DOI: 10.1111/infa.12475

RESEARCH ARTICLE

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

Why the parent's gaze is so powerful in organizing the infant's gaze: The relationship between parental referential cues and infant object looking

Lichao Sun | **Hanako Yoshida**

Department of Psychology, University of Houston, Houston, Texas, USA

Correspondence
Lichao Sun, Department of Psychology, University of Houston, 126 Heyne Building, Houston, TX 77204-5022, USA.
Email: lsun21@central.uh.edu

Abstract

Parental scaffolding such as looking at and showing objects has long been considered to be helpful for early attention and language development. However, relatively little is known about how parental social multimodal cues work alone or together in guiding an infant's attention toward the referent items. The present study aims to document the dynamics of social referential input during an interactive play session and specify the different types of social cues in directing infant attention. Forty-three parent-infant dyads (infants aged from 5.0 to 18.0 months) in the U.S. completed a short play session recorded by head-mounted camera with eye-trackers. The present findings suggest that joint attention between parent and infant toward the same referent item often co-occurred with other referential input. Infants were more likely to maintain sustained attention to an object under the circumstance that the parent looked at the same item and named it explicitly. This was not the case when parent object looking accompanied other utterances, like "Look!" or the child's name. The present study highlights the importance of multimodal referential input, which sets up enriched opportunities for children to become sensitive to social input and develop sustained attention for further learning.

1 | INTRODUCTION

Young children learn what they attend to; thus, directing attention to relevant information is essential for early communication and learning. Accordingly, research suggests that socially scaffolded viewing experiences serve as the basis for early language development (Brook & Meltzoff, 2005, 2008; Carpenter et al., 1998; Kannass & Oakes, 2008; Markus et al., 2000; Morales et al., 2000; Tomasello & Farrar, 1986). Parents use various social scaffolding behaviors, particularly referential cues, to direct infant attention to relevant information (Bakeman & Adamson, 1984; Baldwin, 1993; Senju et al., 2008). Specifically, these referential cues include parental gaze (Brooks & Meltzoff, 2002; Caron et al., 2002), parental speech (Flom & Pick, 2003; Namy et al., 2000; West & Iverson, 2017), object handling by the parent (Deák et al., 2014, 2018; Yu & Smith, 2013, 2017), and the combined use of multiple referential cues such as when the parent looks at the handled object or verbally labels the handled object (Deák et al., 2018; Suarez-Rivera et al., 2019; Zukow-Goldring, 1996).

Parental referential cues have been observed and measured independently in various task contexts and have been found to be strongly associated with the child's visual experiences. For example, recent studies of moment-to-moment infant-centered viewing show that parents actively bring objects into their young children's visual field, thus creating opportunities for children to focus attention on the handled objects (Burling & Yoshida, 2019; Deák et al., 2014; Yu & Smith, 2013, 2017). Parent object looking can also guide infant attention toward an object, prompting attention sharing and sustained attention toward the target item (Gredebäck et al., 2010; Suarez-Rivera et al., 2019; Yu & Smith, 2017).

Despite accumulating evidence of parents' active use of referential cues and the potential impact of such cues on the infant's attention and learning, relatively little is known about underlying mechanisms. Under what circumstances are these referential cues used, and how do they work, either alone or together, to direct an infant's attention? What specific referential cues do parents frequently use when interacting with an infant? Do parents tend to use referential cues to direct attention sharing or to respond to infant object looking? What referential cues are most effective to lead and maintain infant attention to the target object? The present study aims to characterize the dynamic use of referential input and to clarify the contribution of each cue and various combinations of cues to infant attention during parent–child object play.

1.1 | Parent's referential cues

A set of referential cues has been documented to have robust impacts on infant attention to referent objects (Deák et al., 2014; Suarez-Rivera et al., 2019; Zukow-Goldring, 1996). These referential cues include (1) parent's object looking, (2) object handling, (3) object labeling, and (4) combinations of cues, such as parent labeling while handling the object. Effects of these cues on an infant's visual experiences have been found across different experimental task paradigms (Amano et al., 2004; Deák et al., 2008; Flom et al., 2004; Franco et al., 2009). In the case of parent object looking, infants start to respond by looking at an object in the direction of the parent's gaze as young as 2–6 months of age (Bakeman & Adamson, 1984; Butterworth & Cochran, 1980; Gredebäck et al., 2010; Morales et al., 1998; Scaife & Bruner, 1975; Senju et al., 2008; Tomasello, 1995). By 1 year of age, infants become efficient in following parental gaze even when head movement is controlled (Butterworth, 1991; Butterworth & Cochran, 1980; Brooks & Meltzoof, 2002, 2005). However, recent studies that document infants' visual exploration in social contexts reveal that infants rarely look at a parent's face during interactions (Deák et al., 2014; Franchak et al., 2011; Yoshida & Smith, 2008; Yu

& Smith, 2013), which leads to the questions of how the documented early gaze following is generated and whether parental gaze alone directs subsequent infant attention.

The central premise for the present study is that early gaze following is learned through multimodal referential input with or without concurrent attention toward the target object by the parent (Deák et al., 2014; Yu & Smith, 2017). Experiences that are socially coordinated between parents and children are not limited to attention sharing but could consist of multimodal referential cues together (Chang et al., 2016; Suarez-Rivera et al., 2019). To establish social coordination, it is essential to measure the strength of the social cues, individually and together, in relation to infant attention to referents in visually cluttered environments.

As for object handling, it has been shown that parents actively handle and display objects to infants, and this begins when infants are as young as 3 months (Deák et al., 2014). These object manipulations have been closely linked with infant attention to the handled object (Burling & Yoshida, 2019; Deák et al., 2018; West & Iverson, 2017) and are associated with learning of the object names (Yu & Smith, 2013, 2017). For instance, Deák et al. (2014) found that infants aged 3–11 months show attention preferences for the handled object rather than the parent's face or hands in object play. Other developmental research also indicates that parent object handling facilitates children's object name learning on more structured tasks (Rader & Zukow-Goldring, 2010; West & Iverson, 2017). It has been speculated that the act of handling an object links hands and the object in a visually robust reference that can be processed as an entity by young children (Burling & Yoshida, 2019; Yu & Smith, 2017). In addition, previous work on joint attention (JA) suggests a mediational role of the parent's hands in establishing the relation between parental gaze and JA—the shared focus of attention with others on the same target—which has been linked to a number of developmental milestones (Amano et al., 2004; Baldwin, 1995; Striano et al., 2006). Yu and Smith (2017) also found that the coordination of parent gaze with the handled object is an alternative pathway to predict JA. In contrast with parent object looking, object handling is more readily detected and observed, especially in visually complex environments.

Unlike the visual referential cues discussed above, verbal cues (e.g., object labeling, phrasing) are seldomly reported alone and are often studied along with cues such as parent's gaze direction, object handling, or other actions in directing infant attention (Namy & Nolan, 2004; Tamis-LeMonda et al., 2013; Zukow-Goldring, 1996). For example, object labeling is often co-occurred with object handling and thought to promote the learning of object names (Gogate et al., 2000). Additionally, observations of parent–child interaction indicate that object labeling tends to be coordinated with the parent's gaze on the referent object (Tomasello & Farrar, 1986). Studies focusing on object labeling within JA episodes have also shown that multimodal cues are associated with the child's productive vocabulary development (Akhtar et al., 1991; Dunham et al., 1993). These findings demonstrate the importance of simultaneous multimodal input (e.g., labeling while handling the object) for enhancing a child's word learning and later achievements (Chang & Deák, 2019; Deák et al., 2000; Ruffman et al., 2020).

1.2 | Gaps in the literature and objectives of the present study

There are three significant gaps in our knowledge of parents' referential input. One of the gaps is created by the limited context in which referential cues have been studied. Although a diversity of parental referential cue use has been documented, these referential inputs have been studied extensively only within the JA framework, which assumes that social coordination is built primarily on episodes of attention sharing. However, JA is not the precursor of infant object looking but rather an

optimal attentional moment in which simultaneity occurs. Alternatively, infant object looking could be established through other socially coordinated pathways. For example, synchrony between object labeling and handling could also contribute to infant attention toward the target objects without parental gaze (Gogate et al., 2000; Matatyaho & Gogate, 2008; Rader & Zukow-Goldring, 2010). Burling and Yoshida (2019) found that parents actively held and showed the object to support early object viewing for infants aged from 5 to 17 months. Parents held the target object so that it occupied much of the infant's field of view, and this increased the likelihood that the child would maintain sustained attention on the target object. Though social coordination is typically demonstrated within joint attention episodes, infant attention can also be manipulated through alternative means.

Another gap in the literature is the lack of a definitive explanation for the attentional mechanism associated with parental referential input. Although the infant's object looking and associated learning outcomes have been attributed to referential cues (Brooks & Meltzoff, 2008; Morales et al., 1998; Ruffman et al., 2020; Tamis-LeMonda et al., 2014; Yu et al., 2019; Yu & Smith, 2012), we know relatively little about their actual relationship between parental referential cues and infant attention: who is leading whom? Does parental referential input predict infant's object looking, or do parents use referential input as responses? In other words, the temporal relationship between parental referential input and infants' object looking is not fully understood. Infant object looking can be explained from both exogenous and endogenous perspectives. Specifically, infants under 1 year of age are more likely to orient attention to exogenous features, such as color, shape, or other object properties (Colombo, 2001; Ruff & Rothbart, 1996). When parents and infants play together with the object toys, the parents are more likely to share attention and to engage in naming and/or mutual handling (Suarez-Rivera et al., 2019; Yu & Smith, 2017). These multimodal cues could influence infant object looking by making exogenous features of the targets more salient (Deák et al., 2018; Nagai & Rohlfing, 2009; Wass et al., 2018). Of course, infants are not passive referent receivers during social coordination with others. During the second year of life, children progressively develop endogenous attention systems and start to initiate and control their focus of attention, thus increasing active object explorations as well as parents' attention and actions (Burling & Yoshida, 2019; Colombo, 2001; Kannass & Oakes, 2008; Ruff & Rothbart, 2001). The emergence of social intention motivates children to play a more active role in social learning and results in different developmental patterns for initiating and responding to JA (Morales et al., 2005; Mundy et al., 2007). Therefore, a developmental change may occur in the social coordination between parental referential input and infant object looking over time. Referential cues may attract the child's attention toward the target object, or the child's manifest attention to the target may attract the parent to use referential cues as responses.

Furthermore, there is a lack of understanding of the relative strength of individual referential cues and the combined strength of multimodal input in relation to infant object looking. Each referential cue may make a unique contribution to directing infant attention, and multimodal cues might have an additive effect as well. One assumption is that labeling the handled object may guide infant attention to the object more efficiently than each cue alone. Alternatively, since the multimodal input provides a greater variety of referential cues, the child has more chances to find the most pertinent aspects of the combined cue. Either way, the assumption of "more is better" can explain the increasing effectiveness of multimodal cues. The present study aims to evaluate three major referential cues systematically during an early parent–child interaction and to document how each referential cue and its combinations may predict an infant's sustained attention to target objects.

The present study has three specific aims: (1) to describe parents' usage of the three primary referential inputs—object looking, labeling, and handling—during parent–child object play; (2) to examine the temporal relationship between each referential input and the infant's object looking; and (3) to

investigate the relative strength of each referential cue, alone and in combination with other cues, in predicting the infant's sustained attention to target objects.

Three hypotheses correspond to the above aims. Hypothesis 1: Parents will use multiple referential cues actively in dynamic patterns during the object play with the infant. Hypothesis 2: An age effect will be observed in the temporal order of referential cues and infant's object looking during the second year of life (i.e., between 12 and 18 months, cf. Burling & Yoshida, 2019; Mundy et al., 2007). With younger children, parental referential cues are expected to occur in advance of social coordination in order to help direct the infant's attention to the target objects; subsequently, when children gradually begin to lead the play, the parent's referential cues are more likely to be presented in response to the child's object looking. Hypothesis 3: When individual referential cues are aligned (e.g., parent object looking with relevant labels), the multisensory input will be more effective in maintaining an infant's attention on the target objects than will individual referential cues (e.g., object looking alone) or the sum of individual effects (e.g., looking alone + labeling object alone).

