Developmental Psychology

Intersensory Matching of Faces and Voices in Infancy Predicts Language Outcomes in Young Children

Elizabeth V. Edgar, James Torrence Todd, and Lorraine E. Bahrick

Online First Publication, April 21, 2022. http://dx.doi.org/10.1037/dev0001375

CITATION

Edgar, E. V., Todd, J. T., & Bahrick, L. E. (2022, April 21). Intersensory Matching of Faces and Voices in Infancy Predicts Language Outcomes in Young Children. *Developmental Psychology*. Advance online publication. http://dx.doi.org/10.1037/dev0001375

AMERICAN PSYCHOLOGICAL ASSOCIATION

Developmental Psychology

© 2022 American Psychological Association
ISSN: 0012-1649

https://doi.org/10.1037/dev0001375

Intersensory Matching of Faces and Voices in Infancy Predicts Language Outcomes in Young Children

Elizabeth V. Edgar, James Torrence Todd, and Lorraine E. Bahrick
Department of Psychology, Florida International University

Parent language input is a well-established predictor of child language development. Multisensory attention skills (MASks; intersensory matching, shifting and sustaining attention to audiovisual speech) are also known to be foundations for language development. However, due to a lack of appropriate measures, individual differences in these skills have received little research focus. A newly established measure, the Multisensory Attention Assessment Protocol (MAAP), allows researchers to examine predictive relations between early MASks and later outcomes. We hypothesized that, along with parent language input, multisensory attention to social events (faces and voices) in infancy would predict later language outcomes. We collected data from 97 children (predominantly White and Hispanic, 48 males) participating in an ongoing longitudinal study assessing 12-, 18-, and 24-month MASks (MAAP) and parent language input (quality, quantity), and 18- and 24-month language outcomes (child speech production, vocabulary size). Results revealed 12-month intersensory matching (but not maintaining or shifting attention) of faces and voices in the presence of a distractor was a strong predictor of language. It predicted a variety of 18- and 24-month child language outcomes (expressive vocabulary, child speech production), even when holding traditional predictors constant: parent language input and SES (maternal education: 52% bachelor's degree or higher). Further, at each age, parent language input predicted just one outcome, expressive vocabulary, and SES predicted child speech production. These novel findings reveal infant intersensory matching of faces and voices in the presence of a distractor can predict which children might benefit most from parent language input and show better language outcomes.

Keywords: multisensory attention skills, intersensory matching, parent language input, child language outcomes, audiovisual speech perception

Supplemental materials: https://doi.org/10.1037/dev0001375.supp

Parent language input is a well-established predictor of child language, and individual differences in quantity and quality of language input directed to children predict individual differences in child vocabulary development (Hart & Risley, 1995; Huttenlocher et al., 1991; Rowe, 2008). Greater number and diversity of words directed to a child (a measure of language quantity and quality, respectively) is associated with greater vocabulary development (Hoff & Naigles, 2002; Huttenlocher et al., 2010; Weisleder & Fernald, 2013; Weizman & Snow, 2001). On average, children from higher-SES families hear more words than children from lower-SES families (Gilkerson et al., 2017; Purpura, 2019), although there is variability within SES groups (Huttenlocher et al., 1991; Pan et al., 2005). In contrast to these well-established predictors of child language, the role of early developing attention skills in predicting language outcomes has received much less research focus (for exceptions see, Colombo et al., 2004; Rose et al., 2009). This is due, in large part, to the lack of reliable and fine-grained individual difference measures, particularly for assessing attention to dynamic audiovisual events, such as faces and voices during speech, the context most relevant for language acquisition.

To address this need, Bahrick et al. (2018) developed the Multisensory Attention Assessment Protocol (MAAP), a new measure designed to assess individual differences in three "multisensory attention skills" (MASks; sustaining attention, shifting/disengaging attention, and matching synchronous sights and sounds) of both audiovisual social (speech) and nonsocial (object) events. Findings demonstrate that 2- to 5-year-olds with greater sustained attention to faces and better intersensory matching of faces and voices (e.g., selective attention and perceptual processing of

Elizabeth V. Edgar https://orcid.org/0000-0003-4419-1876
Lorraine E. Bahrick https://orcid.org/0000-0002-4126-7065

This study was not preregistered. Data and study materials are not currently available online but will be available in the future. This research was supported by National Institute of Child Health and Human Development Grants R01HD094803 and R01HD053776 awarded to Lorraine E. Bahrick. We thank the children and families who participated in this research. We also thank all of the research assistants in the Infant Development Lab for their help in transcription, including Amanda Sanchez and Melany Gomez.

Correspondence concerning this article should be addressed to Elizabeth V. Edgar or Lorraine E. Bahrick, Department of Psychology, Florida International University, 11200 SW 8th Street, Miami, FL 33199, United States. Email: eedga001@fiu.edu or bahrick@fiu.edu

face–voice synchrony) showed better receptive and expressive language. This study also demonstrated the feasibility of using the MAAP with 12-month-old infants. Together, these findings paved the way for using the MAAP to assess prospective relations between early basic skills in attending to audiovisual events and later language outcomes.

Here, we examine prospective relations between MASks at 12 months and language outcomes at 18 and 24 months. We assess whether MASks predict language outcomes, even after holding constant well-known predictors of language including parent language input (quality and quantity) and socioeconomic status (SES). To what extent does selective attention to faces and voices during natural, synchronous, audiovisual speech predict language outcomes given comparable levels of parent language input and SES? We propose that attention to faces and in particular, intersensory processing of faces and voices, is a critical process through which infants benefit and learn from parent language input and other language learning opportunities. A number of studies indirectly support this proposal (Bahrick et al., 2018; Gogate & Bahrick, 1998; Gogate & Hollich, 2010; Reynolds et al., 2014). However, no studies have assessed relations between parent language input and children's MASks, or explored the unique and overlapping pathways from these attention skills to child language outcomes.

Attention as a Foundation for Child Language Development

Attention is the foundation for all we perceive, learn, and remember (e.g., Bahrick & Lickliter, 2012). What we attend to in a given moment or interaction is shaped by experience and forms the foundation for what we perceive and in turn, learn and remember, and this informs what is attended to in the next cycle of attention and perception. Given that the dynamic, multisensory environment provides far too much stimulation to attend to at any one time, perceivers must learn to selectively attend to a small portion of information—information that is meaningful and relevant to their needs and goals—while filtering out information that is less relevant (Bahrick et al., 2020).

This challenging task is heavily scaffolded by interaction with caretakers during infancy (Gogate et al., 2001; Mundy & Burnette, 2005). Dynamic faces of people speaking are highly salient to young infants (e.g., Bahrick et al., 2016; Courage et al., 2006). Focusing attention on the face of a speaker provides a rich source of language learning opportunities for infants. Infants must learn to detect eye gaze direction and gesture (signaling which object is being labeled), which enhance joint attention and word mapping (for a review, see Flom et al., 2007). They also detect movements of the face and brow during speech which convey affect, prosody, and communicative intent (Bahrick et al., 2019; Shepard et al., 2012).

Moreover, a significant body of research demonstrates the importance of selective attention to face–voice synchrony for promoting language learning. Foundational properties necessary for language learning are detected more easily and earlier in development when they are presented in synchronous audiovisual speech than in auditory speech alone or in asynchronous audiovisual speech, including: speech segmentation and recognition of familiarized words (Hollich et al., 2005), detection of prosodic patterns specifying approval versus prohibition (Bahrick et al., 2019), detection of object–label relations (e.g., Gogate & Bahrick, 1998), social referencing (Vaillant-Molina & Bahrick, 2012), and perception of affect in speech (happy, sad, angry; Flom & Bahrick, 2007; see Walker-Andrews, 1997 for a review). Moreover, neural and heart-rate evidence also support the view that face–voice synchrony leads to better processing and heightened attention. For example, synchronous face–voice events are more salient and processed more efficiently than the voices alone or the same faces and voices presented out of synchrony according to measures of ERPs (Hyde et al., 2011; Reynolds et al., 2014). Further, audiovisual synchrony promotes earlier onset of sustained attention (optimal for learning) and deeper sustained attention according to heart-rate defined measures of attention (Curtindale et al., 2019). Thus, selective attention to the synchronous face and voice of a speaker fosters a variety of skills critical for language learning, as well as deeper processing than attending to the voice alone or to asynchronous face-voice combinations. We thus expect that selective attention to synchronous face-voice events will foster intersensory processing of these events and, in turn, predict language outcomes.

Thus, there are a variety of paths through which selective attention to face–voice events can foster better language development. Longer sustained attention to faces fosters deeper processing (Courage et al., 2006; Shaddy & Colombo, 2004) and, in turn, earlier detection of face–voice synchrony during speech. Detection of face–voice synchrony in turn leads to deeper and more efficient processing of the speech event (Curtindale et al., 2019; Reynolds et al., 2014), optimizing learning from language input, and ultimately leading to more efficient word learning/mapping. Further, better MASks may also promote a social feedback loop (Warlaumont et al., 2014), promoting greater learning from parent language input, as well as eliciting more diverse and frequent parent language input, which in turn foster better MASks. Infants' selective attention to faces and voices in the context of interactions with parents also likely facilitates their engagement in joint attention episodes and object labeling events in both monolingual and bilingual children (Piazza et al., 2020; Ramírez-Esparza et al., 2017; Trueswell et al., 2016). Thus, enhanced multisensory attention to faces and voices should allow infants to both elicit and take advantage of language learning opportunities provided by parent language input.

