Received: 26 February 2017 | Accepted: 13 December 2017

DOI: 10.1111/desc.12735

PAPER

WILEY Developmental Science

Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months

Chen Yu[1] | Sumarga H. Suanda[2] | Linda B. Smith[1]

[1]Department of Psychological and Brain Sciences, Program in Cognitive Science, Indiana University, Bloomington, Indiana

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

Correspondence
Chen Yu, Department of Psychological and Brain Sciences, and Program in Cognitive Science, Indiana University, 1101 East 10th Street, 47405, Bloomington, IN 47405.
Email: chenyu@indiana.edu

Abstract

Vocabulary differences early in development are highly predictive of later language learning as well as achievement in school. Early word learning emerges in the context of tightly coupled social interactions between the early learner and a mature partner. In the present study, we develop and apply a novel paradigm—dual head-mounted eye tracking—to record momentary gaze data from both parents and infants during free-flowing toy-play contexts. With fine-grained sequential patterns extracted from continuous gaze streams, we objectively measure both joint attention and sustained attention as parents and 9-month-old infants played with objects and as parents named objects during play. We show that both joint attention and infant sustained attention predicted vocabulary sizes at 12 and 15 months, but infant sustained attention in the context of joint attention, not joint attention itself, is the stronger unique predictor of later vocabulary size. Joint attention may predict word learning because joint attention supports infant attention to the named object.

KEYWORDS

eye tracking, joint attention, language development, sustained attention, word learning

1 | INTRODUCTION

Children learn words with a social partner. The mature social partner provides the words and chooses the moments when to supply them (Akhtar, Dunham, & Dunham, 1991; Dunham, Dunham, & Curwin, 1993; Hoff & Naigles, 2002; Masur, Flynn, & Eichorst, 2005; Tamis-LeMonda, Kuchirko, & Song, 2014; Tomasello & Todd, 1983). The mature partner also helps guide the young learner's attention to the intended referent, creating moments of joint attention, when the mature partner and learner focus on the same object (Akhtar & Tomasello, 2000; Baldwin, 1995; Deák, Triesch, Krasno, de Barbaro, & Robledo, 2013; Yu & Smith, 2017b). Individual differences in both quantity and quality of parent talk and frequency of parent–infant coordinated attention predict individual differences in early vocabulary size (Cartmill et al., 2013; Golinkoff, Can, Soderstrom, & Hirsh-Pasek, 2015; Rowe, 2012; Weisleder & Fernald, 2013). Individual differences in children's early vocabulary sizes, in turn, are strong predictors of future language and school achievement (Hoff, 2013; Murphy, Rowe, Ramani, & Silverman, 2014). Increasing evidence linking inequalities in developmental environments to inequalities in developmental outcomes, has increased the urgency of a more precise understanding of the pathways through which predictors such as amount of parent talk and joint attention influence vocabulary development (Fernald, Marchman, & Weisleder, 2013; Hart & Risley, 2003; Rowe & Goldin-Meadow, 2009). Determining these pathways is essential if we are to offer prescriptions to parents and policy makers on how to address these differences in developmental outcomes. However, this will not be an easy task. Theorists in developmental psychopathology (Masten & Cicchetti, 2010; Masten et al., 2005) often use the term "developmental cascade" to refer to the complex and multicausal chain of events through which competencies emerge. Individual differences in vocabulary development are clearly the product of such a complex multicausal cascade. Thus, individual differences in vocabulary may be strongly associated with many factors, each of which plays quite different roles in determining individual outcomes (Hirsh-Pasek et al., 2015; Smith, 2013). From the perspective of this complex developmental landscape, there are two critical empirical questions: The first is the determination of

early predictors because these signals are potentially malleable early factors that set different learners on different developmental trajectories. The second empirical question concerns the precise role of different factors and particularly in what way these different factors engage the child's learning mechanisms. The present paper attempts to answer those two questions in the context of the early language environment, and specifically focuses on the quality and quantity of parent object naming when interacting with their 9-month-old infants.

1.1 | Joint attention and quality of naming events

Certainly, an object name cannot be learned if it is not heard. Thus, repetitions of the to-be-learned name is likely to be beneficial to learning. However, the quantity of naming events may not benefit word learning in and of itself; instead, it may be the quantity of high-quality naming events that is the key to growing a vocabulary (Cartmill et al., 2013; Golinkoff et al., 2015; Hirsh-Pasek et al., 2015; Rowe, 2012). Past research has focused on speech and language properties as quality measures or on more global assessments of early communication (Fernald & Marchman, 2010; Golinkoff et al., 2015; Hirsh-Pasek et al., 2015; Newman, Rowe, & Ratner, 2016). Here, we link quality to the attentional states of the mature partner and the infant learner.

These attentional states are potentially relevant indices of quality because past research has shown that naming events that are characterized by the shared attention of both partners to the object lead to more certain learning of object names (Akhtar & Tomasello, 2000; Baldwin, 1995). Moreover, observational studies show that dyad differences in the frequency with which parents and infants engage in episodes of joint attention predict individual differences in child vocabulary size (Tomasello & Todd, 1983). All this suggests that a relevant metric for quality, at least for very young learners, may be whether the mature partner and the young learner coordinate attention to the same object during naming events. By one hypothesis then, the *quantity* of parent *naming that occurs within joint attention episodes* is the relevant property of early word learning environments that determines vocabulary development.

1.2 | The sustained-attention hypothesis

However, there is an alternative hypothesis: The quantity of parent object naming within episodes of joint attention may predict later vocabulary size, not because joint attention is essential to learning object names, but because joint attention coincides with what is essential: infant sustained attention to the named object when it is named. By traditional accounts, joint attention has its own direct effect on word learning because it enables the learner to build an internal model of the mature partner's referential intent, and by these accounts, determining that intent is essential to learning (Akhtar & Tomasello, 2000). However, when a parent and infant jointly attend to a referent, the infant is also attending to the referent. Thus, it may be the infant side of joint attention that is the actual cause of learning an object name. For the learner to link the heard name to the right object, the mature partner must name (and thus is likely to look at) the object and the infant, of course, must look at the object as well. Therefore, effective parent naming is often embedded in a joint attention episode. However, in the end, it is the infant that must do the learning when hearing a name and accordingly the most causally relevant predictor may be infant sustained attention to the named object when it is named.

This sustained-attention hypothesis is suggested by several prior findings. First, infant sustained attention—the stabilization of visual attention to an object for long durations (e.g., greater than 3 s) predicts later language learning and cognitive development (Kannass & Oakes, 2008; Lawson & Ruff, 2004; Ruff, Lawson, Parrinello, & Weissberg, 1990). Second, a growing number of studies indicate that toddlers who visually attend longer to a target object when it is named are more likely to remember the name-object mapping than when visual attention to the named target was briefer (Macroy-Higgins & Montemarano, 2016; Pereira, Smith, & Yu, 2014; Salley, Panneton, & Colombo, 2013). Third, joint attention and sustained attention—when measured independently—have been shown to be strongly associated. More specifically, infant visual attention to an object lasts longer when it occurs within a shared attention episode (Yu & Smith, 2016). This last result suggests that joint attention may not just co-occur with infant sustained attention but may play a key supportive role. A recent study (Yu & Smith, 2016) indeed shows that when the social partner (parent) visually attended to the object to which infant attention was directed, infants, after the parent's look, extended their duration of visual attention to the object. This coincidence between joint attention and sustained attention in parent–infant social interaction raises both a theoretical question and a methodological challenge on how to precisely determine and examine the contributions of the two factors in early word learning.

