Report

Current Biology

# Children with ASD establish joint attention during free-flowing toy play without face looks

## Highlights

- ASD children engage in joint attention at frequent and typical levels during toy play
- Like TD dyads, ASD dyads follow hands (rather than eyes) to establish joint attention
- In both groups, parents name toys more frequently during moments of joint attention
- These results raise questions about the meaning of joint attention deficits in ASD

## Authors

Julia Yurkovic-Harding, Grace Lisandrelli, Rebecca C. Shaffer, ..., Craig A. Erickson, Chen Yu, Daniel P. Kennedy

## Correspondence

jyurkovi@iu.edu (J.Y.-H.),
chen.yu@austin.utexas.edu (C.Y.),
dpk@iu.edu (D.P.K.)

## In brief

Using head-mounted eye tracking, Yurkovic-Harding et al. find that children with ASD engage in joint attention at frequent and typical levels during toy play with parents. Joint attention is achieved via typical behavioral pathways (hand- rather than gaze-following). This work raises questions about the meaning of joint attention deficits in ASD.

Yurkovic-Harding et al., 2022, Current Biology *32*, 2739–2746
June 20, 2022 
https://doi.org/10.1016/j.cub.2022.04.044

CellPress

Current Biology



Report

# Children with ASD establish joint attention during free-flowing toy play without face looks

Julia Yurkovic-Harding,[1,8,10,*] Grace Lisandrelli,[1,8] Rebecca C. Shaffer,[2,4] Kelli C. Dominick,[3,5] Ernest V. Pedapati,[5,6] Craig A. Erickson,[3,4] Chen Yu,[1,7,9,*] and Daniel P. Kennedy[1,9,*]
[1]Department of Psychological and Brain Sciences, Indiana University – Bloomington, E 10th Street, Bloomington, IN 47405, USA
[2]Department of Pediatrics, Cincinnati Children's Hospital, Burnet Avenue, Cincinnati, OH 45229, USA
[3]Department of Psychiatry and Behavioral Neuroscience, Cincinnati Children's Hospital, Burnet Avenue, Cincinnati, OH 45229, USA
[4]School of Medicine, University of Cincinnati, Eden Avenue, Cincinnati, OH 45229, USA
[5]Division of Psychiatry, Cincinnati Children's Hospital Medical Center, Burnet Avenue, Cincinnati, OH 45229, USA
[6]Department of Psychiatry and Behavioral Neuroscience, University of Cincinnati College of Medicine, Eden Avenue, Cincinnati, OH 45229, USA
[7]Department of Psychology, University of Texas at Austin, Dean Keeton Street, Austin, TX 78712, USA
[8]These authors contributed equally
[9]These authors contributed equally
[10]Lead contact
*Correspondence: jyurkovi@iu.edu (J.Y.-H.), chen.yu@austin.utexas.edu (C.Y.), dpk@iu.edu (D.P.K.)
https://doi.org/10.1016/j.cub.2022.04.044

## SUMMARY

Children's ability to share attention with another person (i.e., achieve joint attention) is critical for learning about their environments in general[1–3] and supporting language and object word learning in particular.[1,4–14] While joint attention (JA) as it pertains to autism spectrum disorder (ASD) is often more narrowly operationalized as arising from eye gaze or explicit pointing cues alone,[2,5,10,15–19] recent evidence demonstrates that JA in natural environments can be achieved more broadly through multiple other pathways beyond gaze and gestures.[2,4,20–31] Here, we use dual head-mounted eye tracking to examine pathways into and characteristics of JA episodes during free-flowing parent-child toy play, comparing children with ASD to typically developing (TD) children. Moments of JA were defined objectively as both the child's and parent's gaze directed to the same object at the same time. Consistent with previous work in TD children,[4,21,25,30–32] we found that both TD and ASD children rarely look at their parent's face in this unstructured free play context. Nevertheless, both groups achieved similarly high rates of JA that far exceeded chance, suggesting the use of alternative pathways into JA. We characterize these alternate pathways, find they occur at similar levels across both groups, and achieve similar ends: namely, for both groups, targets of JA are named more frequently by parents in those moments than non-jointly attended objects. These findings broaden the conceptualization of JA abilities and impairment in ASD and raise questions regarding the mechanistic role of the face-gaze-mediated JA pathway in ASD.

## RESULTS

### Overview

Using simultaneously acquired head-mounted eye tracking of parents and children during toy play, we first quantified the rate at which children and parents look at each other's faces (thought to be a primary cue underlying joint attention; JA). We next examined rates and characteristics of JA, operationalized broadly as moments when both dyad members looked at the same toy at the same time, irrespective of the pathway into JA. Face looking and JA are reported as the percent of usable data from the interaction (henceforth "interaction"). Additional analyses (events per minute, median event duration) are included in the supplemental information. Finally, we examined the behavioral cues that precede JA (i.e., face look, touching and naming of toys). We found remarkable similarity between the typically developing (TD) and autism spectrum disorder (ASD) groups.

### Participant characteristics and data quality

Data were acquired from 24- to 48-month-old ASD ($n = 19$) and TD children ($n = 18$) during free-flowing toy play with their parents. Eye movements of both dyad members were recorded using head-mounted eye trackers (Figure 1). There was no difference in child age between groups ($t(35) = -0.30$, $p = 0.77$, $d = -0.10$), and the results did not differ when the effect of age was regressed (all $p > 0.22$). There was a trend for dyads with TD children to play for longer than dyads with children with ASD (henceforth TD dyads and ASD dyads; $t(35) = 1.78$, $p = 0.08$, $d = 0.59$), but this did not impact results. See Tables S1–S3 for additional details of the child participants.

### Face gaze

#### *Child gaze to faces*

Because gaze to eyes is thought to be critical in facilitating JA, we first examined the percent of the session that children gazed