Methods for Assessing Multisensory Attention

A range of methods have been used to assess selective attention skills in preverbal children, primarily appropriate for group-level data analyses rather than for assessing differences across individual children. The intermodal preference method was the earliest and most popular method for assessing attention to dynamic audiovisual events. Infants must selectively attend to a sound-synchronous event while ignoring a concurrent asynchronous event (e.g., Bahrick, 1983, 1988; Spelke, 1976). Bahrick et al. (1981) found that the audiovisual synchrony between the sights and sounds of object movement was so salient to 4-month-old infants that it allowed them to selectively attend to a sound-synchronous event (e.g., hands clapping) while ignoring a visually superimposed event (e.g., hands manipulating a toy slinky). More recently, eye-tracking paradigms have been used to examine selective visual attention to faces of people speaking (Lewkowicz & Hansen-Tift, 2012; Tenenbaum et al., 2015; Tsang et al., 2018). Studies have shown a developmental shift from attention to the eyes to the mouth toward the end of the first year, a time when interest in the native language is emerging (Lewkowicz & Hansen-Tift, 2012). Others, using head-mounted eye trackers to assess individual differences in selective attention to objects children were holding and moving, have successfully predicted word learning in toddlers (Yu et al., 2019).

A variety of indices of selective attention have also been developed for older, verbal children capable of understanding instructions. Examples include the Track-It task, which predicts learning outcomes in kindergartners (Erickson et al., 2015; Fisher et al., 2013), dichotic listening tasks (Isbell et al., 2016), and executive functioning tasks including Card Sort (Zelazo et al., 1996) and Flanker tasks (Eriksen & Eriksen, 1974). However, comparisons across age and studies are challenging given the wide range of methods and stimuli used. Further, few can be used across infancy and early childhood or assess sufficiently fine-grained individual differences capable of revealing developmental trajectories and pathways to later outcomes.

Thus, there is little direct evidence of links between selective attention to face–voice synchrony during audiovisual speech and language outcomes. The infancy research reviewed above was primarily conducted using a group differences approach, contrasting groups of infants of different ages or who received different experimental conditions. Further, many of these studies used "coarse-grained" measures of attention (e.g., total time to reach habituation, novelty preference) in which scores were averaged across a small number of trials. As a result, these measures are not sufficiently fine-grained or reliable for deriving a score for an individual child or for assessing relations with developmental outcomes. Unlike research in the domain of language learning, which has long benefited from measures of individual differences across children, there were no individual difference measures designed to quantify attention to audiovisual speech events, the context most relevant to language development. Direct evidence of relations between attention skills and language outcomes requires assessing how attention skills vary across individual infants and then correlating those skills with individual differences in language outcomes.

Multisensory Attention Assessment Protocol (MAAP)

To address this need, we developed the Multisensory Attention Assessment Protocol (MAAP) which assesses individual differences in the three MASks,[1] sustaining attention, shifting/disengaging attention, and intersensory matching of audiovisual social (faces and voices) and nonsocial (object) events (Bahrick et al., 2018). Each skill is assessed during conditions of high and low competing stimulation (presence or absence of a central visual distractor). It requires no verbal responses or understanding of verbal instructions, and thus is appropriate for assessing nonverbal and preverbal infants as young as 3 months of age. Traditional paradigms, such as the intermodal preference method, are inappropriate for assessing fine-grained individual differences because they include a small number of trials (1 or 2 trials) and their psychometric properties have not been established. In the few studies where they have been assessed, they have shown small to moderate test–retest reliabilities (e.g., see Colombo et al., 2004). Further, they typically include only one condition, and provide just one measure (e.g., preference for the sound-synchronous display). In contrast, the MAAP is more appropriate and useful as a measure of individual differences because it (a) includes many trials (24 trials; 12 social and 12 nonsocial) providing a more stable mean capable of indexing meaningful variability across and within participants, (b) has multiple conditions (social and nonsocial events presented in the context of high and low competing stimulation), (c) provides three different measures of attention, and (d) shows good test–retest reliability and strong internal consistency (Bahrick et al., 2018).

The skills assessed by the MAAP were developed to approximate children's attention to social and nonsocial events in their multisensory, dynamic, natural learning environment. For example, sustaining attention requires focusing on salient events (e.g., faces and voices of caregivers) while ignoring distractors (e.g., TV), shifting/disengaging requires disengaging attention from an immediate attentional focus to shift to a salient event (the face of a person speaking) and intersensory matching requires visually attending to source of a sound in the presence of distractors (competing sounds, events). These skills can reflect both exogenous (bottom up, reflexive) and endogenous (top down control) attention, depending on attention control, competing stimulation, and task demands (see Colombo & Cheatham, 2006, for discussion). By assessing attention in the context of dynamic, audiovisual events, in the presence of competing visual stimulation, measures derived from the MAAP come closer to reflecting real-life social attention skills than many other screen-based measures (using static images, silent events, and no competing stimulation) available for children.

Multisensory Attention Skills Predict Language Outcomes

Only a few studies thus far have investigated intersensory processing as a predictor of language outcomes. For example, the detection of, and longer looking to, audiovisual synchrony was related to greater receptive language in children with autism spectrum disorders (Patten et al., 2014; Todd & Bahrick, 2022). Greater attention to the mouth during audiovisual speech indexed by eye-tracking has been shown to predict greater expressive vocabulary in typically developing children (Tenenbaum et al., 2015; Tsang et al., 2018). In contrast, the MAAP provides fine-grained assessments of individual differences in three MASks simultaneously and can characterize developmental pathways from these three basic attention skills to language outcomes.

We recently found evidence for a model characterizing developmental pathways between MASks for social events and language: sustained attention to audiovisual social events (women speaking) predicted accuracy of intersensory matching (face–voice synchrony) for these events, which in turn predicted receptive and

[1] Although the terms *multisensory* and *intersensory* are often used interchangeably, here we use *multisensory* as a general term to refer to stimulation impacting more than one sensory system (e.g., auditory, visual, proprioceptive, etc.). It serves as a name for our protocol (MAAP) and for the collection of the three attention skills it measures (*multisensory attention skills* (MASks): sustaining attention, shifting/disengaging, intersensory matching). In contrast, the term *intersensory* is more specific and is used here to refer to just one of these skills—intersensory matching. Intersensory matching is the detection of information that is common across auditory and visual stimulation such as synchrony, rhythm, tempo, or intensity patterns. We also refer to intersensory processing as the activity of perceiving, integrating, and further processing this information.

expressive language in 2- to 5-year-old children (Bahrick et al., 2018). In contrast, there was no evidence that attention to the nonsocial events (i.e., objects striking a surface) predicted language outcomes within that age range. Attention to social events (but not nonsocial events) also differentiated typically developing children from those with autism spectrum disorder (ASD), and shorter attention and longer disengagement to faces predicted poorer language outcomes and greater symptom severity in children with ASD (Todd & Bahrick, 2022). Using the MAAP, researchers can thus characterize MASks in individual participants at a sufficiently fine-grained level so that performance can be meaningfully correlated with the participant's achievements in other domains. This allows us to explore how early MASks to social (and/or nonsocial) events serve as building blocks for later language outcomes in individual children.

The Present Study

As demonstrated by the preceding review, both parent language input and child MASks are foundations for child language development. However, child MASks have not been examined in relation to parent language input as predictors of child language outcomes. Thus, it is not known the extent to which individual differences in MASks contribute to child language outcomes independent of parent language input and SES. Further, given evidence from our prior studies using the MAAP (Bahrick et al., 2018; Todd & Bahrick, 2022) that attention to social (but not nonsocial) events predicted language outcomes, we focus specifically on infant attention to social events as assessed by the MAAP in the present study. We predict that better multisensory attention to social events (faces and voices) is an important means by which children benefit from language learning opportunities, enhancing attention and audiovisual processing of linguistic input (Bahrick et al., 2018; Bahrick & Lickliter, 2012).

The primary goal of the present study was to examine the unique contributions of MASks for social events and parent language input at 12, 18, and 24 months in predicting child language outcomes at 18 and 24 months. Since SES also predicts child language outcomes, we assessed maternal education as a proxy for SES and controlled for it in all our analyses. Specifically, we assessed the unique contribution of each variable in predicting language outcomes (while holding all other predictors constant). Table 1 summarizes all the measures used.

Consistent with prior research (e.g., Bahrick et al., 2018), we predicted that child MASks (accuracy of intersensory matching, sustained attention, and reaction time [RT] to shift to social events) and parent language input (quantity and quality) at 12, 18, and 24 months would predict child speech production (quantity and quality) and child vocabulary size (expressive and receptive) within age (e.g., 18- and 24-month MASks predict 18- and 24-month child language) and prospectively (e.g., 12-month MASks predict 18- and 24-month child language). Second, we explored whether there was a relationship between child MASks for social events and parent language input (quantity and quality). Third, given similar levels of parent language input and maternal education, children with greater multisensory attention to social events should benefit most from language learning opportunities provided by parent language input. Thus, we predicted that child MASks for social events would predict language outcomes when controlling for parent language input and maternal education.

Method

Participants

One-hundred six infants who were enrolled in a larger longitudinal study assessing the development of multisensory attention skills (MASks) and language, social, and cognitive outcomes, participated. The longitudinal study, entitled "Development of Intermodal Perception of Social and Nonsocial Events," received IRB approval from the Social and Behavioral Review Board of Florida International University (IRB-13–0448-CR06). The final sample consisted of 97 infants who had data for at least two of the three variables assessed in this study (MASks, parent language input, child language outcomes). Infants were assessed at 12, 18, and 24 months. Demographic information for the sample can be found in Table 2. For a summary of the assessments administered at each age and dependent variables, see Table 1.

