Children with ASD show diminished input statistics for word learning during caregiver-child interaction

Catherine Bianco (cat.bianco@utexas.edu)
Department of Psychology and Center for Perceptual Systems, 108 E Dean Keeton Street
Austin, TX 78712 USA

Chen Yu (chen.yu@austin.utexas.edu)
Department of Psychology and Center for Perceptual Systems, 108 E Dean Keeton Street
Austin, TX 78712 USA

Abstract

This study explored cross-situational word learning in children with and without autism spectrum disorder (ASD). Children learn words by mapping object names in caregiver utterances to objects in their visual field. We examined the confluence of caregiver object naming and child visual attention in children with and without ASD at play. Head-mounted eye-tracking revealed that children with ASD spent less time attending to named objects than typically developing (TD) children. In both groups, learning input improved as children accrued increased looking time to named objects across multiple naming events. However, for objects with high quantity of naming events, TD children had higher quality learning input than children with ASD. These findings suggest that the input statistics of social interaction are less conducive to word learning in children with ASD. This work has important implications for clinical interventions to scaffold word learning.

Keywords: autism spectrum disorder; cross-situational word learning; caregiver-child interaction; eye tracking

Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental condition characterized by atypical social development and repetitive sensorimotor activity that emerges before 3 years of age (American Psychiatric Association, 2013). Language impairment is one of the earliest and most common symptoms of ASD. Approximately 60% of preschoolers with ASD have some form of language delay, and 20%-30% are completely nonverbal (Latrèche et al., 2024; Rose et al., 2016). Early language delays can have cascading effects on social and cognitive development (Fernell et al., 2013; Iverson et al., 2023; Leezenbaum et al., 2014). It is vital to understand the cognitive mechanisms of language impairment in order to fine-tune intervention to the root cause of delay and improve developmental outcomes.

A well-established view suggests that language delay in ASD arises from differences in social cognition. Specifically, children with ASD may not be cognitively equipped to learn words from social cues in day-to-day life (Baron-Cohen et al., 1997; Preissler & Carey, 2005). Empirical evidence for global deficit in social learning is mixed. Some studies suggest that children with ASD show diminished ability to learn words in social contexts, such as interactions with social partners or videos of social scenes (Baron-Cohen et al., 1997; Lund et al., 2025; Preissler & Carey, 2005). However, recent laboratory studies indicate that children with and without ASD have similar abilities to learn words through social interaction (Field et al., 2019; Hartley et al., 2020; Luyster & Lord, 2009), and that disparities in past findings relate to idiosyncrasies in task structure, such as salience of social cue (Liu et al., 2024) and quantity of trials (Bean Ellawadi & McGregor, 2016). As such, the influence of social cognitive deficits may be modulated by learning context.

We examine an alternative explanation for atypical language development in ASD. It is possible that delays emerge from differences in input statistics in everyday social interaction. Statistical learning has been demonstrated to be a powerful mechanism in early language learning. In typically developing (TD) children, the natural dynamics of caregiver-child interaction provide rich statistical input for learning: If a caregiver says the name of an object while the child is simultaneously looking at the object, they have an opportunity to map the name to the object (Clerkin & Smith, 2022; Schroer & Yu, 2023; Yu & Smith, 2012). Caregivers often refer to the same object multiple times within a single play session, and they tend to name objects already in children's field of view (Tamis-LeMonda et al., 2014; Yu & Smith, 2012). As such, TD children glean word-object mappings by tracking the statistical regularity of word-object co-occurrences across multiple exposures, a learning process called cross-situational learning. During repeated naming events, children accumulate more time looking to the named object (the "target") compared to distractor objects and thereby build robust word-object mappings (Figure 1). Indeed, sustained attention to targets during caregiver naming events predicts vocabulary size in TD infants (Yu et al., 2019).

There is a paucity of research examining ASD-related learning differences in naturalistic settings. A handful of studies have used head-mounted eye-tracking to examine visual exploration during social interaction in children with ASD (Yurkovic et al., 2020, 2021; Yurkovic-Harding et al., 2022), but none have explored cross-situational word learning. Thus, it remains unclear whether the input statistics for word learning in everyday environments is comparable in children with and without ASD. It is possible that small differences in visual input during day-to-day interactions may accumulate over time and contribute to developmental language delay. For example, consistently looking at

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

distractor objects during naming events may derail word learning because it can lead children to make incorrect word-object mappings.

