How Parents Impact the Joint Attention-Vocabulary Bond for Children With Autism

Elizsbeth Perkovich
Department of Psychology
University of Houston
Houston, USA
esperkov@cougarnet.uh.edu

Hanako Yoshida
Department of Psychology
University of Houston
Houston, USA
hyoshida@uh.edu

***Abstract*—Joint attention is a form of socially coordinated attention that has been well-documented to predict language development in children with autism spectrum disorder. Computer-based assessment of early JA behaviors limits our understanding of how real-time attention experiences in social and naturalistic dynamics might be related to language learning. The present study used head-mounted eye-tracking methods during parent-child object play to document how moment-to-moment joint attention experiences are related to children's vocabulary achievement and the role of parental input on the relationship. Results indicate that the joint attention measurements positively predicted early vocabulary development and the relation was moderated by parental object-holding and parental sustained attention to their child's face. The results may offer insights into the development of early parent-mediated autism interventions that focus on parental play behaviors.**

Keywords—head-mounted eye-tracking, parent-child interaction, joint attention, autism

I. Introduction

Joint attention (JA) is a critical component of early vocabulary development among children [1] with great individual variability. The present study directly measured parent-child joint attention experiences during an interactive play session and examined the effects of different types of parent input behaviors on the relationship between JA and vocabulary among children with autism. This paper uses a combination of person-first and identity-first language. This intentional decision aligns with recent comments by Vivanti [2] which recognizes the complexities of known and unknown preferences of those in the larger autism community.

A. Joint Attention

JA refers to the ability to coordinate attention with another individual toward a target of interest and emerges during the first year of life [1] and includes the initiation of JA (IJA; an individual's ability to 'initiate' the direction of the JA instance) and the response to JA (RJA; an individual's ability to 'follow' the direction of another person's JA bid). During these moments, more experienced communication partners (i.e., parents) can guide less experienced communication partners (i.e., children) by assisting with problem-solving and providing critical learning opportunities [1]. JA episodes also represent optimal moments for children to find a referent of a corresponding word and early learning as children learn to use their parent's gaze toward a referent during labeling [3], [4]. As such, research investigating the link between JA and language has often focused on early parent-child interaction paradigms [5].

Difficulties in developing JA skills have been identified as one of the earliest indicators of autism, a neurodevelopmental condition characterized by social communication difficulties, including limited attention to referential cues [6]. The ability to attend to and follow up on these referential cues within social interactions has been linked to positive language development among typically developing (TD) children [7]. Thus, researchers have suggested this as a possible explanation as to why autistic children often exhibit delayed communication, language, and speech development [8].

Though difficulty and variations in JA may exist among children with autism, they have also been well-documented to have positive links between their different JA behaviors and their language development. A series of studies suggest that RJA is predictive of vocabulary achievement (receptive and expressive) among children 2 to 5 years old [9], [10], growth rates among 31-64-month-olds [11], and language abilities among 22-93-month-olds [12]. Similar relationships have also been observed with IJA - for instance, IJA at 2 years old has been linked to verbal outcomes at 3 years old [13] with IJA at 3 years old predicting expressive vocabulary trajectories from 3 to 19 years old [14]. Also, more frequent IJA has been associated with better receptive and expressive vocabulary, both concurrently and longitudinally, in this population [15]. Studies have suggested that RJA and IJA represent unique aspects of JA with different underlying processes relevant to developmental trajectories in language development [16]. Therefore, studies must consider these behaviors separately when investigating the contributions of JA to language development [17]. However, little is known about the underlying mechanisms driving the relationship between real-time JA experiences in everyday social activity and language development in this population.

B. Parent Play Behaviors

Previous studies of parent-child social interactions have reported that parents of children with autism use different social scaffolding behaviors than those of parents of TD children. For example, parents of autistic children between the ages of 1.5 and 5 years old were more engaged (e.g., initiating more play schemes) during free play [18]. Studies with children 3 to 5 years old found that mothers of children with autism used more physical contact, more high-intensity behaviors, and fewer social-verbal approaches [19], [20] in addition to producing more gestures and more closely monitoring their children's faces than parents of TD children [21]. Parents of children with autism between 3 and 8 years old also spent more time looking at their children and spent less time looking at objects during object play [22]. A recent meta-analysis also revealed that parents of autistic children

2024 IEEE International Conference on Development and Learning (ICDL) | 979-8-3503-4855-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICDL61372.2024.10644652

showed more controlling and intrusive behaviors within play settings [23]. Taken together, the previous studies show that children with autism experience enriched yet unique parental social scaffolding, which may differentially influence their learning experiences and achievement compared to their same-aged TD peers.