2 | METHOD

2.1 | Participants

The final sample consisted of 43 parent–infant dyads with the infants and toddlers aged from 5.0 to 18.0 months ($M = 10.91$, $SD = 3.98$; 20 males). The age range followed a relatively uniform distribution proved by the Kolmogorov–Smirnov test, $D = 0.19$, $p = 0.099$ (see also the age distribution in Figure 1). Twelve additional dyads were recruited but not included for the current analyses due to incomplete data collection associated with infant fussiness, technical failure, or inadequate recording quality. A sample of 43 dyads was selected according to the size of the effects in previous observational studies using micro-level behavioral approaches (e.g., Deák et al., 2018; Suarez-Rivera et al., 2019; Wass et al., 2018; Yu et al., 2019; Yu & Smith, 2017). The micro-level behavioral approach focuses

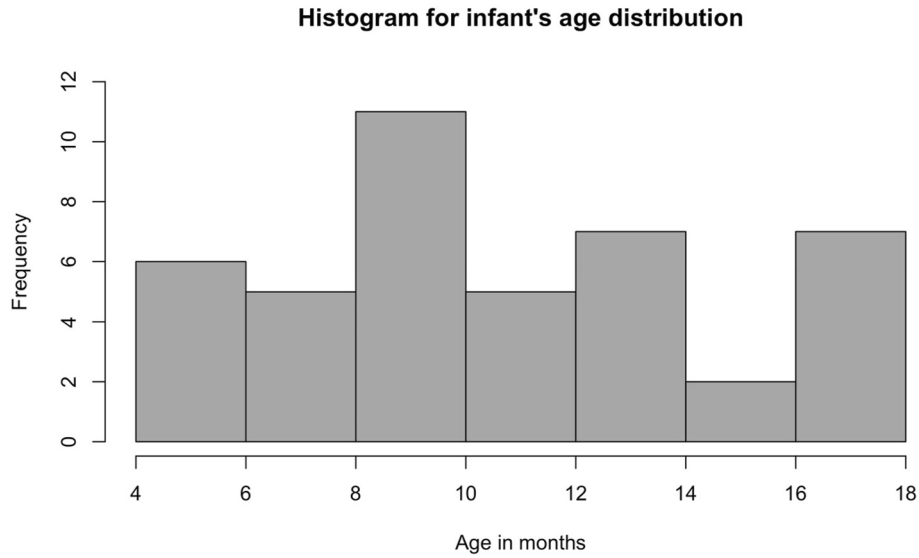

FIGURE 1 A histogram for infant's age distribution

on frame-by-frame gaze and behavioral annotations; hence, around ten-thousand data points per dyad were clustered and included in the following analyses.

Each child was full-term and typically developing with no known cognitive or developmental disorders. All the dyads were recruited from the Greater Houston area with comparable socioeconomic backgrounds that met the following criteria: (1) annual household income above $47,000 (i.e., the median annual household income in Houston; US Census, 2017); and (2) at least one of the parents with a bachelor's or higher degree. The sample of dyads was broadly representative of the ethnicity in the community: non-Hispanic Caucasian (33%), Hispanic (30%), African–American (9%), Asian (12%), bi-racial (9%), and no response (7%). To account for the linguistic diversity in the local community, we categorized the dyads as bilingual or monolingual according to their home language usage. When parents spent over 20% of the time using languages other than English at home, we categorized the child as bilingual. Considering the potential impact of language status on parental referential input (see the variability in parental input in Sun et al., 2022), we treated the child's language status as a covariate in the following analyses.

The present study was conducted according to guidelines laid down in the Declaration of Helsinki and its later amendments. Upon arrival, the experimenter explained the study procedures to parents and obtained written informed consent forms before any assessment or data collection. As a token of appreciation, all the participating dyads were provided a small gift bundle, including a grocery card, a museum pass, a baby t-shirt, and a stuffed animal. All procedures involving human subjects in the present study were approved by the Institutional Review Board at the University of Houston, where the project took place.

2.2 | Procedures

Parent–infant dyads completed a 5-min-20-s object play session in the lab. During the play session, the parent and infant sat across a $60 \times 60 \times 40$ cm table, which was used as a surface for interacting jointly with the objects. Video recordings were started prior to the participants entering the room to minimize distraction from the presence of the camera. Parents were provided with a container of toy objects in advance and told that they would be asked to play with the infant when a word theme was provided via an audio recorder. The parent participants were encouraged to use any of the toys and play as they would typically do at home, yet to incorporate a specified target word (i.e., bunny, eat, cookie, car, put, drink, open, bear) into the play session according to the audio cue. Parent–infant dyads played in eight 40-s-long trials, oriented around 8 target words. Pre-recorded audio instructions were played to direct parents as to the order and duration of each trial. The camera recording lasted around 20 min, including set-up and calibration of the recording equipment prior to the play session.

2.3 | Measures

Watec (WAT-230A) miniature color cameras with supplementary eye trackers were used to record the parent–infant interactive play session. Both the parent and infant would wear a head-mounted camera to record their scenes during the play session. The head-mounted camera provides dynamic visual information from a first-person perspective (e.g., Pereira et al., 2010; Smith et al., 2011; Yoshida & Smith, 2008; Yu & Smith, 2012; also see a review from Smith et al., 2015). In addition to the head-mounted camera system, an eye tracker was used to specify the focus of attention (see Figure 2c,d; Burling & Yoshida, 2019; Sun et al., 2022; Yoshida et al., 2020). Correspondence between images

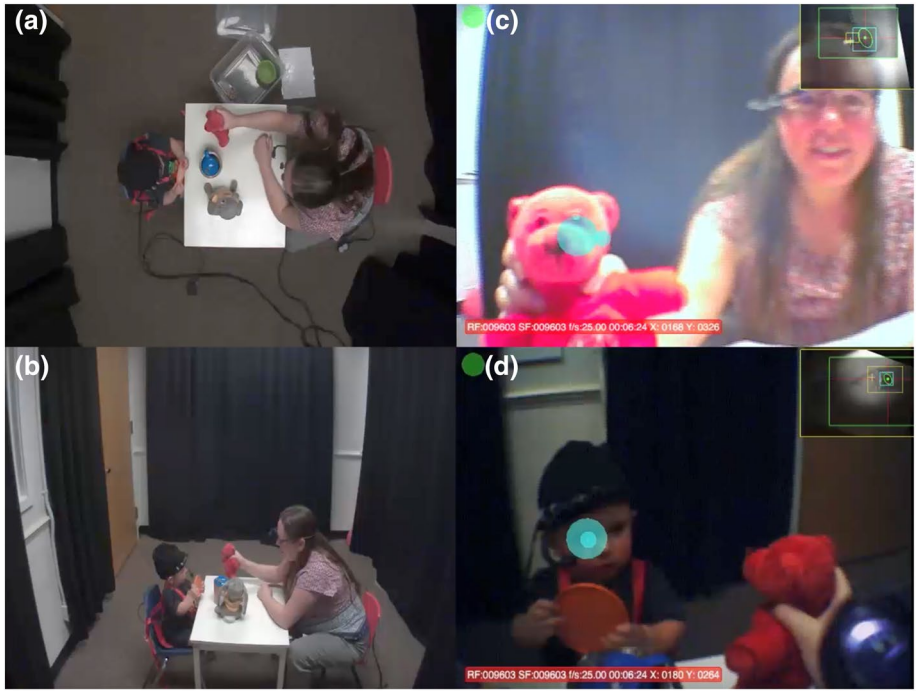

FIGURE 2 The recording structure of a parent–infant play session. (a) Live view; (b) ceiling view; (c) the child's view from the head-mounted camera with eye-tracking; (d) the parent's view from the head-mounted camera with eye-tracking

from the head-mounted camera and eye tracker was achieved using a manual calibration procedure that utilized a 60 × 40 cm board with nine spatially distributed stickers. Before and after the play session, both the parent and infant completed the calibration procedure twice by following the research assistant's pointing to each sticker on the calibration board. To help the infant shift attention on each calibration point, the research assistant would also use an attractive rattle to direct the infant's attention on the board. Subsequent video processing was not undertaken unless a minimum correlation of 0.9 between the camera and eye tracker images was obtained during calibration by Yarbus.

Two additional digital video cameras were mounted on the wall and the ceiling to capture an overall view of the scene in which the play session took place (see Figure 2a,b). Audio recordings were also made. All the videos were recorded at a rate of 33 milliseconds (ms) per frame and synchronized by Adobe Premiere. On average, each parent–infant dyad had 9886 frames (SD = 425) of data for analyses. The inaccessible frames included eye blinks and interruptions due to camera adjustment.

2.4 | Behavioral annotation

We annotated and analyzed each dyad's behaviors during the 5-min-20-s play session only. The referential inputs of interest (i.e., object looking, labeling, and handling) and attentional behaviors were observed and annotated by Datavyu coding software (Datavyu Team, 2014). Two well-trained coders, blind with respect to the experimental condition, annotated the following behavioral variables for each parent–infant dyad: parent's look, object labeling, object handling, and infant's look. After annotating

each referential cue separately, we time-stamped all the annotated behaviors following the timeline of the play session (see an example in Figure 3).

Reliabilities were measured by randomly selecting 25% of the frames for each dyad and checking inter-rater coding agreement for each annotated variable (see Table 1). For instance, the inter-rater reliability of the child's gaze was 88.7% (*SD* = 4.0%, ranging from 82.9% to 94.9%) as assessed by Cohen's kappa of 0.80 (*SD* = 0.07), indicated strong agreement among raters (Cohen, 1968; McHugh, 2012). Additionally, the inter-rater reliability also falls into the reliability range obtained in other eye-tracking studies (e.g., 84% for Yoshida et al., 2020; 82%–95% in Yu & Smith, 2017; 83% in Chang et al., 2016).

2.4.1 | Gaze behaviors, gaze behaviors (i.e., parent and infant object looking)

Previous studies have shown that the most common images captured in person-centered viewing are target objects, individual's hands, partner's hands, and partner's face (e.g., Burling & Yoshida, 2019; Yu & Smith, 2013). Thus, the present study primarily analyzed the gaze data directed to these four regions of interest (ROIs). Specifically, we preliminary focused on the object-looking instances. Both the parent's and the infant's gaze were annotated frame-by-frame according to whether the fixation point was located at one of the four ROIs in the head-mounted camera view.