Child Multisensory Attention Measures: MAAP

Apparatus and Equipment

The MAAP was presented on a 46-in. widescreen monitor (NEC Multisync PV61). An experimenter sat behind the child and presented the stimuli from a computer using a custom MatLab based program. Infants were seated approximately 40-in. from the

Table 1

Constructs, Assessments Used to Index Each Construct, Ages Administered, and Dependent Variables

Construct	Protocol/assessment	Ages	Dependent variables
Parent language input	Parent–Child Interaction (PCI)	12, 18, 24 months	Quantity–tokens Quality–types
Child multisensory attention skills	Multisensory Attention Assessment Protocol (MAAP)	12, 18, 24 months	Sustained attention Intersensory matching Speed of shifting
Child speech production	Parent–Child Interaction (PCI)	18, 24 months	Quantity–tokens Quality–types
Child vocabulary size	Mac-Arthur Bates Communicative Development Inventory (MB-CDI)	18, 24 months	Expressive vocabulary Receptive vocabulary

Table 2
Demographic Information for the Sample (N = 97)

Variable	*N*	Percentage
Gender		
Male	48	49.48%
Female	49	50.52%
Ethnicity		
Hispanic	60	61.86%
Non-Hispanic	35	36.08%
Did not disclose	2	2.06%
Race		
White/European-American	63	64.95%
Black/African-American	16	16.49%
Asian/Pacific Islander	3	3.09%
More than one race	8	8.25%
Other/Did not disclose	7	7.22%
Maternal education		
High school or equivalent	13	13.40%
Some college	14	14.43%
Associate's degree	15	15.46%
Bachelor's degree	25	25.77%
Master's degree or higher	26	26.80%
Did not disclose	4	4.12%
Home language		
English	62	63.92%
Spanish	29	29.90%
Both English and Spanish	2	2.06%
Other	4	4.12%
Age	*M*	*SD*
12-month visit	12.21	1.41
18-month visit	18.05	0.42
24-month visit	24.19	0.37

display on their caregiver's lap. Caregivers were blind to which side of the screen depicted the sound-synchronous event (they wore black-out glasses).

Stimulus Events and Procedure

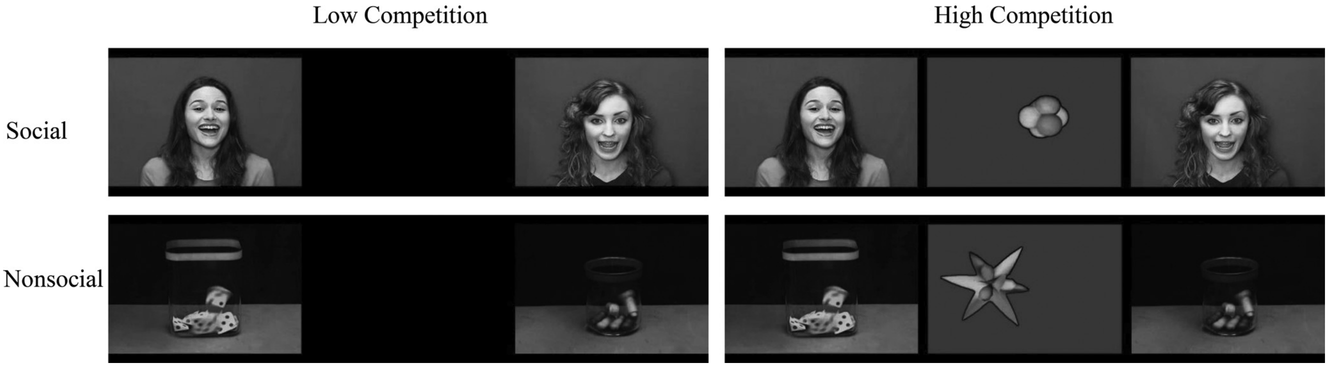
The MAAP (Bahrick et al., 2018) is a three-screen video procedure assessing three basic indices of attention in the context of dynamic, audiovisual social and nonsocial events. Social events depict women telling stories using infant-directed speech, and nonsocial events depict small wooden objects dropping into a container in an erratic temporal pattern (see Figure 1). Example stimulus videos can be seen on Databrary (https://nyu.databrary.org/volume/326). There are two distinct blocks of social and nonsocial events (12 trials each; 24 trials total). The blocks were designed to be used separately depending on the research questions of the study. In the present study, we focus on attention to the social events, given our interest in relations between attention to faces and voices and language outcomes. Each trial begins with a silent, central, 3-s dynamic visual event (distractor event) depicting morphing geometric shapes, followed by two 12-s lateral events. The lateral events (left and right sides of the display) consist of two social or two nonsocial events. The movements of one of the lateral events are synchronous with its natural soundtrack, while the movements of the other are asynchronous. For half of the trials within each block (6 trials), the central distractor event remains on throughout the lateral events providing an additional source of competing stimulation (high-competition trials; see right side, Figure 1). For the other half of the trials (6 trials), it disappears at the onset of the lateral events (low-competition trials; see left side, Figure 1). The visual distractor event simulates the competing stimulation infants and young children experience in the natural multimodal environment of overlapping events.

The experimenter viewed the child through a front-facing camera (SONY FDR-AX33) hidden above the widescreen monitor. Trained observers, hidden behind the monitor via a black curtain, were blind to left-right location of the sound-synchronous event (they could hear the soundtrack but could not see the video displays), and coded infant fixations to the left, center, and right sides of the screen on a game pad in real-time. For additional details on MAAP stimuli, procedure, and counterbalancing, see Bahrick et al. (2018; pp. 2216–2217).

MAAP Measures

The MAAP assesses three MASks, sustained attention (duration), intersensory matching (accuracy), and shifting/disengaging (speed), each in the context of both high and low competing stimulation (i.e., distractor present vs. absent). Sustained attention to

Figure 1
Static Images of the Dynamic Audiovisual Events From the Multisensory Attention Assessment Protocol (MAAP)



Note. On all trials, a 3-second central stimulus (computerized geometric shape) was followed by two side-by-side lateral events (social, nonsocial), one of which was synchronous with its appropriate soundtrack. On low-competition trials (images on left), the central stimulus was turned off during the lateral events, whereas on high-competition trials (images on right), the central stimulus remained on during the lateral event. The individuals whose faces appear here gave signed consent for their likenesses to be published in this article.

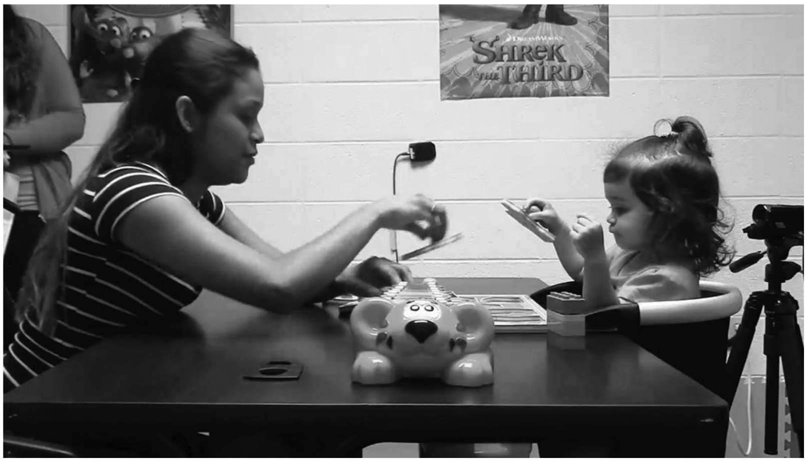
social events (PALT; proportion of available looking time to the audiovisual events) simulates attention to audiovisual speech events in the natural environment. Longer sustained attention to speech events in the face of distraction allows more time for processing audiovisual speech, and, potentially, more time for word learning. PALT was calculated for each trial by dividing the total looking time to both lateral events by the length of the trial. Intersensory matching of social events (PTLT; proportion of total looking time to the sound-synchronous lateral event) assesses the infant's ability to match faces with their synchronous voices and facilitates perceiving the event as a whole and further processing of speech events. PTLT was calculated for each trial by dividing the looking to the audiovisual synchronous event by the total looking time to both the synchronous and asynchronous events. During the high-competition trials, PTLT reflects the ability of children to match audible and visible speech in the face of distraction, much like what occurs in their natural environment. The speed of shifting/disengaging to a social event (RT) assesses how quickly children disengage from the distractor (high-competition trials) or shift their attention (low-competition trials) to look to the face of a person speaking. RT was calculated for each trial as the latency to shift attention in seconds from the central stimulus to either of the two lateral events. Faster RT disengaging from a distracting event to an audiovisual speech event allows the child more time to process the event as a whole. Interobserver reliability was assessed by having a second observer record the looking for a portion of the infants (51% of the sample at 12 months, 41% at 18 months, and 36% at 24 months). Pearson correlation coefficients for the primary and secondary observer were: sustained attention: .92 at 12 months, .95 at 18 months, and .94 at 24 months; intersensory matching: .93 at 12 months, .91 at 18 months, and .90 at 24 months; and speed of shifting: .95 at 12 months, .99 at 18 months, and .98 at 24 months.

