Parent–Child Joint Behaviors in Novel Object Play Create High-Quality Data for Word Learning

Chi-hsin Chen [ORCID]
The Ohio State University

Derek M. Houston
The Ohio State University and Nationwide Children's Hospital

Chen Yu
The University of Texas at Austin

This research takes a dyadic approach to study early word learning and focuses on toddlers' (N = 20, age: 17–23 months) *information seeking* and parents' *information providing* behaviors and the ways the two are coupled in real-time parent–child interactions. Using head-mounted eye tracking, this study provides the first detailed comparison of children's and their parents' behavioral and attentional patterns in two free-play contexts: one with novel objects with to-be-learned names (Learning condition) and the other with familiar objects with known names (Play condition). Children and parents in the Learning condition modified their individual and joint behaviors when encountering novel objects with to-be-learned names, which created clearer signals that reduced referential ambiguity and potentially facilitated word learning.

An intriguing question that has attracted much interest in developmental research is how infants and young children learn words (e.g., Baldwin & Markman, 1989; Gleitman, 1990; Gogate, Bolzani, & Betancourt, 2006; Katz, Baker, & Macnamara, 1974; Markman & Hutchinson, 1984; Smith & Yu, 2008; Tomasello & Todd, 1983; Waxman & Booth, 2001). Most researchers have been trying to answer this question from one of two different perspectives— one from the child's side and the other from the parent's side. Numerous studies in the past few decades have focused on young learners' capabilities of processing or using different types of available information—such as social cues (Baldwin, 1993; Tomasello & Akhtar, 1995), conceptual relations (Markman & Hutchinson, 1984), linguistic structures (Snedeker & Gleitman, 2004; Waxman & Booth, 2001), and statistical information (Smith &

Yu, 2008)—to learn the meanings of novel words. Another line of research focuses on the characteristics of word-learning opportunities by documenting precisely the amount and types of linguistic input provided by parents in early learning environments using extensive audio and video recordings (Bergelson, Amatuni, Dailey, Koorathota, & Tor, 2019; Hurtado, Marchman, & Fernald, 2008; Tamis-LeMonda, Custode, Kuchirko, Escobar, & Lo, 2019). These two lines of studies have yielded fruitful insights into children's and parents' *separate* contributions to early word learning. However, one critical aspect that is largely unknown is how children and parents *jointly* create the visual and linguistic data for word learning and how their *joint* behaviors contribute to early word learning.

In parent–child interaction, the data created for word learning are not just objective linguistic and visual properties of a learning environment, but instead are selected and filtered through children's own actions and sensory systems (Smith, Yu, & Pereira, 2011; Yu & Smith, 2012). Children do not just passively perceive all information in the environment. On the contrary, they explore and attend to different aspects of the environment depending

on their momentary goals and the environment itself (Bloom & Tinker, 2001; Kretch, Franchak, & Adolph, 2014). Furthermore, children's exploratory behaviors and attentional patterns are also accompanied by parents' speech and nonverbal actions, which also change from moment to moment, depending on the dynamics of the interaction (Chen, Castellanos, Yu, & Houston, 2019a; Suarez-Rivera, Smith, & Yu, 2019). Therefore, to have a complete picture of early word learning, it is necessary to go beyond studying children's and parents' individual behaviors separately, but to include the dyad's joint behaviors in real-time interactions (Renzi, Romberg, Bolger, & Newman, 2017; Yurovsky, 2018). In this study, we investigate how children's and parents' real-time multimodal behaviors individually and jointly shape the visual and linguistic input for word learning.

It has been shown that parents modify the way they talk depending on children's language knowledge; and the modification makes it easier for children to identify the referents of novel words and to learn novel word-object mappings (Yurovsky, 2018). Going beyond parent speech, the overarching hypothesis of this study is that both children as language learners and parents as language teachers play complementary roles in creating clear signals for word learning. To test this hypothesis, this study examines young toddlers' real-time exploratory behaviors to *seek information* and their parents' behaviors to *provide information* in free-flowing parent–child interaction. We ask whether certain information seeking and providing behaviors that have been shown to be beneficial for word learning are specific to contexts with learning opportunities or whether these behaviors are general and remain constant across contexts. To experimentally test these two possibilities, we compare children's and parents' individual and joint behaviors in a condition where they play with novel objects with to-be-learned names (subsequently termed Learning condition) and their behaviors in a control condition where they play with familiar objects with known names (subsequently termed Play condition). The direct comparisons of children's and parents' behaviors in these two conditions will shed lights on our understanding of the real-time multimodal behaviors that contribute to learning in free-flowing parent–child interaction.

Behaviors in Object Play That Contribute to Word Learning

Play is an important activity in infants' and young children's daily life and is a key context for their learning of language and various social and cognitive skills (Bornstein, Haynes, O'Reilly, & Painter, 1996; Power, 1985; Tamis-LeMonda et al., 2019; Williams, 2003; Yogman, 1981; Yogman et al., 2018). In this study, we investigate three parent–child interaction components that have been shown to support word learning: children's information seeking behaviors, parents' information providing behaviors, and their joint behaviors with a focus on the synchrony between children's information seeking and parents' information providing behaviors. We compare whether children and parents show behavioral differences in these three aspects in a Learning condition, when they play with novel objects with to-be-learned names, and in a Play condition, when they play with familiar objects with known names.

The first component that contributes to word learning is children's information seeking behaviors, such as their looking behaviors and manual actions during object play (Kannass & Oakes, 2008; Lawson & Ruff, 2004; Yu & Smith, 2012; Yu, Suanda, & Smith, 2019). Prior research has shown that young children increase their visual and exploratory behaviors (e.g., handling, manual manipulation, mouthing) when they encounter novel objects versus familiar objects (e.g., Ruff, 1984, 1986; Ruff, Saltarelli, Capozzoli, & Dubiner, 1992). Increased visual and exploratory behaviors have been suggested to be beneficial for learning about objects and their names. For example, recent modeling and experimental studies show that young infants' manual actions during play create high-quality visual data that facilitate visual object recognition, which is the foundation for the learning of object names (Bambach, Crandall, Smith, & Yu, 2016, 2018; Tsutsui, Chandrasekaran, Reza, Crandall, & Yu, 2020). Furthermore, it has been proposed that much of infants' information processing and learning takes place during sustained attention (Colombo, 2001; Frick & Richards, 2001; Richards, 1997). Infants' sustained attention during object play at 9 months of age has been found to be predictive of their later vocabulary development (Yu et al., 2019). These findings together suggest that more focused attention and more manual actions on toy objects create better data for object learning and word learning. Accordingly, our prediction is that children will show longer looking—including having more sustained attention—and more handling of objects in the Learning condition than the Play condition.

The second component contributing to word learning is parents' information providing behaviors. In this study, we specifically focus on parents'

naming of objects and whether they provide cues to highlight the named object. It has been shown that mothers' naming behaviors are affected by how familiar children are with the objects. When playing with toy animals with their 1-year-olds, mothers are more likely to provide names for animals their child is *familiar with* or comprehend than the names of novel animals (Masur, 1997). Such results seem to suggest that children would have less opportunity to hear an object name when the object is novel, compared to a familiar object. However, one important factor in word learning is how easy it is to identify the object being named (Golinkoff et al., 2000). When parents name a novel object during play, they tend to look at, point, touch, or move the object at the same time (Gogate, Bahrick, & Watson, 2000; Gogate, Maganti, & Bahrick, 2015; Gogate et al., 2006; Lund & Schuele, 2015; Masur, 1997; Suanda, Smith, & Yu, 2016; Yu & Smith, 2012). These multimodal behaviors potentially attract children's attention to the named object and facilitate the learning of object names. From these previous findings, we predict that parents may not name novel objects in the Learning condition as frequently as what they would label familiar objects in the Play condition. However, when they refer to a novel object in the Learning condition, they would be more likely to provide nonverbal cues, such as manual cues, along with their speech.

Third, in the early stage of word learning, hearing a novel object name when the child is paying attention to the named object is crucial for their learning of the word-object mapping (MacRoy-Higgins & Montemarano, 2016; Yu & Smith, 2012). The synchrony between parents' naming of an object and children's attention to the named object during toy play positively correlates with children's learning of object names and later vocabulary development (MacRoy-Higgins & Montemarano, 2016; Tomasello & Farrar, 1986; Yu & Smith, 2012; Yu et al., 2019). Previous studies in controlled laboratory settings suggest that children tend to spend more time looking at and examining novel objects (Oakes, Madole, & Cohen, 1991; Ross, 1980; Ruff, 1984). Therefore, one possibility is that more sustained attention children pay to novel objects in the Learning condition will create better synchrony between parents' naming and children's attention. Alternatively, another possibility is that children's familiarity with the objects being named in the Play condition will create better naming-attention synchrony in the Play condition. Another focus of the current project is the temporal relationship between parents' naming and children's attention, by which

we examine whether parents' naming of an object leads children's attention to the same object or whether children's attention to an object leads to parents' naming of the object. Danis (1997) has found that when playing with familiar objects, mothers are more likely to direct their children's attention; and when playing with novel objects, mothers are more likely to follow children's attention of focus. Therefore, the dynamics of the interaction, in terms of who is leading and who is following, may be different in the Learning and Play conditions.