The experiment had two goals: (a) to show that the attentional states of parents and infants during naming moments were relevant

RESEARCH HIGHLIGHTS

- A dual head-mounted eye tracking paradigm was developed and applied to record momentary gaze data from both parents and infants during free-flowing toy-play contexts.
- Both joint attention and the infant's sustained attention at 9 months predicted vocabulary sizes at 12 and 15 months.
- Infant sustained attention in the context of joint attention, but not joint attention itself, is the stronger unique predictor of later vocabulary size.
- Joint attention may predict word learning because joint attention supports infant attention to the named object.

indices of the quality of parent naming moments and that by this definition the quantity of high quality parent naming events at 9 months predicted later vocabulary size and (b) to disentangle the contributions of joint attention and infant sustained attention in these predictive relations.

2 | METHOD

Most of the evidence linking quantity and quality of parent talk to object name learning has focused on children who are between 1- and 2-years of age (Rollins, 2003). We instead focus on parent interactions with their 9-month-old infants for two reasons: First, infants begin learning object names well before their first birthday (Bergelson & Swingley, 2012). Second, the period around the 9-month mark has been characterized as a key transition period during which infants become more interested in objects, and parents and infants first begin to jointly engage and attend to the same object (Tomasello, Carpenter, Call, Behne, & Moll, 2005). Accordingly, the present study measured individual differences in parent naming of objects and in joint attention to named objects between parents and infants when infants were 9-month-olds, and asked how these properties of early real-time social interactions predicted vocabulary size when the infants were 12 and 15 months.

The usual approach to this kind of study is to use a battery of assessment trial tests (most often individual discrete trials) to measure each of the hypothesized relevant predictors, and then use statistical procedures to link those measures to the to-be-predicted outcome, partitioning variance as a way to determine the unique variance in the outcome measure associated with each predictor (Salley et al., 2013; Yu & Smith, 2017a). We also took this approach but we did so with measures of fine temporal real-time behaviors in free-flowing interaction. Both infants and parents wore head-mounted eye-tracking equipment as shown in Figure 1 when they played with each other with a set of toys. Thus, we recorded the gaze data from both participants with a high temporal and spatial resolution. We objectively measured joint attention and infant sustained attention from the momentary gaze of the two participants and determined those moments within which joint attention and sustained attention did and did not overlap. We determined the frequency of naming events within these two attentional streams. Finally, we asked how well these measures predicted infants' vocabulary sizes. Because we independently measured joint attention and sustained attention in real time, we were able to use analyses that disentangled their contributions as predictors of vocabulary that are more direct and powerful than the statistical partitioning of variance in regression analyses.

2.1 | Participants

The final sample consisted of 26 parent–infant dyads, the mean age of the infants was 9.21 months (*SD* = 0.23) and there were 12 male infants. Parent report measures of vocabulary were collected 3 and 6 months later when the infants were 12 months and 15 months. 8 additional infants began the study but refused to wear the measuring equipment. A sample of 26 dyads was selected given the size of the effects in similar previous studies that also link high-density microlevel behavioral data with early vocabulary (Ramirez-Esparza, Garcia-Sierra, & Kuhl, 2014; Weisleder & Fernald, 2013).

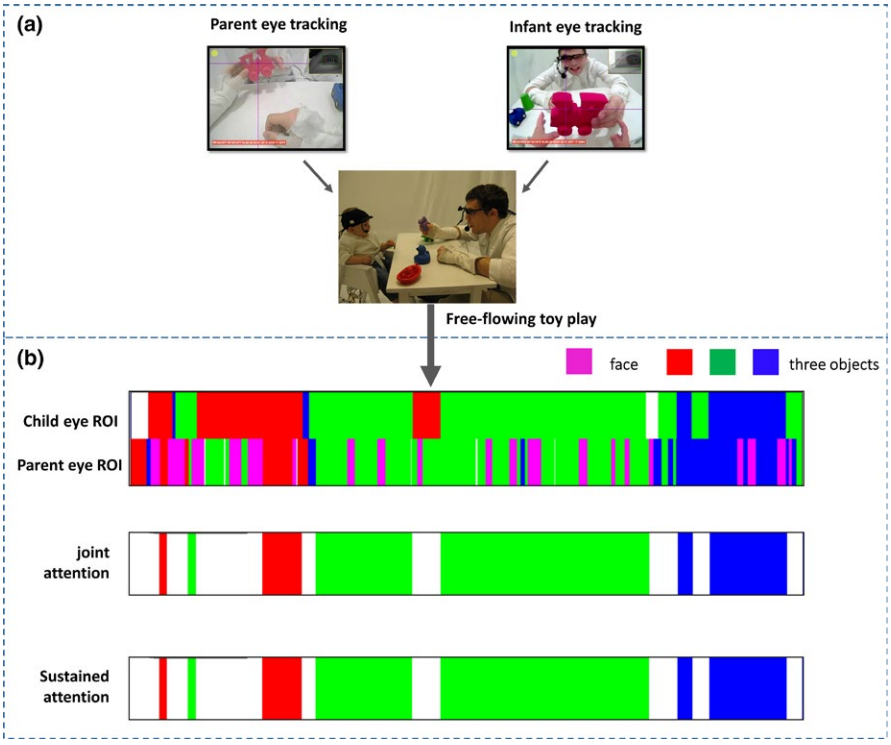

FIGURE 1 (a) A dual eye tracking experimental paradigm wherein infants and parents played with a set of toys on a tabletop in a free-flowing way. Both participants wore a head-mounted eye tracker that recorded their moment-to-moment gaze direction from their egocentric views. (b) The two data streams at the top represent raw regions-of-interest (ROIs) data from infants and their parents in toy play. Based on the two ROI streams, joint attention was derived when parents and infants were looking at the same object at the same time. In addition, sustained attention was derived when infants showed attention on objects for a long period of time

2.2 | Stimuli

The experimental toys were 6 everyday toys, organized into two sets of three (car, cup and train, duck, plane and boat). Each toy in the set had a unique uniform color and all toys were of similar size, on average, 288 cm^3 (see Figure 1). Additional toys were used to engage the child during the placement of the eye-tracker and its calibration.