Parent Language Input and Child Speech Production Measures: PCI

We measured parent language input and child language production during a short lab-based parent–child interaction (PCI) at 12, 18, and 24 months (M = 8.05 minutes, range: 4.05–9.17 minutes). Short, structured interactions have been shown to capture the everyday language experience (Tamis-LeMonda et al., 2017). We obtained measures of quantity (tokens; total number of words) and quality (types; number of different words; diversity) of parent language input at 12, 18, and 24 months, as well as measures of quantity and quality of child speech production at 18 and 24 months. The parent and child were seated across from one another at a table (40 × 28 in., see Figure 2). At 12 and 18 months, children sat in a seat that clamped to the edge of the table. At 24 months, they sat on a booster seat attached to a chair. Three toys were provided (a wooden puzzle, Legos, and a toy piano) to elicit speech and interaction between the parent and child. Parents were instructed to interact with their infant normally as they would at home. For details about camera placement for recordings, please see the online supplemental material, p. 1.

Parent and child speech during the PCI was transcribed by trained research assistants. Transcription reliability was established by having a second trained research assistant check the transcription. Disagreements between the two research assistants were decided by a third research assistant, blind to the topic of the disagreement. Transcriptions were analyzed using the Child Language Data Exchange Systems (CHILDES; MacWhinney, 2000) FREQ program to calculate the quantity (tokens; total number of words spoken) and quality (types; total number of different, or unique, words spoken) of parent language input at 12, 18, and 24 months and child language production at 18 and 24 months (but not at 12 months, as there were too few instances of child word production). To equate across interactions of slightly different

Figure 2
Parent–Child Interaction (PCI): Side View of the Infant Seated Facing Their Parent



Note. At each age, dyads received three toys (a) a wooden puzzle with eight cut-out shapes depicting a farmer and farm animals, (b) a toy piano with eight colored keys, and (c) four large plastic, colored blocks that interconnected (Legos). The authors received signed consent for the caregiver's and child's likenesses to be published in this article.

durations, both types and tokens were divided by the duration of the interaction to generate a per-minute ratio. For details about conceptualizations of quantity and quality of parent language input, please see the online supplemental material, pp. 1–2.

Child Vocabulary Size Measure: MB-CDI

At the 18- and 24-month visits, caregivers completed a parent-report measure of vocabulary, the MacArthur-Bates Communicative Development Inventory (MB-CDI) in English (Jackson-Maldonado et al., 2003), Spanish (Jackson-Maldonado et al., 2003), or both, depending on parental report of the child's primary language (for details, see online supplemental material, p. 2).

This study was not preregistered. Data are currently available online at https://nyu.databrary.org/volume/1410

Results

Data Analysis Overview

Primary analyses consisted of multiple regressions. We first conducted bivariate correlations to inform our regression analyses. With a sample size of $N = 97$, for bivariate correlations there is sufficient power to detect a medium effect size of $r = .28$ or greater (assuming a β of .80 and a two-tailed p value of .05), and for multiple regressions to detect a nonzero path coefficient that accounts for 6% unique variance (assuming a β of .80, a two-tailed p value of .05, four predictors, and an R^2 of .30).

Full information maximum likelihood (FIML) estimation was used for all analyses, using the maximum likelihood estimator in MPlus. Missing data ranged from 21.5% (12-month quantity and quality of parent language input) to 52.3% (18- and 24-month MB-CDI vocabulary). Data were missing primarily because infants did not participate in all longitudinal visits, or because parent-report forms were not returned (see Tables S1 and S2 in the online supplemental material for detail). We tested for mechanisms of missingness to ensure that data were not missing in a systematic way. Various techniques were used (for example, t-tests, logistic regression, Little's MCAR test) in the missing value analyses. Analyses supported the conclusion that data were missing at random (MAR; Rubin, 1976) justifying the use of FIML.[2] Supplemental analyses were also conducted with traditional approaches for dealing with missing data (omitting missing data using listwise deletion; that is, without FIML). Despite the fact that sample size decreased, all major conclusions drawn from the data were identical with and without FIML (for details, see Tables S19 and S20 in the online supplemental material. This further supports our use of FIML.

Secondary analyses were conducted to assess the influence of language spoken at home, gender, race, and ethnicity as covariates in predicting child language outcomes. Gender was not a significant covariate in predicting child language outcomes at any age. Language spoken at home, race, and ethnicity were not significant covariates in predicting child language outcomes except in two of 21 analyses (those for home language and ethnicity in predicting 24-month expressive vocabulary). Importantly, their inclusion did not qualify our overall main findings, or change the strength of the relationship between the main predictors (intersensory matching, parent language input, and maternal education) and child language outcomes (for details see the online supplemental material, pp. 3–6).

All main analyses of MASks were focused on social events (however, a summary of analyses for nonsocial events, where there were few significant relations, is reported in the online supplemental material, pp. 8–9 and Table S5). For the present study, we focused on MASks for social events on high-competition trials because they bring out meaningful individual variability (important for predicting outcomes) at the ages we tested. When competition is high (for example, distractor is present), attentional load is increased and the task becomes more difficult. This leads to both impairments in performance and greater individual variability compared with low-competition trials. Consistent with this proposal, our preliminary analysis revealed lower attention maintenance, slower disengagement, and poorer intersensory matching on high- compared to low-competition trials, $p\text{s} < .05$ (for details, see Table S3 in the online supplemental material). Analyses also revealed greater individual variability on high- than low-competition trials. We calculated the coefficient of variation (CV; a scale independent index of variability) for each MASk at 12, 18, and 24 months. On average, the CV was twice as high for performance on high- (CV: 36.77; range: 15.62 to 62.10) than low-competition trials (CV: 18.48; range: 10.15 to 25.18; for descriptive statistics see Table S3). Accordingly, our results indicated significant correlations between MASks for social events on high-competition trials (that is, in the context of the central visual distractor) and language outcomes (see Table S4). In contrast, we found few significant correlations between MASKs for social events on low-competition trials and child language outcomes (see online supplemental material, pp. 7–8 and Table S3 for descriptive statistics and Table S4 for correlations comparing high- and low-competition trials). Thus, our primary analyses focused on MASks for social events on high-competition trials.

Correlational Analyses

Descriptive statistics for all child multisensory attention skills (MASks) for social events on high-competition trials, parent language input (types and tokens), and child speech production (types and tokens) and child vocabulary size (expressive and receptive) at 18 and 24 months appear in Table 3. We first calculated first-order, bivariate correlations between predictors (child MASks, parent language input: types and tokens) and child language outcomes (speech production: types and tokens; vocabulary size: expressive and receptive; see Table 4 for a summary of significant relations between predictors and outcomes and Tables S4 through S7 in the online supplemental material for all possible correlations

[2] FIML is generally appropriate for dominant missing data rates around 50% (see Enders, 2010; Graham & Schafer, 1999) and has been shown to yield unbiased parameter estimates for data that are missing at random. FIML produces unbiased parameter estimates where other traditional approaches (e.g., deletion methods) fail, because it maximizes statistical power by borrowing information from observed data (Enders, 2010). Further, in addition to conducting correlations (Table 4) and multiple regressions (Table 5) using FIML, we also conducted the same analyses without FIML using traditional Pearson correlations (Table S19) and OLS multiple regressions (Table S20) with the available data and compared their results. Both approaches yielded similar correlation and regression estimates.

Table 3

Mean (M), Standard Deviation (SD), and Sample Size (n) of Child Multisensory Attention Skills for Social Events on High-Competition Trials, Parent Language Input, and Child Language Outcomes at 12, 18, and 24 Months (N = 97)

	12 months			18 months			24 months		
Measure	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
Multisensory attention skills									
Intersensory matching	0.50	0.11	75	0.49	0.11	71	0.51	0.08	54
Sustained attention	0.45	0.20	78	0.47	0.20	74	0.51	0.16	54
Speed of shifting	1.34	0.65	76	1.46	0.91	71	1.32	0.53	53
Parent language input									
Quality–Types	13.31	4.57	84	14.96	4.77	76	18.00	5.55	70
Quantity–Tokens	49.65	21.97	84	57.01	23.35	76	66.61	21.00	70
Child speech production									
Quality–Types	—	—	—	0.65	0.75	76	2.70	2.09	70
Quantity–Tokens	—	—	—	1.50	1.80	76	6.13	5.24	70
Child vocabulary size									
Receptive vocabulary	—	—	—	231.67	148.90	51	—	—	—
Expressive vocabulary	—	—	—	61.75	77.93	51	275.37	180.00	51

conducted among predictors and outcomes). Pearson's *r* correlation coefficients were conducted using FIML,[3] and we corrected for familywise error rate.[4] The purpose of our correlational analyses was to narrow down and guide our decisions about which variables to include in the regression models, which tested our main research questions (Cohen et al., 2003; Kline, 2005).

Overall, of the three MASks, accuracy of intersensory matching for social events in the presence of a distracting event at 12 months predicted multiple child language outcomes, including quality and quantity of child speech and expressive vocabulary at 18 and 24 months (5 of 7; *r*-range: .30–.50, *p*s $<$ .01; see Table 4). In contrast, sustained attention and speed of shifting at 12, 18, and 24 months, and accuracy of intersensory matching at 18 and 24 months, did not predict child speech production and vocabulary size at 18 and 24 months after correcting for familywise error (*p*s $>$ .05). Parent language input (quality and quantity) at 18 and 24 months also predicted multiple child language outcomes, including quality and quantity of child speech production and expressive vocabulary at 18 months (*r*-range: .25–.40, *p*s $<$ .01; see Table 4). However, parent language input at 12 months predicted fewer child language outcomes. It predicted receptive vocabulary at 18 months and expressive vocabulary at 24 months (*r*-range: .26–.51, *p*s $<$ .01). Maternal education predicted quality and quantity of child speech production at 18 and 24 months (*r*-range: .27–.44, *p*s $<$ .01). For more detail about correlational analyses, please see the online supplemental material, pp. 6–10 and Tables S4 through S7).