To understand whether language delays in ASD are related to input statistics, we examined the confluence of caregiver naming and child visual attention during free-flowing play sessions in children with and without ASD. If language differences reflect the quantity and quality of input statistics, we predict that children with ASD will spend less time attending to target objects and more time attending to distractors during naming events compared to TD children. We expect to find this difference in learning input in individual naming events and cross-situational learning opportunities, where children are exposed to multiple naming events for a single target.

Methods

Dataset

Study procedures match those in Yurkovic-Harding et al. (2022). Children with and without ASD engaged in free-flowing play with a caregiver. Dyads were provided with 24 age-appropriate toys, and children wore a Positive Science head-mounted eye-tracker (www.positivescience.com) that recorded where they looked during play (Figure 2). Fourteen children with ASD (M = 38.1 months, SD = 7.34 months, 5 girls) and 36 TD children (M = 18.8 months, SD = 3.19 months, 16 girls) contributed head-mounted eye-tracking data (see Table 1 for demographic breakdown). Sample size is comparable to previous research examining cross-situational word learning in children with ASD (Bean Ellawadi & McGregor, 2016; Rothwell et al., 2024).

Table 1: Sample demographic information

	TD (N=36)	ASD (N=14)
Age (months)		
Mean (SD)	18.8 (3.19)	38.1 (7.34)
Sex		
Female	16 (44.4%)	5 (35.7%)
Male	20 (55.6%)	9 (64.3%)

In order to control for cognitive ability and language level, we matched the ASD and TD groups based on language ability instead of chronological age, following convention in prior work (Bean Ellawadi & McGregor, 2016; Rothwell et al., 2024; Venker, 2019). The age range in the TD group is comparable to range of Mullen Scales of Early Learning (MSEL; Mullen, 1995) age equivalence (AE) scores for receptive and expressive language in the ASD group. See Table 2 for ASD group characterization information.

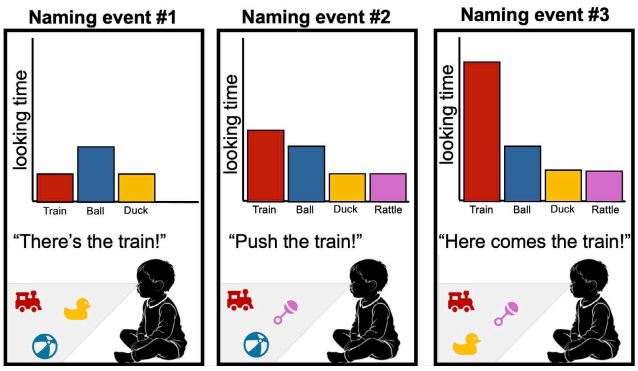

Figure 1: Example of looking time across multiple naming events. Quoted text indicates caregiver speech. Child accumulates increased looking time to target object (train) compared to distractors in their visual field across three caregiver naming events.

Table 2: ASD group characterization information

	(N=14)
ADOS RRB Severity Score	
Mean (SD)	7.43 (1.34)
ADOS Social Affect Severity Score	
Mean (SD)	7.64 (1.60)
ADOS Total Severity Score	
Mean (SD)	7.64 (1.34)
MSEL Expressive Language AE (months)	
Mean (SD)	15.0 (7.09)
MSEL Receptive Language AE (months)	
Mean (SD)	13.4 (4.94)

* ADOS = Autism Diagnostic Observation Schedule, Second Edition (Lord et al., 2012)

Figure 2: Caregiver-child interaction and child's first-person view from head-mounted eye-tracker. Pink crosshair indicates child's gaze location.

Data Coding and Processing

Caregiver speech was annotated by human coders using Audacity (www.audacityteam.org). Coders scored the onset and offset of utterances separated by 400ms or more. Then, they identified naming events, or utterances that included the word for one of the 24 free-play toys.

We annotated children's visual attention during naming events to assess visual input for word learning. A custom algorithm was used to identify the timing of individual looks for each participant based on their eye movements and blinks. Then, highly trained human coders manually annotated the locations of fixations within participants' field of view using a custom coding software. Location was coded as one of the 24 toys or as an irrelevant area of the room, e.g., the carpet. A primary coder annotated 100% of each participant's video, and a secondary coder annotated 10% of each video to establish interrater reliability. Average interrater agreement was > 90% and Cohen's kappa was > .85.

We chose to examine children's visual attention within a 3-second interval after the onset of naming events because sustained visual attention to named objects lasting 3 seconds or more has been shown to support word learning in prior work (Yu et al., 2019). This interval also controls for variability in caregiver utterance length. Within each 3-second interval, we calculated the time children spent looking at each free-play toy. Duration of attention to individual toys ranged from 0 to 3 seconds, and children often looked at multiple toys within a single interval (Figure 3).