These differential parental behaviors have been linked to the developmental outcomes of children with autism. One study documented that parental responsiveness to their children's attention and play behaviors is predictive of their child's rate of language growth [11]. Other studies recording parental speech towards children with autism - such as responsiveness, child-directed speech, and mean length of utterances - suggest the relation to language development [24], [25]. Further, parental synchronization during play also predicted language longitudinally over 1, 10, and 16 years [26]. Though this line of research suggests that parental social scaffolding behaviors may differentially influence language development, it is not clear how the impact may be generated – which specific child behaviors are impacted, and which parental behaviors are most impactful.

Recent studies exploring this question within live social interaction paradigms through head-mounted eye-tracking technology is encouraging [27]. By using this technology, researchers have identified several parental social scaffolding strategies (e.g., object handling, object looking, and object labeling) that support their children's sustained attention experiences and their subsequent word learning [27]. Though there is still limited work using head-mounted eye-tracking methods with children with autism, there are some initial efforts [22], [28], [29]. A study of 24 to 48-month-old autistic children found similar pathways of establishing JA (through following the hands of their parents rather than their parent's faces) compared to children without autism [28]. In another study of children with autism aged 3 to 8 years old, parental attention to their child predicted the frequency and duration of JA episodes [22]. These studies document that head-mounted eye-tracking can be successfully used with young autistic children and that parents utilize different types of scaffolding strategies to support JA. However, the current knowledge is still limited to the description of synchrony, as the field awaits more work to address the developmental significance – how social scaffolding behaviors may be relevant to social attention and language development. The present study investigated the moderating role of parental input in the relationship between JA experiences and vocabulary scores.

C. Hypotheses

The hypotheses are: (1) IJA and RJA will positively predict receptive and expressive vocabulary and (2) parental input behaviors (attention to child's face, attention to objects, object handling, and object labeling) will positively moderate the relationship between JA experiences and vocabulary.

II. METHOD

A. Participants

40 children diagnosed with autism between 3 and 8 years old (M = 6.00 years, SD = 1.67 years, 33 males, 38 mothers, 15 Hispanic, 11 White, 6 Asian, 3 Black, 2 Biracial, and 1 American Indian) and their parents participated in the current study. Children met the following inclusion criteria to participate: a prior medical or school-based diagnosis of autism, full-term birth (i.e., 38 weeks or weighed >2.41 kg at birth), ambulatory (i.e., able to walk), no documented hearing or visual impairments, and came from English-speaking households.

B. Procedures

Upon arrival at the lab at which the present study took place, parent participants completed informed consent forms. After form completion, parent-child dyads were directed to the experiment room, where they participated in a 5-minute and 20-second semi-naturalistic object play session with eight unique toy objects (four familiar: bottle, bunny, car, and cookie; four unfamiliar: caliper, nylon, pipette, and strainer). The unfamiliar objects were added to increase task complexity to maintain the child's interest in the play session, as children are more likely to select and attend to unfamiliar objects [30]. During the play session, parents and children sat across from each other at a child-sized table (60 cm × 60 cm × 40 cm) while wearing head-mounted eye trackers. Calibration procedures were administered to the parent-child dyads both before and after the play session. Parent participants were instructed to play as naturally as possible during the object play session as if they were at home. This portion of the visit - including set-up, calibrations, and the play session itself - lasted approximately 20 minutes.

Following the play session, child participants completed Form A of both the Peabody Picture Vocabulary Test Fourth Edition (PPVT-4) [31] for a respective vocabulary score and the Expressive Vocabulary Test Second Edition (EVT-2) [32] for an expressive vocabulary score. Standardized scores using child age in months were used for the current analysis. These tests are co-normed and have been deemed suitable for testing with autistic children [33]. Each assessment takes approximately 20 minutes, for a total of 40 minutes for the language assessment portion of the visit.

All dyads received a gift card, a family pass to a local children's museum, and a toy after participation. The study and its procedures were approved by the university's Institutional Review Board where the research took place.

C. Equipment

Positive Science, Inc. [34] head-mounted eye-trackers were used for the current study, which consisted of Watec (WAT-230A) miniature color cameras with supplementary eye-trackers (weighing 51g in total). A minimum inter-correlation of 0.9 between the scene camera and the eye camera was obtained for each participant through the Yarbus software program, which estimates the participant's eye gaze location on the scenery image captured from the participant's forehead camera [34]. Both child and parent eye-tracking videos were synchronized with two additional views of the play session (from a wall-mounted camera and a ceiling-mounted camera) and an audio recording before being rendered at 30 frames per second using the Adobe Premiere software program. On average, each parent-infant dyad has 9,377 frames (SD = 435) recorded during the play session which were annotated and used for the current analysis.