Multiple Regression Analyses

Results from correlational analyses revealed that 12-month intersensory matching of social events in the presence of a distractor predicted a variety of child language outcomes (child speech quantity and quality at 18 and 24 months, and expressive vocabulary at 18 months; see Table 4). Might 12-month intersensory matching of social events also remain a significant predictor of these child language outcomes even when both quality and quantity of parent language input and maternal education are held constant? If so, this would demonstrate the importance of intersensory matching as a unique predictor of child language outcomes.

Given the pattern of significant findings from the correlations, and our interest in early attention skills as predictors of later language outcomes, our primary multiple regression models focus on assessing relations between 12-month accuracy of intersensory matching for social events in the presence of a distractor and child language outcomes (child speech production and vocabulary) at 18 and 24 months. We also included both quality and quantity of parent language input at 12 months (rather than at 18 or 24 months), as well as maternal education, as predictors because we were interested in the extent to which intersensory matching at 12 months predicted language outcomes holding constant other factors at that age. However, we also provide supplemental analyses for 18- and 24-month parent language input in Tables S12 through S18 (see online supplemental material, pp. 31–37) given that they predicted language outcomes as well.

For each outcome variable, we conducted four multiple regression models to assess the amount of unique variance (change in R^2) attributable to each 12-month predictor when all other predictors were held constant. To accomplish this, each of the four predictors at 12 months (intersensory matching, quantity of parent language input, quality of parent language input, maternal education) was entered into the regression model in a different order (first, second, third, fourth; see Tables S8 through S11 for

[3] All correlations conducted using FIML were also compared to those conducted with both the traditional bivariate pairwise Pearson's *r* correlations (using participants with data for both variables and excluding those with missing data) and percentage bend robust correlations to assure that the findings derived from FIML estimation were similar to the general pattern of findings from participants with complete data. All findings using FIML were similar in direction and magnitude to those from the bivariate pairwise correlations and percentage bend correlations for all results.

[4] At 18 months, there were four child language outcomes (child speech production: quantity and quality; child vocabulary size: receptive and expressive) and thus we used a familywise significance level of $p < .0125$ (.05/4; two-tailed) to evaluate results. At 24 months, there were three child language outcomes (child speech production: quantity and quality; child expressive vocabulary size) and thus we used a familywise significance level of $p < .0167$ (.05/3; two-tailed) to evaluate results.

Table 4

Estimates Using FIML: Bivariate Correlations Among Child Intersensory Matching for Social Events During Competing Stimulation, Parent Language Input (Quantity, Quality), Maternal Education, Child Speech Production (Quantity, Quality) and Child Vocabulary Size (Expressive, Receptive; N = 97)

	IM	Parent language input						Child language outcomes						
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Child intersensory matching (IM)														
1. 12-month IM	—													
Parent language input														
2. Quality (Types): 12 months	−.06	—												
3. Quantity (Tokens): 12 months	−.07	.86***	—											
4. Quality (Types): 18 months	.08	.62***	.55***	—										
5. Quantity (Tokens): 18 months	−.01	.58***	.63***	.86***	—									
6. Quality (Types): 24 months	.12	.55***	.46***	.52***	.49***	—								
7. Quantity (Tokens): 24 months	0	.61***	.59***	.56***	.65***	.83***	—							
Child language outcomes														
8. Quality (Types): 18 months	.50***	.06	.07	.29**	.23	—	—	—						
9. Quantity (Tokens): 18 months	.40***	.07	0	.26**	.24	—	—	.87***	—					
10. Receptive vocabulary: 18 months	.02	.20	.28**	.34***	.30**	—	—	.09	.09	—				
11. Expressive vocabulary: 18 months	.45***	.08	.17	.29**	.30**	—	—	.67***	.51***	.40***	—			
12. Quality (Types): 24 months	.35***	.18	.13	.33***	.21	.40***	.32**	.65***	.60***	.001	.50***	—		
13. Quantity (Tokens): 24 months	.30**	.01	.02	.15	.08	.29**	.23	.59***	.48***	−.04	.48***	.92***	—	
14. Expressive vocabulary: 24 months	−.01	.35***	.34***	.30**	.35***	.43***	.51***	.38***	.29**	.25	.50***	.58***	.59***	—
SES														
15. Maternal Education	.21	.40***	.30**	.31**	.19	.36***	.27**	.27**	.20	.03	.14	.44***	.32**	.22

Note. All significant values meet criteria of significance for familywise error, $p < .025$ for parent language input, $p < .0125$ at 18 months and $p < .0167$ at 24-months for child language outcomes. FIML = Full information maximum likelihood; SES = socioeconomic status.
$^{**}p < .01$. $^{***}p < .001$.

regression coefficients and R^2 for each outcome variable). Thus, in Model 1, maternal education was entered first followed by quality of parent language input, quantity of parent language input, and finally intersensory matching, and so forth for Models 2 to 4. The unique variance attributable to a given predictor is the change in R^2 when the predictor is entered last in the model (i.e., holding other predictors constant; for details see Tables S8 through S11).

The amount of variance uniquely attributable to intersensory matching of social events in the presence of a distractor in predicting each outcome along with the total variance accounted for by all predictors is shown in Table 5. Overall, the 12-month predictors taken together accounted for a significant amount of total variance in six of the seven child language outcomes, including quality of child speech, quantity of child speech, and expressive vocabulary at 18 and 24 months (range: 28 to 35%, $ps < .05$). Remarkably, 12-month accuracy of intersensory matching of social events on high-competition trials was a significant predictor and accounted for the largest amount of unique variance of all predictors in these six child language outcomes (range: 8 to 27%, $ps < .05$; see Table 5). These were primarily moderate to strong effects (with greater than 9% explained variance for moderate effects and greater than 25% for large effects, according to Cohen, 1988). In other words, when children receive equal amounts of parent language input and have similar levels of SES (i.e., holding these variables constant), there is leftover variability in predicting both quantity and quality of child speech production and expressive vocabulary size. Thus, intersensory matching of audiovisual speech in the presence of a distracting event at 12 months explains a significant proportion of this leftover variability in child language outcomes at 18 and 24 months. Further, individual differences in intersensory processing are associated with meaningful change in child language outcomes, particularly for 18- and 24-month expressive vocabulary (e.g., a 5% increase in intersensory matching predicts a 16.25 word increase in expressive vocabulary at 18 months, and a 23.55 word increase at 24 months, $ps < .05$; for details, see online supplemental material, pp. 12–13).

In contrast, 12-month parent language input (both quality and quantity) and maternal education accounted for a smaller and nonsignificant amount of unique variance in most child language outcomes. There were only a few exceptions: both 12-month quality and quantity of parent language input accounted for a significant amount of variance in 18-month expressive vocabulary size (quantity of parent language: 5%; quality of parent language: 9%, $ps < .05$), and maternal education accounted for a significant amount of unique variance in 24-month child speech production quality (12%, $p < .05$) and quantity (9%, $p < .05$). In sum, at 12 months, intersensory matching for social events accounted for the largest and significant amount of unique variance in child language outcomes when other predictors at 12 months (parent language input and maternal education) were held constant. Details for effects of 12-month predictors on each outcome variable can be found in the online supplemental material, pp. 10–13 and Tables S8 through S11.

Supplementary Analyses: Parent Language Input at Older Ages

Parent language input (quantity and quality) at 12 months was a weaker predictor of child language outcomes. However, parent language input at older ages—18 and 24 months—was moderately correlated with child language outcomes (see Table 4). When parent language input at these older ages was substituted for 12-month parent language input in our multiple regression models, analyses indicated that by 24 months, quantity and quality of parent language input became a somewhat stronger predictor of 24-month expressive vocabulary, predicting a greater amount of unique variance than it did at 12 or 18 months (see online supplemental material, pp. 13–16 and Tables S12 through S18). Specifically, 24-month quantity and quality of parent language input explained 6% and 15% (respectively), of the unique variance in 24-month expressive vocabulary after holding intersensory matching at 12 months constant. Importantly, intersensory matching of social events in the presence of a distractor remained a significant predictor of language outcomes, even after holding constant quantity and quality of parent language input at 18 or 24 months. It still predicted a moderate to large amount of unique variance in six out of the seven child language outcomes. Thus, when children receive equal amounts of parent language input at 12, 18, and 24 months, intersensory matching of social events explains a significant proportion of leftover variability in child language outcomes.

Discussion

In this study, we assessed relations among child multisensory attention skills (MASks) to social events (at 12, 18, and 24 months),

Table 5

Estimates Using FIML: Amount of Unique Variance Accounted for by Each Variable in Predicting Child Language Outcomes at 18 and 24 Months While Holding Constant All Other Predictors (N = 97)

	18-Month child language outcomes				24-Month child language outcomes		
12-Month predictors	Speech production		Vocabulary size		Speech production		Vocabulary size
Variance	Quantity	Quality	Expressive	Receptive	Quantity	Quality	Expressive
Total variance	.19**	.35***	.29***	.06	.22***	.30***	.18**
Unique variance							
Intersensory matching	.14**	.27***	.26***	.00	.09*	.10*	.08*
Maternal education	.02	.03	.00	.00	.09*	.12*	.01
Parent language quantity	.01	.02	.09*	.02	.01	.00	.01
Parent language quality	.01	.03	.05*	.00	.02	.00	.00

Note. FIML = Full information maximum likelihood.
$*p < .05$. $**p < .01$. $***p < .001$.

parent language input (at 12, 18, and 24 months), and child language outcomes (at 18 and 24 months) to examine the contributions of child MASks to well-established relations between parent language input and child language outcomes. Correlational analyses revealed that of the three MASks (sustained attention, intersensory matching, and shifting/disengaging), only intersensory matching of social events in the context of competing stimulation was associated with child language outcomes. We found a variety of novel relations revealing that intersensory matching of social events in the context of competing stimulation, along with parent language input and SES, play a significant role in predicting child language outcomes. Moreover, intersensory matching of social events contributed significantly to child language outcomes when controlling for parent language input and SES. We discuss each finding in turn.