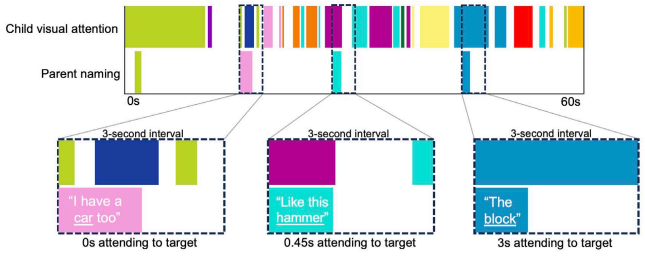

Figure 3: Time series of child visual attention and caregiver object naming during 60 seconds of free-flowing interaction. Bar colors denote the identity of attended and named objects. Quoted text is transcribed caregiver utterance with target object name underlined. Magnified panels depict the 3-second interval after onset of naming which was used to calculate child looking time to target and distractor objects. Looking time to target object for each magnified naming event is indicated below each panel.

Child Attention Within Individual Naming Events

In order to map heard words to objects in the visual field, children must attend to the target object and disregard distractor objects. If children attend to distractors for a large proportion of the naming event, they may mistakenly map the target name to a distractor object. For our dependent variable, we sought to create a metric for the input statistics of word learning that accounted for both attention to target and attention to distractors. For each 3-second interval from the onset of the naming event, we calculated the time children spent attending to the target object. Then, we divided target looking time by the total looking time for all 24 free-play toys. Event-level metrics ranged from 0 to 1, with higher values indicating proportionately greater attention to target object and therefore better visual input for word-object mapping. Lower values indicated proportionately greater attention to distractor objects than to the target and suboptimal input for word-object mapping.

Child Attention Across Multiple Naming Events

The dynamics of caregiver-child play afford children multiple opportunities to map a name to an object, as caregivers often name the same object multiple times. To capture how repeated naming events confer benefit for word learning, we created a second metric that captures accumulated looking across multiple naming events for the same object. We summed the looking time for each of the 24 toys across all naming events for a given target. Then, we divided total time looking to the target by overall looking time for all objects. Cross-situational metrics ranged from 0 to 1, with positive values indicating better input for word learning and negative values indicating worse input for word learning.

Statistical Analysis

We used the lme4 package in R to conduct linear mixed models on event-level word learning metric and cross-situational word learning metric with fixed effects of group (TD vs. ASD) and random effects of participant and object. To explore whether quantity of naming events influences learning input for TD and ASD participants, we did a median split on the number of naming events and compared metrics for each diagnostic group within objects with higher and lower event quantity. Group differences in learning metric for objects with high and low event quantity were evaluated separately using linear mixed models with fixed effects of group (TD vs. ASD) and random effects of participant and object.

Play Session Information

TD children engaged in longer free-play sessions than children with ASD (ASD = 4.87 minutes, TD = 7.14 minutes; $t(45.91) = 4.83$, $p < .001$). Session times in the ASD group were shorter because children were relatively less tolerant of the head-mounted eye-tracker. Overall, the TD group contributed 1508 naming events, and the ASD group contributed 391 naming events to the event-level word learning metric.

Results

Caregiver Naming and Child Looking

We started by examining descriptive statistics of caregiver naming and child attention. TD children had more naming events during play sessions (ASD = 27.93 events, TD = 41.89 events; $t(32.18) = 2.24$, $p = .03$) and more naming events per object (ASD = 2.21 events, TD = 2.96 events; $t(26.44) = 2.75$, $p = .01$). This discrepancy is related to the difference in overall session length, where TD children played for longer periods of time. However, the rate of naming events per minute was similar between groups (ASD = 5.72 events/minute, TD = 5.69 events/minute; $t(17.99) = .03$, $p = .98$), and there was no significant difference in the rate that children looked at objects overall throughout their play sessions (ASD = 28.40 looks/minute, TD = 24.04 looks/minute; $t(19.14) = 2.03$, $p = .056$). Also, the number of objects with 1 or more naming events was not different between groups (ASD = 11.79 objects, TD = 13.50 objects; $t(18.82) = 1.00$, $p = .33$). In sum, the rate that TD and ASD children heard object names and visually attended to objects during play was comparable, and they were exposed to a similar variety of object names. See Table 3 for a summary of play session information. Next, we examined how this visual input created opportunities for word-object mapping during play by comparing word learning metrics between groups.