Inaccessible frames included eye blinks and play-session interruptions.

D. Behavioral Annotation

Behavioral annotation took place in the Datavyu software program [35] by trained research assistants, who manually annotated each frame of the 5-minute and 20-second play session. Child and parent attention patterns were annotated for attention to toy objects, while parent gaze patterns were additionally annotated for attention to their child's face. Child and parent attention to the toy objects was then synchronized to annotate JA experiences (moments when both the parent and the child looked at the same toy object simultaneously). These JA moments were further annotated to identify IJA moments (when children looked at the object first) and RJA moments (when parents looked at the object first). Four parental input behaviors were measured: (1) parent sustained attention to their child's face for at least 2000 milliseconds ("SA to child's face"), (2) parent sustained attention to a toy object for at least 2000 milliseconds ("SA to objects"), (3) parent object handling, and (4) parent object labeling. Parent object handling was annotated for each hand (i.e., right and left hand) separately, with each moment beginning when the parent started to touch any of the toy objects and ending when the parent no longer touched the object. Parent object labels were annotated and transcribed. We calculated the frequency per minute and total duration per minute for each measurement.

Reliability was measured by randomly selecting 25% of the frames for 10 randomly selected dyads and assessing inter-rater coding agreement. Inter-rater reliability averaged 87% across the behaviors of interest - this aligns with reliability rates reported in other head-mounted eye-tracking studies with children with autism [21], [22].

E. Data Analysis Approach

To test the first hypothesis; (1) IJA and RJA will positively predict receptive and expressive vocabulary, a series of multiple regression models with child gender and parent gender as dummy-coded covariates were conducted in SPSS. To test the second hypothesis; (2) parental input behaviors (SA to child's face, SA to objects, object handling, and object labeling) will positively moderate the relationship between JA experiences and vocabulary, a series of moderation models with child gender and parent gender as dummy-coded covariates were conducted using the PROCESS v4.2 macro in SPSS [36] with a 95% confidence interval and 5,000 bootstrap samples. Conditioning values were set at -1 SD of the mean ("low"), the mean, and +1 SD of the mean ("high"). We include child gender and parent gender as covariates given recent research documenting gender-related differences in child attention [37] and parent play behaviors [38] respectively.

III. RESULTS

A. Descriptive Statistics

The average standardized PPVT-4 receptive vocabulary score of the current sample was 87.40 (*SD* = 24.92) and the average standardized EVT-2 expressive vocabulary score was 85.35 (*SD* = 24.84). Parent-child dyads experienced an average of 5.02 (*SD* = 3.05) IJA instances per minute, for an average of 1.95 (*SD* = 1.22) seconds per minute. Parent-child dyads also experienced an average of 5.36 (*SD* = 2.65) RJA instances per minute, for an average of 2.08 (*SD* = 1.57) seconds per minute. Parents averaged 0.73 (*SD* = 0.63) SA instances to objects per minute, for an average of 2.11 (*SD* = 2.00) seconds per minute. Parents also averaged 1.74 (*SD* = 1.48) SA instances to their child's face per minute, for an average of 6.99 (*SD* = 7.75) seconds per minute. Parents averaged 20.31 (*SD* = 10.09) unique instances of object handling per minute, for an average of 37.27 (*SD* = 11.85) seconds per minute. Parents averaged 3.73 (*SD* = 2.20) instances of labeling objects per minute, for an average of 2.72 (*SD* = 2.15) seconds per minute.

A. Summary of Results

The series of multiple linear regression and moderation models revealed three key results concerning (1) the relationship between JA behaviors and vocabulary scores and (2) the moderating effect of parental behaviors (SA to objects, SA to child's face, object handling, and object labeling). First, RJA/IJA frequency per minute and duration per minute positively predicted both receptive and expressive vocabulary (all p-values < .05). Second, parent SA to child's face overall negatively moderates the relationship between RJA frequency/duration per minute and receptive/expressive vocabulary, though simple slopes analysis revealed positive moderation at low and mean levels of parent SA to child's face. Third, parent object handling positively moderates IJA/RJA frequency per minute and receptive/expressive vocabulary.