Intersensory Processing of Social Events at 12 Months Predicts Unique Variance in Child Language Outcomes

Our primary regression models indicated that intersensory matching of faces and voices at 12 months was a strong and significant predictor of multiple measures of 18- and 24-month child speech production and vocabulary size even when 12-month parent language input (both quantity and quality) and SES were held constant. In particular, it predicted unique variance in both quality and quantity of child speech production, and expressive (but not receptive) vocabulary at 18 months, whereas it predicted a smaller amount of unique variance in both quality and quantity of child speech production, and expressive vocabulary at 24 months. Thus, at 12 months, given the same amount of parent language input and level of SES, intersensory processing skills can predict which children will benefit most from language learning opportunities provided by parent language input. Those children then go on to show better language functioning at 18 and 24 months of age. These novel findings are the first to establish infant intersensory processing of social events as an independent predictor of later language development in children. They extend previous findings using the MAAP with 2- to 5-year-old children (Bahrick et al., 2018) to younger infants, and demonstrate that relations between intersensory processing of social events and language outcomes are evident even after important predictors such as parent language input and maternal education are controlled (variables not controlled in the Bahrick et al., 2018 study). Findings highlight the importance of assessing intersensory processing of social events at 12 months of age along with well-established predictors of language.

In contrast with intersensory matching, parent language input and maternal education were weaker predictors of child language outcomes at 12 months of age. Quantity and quality of parent language input at 12 months accounted for unique variance in just expressive vocabulary at 18 months when intersensory matching and maternal education were held constant. Maternal education predicted unique variance in two outcomes—quantity and quality of child speech production at 24 months—when intersensory matching and parent language input were held constant. Thus, parent language input at 12 months and maternal education predicted fewer child language outcomes than intersensory matching.

Moreover, we also found that 12-month intersensory matching predicted language outcomes across development along with 18- and 24-month parent language input. Intersensory matching of faces and voices at 12 months continued to predict both child speech production and vocabulary size, even after holding constant parent language input at 18 and 24 months. Thus, given equivalent amounts of parent language input at 18 and 24 months, intersensory processing of faces and voices at 12 months still predicts which children will benefit most from parent language input and show better language outcomes.

Parent Language Input Predicts Unique Variance in Child Language Outcomes at All Ages

Our main regression analyses assessing 12-month predictors of language outcomes along with those from our online supplemental material (focusing on 18- and 24-month predictors) together revealed that parent language input at 12, 18, and 24 months predicted unique variance in one type of child language outcome at 18 and 24 months, expressive vocabulary size. Both quantity and quality of parent language input at 12 months, and quantity of parent language input at 18 months, predicted expressive vocabulary size at 18 months, holding other predictors constant. Also, both quantity and quality of parent language input at 18 and 24 months (but not 12 months) predicted unique variance in expressive vocabulary size at 24 months, holding other predictors constant. In contrast with findings from analyses of intersensory processing at 12 months, there was no evidence that parent language input at any of the three ages predicted unique variance in child speech production, quantity, or quality. Our findings of relations with child vocabulary are consistent with previous literature demonstrating that parent language input at older ages (18 and 24 months) predict children's language outcomes (e.g., Gilkerson et al., 2018; Hoff & Naigles, 2002; Jones & Rowland, 2017; Pan et al., 2005; Weisleder & Fernald, 2013). Few studies, in contrast, have assessed the role of parent language input (quality and/or quantity) in predicting language outcomes in infants of 12 months or younger (but see Newman et al., 2016, for a similar finding using a measure of parent repetition). Our findings extend this literature, demonstrating that parent language input (quality and quantity of input) at 12, 18, and 24 months predicts vocabulary size (but not child speech production) at 18 and 24 months. Thus, given similar levels of intersensory processing of faces and voices at 12 months, parents who provided more speech (quantity) and more diverse speech (quality) had children with larger vocabulary sizes at 18 and 24 months.

Developmental Changes in Foundations for Language Learning

Twelve months (and possibly earlier) may thus be an important time in development for assessing the role of intersensory processing skills in predicting outcomes, in part because these skills are still undergoing significant development during this period. As a result, at 12 months of age, infants show meaningful individual differences in intersensory matching skills that may correlate with later language. Across the first year, infants learn to efficiently locate a speaker based on audiovisual synchrony and selectively attend to her face and voice while filtering out other concurrent auditory and visual stimulation. Performance on the MAAP during trials with competing stimulation likely reflects these skills. In contrast, by 18 and 24 months of age, infants may be sufficiently skilled to pick out a speaker and efficiently attend to her face and voice while filtering out competing stimulation, leaving more time for processing other properties of the speech event (e.g., communicative intent, affect, linguistic content). Consistent with this interpretation, our correlations revealed that at 12 months, intersensory matching was a strong predictor of later language outcomes, but at the older ages, it was no longer predictive of language outcomes under these conditions. Instead, at older ages, parent language input and maternal education predicted language outcomes, particularly expressive vocabulary size. Taken together, the findings reviewed above suggest a changing pattern across development. That is, in early development, at 12 months, intersensory processing skills are most important in enabling infants to take advantage of language input. By 18 and 24 months, the effects of the quality and quantity of parent language input on child language outcomes appear to be increasingly important.

Why might intersensory processing at 12 months be a stronger predictor of later child language outcomes than parent language input at 12 months? We propose that, given equal amounts and diversity of parent language input at 12 months, intersensory processing skills determine the extent to which children benefit from opportunities for word learning. That is, infants who have better intersensory matching skills can more quickly locate a speaker and filter out competing stimulation. This leaves more time and attentional resources for further processing the speech event, including parsing the speech stream, following eye gaze direction, facilitating word mapping, and detecting facial and vocal affect signaling communicative intent, all skills that are built on intersensory processing (Bahrick et al., 2018; Gogate & Bahrick, 1998; Gogate & Hollich, 2010). Thus, when language input is equivalent, children's intersensory processing efficiency allows them to abstract more information from available input. It likely acts alongside other learning processes including joint attention and statistical learning of language.

Our findings indicate that at older ages, infants' level of intersensory processing at 12 months still impacts their ability to abstract information from the language input at 18 and 24 months, but not as strongly as it did at 12 months. Instead, the quantity and quality of parent language input at 18 and 24 months predicts an increasingly greater percent of unique variance in language outcomes. By 18 and 24 months of age, intersensory processing skills may be sufficiently refined so that once infants detect the speaker, they have more time and attentional resources available for further processing the language input. Thus, infants appear to show a shifting reliance on different sources of information across development for supporting language outcomes, from intersensory matching in early development to parent language input in later development. Future research should explore this possibility.

Intersensory Matching of Social Events, but Not Nonsocial Events at 12 Months is a Strong Predictor of Language Outcomes

In this article, we focused on intersensory processing of social events for predicting language outcomes because speaking faces provide a rich source of language learning opportunities for infants. As noted earlier, looking to the face of a speaker can provide information about affect, communicative intent, emphasis, and word-referent relations such as eye gaze direction, as well as facilitating deeper processing of the event as a whole (Gogate & Hollich, 2010; Reynolds et al., 2014). Further, intersensory matching of social, but not nonsocial, events predicted receptive and expressive language in 2- to 5-year-old children in our prior study using the MAAP (without controlling for quantity and quality of parent language input and SES; Bahrick et al., 2018). In the present study, analyses of attention to the nonsocial events reported in the online supplemental material also replicated this finding (pp. 8–9).

Why might intersensory matching of social events, in particular, be a better predictor of language outcomes than nonsocial events, given that intersensory processing is a basic skill involving audiovisual synchrony detection, a skill necessary for perceiving both social and nonsocial events? First, attention to social information provides the input for language development. Further, research has shown that attention to audiovisual speech events increases gradually across infancy whereas attention to nonsocial audiovisual events declines (Bahrick et al., 2016). This increase across age in attention to social events may be due to several factors including social scaffolding of language by caretakers and social interaction, which in turn, leads infants to develop increasingly greater expertise in the domain of social events.

Second, social events are typically more complex and variable than nonsocial events (Adolphs, 2001; Dawson et al., 2004). They provide an extraordinary amount of intersensory redundancy from rapidly changing coordinated patterns across face, voice, and gesture, making them more demanding of attentional resources than typical nonsocial events (Bahrick et al., 2016; Bahrick & Todd, 2012). Thus, task difficulty/complexity may be a significant factor in determining which contexts or protocols best predict outcomes at different ages across development. Protocols would be expected to best predict outcomes when their difficulty/complexity and/or task demands are optimally matched to the skills of the perceiver (Bahrick et al., 2010). The greater difficulty/complexity of processing social events (as compared with nonsocial events) presented by the MAAP may be optimal for infants at 12 months of age, and thus be more predictive of outcomes in later infancy. However, the finding that intersensory matching of social events at 12 months best predicted language outcomes in the current study, does not imply that intersensory matching of nonsocial events is irrelevant as a foundation for later language development. In earlier or later development, task difficulty/complexity of the nonsocial events may be better matched to the skills and attentional resources of younger infants (e.g., 6 months), and predict language outcomes.