Current Study

In this study, we recruited parents and their toddlers into the labortaory to participate in free play with toys. During the free play, both participants wore a head-mounted eye-tracker that recorded their eye movements. Head-mounted eye-tracking provided high-resolution data from toddler's and parent's first-person view and allowed us to examine children's attention in real-time interactions. We also recorded parents' speech as well as each participants' manual actions on the objects during play. Half of the parent–child dyads played with sets of novel objects made in the lab (Learning condition), whereas the other half with objects that were familiar to children (Play condition). We asked three specific questions: (a) do children show different information seeking behaviors when they play with novel objects compared to playing with familiar objects? (b) do parents show different information providing behaviors in the two conditions? (c) how frequently and in what ways do they achieve the synchrony between children's information seeking and parents' information providing behaviors? To answer the first question, we examined children's looking and touching of objects in play and compared their looking and touching behaviors in the two conditions. To address the second question, we analyzed the base rates of parents' naming and touching of objects and the synchrony between these two types of events in the two conditions. With regard to the third question, we focused on the synchrony between parents' naming and children's attention to objects and investigated whether parents' naming of an object led to children's attention on the same object or whether it was the other way around—children's attention on an object led to parents' naming of it. We used two synchrony measures: (a) the temporal overlap between parents' naming of an object and children's look, of any duration, to the target object, and (b) the

temporal overlap between parents' naming and children's sustained attention to the named object, which is defined as object looks lasting 3 s or longer (e.g., Chen, Castellanos, Yu, & Houston, 2019b; Kannass & Oakes, 2008; Lawson & Ruff, 2004; Yu et al., 2019). The first measure can be viewed as a broader, more lenient, definition of synchrony between children's information seeking and parents' information providing behaviors. The second measure can be viewed as a more stringent definition, as previous studies suggest that infants' sustained attention was crucial for their concept and language learning (Colombo, 2001; Frick & Richards, 2001; Richards, 1997; Yu et al., 2019).

Method

Participants

Participants were 20 toddlers (11 girls, M_{age} = 19.8 months, range = 17–23 months) and their parents (17 moms and 3 dads) recruited in the State of Indiana. Half of the participants were randomly assigned to play with sets of novel objects (Learning condition), whereas the other half played with sets of familiar objects (Play condition) in the experiment. Children in these two conditions had comparable receptive vocabulary size measured by the

MacArthur-Bates Communicative Development Inventories: Words and Gestures (Learning: MacArthur Communicative Development Inventory [MCDI] mean = 112.5, Play: MCDI mean = 98.6, $t(18)$ = 0.267, p = .792). All parents provided informed consent prior to their participation in the study. All procedures in the study were approved by the Institutional Review Board. The entire sample of participants was broadly representative of the State of Indiana (85% White, 10% Black, 5% Asian, Hispanic, and other), and consisted of predominantly working and middle-class families. The primary language used in the families was English. Data collection and coding were conducted between 2013 and 2016.

Design

Parents and their toddlers participated in a toy play session, in which they wore head-mounted eye-trackers (Positive Science, http://www.positive science.com/, also see Franchak, Kretch, Soska, & Adolph, 2011) while playing with each other across a small table (Figure 1A). Each eye-tracker was composed of an eye camera that pointed to the participant's right eye to record eye movements and a scene camera that was placed on the participant's forehead to record first-person view. In addition to
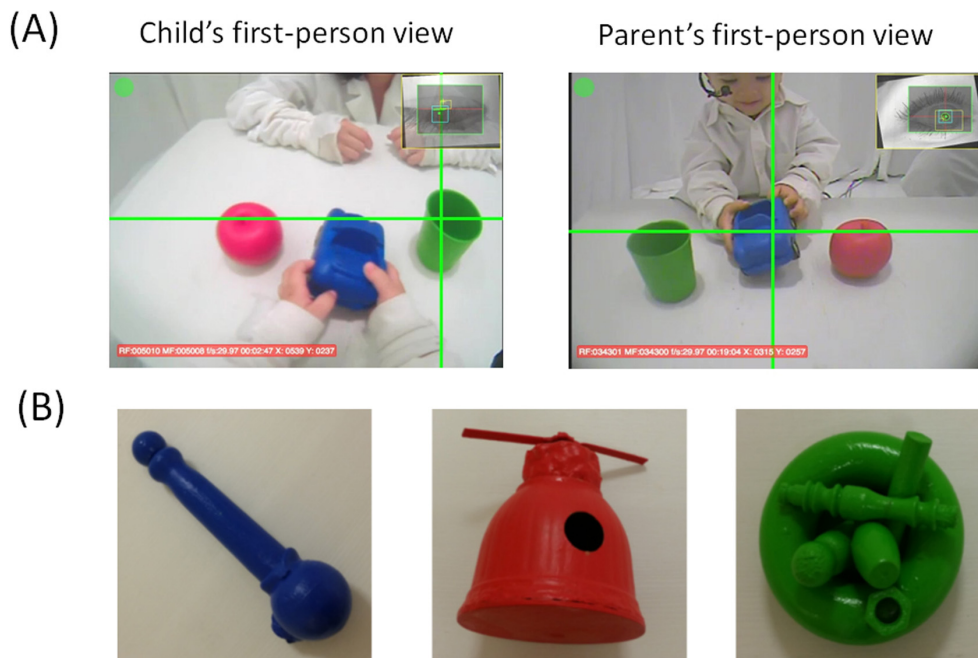

Figure 1. (A) Parent and child sat across from each other at a small table and played with a set of three objects. Both participants wore a head-mounted eye-tracker, which recorded, respectively, where they attended (indicated by a cross-hair in each image) in their egocentric view. (B) Example toy set in the Learning condition. [Color figure can be viewed at wileyonlinelibrary.com]

the eye-trackers worn by the participants, two third-person view cameras were used to record the interaction. Parents' speech was also recorded through a microphone incorporated in the parents' eye-tracker. Half of the parent–child dyads participated in the Learning condition, whereas the other half participated in the Play condition. In each condition, they played with two sets of three objects in an alternating order.

In the Learning condition, the toys consisted of two sets of three novel objects, one blue, one red, and one green, constructed in the lab (examples see Figure 1B). The novel objects were made from wood and plastic and were of comparable overall size (average size: 288 cm^3). Each object was assigned a novel name that followed the phonological rules of English (*dodi*, *habble*, *mapoo*, *tema*, *wawa*, and *zeebee*). Parents were instructed on the names of the objects before playing with each set. They were also provided with a cheat sheet containing the object names on the side of the table for their reference.

In the Play condition, participants played with two sets of three familiar toys. The two toy sets consisted of an apple, a block, a boat, a car, a cup, and a duck. Each set contained one blue, one red, and one green object (example of toys see Figure 1A). Based on the MCDI norms, these objects are familiar and their names are comprehended by children of the tested age (i.e., 17- to 23-month-olds, data retrieved from the open repository WordBank: http://wordbank.stanford.edu/analyses?name=ite m_data, see Frank, Braginsky, Yurovsky, & Marchman, 2017). In addition, we went to each child's MCDI forms and checked whether they knew all the words used in the toy sets. There was one child that did not know one of the six words (*cup*) and another child that did not know two of the words (*boat* and *duck*). In the analyses reported below, we conducted two rounds of testing, one with all data included and the other with those items removed for those two children. The two rounds of analyses yielded the same conclusions. We will report the tests with all data included in the Results section.

Procedure

Prior to the experiment, two experimenters helped the parent and toddler put on eye-trackers. Following that, one experimenter calibrated the eye-trackers by directing the participants' attention to a small toy on the table and moved the toy to several predetermined locations on the table (corners of the table, the center, and a few locations in between).

The procedure was repeated until 15 calibration points that covered the whole table were obtained. Parents in both conditions were instructed to play with their child as they normally would. The parents in the Learning condition were instructed to use the novel names assigned to the objects if they wanted to name the objects. It is noteworthy that we did not ask the parents to teach the names to children, because we wanted them to play naturally with their children as they would in daily life.

The participants played with two sets of toys in an alternating order. They played with each set twice, each time for 1.5 min. This resulted in 6 min of play data for each dyad.

Data Coding

Gaze Data

The eye-trackers recorded at a sampling rate of 30 Hz, which resulted in approximately 10,800 frames of gaze data from each participant. Gaze data were coded frame by frame. We used an in-house-built program that allowed coders to simultaneously play all recorded videos frame by frame. Based on whether the coder was coding the child view or the parent view, the coder can choose the corresponding video(s) accordingly. Four regions of interest (ROIs) were identified, the play partner's face and the three objects participants played with at each moment. Trained coders went through each frame and coded whether participants' gaze direction overlapped with any of the ROIs, and if so, which one (detailed information about ROI coding can be seen in appendix B in Yu & Smith, 2017). In this study, gaze was defined as one (continuous) look to an ROI. We did not use predetermined duration criteria, such as only counting looks being stable for at least a certain number of frames or milliseconds. In theory, a look could be as short as lasting only one frame. However, in practice, only 2.3% of looks in our analyses lasted fewer than three frames. The majority of looks included in our analyses lasted longer than 100 ms. In total, children in the two conditions generated 3,381 looks to the four ROIs while parents generated 8,340 looks to the ROIs. Children's ROI looks served as the gaze data in the following analyses.