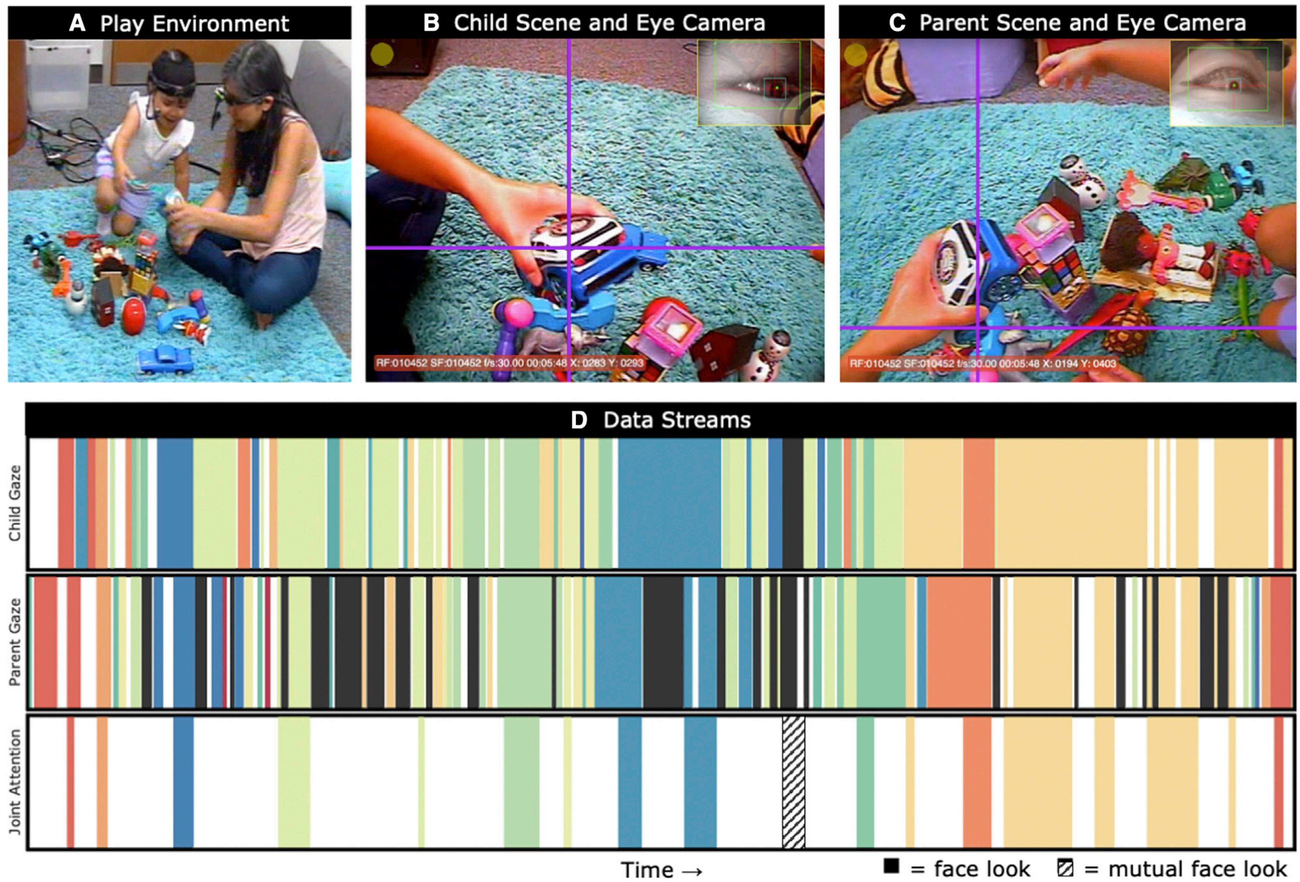


**Figure 1. Experimental set-up and data visualization**

(A) A child and parent play together freely with 24 toys while wearing head-mounted eye trackers.

(B and C) The first-person videos of child (B) and parent (C) feature fixation crosshairs that indicate the object to which each dyad member looked at any given time. These selected images capture an example of JA between the child and parent.

(D) Data streams showing 2 min of the child's gaze behavior (top), the parent's gaze behavior (middle), and the periods of JA and mutual face looks (bottom) during the interaction. The colored bars represent the 25 regions of interest: the 24 toys (each a different color) and the social partner's face (dark gray bars). The white space in the top two streams represents gaze toward other aspects of the scene, while white space in the bottom stream represents non-JA moments and non-mutual face look moments.

to their parent's face. Children rarely looked at the parents' faces during the interaction (mean [SD]; TD = 1.1% [±1.4%]; ASD = 0.8% [±1.0%]; Figure 2A), with no difference between groups ($t(35) = 0.88$, $p = 0.39$, $d = 0.29$). Additionally, a similar number of children in both groups (4 TD and 6 ASD) never looked at the parent's face during the play session ($\chi^2(1, n = 37) = 0.41$, $p = 0.52$).

### *Parent gaze to faces*

In contrast, parents in both groups spent a greater percentage of the interaction looking at their children's faces than did the children at their parents' faces (TD, parent = 10.3% [±6.6%], $t(17) = 6.57$, $p < 0.001$, $d = 1.39$; ASD, parent = 13.1% [±7.0%], $t(18) = 7.91$, $p < 0.001$, $d = 1.55$). This was true for every dyad in both the TD and ASD groups. There was no group difference in the time that parents spent looking at their children's faces ($t(35) = -1.26$, $p = 0.22$, $d = 0.41$; Figure 2B).

### *Mutual face looking*

Moments when both parent and child looked at each other's faces simultaneously were rare (constrained by low levels of face gaze overall) and not different between groups (TD = 0.5% [±1.0%]; ASD = 0.3% [±0.6%]; $t(35) = 1.01$, $p = 0.32$, $d = 0.33$; Figure 2C). A similar number of dyads from each group (10 TD and 11 ASD) established at least one mutual face look ($\chi^2(1, n = 37) = 0.02$, $p = 0.89$); the remaining 8 TD and 8 ASD dyads never established a mutual face look. Nevertheless, mutual face looks still occurred within dyads to a greater extent than if gaze was randomly distributed in time (TD, permuted mean = 0.2% [SD ± 0.1%], $p < 0.001$; ASD, permuted mean = 0.1% [±0.0%], $p = 0.03$). This result suggests a coordination to the social partner's face in both groups.

## Joint attention

Next, we assessed to what extent JA was achieved in this free play context. In concordance with research on JA in TD children,[2,4,20–28] we implement a broad definition of JA: namely, that both social partners look at the same object at the same time, regardless of the specific behavioral cues (e.g., gaze, pointing) used to establish JA. Importantly, we find that dyads

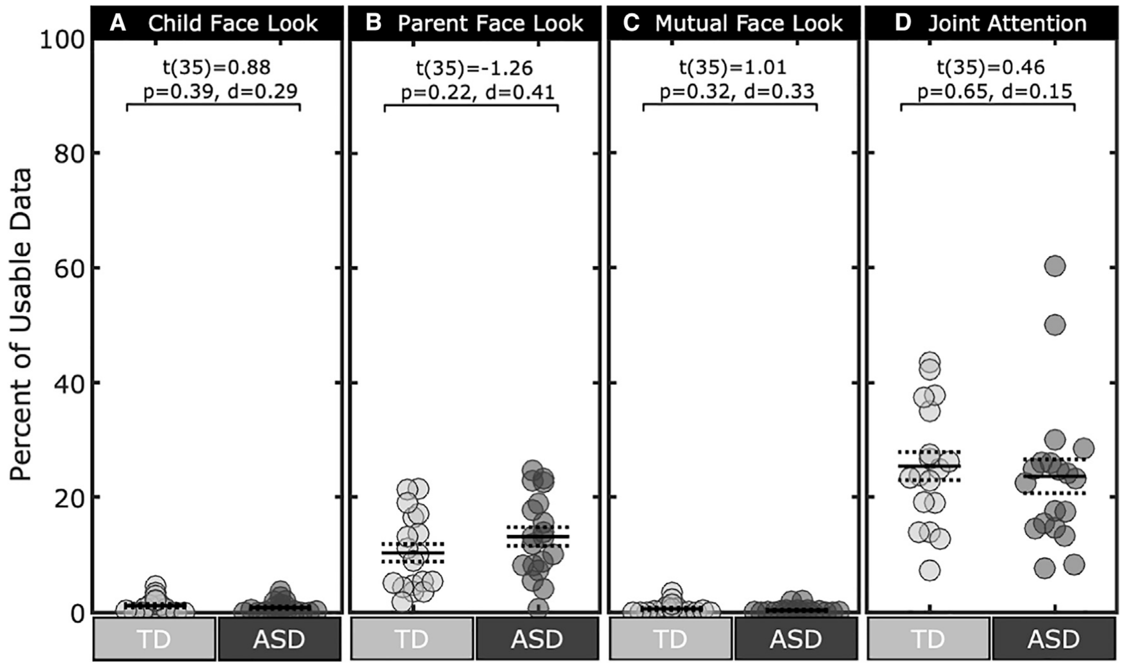


**Figure 2. Attention to faces, mutual looking at faces, and JA among TD children and children with ASD, and their parents**