2.4.2 | Parent object labeling

The parental speech was annotated at the utterance level, with a new utterance defined as an utterance beginning after 400 ms of silence (Pereira et al., 2013; Suanda et al., 2016; Sun et al., 2022; Yu & Smith, 2012). An utterance includes all meaningful phrases, words, and word-like sounds. Phrases

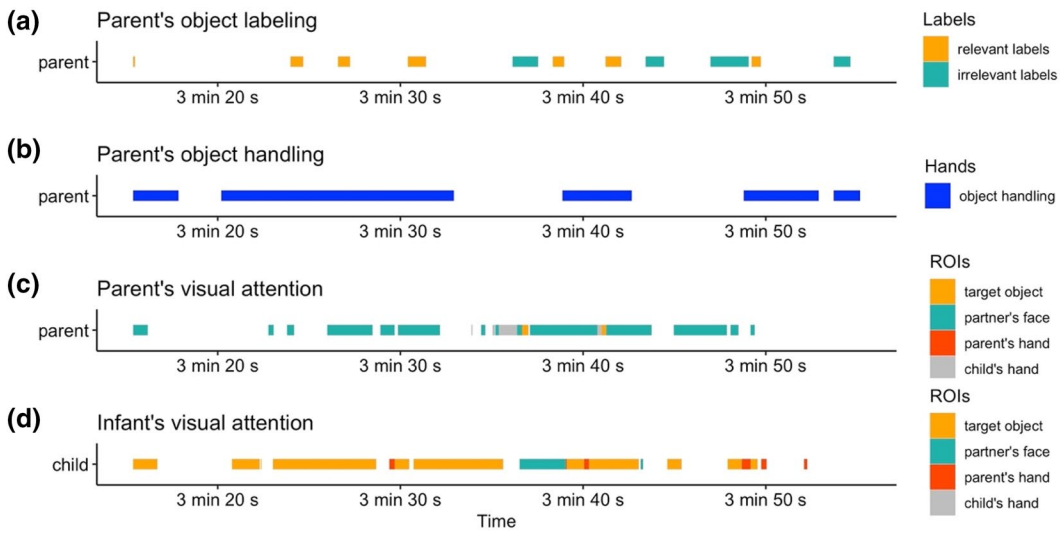

FIGURE 3 The multimodal behavioral annotation of a video clip of a parent–infant play episode. (a) Parent object labeling; parental phrases were categorized into phrases with relevant or irrelevant labels given the object toys they used in the same moment; (b) parent object handling; (c) the parent's gaze behaviors; (d) the infant's gaze behaviors; both the parent's and infant's gaze behaviors were annotated according to the four ROIs

TABLE 1 Reliabilities of target annotated behaviors

	Inter-rater reliability			Cohen's Kappa	
Targeted behaviors	**Mean (%)**	***SD* (%)**	**Range (%)**	**Mean**	***SD***
Infant's gaze patterns	88.7	4.0	82.9–94.9	0.80	0.07
Parent's gaze patterns	93.8	4.9	89.1–99.8	0.83	0.14
Parent object labeling	96.3	6.7	80–100	0.86	0.11
Parent object handling	96.2	3.6	85.2–100	0.95	0.04

were further categorized as relevant labels that contained the target words (e.g., "This is bunny!", "Do you like this bunny?") or irrelevant labels (e.g., "That's so fun!", "Do you like this?"). We ascertained the degree to which labeling overlapped with the child's attention toward the referent items, that is, the correspondence between parent's object labeling and infant object looking.

2.4.3 | Parent object handling

The parent's hand actions were coded frame-by-frame from both the wall and ceiling cameras. We annotated each parent's hand usage separately and included all the episodes of object handling, including right hand alone, left hand alone, and both hands together. Object handling was counted only after the parent started touching the target toy.

2.5 | Analytic approach

For testing the first hypothesis, that is, the prediction that parents use multiple referential cues in dynamic patterns during object play with the infant, we presented the distributions of time devoted to each referential cue and to each instance of mutual referential cue usage as a proportion of the entire play session.

The second hypothesis concerns the temporal relationship between parental referential cue use and infant's object looking: do changes in parent's referential cue use (e.g., parent's object looking) tend to occur before or after changes in infant's object looking? We expected a developmental shift such that parental referential cues occur in advance of social coordination and help direct infant attention to the target object during the second year of life, whereas the referential cues are more likely to be responses to the infant's object looking when the infant is older and more capable of actively exploring the outside world. We applied time-lagged cross-correlation analyses to each of the referential cues and infant's object looking. Cross-correlation analysis is widely used for tracking multiple sets of time-series data. It assesses behavioral correspondences without requiring any assumptions about a specific time frame in which behaviors should happen (e.g., Podobnik & Stanley, 2008). The cross-correlations between all lagged pairs of data within ± 10-s windows were calculated. General mixed-effect models were used, and all the lagged cross-correlations were clustered by parent–infant dyads to examine whether the temporal relationship between a parent's referential cue use and the infant's object looking changes with age.

The third hypothesis concerns the effects of multimodal input on an infant's attention. We used generalized mixed-effect models to examine how well the infant's sustained attention could be predicted on the basis of each referential cue and different combinations of multimodal cues. All

the data organization and analyses were conducted in the R environment (version 4.1.0; RStudio Team, 2021). In specific, lmer and glmer functions of the R package lme4 (Ver 1.1-27.1, Bates et al., 2015) were used for estimating the present mixed-effect models, and ggeffects was used for computing the estimated predicted probabilities (Ver 1.1.1, Lüdecke, 2018).

3 | RESULTS

3.1 | The distribution of parental referential cues

There were three main referential cues in the present study: (1) parent object looking, (2) parent object labeling, and (3) parent object handling. Table 2 summarizes the usage of each kind of cue during the play sessions. The proportion of time in which the parents used at least one of the three referential cues was 90.9%, which indicates extensive scaffolding. The parents spent 81.2% (*SD* = 0.13) of the time handling target objects and 46.1% (*SD*: 0.17) of the time speaking. In particular, the parents spent 26.4% (*SD* = 0.12) of the time using labels relevant to target objects and 20.2% (*SD* = 0.12) of the time using irrelevant labels. Object looking by the parents accounted for 19.3% (*SD* = 0.11) of the play session.

We then identified how often the parents used these referential cues alone or together (see the proportional distribution of mutual referential cue uses in Figure 4). Object handling alone accounted for the majority of time in the play session (34.7%, *SD* = 0.15), followed by object handling with relevant labels (17.5%, *SD* = 0.09) and object handling with irrelevant labels (13.1%, *SD* = 0.09). The combination of the three referential cues (looking at the handled object with labels) accounted for 7.6% of the time, which could be decomposed into looking at the handled object with relevant labels (4.6%, *SD* = 0.03) and looking at the handled object with irrelevant labels (3.1%, *SD* = 0.03).

Figure 5 illustrates the distribution of each referential cue and its relation to the other two. Figure 5a shows that 82.9% of parent object looking was accompanied by object handling, whereas 46.1% occurred with object labeling, regardless of whether the labels were relevant or irrelevant to the target objects. In addition, object handling alone accounted for 42.9% of the overall object handling, whereas 46.8% of object handling accompanied object labeling (27.1% with relevant labels), and

TABLE 2 Descriptive statistics of parental referential input

Parental referential cues	Mean duration (s)	*SD* duration (s)	Median duration (s)	Range in duration (s)	Mean prop (%)	*SD* prop (%)	Median prop (%)	Range prop (%)
Parent object handling	264.93	44.40	265.19	166.02–332.67	81.22	13.34	81.26	52.47–100
Parental phrases	150.82	58.21	143.81	39.27–265.29	46.08	17.21	43.83	12.01–79.96
(1) parental phrases with relevant labels	86.14	40.00	82.14	11.09–172.19	26.36	12.09	26.53	3.34–54.20
(2) parental phrases with irrelevant labels	66.22	38.25	60.44	9.11–185.16	20.19	11.54	18.24	2.75–55.81
Parent object looking	63.09	37.74	57.42	8.65–161.53	19.30	11.45	16.94	2.61–48.36

Note: Parental referential inputs have been summarized with respect to parent object looking, parent object labeling, and parent handling. To be noted, parental phrases have been categorized into two types according to the contained target words on the object toys: (1) parental phrases with irrelevant labels and (2) parental phrases with relevant labels.

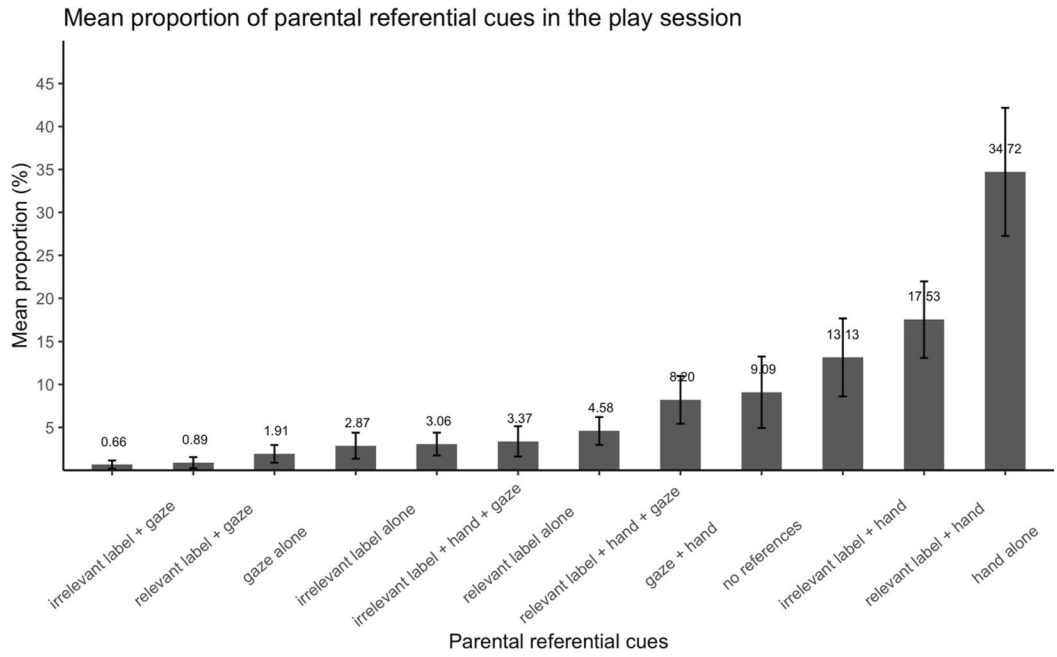

FIGURE 4 Mean proportion of time of mutual referential cues in the play session. A bar plot of the mean proportion of time of mutual referential cues, including three major referential cues (i.e., object looking alone (relevant/irrelevant) labeling alone, and object handling alone) and their combinations (e.g., parent looking at the handled object with relevant labels)

19.7% occurred when the parent was looking at the objects (see Figure 5b). Figure 5c,d represents the distributions of parental phrases with relevant labels and irrelevant labels separately, and the overlapping proportions show the contingency of the co-occurrence of other referential cues, regardless of the phrase content (relevant or irrelevant). Of all the moments the parent talked to the child, 82.6% of object labeling accompanied object handling, and 19.0% accompanied object looking (see also the proportion of multimodal input in each referential cue in Table A1 in Appendix).

3.2 | The distribution of infant object looking

Among the four target ROIs, infants actively looked at the ROI containing target objects 72.61% ($SD = 12.53$) of the time in the play session. The duration of looking time ranged from 10 to 23,990 ms. Brief looks (<3 s) dominated infant object looking (90.7%) and sustained attention (>3 s) accounted for only 9.3% of object looking.

3.3 | The role of referential cues in guiding infant object looking

We first applied time-lagged cross-correlation analyses to determine the temporal relationship between the parent's referential cues and infant object looking (see Figure 6). The cross-correlation tests using portmanteau statistics demonstrated that the time series of parent object looking, object labeling, and object handling all correlated significantly, at $\alpha = 0.05$, with infant object looking, and the correlation