To assess the reliability of the gaze coding, a second coder independently coded data for six participants. We compared the coding frame by frame and calculated Cohen's κ for each comparison. Intercoder reliability was good with an average Cohen's κ of .75 (range = .69–.87, Landis & Koch, 1977).

Speech Data

Parents' speech was divided into utterances and transcribed using the open source program Audacity (https://www.audacityteam.org/). The utterance boundaries were defined based on a pause of 400 ms or longer (Suanda et al., 2016; Suarez-Rivera et al., 2019; Yu & Smith, 2012). Following utterance transcription, we then identified the utterances containing a name for an object in play (e.g., *dodi* in the Learning condition or *car* in the Play condition). These utterances were termed *naming utterances* and served as the speech data in the analyses.

A second coder transcribed the speech for six randomly selected parents. We compared the timing and referent of the naming utterances and calculated Cohen's κ for these participants. The average κ score was .89 (range = .82–.94), indicating near-perfect intercoder agreement according to the guidelines of Landis and Koch (1977).

Manual Data

Parents' and children's hand contact with the objects were coded separately. Using the same in-house-built program as described previously, hand contact was coded frame-by-frame from the participants' first-person view and the two third-person view cameras. Participants' left and right hands were coded separately. Trained coders coded whether the participants' left or right hand were in contact with any of the objects, and if so, which one. The data from the two hands' contact were then combined as the manual data for the following analyses.

A second coder coded the manual data for 10 randomly selected participants. Cohen's κ was calculated based on frame-by-frame comparison. Intercoder reliability was near perfect with an average κ score of .95 (range = .90–.98, Landis & Koch, 1977).

Data Analyses

We conducted three sets of analyses. We first examined children's information seeking behaviors by focusing on their looking and touching of the objects (Figure 2A). We compared how often and how long children looked at and touched the objects in the Learning and Play conditions. Previous studies suggest that children may show longer looking and more touching in the Learning condition (Danis, 1997; Oakes et al., 1991; Ross, 1980; Ruff, 1984).

Second, we compared parents' information providing behaviors by calculating their naming and touching rates in the Learning and Play conditions

and then investigating whether there were differences in how often parents' naming of an object overlapped with their touching of the same object (Figure 2B). Parents may be more likely to name the objects in the Play condition (Masur, 1997), but have better naming and touching synchrony in the Learning condition (Gogate et al., 2000, 2006, 2015; Lund & Schuele, 2015; Masur, 1997; Suanda et al., 2016; Yu & Smith, 2012).

Third, we looked at the synchrony between parents' naming of an object and children's attention on the same object. We investigated whether parents' naming of an object overlapped with children's look to the same object, and if so, which event happened first (Figure 2C). We used two synchrony measures: (a) the temporal overlap between parents' naming of an object and children's look, of any duration, to the target object, and (b) the temporal overlap between parents' naming and children's sustained attention to the named object, which is defined as object looks lasting 3 s or longer (e.g., Chen et al., 2019b; Kannass & Oakes, 2008; Lawson & Ruff, 2004; Yu et al., 2019). One possibility is that children may look longer to the novel objects, and this allows parents to follow in and name the object of children's interest and thus creates better synchrony in the Learning condition. However, given that children are familiar with the objects and their names in the Play condition, another possible outcomes is that there is better synchrony between parents' naming and children's attention in the Play condition, because children can follow parents' naming. A third possibility is that there may be no difference in the overall synchrony measures, but the synchrony is achieved by different mechanisms—one by children following parents' naming and the other by parents following children's attention.

These three sets of analyses provide a detailed comparison of children's and their parents' behavioral and attentional patterns in two free-play contexts. The first two sets of analyses were informed by previous literature and were of a more hypothesis-driven nature. The third set of analyses focused on the dynamics of children's information seeking and parents' information providing behaviors. This type of joint behaviors is relatively unexplored in prior literature; and therefore, the third set of analyses is more exploratory in nature.

Results

We used generalized estimating equations (GEE) for all following analyses to account for
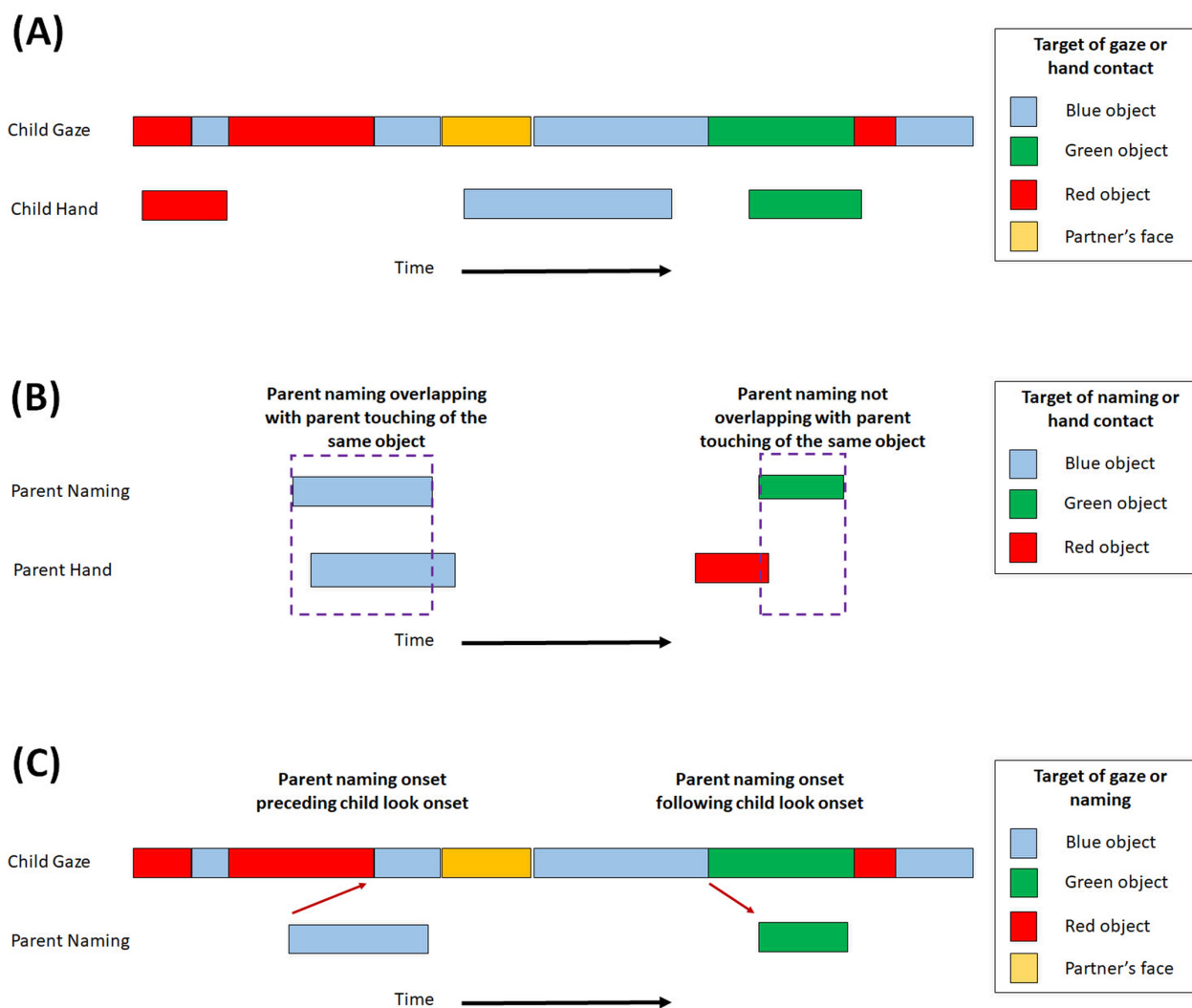
Figure 2. (A) The first set of analyses examines the frequencies and durations of children's looking and touching of the objects. (B) The second set of analyses investigated the frequencies of parents' naming and touching behaviors as well as the overlapping between these two events. (C) The third set of analyses tests whether parent's naming of an object overlaps children's look to the same object, and if so, which event starts first. [Color figure can be viewed at wileyonlinelibrary.com]

nonindependence and non-normal distributions of the data (Liang & Zeger, 1986). Condition (Learning vs. Play) was taken as the independent variable. Depending on different research questions, frequencies, durations, or proportions of gaze, manual actions, or speech measures were taken as the dependent variables in different analyses. The raw data and analyses scripts can be found at OSF site (https://osf.io/d4aev/?view_only=fa173495267f420 2bf25360f59167ce3).