Table 3: Frequencies of Caregiver Naming and Child Looking Behaviors

	TD	ASD	*p*
Naming events per minute			
Mean (SD)	5.69 (2.08)	5.72 (3.02)	n.s.
Naming events per session			
Mean (SD)	41.89 (24.2)	27.93 (17.8)	0.03
Naming events per object			
Mean (SD)	2.96 (0.95)	2.21 (0.84)	0.01
Objects named per session			
Mean (SD)	13.5 (4.32)	11.79 (5.83)	n.s.
Object looks per minute			
Mean (SD)	24.04 (5.51)	28.4 (7.25)	n.s.

Difference in Child Attention Within Individual Naming Events

Similar to past work, the distribution of event-level word learning metrics across individual naming events was bimodal (Yu et al., 2021). A substantial proportion of naming events had either very low target attention (first peak) or very high target attention (second peak) in both the TD and ASD groups. However, the distribution differed in relation to ASD diagnosis such that children with ASD had a greater proportion of naming events where they did not attend to the target object at all, indicated by the height of the first peak (Figure 4). A linear mixed effects model confirmed that the TD group had significantly higher average event-level learning metric than the ASD group (TD $M = 0.46$, $Mdn = 0.43$, $SD = 0.41$, ASD $M = 0.35$, $Mdn = 0.17$, $SD = 0.39$; $b = 0.09$, $SE = 0.04$, $t = 2.49$, $p = .02$; Figure 4). As such, TD children spent a greater proportion of time attending to target objects during naming events than children with ASD. These results suggest that the visual input for word learning during naturalistic social interaction in TD participants is more conducive to correct word-object mapping than in ASD participants.

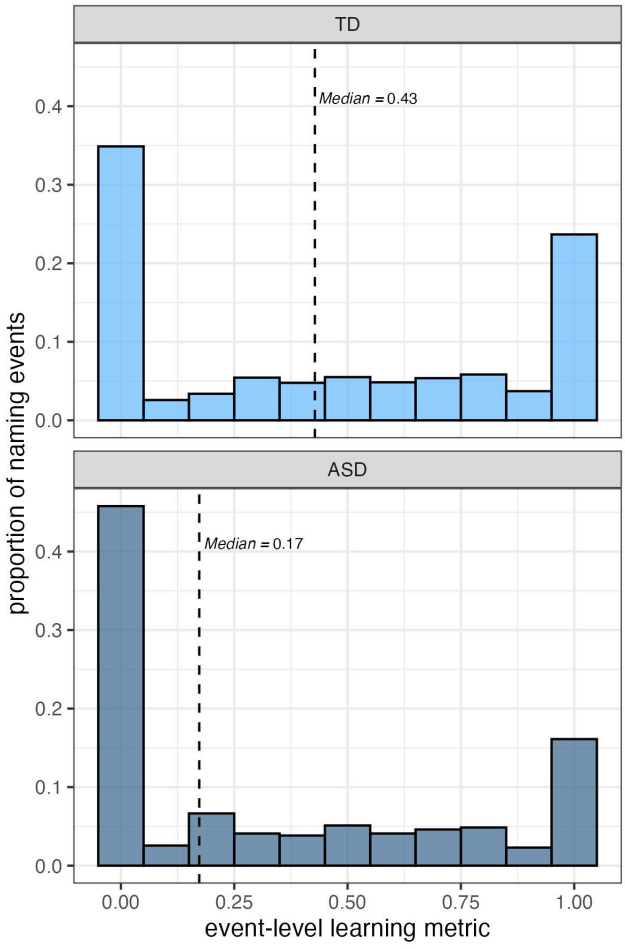

Figure 4: Distributions of event-level word learning metrics for TD children and children with ASD. Bar height indicates proportion of events in each bin for each group. Dashed line indicates median of each group. The TD group had higher average event-level learning metric than the ASD group, indicating more optimal input for word learning ($p = .02$).

Difference in Child Attention Across Multiple Naming Events

The distributions of cross-situational word learning metric were also bimodal in both TD and ASD groups (Figure 5). Crucially, the median values for the cross-situational learning metric were higher than the median values for the event-level learning metric. This suggests that learning input improved as children had more chances to attend to the target object across multiple naming events. Average cross-situational word learning metric was higher in the TD group (TD M = 0.47, Mdn = 0.48, SD = 0.33, ASD M = 0.38, Mdn = 0.33, SD = 0.35; b = 0.08, SE = 0.04, t = 2.13, p = .04; Figure 5). As such, TD children accumulated more optimal input for word-object mapping throughout their play sessions than children with ASD.