B. Joint Attention to Vocabulary

We first examined the relationship between RJA *frequency* per minute and the two types of vocabulary. RJA frequency per minute was positively related to receptive vocabulary (β = .448, p = .003). RJA frequency per minute was positively related to expressive vocabulary (β = .418, p = .009). We then examined the relationship between IJA *frequency* per minute and the two types of vocabulary. IJA frequency per minute was positively related to receptive vocabulary (β = .410, p = .008). IJA frequency per minute was positively related to expressive vocabulary (β = .423, p = .008). Third, we examined the relationship between RJA *duration* per minute and the two types of vocabulary. RJA duration per minute was positively related to receptive vocabulary (β = .388, p = .012). RJA duration per minute was positively related to expressive vocabulary (β = .420, p = .008). Finally, we tested the relationship between IJA *duration* per minute and the two types of vocabulary. IJA duration per minute was positively related to receptive vocabulary (β = .376, p = .014). IJA duration per minute was positively related to expressive vocabulary (β = .398, p = .012).

C. Parent SA to Child's Face

Both models exploring the moderating effect of parent SA to child's face *frequency* per minute on the relationship between IJA frequency and receptive/expressive vocabulary were nonsignificant. Both models exploring the moderating effect of parent SA to child's face *duration* per minute on the relationship between IJA duration per minute and receptive/expressive vocabulary were nonsignificant.

Parent SA to child's face *frequency* per minute had a negative moderating effect on the relationship between RJA frequency per minute and receptive vocabulary (b = -2.357, t = -2.186, p = .036). A post hoc simple slopes analysis revealed that when parents exhibited low and mean levels of SA to child's face, the positive association between RJA frequency per minute and receptive vocabulary was significant (b = 7.652, se = 2.092, t = 3.657, p < .001; b = 4.156, se = 1.341, t = 3.098, p = .004, respectively). However, when parents exhibited high levels of SA to child's face, the association between RJA frequency per minute and receptive vocabulary was non-significant (b = .660, se = 2.082, t = .317, p = .753).

Parental SA to child's face *frequency* per minute had a significant negative moderating effect on the relationship between RJA frequency per minute and expressive vocabulary (b = -2.664, t = -2.418, p = .022). A post hoc simple slopes analysis revealed that when parents exhibited low and mean levels of SA to child's face, the positive association between RJA frequency per minute and expressive vocabulary was significant (b = 7.640, se = 2.138, t = 3.573, p = .001; b = 3.690, se = 1.371, t = 2.690, p = .011, respectively). However, when parents exhibited high levels of child-face-looking SA, the association between RJA frequency per minute and expressive vocabulary was non-significant (b = -.264, se = 2.128, t = -.124, p = .902).

Parental SA to child's face *duration* per minute had a significant negative moderating effect on the relationship between RJA duration per minute and receptive vocabulary (b = -.001, t = -2.063, p = .047). A post hoc simple slopes analysis revealed that when parents exhibited low levels of SA to child's face, the positive association between RJA duration per minute and receptive vocabulary was significant (b = .015, se = .005, t = 3.273, p = .002). However, when parents exhibited mean and high levels of SA to child's face, the association between RJA duration per minute and receptive vocabulary was non-significant (b = .005, se = .003, t = 1.745, p = .090; b = -.005, se = .007, t = -.735, p = .468, respectively).

Parental SA to child's face *duration* per minute had a significant negative moderating effect on the relationship between RJA duration per minute and expressive vocabulary (b = -.002, t = -2.252, p = .031). A post hoc simple slopes analysis revealed that when parents exhibited low levels of SA to child's face, the positive association between RJA duration per minute and expressive vocabulary was significant (b = .016, se = .005, t = 3.528, p = .001). However, when parents exhibited mean and high levels of SA to child's face, the association between RJA duration per minute and expressive vocabulary was non-significant (b = .006, se = .003, t = 1.839, p = .075; b = -.006, se = .007, t = -.832, p = .411, respectively).

D. Parent SA to Objects

All four models exploring the moderating effect of parent SA to object *frequency* per minute on the relationship between IJA/RJA frequency per minute and receptive/expressive vocabulary were nonsignificant. All four models exploring the moderating effect of parent SA to object *duration* per minute on the relationship between IJA/RJA duration per minute and receptive/expressive vocabulary were nonsignificant.

F. **Parent Object Handling**

Parental object handling frequency per minute had a significant positive moderating effect on the relationship between IJA frequency per minute and receptive vocabulary (b = .261, t = 2.468, p = .019). A post hoc simple slopes analysis revealed that when parents exhibited mean and high levels of object handling, the positive association between IJA frequency per minute and receptive vocabulary was significant (b = 3.152, se = 1.120, t = 2.815, p = .008; b = 5.789, se = 1.553, t = 3.729, p < .001, respectively). However, when parents exhibited low levels of object handling, the association between IJA frequency per minute and receptive vocabulary was non-significant (b = .515, se = 1.543, t = .334, p = .741).