Attention in the Context of Competing Stimulation Predicts Language Outcomes

Similarly, perceiving events in the context of competing stimulation also challenges attentional resources, making the task more difficult. Results of the present study revealed that intersensory matching of social events in the context of competing stimulation best predicted child language outcomes and thus, main analyses were conducted using these measures. Children who better matched the synchronous faces and voices of the woman speaking in the context of a central distractor event (morphing geometric forms) at 12 months had greater language outcomes at 18 and 24 months. Thus, the social events with high levels of competing stimulation may provide optimal task difficulty for 12-month-olds, and thus reveal sufficient variability across individuals for predicting language outcomes at 18 and 24 months. In fact, analyses revealed greater variability in intersensory matching for high- as compared with low-competition conditions. Our findings are also consistent with prior findings linking sustained attention during a distractor condition with later language outcomes (Salley et al., 2013). Given that competing stimulation increases task difficulty, it is possible that for younger infants, events with low levels of competing stimulation may be more optimally matched to their skills and thus may be most predictive of later language outcomes. This remains an important topic for future research.

Limitations and Future Research Directions

There are several limitations to the present study. Parent language input was coded from a semistructured lab-based interaction, which may present demand characteristics. Although previous literature indicates that semistructured lab-based interactions provide data similar to that of the periods when infants receive most language input in their home environment, it is nonetheless important to incorporate more naturalistic measures of parent language input in future research. Further, richer measures of quality of parent language input should also be incorporated along with the quantitative measure of diversity used in the present study. For example, fluency and connectedness, contingency of parent responses, and joint attention could be assessed in future research. The present study also did not incorporate measures of receptive, expressive, or productive language at 12 months of age. Parent language input at 12 months of age could be related to these measures of child language at 12 months of age. Finally, some of the child language outcomes used were parent report measures, which may over- or underestimate performance. Although we included both parent report and observational measures in our study, standardized observational measures of language could be incorporated in future research.

The present study also reveals a number of other important future research directions. Future research should examine the relations among MASks in infancy (intersensory processing, sustained attention, speed of shifting) and how they together may cascade into later child language skills. For example, our prior study found that basic attention skills (i.e., sustained attention) predicted intersensory matching, which in turn predicted language outcomes (Bahrick et al., 2018). Second, future research should examine other possible predictors of child language outcomes along with intersensory processing of social events. Our primary analyses revealed that intersensory processing of social events accounted for up to 27% of unique variance in child language outcomes, after controlling for parent language input and SES, and our secondary analyses found that gender, race and ethnicity had limited effects on language outcomes. Other possible predictors include birth order, gestational age, the infant's own vocal production and speech-like utterances, and child cognitive factors such as working memory, processing speed, early visual reception abilities, executive function, and so forth. Third, the present findings also suggest that an important avenue for intervention to optimize child language outcomes may be to improve the child's intersensory processing skills. This could be approached through training better audiovisual synchrony detection for audiovisual events. This, in turn, could cascade to later language outcomes. Together with existing interventions targeted at increasing parent language input, focusing on enhancing intersensory processing skills in children at risk for language delays may yield even greater benefits for language outcomes.

Conclusions

The present study is the first to examine relations among intersensory processing, parent language input, and child language outcomes across the period when language first emerges. Consistent with prior research, we found that parent language input at 12, 18 and 24 months predicted some types of child language outcomes (e.g., expressive vocabulary but not child speech production). Moreover, our findings revealed that intersensory matching of social events (in the presence of competing stimulation) at 12 months was a remarkably strong predictor of child language outcomes at 18 and 24 months, even when controlling for traditional predictors, including parent language input (quantity and quality), and SES (maternal education). Intersensory processing predicted a moderate to large and significant percentage of variance for a variety of child language outcomes at 18 and 24 months, including child speech production quality and quantity (assessed in a lab-based interaction) and child expressive vocabulary (a parent-report measure). In particular, for 18-month outcomes, it predicted more variance than either parent language input and maternal education. Thus, given equal levels of parent language input and SES at 12 months, it is the children with better intersensory matching skills who can take greater advantage of the language learning opportunities provided by the input. They may extract more meaning, affect, prosodic information, and move their attention more quickly and accurately in word naming contexts. Thus, for children with lower levels of parent language input and SES, good intersensory processing skills may be a protective factor, enhancing their efficiency in abstracting information from the linguistic environment, and in turn, enhancing their later language outcomes.

References

Adamson, L. B., Bakeman, R., & Deckner, D. F. (2004). The development of symbol-infused joint engagement. *Child Development*, *75*(4), 1171–1187. https://doi.org/10.1111/j.1467-8624.2004.00732.x

Adolphs, R. (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology*, *11*(2), 231–239. https://doi.org/10.1016/S0959-4388(00)00202-6

Bahrick, L. E. (1983). Infants' perception of substance and temporal synchrony in multimodal events. *Infant Behavior & Development*, *6*(4), 429–451. https://doi.org/10.1016/S0163-6383(83)90241-2

Bahrick, L. E. (1988). Intermodal learning in infancy: Learning on the basis of two kinds of invariant relations in audible and visible events. *Child Development*, *59*(1), 197–209. https://doi.org/10.2307/1130402

Bahrick, L. E., & Lickliter, R. (2012). The role of intersensory redundancy in early perceptual, cognitive, and social development. In A. Bremner, D. J. Lewkowicz, & C. Spence (Eds.), *Multisensory development* (pp. 183–206). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199586059.003.0008

Bahrick, L. E., Lickliter, R., Castellanos, I., & Vaillant-Molina, M. (2010). Increasing task difficulty enhances effects of intersensory redundancy: Testing a new prediction of the Intersensory Redundancy Hypothesis. *Developmental Science*, *13*(5), 731–737. https://doi.org/10.1111/j.1467-7687.2009.00928.x

Bahrick, L. E., Lickliter, R., & Todd, J. T. (2020). The development of multisensory attention skills: Individual differences, developmental outcomes, and applications. In J. J. Lockman & C. S. Tamis-LeMonda (Eds.), *The Cambridge handbook of infant development* (pp. 303–338). Cambridge University Press. https://doi.org/10.1017/9781108351959.011

Bahrick, L. E., McNew, M. E., Pruden, S. M., & Castellanos, I. (2019). Intersensory redundancy promotes infant detection of prosody in infant-directed speech. *Journal of Experimental Child Psychology*, *183*, 295–309. https://doi.org/10.1016/j.jecp.2019.02.008

Bahrick, L. E., & Todd, J. T. (2012). Multisensory processing in autism spectrum disorders: Intersensory processing disturbance as a basis for atypical development. In B. E. Stein (Ed.), *The new handbook of multisensory processes* (pp. 657–674). MIT Press.

Bahrick, L. E., Todd, J. T., Castellanos, I., & Sorondo, B. M. (2016). Enhanced attention to speaking faces versus other event types emerges gradually across infancy. *Developmental Psychology*, *52*(11), 1705–1720. https://doi.org/10.1037/dev0000157

Bahrick, L. E., Todd, J. T., & Soska, K. C. (2018). The Multisensory Attention Assessment Protocol (MAAP): Characterizing individual differences in multisensory attention skills in infants and children and relations with language and cognition. *Developmental Psychology*, *54*(12), 2207–2225. https://doi.org/10.1037/dev0000594

Bahrick, L. E., Walker, A. S., & Neisser, U. (1981). Selective looking by infants. *Cognitive Psychology*, *13*(3), 377–390. https://doi.org/10.1016/0010-0285(81)90014-1

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analyses for the behavioral sciences* (3rd ed.). Erlbaum.

Colombo, J., Shaddy, D. J., Richman, W. A., Maikranz, J. M., & Blaga, O. M. (2004). The developmental course of habituation in infancy and preschool outcome. *Infancy*, *5*(1), 1–38. https://doi.org/10.1207/s15327078in0501_1

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

Courage, M. L., Reynolds, G. D., & Richards, J. E. (2006). Infants' attention to patterned stimuli: Developmental change from 3 to 12 months of age. *Child Development*, *77*(3), 680–695. https://doi.org/10.1111/j.1467-8624.2006.00897.x

Curtindale, L. M., Bahrick, L. E., Lickliter, R., & Colombo, J. (2019). Effects of multimodal synchrony on infant attention and heart rate during events with social and nonsocial stimuli. *Journal of Experimental Child Psychology*, *178*, 283–294. https://doi.org/10.1016/j.jecp.2018.10.006

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early social attention impairments in autism: Social orienting, joint attention, and attention to distress. *Developmental Psychology*, *40*(2), 271–283. https://doi.org/10.1037/0012-1649.40.2.271

Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.