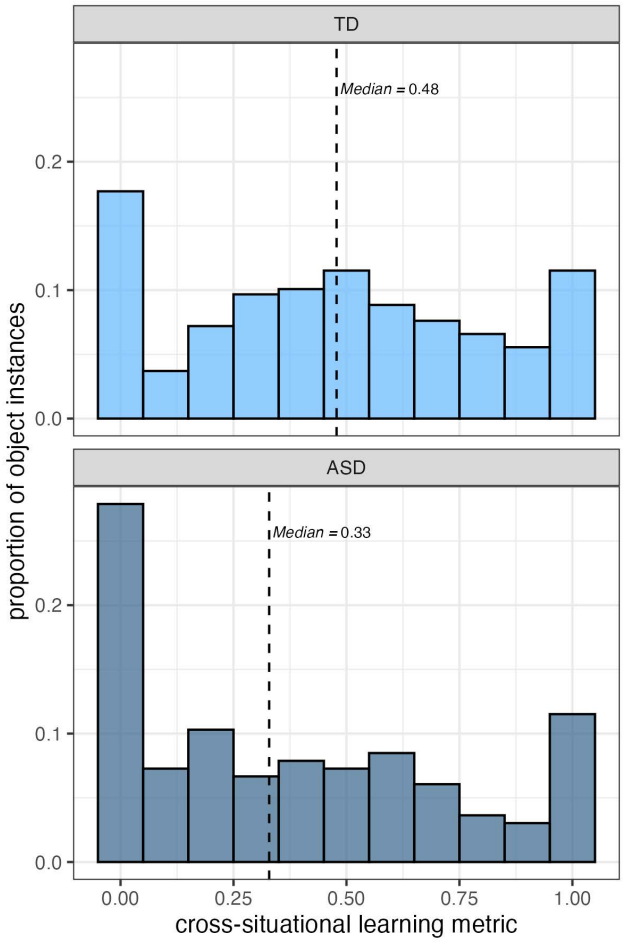

Figure 5: Distributions of cross-situational learning metrics. Bar height indicates the proportion of object instances in each bin. Dashed lines denote the median metric for each group. On average, TD children had higher cross-situational learning metrics than children with ASD (p = .04).

Relation Between Frequency of Naming Events and Cross-situational Word Learning Metric

Do ASD-related differences in learning metric vary in relation to the quantity of naming events? To answer this question, we compared average cross-situational learning metrics for objects with naming events above and below the median number of events per object. Both ASD and TD groups had a median of two events per object. Across all of the TD children, there were 255 object instances with two or less events, and 231 object instances with greater than 2 events. In the ASD group, there were 110 object instances with two or less events and 55 object instances with greater than 2 events. Group differences were modulated by event quantity. For objects with two or less events, the ASD group had lower average cross-situational learning metric than TD children, but the difference was not significant (TD M = 0.46, Mdn = 0.47, SD = 0.37, ASD M = 0.40, Mdn = 0.33, SD = 0.39; b = 0.06, SE = 0.05, t = 1.27, p = .21; Figure 6). For objects with greater than two events, the ASD group had significantly lower average cross-situational learning metric than the TD group (TD M = 0.48, Mdn = 0.49, SD = 0.27, ASD M = 0.34, Mdn = 0.31, SD = 0.26; b = .12, SE = .05, t = 2.39, p = .02; Figure 6). Thus, ASD children have less optimal learning input than TD children for objects with high quantity of naming events. This disparity could reflect underlying differences in statistical learning ability.

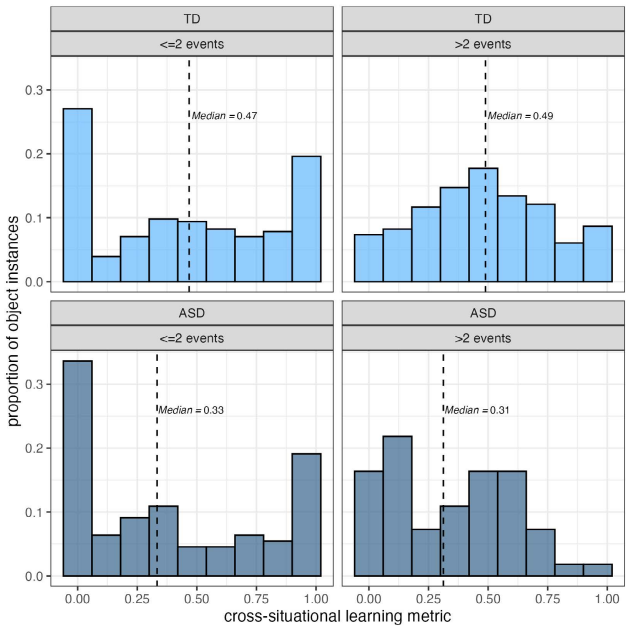

Figure 6: Cross-situational learning metric distributions for objects with two or less events and greater than two events. Bar heights indicate proportion of instances in each bin, and dashed lines denote median. Group differences were modulated by event quantity, where children with ASD had significantly lower average learning metric than TD children for objects with more than two events (p = .02).