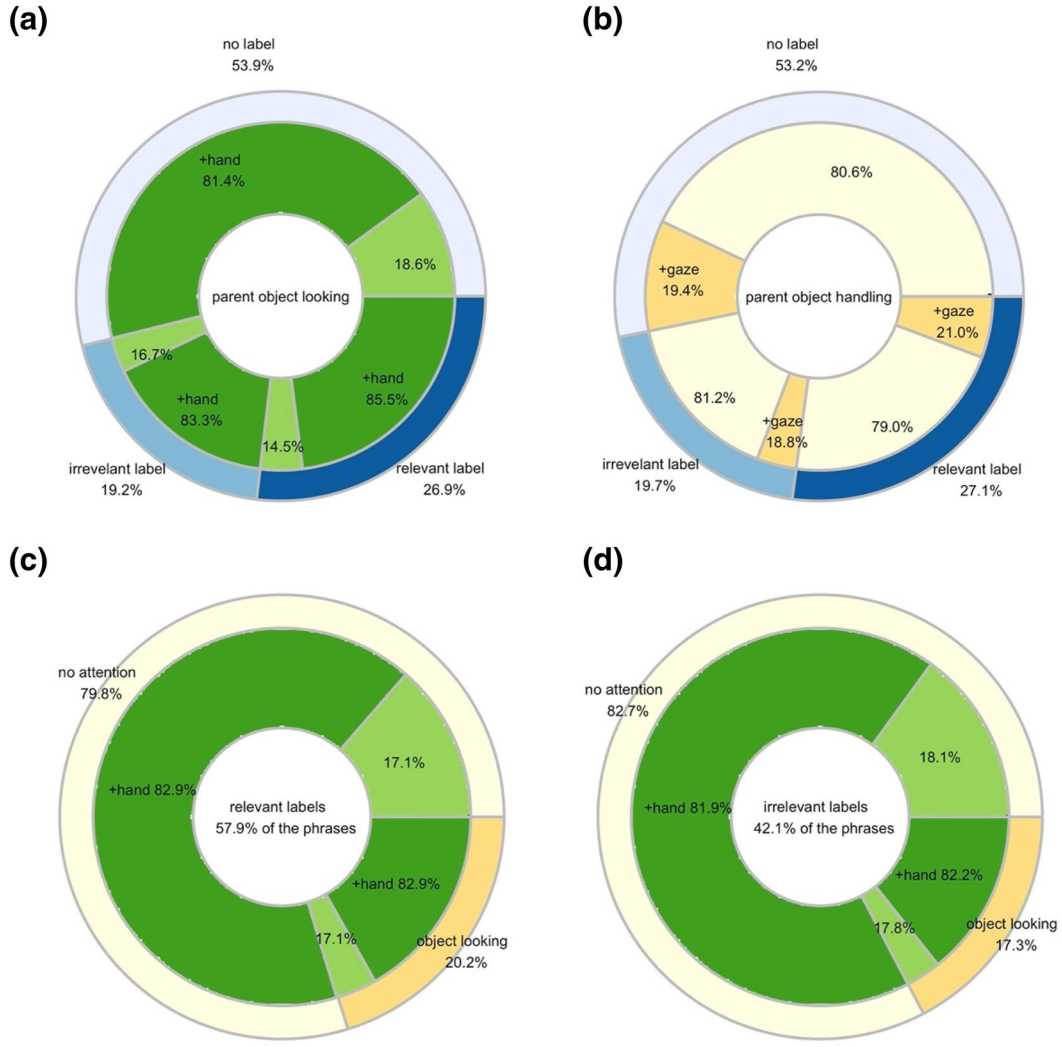

FIGURE 5 Proportion of multimodal input in each parental referential cue. The double-ring represents the overlapping among the three parental referential usages: the inner ring refers to the conditional probabilities under the other two combined cues (light yellow: no object looking; yellow: object looking; light green: no handling; green: object handling; light blue: no phrases; blue: parental phrases with irrelevant labels; dark blue: parental phrases with relevant labels). (a) Parent object looking; (b) parent object handling; (c) parental phrases with relevant labels; (d) parental phrases with irrelevant labels

reached a maximum when the lag was set to ± 10 s (see the Haugh–Box standard and robust methods in Dalla et al., 2020). In the following cross-correlation analyses, two key components—directionality and lag—indicate the temporal relationship. At the parent–infant lag time of 0, the cross-correlations showed that all three of the parental referential cues were positively correlated with infant object looking. In other words, an infant was more likely to maintain a longer look toward target objects when his or her look was accompanied by (1) longer object looking by the parent, (2) longer parental utterances with labels, and (3) longer object handling by the parent.

Lag, on the other hand, represents the interval between the parental referential cue and infant object looking. The highest peak of correlation coefficients indicates the overall direction of signal

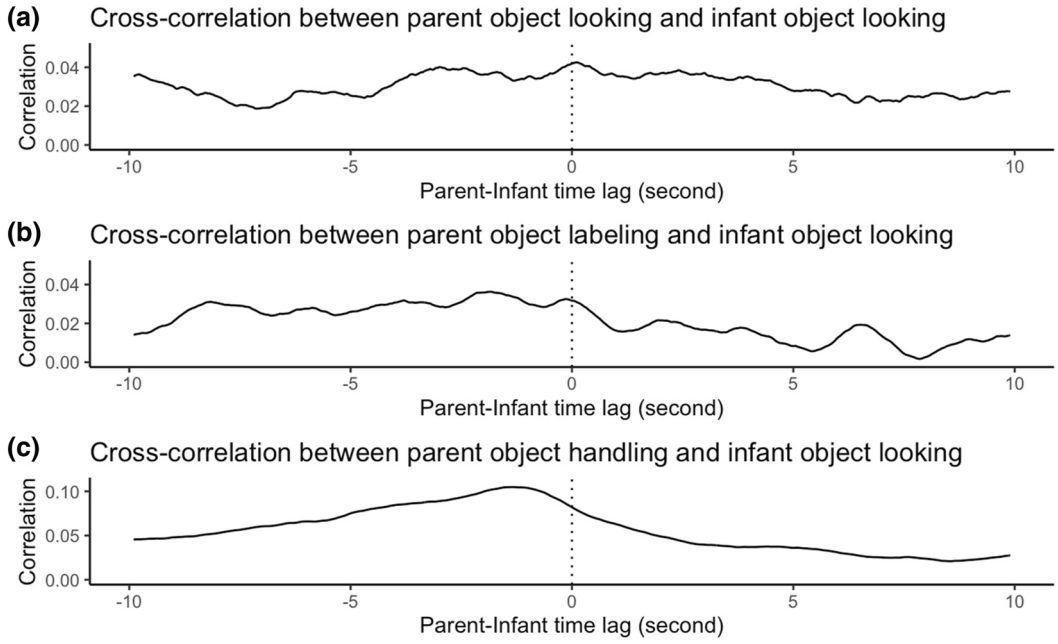

FIGURE 6 Cross-correlations between each of the parental referential cues and infant object looking. (a) Parent object looking and infant object looking; (b) parent object labeling and infant object looking; (c) parent object handling and infant object looking. The cross-correlation plot is asymmetric around time lag = 0 with higher correlation values at positive lags than negative lags, suggesting that parent object looking tend to predict the infant's subsequent attention to the objects rather than vice versa; The cross-correlations for object labeling and object handling with infant object looking are also asymmetric but with higher correlation values at negative lags than positive lags, indicating that both changes in parent object labeling and object handling tend to occur in response to changes in infant object looking rather than vice versa

change within the lagged window. For example, the cross-correlation between parent object looking and infant object looking had the highest peak of $r = 0.04$ when lag = 0.10 s, $t = 27.75$, $p < 0.001$. Finding that the peak occurred with a positive lag indicates that changes in parent object looking preceded changes in infant object looking. In contrast, the cross-correlation between parent object labeling and infant object looking had the highest peak of $r = 0.04$ when lag = −1.85 s, $t = 23.66$, $p < 0.001$. Similarly, the cross-correlation between parent object handling and infant object looking had the highest peak of $r = 0.11$ when lag = −1.35 s, $t = 68.34$, $p < 0.001$.

The highest peaks with negative lags in both asymmetric distributions reveal that changes in parent object labeling and object handling alike tend to occur in response to changes in infant object looking, suggesting the importance of parents' timely responses to infant attention of interests. Although the correlation between parent object looking and infant object looking was significant, the correlation coefficient was relatively small (cf. $r = 0.2$ in Wass et al., 2018). The relatively weak relationship may be related to the difficulty of generating attention synchrony between two agents, requiring the direction of parent's gaze has to be identified and followed by the child, and such gaze following may be still under development (e.g., Deák et al., 2000, 2014). In contrast, parent object handling was more tightly related to infant object looking (Burling & Yoshida, 2019; Chang et al., 2016; Yu & Smith, 2013, 2017). The more robust synchrony between parent object handling and infant object looking can be attributed to the perceptual saliency in the visual scenes, that infants are sensitive to

changes in the motion contrast created by parent's hand actions (Deák et al., 2014; Nagai & Rohlfing, 2009; Yu & Smith, 2013).

3.4 | Temporal relationship between parent object looking and infant object looking

Given the asymmetric distributions of cross-correlations between parental referential cues and infant object looking (see also Figure 6), it is essential to test whether parents lead or follow the infant's attention. General mixed-effects models were used to evaluate changes in the temporal relationship between the parent's referential cues and infant object looking within the 10-s window. Consideration of age is essential. Given the dramatic changes in the role of parental scaffolding in directing infant object looking in the first 2 years of life (Burling & Yoshida, 2019; Mundy et al., 2007), we expected to see a main effect of age on the temporal order between each referential cue and infant object looking. For this set of analyses, the correlations between parent object looking and all lagged pairs of infant's look epochs within the 10-s window were dichotomized into *positive* epochs—instances in which "the infant's look leads the parent's look" (i.e., parent post infant; parent at time x and infant at time $x - t$) and *negative* epochs—instances in which "the parent's look leads the infant's look" (i.e., parent pre infant; parent at time x and infant at time $x + t$). Considering the variability in referential cue usage within and between dyads, all the correlation coefficients between parent and infant object looking were nested by parent–infant dyads and by the time lag (±10 s). In sum, temporal order, infant's age, and the interaction (temporal order × age) were the predictors, and language status was included as a covariate in the models.

The general mixed-effects model revealed that temporal order, $\beta = 0.0056$, $p < 0.001$, and the interaction between temporal order and age, $\beta = -0.0004$, $p = 0.002$, both significantly predicted the correlation between parent and infant object looking with $R^2 = 0.57$ (see also Figure 7 and model 1 in the Appendix). The effect for temporal order indicates that the correlation for "parent object looking leads infant object looking" was significantly stronger than for "infant object looking leads parent object looking," and the interaction reveals that this temporal order tends to reverse with age. After the age of 15.3 months, infant object looking was more likely to predict the parent's subsequent object looking rather than vice versa.

3.5 | Temporal relationship between parent object labeling and infant object looking

An identical general mixed-effects model was used to examine the temporal relationship between parent object labeling and infant object looking. The model explained 59% of variance and revealed a significant main effect of temporal order, $\beta = -0.0138$, $p = 0.021$, which indicates that "infant object looking leads the parent object labeling" was more prevalent than "parent object labeling leads infant object looking," and the difference significantly increased with age (see Model 2 in the Appendix). In other words, it was more common for parents to respond to an infant's object looking with relevant labels on the infant's attended object rather than for an infant to shift attention toward the target object after the parent named the object.

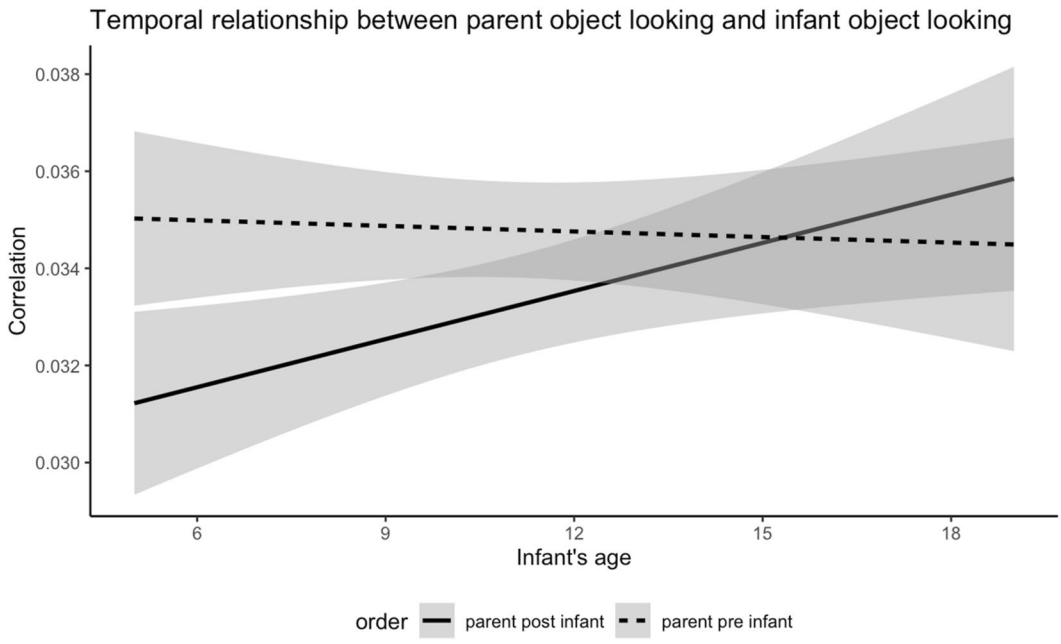

FIGURE 7 Temporal relationship between parent object looking and infant object looking. Changes in the temporal relationship between parent object looking and infant object looking with age. Prior to the developmental shift at the age of 15.3 months, the correlation on "parent leads infant" (parent pre infant) was stronger than "infant leads parent" (parent post infant), and the tendency reverses over developmental time

3.6 | Temporal relationship between parent object handling and infant object looking

Similar to the analysis for the correlation between parent object looking and infant object looking, the general mixed-effect model for the correlation between parent object handling and infant object looking explained 86% of variance and revealed a significant fixed effect of temporal order, $\beta = -0.0260$, $p = 0.029$ (see Model 3 in the Appendix). In other words, the correlation for "infant object looking leads parent object handling" was significantly stronger than for "parent object handling leads infant object looking," and the difference significantly increased with age. Parents were more likely to handle the object that the infant was already attending to rather than leading the infant to shift attention to the handled object.