Parental object handling frequency per minute had a significant positive moderating effect on the relationship between IJA frequency per minute and expressive vocabulary (b = .328, t = 3.119, p = .004). A post hoc simple slopes analysis revealed that when parents exhibited mean and high levels of object handling, the positive association between IJA frequency per minute and receptive vocabulary was significant (b = 3.284, se = 1.112, t = 2.952, p = .006; b = 6.595, se = 1.543, t = 4.275, p < .001, respectively). However, when parents exhibited low levels of object handling, the association between IJA frequency per minute and expressive vocabulary was non-significant (b = -.028, se = 1.533, t = -.018, p = .986).

Parental object handling frequency per minute had a significant positive moderating effect on the relationship between RJA frequency per minute and receptive vocabulary (b = .356, t = 2.296, p = .028). A post hoc simple slopes analysis revealed that when parents exhibited mean and high levels of object handling, the positive association between RJA frequency per minute and receptive vocabulary was significant (b = 3.791, se = 1.345, t = 2.811, p = .008; b = 7.377, se = 1.969, t = 3.747, p < .001, respectively). However, when parents exhibited low levels of object handling, the association between RJA frequency per minute and receptive vocabulary was non-significant (b = .204, se = 2.155, t = .095, p = .925).

Parental object handling frequency per minute had a significant positive moderating effect on the relationship between RJA frequency per minute and expressive vocabulary (b = .432, t = 2.696, p = .011). A post hoc simple slopes analysis revealed that when parents exhibited mean and high levels of object handling, the positive association between RJA frequency per minute and receptive vocabulary was significant (b = 3.456, se = 1.396, t = 2.475, p = .019; b = 7.815, se = 2.038, t = 3.834, p < .001, respectively). However, when parents exhibited low levels of object handling the association between RJA frequency per minute and expressive vocabulary was non-significant (b = -.903, se = 2.230, t = -.405, p = .688).

E. Parent Object Labeling

All four models exploring the moderating effect of parent object labeling *frequency* per minute on the relationship between IJA/RJA frequency per minute and

receptive/expressive vocabulary were not significant. All four models exploring the moderating effect of parent object labeling *duration* per minute on the relationship between IJA/RJA duration per minute and receptive/expressive vocabulary were not significant.

IV. DISCUSSION

The present study recorded naturally occurring JA experiences between children with autism and their parents during an interactive object play and examined the predictiveness of JA experiences on vocabulary and the influence of parental input on these relationships. There are two discussion points regarding the nature of the relationship between early social attention and vocabulary measures.

A. The Role of JA in Language Development

The results indicated the significant positive relationships between JA measure (RJA and IJA) and vocabulary measures (receptive and expressive), which is consistent with previous literature using well-controlled attention tasks [7], [10], [11], [12], [13], [14], [15]. These results further suggest that some autistic children may continue to develop vocabulary contingent upon their JA abilities. Indeed, previous researchers have speculated that there may be a 'threshold' of JA, that when met, results in children's vocabulary no longer being directly tied to their JA experiences - rather, more advanced developmental behaviors drive further language growth [7], [39]. Our findings of slightly stronger RJA-vocabulary relationships are supported by a recent meta-analysis suggesting that RJA may play a stronger role in language development than IJA [7] - that is, children who do not respond to JA bids may miss out on the parental social input that would supplement or follow a successful bid (e.g., a vocal utterance about the object in question). Additionally, children who exhibit greater RJA abilities may better process varied linguistic input that is not directly related to the object of focus, contributing to greater growth in vocabulary. This difference may also be attributed to the differences in how each JA behavior is generated - RJA is initiated by a secondary party, such as an examiner or a parent, while IJA is spontaneous by the child.