Discussion

The present study examined word learning during free-flowing play in children with and without ASD. Although visual exploration and exposure to object names was similar between groups, we found significant differences in learning input. During individual caregiver naming events, children with ASD spent less time attending to target objects than TD children. In both ASD and TD groups, learning input improved when young learners aggregated statistical evidence across multiple naming events. When caregivers named the same object multiple times, the proportion of time children spent attending to the target increased. This suggests that children in both groups were able to retain and learn from statistical information in real time. However, group disparity in input statistics for objects with high quantity of naming events suggests that TD children may have a greater capacity to harness statistical information for word learning than children with ASD.

Our findings underscore the critical role of input statistics in early language learning. During free-flowing play, children with ASD receive less optimal visual input for word learning than TD children. It is possible that these small variations in moment-to-moment word learning input accumulate and contribute to larger disparity in developmental outcome.

Future work should examine why children with ASD spend relatively less time attending to target objects and more time attending to distractors during naming events. It is possible that differences in learning input reflect diminished statistical learning ability in children with ASD. Several studies demonstrate that children with ASD have intact statistical learning ability in structured experimental tasks (Bean Ellawadi & McGregor, 2016; Field et al., 2019; Haebig et al., 2017; Hartley et al., 2020; Venker, 2019). However, real-world environments have less statistical regularity and more uncertainty than laboratory tasks. For instance, cross-situational learning experiments typically only include one or two distractor objects for each word-target pairing. In contrast, children may encounter dozens of distractors in their visual field during each naming event in everyday activity. As such, ASD-related differences in statistical learning may only manifest in laboratory tasks that approximate their day-to-day environment, as in the current study.

ASD-related differences in input statistics could also be associated with the contingency of child and caregiver behavior during free-flowing interaction. Children are in charge of where they look, and by extension, what they learn (Smith et al., 2018; Yu & Smith, 2011). Caregivers' propensity to follow children's lead during play (e.g., naming an object that is already in their child's field of view) also modulates the quality of individual naming events (Tamis-LeMonda et al., 1996, 2014; Yu & Smith, 2012). As such, both children's and caregivers' actions jointly constrain the quality of learning input.

Several studies suggest that parents of children with ASD spend less time following children's lead during play. Steiner et al. (2018) examined the contingency of dyadic play in 12-month-olds with and without older siblings with ASD (indicating elevated likelihood of diagnosis due to genetic predisposition). They demonstrated that parents of infants with ASD siblings tended to be more directive than parents of infants with TD siblings. For example, parents were more likely to redirect infants' attention to new toys instead of engaging with toys already in infants' hands. Longitudinal work shows that parents of infant siblings subsequently diagnosed with ASD are more directive during free-play at 12 and 15 months of age than parents of TD infants (Wan et al., 2013). Researchers suggest that parents of infants with elevated likelihood for ASD diagnosis may be especially attuned to early signs of developmental delay and engage in directive parenting in an effort to bolster learning (Ozonoff et al., 2009; Steiner et al., 2018; Wan et al., 2013). However, some directive parenting behaviors may not be ideal for word learning. For instance, referring to a toy outside of children's focus necessitates a shift in attention to look at the toy and learn its name. Future work should examine how parent directiveness relates to word learning input in children with and without ASD.

There are two important caveats to our results. First, we did not measure children's familiarity with the free play toys before they participated in the study. As such, it is possible that prior knowledge of the objects and their names influenced children's looking patterns. Additionally, we did not include a learning assessment at the end of the play session to verify that proportion of time attending to the target object influenced children's knowledge of word-object mappings as in other work (Schroer & Yu, 2023; Yu et al., 2019). Future research should directly assess children's prior knowledge and real-time learning to further explore the mechanisms of word learning in social interaction.