3.7 | The strengths of referential cues on an infant's sustained attention to objects

To investigate the relative strength of different referential cues on an infant's sustained attention (SA), the three referential cues and their combinations were treated as predictors of an infant's SA on the target objects. Specifically, the individual and combined cues were classified into 12 conditions: (1) no referential cues, (2) parent object looking alone, (3) parent object handling alone, (4) relevant labeling alone, (5) irrelevant labeling alone, (6) parent looking at the handled object, (7) parent object looking with relevant labels, (8) parent object looking with irrelevant labels, (9) object handling with relevant labels, (10) object handling with irrelevant labels, (11) parent looking at the handled object

with relevant labels, and (12) parent looking at the handled object with irrelevant labels. We examined the impacts of individual and multimodal cues in directing infant attention and then compared the impacts among these 12 conditions.

A generalized mixed-effect model was selected to examine the impacts of the three targeted referential cues as well as the multimodal cues on the infant's SA on target objects during the play session. The SA measure consisted of instances in which the infant maintained his/her look at the target objects for more than 3000 ms (Campbell et al., 2014; Kannass & Oakes, 2008; Ruff & Lawson, 1990; Yu et al., 2019). Parent object looking, object labeling, object handling, and the interactions among the three variables were the predictors. SA was clustered by parent–infant dyad and served as the dependent variable. Two demographic factors (i.e., age and language status) were also included as covariates at the dyad level in the model.

The generalized mixed-effect model revealed that all the referential cues and their interactions were statistically significant, including parent object looking ($\beta = 0.17$, $p < 0.001$), relevant labeling ($\beta = 0.15$ $p < 0.001$), irrelevant labeling ($\beta = -0.54$, $p < 0.001$), object handling ($\beta = -0.06$, $p < 0.001$), the interaction of parent object handling × object looking ($\beta = 0.08$, $p < 0.01$), the interaction of parent object handling × relevant labeling ($\beta = 0.12$, $p < 0.001$), the interaction of parent object handling × irrelevant labeling ($\beta = 0.37$, $p < 0.001$), the interaction of parent object looking × relevant labeling ($\beta = 0.16$, $p < 0.001$), the interaction of parent object looking × irrelevant labeling ($\beta = 0.21$, $p < 0.001$), the interaction of parent object handling × parent object looking × relevant labeling ($\beta = -0.31$, $p < 0.001$), and the interaction of parent object handling × parent object looking × irrelevant labeling, $\beta = -0.13$, $p < 0.01$ (see Model 4 summary in the Appendix).

The effects of referential cues on SA were represented in terms of the predicted probabilities (see Figure 8 and Table 3) and were compared through a set of Tukey post hoc analyses. The predicted

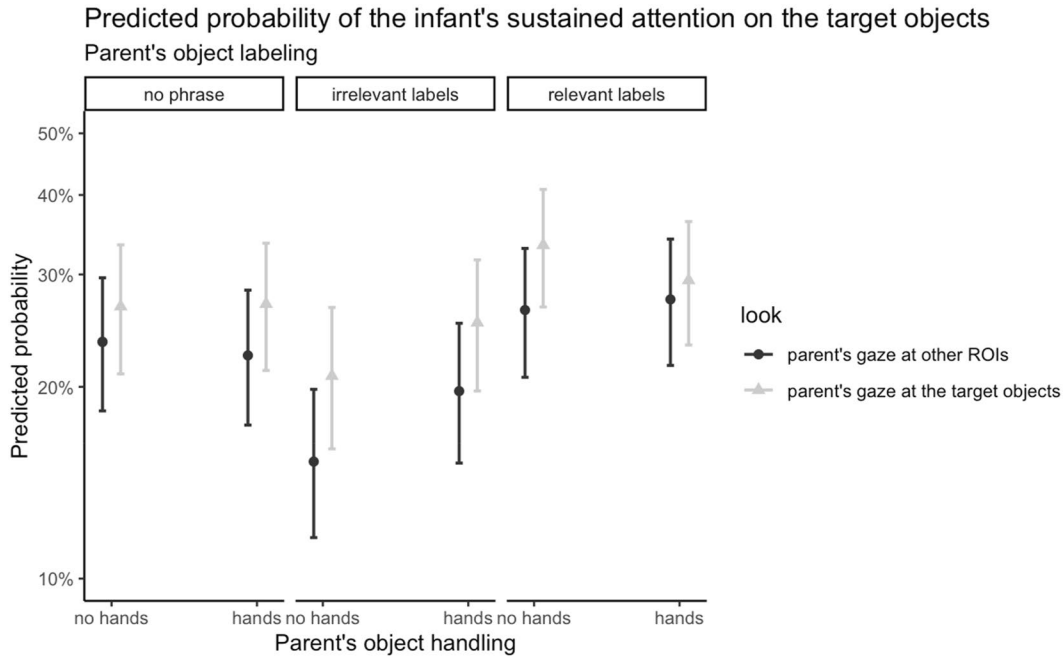

FIGURE 8 The strength of parental referential cues on infant's sustained attention on the target object. The predicted probability of the infant's sustained attention on the target objects with each referential cue and their combinations with 95% confidence intervals around the estimated means

TABLE 3 Predicted probabilities of infant sustained attention on objects

Type of cues	Predicted probability	Standard error	Lower CI	Higher CI
No referential cues	0.24	0.16	0.18	0.30
Parent object looking alone	0.27	0.16	0.21	0.33
Parent object handling alone	0.22	0.16	0.17	0.28
Relevant labeling alone	0.26	0.16	0.21	0.33
Irrelevant labeling alone	0.15	0.16	0.12	0.20
Parent looking at the handled object	0.27	0.16	0.21	0.34
Parent object looking with relevant labels	0.33	0.16	0.27	0.41
Parent object looking with irrelevant labels	0.21	0.17	0.16	0.27
Object handling with relevant labels	0.27	0.16	0.22	0.34
Object handling with irrelevant labels	0.20	0.16	0.15	0.25
Parent looking at the handled object with relevant labels	0.29	0.16	0.23	0.36
Parent looking at the handled object with irrelevant labels	0.25	0.16	0.20	0.32

probability of SA with no referential cues was 24%, which served as a baseline. In line with previous studies on joint attention, we found that parent object looking played a predominant role in the infant's SA under any referential circumstances. For instance, the predicted probability of SA was increased to 27% when the parent looked at the same target object and created a joint attention moment, $Z = 6.06$, $p < 0.01$. There were similar additive effects of parent object looking on the infant's SA when comparing (1) parent looking at the handled object versus parent object handling alone (additional 5%), $Z = 16.94$, $p < 0.01$, (2) parent looking at the target object with irrelevant labels versus irrelevant labeling alone (additional 6%), $Z = 7.87$, $p < 0.01$, (3) parent looking at the handled object with irrelevant labels versus object handling with irrelevant labels (additional 5%), $Z = 13.65$, $p < 0.01$, (4) parent looking at the target object with relevant labels versus relevant labeling alone (additional 7%), $Z = 8.28$, $p < 0.01$, and (5) parent looking at the handled object with relevant labels versus object handling with relevant labels (additional 2%), $Z = 5.18$, $p < 0.01$.

Furthermore, the type of object labeling (relevant vs. irrelevant) was crucial in determining the chances of an infant's SA on target objects. On the one hand, irrelevant labeling had a significantly negative impact on maintaining the child's SA on objects. For example, the predicted probability of SA was decreased to 15% when irrelevant labeling was added, $Z = -20.71$, $p < 0.01$. Similar subtractive effects of irrelevant labeling were found when comparing (1) parent object looking with irrelevant labels versus parent object looking alone (subtractive 6%), $Z = -6.66$, $p < 0.01$, (2) object handling with irrelevant labels versus object handling alone (subtractive 2%), $Z = -12.90$, $p < 0.01$, and (3) parent looking at the handled object with irrelevant labels versus parent looking at the handled object with no labeling (subtractive 2%), $Z = -3.75$, $p < 0.01$. On the other hand, relevant labeling was beneficial in maintaining infant SA on target objects. When the parent used the relevant labels only, the predicted probability of SA jumped to 26% and was increased relative to the absence of referential cues, $Z = 6.71$, $p < 0.01$. The additive impact of relevant labeling was also found in the following comparisons: (1) parent object looking with relevant labels versus parent object looking alone (additional 6%), $Z = 7.29$, $p < 0.01$; (2) object handling with relevant labels versus object handling alone (additional 5%), $Z = 24.42$, $p < 0.01$, and (3) parent looking at the handled object with relevant labels versus parent looking at the handled object (additional 2%), $Z = 5.73$, $p < 0.01$.

In addition, the impact of parent object handling on the infant's SA was dependent on the co-occurrence of the other two cues. The addition of object handling by the parent produced no change in SA relative to object looking alone, $Z = 0.43$, $p = 1.00$, or to relevant labeling alone, $Z = 2.47$, $p = 0.31$. Moreover, there was a significant decline from baseline in the predicted probability of SA when parent object handling occurred alone (subtractive 2%), $Z = -4.54$, $p < 0.01$. A similar difference was also found when parent looking at the handled object with relevant labels was compared with parent object looking with relevant labels (subtractive 4%), $Z = -4.74$, $p < 0.01$. The benefit of object handling was found only when the parent used irrelevant labeling. There were significant increases in the predicted probability of an infant's SA for (1) parent object handling with irrelevant labels versus irrelevant labeling alone (additional 5%), $Z = 12.18$, $p < 0.01$, and (2) parent looking at the handled object with irrelevant labels versus parent object looking with irrelevant labels (additional 4%), $Z = 5.31$, $p < 0.01$.

In sum, the most effective referential cue for predicting an infant's SA on the target objects was parent object looking with relevant labels. The predicted probability associated with this combination of cues reached 33%, followed by the 29% probability associated with parent looking at the handled object with relevant labels. It is important to note that parent object looking had a consistent additive effect on an infant's SA on target objects, which highlights the association between parent–infant object sharing and the consistency of infant object attention. Moreover, the impacts of parent object labeling varied with the type of labeling, and the contribution of parent object handling depended on an interaction between parent object looking and the type of object labeling.

4 | DISCUSSION

The present study confirms a significant relationship between parents' referential input and infants' early object viewing experiences. Parents not only initiate the social coordination, but they also adapt and accommodate their behaviors to the infants' needs. We found that parents constantly support the infants' visual exploration by providing diverse referential input as different sequences of action during the object play. Of the three referential inputs monitored in the present study, parent object handling was the most frequent cue use, and it often co-occurred with the other two. For example, parent object labeling was found to coincide frequently with object handling. The combination of object labeling and handling has been shown to help infants maintain SA on objects (Chang et al., 2016) and to optimize the formation of word-meaning linkages (Yu & Smith, 2012). In addition, the frequent co-occurrence of object viewing and labeling within a brief time window has been shown to benefit word learning (Pereira et al., 2013; Yu & Smith, 2012) and language development (also see the review by Tamis-LeMonda et al., 2014).