Children's Looking and Touching of Objects

In the first set of analyses, we examined the overall patterns of children's looks to the 4 ROIs (i.e., face and objects in play) and their touching of

the objects in play. Children in the two conditions generated similar frequencies of looks to the 4 ROIs (Learning: $M = 155.6$, $SD = 36.95$; Play: $M = 182.5$, $SD = 63.86$; Wald $\chi^2 = 1.48$, $p = .22$). Children in both groups generated approximately 80% of ROI looks to the objects (Learning: 77%, Play: 79%) and the rest were looks to parents' face. There was no significant group difference in the proportion of looks to the objects (as opposed to parents' face) in the two conditions (Wald $\chi^2 = 0.29$, $p = .59$). As illustrated in Figure 3A, children's object looks were overall longer in the Learning condition than in the Play condition (Learning: $M = 2.10$ s, $SD = 2.84$; Play: $M = 1.39$ s, $SD = 1.65$; Wald $\chi^2 = 17.73$, $p < .001$). Children in the Learning condition (Figure 3B) also had a higher proportion of
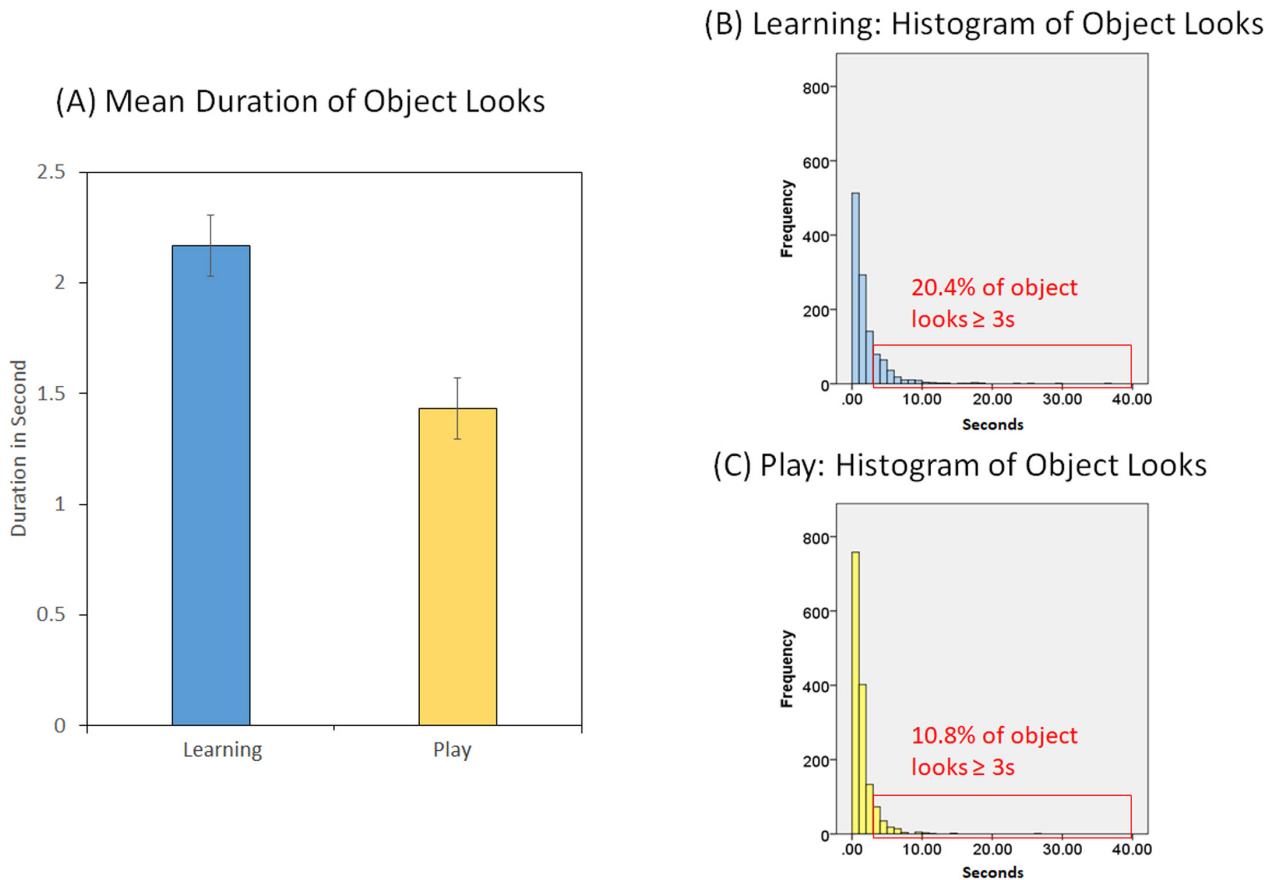
(A) Mean Duration of Object Looks

(B) Learning: Histogram of Object Looks

(C) Play: Histogram of Object Looks

Figure 3. (A) Mean duration (and *SE*) of object looks in the Learning and Play conditions. (B) Histogram and proportion of object looks lasting 3 s or longer in the Learning and Play conditions. [Color figure can be viewed at wileyonlinelibrary.com]

object looks lasting 3 s or longer, which are viewed as sustained attention by prior studies (e.g., Chen et al., 2019b; Kannass & Oakes, 2008; Lawson & Ruff, 2004; Yu et al., 2019), than children in the Play condition (Figure 3C, Wald $\chi^2 = 13.32$, $p < .001$).

In addition to longer object looks, children in the Learning condition were also more likely to touch the objects than children in the Play condition (Learning: $M = 109.4$, $SD = 21.47$; Play: $M = 84.4$, $SD = 22.28$; Wald $\chi^2 = 7.26$, $p = .01$). However, there was no group difference in their touch durations (Learning: $M = 3.03$ s, $SD = 6.34$; Play: $M = 3.83$ s, $SD = 5.58$; Wald $\chi^2 = 2.65$, $p = .10$). These results together suggest that object novelty affected children's looking and touching behaviors in play. Specifically, children in the Learning conditions produced more exploratory behaviors with longer looks and more touches.

Parents' Naming and Touching of Objects

In the second set of analyses, we examined parents' information providing behaviors by focusing on the overall patterns of parents' naming and touching of the objects and the synchrony between these two types of events. On average, parents in the Play condition produced over twice as many utterances containing the object names than parents in the Learning condition (Learning: $M = 26.30$, $SD = 22.96$; Play: $M = 58.40$, $SD = 28.44$; Wald $\chi^2 = 8.568$, $p < .001$). Their naming utterances also tended to be longer in the Play condition than in the Learning condition (Learning: $M = 1.27$ s, $SD = 0.73$; Play: $M = 1.94$ s, $SD = 1.33$; Wald $\chi^2 = 15.86$, $p = .001$).

With regard to parents' touching, parents in the two conditions touched the objects equally frequently throughout the interaction (Learning: $M = 130.70$, $SD = 53.31$; Play: $M = 103.10$, $SD = 41.56$; Wald $\chi^2 = 1.85$, $p = .17$) and there was no difference in their mean touch durations (Learning: $M = 2.60$ s; $SD = 3.86$; Play: $M = 2.38$ s, $SD = 3.48$; Wald $\chi^2 = 0.52$, $p = .47$). However, there was a higher proportion of parents' naming instances that overlapped with their touching of objects in the Learning condition than the Play

condition (Learning: $M = 0.68$, $SD = 0.13$; Play: $M = 0.54$, $SD = 0.15$; Wald $\chi^2 = 4.74$, $p = .03$). This suggests that parents were more likely to touch an object when they named the object in the Learning condition than in the Play condition. Interestingly, even though children were more likely to touch the objects in the Learning condition, parents' naming of an object did *not* overlap with children's touching more frequently in the Learning condition (Learning: 56%, Play: 50%, Wald $\chi^2 = 1.02$, $p = .31$). This suggests that parents' naming (or not naming) of the object touched by children did not differ as a function of object novelty. The synchrony between parents' naming and their own touching of objects suggest that they were more likely to use their manual actions to highlight the referent of their naming utterance in the Learning condition. This behavior reduced the ambiguity of their naming utterances and potentially made it easier for children to learn the word-object mappings.

Synchrony Between Parents' Naming and Children's Attention

We next asked how many of parents' naming utterances of an object overlapped with children's attention to the target object and, when they overlapped, which event occurred first. We used two synchrony measures. The first one can be viewed as a "broader", or more "lenient" measure, which