Group means are represented by the solid line and standard errors are represented by the dotted lines. Each colored circle represents one individual. t tests are reported at the top of each figure.

(A and B) The percentage of the session that (A) each child spent looking at the parent's face and (B) each parent spent looking at the child's face.

(C) The percentage of the session that dyads spent engaging in mutual face looking.

(D) The percentage of the session that dyads spent in JA (defined as instances in which the child and parent looked at the same object at the same time for at least 500 ms).

See Figure S1 for additional analysis of the number of events per minute and median event duration for child face looks, parent face looks, mutual face looks, and JA.

in both groups similarly engaged in JA for a substantial percentage of their interactions (TD, 25.4% [±10.5%]; ASD, 23.6% [±13.0%]; $t(35) = 0.46$, $p = 0.65$, $d = 0.15$; Figure 2D). JA occurred within dyads to a greater extent than if gaze was randomly distributed in time (TD, permuted mean = 3.5% [SD ± 1.1%], $p < 0.001$; ASD, permuted mean = 4.6% [±1.2%], $p < 0.001$). Thus, both TD and ASD dyads engaged in JA at similarly high and above-chance rates.

We next assessed what proportion of JA bouts were child-led and parent-led, defined by the dyad member that first looked at the jointly attended object. Both groups exhibited similar proportions of child-led (TD = 48.1% [±12.0%]; ASD = 50.5% [±13.9%], $t(35) = 0.56$, $p = 0.58$, $d = 0.19$) and parent-led JA (TD = 51.9% [±12.0%]; ASD = 49.5% [±13.9%], $t(35) = 0.56$, $p = 0.58$, $d = 0.19$). Thus, the children and parents in both groups contributed roughly equally to the "success" of JA.

### *Parent naming behavior is different during JA moments*

To assess if JA is a differentiated context within toy play, we compared the rate at which parents named the toy of their child's attention during child looks *with* JA compared to looks *without* JA. An ANOVA on a linear mixed-effects (LME) model revealed a significant main effect of JA on parent naming ($F(1,70) = 66.15$, $p < 0.001$), but not group ($F(1,70) = 0.70$, $p = 0.41$) or interaction between JA and group ($F(1,70) = 0.95$, $p = 0.33$). Post hoc t tests revealed higher rates of naming during looks with JA compared to those without JA for both TD and ASD dyads (TD with JA, 15.2% [±11.5%]; TD without JA, 3.4% [±3.0%]; $t(17) = 5.22$, $p < 0.001$, $d = 1.29$; ASD with JA, 18.7% [±12.3%]; ASD without JA, 3.7% [±3.3%]; $t(18) = 5.99$, $p < 0.001$, $d = 1.28$; Figure 3). Thus, JA is a unique context that elicits high rates of parent naming for both groups.

## Behavioral pathways into JA

Given that both TD and ASD dyads achieved JA at similar rates, we next examined if the *pathways* into JA were similar between groups. We quantified the rates at which face looking, manual action, and toy naming occurred in the 1 s preceding JA onset (Figure 4; note that 1 s window was chosen to align with previously published values,[28] and findings were largely similar when 0.5–5 s windows were used).

### *Cues for parents to follow into JA (look at child face, child touch)*

An ANOVA on an LME revealed a significant main effect of cue ($F(1,70) = 90.622$, $p < 0.001$), but not group ($F(1,70) = 2.57$, $p = 0.11$) or the interaction between cue and group ($F(1,70) = 0.34$, $p = 0.56$). A post hoc t test revealed that the significant effect of cue was driven by more child touch than parent look at child face before JA for both the ASD (touch, 52.3% [±18.9%]; face, 25.3% [±13.1%]; $t(18) = -5.23$, $p < 0.001$, $d = 1.67$) and TD groups (touch, 48.8% [±10.8%]; face, 18.1% [±11.9%]; $t(17) = -9.17$, $p < 0.001$, $d = 2.70$). Further post hoc t tests examined the near-trending effect of group. Both TD and ASD children

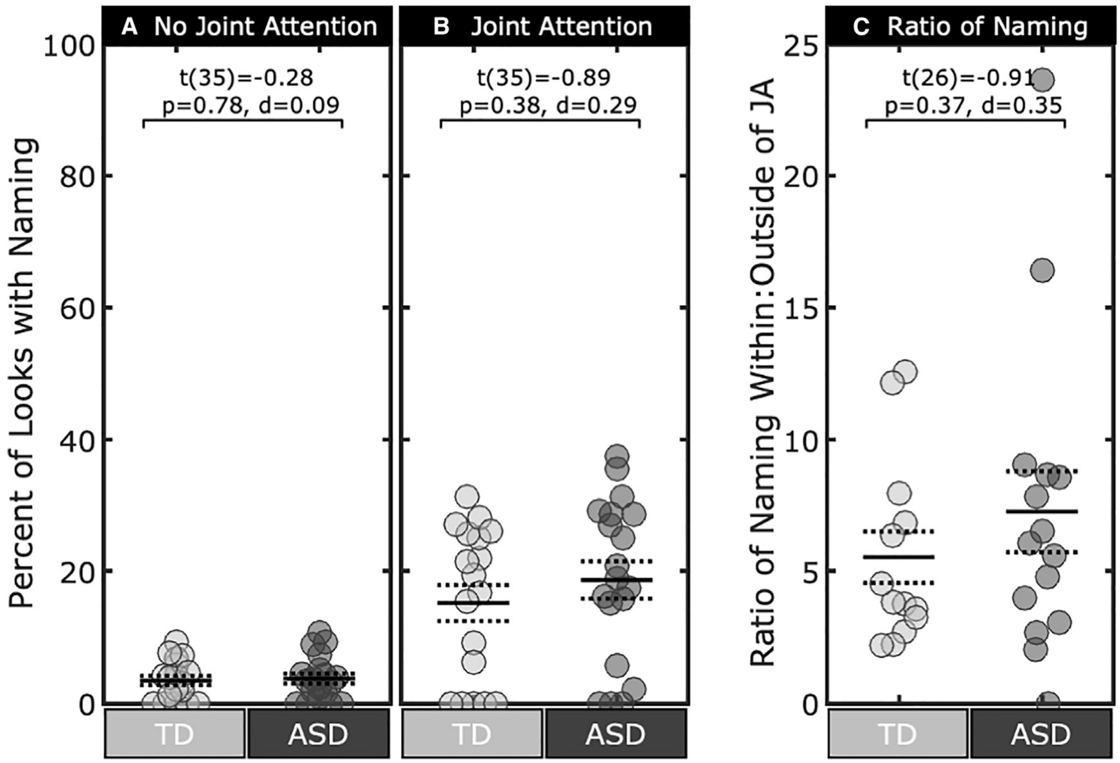


**Figure 3. Rate of parent naming during child looks with and without JA**

Group means are represented by the solid line and standard errors are represented by the dotted lines. Each colored circle represents one individual. t tests are reported at the top of each figure.