Erickson, L. C., Thiessen, E. D., Godwin, K. E., Dickerson, J. P., & Fisher, A. V. (2015). Endogenously and exogenously driven selective sustained attention: Contributions to learning in kindergarten children. *Journal of Experimental Child Psychology*, *138*, 126–134. https://doi.org/10.1016/j.jecp.2015.04.011

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Fisher, A., Thiessen, E., Godwin, K., Kloos, H., & Dickerson, J. (2013). Assessing selective sustained attention in 3- to 5-year-old children: Evidence from a new paradigm. *Journal of Experimental Child Psychology*, *114*(2), 275–294. https://doi.org/10.1016/j.jecp.2012.07.006

Flom, R., Lee, K., & Muir, D. (Eds.). (2007). *Gaze-following: Its development and significance*. Erlbaum. https://doi.org/10.4324/9781315093741

Flom, R., & Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology*, *43*(1), 238–252. https://doi.org/10.1037/0012-1649.43.1.238

Gilkerson, J., Richards, J. A., Warren, S. F., Montgomery, J. K., Greenwood, C. R., Kimbrough Oller, D., Hansen, J. H. L., & Paul, T. D. (2017). Mapping the early language environment using all-day recordings and automated analysis. *American Journal of Speech-Language Pathology*, *26*(2), 248–265. https://doi.org/10.1044/2016_AJSLP-15-0169

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

Gogate, L. J., Walker-Andrews, A. S., & Bahrick, L. E. (2001). The intersensory origins of word comprehension: An ecological-dynamic systems view. *Developmental Science*, *4*(1), 1–18. https://doi.org/10.1111/1467-7687.00143

Gogate, L. J., & Bahrick, L. E. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology*, *69*(2), 133–149. https://doi.org/10.1006/jecp.1998.2438

Gogate, L. J., & Hollich, G. (2010). Invariance detection within an interactive system: A perceptual gateway to language development. *Psychological Review*, *117*(2), 496–516. https://doi.org/10.1037/a0019049

Graham, J. W., & Schafer, J. L. (1999). On the performance of multiple imputation for multivariate data with small sample size. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 1–29). Sage.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Brookes Publishing.

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., Yust, P. K. S., & Suma, K. (2015). The contribution of early communication quality to low-income children's language success. *Psychological Science*, *26*(7), 1071–1083. https://doi.org/10.1177/0956797615581493

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

Hollich, G., Newman, R. S., & Jusczyk, P. W. (2005). Infants' use of synchronized visual information to separate streams of speech. *Child Development*, *76*(3), 598–613. https://doi.org/10.1111/j.1467-8624.2005.00866.x

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, *27*(2), 236–248. https://doi.org/10.1037/0012-1649.27.2.236

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, *61*(4), 343–365. https://doi.org/10.1016/j.cogpsych.2010.08.002

Hyde, D. C., Jones, B. L., Flom, R., & Porter, C. L. (2011). Neural signatures of face-voice synchrony in 5-month-old human infants. *Developmental Psychobiology*, *53*(4), 359–370. https://doi.org/10.1002/dev.20525

Isbell, E., Wray, A. H., & Neville, H. J. (2016). Individual differences in neural mechanisms of selective auditory attention in preschoolers from lower socioeconomic status backgrounds: An event-related potentials study. *Developmental Science*, *19*(6), 865–880. https://doi.org/10.1111/desc.12334

Jackson-Maldonado, D., Thal, D., Marchman, V. A., Newton, T., Fenson, L., & Conboy, B. (2003). *MacArthur inventorios del desarollo de habilidades communicativas: User's guide and technical manual*. Brookes.

Jones, G., & Rowland, C. F. (2017). Diversity not quantity in caregiver speech: Using computational modeling to isolate the effects of the quantity and the diversity of the input on vocabulary growth. *Cognitive Psychology*, *98*, 1–21. https://doi.org/10.1016/j.cogpsych.2017.07.002

Kline, R. B. (2005). *Methodology in the social sciences: Principles and practice of structural equation modeling* (2nd ed.). Guilford Press.

Leech, K. A., Salo, V. C., Rowe, M. L., & Cabrera, N. J. (2013). Father input and child vocabulary development: The importance of Wh questions and clarification requests. *Seminars in Speech and Language*, *34*(4), 249–259. https://doi.org/10.1055/s-0033-1353445

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(5), 1431–1436. https://doi.org/10.1073/pnas.1114783109

MacWhinney, B. (2000). *The CHILDES Project: Volume 1: Tools for analyzing talk: Transcription format and programs* (3rd ed.). Erlbaum. https://doi.org/10.1162/coli.2000.26.4.657

Malvern, D., & Richards, B. (2012). Measures of lexical richness. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 3622–3627). Wiley. https://doi.org/10.1002/9781405198431.wbeal0755

Marchman, V. A., & Martine-Sussmann, C. (2002). Concurrent validity of caregiver/parent report measures of language for children who are learning both English and Spanish. *Journal of Speech, Language, and Hearing Research*, *45*(5), 983–997. https://doi.org/10.1044/1092-4388(2002/080

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381–392. https://doi.org/10.3758/BRM.42.2.381

Mundy, P., & Burnette, C. (2005). Joint attention and neurodevelopmental models of autism. In F. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders* (Vol. 1, 3rd ed., pp. 650–681). Wiley. https://doi.org/10.1002/9780470939345.ch25

Newman, R. S., Rowe, M. L., & Bernstein Ratner, N. (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

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development*, *76*(4), 763–782. https://doi.org/10.1111/j.1467-8624.2005.00876.x

Patten, E., Watson, L. R., & Baranek, G. T. (2014). Temporal synchrony detection and associations with language in young children with ASD. *Autism Research and Treatment*, *2014*, 678346. https://doi.org/10.1155/2014/678346

Pearson, B. Z., Fernandez, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics*, *18*(1), 41–58. https://doi.org/10.1017/S0142716400009863

Piazza, E. A., Hasenfratz, L., Hasson, U., & Lew-Williams, C. (2020). Infant and adult brains are coupled to the dynamics of natural communication. *Psychological Science*, *31*(1), 6–17. https://doi.org/10.1177/0956797619878698

Purpura, D. J. (2019). Language clearly matters; Methods matter too. *Child Development*, *90*(6), 1839–1846. https://doi.org/10.1111/cdev.13327

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2017). The impact of early social interactions on later language development in Spanish–English bilingual infants. *Child Development*, *88*(4), 1216–1234. https://doi.org/10.1111/cdev.12648

Reynolds, E., Vernon-Feagans, L., Bratsch-Hines, M., & Baker, C. E. (2019). Mothers' and fathers' language input from 6 to 36 months in rural two-parent-families: Relations to children's kindergarten achievement. *Early Childhood Research Quarterly*, *47*, 385–395. https://doi.org/10.1016/j.ecresq.2018.09.002

Reynolds, G. D., Bahrick, L. E., Lickliter, R., & Guy, M. W. (2014). Neural correlates of intersensory processing in 5-month-old infants. *Developmental Psychobiology*, *56*(3), 355–372. https://doi.org/10.1002/dev.21104

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2009). A cognitive approach to the development of early language. *Child Development*, *80*(1), 134–150. https://doi.org/10.1111/j.1467-8624.2008.01250.x

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language*, *35*(1), 185–205. https://doi.org/10.1017/S0305000907008343

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

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. https://doi.org/10.1093/biomet/63.3.581

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

Shaddy, D. J., & Colombo, J. (2004). Developmental changes in infant attention to dynamic and static stimuli. *Infancy*, *5*(3), 355–365. https://doi.org/10.1207/s15327078in0503_6

Shepard, K. G., Spence, M. J., & Sasson, N. J. (2012). Distinct facial characteristics differentiate communicative intent of infant-directed speech. *Infant and Child Development*, *21*(6), 555–578. https://doi.org/10.1002/icd.1757

Spelke, E. (1976). Infants' intermodal perception of events. *Cognitive Psychology*, *8*(4), 553–560. https://doi.org/10.1016/0010-0285(76)90018-9

Tamis-LeMonda, C. S., Kuchirko, Y., Luo, R., Escobar, K., & Bornstein, M. H. (2017). Power in methods: Language to infants in structured and naturalistic contexts. *Developmental Science*, *20*(6), e12456. https://doi.org/10.1111/desc.12456

Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Shah, R. J., Malle, B. F., & Morgan, J. L. (2015). Attention to the mouth and gaze following in infancy predict language development. *Journal of Child Language*, *42*(6), 1173–1190. https://doi.org/10.1017/S0305000914000725

Todd, J. T., & Bahrick, L. E. (2022). *Individual differences in multisensory attention skills in children with autism spectrum disorder predict language functioning and symptom severity: Evidence from the Multisensory Attention Assessment Protocol (MAAP)* [Manuscript submitted for publication]. Department of Psychology, Florida International University.

Trueswell, J. C., Lin, Y., Armstrong, B., III Cartmill, E. A., Goldin-Meadow, S., & Gleitman, L. R. (2016). Perceiving referential intent: Dynamics of reference in natural parent–child interactions. *Cognition*, *148*, 117–135. https://doi.org/10.1016/j.cognition.2015.11.002

Tsang, T., Atagi, N., & Johnson, S. P. (2018). Selective attention to the mouth is associated with expressive language skills in monolingual and bilingual infants. *Journal of Experimental Child Psychology*, *169*, 93–109. https://doi.org/10.1016/j.jecp.2018.01.002

Vaillant-Molina, M., & Bahrick, L. E. (2012). The role of intersensory redundancy in the emergence of social referencing in 5 1/2-month-old infants. *Developmental Psychology*, *48*(1), 1–9. https://doi.org/10.1037/a0025263

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin*, *121*(3), 437–456. https://doi.org/10.1037/0033-2909.121.3.437

Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychological Science*, *25*(7), 1314–1324. https://doi.org/10.1177/0956797614531023

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

Weizman, Z. O., & Snow, C. E. (2001). Lexical input as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology*, *37*(2), 265–279. https://doi.org/10.1037/0012-1649.37.2.265

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

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*(1), 37–63. https://doi.org/10.1016/S0885-2014(96)90027-1

Received January 27, 2020
Revision received February 21, 2022
Accepted March 4, 2022 ■