2.3 | Experimental setup

Parents and infants sat across from each other at a small table (61 cm × 91 cm × 64 cm). The infants sat in a customized high-chair that supported sitting steadily. Parents sat on the floor such that their eyes and heads were at approximately the same distance from the tabletop as those of the infants, a posture that parents reported to be natural and comfortable. Both participants wore head-mounted eye trackers from Positive Science, LLC (Franchak, Kretch, Soska, Babcock, & Adolph, 2010; Yu & Smith, 2013). Each eye-tracking system includes an infrared camera—mounted on the head and pointed to the right eye of the participant that records eye images, and a scene camera (see in Figure 1) capturing the first-person view from the participant's perspective. The scene camera's visual field is 90°, providing a broad view but one less than the full visual field (approximately 170°). Each eye tracking system recorded both the egocentric-view video and gaze direction (x and y) in that view, with a sampling rate of 30 Hz. Another high-resolution camera (recording rate 30 frames per second) was mounted above the table and provided a bird's eye view that was independent of participants' movements. Parent speech was recorded from an onboard microphone in the parent eye tracker.

2.4 | Procedure

Three experimenters worked together during the experiment. One experimenter played with the infant while another placed the eye-tracking gear low on the forehead of the infant at a moment when the child was engaged with a toy used only for this phase of the experiment. The third experimenter controlled the computer to ensure data recording. To collect calibration points for eye tracking, the first experimenter directed the infant's attention toward an attractive toy used only for calibration while the second experimenter recorded the attended moment that was used in later eye tracking calibration. This procedure was repeated 15 times with the calibration toy placed in various locations on the tabletop. To calibrate the parent's eye tracker, the experimenter asked the parent to look at one of the calibration toys on the table, placed close to the infant, and then repeated the same procedure to obtain 9 to 15 calibration points from the parent. Parents were told that the goal of the experiments was to study how parents and infants interacted with objects during play. Therefore, they were asked to engage their infants with the toys and to do so as naturally as possible. They were not told that we were interested in naming events, nor were they instructed to name the objects. Each of the two sets of toys was played with twice for 1.5 min, resulting in 6 min of play data from each dyad. Order of sets (ABAB or BABA) was counterbalanced across dyads.

2.5 | Data and data processing

Each eye-tracker collected at a rate of 30 frames per second for approximately 240 s (four trials with 1 min per trial) of interaction, yielding potentially 7,200 data points per measure for each participant. Not all participants provided eye-tracking data for the entire session, the mean number of good eye-tracking frames was 5,735 (*SD* = 568) for infants and 5,617 (*SD* = 521) for adults. Together, there were 5,122 frames (*SD* = 537) available simultaneously from both social partners in each interaction, which accounted for 71% of the time. Roughly 25% of frames from infants that were not codable with respect to regions of interest (ROIs, defined in the next paragraph); this was due to 13% eye-tracking failure and the rest due to the infant's being off task (looking elsewhere than defined regions of interest).

In total, the method uses microbehavioral analyses with over 10,000 gaze data points from each interaction. We annotated gaze and speech data during toy play, from which we derived measures of both quantity and quality of parent talk.

2.5.1 | Gaze data

The three regions-of-interest (ROIs) were the three toy objects in play at a time. These ROIs were coded manually by coders who watched the first-person view video, frame by frame, with a cross-hair indicating gaze direction and annotated when the cross-hairs overlapped any portion of an object and if so, on which object. Thus, each dyad provided two gaze data streams as shown in Figure 1B. The second coder independently coded a randomly selected 10% of the frames with 95% agreement.

2.5.2 | Naming events

Parental speech was transcribed into spoken utterances, among which those that contained names of the toys were designated as naming events. Each naming event is coded as a triplet <onset, offset, name>.

2.6 | Joint attention

Joint attention (JA) was defined as periods during which parents and infants were jointly fixated on the same object at the same time (Yu & Smith, 2013, 2017b). Previous research has shown that parents, but not infants, often glance very briefly to other objects or the infant's face—monitoring the whole scene—even while more generally attending to the same object as their infant (Yu & Smith, 2013). Further, meaningful shared attention should last some amount of time longer than a single video frame (33 ms). Accordingly, we defined a joint attention bout as a continuous alignment of parent and infant fixation that lasted longer than 500 ms but that could include

TABLE 1 Means, standard deviations (*SD*), and ranges of naming, SA, JA, the interaction term of naming and SA, and the interaction term of naming and JA at 9 months, and vocabulary measured at 12 and 15 months

Measure	*M*	*SD*	Range
Naming (frequency per min)	9.62	2.93	4.43–16.42
Sustained attention (SA)	25.78%	12.22%	3.62%–60.67%
Joint attention (JA)	21.07%	11.12%	2.63%–54.88%
Naming × SA	2.31	1.2	0.38–4.71
Naming × JA	1.85	0.99	0.21–4.21
MCDI at 12 months	83.73	49.94	10–185
MCDI at 15 months	133.96	53.57	18–236

looks briefer than 300 ms elsewhere. Given that humans generate three saccades per minute, this threshold of 300 ms allowed one brief look away before switching back to the target. Examples of joint attention bouts are shown in Figure 1B. To determine the quality of naming events with respect to JA, we calculated the proportion of time within a naming event that parents and infants also jointly attended to the named object.

2.7 | Sustained attention

Sustained attention (SA) was defined by consideration of the infant gaze alone. A 3 s of consistent looking by the infant within the ROI for a single object *without any looks elsewhere* was counted as the threshold for sustained attention on that object by the infant. The 3 s was chosen as the threshold because it was the average duration of concentrated attention for 1-year-olds reported in a recent study using head-mounted eye tracking (Yu & Smith, 2016) and because this is the same threshold used by other researchers as defining a period of sustained attention (Ruff & Lawson, 1990). To measure the quality of naming events with respect to SA, we calculated the proportion of time within a naming event that infants showed sustained attention on the target object.

2.8 | Vocabulary growth

Infants returned twice at age 12 and 15 months ($SD_{12\text{-month}}$ = 9.62 days; $SD_{12\text{-month}}$ = 11.67 days) after the first visit. At those two visits, parents completed the MacArthur-Bates Communicative Development Inventory (Fenson et al., 1993) which asks parents to indicate on a checklist the words that their infants comprehend.

3 | RESULTS

3.1 | Descriptive statistics of individual measures

The to-be-predicted outcomes, infants' vocabularies at 12 and 15 months, varied considerably in the sample, as is characteristic for infants at these ages. The Mean MCDI score at 12 months was 83.73 words (*SD* = 49.74, ranging from 10 to 185), and mean MCDI score at 15 months was 133.96 words (*SD* = 53.57, ranging from 18 to 236). These ranges are comparable to MCDI data collected from a large sample of children at similar ages from wordbank (Frank, Braginsky, Yurovsky, & Marchman, 2016). There was a strong correlation between MCDI scores at 12 and 15 months (Pearson correlation, $r = 0.83$, $p < 0.001$).