B. Parental Role in the JA-Vocabulary Bond

One of the interesting yet puzzling findings was that there was a negative moderating effect of parental face monitoring (parent SA to child's face) on the relationship between RJA and receptive/expressive vocabulary, with significant positive associations at the mean and low levels of parent's SA to child's face. One explanation is that parents who look 'too much' at their child's face (i.e., high levels of SA to child's face) have fewer opportunities to establish moments of JA with their children. These children may unintentionally have fewer opportunities to respond to their parent's JA bids than children with parents with lower frequencies of SA to child's face. This may lead to an over-representation of IJA experiences compared to RJA experiences amongst these children, as parents are more frequently able to follow their child's gaze because of looking at their face more, which may weaken the subsequent JA-vocabulary bond. A second explanation is that these parents are dominant in creating JA experiences by monitoring their child's face (and therefore their gaze) more closely – this could be unique to the social interactive context used in the current study since JA can be driven primarily by one social partner (i.e., the parent). In contrast, computer-based tasks test the child's capacity to engage in JA. This raises the important question of the connection between children's JA capacity during non-social contexts compared to their naturally occurring JA experiences within social contexts. We also found a positive moderating effect of parental object handling on the relationship between IJA/RJA and receptive/expressive vocabulary among parents who frequently handled objects. This could mean that parents who engage in more frequent bouts of object handling facilitate increased and enriched JA experiences with their children. In contrast, parents who handle toys less frequently may inadvertently limit opportunities for shared attention by taking a more back-seat role in the interaction. These moderating effects along with the speculations underscore the importance of investigating the influence of parental input on the relationship between JA and vocabulary, for which their unique contributions can be leveraged for parent-mediated autism interventions.

Some limitations should be noted to further address the implications. First, the significance of parental input and naturally occurring JA experiences may differ among autistic children who are younger than the current sample. Studies of 2-3-year-olds using computer-based tasks indicate that JA experiences may show more atypicality earlier in autism development [40], which could significantly impact the influence of parental behaviors on the JA-vocabulary relationship, compared to when their JA experiences are more similar to children without autism. However, these studies were conducted with computer-based tasks, which differs from the social interaction in the current study - this suggests that not only age but also context may influence children's JA experiences. For example, contextual factors such as the novelty of toys and the lab setting may differentially impact the JA and parental social scaffolding behaviors measured in the current paper. The present study also exclusively focused on *parental* social scaffolding, one of the least studied variables within early attention and language literature. However, studies indicate a strong reciprocal relationship between parents and children during social interaction [41]. Further investigation of child-centered measurements (e.g., child object handling) will help increase our understanding of how the observed relationships may be mutually driven.

In conclusion, the present study revealed the specific role of parental scaffolding behaviors on JA behaviors among children with autism and how this relates to their vocabulary. Both JA measurements positively predicted vocabulary measures and parent object handling and face monitoring had significant moderating effects on these relationships. These results add to our understanding of the potential mechanisms underlying how social attention is related to language development and have potential applications in the development of early parent-mediated autism interventions that focus on parental play behaviors.

REFERENCES

[1] M. Tomasello and M. J. Farrar, "Joint attention and early language," *Child Dev.*, vol. 57, no. 6, pp. 1454–1463, 1986, doi: 10.2307/1130423.

[2] G. Vivanti, "Ask the Editor: What is the Most Appropriate Way to Talk About Individuals with a Diagnosis of Autism?," *J. Autism Dev. isord.*,

vol. 50, no. 2, pp. 691–693, Feb. 2020, doi: 10.1007/s10803-019-04280-x.

[3] M. Morales *et al.*, "Responding to Joint Attention Across the 6-Through 24-Month Age Period and Early Language Acquisition," *J. Appl. Dev. Psychol.*, vol. 21, no. 3, pp. 283–298, May 2000, doi: 10.1016/S0193-3973(99)00040-4.

[4] N. Akhtar, F. Dunham, and P. J. Dunham, "Directive interactions and early vocabulary development: the role of joint attentional focus," *J. Child Lang.*, vol. 18, no. 1, pp. 41–49, Feb. 1991, doi: 10.1017/s0305000900013283.

[5] P. Mundy and A. Gomes, "Individual differences in joint attention skill development in the second year," *Infant Behav. Dev.*, vol. 21, no. 3, pp. 469–482, Jan. 1998, doi: 10.1016/S0163-6383(98)90020-0.

[6] T. Jellema *et al.*, "Involuntary interpretation of social cues is compromised in autism spectrum disorders," *Autism Res.*, vol. 2, pp. 192–204, 2009.

[7] K. Bottema-Beutel, "Associations between joint attention and language in autism spectrum disorder and typical development: A systematic review and meta-regression analysis," *Autism Res.*, vol. 9, no. 10, pp. 1021–1035, 2016, doi: 10.1002/aur.1624.

[8] H. Tager-Flusberg, R. Paul, and C. Lord, *Language and communication in autism.*, vol. 1. in Handbook of autism and pervasive developmental disorder., no. 3, vol. 1. New York: Wiley, 2005.

[9] R. J. Luyster, M. B. Kadlec, A. Carter, and H. Tager-Flusberg, "Language assessment and development in toddlers with autism spectrum disorders," *J. Autism Dev. Disord.*, vol. 38, no. 8, pp. 1426–1438, Sep. 2008, doi: 10.1007/s10803-007-0510-1.