(A) The percentage of child looks that do not contain any JA with a parent that overlap with the parent naming the toy at which the child is looking.

(B) The percentage of child looks that do contain JA with a parent that overlap with the parent naming the toy at which the child is looking.

(C) For each participant, we calculated the ratio of naming during looks with JA to those without JA. Parents name objects during looks with JA ~5–7x more frequently than they name objects during looks without JA.

touched toys at similarly high rates preceding JA ($t(35) = -0.70$, $p = 0.49$, $d = 0.23$), consistent with prior work in typical development.[20,25–28] There was a trend for parents of children with ASD to more frequently look at their child's face before JA than parents of TD children ($t(35) = -1.73$, $p = 0.09$, $d = 0.55$).

### *Cues for children to follow into JA (look at parent face, parent touch, parent naming)*

An ANOVA on an LME revealed a main effect of cue ($F(1,105) = 118.15$, $p < 0.001$), but not group ($F(1,105) = 2.15$, $p = 0.15$) or interaction between cue and group ($F(1,105) = 1.07$, $p = 0.35$). Post hoc t tests revealed that children in both groups rarely looked at their parent's face before JA (TD = 1.2% [±2.2%]; ASD = 1.6% [±3.4%]; $t(35) = -0.46$, $p = 0.65$, $d = 0.15$). Parents in both groups frequently touched the object before JA (TD = 43.9% [±14.0%]; ASD = 36.6% [±20.7%]; $t(35) = 1.24$, $p = 0.22$, $d = 0.40$). Finally, parents in both groups named toys at similar frequencies preceding JA (TD = 12.2% [±10.6%]; ASD = 9.5% [±7.6%]; $t(35) = 0.89$, $p = 0.38$, $d = 0.29$). Child look at parent face was less frequent than parent naming of toys (ASD, $t(18) = -4.24$, $p < 0.001$, $d = 1.33$; TD, $t(17) = -4.29$, $p < 0.001$, $d = 1.44$) and parent touch (ASD, $t(18) = -6.96$, $p < 0.001$, $d = 2.36$; TD, $t(17) = -12.81$, $p < 0.001$, $d = 4.26$), suggesting that touch and naming pathways may be more available when establishing JA.

## DISCUSSION

This study provides a number of insights into how young children with ASD effectively coordinate their attention with their parents during natural social interaction. In doing so, we replicated several important findings in TD children and extended them to children with ASD: dyads achieved high levels of JA during toy play[21,25,27,28] despite children allocating very little attention to their parents' faces[21,25,27,28,32]; instead, dyads achieved JA more frequently via behavioral hand-following pathways;[20,25–28] finally, the rate of object naming was higher during JA moments,[4,13] thus setting up the conditions that can facilitate word learning. These results raise important questions regarding the nature of JA deficits in ASD.

JA, as often narrowly operationalized in the literature on ASD, refers to achieving coordinated visual attention on an object or event via a specific gaze- or point-following pathway.[2,5,10,15–19] Deficits in these specific pathways into JA are clear, robust, and replicable in children with ASD[7,18,33–35] and are used effectively in diagnosis.[36–38] Indeed, this work does not challenge these previous findings or their diagnostic utility. However, our findings of typical levels of JA in children with ASD in a free-play paradigm and irrespective of the behavioral pathways used to get into JA raise an important question: do the deficits

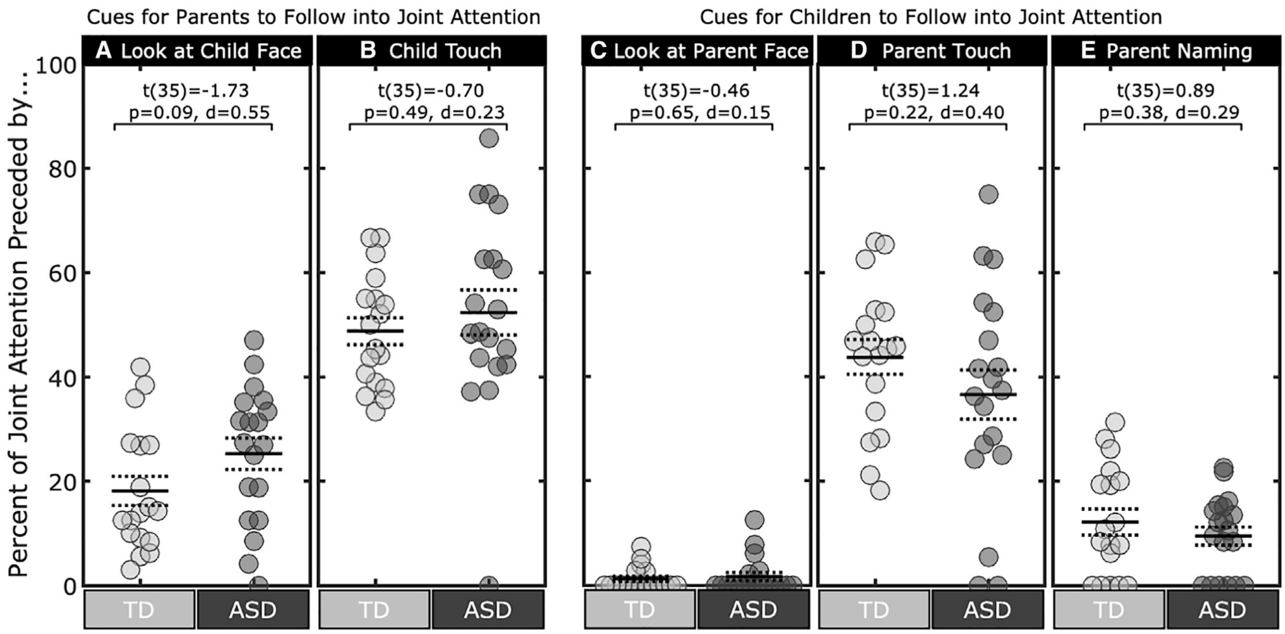


**Figure 4. Behaviors preceding JA**

Group means are represented by the solid line and standard errors are represented by the dotted lines. Each colored circle represents one individual. t tests are reported at the top of each figure. Behaviors were considered as preceding JA if they occurred in the 1 s window prior to the onset of JA.