To acquire object names, infants need to hear object labels; therefore, the frequency of parent naming of objects during the play would seem a likely predictor of later vocabulary development. We calculated the number of naming utterances produced by parents during toy play. These ranged from rates of 4.43 to 16.42 naming instances per minute (*M* = 9.62, *SD* = 2.31). This variation, also shown in Table 1, is consistent with the literature showing marked individual differences in the amount of speech directed to children in everyday learning environments (Weisleder & Fernald, 2013). However, the quantity of naming events during this one-time play session when the infant was 9 months old did not, *by itself*, predict vocabulary size at ages 12 and 15 months ($r_{12\text{-mo}} = 0.21$, $p = 0.29$; $r_{15\text{-mo}} = 0.13$, $p = 0.53$; see Figure 2). As shown in Table 1, the proportion of time that parents and infants jointly attended to the named object during naming moments also varied widely (2.63%–54.88%, *M* = 21.07%, *SD* = 11.12%). Some dyads spent the majority of their time jointly attending to named objects and others rarely did so. Similar to the JA measure, we found considerable variation in infant sustained attention (SA) within naming moments, ranging from 3.62% to 60.67% (*M* = 25.78%, *SD* = 12.22%). For some infants, parent naming of objects overlapped with infant sustained attention to the object, but for other naming events, infants merely glanced at the object or did not look at all. As expected, the proportion of time in JA and SA during naming moments were strongly correlated with each other (Pearson correlation, $r = 0.69$, $p < 0.001$). When parents and infants jointly attended to the same object while parents named it, infants were also likely to show sustained attention on the target object.

Critically, neither the proportion of time in JA nor in SA during naming moments were significantly correlated with the frequency of naming ($r_{JA} = -0.30$, $p = 0.12$, $r_{SA} = -0.23$, $p = 0.24$). Thus, quantity of parent talk was not correlated with either aspect of quality. Parents who produced more toy names during play did not, as a group, provide proportionally more or less high-quality learning instances. However, parents who talk more are, by definition, offering their children more words, and the more words an infant hears, the more likely some naming instances have high quality.

3.2 | Combining quality and quantity

The pathway to learning likely depends on both quality and quantity with the key predictor being the frequency of high-quality naming events. But what counts as a high quality naming event, one characterized by joint attention or by infant sustained attention? We addressed this question by calculating the interaction term of quality times quantity, using the number of naming instances as the quantity

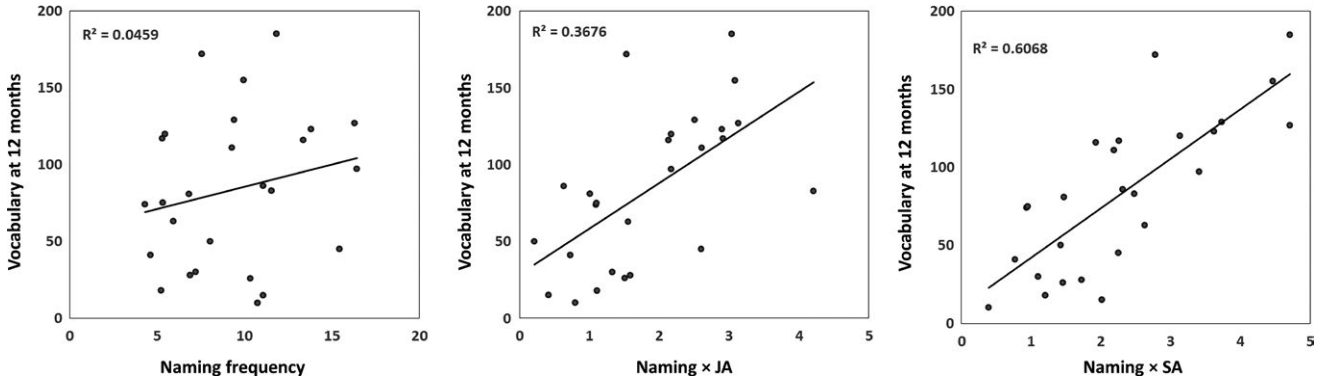

FIGURE 2 The three scatter plots (with the best-fitting regression lines) show correlations between (a) naming frequency during toy play at 9 months and vocabulary size at 12 months, (b) naming frequency (quantity measure of parent talk) times joint attention at naming moments (quality measure) at 9 months and vocabulary size at 12 months, and (c) naming frequency (quantity measure) times sustained attention at naming moments (quantity measure) at 9 months and vocabulary size at 12 months

measure and the proportion of JA or SA time within naming events as the two quality measures. The interaction term is calculated by multiplying the quantity measure (number of naming instances) with a quality measure (either using JA or SA within naming events as quality measure). This metric combines the quantity of naming instances with quality of those individual instances, which can be conceptualized as aggregating information from multiple naming instances, each of which contributes more or less to learning dependent on its quality. The two interaction terms are denoted as "naming × JA" and "naming × SA" and descriptive statistics of the two are reported in Table 1. As shown in Table 2, the interaction terms of the frequency of naming with JA and with SA each reliably predicted vocabulary outcomes at both 12 and 15 months. We verified statistical significances using robust regression (robustfit function in the MATLAB Statistics Toolbox) which prevents undue influence of outliers, and found again both JA and SA at naming moments predicted MCDI scores at 12 and 15 months. For MCDI scores at age 12 months, the quantity-times-quality measure, using SA explained 58.9% of the variation ($p < 0.001$) and the quality-times-quantity measure using JA explained 39.7% ($p < 0.001$). Moreover, quantity-times-quality SA measure and the quantity-times-quality JA measure accounted for, respectively, 32.2% ($p < 0.005$) and 23.5% ($p < 0.01$) of the variation in child vocabulary at age 15 months. However, when we correlated vocabulary measures at 12 and 15 months with the average of quality measures from individual infants, neither JA- or SA-based measures predicted later vocabulary ($r_{JA\text{-}12mo} = 0.12$, n.s.; $r_{JA\text{-}15mo} = 0.16$, n.s.; $r_{SA\text{-}12mo} = 0.21$, n.s.; $r_{SA\text{-}15mo} = 0.17$, n.s.). Thus, neither quantity nor quality alone is predicative but the relevant quantity-by-quality measure involves the combination of both measures from all the naming events in toy play. These results highlight the importance of quality–quantity combination in early parent–infant interaction, and by doing so, provide additional evidence on the role of quality and quantity on language learning (Cartmill et al., 2013; Weisleder & Fernald, 2013). Moreover, past research has focused on speech and language properties as quality measures or on more global assessments of early communication (Fernald & Marchman, 2010; Golinkoff et al., 2015; Hirsh-Pasek et al., 2015; Newman et al., 2016). Here, we link quality to real-time attentional states of the participants in learning—joint attention by the social partners and sustained attention by the infant..

TABLE 2 Correlations between MCDI scores at age 12 months and 15 months with three measures derived from toy play at 9 months, naming, the interaction term of naming and JA, and the interaction term of naming and JA. The numbers in [] show 95% confidence intervals

	MCDI at 12 months	MCDI at 15 months
Naming	0.21 [−0.18, 0.55]	0.13 [−0.27, 0.49]
Naming × SA	0.61***[0.28, 0.81]	0.48** [0.12, 0.73]
Naming × JA	0.78***[0.56, 0.90]	0.58*** [0.26, 0.79]

Notes. **$p < 0.01$, ***$p < 0.005$.