focused on the overlapping between parents' naming utterances and children's look to the named object, regardless of the gaze length. This measure included children's brief glance toward the object. The second measure focuses on parents' naming and children's sustained attention, which has been viewed as important for information processing and learning (Colombo, 2001; Frick & Richards, 2001; Richards, 1997). With regard to the first synchrony measure, interestingly, in both conditions, approximately 70% of parents' naming of an object overlapped with children's look to the same object and there was no significant difference between groups (Figure 4A; Learning: 70%, Play: 68%, Wald $\chi^2 = 0.33$, $p = .57$). We then further examined whether parents' naming of an object preceded or followed children's look to the same object. As can be seen in Figure 4B, in the Learning condition, of those naming utterances overlapping with children's look on the same object, a larger proportion of parents' naming utterances followed children's look to the same object (parent naming preceding child look: 35%, parent naming following child look: 65%). In contrast, in the Play condition, parents' naming utterance onsets were more likely to start before children's look to the same object (parent naming preceding child look: 56%, parent naming following child look: 43%). The distributions of these two types of temporal overlapping relationships were significantly different across conditions (Wald $\chi^2 = 19.10$, $p < .001$).
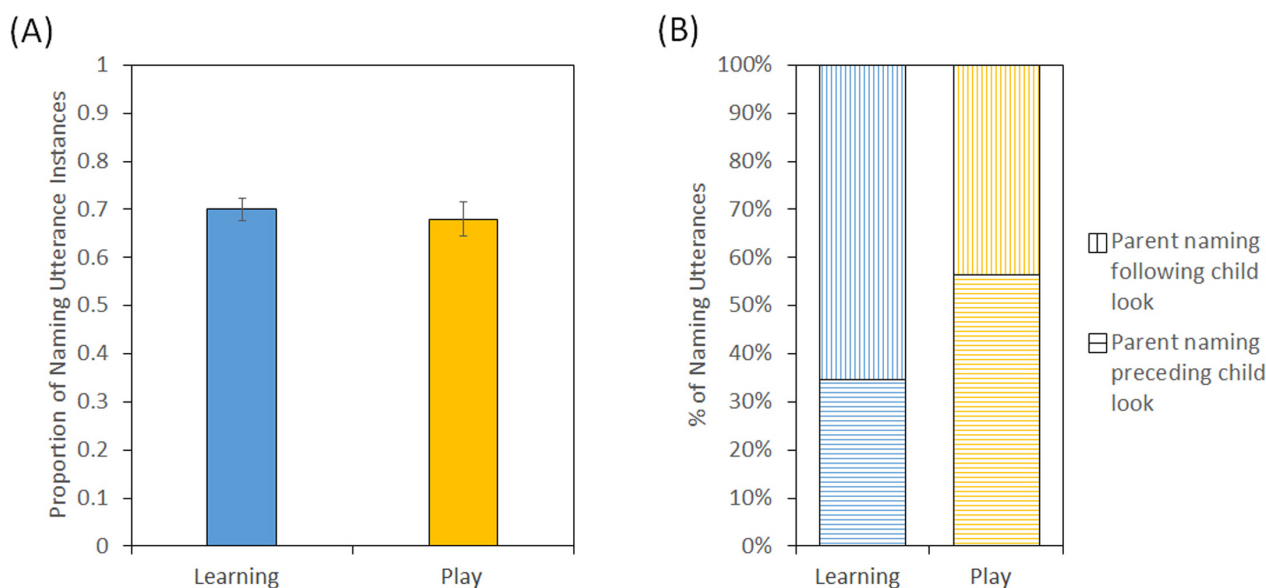

Figure 4. (A) Proportion (and *SE*) of parents' naming utterances overlapping with children's look to the same object. (B) Percentages of naming utterances following or preceding children's look. [Color figure can be viewed at wileyonlinelibrary.com]

We next examined the synchrony between parents' naming and children's sustained attention. Parents' naming utterances overlapped with children's sustained attention over twice as often in the Learning condition than in the Play condition (Figure 5A, Learning: 48%, Play: 22%, Wald χ^2 = 23.89, p < .001). Interestingly, in both Learning and Play conditions, parents' naming onset tended to follow children's sustained attention onset and there was no significant group difference (Figure 5B, Wald χ^2 = 2.15, p = .14).

As mentioned in the Method section, the participants in both conditions played with two sets of objects, each set twice. One question to ask is whether there is any difference in the naming-attention synchrony between children's first and second encounters of the objects within each condition. If so, the differences will reflect how their familiarity and learning of the objects affect their subsequent interaction with the relevant objects. To answer this question, we conducted two additional sets of analyses that investigated (a) whether the naming-attention measures changed between children's first and second encounters of the objects, and (b) whether children's look duration to a named object changed between their first and second encounters. In the Learning condition, naming-attention synchrony did not differ between the first and second encounters of the objects, regardless of whether we used the lenient or strict measures

(Naming-overall attention synchrony: Wald χ^2 = 1.21, p = .27, Naming-sustained attention synchrony: Wald χ^2 = 2.19, p = .14). In the Play condition, encounter order also had no significant effect on the naming-attention synchrony measures (Naming-overall attention synchrony: Wald χ^2 = 0.64, p = .43, Naming-sustained attention synchrony: Wald χ^2 = 0.09, p = .76). We then examined whether children's look duration to a named object changed between the first and second encounters. In the Learning condition, children's looks to the named objects were significantly longer during the first encounter than the second encounter (First encounter: M = 3.94 s, SD = 4.27; Second encounter: M = 2.38 s, SD = 3.20; Wald χ^2 = 11.04, p = .001). The patterns were different in the Play condition. Children had slightly longer looks to the familiar named objects in their second encounter (First encounter: M = 1.91 s, SD = 1.58; Second encounter: M = 2.52 s, SD = 3.23; Wald χ^2 = 5.18, p = .02). It is interesting to note that children's look durations to the novel named objects in their second encounter in the Learning condition were comparable to children's overall look durations to the familiar named objects in the Play condition (Wald χ^2 = 0.32, p = .57). This (indirectly) suggests that children's decreased look durations at the second encounter in the Learning condition were likely due to their learning or increased familiarity with the novel objects.
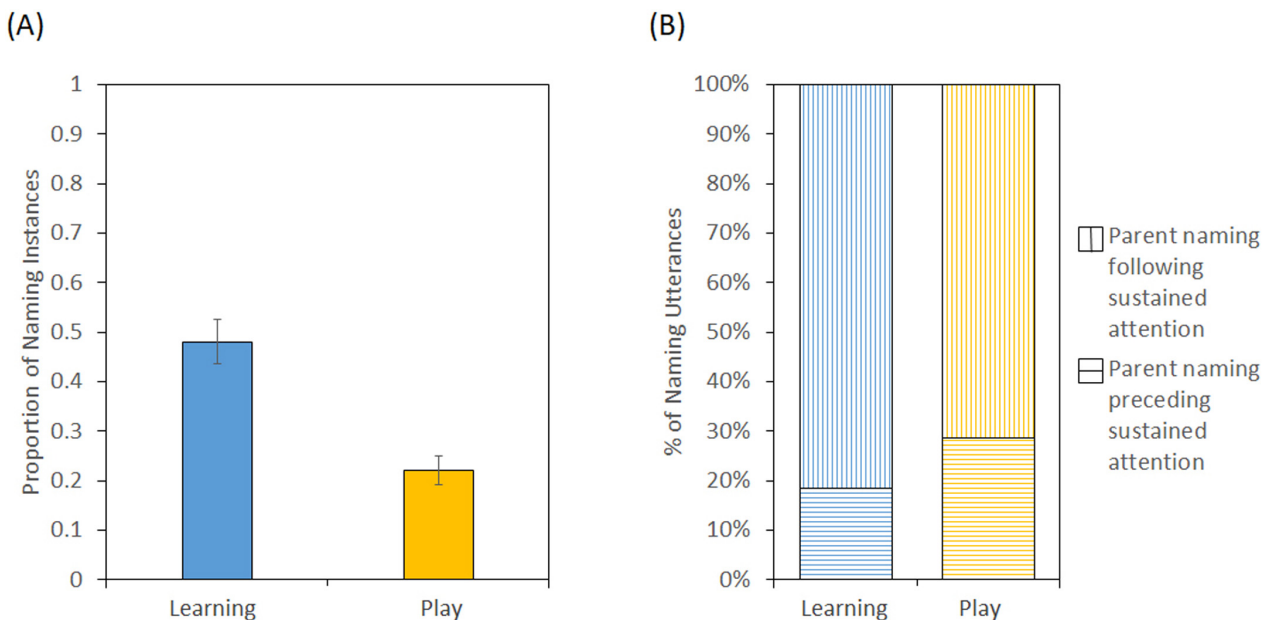

Figure 5. (A) Proportion (and *SE*) of parents' naming utterances overlapping with children's sustained attention to the same object. (B) Percentages of naming utterances following or preceding children's sustained attention. [Color figure can be viewed at wileyonlinelibrary.com]

Another question to ask is whether parents' synchronous multimodal cues affect children's look durations. To answer this question, we examined whether parents' touch of a named object affected children's look duration to the object. In the Learning condition, there was a significant multimodal effect (Wald χ^2 = 27.10, p < .001), in that parents' touch of a named object significantly extended children's look to the object in both the first and second encounters (First encounter: with touch M = 4.48 s, without touch M = 2.53 s; Second encounter: with touch M = 2.89 s, without touch M = 1.10 s). In contrast, parents' touch of a familiar named object in the Play condition did not have a significant effect on children's look duration (Wald χ^2 = 2.03, p = .15), even though children's look duration to a touched object was numerically longer than an untouched object in both the first and second encounters (First encounter: with touch M = 2.13 s, without touch M = 1.65 s; Second encounter: with touch M = 2.93 s, without touch M = 2.07 s).

Together, these different sets of results show that, the overall synchrony between parents' naming and children's looks toward objects, regardless of duration, were comparable in the two conditions. However, in the Learning condition, parents' naming was more synchronous with children's sustained attention, which is associated with information processing and learning (Colombo, 2001; Frick & Richards, 2001; Richards, 1997). Parents' naming tended to follow children's sustained attention on objects in both conditions. This suggests that parents' naming is contingent on children's sustained attention, a behavior indicative of parents' responsiveness to children's attentional state (Tamis-LeMonda, Kuchirko, & Song, 2014). Children in the Learning condition reduced their look durations to the named objects in their second encounter to a degree comparable to the patterns seen in the Play condition. This suggests that their look durations changed as a function of learning and increased familiarity with the objects (Oakes et al., 1991; Ross, 1980; Ruff, 1984). In addition, parents' touch of a named object significantly extended children's look to the named object in the Learning condition. This provides evidence that multimodal cues can further sustain children's attention to novel objects (Suarez-Rivera et al., 2019).