(A and C) Children rarely looked at their parent's face (A), but parents frequently looked at their child's face before JA onset (C).

(B, D, and E) Child (B) and parent (D) touch behaviors also needed to be generated to the same toy as the following JA instance. Both children and parents frequently touched the toys before JA, (E) and parents often named toys before JA, suggesting alternative pathways into JA.

in gaze- and point-specific JA reported in the literature and used in clinical settings represent a causal behavioral *mechanism* underlying ASD or instead do they simply describe a robust and reliable *marker* of ASD? If mechanistic, the assumption is that the reduced amount of JA arising from the reduction in gaze- and point-following is critical for the development of ASD; if a marker, it may be a useful way to describe an atypical behavioral phenomenon but one that does not play a causal role. Of course, elucidating both markers and mechanism are important endeavors, but there is a critical difference between them: if the goal is to understand and intervene, then markers will ultimately prove less effective than mechanisms. A biomarker or behavioral marker that plays no causal role, regardless of whether it can effectively distinguish the groups (e.g., ash-leaf spots in tuberous sclerosis), would not be a good target for interventions. If the mechanistic value of JA is the coordination of attention on an object or event,[2,5,10,15–19] irrespective of the pathway into JA, then this work shows that such coordination can be achieved by children with ASD through non-gaze-based pathways (e.g., manual action), at least in certain contexts, as is the case in TD children.[4,21,25,26,28–31] This work therefore raises the provocative idea that gaze-following pathways into JA index atypical development in ASD, but perhaps not because this particular pathway serves as a fundamental causal behavioral mechanism underlying ASD—rather, it is possible that such difficulties in JA relate to other contextual or cognitive demands rather than a deficit in the ability to achieve shared attention with another person. This remains a testable hypothesis at this point.

Much of the extant knowledge on JA has been drawn from studies that test the use of explicit and pre-determined cues in more constrained tasks (e.g., screen-based eye tracking studies) or during semi-structured clinical assessments that are specifically designed to elicit overt, readily measurable behavioral markers, and not necessarily to elicit natural social behavior. While these studies have offered a wealth of insight into social attention in ASD, it remains understudied how the findings from such contexts translate to the less constrained, three-dimensional, real-world contexts where the cues used to achieve JA may be more subtle, dynamic, multimodal, and implicit.[21,29,39] For example, in constrained contexts, faces are often highly attended by children—but they are also prominently featured and may be one of few social cues available to guide attention. Indeed, face looking seems to be much less frequent in contexts where children are free to play and move naturally,[4,21,25,30,32,40] in part due to postural and other physical constraints that limit the child's ability to simultaneously look at their parent's face and engage with toys (e.g., heads tend to be positioned off the floor, while toys tend to be lower due to gravity and arm/hand position).[21,32] However, the high concordance between hand, body, head, and eye movements provides redundant signals of the social partner's attention,[25,27,41–47] which makes sole reliance on any one cue (e.g., faces) less necessary. Face looking is also less frequent in both ASD and TD children when playing with parents compared to an experimenter,[48] suggesting that familiarity may further reduce the reliance on explicit gaze-following cues. Thus, while individuals with ASD exhibit reduced gaze to faces in more constrained contexts,[49–53] our results suggest that this reduced looking may be less relevant to the child's social behavior during toy play where faces are neither easily attended to, frequently attended to, novel, nor necessary for JA. Instead, like TD children,[4,20,24,27–30] children with ASD in our study primarily relied on manual actions on toys (e.g., touching, holding) as a direct pathway into JA. These findings highlight the value of carrying out research in increasingly more naturalistic and ecologically valid contexts. However, both experimental and naturalistic approaches must be viewed as complementary to one another: free-flowing naturalistic paradigms can offer insight into the real-world behaviors of interest, and more constrained experimental paradigms can strategically manipulate the environment to help to efficiently and precisely disentangle possible mechanisms by which such behaviors manifest.

Finally, as has been described in TD dyads,[4] we found that particularly clear and more frequent naming moments occur during JA in ASD dyads. This is important because it suggests that the moments of coordinated attention that are known to be particularly conducive for word learning in TD dyads[23,54–58] are also achieved in ASD dyads. Further, joint attention, both broadly and narrowly operationalized, is linked to language learning abilities in TD children,[4–6,13] suggesting that the common component of both definitions—i.e., shared attention on an object or event at the same time—may be the fundamental component that supports language development, rather than the precise pathway used to achieve it. Research on joint engagement—a global measure of coordinated attention on a toy either with or without face looks during parent-child toy play[5]—has found that joint engagement *without* face looking is actually more predictive of later language abilities in children with ASD than joint engagement *with* face looking,[59,60] and the former is associated with hearing more parent object labels in infants at elevated risk for ASD.[61] When children are fully focused on a jointly attended object, their parents can provide labels that are clearly mapped to their target, thereby supporting word learning.[23,54–58] It remains an open question, then, why the children with ASD in the current study have profoundly reduced language abilities. At least three possibilities should be considered. First, children with ASD may not be able to process real-time information as efficiently as their TD peers, such that similar input does not translate into similar learning.[62,63] Second, the typical rates of object naming observed during toy play may not extend to other socially interactive contexts (e.g., mealtime, bath time). Finally, recent evidence suggests that JA is associated with distinct first-person visual perspectives: computer vision models reliably discriminate between JA and non-JA moments in TD children when trained on the child's first-person views.[64] Whether similar visual input is created during JA in children with ASD is currently unknown.

The present work raises many potential future directions. While the current study of 2- to 4-year-old children identified remarkable similarity between ASD and TD children, it is possible that differences manifest earlier and set the stage for atypical development. Therefore, studies of even younger children with ASD that examine whether and how JA (and the use of pathways into JA) changes over development, or with behavioral interventions, are warranted. A more thorough understanding of JA in children with ASD would also include varied activities, contexts, and social partners (e.g., siblings, same-aged children). Further, longer segments of data acquired in increasingly natural and varied contexts may reveal additional typical or atypical social, behavioral, or contextual influences on JA than the shorter, semi-natural interactions in the current study. Richer descriptions of behavior that emerge within the complexity of naturalistic dyadic interactions may continue to reveal unexpected patterns of both similarities and differences in children with ASD, help to illuminate causal mechanisms, and identify novel targets of intervention for children with ASD.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Inclusion and exclusion criteria
  - Experimental set-up
  - Paradigm
  - Assessments and questionnaires
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Gaze calibration
  - Eye gaze region of interest coding
  - Coding touch and naming during play
  - Usable data
  - Age
  - Attention to faces
  - Mutual face looking
  - Joint attention
  - Permutation testing procedures
  - Naming during looks with and without joint attention
  - Behaviors preceding joint attention
  - Group comparisons

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2022.04.044.