[10] P. Yoder, L. R. Watson, and W. Lambert, "Value-Added Predictors of Expressive and Receptive Language Growth in Initially Nonverbal Preschoolers with Autism Spectrum Disorders," *J. Autism Dev. Disord.*, vol. 45, no. 5, pp. 1254–1270, May 2015, doi: 10.1007/s10803-014-2286-4.

[11] [M. Siller and M. Sigman, "Modeling longitudinal change in the language abilities of children with autism: Parent behaviors and child characteristics as predictors of change," *Dev. Psychol.*, vol. 44, no. 6, pp. 1691–1704, 2008, doi: 10.1037/a0013771.

[12] K. E. Pickard and B. R. Ingersoll, "Brief report: High and low level initiations of joint attention, and response to joint attention: Differential relationships with language and imitation," *J. Autism Dev. Disord.*, vol. 45, no. 1, pp. 262–268, 2015, doi: 10.1007/s10803-014-2193-8.

[13] S. Shumway and A. M. Wetherby, "Communicative acts of children with autism spectrum disorders in the second year of life," *J. Speech Lang. Hear. Res. JSLHR*, vol. 52, no. 5, pp. 1139–1156, Oct. 2009, doi: 10.1044/1092-4388(2009/07-0280).

[14] V. H. Bal *et al.*, "Predictors of longer-term development of expressive language in two independent longitudinal cohorts of language-delayed preschoolers with Autism Spectrum Disorder," *J. Child Psychol. Psychiatry*, vol. 61, no. 7, pp. 826–835, 2020, doi: 10.1111/jcpp.13117.

[15] P. Mundy, M. Sigman, J. Ungerer, and T. Sherman, "Defining the social deficits of autism: The contribution of non-verbal communication measures," *J. Child Psychol. Psychiatry*, vol. 27, no. 5, pp. 657–669, 1986.

[16] P. Mundy, J. Block, C. Delgado, Y. Pomares, A. V. Van Hecke, and M. V. Parlade, "Individual Differences and the Development of Joint Attention in Infancy," *Child Dev.*, vol. 78, no. 3, pp. 938–954, 2007, doi: 10.1111/j.1467-8624.2007.01042.x.

[17] R. Bakeman and L. B. Adamson, "Coordinating attention to people and objects in mother-infant and peer-infant interaction," *Child Dev.*, vol. 55, no. 4, pp. 1278–1289, Aug. 1984.

[18] S. Freeman and C. Kasari, "Parent–child interactions in autism: Characteristics of play," *Autism*, vol. 17, no. 2, pp. 147–161, Mar. 2013, doi: 10.1177/1362361312469269.

[19] J. A. Doussard-Roosevelt, C. M. Joe, O. V. Bazhenova, and S. W. Porges, "Mother-child interaction in autistic and nonautistic children: characteristics of maternal approach behaviors and child social responses," *Dev. Psychopathol.*, vol. 15, no. 2, pp. 277–295, 2003, doi: 10.1017/s0954579403000154.

[20] C. Kasari, M. Sigman, P. Mundy, and N. Yirmiya, "Caregiver interactions with autistic children," *J. Abnorm. Child Psychol.*, vol. 16, no. 1, pp. 45–56, Feb. 1988, doi: 10.1007/BF00910499.

[21] H. Yoshida, P. Cirino, S. Mire, and J. Burling, "Parents' gesture adaptations to children with autism spectrum disorders," *J. Child Lang. 1–20*, 2019, doi: 10.1017/S0305000919000497.

[22] E. Perkovich, L. Sun, S. Mire, A. Laakman, U. Sakhuja, and H. Yoshida, "What children with and without ASD see: Similar visual experiences with different pathways through parental attention strategies," *Autism Dev. Lang. Impair.*, vol. 7, p. 23969415221137293, 2022, doi: 10.1177/23969415221137293.

[23] B. Ku, J. D. Stinson, and M. MacDonald, "Parental Behavior Comparisons Between Parents of Children with Autism Spectrum Disorder and Parents of Children Without Autism Spectrum Disorder: A Meta-analysis," *J. Child Fam. Stud.*, vol. 28, no. 6, pp. 1445–1460, Jun. 2019, doi: 10.1007/s10826-019-01412-w.

[24] A. McDuffie and P. Yoder, "Types of Parent Verbal Responsiveness That Predict Language in Young Children With Autism Spectrum Disorder," *J. Speech Lang. Hear. Res.*, vol. 53, no. 4, pp. 1026–1039, Aug. 2010, doi: 10.1044/1092-4388(2009/09-0023).