By examining the temporal relationship between parental use of referential cues and infant object looking—"who leads whom?"—the present study revealed a significant developmental shift in the temporal order of parent and infant attention toward the target object. Parent object looking tended to lead to the infant's subsequent attention toward the target objects, but this trend would be reversed with age. Around the age of 15 months, infant object looking was likely to lead parent object looking rather than vice versa. The developmental shift in the temporal order reflects variations in parents' referential input as children grow up (Adamson & Bakeman, 1984; Bornstein et al., 2008). This resembles changes in two types of socially coordinated attention as discussed in the JA literature, that is, responding joint attention (RJA) and initiating joint attention (IJA; Bakeman & Adamson, 1984; Saxon et al., 2000; see also Mundy et al., 2007 for a review). This literature suggests that infants as young as 9 months commonly follow parents' gaze to create coordinated attention (e.g., RJA; Mundy et al., 2007; Flom et al., 2004), and they gradually play more active roles in initiating self-control

attention (e.g., IJA) during the second year of life (Deák et al., 2014; Suarez-Rivera et al., 2019; Wass et al., 2018).

In addition, the socially mediated developmental shift—from "parent leads infant" to "infant leads parent"—in the present study, along with the JA literature, could be related to the development of endogenous attention and executive functioning skills through the first 2 years of life (Mundy et al., 2007; also see a review by Colombo & Cheatham, 2006). The dramatic development in the prefrontal cortex between 6 and 12 months of age facilitates the emergence of endogenous attention around 9 months of age (Holmboe et al., 2018; Moscovitch & Winocur, 2002; Oakes et al., 2002), and further contributes to the functional maturation of endogenous attention, which enables infants to initiate and control their attention and to better inhibit responses to distractors in the second year (Paterson et al., 2006; Raz & Saxe, 2020).

The developmental shift between parent and infant object looking in the middle of the second year corresponds to advances in young children's physical capacities with respect to motor competence and language skills. When children gradually become capable of reaching and manipulating objects and start to vocalize, they acquire more opportunities to initiate social interchanges and to maintain social coordination with others (Hilbrink et al., 2015; Rutter & Durkin, 1987; West & Iverson, 2017). As such, young children are not merely passive receivers of learning signals from the outside environment, but they also develop their visual exploration competencies and means of sharing their own interests and goals (Bakeman & Adamson, 1984; Burling & Yoshida, 2019; Chang et al., 2016).

Furthermore, the present finding adds to the extant evidence that the developmental shift in attention sharing is *also* supported by parents' adaptions and accommodations with respect to multimodal referential input (Bornstein et al., 2008; Iverson et al., 1999; Namy & Nolan, 2004; Yoshida et al., 2020). According to our hypothesis, the socially coordinated experience directed toward a target object is not built only upon the repetition of attention sharing. Before parents shift attention to a target object, their object labeling or handling can drive infants' interest and stabilize their attention toward the target objects. We expected the multimodal input presented in different combinations to nurture infant visual experience throughout the interactive play. Consistent with the recent evidence of multimodal input by parents (Deák et al., 2018; Suarez-Rivera et al., 2019), JA can be considered to be a proxy for a suite of multimodal referential input that is temporally linked to attention sharing. For example, the present findings revealed that 91.81% of JA moments consist of multimodal cues, while 8.19% involved mutual gaze sharing only (see Figure 9). Consistent with this finding, a growing number of studies suggest that infants learn effectively from contingently presented referential cues (Chang et al., 2016; Tamis-LeMonda et al., 2014; Yu & Smith, 2012). For example, Goldstin et al. (2010) documented that a parent can facilitate word-referent mapping during novel word learning by responding with a contingent label when the infant babbles and attends to a handled item. Coordination between multimodal input and infant object looking helps infants associate relevant labels with patterns of sensorimotor experience in a timely manner and creates opportunities for word acquisition (Chang et al., 2016; Ruffman et al., 2020). Therefore, developmentally meaningful responsiveness requires parents to provide prompt and appropriate referential input in response to infant object looking.

To examine the specific strength of each of the three major referential cues and their combinations, the present study focused on infant SA on objects, which was defined as the infant maintaining stable attention on the target object for over 3 s. Different from other object looking behaviors (i.e., saccade, fixation), SA has been shown to be a strong predictor of cognitive skills (Kopp & Vaughn, 1982; Ruff & Lawson, 1990; Sigman et al., 1987) and word learning outcomes (Kannass & Oakes, 2008; Ruffman et al., 2020; Welsh et al., 2010; Yu et al., 2019). Of all types of referential cues, we found that parent object looking is the only cue that consistently promotes infant SA under any circumstances. One

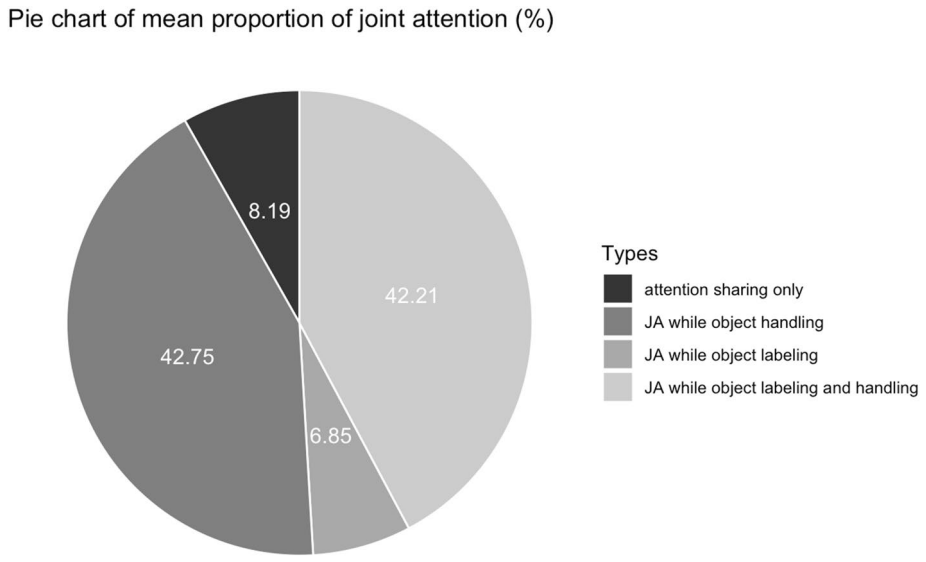

FIGURE 9 The distribution of joint attention with referential input. Proportion of joint attention between the parent and the infant to the same target object that was classified into four types of joint attention: (1) attention sharing alone, (2) attention sharing with parent object handling, (3) attention sharing with parent object labeling, and (4) attention sharing with parent object handling and labeling

may argue that parent gaze alone is difficult for children to notice clearly and to keep track of among visually complex and cluttered scenes. The consistent and robust benefit of parent object looking might be due to *its correlation with other referential cues*. Thus, infants could quickly shift attention toward target objects through multiple pathways that accompany parent object looking. Social gaze has long been documented as a strong goal-directed signal of sharing interests, and the present results reveal the additive effect of combined referential cues as an explanation of the powerful role of parent object looking in directing infant attention in an interactive context. In other words, referential cues may not be used or learned in isolation but instead by being coupled with other more easily observed and obvious cues and actions. Through the process of receiving redundant cues in dynamic ways, infants may gradually become more sensitive to, and pay more attention to, the gaze cue by itself.

There are a few limitations to the conclusions that we can draw from the present findings, and those limitations point to some potential directions for further work. First, the impact of object labeling and object handling may depend upon their contribution to the social synchrony. The present study tabulated moments of parent object handling irrespective of the underlying meaning. For example, the direction of hand movement may convey vague social meanings: the parent can move the object closer to the infant to denote turn-taking or demonstrate the object's functions along with gestures. Also, the present study did not fully account for all the deictic gestures on attention orientation. In addition to showing or giving, which requires parents to touch the objects by hand, pointing is also strongly associated with attention shifting and language learning (Leung & Rheingold, 1981; Morissette et al., 1995; Rader & Zukow-Goldring, 2012; Rohlfing et al., 2012), but it was excluded in the present study. Future studies should take pointing and other sophisticated gestures (e.g., representational or symbolic gestures) and their underlying semantic meanings into consideration.

Another limitation is that the temporal relations among the combined referential cues and infant attention could be more complicated than we described in the present work. The temporal order of multimodal cues may contribute to different sequential patterns in supporting visual selection from

complex scenes. Future studies can address this possibility by considering the impacts of referential patterns on infant free viewing as long time series and then link infant attention ability to the subsequent learning of relevant labels. Machine learning algorithms can possibly be used to explore the possible attentional mechanism that elicits referential input from the outside world (Gottlib et al., 2013; Messinger et al., 2010).

In conclusion, our work reinforces an important idea about the early stages of child development, viz., that language learning emerges in a social context and contributes to building a perceptual foundation for object exploration. In turn, this generates a positive cascading effect on subsequent word learning. The present findings suggest that each kind of referential cue and its various combinations help to orient infant attention toward the referent items. The multimodal referential input infrequently occurs simultaneously, but it nonetheless is capable of effectively directing infant object looking. The overlap among redundant referential cues may provide enriched opportunities for infants to become gradually more sensitive to each kind of referential input respectively. The present work is a first step in characterizing the strength of different referential cues in relation to an infant's visual experiences. It is essential to continue extending this line of work to reveal the attention mechanisms that underlie early social scaffolding in the context of various background factors (e.g., age, ethnicity, culture, socioeconomic status, language), atypical learning experiences (e.g., learning difficulties or disabilities), and diverse educational circumstances (e.g., daycares or schools), and to demonstrate the implications of early scaffolding for effective learning.

ACKNOWLEDGMENTS

Parts of the data for the present work were presented at the SRCD Biennial Meeting (2021) and the Annual Meeting of the Southwest Cognition and Cognitive Neuroscience Society (2020). This research was supported in part by a National Institutes of Health grant awarded to the University of Houston and by the University of Houston Small Grant Program and Research Progress Grant. We give our special thanks to Dr. Merrill Hiscock for his thorough editing. We also thank all the parents and children in our community who supported the research and participated in the study, as well as the research assistants in the Cognitive Development Lab for the data collection and annotation.

CONFLICT OF INTEREST

All the authors declare no conflicts of interest with regard to the funding source for the present study.

ORCID

Lichao Sun https://orcid.org/0000-0003-0166-9278

REFERENCES

Adamson, L. B., & Bakeman, R. (1984). Mothers' communicative acts: Changes during infancy. *Infant Behavior and Development*, *7*(4), 467–478. https://doi.org/10.1016/S0163-6383(84)80006-5

Akhtar, N., Dunham, F., & Dunham, P. J. (1991). Directive interactions and early vocabulary development: The role of joint attentional focus. *Journal of Child Language*, *18*(1), 41–49. https://doi.org/10.1017/S0305000900013283

Amano, S., Kezuka, E., & Yamamoto, A. (2004). Infant shifting attention from an adult's face to an adult's hand: A precursor of joint attention. *Infant Behavior and Development*, *27*(1), 64–80. https://doi.org/10.1016/j.infbeh.2003.06.005

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development*, *55*(4), 1278–1289. https://doi.org/10.2307/1129997

Baldwin, D. A. (1993). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology*, *29*(5), 832–843. https://doi.org/10.1037/0012-1649.29.5.832

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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bornstein, M. H., Tamis-LeMonda, C. S., Hahn, C.-S., & Haynes, O. M. (2008). Maternal responsiveness to young children at three ages: Longitudinal analysis of a multidimensional, modular, and specific parenting construct. *Developmental Psychology*, *44*(3), 867–874. https://doi.org/10.1037/0012-1649.44.3.867

Brooks, R., & Meltzoff, A. N. (2002). The importance of eyes: How infants interpret adult looking behavior. *Developmental Psychology*, *38*(6), 958–966. https://doi.org/10.1037/0012-1649.38.6.958

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, *8*(6), 535–543. https://doi.org/10.1111/j.1467-7687.2005.00445.x

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, *35*(1), 207–220. https://doi.org/10.1017/S030500090700829X

Burling, J. M., & Yoshida, H. (2019). Visual constancies amidst changes in handled objects for 5- to 24-month-old infants. *Child Development*, *90*(2), 452–461. https://doi.org/10.1111/cdev.13201

Butterworth, G. (1991). The ontogeny and phylogeny of joint visual attention. In A. Whiten (Ed.), *Natural theories of mind: Evolution, development and simulation of everyday mindreading* (pp. 223–232). Basil Blackwell.