## ACKNOWLEDGMENTS

This work is supported in part by Indiana University Research Funding (D.P.K.), the Cincinnati Children's Hospital Research Foundation (C.A.E.), and NICHD R01 HD093792 and R01 HD074601 (C.Y.). J.Y.-H. and G.L. were supported by NIH institutional training grant T32 HD07475, and J.Y.-H. was additionally supported by NIH institutional training grant T32 MH103213 and an Autism Science Foundation Predoctoral Training Fellowship. We thank the following individuals for their contributions to the study: Margo Moore, Hilary Rosselot, and Sophia Melnyk for recruitment and scheduling of child participants with ASD and their parents; Anting Chen, Seth Foster, Daniel Pearcy, Arispa Weigold, Alyssa Yarbrough, and Christian Jerry for assistance with data collection and data management; Seth Foster for algorithm development; and Tracy Burns-Yocum, Lisa Byrge, and Samuel Harding for feedback on data analysis. Finally, we deeply appreciate the contributions of all our participating families.

## AUTHOR CONTRIBUTIONS

J.Y.-H., G.L., C.A.E., C.Y., and D.P.K. conceived of the experiment; J.Y.-H. and G.L. conducted the eye tracking experiment; R.C.S., K.C.D., E.V.P., and C.A.E. performed clinical characterization; J.Y.-H. and G.L. analyzed the results and wrote the manuscript with input from D.P.K. and C.Y.; and all authors reviewed the final version.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 20, 2021
Revised: March 2, 2022
Accepted: April 14, 2022
Published: May 12, 2022

## REFERENCES

1. Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., and Drew, A. (2000). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. Cogn. Dev. *15*, 481–498. https://doi.org/10.1016/s0885-2014(01)00037-5.
2. Corkum, V., and Moore, C. (1998). The origins of joint visual attention in infants. Dev. Psychol. *34*, 28–38. https://doi.org/10.1037/0012-1649.34.1.28.
3. Morales, M., Mundy, P., Crowson, M., Neal, A.R., and Delgado, C. (2005). Individual differences in infant attention skills, joint attention, and emotion regulation behaviour. Int. J. Behav. Dev. *29*, 259–263. https://doi.org/10.1080/01650250444000432.
4. Abney, D.H., Suanda, S.H., Smith, L.B., and Yu, C. (2020). What are the building blocks of parent–infant coordinated attention in free-flowing interaction? Infancy *25*, 871–887. https://doi.org/10.1111/infa.12365.
5. Bakeman, R., and Adamson, L.B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. Child. Dev. *55*, 1278. https://doi.org/10.2307/1129997.
6. Brooks, R., and Meltzoff, A.N. (2005). The development of gaze following and its relation to language. Dev. Sci. *8*, 535–543. https://doi.org/10.1111/j.1467-7687.2005.00445.x.
7. Charman, T. (2003). Why is joint attention a pivotal skill in autism? Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. *358*, 315–324. https://doi.org/10.1098/rstb.2002.1199.
8. Chevallier, C., Kohls, G., Troiani, V., Brodkin, E.S., and Schultz, R.T. (2012). The social motivation theory of autism. Trends Cogn. Sci. *16*, 231–239. https://doi.org/10.1016/j.tics.2012.02.007.
9. Morales, M., Mundy, P., Delgado, C.E.F., Yale, M., Messinger, D., Neal, R., and Schwartz, H.K. (2000). Responding to joint attention across the 6- through 24-month age period and early language acquisition. J. Appl. Dev. Psychol. *21*, 283–298. https://doi.org/10.1016/s0193-3973(99)00040-4.
10. Mundy, P., Sigman, M., and Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. J. Autism Dev. Disord. *20*, 115–128. https://doi.org/10.1007/bf02206861.
11. Sigman, M., and Ruskin, E. (1999). Continuity and change in the social competence of children with autism, Down syndrome, and developmental delays. Monogr. Soc. Res. Child Dev. *64*, 109–113. https://doi.org/10.1111/1540-5834.00009.
12. Siller, M., and Sigman, M. (2008). Modeling longitudinal change in the language abilities of children with autism: parent behaviors and child characteristics as predictors of change. Dev. Psychol. *44*, 1691–1704. https://doi.org/10.1037/a0013771.
13. Tomasello, M., and Farrar, M.J. (1986). Joint attention and early language. Child. Dev. *57*, 1454. https://doi.org/10.2307/1130423.
14. Toth, K., Munson, J.N., Meltzoff, A., and Dawson, G. (2006). Early predictors of communication development in young children with autism spectrum disorder: joint attention, imitation, and toy play. J. Autism Dev. Disord. *36*, 993–1005. https://doi.org/10.1007/s10803-006-0137-7.
15. Bates, E. (1979). In The Emergence of Symbols: Cognition and Communication in Infancy, E.A. Hammel, ed. (Academic Press).
16. Mundy, P.C., Sigman, M., and Kasari, C. (1994). Joint attention, developmental level, and symptom presentation in autism. Dev. Psychopathol. *6*, 389–401. https://doi.org/10.1017/s0954579400006003.
17. Mundy, P., Block, J., Delgado, C., Pomares, Y., van Hecke, A.V., and Parlade, M.V. (2007). Individual differences and the development of joint attention in infancy. Child. Dev. *78*, 938–954. https://doi.org/10.1111/j.1467-8624.2007.01042.x.
18. Mundy, P., and Newell, L. (2007). Attention, joint attention, and social cognition. Curr. Dir. Psychol. Sci. *16*, 269–274. https://doi.org/10.1111/j.1467-8721.2007.00518.x.
19. Mundy, P. (2018). A review of joint attention and social-cognitive brain systems in typical development and autism spectrum disorder. Eur. J. Neurosci. *47*, 497–514. https://doi.org/10.1111/ejn.13720.
20. Elmlinger, S.L., Suanda, S.H., Smith, L.B., and Yu, C. (2019). Toddlers' hands organize parent-toddler attention across different social contexts. In 2019 Joint IEEE 9th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob) (IEEE), pp. 296–301.
21. Franchak, J.M., Kretch, K.S., and Adolph, K.E. (2018). See and be seen: infant–caregiver social looking during locomotor free play. Dev. Sci. *21*, e12626. https://doi.org/10.1111/desc.12626.
22. Gabouer, A., and Bortfeld, H. (2021). Revisiting how we operationalize joint attention. Infant Behav. Dev. *63*, 101566. https://doi.org/10.1016/j.infbeh.2021.101566.
23. Suanda, S.H., Smith, L.B., and Yu, C. (2016). The multisensory nature of verbal discourse in parent–toddler interactions. Dev. Neuropsychol. *41*, 324–341. https://doi.org/10.1080/87565641.2016.1256403.
24. Suarez-Rivera, C., Smith, L.B., and Yu, C. (2019). Multimodal parent behaviors within joint attention support sustained attention in infants. Dev. Psychol. *55*, 96–109. https://doi.org/10.1037/dev0000628.
25. Yu, C., and Smith, L.B. (2013). Joint attention without gaze following: human infants and their parents coordinate visual attention to objects through eye-hand coordination. PLoS One *8*, e79659. https://doi.org/10.1371/journal.pone.0079659.
26. Yu, C., and Smith, L.B. (2015). Linking joint attention with hand-eye coordination: a sensorimotor approach to understanding child-parent social interaction, pp. 2763–2768.
27. Yu, C., and Smith, L.B. (2017). Hand-eye coordination predicts joint attention. Child. Dev. *88*, 2060–2078. https://doi.org/10.1111/cdev.12730.
28. Yu, C., and Smith, L.B. (2017). Multiple sensory-motor pathways lead to coordinated visual attention. Cogn. Sci. *41*, 5–31. https://doi.org/10.1111/cogs.12366.
29. Suarez Rivera, C., Schatz, J.L., Herzberg, O., and Tamis LeMonda, C.S. (2022). Joint engagement in the home environment is frequent, multimodal, timely, and structured. Infancy *27*, 232–254. https://doi.org/10.1111/infa.12446.
30. Deák, G.O., Krasno, A.M., Triesch, J., Lewis, J., and Sepeta, L. (2014). Watch the hands: infants can learn to follow gaze by seeing adults manipulate objects. Dev. Sci. *17*, 270–281. https://doi.org/10.1111/desc.12122.
31. Deák, G.O., Krasno, A.M., Jasso, H., and Triesch, J. (2018). What leads to shared attention? Maternal cues and infant responses during object play. Infancy *23*, 4–28. https://doi.org/10.1111/infa.12204.
32. Franchak, J.M., Kretch, K.S., Soska, K.C., and Adolph, K.E. (2011). Head-mounted eye tracking: a new method to describe infant looking. Child. Dev. *82*, 1738–1750. https://doi.org/10.1111/j.1467-8624.2011.01670.x.
33. Baron-Cohen, S., Cox, A., Baird, G., Swettenham, J., Nightingale, N., Morgan, K., Drew, A., and Charman, T. (1996). Psychological markers in the detection of autism in infancy in a large population. Br. J. Psychiatry *168*, 158–163. https://doi.org/10.1192/bjp.168.2.158.