Discussion

This study investigated the dynamics of parent–child interactions in a condition where toddlers encountered novel objects and their names for the first time (i.e., Learning condition) and in a control condition where the objects were familiar and their names were known (i.e., Play condition). The results suggest that children and parents adapted their information seeking and providing behaviors depending on the nature of the learning opportunities. Importantly, their behaviors in the Learning condition created better and clearer signals for word learning. In the following, we first discuss children's information seeking behaviors and then parents' information providing behaviors in the Learning and Play conditions. After that, we focus on their joint behaviors by looking at the synchrony between parents' naming and children's attention. In the last two sections, we consider the developmental implications of the results as well as the limitations of this study and future directions.

The Role of Children as Word Learners—Information Seeking

This study revealed several differences in children's information seeking behaviors in the Learning and Play conditions. Children looked longer at novel objects and were more likely to touch them in the Learning condition, which is consistent with previous findings that novel objects tend to attract children's attention and facilitate object exploration (Danis, 1997; Oakes et al., 1991; Ross, 1980; Ruff, 1984). Within-condition analyses on the effect of object encounters provide (indirect) evidence that children's overall longer looking behaviors in the Learning condition were due to object novelty, and not because children in this condition were "longlookers" by nature. Their longer looking and examining suggest a larger amount of visual information being processed in the Learning condition (Oakes & Tellinghuisen, 1994). It has been argued that this kind of behaviors involves more active cognitive processing and can lead to better learning of object characteristics, which is in turn beneficial for the learning and memory of object names (Kucker & Samuelson, 2012; Oakes & Tellinghuisen, 1994; Oakes et al., 1991; Ruff, 1984, 1986).

Why did children show different behaviors in these two conditions? More specifically, why did they show longer looking and more exploration of the novel objects in the Learning condition? One possibility is that they were, either explicitly or implicitly, aware of the word or object learning opportunities and tried to gather as much information as possible to maximize learning. Studies using computational approaches have suggested that

infants allocate their attention to sample stimuli with maximal subjective novelty to maximize visual category learning (Twomey & Westermann, 2018). It is thus possible that infants in our study actively explored the novel objects in the Learning condition to maximize learning. Another possibility is related to a more general learning mechanism that has been observed in both humans and nonhuman species—to reduce uncertainty about objects and events in the world (Gottlieb, Oudeyer, Lopes, & Baranes, 2013; Kidd & Hayden, 2015; Oudeyer & Smith, 2016; Zettersten & Saffran, 2020). Studies have shown that infants and children would allocate their attention and structure their play to maintain information absorption and to reduce uncertainty (Kidd, Piantadosi, & Aslin, 2012; Schulz & Bonawitz, 2007). These two possibilities are related and not mutually exclusive. It is possible that both play a role in influencing toddlers' behaviors in this study.

Children's information seeking or active exploration behaviors have been associated with curiosity, which has been proposed to be the motivator for learning (Gottlieb et al., 2013; Kidd & Hayden, 2015; Oudeyer & Smith, 2016). The progress in learning in and for itself is fun and generates intrinsic rewards for learners (Oudeyer & Smith, 2016). Information gathered through these active exploration behaviors can be used not only for immediate purposes, but also be stored for future tasks or events (Gottlieb et al., 2013). Curiosity-driven learning promotes the acquisition of language and a wide-range of domain-specific knowledge in both humans and other species (Kidd & Hayden, 2015; Oudeyer & Smith, 2016).

In this study, we focused on the coupling of children's information seeking behaviors and parents' information providing behaviors. Children's longer or sustained attention on objects in the Learning condition provided parents with the opportunities to join in and name the object of interest at the optimal time. In this way, children's curiosity-driven information seeking behaviors did not only create opportunities for them to learn; but also opportunities for parents to help their children learn. Naming an object attended by a child has been shown to be beneficial for early word learning (MacRoy-Higgins & Montemarano, 2016; Tomasello & Farrar, 1986; Yu & Smith, 2012; Yu et al., 2019). In this research, we did not directly test how they relate to the learning of novel words. One future direction is to examine how different information seeking behaviors are associated with children's word learning outcomes and long-term vocabulary development.

The Role of Parents as Word Teachers—Information Providing

Parents showed different information providing behaviors in the Learning and Play conditions. Parents were more likely to name familiar objects than novel objects, a finding consistent with those in Masur (1997). Even though parents were instructed of the names of the novel objects before playing with each toy set and also provided with a cheat sheet for their reference in the Learning condition, parents' own familiarity with the toys (and their names) used in the Play condition likely made them more comfortable and confident in referring to those objects by name in the interaction. Interestingly, parents' naming of an object overlapped with children's overall looks, regardless of gaze length, on the named object equally frequently in both conditions. Yet, the synchrony was more likely to be established by parents naming the target of children's attentional focus in the Learning condition than in the Play condition. This pattern suggests that parents, either consciously or unconsciously, changed their naming behaviors depending on the momentary dynamics of children's attentional states. They provided fewer but higher quality of word-learning opportunities—as naming the object the child is looking at greatly reduces the ambiguity of the naming utterance—in the Learning condition. This type of contingent and contiguous naming behavior indicates parents' responsiveness and have been shown to facilitate children's word learning (Tamis-LeMonda et al., 2014; Yu & Smith, 2012).

Parents' *overall* manual action frequencies or durations did not differ in the Learning and Play conditions. However, they were more likely to touch an object *when naming it* in the Learning condition. This behavior likely increased the saliency of the novel object being named and provided additional cues for children to discern the referent of parent's utterance (Gogate et al., 2000, 2006, 2015; Lund & Schuele, 2015; Masur, 1997; Suanda et al., 2016; Yu & Smith, 2012). This finding also underlines the importance of taking timing into account (Xu, de Barbaro, Abney, & Cox, 2020). At the macro-level, parents' overall manual actions did not differ across conditions. However, at the micro-level, the temporal relationships between parents' naming and touching of a novel object were more synchronized in the Learning condition. This synchrony may serve as an important cue for children's learning of the object labels (Gogate et al., 2006).

The Coupling of Parents' and Children's Joint Behaviors as a Coordinated System

As mentioned previously, there were two synchrony measures: (a) synchrony between parents' naming and children's look, of any duration, toward the named object, and (b) synchrony between parents' naming and children's sustained attention toward the named object. Regarding the first synchrony measure, parents' naming of an object overlapped with children's looks, of any duration, on the named object equally frequently in both conditions. However, parents' naming tended to follow children's attention in the Learning condition. In contrast, their naming was more likely to lead children's attention in the Play condition than in the Learning condition. Previous studies suggest that the synchrony between parents' naming and children's attention is crucial for their learning of object names (MacRoy-Higgins & Montemarano, 2016; Tomasello & Farrar, 1986; Yu & Smith, 2012). Our study adds to the literature by showing that the synchrony maybe more likely to be established by parents' following children's attention when things are novel. Prior research indicates that parents' naming following children's attention results in better word learning and larger vocabulary size than parents' naming directing children's attention (Tomasello & Farrar, 1986; Tomasello & Todd, 1983; Yu & Smith, 2012). The parent-following pattern in the Learning condition is likely beneficial for children's learning of the novel object names. As children get more familiar with the objects and their names, there may be a gradual shift, in that the naming-attention synchrony becomes more likely to be established through children following parents' naming, as is evident from the Play condition.

Our second synchrony measure showed that parents' naming was more likely to overlap with children's sustained attention in the Learning condition. As children's looks were longer in the Learning condition, parents might have more opportunity to follow in and talk about the object of their child's interest. This interpretation was supported by the results that, in both conditions, parents' naming tended to follow the onset of children's sustained attention (Figure 5B). This following behavior indicates parents' contingent responsiveness to children's attentional state in toy play, which has been found to facilitate children's word learning and language development (Tamis-LeMonda et al., 2014). As mentioned earlier, parents were more likely to touch an object while naming it in the Learning condition. It has been

proposed that multimodal cues can sustain children's attention (Suarez-Rivera et al., 2019). We found that parents' touch of a named object extended children's look to the object in the Learning condition, but not in the Play condition. This suggests that there is a positive feedback loop in the Learning condition, in that parents' naming followed children's sustained attention on an object, and their multimodal naming-touching further sustained children's attention to the named object.

Together, these results all point to the conclusion that novel objects did not only attract children's attention and change children's exploratory behaviors, they also changed the synchrony between parents' naming and manual actions. These changes from the parents' part further sustained children's attention on the objects and facilitated the coupling between parents' naming and children's attention. The intense and sustained object looks and examination from the children's part, the synchronous multimodal behaviors from the parents' part, and the coupling between children's attention and parents' naming all contribute to creating better and clearer signals for word learning in the Learning condition.

Developmental Implications

It has been proposed that word learning should be understood in the context of dyadic communication, in that the parents' goal is to successfully communicate information with their children (Renzi et al., 2017; Yurovsky, 2018). With this goal in mind, parents modify their speech and behaviors, depending on the momentary dynamics and their children's feedback, to make the information easier for children to understand (Smith & Trainor, 2008; Yurovsky, 2018; Yurovsky, Doyle, & Frank, 2016). Our results are largely in line with this proposal. Using two experimental toy play conditions for direct comparisons, we found that children increased their visual and manual exploration when encountering novel objects with to-be-learned names. Their sustained attention and intense examination provided the information necessary for object recognition and discrimination (Bambach, Crandall, Smith, & Yu, 2016, 2018; Oakes et al., 1991). Parents' behaviors suggest that they, either consciously or unconsciously, provided (additional) scaffolding in various ways when they played with novel objects with their children. The changes in parents' and children's joint behaviors further created better synchrony between parents' naming and children's sustained attention to the named object

(Oakes & Tellinghuisen, 1994; Oakes et al., 1991; Ruff, 1984; Tomasello & Farrar, 1986).