Butterworth, G., & Cochran, E. (1980). Towards a mechanism of joint visual attention in human infancy. *International Journal of Behavioral Development*, *3*(3), 253–272. https://doi.org/10.1177/016502548000300303

Campbell, B. A., Hayne, H., Richardson, R., & Campbell, B. A. (2014). *Attention and information processing in infants and adults: Perspectives from human and animal research*. Psychology Press. https://doi.org/10.4324/9781315807355

Caron, A. J., Kiel, E. J., Dayton, M., & Butler, S. C. (2002). Comprehension of the referential intent of looking and pointing between 12 and 15 months. *Journal of Cognition and Development*, *3*(4), 445–464. https://doi.org/10.1207/S15327647JCD3,4-04

Carpenter, M., Nagell, K., Tomasello, M., Moore, C., & Butterworth, G. (1998). *Social cognition, joint attention, and communicative competence from 9 to 15 months of age*. University of Chicago Press.

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

Chang, L., de Barbaro, K., & Deák, G. (2016). Contingencies between infants' gaze, vocal, and manual actions and mothers' object-naming: Longitudinal changes from 4 to 9 months. *Developmental Neuropsychology*, *41*(5–8), 342–361. https://doi.org/10.1080/87565641.2016.1274313

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*(4), 213–220. https://doi.org/10.1037/h0026256

Colombo, J. (2001). The development of visual attention in infancy. *Annual Review of Psychology*, *52*(1), 337–367. https://doi.org/10.1146/annurev.psych.52.1.337

Colombo, J., & Cheatham, C. L. (2006). The emergence and basis of endogenous attention in infancy and early childhood. *Advances in Child Development and Behavior*, *34*, 283–322. https://doi.org/10.1016/S0065-2407(06)80010-8

Dalla, V., Giraitis, L., & Phillips, P. (2020). Robust tests for white noise and cross-correlation. *Econometric Theory*, 1–29. https://doi.org/10.1017/S0266466620000341

Datavyu Team. (2014). *Datavyu: A video coding tool*. Databrary Project, New York University. http://datavyu.org

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

Deák, G. O., Krasno, A. M., Jasso, H., & Triesch, J. (2018). What leads to shared attention? Maternal cues and infant responses during object play. *Infancy*, *23*(1), 4–28. https://doi.org/10.1111/infa.12204

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

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

Dunham, P. J., Dunham, F., & Curwin, A. (1993). Joint-attentional states and lexical acquisition at 18 months. *Developmental Psychology*, *29*(5), 827–831. https://doi.org/10.1037/0012-1649.29.5.827

Flom, R., Deák, G. O., Phill, C. G., & Pick, A. D. (2004). Nine-month-olds' shared visual attention as a function of gesture and object location. *Infant Behavior and Development*, *27*(2), 181–194. https://doi.org/10.1016/j.infbeh.2003.09.007

Flom, R., & Pick, A. D. (2003). Verbal encouragement and joint attention in 18-month-old infants. *Infant Behavior and Development*, *26*(2), 121–134. https://doi.org/10.1016/s0163-6383(03)00012-2

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-mounted eye tracking: A new method to describe infant looking. *Child Development*, *82*(6), 1738–1750. https://doi.org/10.1111/j.1467-8624.2011.01670.x

Franco, F., Perucchini, P., & March, B. (2009). Is infant initiation of joint attention by pointing affected by type of interaction? *Social Development*, *18*(1), 51–76. https://doi.org/10.1111/j.1467-9507.2008.00464.x

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, *71*(4), 878–894. https://doi.org/10.1111/1467-8624.00197

Goldstin, M. H., Schwade, J., Briesch, J., & Syal, S. (2010). Learning while babbling: Prelinguistic object-directed vocalizations indicate a readiness to learn. *Infancy*, *15*(4), 362–391. https://doi.org/10.1111/j.1532-7078.2009.00020.x

Gottlib, J., Oudeyer, P. Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: Computational and neural mechanisms. *Trends in cognitive sciences*, *17*(11), 585–593. https://doi.org/10.1016/j.tics.2013.09.001

Gredebäck, G., Fikke, L., & Melinder, A. (2010). The development of joint visual attention: A longitudinal study of gaze following during interactions with mothers and strangers: The development of joint visual attention. *Developmental Science*, *13*(6), 839–848. https://doi.org/10.1111/j.1467-7687.2009.00945.x

Hilbrink, E. E., Gattis, M., & Levinson, S. C. (2015). Early developmental changes in the timing of turn-taking: A longitudinal study of mother–infant interaction. *Frontiers in Psychology*, *6*, 1492. https://doi.org/10.3389/fpsyg.2015.01492

Holmboe, K., Bonneville-Roussy, A., Csibra, G., & Johnson, M. H. (2018). Longitudinal development of attention and inhibitory control during the first year of life. *Developmental Science*, *21*(6), e12690. https://doi.org/10.1111/desc.12690

Iverson, J. M., Capirci, O., Longobardi, E., & Caselli, M. C. (1999). Gesturing in mother-child interactions. *Cognitive Development*, *14*(1), 57–75. https://doi.org/10.1016/S0885-2014(99)80018-5

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

Kopp, C. B., & Vaughn, B. E. (1982). Sustained attention during exploratory manipulation as a predictor of cognitive competence in preterm infants. *Child Development*, *53*(1), 174–182. https://doi.org/10.1111/j.1467-8624.1982.tb01305.x

Leung, E. H., & Rheingold, H. L. (1981). Development of pointing as a social gesture. *Developmental Psychology*, *17*(2), 215–220. https://doi.org/10.1037/0012-1649.17.2.215

Lüdecke, D. (2018). ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software*, *3*(26), 772. https://doi.org/10.21105/joss.00772

Markus, J., Mundy, P., Morales, M., Delgado, C. E., & Yale, M. (2000). Individual differences in infant skills as predictors of child-caregiver joint attention and language. *Social Development*, *9*(3), 302–315. https://doi.org/10.1111/1467-9507.00127

Matatyaho, D. J., & Gogate, L. J. (2008). Type of maternal object motion during synchronous naming predicts preverbal infants' learning of word-object relations. *Infancy*, *13*(2), 172–184. https://doi.org/10.1080/15250000701795655

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282. https://doi.org/10.11613/bm.2012.031

Messinger, D. M., Ruvolo, P., Ekas, N. V., & Fogel, A. (2010). Applying machine learning to infant interaction: The development is in the details. *Neural Networks*, *23*(8–9), 1004–1016. https://doi.org/10.1016/j.neunet.2010.08.008

Morales, M., Mundy, P., Crowson, M., Neal, A. R., & Delgado, C. (2005). Individual differences in infant attention skills, joint attention, and emotion regulation behaviour. *International Journal of Behavioral Development*, *29*(3), 259–263. https://doi.org/10.1080/01650250444000432

Morales, M., Mundy, P., Delgado, C. E., Yale, M., Neal, R., & Schwartz, H. K. (2000). Gaze following, temperament, and language development in 6-month-olds: A replication and extension. *Infant Behavior and Development*, *23*(2), 231–236. https://doi.org/10.1016/S0163-6383(01)00038-8

Morales, M., Mundy, P., & Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behavior and Development*, *21*(2), 373–377. https://doi.org/10.1016/S0163-6383(98)90014-5

Morissette, P., Ricard, M., & Décarie, T. G. (1995). Joint visual attention and pointing in infancy: A longitudinal study of comprehension. *British Journal of Developmental Psychology*, *13*(2), 163–175. https://doi.org/10.1111/j.2044-835X.1995.tb00671.x

Moscovitch, M., & Winocur, G. (2002). The frontal cortex and working with memory. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 188–209). Oxford University Press.

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. https://doi.org/10.1111/j.1467-8624.2007.01042.x

Nagai, Y., & Rohlfing, K. J. (2009). Computational analysis of motionese toward scaffolding robot action learning. *IEEE Transactions on Autonomous Mental Development*, *1*(1), 44–54. https://doi.org/10.1109/TAMD.2009.2021090

Namy, L. L., Acredolo, L., & Goodwyn, S. (2000). Verbal labels and gestural routines in parental communication with young children. *Journal of Nonverbal Behavior*, *24*(2), 63–79. https://doi.org/10.1023/a:1006601812056

Namy, L. L., & Nolan, S. A. (2004). Characterizing changes in parent labelling and gesturing and their relation to early communicative development. *Journal of Child Language*, *31*(4), 821–835. https://doi.org/10.1017/S0305000904006543

Oakes, L. M., Kannass, K. N., & Shaddy, D. J. (2002). Developmental changes in endogenous control of attention: The role of target familiarity on infants' distraction latency. *Child Development*, *73*(6), 1644–1655. https://doi.org/10.1111/1467-8624.00496

Paterson, S. J., Heim, S., Friedman, J. T., Choudhury, N., & Benasich, A. A. (2006). Development of structure and function in the infant brain: Implications for cognition, language and social behaviour. *Neuroscience & Biobehavioral Reviews*, *30*(8), 1087–1105. https://doi.org/10.1016/j.neubiorev.2006.05.001

Pereira, A. F., James, K. H., Jones, S. S., & Smith, L. B. (2010). Early biases and developmental changes in self-generated object views. *Journal of Vision*, *10*(11), 22. https://doi.org/10.1167/10.11.22

Pereira, A. F., Smith, L. B., & Yu, C. (2013). A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review*, *21*(1), 178–185. https://doi.org/10.3758/s13423-013-0466-4

Podobnik, B., & Stanley, H. E. (2008). Detrended cross-correlation analysis: A new method for analyzing two nonstationary time series. *Physical Review Letters*, *100*(8), 084102084102. https://doi.org/10.1103/PhysRevLett.100.084102

Rader, N., & Zukow-Goldring, P. (2010). How the hands control attention during early word learning. *Gesture*, *10*(2–3), 202–221. https://doi.org/10.1075/gest.10.2-3.05rad

Rader, N., & Zukow-Goldring, P. (2012). Caregivers' gestures direct infant attention during early word learning: The importance of dynamic synchrony. *Language Sciences*, *34*(5), 559–568. https://doi.org/10.1016/j.langsci.2012.03.011

Raz, G., & Saxe, R. (2020). Learning in infancy is active, endogenously motivated, and depends on the prefrontal cortices. *Annual Review of Developmental Psychology*, *2*(1), 247–268. https://doi.org/10.1146/annurev-devpsych-121318-084841

Rohlfing, K. J., Longo, M. R., & Bertenthal, B. I. (2012). Dynamic pointing triggers shifts of visual attention in young infants. *Developmental Science*, *15*(3), 426–435. https://doi.org/10.1111/j.1467-7687.2012.01139.x

RStudio Team. (2021). *RStudio: Integrated Development for R*. RStudio, PBC. http://www.rstudio.com/

Ruff, H. A., & Lawson, K. R. (1990). Development of sustained, focused attention in young children during free play. *Developmental Psychology*, *26*(1), 85–93. https://doi.org/10.1037/0012-1649.26.1.85

Ruff, H. A., & Rothbart, M. K. (1996). *Attention in early development: Themes and variations*. Oxford University Press.

Ruff, H. A., & Rothbart, M. K. (2001). *Attention in early development: Themes and variations*. Oxford University Press.