34. Hobson, J.A., and Hobson, R.P. (2007). Identification: the missing link between joint attention and imitation? Dev. Psychopathol. *19*, 411–431. https://doi.org/10.1017/s0954579407070204.

35. Mundy, P., Sigman, M., Ungerer, J., and Sherman, T. (1986). Defining the social deficits of autism: the contribution of non-verbal communication measures. J. Child Psychol. Psychiatry *27*, 657–669. https://doi.org/10.1111/j.1469-7610.1986.tb00190.x.

36. Gotham, K., Risi, S., Pickles, A., and Lord, C. (2007). The Autism Diagnostic Observation Schedule: revised algorithms for improved diagnostic validity. J. Autism Dev. Disord. *37*, 613–627. https://doi.org/10.1007/s10803-006-0280-1.

37. Lord, C., Rutter, M., and Le Couteur, A. (1994). Autism Diagnostic Interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J. Autism Dev. Disord. *24*, 659–685. https://doi.org/10.1007/bf02172145.

38. Lord, C., Risi, S., Lambrecht, L., Cook Jr, E.H., Leventhal, B.L., Dilavore, P.C., Pickles, A., and Rutter, M. (2000). The Autism Diagnostic Observation Schedule–generic: a standard measure of social and communication deficits associated with the spectrum of autism. J. Autism Dev. Disord. *30*, 205–223. https://doi.org/10.1023/a:1005592401947.

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

40. Jones, R.M., Southerland, A., Hamo, A., Carberry, C., Bridges, C., Nay, S., Stubbs, E., Komarow, E., Washington, C., Rehg, J.M., et al. (2017). Increased eye contact during conversation compared to play in children with autism. J. Autism Dev. Disord. *47*, 607–614. https://doi.org/10.1007/s10803-016-2981-4.

41. Amso, D., and Johnson, S.P. (2006). Learning by selection: visual search and object perception in young infants. Dev. Psychol. *42*, 1236–1245. https://doi.org/10.1037/0012-1649.42.6.1236.

42. Canfield, R.L., and Kirkham, N.Z. (2001). Infant cortical development and the prospective control of saccadic eye movements. Infancy *2*, 197–211. https://doi.org/10.1207/s15327078in0202_5.

43. Corbetta, D., Thelen, E., and Johnson, K. (2000). Motor constraints on the development of perception-action matching in infant reaching. Infant Behav. Dev. *23*, 351–374. https://doi.org/10.1016/s0163-6383(01)00049-2.

44. Hayhoe, M., and Ballard, D. (2005). Eye movements in natural behavior. Trends Cogn. Sci. *9*, 188–194. https://doi.org/10.1016/j.tics.2005.02.009.

45. Pelz, J., Hayhoe, M., and Loeber, R. (2001). The coordination of eye, head, and hand movements in a natural task. Exp. Brain Res. *139*, 266–277. https://doi.org/10.1007/s002210100745.

46. Smith, L.B., Thelen, E., Titzer, R., and McLin, D. (1999). Knowing in the context of acting: the task dynamics of the A-not-B error. Psychol. Rev. *106*, 235–260. https://doi.org/10.1037/0033-295x.106.2.235.

47. Smith, L.B. (2010). Action as developmental process – a commentary on Iverson's Developing language in a developing body: The relationship between motor development and language development. J. Child Lang. *37*, 263–267. https://doi.org/10.1017/s0305000909990535.

48. Kasari, C., Sigman, M., and Yirmiya, N. (1993). Focused and social attention of autistic children in interactions with familiar and unfamiliar adults: a comparison of autistic, mentally retarded, and normal children. Dev. Psychopathol. *5*, 403–414. https://doi.org/10.1017/s0954579400004491.

49. Riby, D.M., and Hancock, P.J.B. (2008). Viewing it differently: social scene perception in Williams syndrome and autism. Neuropsychologia *46*, 2855–2860. https://doi.org/10.1016/j.neuropsychologia.2008.05.003.

50. Nakano, T., Tanaka, K., Endo, Y., Yamane, Y., Yamamoto, T., Nakano, Y., Ohta, H., Kato, N., and Kitazawa, S. (2010). Atypical gaze patterns in children and adults with autism spectrum disorders dissociated from developmental changes in gaze behaviour. Proc. R. Soc. B Biol. Sci. *277*, 2935–2943. https://doi.org/10.1098/rspb.2010.0587.

51. Hosozawa, M., Tanaka, K., Shimizu, T., Nakano, T., and Kitazawa, S. (2012). How children with specific language impairment view social situations: an eye tracking study. Pediatrics *129*, e1453–e1460. https://doi.org/10.1542/peds.2011-2278.

52. Dalton, K.M., Nacewicz, B.M., Johnstone, T., Schaefer, H.S., Gernsbacher, M.A., Goldsmith, H.H., Alexander, A.L., and Davidson, R.J. (2005). Gaze fixation and the neural circuitry of face processing in autism. Nat. Neurosci. *8*, 519–526. https://doi.org/10.1038/nn1421.