[25] T. Y. Perryman, A. S. Carter, D. S. Messinger, W. L. Stone, A. E. Ivanescu, and P. J. Yoder, "Brief Report: Parental Child-Directed Speech as a Predictor of Receptive Language in Children with Autism Symptomatology," *J. Autism Dev. Disord.*, vol. 43, no. 8, pp. 1983–1987, Aug. 2013, doi: 10.1007/s10803-012-1725-3.

[26] M. Siller and M. Sigman, "The Behaviors of Parents of Children with Autism Predict the Subsequent Development of Their Children's Communication," *J. Autism Dev. Disord.*, vol. 32, no. 2, pp. 77–89, Apr. 2002, doi: 10.1023/A:1014884404276.

[27] L. Sun and H. Yoshida, "Why the parent's gaze is so powerful in organizing the infant's gaze: The relationship between parental referential cues and infant object looking," *Infancy*, vol. 27, no. 4, pp. 780–808, 2022, doi: 10.1111/infa.12475.

[28] J. Yurkovic-Harding *et al.*, "Children with ASD establish joint attention during free-flowing toy play without face looks," *Curr. Biol.*, vol. 32, May 2022, doi: 10.1016/j.cub.2022.04.044.

[29] J. R. Yurkovic *et al.*, "Using head-mounted eye tracking to examine visual and manual exploration during naturalistic toy play in children with and without autism spectrum disorder," *Sci. Rep.*, vol. 11, no. 1, p. 3578, Feb. 2021, doi: 10.1038/s41598-021-81102-0.

[30] J. S. Horst, L. K. Samuelson, S. C. Kucker, and B. McMurray, "What's new? Children prefer novelty in referent selection," *Cognition*, vol. 118, no. 2, pp. 234–244, Feb. 2011, doi: 10.1016/j.cognition.2010.10.015.

[31] L. M. Dunn and D. M. Dunn, *Peabody Picture Vocabulary Test, Fourth Edition*. Minneapolis, MN: NCS Pearson, 2007.

[32] K. Williams, *Expressive Vocabulary Test, Second Edition (EVT-2)*. Circle Pines, MN: AGS, 2007.

[33] C. Kasari, N. Brady, C. Lord, and H. Tager-Flusberg, "Assessing the Minimally Verbal School-Aged Child with Autism Spectrum Disorder," *Autism Res. Off. J. Int. Soc. Autism Res.*, vol. 6, no. 6, pp. 479–493, Dec. 2013, doi: 10.1002/aur.1334.

[34] [34] J. M. Franchak, K. S. Kretch, K. C. Soska, and K. E. Adolph, "Head-mounted eye tracking: a new method to describe infant looking," *Child Dev.*, vol. 82, no. 6, pp. 1738–1750, 2011, doi: 10.1111/j.1467-8624.2011.01670.x.

[35] Datavyu Team, "Datavyu: a video coding tool." Databrary Project, New York University, 2014.

[36] A. F. Hayes, "Introduction to Mediation, Moderation, and Conditional Process Analysis: Third Edition: A Regression-Based Approach," Guilford Press.

[37] C. Harrop, D. Jones, S. Zheng, S. W. Nowell, B. A. Boyd, and N. Sasson, "Sex differences in social attention in autism spectrum disorder," *Autism Res.*, vol. 11, no. 9, pp. 1264–1275, 2018, doi: 10.1002/aur.1997.

[38] S. Perzolli, A. Bentenuto, G. Bertamini, and P. Venuti, "Play with Me: How Fathers and Mothers Play with Their Preschoolers with Autism," *Brain Sci.*, vol. 13, no. 1, Art. no. 1, Jan. 2023, doi: 10.3390/brainsci13010120.

[39] K. Toth, J. Munson, A. N. Meltzoff, and G. Dawson, "Early Predictors of Communication Development in Young Children with Autism Spectrum Disorder: Joint Attention, Imitation, and Toy Play," *J. Autism Dev. Disord.*, vol. 36, no. 8, pp. 993–1005, Nov. 2006, doi: 10.1007/s10803-006-0137-7.

[40] P. Nyström, E. Thorup, S. Bölte, and T. Falck-Ytter, "Joint Attention in Infancy and the Emergence of Autism," *Biol. Psychiatry*, vol. 86, no. 8, pp. 631–638, Oct. 2019, doi: 10.1016/j.biopsych.2019.05.006.

[41] J. Burling and H. Yoshida, "Visual constancies amidst changes in handled objects for 5- to 24-month-old infants," *Child Dev. Child Dev.*, vol. 90, no. 2, pp. 452–461, 2019.