It has been argued that the data in young children's word-learning environment is too messy and posts great challenges for young learners to learn word-object mappings (e.g., Medina, Snedeker, Trueswell, & Gleitman, 2011). At naming moments wherein children hear object labels, high-quality data should contain unambiguous or easy-to-identify referents and clear visual signals of the named objects. However, there are usually many potential referents in view every time a child hears a novel word (e.g., Medina et al., 2011; Smith & Yu, 2008). As a result, many studies have focused on top-down cognitive or sociocognitive learning mechanisms infants use to find the referent of a novel word and to learn the word-object mappings (Baldwin, 1993; Markman & Hutchinson, 1984; Smith & Yu, 2008; Snedeker & Gleitman, 2004; Tomasello & Akhtar, 1995; Waxman & Booth, 2001). Yet, recent studies have shown that children's own manual actions, such as touch, handling, manipulation, often "declutter" the visual input (Bambach et al., 2016, 2018 Suanda, Barnhart, Smith, & Yu, 2019; Tsutsui et al., 2020). Here, we show that children's own bottom-up information seeking behaviors and parents' contingent information providing behaviors together reduce the referential ambiguity of utterances containing novel object names. Their joint behaviors in real-time interactions create a tight coupling between children's sustained attention on objects and parents' naming; and such synchrony provides optimal opportunities for toddlers to learn novel object names. These results underscore the importance of taking an embodied view of early word learning. Our findings also indicate that the visual and linguistic data created for word learning during parent–child interaction may not be as messy as previously assumed. It is thus necessary to re-think the nature of the input in young learners' environment and how they may relate to children's use of top-down learning mechanisms.

Limitations and Future Directions

Our sample size (10 dyads per condition) is relatively small compared to some previous studies on infants' information seeking behaviors using fixed-trial designs (e.g., Oakes & Tellinghuisen, 1994; Oakes et al., 1991; Ruff, 1986). However, it is important to note that we collected micro-level sensorimotor data at high-density and coded data frame by frame. On average, each child generated 169 looks and over 100 touches on the objects during the experiment. Compared to studies using fixed trials, which usually consist of no more than 20 trials, the amount of data we collected for each child was extremely large. Importantly, previous studies on infants' micro-level sensorimotor behaviors have shown that high-density data collected within individuals can produce reliable and generalizable results, even with a small sample size (Yoshida & Smith, 2008; Yu & Smith, 2012). One future direction is to increase the sample size and test the generalizability of our results and examine how individual differences may affect the dynamics of interaction.

It is also important to note that the participants in our study are fairly homogeneous in terms of cultural and socioeconomic backgrounds. Previous studies have shown that cultural and socioeconomic differences can affect how parents interact with their children (Bornstein, Cote, & Kwak, 2019; Bornstein et al., 1992; Gogate et al., 2015; Rowe, 2008; Tamis-LeMonda, Song, Leavell, Kahana-Kalman, & Yoshikawa, 2012). For example, parents from different cultural backgrounds respond differently to their children's looking behaviors and they differ in their use of multimodal cues when providing information (e.g., Bornstein et al., 1992; Gogate et al., 2015). Parents of higher socioeconomic status talk more and use more diverse vocabulary and longer utterances when they interact with their children (Rowe, 2008). It is thus important for future research to include a more diverse sample and examine how cultural and socioeconomic backgrounds affect parents' and children's individual and joint behaviors in real-time interactions.

Finally, our study has shown that parents and children change their individual and joint behaviors when encountering novel objects with to-be-learned names. One question to ask is whether these findings can generalize to the learning of other types of words, such as verbs, adjectives, or prepositions. It would be interesting for future work to investigate whether parents provide more multimodal cues to highlight the actions or properties to be learned and use more explicit explanations when they introduce novel words to children.

Conclusions

Everyday learning contexts contain various degrees of ambiguity and uncertainty. When hearing a novel name, there are often a few to many potential referents in the learning environment. To successfully learn object names, children need to

solve the problem of referential ambiguity, which has long been deemed as the key problem in early word learning (Quine, 1960). This study provided the first detailed evidence on how the real-time behavioral and attentional patterns from both children and parents may reduce the degree of referential ambiguity and potentially facilitate word learning during parent–child social interaction. We found that when encountering novel objects with to-be-learned names, the dyad's individual and joint behaviors together simplify the referential ambiguity problem by making the target of parent's naming utterance easier to discern. Understanding how children and parents adjust their real-time behaviors in the learning environment to create better data for learning is not only critical for understanding how typically developing children successfully acquire their early vocabulary but also for providing guiding principles to create better learning environments for children with delays in language development.

References

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language*, *20*, 395–418. https://doi.org/10.1017/S030500090 0008345

Baldwin, D. A., & Markman, E. M. (1989). Establishing word-object relations: A first step. *Child Development*, *60*, 381–398. https://doi.org/10.2307/1130984

Bambach, S., Crandall, D. J., Smith, L. B., & Yu, C. (2016). Active viewing in toddlers facilitates visual object learning: An egocentric vision approach. In D. Grodner, D. Mirman, A. Papafragou, & J. Trueswell (Eds.), *Proceedings of the 38th annual conference of the Cognitive Science Society* (pp. 1631–1636). Austin, TX: Cognitive Science Society.

Bambach, S., Crandall, D. J., Smith, L. B., & Yu, C. (2018). Toddler-inspired visual object learning. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Paper presented at Neural Information Processing Systems (NIPS)*. Thirty-second Conference on Neural Information Processing Systems (pp. 1209–1218). Montreal, Canada: Neural Information Processing Systems Foundation.

Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2019). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science*, *22*, e12715. https://doi.org/10.1111/desc.12715

Bloom, L., & Tinker, E. (2001). *The intentionality model and language acquisition: Engagement, effort, and the essential tension in development*. Monographs of the Society for Research in Child Development (Serial No. 270), (Vol. *66*, pp. i–101). Hoboken, NJ: Wiley-Blackwell.

Bornstein, M. H., Cote, L. R., & Kwak, K. (2019). Comparative and individual perspectives on mother–infant interactions with people and objects among South Koreans, Korean Americans, and European Americans. *Infancy*, *24*, 526–546. https://doi.org/10.1111/infa.12288

Bornstein, M. H., Haynes, O. M., O'Reilly, A. W., & Painter, K. M. (1996). Solitary and collaborative pretense play in early childhood: Sources of individual variation in the development of representational competence. *Child Development*, *67*, 2910–2929. https://doi.org/10.2307/1131759

Bornstein, M. H., Tamis-LeMonda, C. S., Tal, J., Ludemann, P., Toda, S., Rahn, C. W., . . . Vardi, D. (1992). Maternal responsiveness to infants in three societies: The United States, France, and Japan. *Child Development*, *63*, 808–821. https://doi.org/10.2307/1131235

Chen, C., Castellanos, I., Yu, C., & Houston, D. M. (2019a). Parental linguistic input and its relation to toddlers' visual attention in joint object play: A comparison between children with normal hearing and children with hearing loss. *Infancy*, *24*, 589–612. https://doi.org/10.1111/infa.12291

Chen, C., Castellanos, I., Yu, C., & Houston, D. M. (2019b). Effects of children's hearing loss on the synchrony between parents' object naming and children's attention. *Infant Behavior and Development*, *57*, 101322. https://doi.org/10.1016/j.infbeh.2019.04.004

Colombo, J. (2001). The development of visual attention in infancy. *Annual Review of Psychology*, *52*, 337–367. https://doi.org/10.1146/annurev.psych.52.1.337

Danis, A. (1997). Effects of familiar and unfamiliar objects on mother-infant interaction. *European Journal of Psychology of Education*, *12*, 261–272.

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-mounted eye tracking: A new method to describe infant looking. *Child Development*, *82*, 1738–1750. https://doi.org/10.1111/j.1467-8624.2011.01670.x

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*, 677–694.

Frick, J. E., & Richards, J. E. (2001). Individual differences in infants' recognition of briefly presented visual stimuli. *Infancy*, *2*, 331–352. https://doi.org/10.1207/S15327078IN0203_3

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*, 3–55. https://doi.org/10.1207/s15327817la0101_2

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, *71*, 878–894. https://doi.org/10.1111/1467-8624.00197

Gogate, L. J., Bolzani, L. H., & Betancourt, E. A. (2006). Attention to maternal multimodal naming by 6-to 8-month-old infants and learning of word–object relations. *Infancy*, *9*, 259–288. https://doi.org/10.1207/s15327078in0903_1

Gogate, L., Maganti, M., & Bahrick, L. E. (2015). Cross-cultural evidence for multimodal motherese: Asian Indian mothers' adaptive use of synchronous words and gestures. *Journal of Experimental Child Psychology*, *129*, 110–126. https://doi.org/10.1016/j.jecp.2014.09.002

Golinkoff, R., Hirsh-Pasek, K., Bloom, L., Smith, L. B., Woodward, A., Akhtar, N., . . . Hollich, G. (2000). *Becoming a word learner: A debate on lexical acquisition.* London, UK: Oxford University Press.