53. Riby, D.M., and Hancock, P.J.B. (2009). Do faces capture the attention of individuals with Williams syndrome or autism? Evidence from tracking eye movements. J. Autism Dev. Disord. *39*, 421–431. https://doi.org/10.1007/s10803-008-0641-z.

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

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

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

57. Yu, C., Smith, L.B., Shen, H., Pereira, A.F., and Smith, T. (2009). Active information selection: visual attention through the hands. IEEE Trans. Auton. Ment. Dev. *1*, 141–151. https://doi.org/10.1109/tamd.2009.2031513.

58. Bambach, S., Smith, L.B., Crandall, D.J., and Yu, C. (2017). Objects in the center: how the infant's body constrains infant scenes. In 2016 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics, ICDL-EpiRob (IEEE), pp. 132–137.

59. Adamson, L.B., Bakeman, R., Deckner, D.F., and Romski, M. (2009). Joint engagement and the emergence of language in children with autism and Down syndrome. J. Autism Dev. Disord. *39*, 84–96. https://doi.org/10.1007/s10803-008-0601-7.

60. Adamson, L.B., Bakeman, R., Suma, K., and Robins, D.L. (2019). An expanded view of joint attention: skill, engagement, and language in typical development and autism. Child. Dev. *90*, e1–e18. https://doi.org/10.1111/cdev.12973.

61. Roemer, E.J., Kushner, E.H., and Iverson, J.M. (2021). Joint engagement, parent labels, and language development: Examining everyday interactions in infant siblings of children with autism. J. Autism Dev. Disord. *52*, 1984–2003. https://doi.org/10.1007/s10803-021-05099-1.

62. Mitchell, S., Brian, J., Zwaigenbaum, L., Roberts, W., Szatmari, P., Smith, I., and Bryson, S. (2006). Early language and communication development of infants later diagnosed with autism spectrum disorder. J. Dev. Behav. Pediatr. *27*, S69–S78. https://doi.org/10.1097/00004703-200604002-00004.

63. Swanson, M.R., Shen, M.D., Wolff, J.J., Elison, J.T., Emerson, R., Styner, M.A., Hazlett, H.C., Truong, K., Watson, L.R., Paterson, S.J., et al. (2017). Subcortical brian and behavior phenoypes differentiate infants with autism versus language delay. Biol. Psychiatry Cogn. Neurosci. Neuroimaging 2, 664–672.

64. Peters, R., Amatuni, A., Schroer, S., Naha, S., Crandall, D.J., and Yu, C. (2021). Modeling joint attention from egocentric vision. In Proceedings of the 43rd Annual Meeting of the Cognitive Science Society.

65. Lord, C., Rutter, M., Dilavore, P.S., Risi, S., Gotham, K., and Bishop, S. (2012). Autism Diagnostic Observation Schedule, Second Edition (ADOS-2).

66. Mullen, E.M. (1995). Mullen Scales of Early Learning (American Guidance Service), pp. 55–64.

CellPress

# STAR★METHODS

## KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Deposited data | | |
| Head-mounted eye tracking data | This paper | https://osf.io/fw7tv/ |
| Exemplar participant videos | This paper | https://nyu.databrary.org/volume/1426 |
| Software and algorithms | | |
| Yarbus head-mounted eye tracking calibration software | Positive Science, LLC | https://www.positivescience.com/software/ |
| GV-Center V2 Pro Central Monitoring Station | GeoVision | https://www.geovision.com.tw/product/GV-Center%20V2%20Pro |
| Custom scripts to analyze eye tracking data | This paper (shared by J. Yurkovic-Harding) | https://github.com/JuliaYurkovicHarding/Joint-Attention-without-Face-Looking |
| MATLAB 2021a | The Mathworks | https://www.mathworks.com/products/matlab.html |
| Other | | |
| Positive Science head-mounted eye tracker | Positive Science, LLC | https://www.positivescience.com/hardware/ |

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Julia Yurkovic-Harding (jyurkovi@iu.edu).

### Materials availability

An exemplar video of the head-mounted eye tracking session is available in supplemental information and to authorized users on Databrary (https://nyu.databrary.org/volume/1426).

### Data and code availability

De-identified data have been deposited on Open Science Framework (https://osf.io/fw7tv/) and are publicly available as of the date of publication. All original code is available on GitHub (https://github.com/JuliaYurkovicHarding/Joint-Attention-without-Face-Looking). These resources are listed in the key resources table. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

Sixty children with ASD (45 males) and 22 age-matched typically developing (TD) children (14 males) between the ages of 24 and 48 months participated in the study with their parents. Nineteen child-parent dyads in the ASD group (13 males) and 18 dyads in the TD group (13 males) contributed eye-tracking data. (For characterization details, see Tables S1–S3.) Twenty-six child participants with ASD and two TD child participants (1 male) refused to wear the eye-tracking equipment or to allow the experimenter to make the necessary adjustments. An additional six children with ASD wore the eye tracker but the data were unusable (i.e., unable to generate a reliable calibration), four children with ASD experienced equipment malfunctions, four children with ASD were out of the age range of the study, and the data from one ASD dyad were lost. Finally, one TD child was excluded due to vision concerns, and one was excluded because the parent eye image was not usable. There were no substantial differences between the participants who contributed data and those that did not (see Table S4). The study aims were explained to all parents before participating and all parents provided informed consent. Parental consent was also obtained for the use of images collected during the study in scientific papers and presentations. Data collection with the ASD children and their parents took place at Cincinnati Children's Hospital Medical Center (Cincinnati, OH), and with the TD children and their parents at Indiana University – Bloomington (Bloomington, IN). Children with ASD were recruited from Cincinnati Children's Hospital where their diagnosis was confirmed by a trained clinician according to the DSM-V criteria. TD children were recruited from the Bloomington, Indiana community. The two locations involved the same personnel, testing equipment, and materials (described below). The experimental procedure was approved by the Institutional Review Boards at Cincinnati Children's Hospital Medical Center and Indiana University – Bloomington.

## METHOD DETAILS

### Inclusion and exclusion criteria

Our control sample was comprised of TD children who, according to parent report, did not have a diagnosis of ASD. We included children with ASD for whom their ASD diagnosis did not have a known genetic cause and who did not have complex medical histories. All ASD diagnoses were confirmed by clinical staff at Cincinnati Children's Hospital.

### Experimental set-up

Participating dyads played with 24 toys with which they were likely to be familiar (e.g., cars, animal figurines, blocks) in a space that resembled a playroom (Figure 1). Children and parents sat with each other on a plush, carpeted floor with the toys spread out randomly in the space between them. Each member of the dyad was equipped with a head-mounted eye tracker (Positive Science, LLC). The eye-tracking equipment included an infrared camera pointed at the participant's right eye to capture fixations, and a small video camera on the participant's forehead to capture the scene in front of him/her. The child eye-tracker was attached to a soft, adjustable cap placed on the child's head. The eye trackers sampled data at 30 Hz, and the eye-tracker's scene camera had a visual angle of 90 degrees. One experimenter placed the eye-trackers on the dyad members, while a second experimenter distracted the child with snacks and/or enticing toys that differed from the toys used in the experiment. