (pp. 57–86). New York, NY: Academic Press.

Lavelli, M., & Fogel, A. (2005). Developmental changes in the relationship between the infant's attention and emotion during early face-to-face communication: The 2-month transition. *Developmental Psychology*, *41*(1), 265–280.

Leong, V., Byrne, E., Clackson, K., Georgieva, S., Lam, S., & Wass, S. (2017). Speaker gaze increases information coupling between infant and adult brains. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(50), 13290–13295.

MacLean, P. C., Rynes, K. N., Aragón, C., Caprihan, A., Phillips, J. P., & Lowe, J. R. (2014). Mother–infant mutual eye gaze supports emotion regulation in infancy during the still-face paradigm. *Infant Behavior and Development*, *37*(4), 512–522.

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, *11*(3), F9–F16.

McCathren, R. B., Yoder, P. J., & Warren, S. F. (1995). The role of directives in early language intervention. *Journal of Early Intervention*, *19*(2), 91–101.

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.

Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A. V., & Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy. *Child Development*, *78*(3), 938–954.

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current Directions in Psychological Science*, *16*(5), 269–274.

Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders*, *20*(1), 115–128.

Murphy, P. K., Rowe, M. L., Ramani, G., & Silverman, R. (2014). Promoting critical- analytic thinking in children and adolescents at home and in school. *Educational Psychology Review*, *26*(4), 561–578.

Niedźwiecka, A., Ramotowska, S., & Tomalski, P. (2018). Mutual gaze during early mother–infant interactions promotes attention control development. *Child Development*, *89*(6), 2230–2244.

Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin and Review*, *21*(1), 178–185.

Piazza, E. A., Hasenfratz, L., Hasson, U., & Lew-Williams, C. (2018). Infant and adult brains are coupled to the dynamics of natural communication. *bioRxiv*, 359810.

Ruff, H. A., & Lawson, K. R. (1990). Development of sustained, focused attention in young children during free play. *Developmental Psychology*, *26*(1), 85–93.

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, *18*(9), 668–671.

Senju, A., Csibra, G., & Johnson, M. H. (2008). Understanding the referential nature of looking: Infants' preference for object-directed gaze. *Cognition*, *108*(2), 303–319.

Slone, L. K., Abney, D. H., Borjon, J. I., Chen, C.-H., Franchak, J. M., Pearcy, D., … Yu, C. (2018). Gaze in action: Head-mounted eye tracking of children's dynamic visual attention during naturalistic behavior. *JoVE (Journal of Visualized Experiments)*, *141*, e58496.

Striano, T., Reid, V. M., & Hoehl, S. (2006). Neural mechanisms of joint attention in infancy. *European Journal of Neuroscience*, *23*(10), 2819–2823.

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.

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.

Tamis-LeMonda, C. S., & Bornstein, M. H. (1994). Specificity in mother-toddler language-play relations across the second year. *Developmental Psychology*, *30*(2), 283.

Tamis-Lemonda, C. S., Bornstein, M. H., Kahana-Kalman, R., Baumwell, L., & Cyphers, L. (1998). Predicting variation in the timing of language milestones in the second year: An events history approach. *Journal of Child Language*, *25*(3), 675–700.

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*, 1–14.

Tamis-LeMonda, C. S., Kuchirko, Y., & Song, L. (2014). Why is infant language learning facilitated by parental responsiveness? *Current Directions in Psychological Science*, *23*(2), 121–126.

Tamis-LeMonda, C. S., Kuchirko, Y., & Tafuro, L. (2013). From action to interaction: Infant object exploration and mothers' contingent responsiveness. *IEEE Transactions on Autonomous Mental Development*, *5*(3), 202–209.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, 1454–1463.

Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First Language*, *4*(12), 197–211.

Tronick, E. Z., & Cohn, J. F. (1989). Infant-mother face-to-face interaction: Age and gender differences in coordination and the occurrence of miscoordination. *Child Development*, *60*, 85–92.

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.

Yamamoto, H., Sato, A., & Itakura, S. (2019). Eye tracking in an everyday environment reveals the interpersonal distance that affords infant-parent gaze communication. *Scientific Reports*, *9*(1), 1–9.

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*(2), 244–262.

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.

Yu, C., & Smith, L. B. (2016a). Multiple sensory-motor pathways lead to coordinated visual attention. *Cognitive Science*, *41*, 5–31.

Yu, C., & Smith, L. B. (2016b). The social origins of sustained attention in one-year-old human infants. *Current Biology*, *26*(9), 1235–1240.

Yu, C., & Smith, L. B. (2017). Hand-eye coordination predicts joint attention. *Child Development*, *88*(6), 2060–2078.

Yu, C., Suanda, S. H., & Smith, L. B. (2019). Infant sustained attention but not joint attention to object at 9 months predicts vocabulary at 12 and 15 months. *Developmental Science*, *22*(1), e12735.

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