3.3 | Joint attention or infant sustained attention?

JA and SA are strongly correlated with each other. They could contribute to word learning through potentially different pathways or by hypothesis through a single pathway in which joint attention is often the context for infant sustained attention, but wherein child sustained attention is the key factor. The present high-density gaze data provides a way to derive precise independent measures of both JA and SA allowing for the direct evaluation of their contributions. As shown in Figure 3, we defined three new measures of naming quality based on the temporal relation between JA and SA: (a) SA with JA: moments in which infants sustained attention to the named object within a joint attention bout; (b) SA w/o JA: moments during which infants sustained attention to the named referent that did not coincide with a joint attention bout; and (c) JA w/o SA: moments that parents and infants jointly attended to the same object but infants' attention did not pass the duration threshold for "sustained." In addition, we also calculated three interaction terms: naming × (SA with

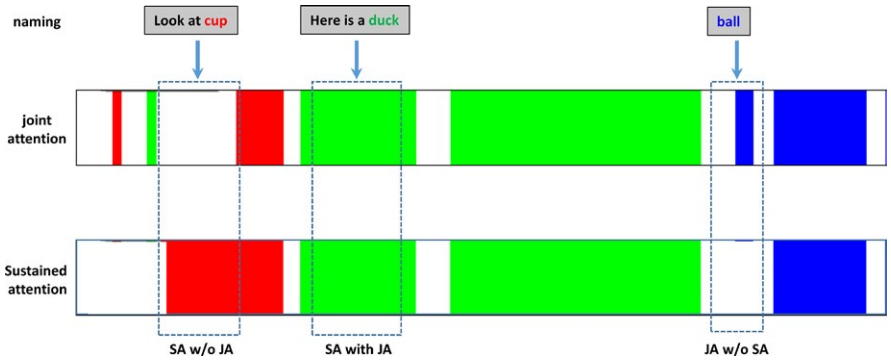

FIGURE 3 Three measures of naming quality are fined based on the temporal relation between JA and SA at naming moments: (1) SA with JA: moments that infants showed sustained attention on the named object while parents also attended to the same object; (2) SA w/o JA: moments that infants showed sustained attention on the target without parents looking at the target object; and (3) JA w/o SA: moments that parents and infants jointly attended to the same object but infants did not show sustained attention to the object

JA), naming × (SA w/o JA), and naming × (JA w/o SA), by multiplying the frequency of naming instances (per minute) with the proportion of time of SA with JA, SA w/o JA, and JA w/o SA respectively. Descriptive statistics of the three measures and their corresponding interaction terms with naming frequency are reported in Table 3.

Overall, SA with JA measures the proportion of play time that SA and JA co-occurred together, a fact that underlies their overall correlation. In contrast, SA w/o JA and JA w/o SA, respectively, capture two kinds of situations in which JA or SA happened alone and therefore were not correlated. Accordingly, we expected that SA with JA should positively correlate with MCDI scores. The key question concerns the two uncorrelated cases. As shown Table 4, as expected SA moments accompanying JA predict vocabulary. Critically, however, moments of SA *without* accompanying JA also positively predict MCDI scores. Thus, two of the three measures containing sustained attention during naming events are predictive of later language development. However, the third measure—JA without accompanying SA—does not predict vocabulary. Taken together, our results suggest that sustained attention, but not joint attention, is a key predictor of later vocabulary size.

To provide additional support for this conclusion, we also conducted more traditional hierarchical regression analyses to use naming × JA and naming × SA measures at 9 months to predict MCDI scores at 12 and 15 months. This type of analysis—which has been used in developmental studies (Hirsh-Pasek et al., 2015)—allows us to specify an order of entering variables in regression to test the effects of newly added variables in addition to the influence of the existing ones. For example, in Model 1, we first fitted naming × SA to evaluate its contribution to predict MCDI scores. Next, we fitted a second variable of naming × JA which provided information on how much more this additional variable contributed to the prediction given the first variable. As shown in Models 1 and 2 in Table 5, joint attention and sustained attention during parent-infant play when infants were 9 months of age jointly accounted for 61.5% of the variance in MCDI scores at age 12 months. Alone, sustained attention accounted for 60.2%. Adding joint attention increased this by only 0.8%. Reversing the order, joint attention alone accounted for 36.8% of the variance, and adding sustained attention increased this by 24.7%. Thus, sustained attention accounted uniquely for 24.7% of the variance, joint attention uniquely for 0.8%, with 36.1% accounted for by both jointly (because joint attention was correlated with sustained attention, $r = 0.69$, $p < 0.001$). As shown in Models 3 and 4 in Table 5, the same results were obtained when linking JA and SA measurements with MCDI scores at age 15 months. In summary,

TABLE 3 Means, standard deviations (*SD*), and ranges of naming, SA with JA, SA w/o JA and JA w/o SA measures as well as basic statistics of the interaction terms of the three quality measures

Measure	*M*	*SD*	Range
SA with JA	14.73%	8.72%	1.23%–36.77%
SA w/o JA	11.08%	6.04%	1.92%–20.81%
JA w/o SA	6.73%	5.54%	0%–24.04%
Naming × (SA with JA)	1.27	0.74	0.13–2.89
Naming × (SA w/o JA)	1.04	0.71	0.10–2.76
Naming × (JA w/o SA)	0.61	0.51	0–2.27

TABLE 4 Correlations between MCDI scores at age 12 months and 15 months with the three interaction terms based on the three quality measures from toy play at 9 months. The numbers in [] show 95% confidence intervals

	MCDI at 12 months	MCDI at 15 months
Naming × (SA with JA)	0.76*** [0.53, 0.89]	0.61*** [0.29, 0.81]
Naming × (SA w/o JA)	0.53*** [0.19, 0.77]	0.42* [0.12, 0.65]
Naming × (JA w/o SA)	0.09 [−0.30, 0.46]	0.07 [−0.32, 0.45]

Notes. *$p < .05$, **$p < .01$, ***$p < 0.005$.

both sustained attention and joint attention matter for later vocabulary growth, but, supporting our hypothesis, the attentional state of the learner—sustained attention—matters more. This overall pattern is consistent with the hypothesis that while joint attention is a context in which infant sustained attention to object often occurs, it is the learner's sustained visual attention to the named referent that is the proximal cause of learning.

4 | GENERAL DISCUSSION

Consequential differences in early word learning environments begin early. The present experiment provides three findings relevant to these early differences. First, the results indicate that early differences in word learning are related to the quality, not just the quantity, of parent naming events. Second, the results show that for early vocabulary growth, the attentional states of the mature partner who does the naming and of the infant who does the learning are both predictive factors. Finally, the results show that the most relevant measure of the quality of an object-naming event is the infant's sustained attention to the object during naming, not whether the parent and infant jointly attend to that object. The overall pattern of results suggests that parent naming within joint attention episodes predicts infant vocabulary because joint attention and infant sustained attention often coincide. But joint attention that does *not* include infant *sustained* attention to the object does not predict later vocabulary. In contrast, infant sustained attention to the named object when it is named—either with or without shared attention with the parent—does predict later vocabulary. The findings challenge well-accepted ideas about the role of joint attention in early vocabulary development and provide new insights into the role of the quantity and quality of input in the language learning environment. Finally, the results point to infant sustained attention—and the social and endogenous factors that underlie it—as an important source of early individual differences in word learning.