Gottlieb, J., Oudeyer, P. Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: Computational and neural mechanisms. *Trends in Cognitive Sciences*, *17*, 585–593. https://doi.org/10.1016/j.tics.2013.09.001

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science*, *11*, F31–F39. https://doi.org/10.1111/j.1467-7687.2008.00768.x

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*, 222–246. https://doi.org/10.1080/15248370802022696

Katz, N., Baker, E., & Macnamara, J. (1974). What's in a name? A study of how children learn common and proper names. *Child Development*, *45*, 469–473. https://doi.org/10.2307/1127970

Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron*, *88*, 449–460. https://doi.org/10.1016/j.neuron.2015.09.010

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS One*, *7*, e36399. https://doi.org/10.1371/journal.pone.0036399

Kretch, K. S., Franchak, J. M., & Adolph, K. E. (2014). Crawling and walking infants see the world differently. *Child Development*, *85*, 1503–1518. https://doi.org/10.1111/cdev.12206

Kucker, S. C., & Samuelson, L. K. (2012). The first slow step: Differential effects of object and word-form familiarization on retention of fast-mapped words. *Infancy*, *17*, 295–323. https://doi.org/10.1111/j.1532-7078.2011.00081.x

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174. https://doi.org/10.2307/2529310

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

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*, 13–22. https://doi.org/10.1093/biomet/73.1.13

Lund, E., & Schuele, C. M. (2015). Synchrony of maternal auditory and visual cues about unknown words to children with and without cochlear implants. *Ear and Hearing*, *36*, 229–238. https://doi.org/10.1097/AUD.0000000000000104

MacRoy-Higgins, M., & Montemarano, E. A. (2016). Attention and word learning in toddlers who are late talkers. *Journal of Child Language*, *43*, 1020–1037. https://doi.org/10.1017/S0305000915000379

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology*, *16*, 1–27. https://doi.org/10.1016/0010-0285(84)90002-1

Masur, E. F. (1997). Maternal labelling of novel and familiar objects: Implications for children's development of lexical constraints. *Journal of Child Language*, *24*, 427–439.

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

Oakes, L. M., Madole, K. L., & Cohen, L. B. (1991). Infants' object examining: Habituation and categorization. *Cognitive Development*, *6*, 377–392. https://doi.org/10.1016/0885-2014(91)90045-F

Oakes, L. M., & Tellinghuisen, D. J. (1994). Examining in infancy: Does it reflect active processing? *Developmental Psychology*, *30*, 748–756. https://doi.org/10.1037/0012-1649.30.5.748

Oudeyer, P. Y., & Smith, L. B. (2016). How evolution may work through curiosity-driven developmental process. *Topics in Cognitive Science*, *8*, 492–502. https://doi.org/10.1111/tops.12196

Power, T. G. (1985). Mother-and father-infant play: A developmental analysis. *Child Development*, *56*, 1514–1524. https://doi.org/10.2307/1130470

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Renzi, D. T., Romberg, A. R., Bolger, D. J., & Newman, R. S. (2017). Two minds are better than one: Cooperative communication as a new framework for understanding infant language learning. *Translational Issues in Psychological Science*, *3*, 19–33. https://doi.org/10.1037/tps0000088

Richards, J. E. (1997). Effects of attention on infants' preference for briefly exposed visual stimuli in the paired-comparison recognition-memory paradigm. *Developmental Psychology*, *33*, 22–31. https://doi.org/10.1037/0012-1649.33.1.22

Ross, G. S. (1980). Categorization in 1-to 2-yr-olds. *Developmental Psychology*, *16*, 391–396. https://doi.org/10.1037/0012-1649.16.5.391

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language*, *35*, 185–205. https://doi.org/10.1017/S0305000907008343

Ruff, H. A. (1984). Infants' manipulative exploration of objects: Effects of age and object characteristics. *Developmental Psychology*, *20*, 9–20. https://doi.org/10.1037/0012-1649.20.1.9

Ruff, H. A. (1986). Components of attention during infants' manipulative exploration. *Child Development*, *57*, 105–114. https://doi.org/10.2307/1130642

Ruff, H. A., Saltarelli, L. M., Capozzoli, M., & Dubiner, K. (1992). The differentiation of activity in infants' exploration of objects. *Developmental Psychology*, *28*, 851–861. https://doi.org/10.1037/0012-1649.28.5.851

Schulz, L., & Bonawitz, E. B. (2007). Serious fun: Preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology*, *43*, 1045–1050. https://doi.org/10.1037/0012-1649.43.4.1045

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*, 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

Smith, L. B., Yu, C., & Pereira, A. F. (2011). Not your mother's view: The dynamics of toddler visual experience. *Developmental Science*, *14*, 9–17. https://doi.org/10.1111/j.1467-7687.2009.00947.x

Smith, N. A., & Trainor, L. J. (2008). Infant-directed speech is modulated by infant feedback. *Infancy*, *13*, 410–420. https://doi.org/10.1080/15250000802188719

Snedeker, J., & Gleitman, L. (2004). Why it is hard to label our concepts. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (pp. 257–294). Cambridge, MA: MIT Press.

Suanda, S. H., Barnhart, M., Smith, L. B., & Yu, C. (2019). The signal in the noise: The visual ecology of parents' object naming. *Infancy*, *24*(3), 455–476.

Suanda, S. H., Smith, L. B., & Yu, C. (2016). The multisensory nature of verbal discourse in parent–toddler interactions. *Developmental Neuropsychology*, *41*, 324–341. https://doi.org/10.1080/87565641.2016.1256403

Suarez-Rivera, C., Smith, L. B., & Yu, C. (2019). Multimodal parent behaviors within joint attention support sustained attention in infants. *Developmental Psychology*, *55*, 96–109. https://doi.org/10.1037/dev0000628

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*, 2135–2152. https://doi.org/10.1111/cdev.13089

Tamis-LeMonda, C. S., Kuchirko, Y., & Song, L. (2014). Why is infant language learning facilitated by parental responsiveness? *Current Directions in Psychological Science*, *23*, 121–126. https://doi.org/10.1177/0963721414522813

Tamis-LeMonda, C. S., Song, L., Leavell, A. S., Kahana-Kalman, R., & Yoshikawa, H. (2012). Ethnic differences in mother–infant language and gestural communications are associated with specific skills in infants. *Developmental Science*, *15*, 384–397. https://doi.org/10.1111/j.1467-7687.2012.01136.x

Tomasello, M., & Akhtar, N. (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. *Cognitive Development*, *10*, 201–224. https://doi.org/10.1016/0885-2014(95)90009-8

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *57*, 1454–1463. https://doi.org/10.2307/1130423

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

Tsutsui, S., Chandrasekaran, A., Reza, M. A., Crandall, D., & Yu, C. (2020). *A computational model of early word learning from the infant's point of view*. arXiv preprint arXiv:2006.02802.

Twomey, K. E., & Westermann, G. (2018). Curiosity-based learning in infants: A neurocomputational approach. *Developmental Science*, *21*, e12629. https://doi.org/10.1111/desc.12629

Waxman, S. R., & Booth, A. E. (2001). Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology*, *43*, 217–242. https://doi.org/10.1006/cogp.2001.0764

Williams, E. (2003). A comparative review of early forms of object-directed play and parent-infant play in typical infants and young children with autism. *Autism*, *7*, 361–374. https://doi.org/10.1177/1362361303007004003

Xu, T., de Barbaro, K., Abney, D., & Cox, R. (2020). Finding structure in time: Visualizing and analyzing behavioral time series. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2020.01457

Yogman, M. W. (1981). Games fathers and mothers play with their infants. *Infant Mental Health Journal*, *2*, 241–248. https://doi.org/10.1002/1097-0355(198124)2:4<241:AID-IMHJ2280020406>3.0.CO;2-8

Yogman, M., Garner, A., Hutchinson, J., Hirsh-Pasek, K., Golinkoff, R. M., & Committee on Psychosocial Aspects of Child and Family Health. (2018). The power of play: A pediatric role in enhancing development in young children. *Pediatrics*, *142*, e20182058. https://doi.org/10.1542/peds.2018-2058

Yoshida, H., & Smith, L. B. (2008). What's in view for toddlers? Using a head camera to study visual experience. *Infancy*, *13*(3), 229–248.

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*, 244–262. https://doi.org/10.1016/j.cognition.2012.06.016

Yu, C., & Smith, L. B. (2017). Hand-eye coordination predicts joint attention. *Child Development*, *88*, 2060–2078. https://doi.org/10.1111/cdev.12730

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

Yurovsky, D. (2018). A communicative approach to early word learning. *New Ideas in Psychology*, *50*, 73–79. https://doi.org/10.1016/j.newideapsych.2017.09.001

Yurovsky, D., Doyle, G., & Frank, M. C. (2016). Linguistic input is tuned to children's developmental level. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th annual meeting of the Cognitive Science Society* (pp. 2093–2098). Austin, TX: Cognitive Science Society.

Zettersten, M., & Saffran, J. R. (2020). Sampling to learn words: Adults and children sample words that reduce referential ambiguity. *Developmental Science*. https://doi.org/10.1111/desc.13064