Conclusion

The current study builds on previous research to illuminate the underlying mechanisms of language variability in ASD. We demonstrated significant differences in visual input for word learning between diagnostic groups. Compared to TD children, children with ASD spend proportionately less time attending to named objects and more time attending to distractors during instances where their caregiver says the name of an object. If these small differences in input statistics pervade throughout day-to-day social interaction, they may contribute to language delay. Future work will examine differences in contingency of dyadic behaviors related to input statistics during free-flowing play. This line of inquiry has important implications for the development of targeted intervention that addresses the root cause of language delay.

Acknowledgments

This work was supported by NIH R01DC017925 and NIH R01HD104624 to CY. CB was supported by an NSF GRFP. We thank the members of the Developing Intelligence Lab for their contributions to data collection and coding and for their helpful feedback throughout the research process. Our sincerest gratitude to participating families for their time.

References

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). American Psychiatric Association.

Baron-Cohen, S., Baldwin, D. A., & Crowson, M. (1997). Do Children with Autism Use the Speaker's Direction of Gaze Strategy to Crack the Code of Language? *Child Development*, *68*(1), 48–57. https://doi.org/10.1111/j.1467-8624.1997.tb01924.x

Bean Ellawadi, A., & McGregor, K. K. (2016). Children with ASD can use gaze to map new words. *International Journal of Language & Communication Disorders*, *51*(2), 212–218. https://doi.org/10.1111/1460-6984.12192

Clerkin, E. M., & Smith, L. B. (2022). Real-world statistics at two timescales and a mechanism for infant learning of object names. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(18), e2123239119. https://doi.org/10.1073/pnas.2123239119

Fernell, E., Eriksson, M. A., & Gillberg, C. (2013). Early diagnosis of autism and impact on prognosis: A narrative review. *Clinical Epidemiology*, *5*, 33–43. https://doi.org/10.2147/CLEP.S41714

Field, C., Lewis, C., & Allen, M. L. (2019). Referent selection in children with Autism Spectrum Condition and intellectual disabilities: Do social cues affect word-to-object or word-to-location mappings? *Research in Developmental Disabilities*, *91*, 103425. https://doi.org/10.1016/j.ridd.2019.05.004

Haebig, E., Saffran, J. R., & Ellis Weismer, S. (2017). Statistical word learning in children with autism spectrum disorder and specific language impairment. *Journal of Child Psychology and Psychiatry*, *58*(11), 1251–1263. https://doi.org/10.1111/jcpp.12734

Hartley, C., Bird, L.-A., & Monaghan, P. (2020). Comparing cross-situational word learning, retention, and generalisation in children with autism and typical development. *Cognition*, *200*, 104265. https://doi.org/10.1016/j.cognition.2020.104265

Iverson, J. M., West, K. L., Schneider, J. L., Plate, S. N., Northrup, J. B., & Roemer Britsch, E. (2023). Early development in autism: How developmental cascades help us understand the emergence of developmental differences. In *Advances in Child Development and Behavior* (Vol. 64, pp. 109–134). Elsevier. https://doi.org/10.1016/bs.acdb.2022.10.005

Latrèche, K., Godel, M., Franchini, M., Journal, F., Kojovic, N., & Schaer, M. (2024). Early trajectories and moderators of autistic language profiles: A longitudinal study in preschoolers. *Autism*, *28*(12), 3043–3062. https://doi.org/10.1177/13623613241253015

Leezenbaum, N. B., Campbell, S. B., Butler, D., & Iverson, J. M. (2014). Maternal verbal responses to communication of infants at low and heightened risk of autism. *Autism*, *18*(6), 694–703. https://doi.org/10.1177/1362361313491327

Liu, L., Li, S., Tian, L., Yao, X., Ling, Y., Chen, J., Wang, G., & Yang, Y. (2024). The Impact of Cues on Joint Attention in Children with Autism Spectrum Disorder: An Eye-Tracking Study in Virtual Games. *Behavioral Sciences*, *14*(10), Article 10. https://doi.org/10.3390/bs14100871

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). Autism diagnostic observation schedule (ADOS-2) modules 1–4. *Western Psychological Services*.

Lund, S., Rothwell, C., Monaghan, P., & Hartley, C. (2025). A meta-analysis of word learning in autistic and neurotypical children: Distinguishing noun-referent mapping, retention, and generalisation. *Developmental Review*, *75*, 101171. https://doi.org/10.1016/j.dr.2024.101171

Luyster, R., & Lord, C. (2009). Word Learning in Children with Autism Spectrum Disorders. *Developmental Psychology*, *45*(6), 1774–1786. https://doi.org/10.1037/a0016223

Mullen, E. (1995). Mullen scales of early learning manual. *American Guidance Service*.