TABLE 5 Changes in variance accounted for depending on the order in which sustained attention (SA) and joint attention (JA) are added to regression models. Models 1 and 3 fit Naming × SA first, followed by Naming × JA. Models 2 and 4 fit Naming × JA first, followed by Naming × SA

	Additional variance account for			
	R^2	ΔR^2	*df*	*p*
Model 1 with 12-month MCDI				
Naming × SA	0.61	0.61	1,24	<0.001
Naming × JA	0.62	0.008	1,23	0.49
Model 2 with 12-month MCDI				
Naming × JA	0.37	0.37	1,24	0.001
Naming × SA	0.62	0.25	1,23	<0.001
Model 3 with 15-month MCDI				
Naming × SA	0.35	0.35	1,24	0.002
Naming × JA	0.36	0.01	1,23	0.53
Model 4 with 15-month MCDI				
Naming × JA	0.24	0.24	1,24	0.011
Naming × SA	0.36	0.12	1,23	0.035

4.1 | Early word learning in social contexts

Early word learning emerges in the context of tightly coupled social interactions between the early learner and a mature partner. Past work has shown that the social context plays a key role in young learners' prowess in early word learning. For example, joint attention between infants and parents has been extensively studied in the developmental literature because of overwhelming evidence that the ability to socially coordinate visual attention to an object is essential to language learning (Tomasello & Farrar, 1986; Tomasello & Todd, 1983). In this literature, the focus has been on the child side of joint attention with many powerful demonstrations of how social–interactional cues from parents guide infants' word learning. Most prior paradigms have used experiments with discrete trials to measure infants' ability to "read" the meaning of social cues such as eye gaze, head orientation, or pointing (Baldwin, 1993; Brooks & Meltzoff, 2002; Mundy & Newell, 2007; Woodward & Guajardo, 2002). In those experiments, the adult partner (usually the experimenter) is instructed to focus on the child and on effective teaching, and to provide clear and repeated signals of her attention to the object being named. Thus, the experimental paradigms focus on clean ways to assess infants' social skills, their sensitivity to social cues and their usage of social cues to link a label with its target referent. Using such paradigms, studies have shown that very young learners map nouns to objects only when the speaker is intentionally attending to the named object and not, for example, when there is an accidental co-occurrence of objects and names (Tomasello, 2000). Often the importance of social cues is interpreted in terms of children's understanding that words are used with an "intent to refer." Thus, children's early dependence on social cues is seen as a diagnostic marker of their ability to infer the intentions of the speaker. This kind of social cognition is called "mind reading" by Baron-Cohen (1995) or more generally "Theory of Mind" (Wellman & Liu, 2004) which is deemed to be central to language learning by some theorists (Tomasello, 2000).

The present findings suggest a different account. Parents and infants have different roles in supporting word learning. The parent's role is to name the object at moments conducive to infant learning. Hence, the degree to which the parent jointly attends to and names an object that the infant is attending is critical to word learning and vocabulary development. The infant's role is not to read the mental state and/or intention of the social partner but to learn the name. To do so, sustained attention to the named object when hearing the name may be most critical to associate and memorize the seen object with the heard word. Thus, for the infant, reading the mental state and/or intention of the social partner who label objects may not be the key factor. However, for the parent, reading the mental state of the learner to provide object names at the right moments is a key factor. This idea is well supported by research showing that parents' attentiveness and

responsivity to their infants' interest and attentional states supports language learning (Goldstein & Schwade, 2008; Tamis-LeMonda et al., 2014). Researchers of early word learning have distinguished two pathways to learning (Akhtar et al., 1991; Yu & Smith, 2017a, 2017b). One pathway termed "lead in" is by the parent naming objects of one's own interest and the infant reading the cues in the parent's behaviors to follow the parent's lead. The alternative "follow-in" pathway is for the parent to follow the infant's attention and name objects that infants already show interest. Even through different pathways, successful word learning requires the final state of the child's attention on target objects when hearing labels (Shimpi & Huttenlocher, 2007). Considerable research (Yu & Smith, 2012, 2016) suggests the second pathway—the parent follows the child's attention—is a more robust way to achieve the final learning state in free flowing parent–child interaction and therefore leads to more robust early word learning.

4.2 | Quantity and quality

Recent evidence shows that early differences in vocabulary growth among otherwise typically developing children are strongly related to differences in their language learning environments (Hart & Risley, 1999; Hoff, 2013; Newman et al., 2016; Ramirez-Esparza et al., 2014; Rowe, 2012; Weisleder & Fernald, 2013). One particularly strong predictor is the total words per unit time in child-directed speech at home (Golinkoff et al., 2015); a finding that has led to public health efforts directed to getting the parents of infants and young children to talk more (Reese, Sparks, & Leyva, 2010; Roberts & Kaiser, 2011), and public health initiatives such as Providence Talks, First 5 California, and Too Small to Fail. A body of recent research has also shown that the quality of parent speech is also critical for children's vocabulary. When parents talk to their children, they tend to speak differently compared with when talking to adults. At the speech level, child-directed speech contains unique acoustic properties, such as high pitch, short utterance and vowel alterations (Fernald, 1989; Golinkoff et al., 2015). At the language level, characteristics such as diversity of vocabulary and syntactical complexity (Hoff, 2003) as well as vocabulary sophistication (Rowe, 2012), have been shown to be predictive of later language outcomes.

Here, we show quality of a different kind matters, one that centers on the learner's visual attention. For very early learners who are acquiring first words, the attentional state of the learner may be particularly important. Early word learning requires infants to link what they hear with what they see and this requires that they attend to the target object when parents name it. Sustained attention—beyond a mere look to the target object—may be critical in building a robust memory of the object and its link to the word, memories that are strong enough to last and to be later retrieved from memory. Thus, more than the quality of parent talk may matter; the quality of the visual information as it engages the infant's real-time attention and memory processes will also be essential to learning words (Smith, Suanda, & Yu, 2014; Smith & Yu, 2013; Vlach & Johnson, 2013; Yu & Smith, 2011). The present results show that infant sustained attention at target objects named by parents at free play is the strongest predictor of later vocabulary size. More generally, our findings suggest that quality of the language learning input includes not only quality of parent talk, but quality of the attentional contexts wherein infants learn words. The parent's behavior—and joint attention—may be relevant because infant sustained attention to an object often occurs when parents show interest in and attend to that object as well. Parent interest and joint attention may actively extend infant attention as well as be the context in which parents name objects for their infants (Yu & Smith, 2016). These hypotheses are underdetermined by the extant data. However, the present results clearly indicate that we need to know how parents support sustained attention and how they can exploit it by choosing the optimal moments for naming objects. We also need to go beyond speech and language properties in child-directed speech and examine, more broadly, the nonlinguistic properties of the language learning environment.