Ruffman, T., Lorimer, B., Vanier, S., Scarf, D., Du, K., & Taumoepeau, M. (2020). Use of a head camera to examine maternal input and its relation to 10- to 26-month-olds' acquisition of mental and non-mental state vocabulary. *Journal of Child Language*, *47*(6), 1228–1243. https://doi.org/10.1017/S0305000920000240

Rutter, D. R., & Durkin, K. (1987). Turn-taking in mother–infant interaction: An examination of vocalizations and gaze. *Developmental Psychology*, *23*(1), 54–61. https://doi.org/10.1037/0012-1649.23.1.54

Saxon, T. F., Colombo, J., Robinson, E. L., & Frick, J. E. (2000). Dyadic interaction profiles in infancy and preschool intelligence. *Journal of School Psychology*, *38*(1), 9–25. https://doi.org/10.1016/S0022-4405(99)00034-5

Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature (London)*, *253*(5489), 265–266. https://doi.org/10.1038/253265a0

Senju, A., Csibra, G., & Johnson, M. H. (2008). Understanding the referential nature of looking: Infants' preference for object-directed gaze. *Cognition*, *108*(2), 303–319. https://doi.org/10.1016/j.cognition.2008.02.009

Sigman, M., Cohen, S. E., Beckwith, L., & Topinka, C. (1987). Task persistence in 2-year-old preterm infants in relation to subsequent attentiveness and intelligence. *Infant Behavior and Development*, *10*(3), 295–305. https://doi.org/10.1016/0163-6383(87)90018-X

Smith, L. B., Yu, C., & Pereira, A. F. (2011). Not your mother's view: The dynamics of toddler visual experience. *Developmental Science*, *14*(1), 9–17. https://doi.org/10.1111/j.1467-7687.2009.00947.x

Smith, L. B., Yu, C., Yoshida, H., & Fausey, C. M. (2015). Contributions of head-mounted cameras to studying the visual environments of infants and young children. *Journal of Cognition and Development*, *16*(3), 407–419. https://doi.org/10.1080/15248372.2014.933430

Striano, T., Chen, X., Cleveland, A., & Bradshaw, S. (2006). Joint attention social cues influence infant learning. *European Journal of Developmental Psychology*, *3*(3), 289–299. https://doi.org/10.1080/17405620600879779

Suanda, S. H., Smith, L. B., & Yu, C. (2016). The multisensory nature of verbal discourse in parent-toddler interactions. *Developmental Neuropsychology*, *41*(5–8), 324–341. https://doi.org/10.1080/87565641.2016.1256403

Suarez-Rivera, C., Smith, L. B., & Yu, C. (2019). Multimodal parent behaviors within joint attention support sustained attention in infants. *Developmental Psychology*, *55*(1), 96–109. https://doi.org/10.1037/dev0000628

Sun, L., Griep, C. D, & Yoshida, H. (2022). Shared multimodal input through social coordination: Infants with monolingual and bilingual learning experiences. *Frontiers in Psychology, Developmental Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.745904

Tamis-LeMonda, C. S., Kuchirko, Y., & Song, L. (2014). Why is infant language learning facilitated by parental responsiveness? *Current Directions in Psychological Science: A Journal of the American Psychological Society*, *23*(2), 121–126. https://doi.org/10.1177/0963721414522813

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. https://doi.org/10.1109/TAMD.2013.2269905

Tomasello, M. (1995). Joint attention as social cognition. *Joint attention: Its origins and role in development*, *103130*, 103–130.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *57*(6), 1454–1463. https://doi.org/10.1111/j.1467-8624.1986.tb00470.x

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

Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of Educational Psychology*, *102*(1), 43–53. https://doi.org/10.1037/a0016738

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

Yoshida, H., Cirino, P., Burling, J. M., Sunbok, L., & Lee, S. (2020). Parents' gesture adaptations to children with autism spectrum disorder. *Journal of Child Language*, *47*(1), 205–224. https://doi.org/10.1017/S0305000919000497

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. https://doi.org/10.1080/15250000802004437

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*(2), 244–262. https://doi.org/10.1016/j.cognition.2012.06.016

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLoS One*, *8*(11), e79659. https://doi.org/10.1371/journal.pone.0079659

Yu, C., & Smith, L. B. (2017). Hand–eye coordination predicts joint attention. *Child Development*, *88*(6), 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*(1), e12735. https://doi.org/10.1111/desc.12735

Zukow-Goldring, P. (1996). Sensitive caregiving fosters the comprehension of speech: When gestures speak louder than words. *Early Development and Parenting: An International Journal of Research and Practice*, *5*(4), 195–211. https://doi.org/10.1002/(SICI)1099-0917(199612)5:4<195::AID-EDP133>3.0.CO;2-H

How to cite this article: Sun, L., & Yoshida, H. (2022). Why the parent's gaze is so powerful in organizing the infant's gaze: The relationship between parental referential cues and infant object looking. *Infancy*, *27*(4), 780–808. https://doi.org/10.1111/infa.12475

APPENDIX

Model comparisons and selection process for Model 1

		Effects			Model fit			Likelihood ratio test		
Model	Nested model	Fixed effect	Random by dyad	Random by time lag	AIC	BIC	Loglik	Npar	*df*	Chi-squares
m00			Intercept	-	−94,791	−94,766	47,398	3		
m01	m00		"	Intercept	−94,824	−94,792	47,416	4	1	35.54***
m2	m01	Order	"	"	−94,830	−94,790	47,420	5	1	7.90**
m3	m2	Order + age	"	"	−94,828	−94,779	47,420	6	1	0.01
m3a	-	Order + age	Order	"	Convergence warning and removed random slope.			-	-	-
m4	m3	Order × age	Intercept	"	−94,836	−94,779	47,425	7	1	9.49**
m5	m4	Order × age, language group	"	"	−94,834	−94,769	47,425	8	1	0.07

Significance levels: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Model 1 summary

Estimates of random effects for the general mixed-effect model on the correlation between parent object looking and infant object looking

Parameter	Variance	*SD*
Subject (intercept)	0.0019	0.044
Time lag (intercept)	0.0001	0.004
Residual	0.0015	0.038

Estimates of fixed effects for the general mixed-effect model on the correlation between parent object looking and infant object looking

Parameter	Estimate	SE	*t*	*p*	95% CI
(Intercept)	0.0296	0.0200	1.48	<0.001	[−0.0107, 0.0698]
Age	0.0003	0.0017	0.19	0.685	[−0.0031, 0.0038]
Temporal order: Parent pre infant	0.0056	0.0014	3.93	<0.001***	[0.0028, 0.0085]
Temporal order: Parent pre infant × age	−0.0003	0.0001	−3.08	<0.001***	[−0.0006, −0.0001]

Significance levels: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Model comparisons and selection process for Model 2

		Effects			Model fit			Likelihood ratio test		
Model	Nested model	Fixed effect	Random by dyad	Random by time lag	AIC	BIC	Loglik	Npar	*df*	Chi-squares
m00		-	Intercept	-	−89,491	−89,467	44,749	3		
m01	m00	-	"	Intercept	−89,667	−89,634	44,837	4	1	177.54***
m2	-	Order	"	"	Convergence warning, variance in time lag closed to zero and removed.			-	-	-
m2a	m00	Order	"	-	−90,176	−90,143	45,092	4	1	686.98***
m2b	m2a	Order	Order	-	−95,792	−95,744	47,902	6	2	5620.36***
m3	m2b	Order + age	"	-	−95,793	−95,736	47,904	7	1	2.78
m4	m2b	Order × age			−95,792	−95,727	47,904	8	2	3.33
m5	m2b	Order × age, language group	"	-	−95,790	−95,716	47,904	9	3	3.37

Significance levels: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Model 2 summary

Estimates of random effects for the general mixed-effect model on the correlation between parent object labeling and infant object looking

Parameter	Variance	*SD*	Correlation
Subject (intercept)	0.0019	0.043	
Order | subject	0.0014	0.038	−0.39
Residual	0.0014	0.037	

Estimates of fixed effects for the general mixed-effect model on the correlation between parent object labeling and infant object looking

Parameter	Estimate	SE	*t*	*p*	95% CI
(Intercept)	0.0137	0.0066	2.09	0.043	[0.0005, 0.0270]
Temporal order: Parent pre infant	−0.0138	0.0058	−2.39	0.021*	[−0.0254, −0.0021]

Significance levels: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Model comparisons and selection process for Model 3

		Effects			Model fit			Likelihood ratio test		
Model	Nested model	Fixed effect	Random by dyad	Random by time lag	AIC	BIC	Loglik	Npar	*df*	Chi-squares
m00			Intercept	-	−70,537	−70,512	35,271	3		
m01	m00		"	Intercept	−73,375	−73,342	36,691	4	1	2840.4***
m2	m01	Order	"	"	−73,570	−73,530	36,790	5	1	197.42***
m3	m2	Order + age	"	"	−73,569	−73,520	36,790	6	1	0.1908
m3a	m3	Order + age	Order	"	−88,878	−88,813	44,447	8	2	15,313.32***
m4	-	Order × age	"	"	Convergence warning and removed random slope.			-	-	-

Significance levels: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Model 3 summary

Estimates of random effects for the general mixed-effect model on the correlation between parent object handling and infant object looking

Parameter	Variance	*SD*	Correlation
Subject (intercept)	0.0097	0.098	
Order | subject	0.0055	0.074	−0.44
Time lag (intercept)	0.0004	0.020	
Residual	0.0016	0.040	

Estimates of fixed effects for the general mixed-effect model on the correlation between parent object handling and infant object looking

Parameter	Estimate	SE	*t*	*p*	95% CI
(Intercept)	0.0500	0.0406	1.23	0.225	[−0.0319, 0.1319]
Temporal order: Parent pre infant	−0.0260	0.0115	−2.25	0.029*	[−0.0493, −0.0027]
Age	0.0014	0.0033	0.40	0.692	[−0.0055, 0.0082]

Significance levels: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Model 4 summary

Estimates of random effects for the generalized mixed-effect model on an infant's sustained attention on the target objects

Source	Variance	*SD*
Subject (intercept)	0.86	0.93

Estimates of fixed effects for the generalized mixed-effect model on an infant's sustained attention on the target objects

Source	Estimate	SE	Z	p
(Intercept)	−1.24	0.17	−7.48	<0.001
Parent object handling	−0.06	0.01	−4.58	<0.001***
Parent object looking	0.17	0.03	6.66	<0.001***
Irrelevant labeling	−0.54	0.02	−22.04	<0.001***
Relevant labeling	0.15	0.02	7.06	<0.001***
Parent object handling × object looking	0.08	0.03	2.66	0.008**
Parent object handling × irrelevant labeling	0.37	0.03	13.93	<0.001***
Parent object handling × relevant labeling	0.12	0.02	4.91	<0.001***
Parent object looking × irrelevant labeling	0.21	0.05	4.91	<0.001***
Parent object looking × relevant labeling	0.16	0.04	3.88	<0.001***
Parent object handling × parent object looking × irrelevant labeling	−0.13	0.05	−2.68	0.007**
Parent object handling × parent object looking × relevant labeling	−0.31	0.05	−6.75	<0.001***
Age	0.005	0.02	0.30	0.762
Language group: ML	−0.08	0.22	−0.37	0.713

Significance levels: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

TABLE A1 Proportions of multimodal input in each parental referential cue

Parental referential cue	Overlapping	Mean (%)	Std (%)
Object looking	Alone	10.1	8.3
	+hand	43.9	18.7
	+irrelevant label	3.2	3.7
	+irrelevant label + hand	16.0	11.2
	+relevant label	3.9	4.4
	+relevant label + hand	23.0	10.1
Parental phrases			
Irrelevant label	Alone	6.3	6.1
	+gaze	1.3	1.7
	+hand	28.5	16.1
	+gaze + hand	6.0	4.5
Relevant label on target objects	Alone	7.9	8.8
	+gaze	2.0	2.9
	+hand	38.3	14.6
	+gaze + hand	9.7	6.6
Object handling	Alone	42.9	16.8
	+gaze	10.3	7.0
	+irrelevant label	16.0	10.5
	+irrelevant label + gaze	3.7	3.1
	+relevant label	21.4	9.9
	+relevant label + gaze	5.7	4.3