Ozonoff, S., Young, G. S., Steinfeld, M. B., Hill, M. M., Cook, I., Hutman, T., Macari, S., Rogers, S. J., & Sigman, M. (2009). How Early Do Parent Concerns Predict Later Autism Diagnosis? *Journal of Developmental & Behavioral Pediatrics*, *30*(5), 367. https://doi.org/10.1097/DBP.0b013e3181ba0fcf

Preissler, M. A., & Carey, S. (2005). The role of inferences about referential intent in word learning: Evidence from autism. *Cognition*, *97*(1), B13–B23. https://doi.org/10.1016/j.cognition.2005.01.008

Rose, V., Trembath, D., Keen, D., & Paynter, J. (2016). The proportion of minimally verbal children with autism spectrum disorder in a community-based early intervention programme. *Journal of Intellectual Disability Research*, *60*(5), 464–477. https://doi.org/10.1111/jir.12284

Rothwell, C., Westermann, G., & Hartley, C. (2024). How do Autistic and Neurotypical Children's Interests Influence their Accuracy During Novel Word Learning? *Journal of Autism and Developmental Disorders*, *54*(9), 3301–3315. https://doi.org/10.1007/s10803-023-06066-8

Schroer, S. E., & Yu, C. (2023). Looking is not enough: Multimodal attention supports the real-time learning of new words. *Developmental Science*, *26*(2), e13290. https://doi.org/10.1111/desc.13290

Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The Developing Infant Creates a Curriculum for Statistical Learning. *Trends in Cognitive Sciences*, *22*(4), 325–336. https://doi.org/10.1016/j.tics.2018.02.004

Steiner, A. M., Gengoux, G. W., Smith, A., & Chawarska, K. (2018). Parent–Child Interaction Synchrony for Infants At-Risk for Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *48*(10), 3562–3572. https://doi.org/10.1007/s10803-018-3624-8

Tamis-LeMonda, C. S., Bornstein, M. H., Baumwell, L., & Melstein Damast, A. (1996). Responsive Parenting in the Second Year: Specific Influences on Children's Language and Play. *Early Development and Parenting*, *5*(4), 173–183.

Tamis-LeMonda, C. S., Kuchirko, Y., & Song, L. (2014). Why Is Infant Language Learning Facilitated by Parental Responsiveness? *Current Directions in Psychological Science*, *23*(2), 121–126. https://doi.org/10.1177/0963721414522813

Venker, C. E. (2019). Cross-situational and ostensive word learning in children with and without autism spectrum disorder. *Cognition*, *183*, 181–191. https://doi.org/10.1016/j.cognition.2018.10.025

Wan, M. W., Green, J., Elsabbagh, M., Johnson, M., Charman, T., Plummer, F., & BASIS Team. (2013). Quality of interaction between at-risk infants and caregiver at 12-15 months is associated with 3-year autism outcome. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *54*(7), 763–771. https://doi.org/10.1111/jcpp.12032

Yu, C., & Smith, L. (2011). What You Learn is What You See: Using Eye Movements to Study Infant Cross-Situational Word Learning. *Developmental Science*, *14*(2), 165–180.

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

Yu, C., 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

Yu, C., Zhang, Y., Slone, L. K., & Smith, L. B. (2021). The infant's view redefines the problem of referential uncertainty in early word learning. *Proceedings of the National Academy of Sciences*, *118*(52), e2107019118. https://doi.org/10.1073/pnas.2107019118

Yurkovic, J. R., Lisandrelli, G., Shaffer, R. C., Dominick, K. C., Pedapati, E. V., Erickson, C. A., Kennedy, D. P., & Yu, C. (2020). *Examining Sustained Attention in Child-Parent Interaction: A Comparative Study of Typically Developing Children and Children with Autism Spectrum Disorder*.

Yurkovic, J. R., Lisandrelli, G., Shaffer, R. C., Dominick, K. C., Pedapati, E. V., Erickson, C. A., Kennedy, D. P., & Yu, C. (2021). Using head-mounted eye tracking to examine visual and manual exploration during naturalistic toy play in children with and without autism spectrum disorder. *Scientific Reports*, *11*(1), 3578. https://doi.org/10.1038/s41598-021-81102-0

Yurkovic-Harding, J., Lisandrelli, G., Shaffer, R. C., Dominick, K. C., Pedapati, E. V., Erickson, C. A., Yu, C., & Kennedy, D. P. (2022). Children with ASD establish joint attention during free-flowing toy play without face looks. *Current Biology*, *32*(12), 2739-2746.e4. https://doi.org/10.1016/j.cub.2022.04.044