Within a single naming event, learners must build a robust representation of the named object and its name. In light of the strong predictive relation between infant sustained attention during naming events and later vocabulary growth, we have suggested that infant sustained attention to the named object is a critical component in early word learning. Moreover, the predictors used are the interaction term combining the quantity of naming instances with quality of those individual instances, which can be conceptualized as aggregating information from multiple naming instances, each of which contributes more or less to learning dependent on its quality. Given the same object name heard in multiple times and multiple contexts, how do infants aggregate that information? Do they just use the high-quality information or the whole mix of higher and lower quality naming events? In our measures, naming instances with high quality contribute more to the final metric and therefore it is reasonable to assume that those naming instances are critical for learning. A relevant question, then, is whether a single or just a few high-quality naming events would be sufficient? This is a critical question for the quality–quantity issue because one parent who talks a lot might in some unit time have just as many high-quality naming events as a parent who talks much less, but the parent who talks more might have many low-quality naming events as well. In addition, the present results are based on a parent-report vocabulary measure (MCDI, etc.), which has been widely used in developmental research. Even though such parent report can provide valuable information of an overall assessment of child language, its usage is also limited by the parents' ability to provide an accurate report, compared with more direct observational assessment of children's lexical knowledge (Dale, 1991). One possibility is to use the looking-while-listening paradigm (Marchman & Fernald, 2008). With eye tracking, this paradigm can be used to measure not only whether children prefer to look at a target after hearing its label but also how fast they switch their attention as a measure of the processing speed of spoken word recognition.

4.3 | Sustained attention in early word learning

Sustained attention was traditionally thought of as an intrinsic property of the infant (Ruff et al., 1990) and by this traditional view, the

present results might be interpreted as showing that properties of the learner and not the learning environment are a major source of individual differences in early vocabulary. A complete account, however, is likely to be more complicated. Although infant intrinsic properties surely play a role (and can do so by influencing parent behavior), new findings suggest that infants' sustained attention to an object lasts longer when a mature social partner also attends to that object (Yu & Smith, 2016). Thus, the quality of social interactions may influence infant sustained attention in real time and be a causal factor in individual differences in infant sustained attention over developmental time.

Considerable evidence indicates that parent–infant social interactions with objects is a complex real-time dynamic system with each partner's behaviors determining the behaviors of the other (Chang, de Barbaro, & Deák, 2016; Suanda, Smith, & Yu, 2016; Yu & Smith, 2013, 2017b). The mature partner may lead, follow, and help infants sustain attention to an object (Yu & Smith, 2016). The immature partner may show clearer or less clear indicators of their own interests to the parent (Yu & Smith, 2017a). In the course of these real-time events, the parent chooses momentarily when to name objects and how often to name objects. Infant sustained attention to an object within a naming event elicits parent naming (Tamis-LeMonda, Kuchirko, & Tafuro, 2013). Many aspects of parent–infant joint behavior may conspire to support high-quality naming moments in which the infant sustains attention to the named object. Thus, many of the properties of the social interaction are likely to predict object name learning and vocabulary. However, sustained attention to the named object by the infant may be the final step that enables the *young* infant to build a memory for an individual object and its name.

Over developmental time, the complete story is likely more complex. Infants' *past* experiences of being engaged in joint attention with a parent may drive their ability to sustain attention within an individual naming moment (Yu & Smith, 2012). Parents' *histories* of attentiveness and responsiveness to their infants' interest and attentional states may determine how well they can judge and exploit their infant's current interest so as to name an object at the right moment for learning (Tamis-LeMonda et al., 2014). Infants are also learning the sentential frames surrounding different syntactic categories, which will become critical to the word learning process (Mintz & Gleitman, 2002; Waxman & Booth, 2001; Waxman & Leddon, 2002). When the developmental interdependencies are considered, joint attention and sustained attention to objects during naming events are unlikely to be independent or easily separable predictors to object name learning specifically, or to vocabulary development more generally. Given this, the value of focusing on precisely defined real-time microbehaviors such as gaze as potential predictors is twofold. First, the precise objective definitions allow for distinguishing behaviors that may often but not always coincide. In this way, we can separate causal pathways more directly, than through statistical approaches to co-varying predictors. In the long run, if we take objective measures of real-time behaviors longitudinally, we may be able to pull apart the developmental interdependencies and track with precision their unique contributions to different outcomes. Second, the focus on real-time behaviors, which bring us closer to real-time causes and effects, may help us locate actionable targets for intervention.

ACKNOWLEDGEMENTS

We thank Melissa Elston, Steven Elmlinger, Charlotte Wozniak, Melissa Hall, Charlene Tay and Seth Foster for collection of the data, Seth Foster and Tian (Linger) Xu for developing data management and processing software. Lauren Slone, Tian Xu, Drew Abney, Catalina Suarez, Sven Bambach, Yayun Zhang, and Maureen McQuillan for fruitful discussions. This work was funded by National Institutes of Health Grant R01HD074601 and R01HD093792 to CY and LBS, and also K99HD082358 to SHS.

REFERENCES

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

Akhtar, N., & Tomasello, M. (2000). The social nature of words and word learning. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello & G. Hollich (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 115–135). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195130324.001.0001

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. (1995). Understanding the link between joint attention and language. In: Joint Attention: Its origin and role in development, ed. C. Moore & P. Dunham pp. 131–158. Erlbaum.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the USA*, *109*(9), 3253–3258. https://doi.org/10.1073/pnas.1113380109

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

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences of the USA*, *110*(28), 11278–11283. https://doi.org/10.1073/pnas.1309518110

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

Dale, P. S. (1991). The validity of a parent report measure of vocabulary and syntax at 24 months. *Journal of Speech, Language, and Hearing Research*, *34*(3), 565–571. https://doi.org/10.1044/jshr.3403.565

Deák, G. O., Triesch, J., Krasno, A., de Barbaro, K., & Robledo, M. (2013). Learning to share: The emergence of joint attention in human infancy. In B. R. Kar (Ed.), *Cognition and brain development: Converging evidence from various methodologies* (pp. 173–210). Washington, DC: American Psychological Association. https://doi.org/10.1037/14043-000

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

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., & Reilly, J. S. (1993). *MacArthur Communicative Development Inventories: User's guide and technical manual*. Baltimore, WV: Paul H. Brookes.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, *60*, 1497–1510. https://doi.org/10.2307/1130938

Fernald, A., & Marchman, V. A. (2010). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and later-talking toddlers. *Child Development*, *83*(1), 203–222.

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, *16*(2), 234–248. https://doi.org/10.1111/desc.12019

Franchak, J., Kretch, K., Soska, K., Babcock, J., & Adolph, K. (2010). *Head-mounted eye-tracking of infants' natural interactions: A new method*. Paper presented at the Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, Austin, TX. https://doi.org/10.1145/1743666

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2016). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*, 677–694. https://doi.org/10.3758/s13423-013-0466-4

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

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby) Talk to me the social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science*, *24*(5), 339–344. https://doi.org/10.1177/0963721415595345

Hart, B., & Risley, T. R. (1999). *The social world of children: Learning to talk*. Baltimore, MD: P.H. Brookes.

Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator*, *27*(1), 4–9.

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., & Suma, K. (2015). The contribution of early communication quality to low-income children's language success. *Psychological Science*, *26*, 1071–1083. https://doi.org/10.1177/0956797615581493

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*(5), 1368–1378. https://doi.org/10.1111/1467-8624.00612

Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: Implications for closing achievement gaps. *Developmental Psychology*, *49*(1), 4–14. https://doi.org/10.1037/a0027238

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development*, *73*(2), 418–433. https://doi.org/10.1111/1467-8624.00415

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

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

Macroy-Higgins, M., & Montemarano, E. A. (2016). Attention and word learning in toddlers who are late talkers. *Journal of Child Language*, *43*(05), 1020–1037. https://doi.org/10.1017/S0305000915000379

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, *11*(3), F9–F16. https://doi.org/10.1111/j.1467-7687.2008.00671.x

Masten, A. S., & Cicchetti, D. (2010). Developmental cascades. *Development and Psychopathology*, *22*(3), 491–495. https://doi.org/10.1017/S0954579410000222

Masten, A. S., Roisman, G. I., Long, J. D., Burt, K. B., Obradović, J., Riley, J. R., & Tellegen, A. (2005). Developmental cascades: Linking academic achievement and externalizing and internalizing symptoms over 20 years. *Developmental Psychology*, *41*(5), 733–746. https://doi.org/10.1037/0012-1649.41.5.733

Masur, E. F., Flynn, V., & Eichorst, D. L. (2005). Maternal responsive and directive behaviours and utterances as predictors of children's lexical development. *Journal of Child Language*, *32*(1), 63–91. https://doi.org/10.1017/S0305000904006634

Mintz, T. H., & Gleitman, L. R. (2002). Adjectives really do modify nouns: The incremental and restricted nature of early adjective acquisition. *Cognition*, *84*(3), 267–293. https://doi.org/10.1016/S0010-0277(02)00047-1

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current Directions in Psychological Science*, *16*(5), 269–274. https://doi.org/10.1111/j.1467-8721.2007.00518.x

Murphy, P. K., Rowe, M. L., Ramani, G., & Silverman, R. (2014). Promoting critical-analytic thinking in children and adolescents at home and in school. *Educational Psychology Review*, *26*(4), 561–578. https://doi.org/10.1007/s10648-014-9281-3

Newman, R. S., Rowe, M. L., & Ratner, N. B. (2016). Input and uptake at 7 months predicts toddler vocabulary: The role of child-directed speech and infant processing skills in language development. *Journal of Child Language*, *43*(5), 1158–1173. https://doi.org/10.1017/S0305000915000446

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

Ramirez-Esparza, N., Garcia-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science*, *17*(6), 880–891. https://doi.org/10.1111/desc.12172

Reese, E., Sparks, A., & Leyva, D. (2010). A review of parent interventions for preschool children's language and emergent literacy. *Journal of Early Childhood Literacy*, *10*(1), 97–117. https://doi.org/10.1177/1468798409356987

Roberts, M. Y., & Kaiser, A. P. (2011). The effectiveness of parent-implemented language interventions: A meta-analysis. *American Journal of Speech-Language Pathology*, *20*(3), 180–199. https://doi.org/10.1044/1058-0360(2011/10-0055)

Rollins, P. R. (2003). Caregivers' contingent comments to 9-month-old infants: Relationships with later language. *Applied Psycholinguistics*, *24*(2), 221–234.

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

Rowe, M. L., & Goldin-Meadow, S. (2009). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science*, *323*(5916), 951–953. https://doi.org/10.1126/science.1167025

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

Salley, B., Panneton, R. K., & Colombo, J. (2013). Separable attentional predictors of language outcome. *Infancy*, *18*(4), 462–489. https://doi.org/10.1111/j.1532-7078.2012.00138.x

Shimpi, P. M., & Huttenlocher, J. (2007). Redirective labels and early vocabulary development. *Journal of Child Language*, *34*(4), 845–859.

Smith, L. B. (2013). It's all connected: Pathways in visual object recognition and early noun learning. *American Psychologist*, *68*(8), 618–629. https://doi.org/10.1037/a0034185

Smith, L. B., Suanda, S. H., & Yu, C. (2014). The unrealized promise of infant statistical word–referent learning. *Trends in Cognitive Sciences*, *18*(5), 251–258. https://doi.org/10.1016/j.tics.2014.02.007

Smith, L. B., & Yu, C. (2013). Visual attention is not enough: Individual differences in statistical word-referent learning in infants. *Language Learning and Development*, *9*(1), 25–49. https://doi.org/10.1080/15475441.2012.707104

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

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

Tomasello, M. (2000). The social-pragmatic theory of word learning. *Pragmatics*, *10*(4), 401–413. https://doi.org/10.1075/prag

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). In search of the uniquely human. *Behavioral and Brain Sciences*, *28*(5), 721–727.

Tomasello, M., & Farrar, M. (1986). Joint attention and early language. *Child Development*, *57*(6), 1454–1463. https://doi.org/10.2307/1130423

Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First Language*, *4*(12), 197–212. https://doi.org/10.1177/014272378300401202

Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, *127*(3), 375–382. https://doi.org/10.1016/j.cognition.2013.02.015

Waxman, S. R., & Booth, A. E. (2001). Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology*, *43*(3), 217–242. https://doi.org/10.1006/cogp.2001.0764

Waxman, S. R., & Leddon, E. M. (2010). Early word-learning and conceptual development. In U. Goswami (Ed.), The WileyBlackwell handbook of childhood cognitive development (pp. 180–208). Wiley-Blackwell.

Weisleder, A., & Fernald, A. (2013). Talking to children matters early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–2152. https://doi.org/10.1177/0956797613488145

Woodward, A. L., & Guajardo, J. J. (2002). Infants' understanding of the point gesture as an object-directed action. *Cognitive Development*, *17*(1), 1061–1084. https://doi.org/10.1016/S0885-2014(02)00074-6

Yu, C., & Smith, L. B. (2011). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental Science*, *16*(2), 165–180. https://doi.org/10.1111/j.1467-7687.2010.00958.x

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

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

Yu, C., & Smith, L. B. (2016). The social origins of sustained attention in one-year-old human infants. *Current Biology*, *26*(9), 1235–1240. https://doi.org/10.1016/j.cub.2016.03.026

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

Yu, C., & Smith, L. B. (2017b). Multiple sensory-motor pathways lead to coordinated visual attention. *Cognitive Science* , *41*, 5–31. https://doi.org/10.1111/cogs.12366

How to cite this article: Yu C, Suanda SH, Smith LB. Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Dev Sci*. 2018;e12735. https://doi.org/10.1111/desc.12735

Graphical Abstract

The contents of this page will be used as part of the graphical abstract of html only. It will not be published as part of main.

XXX.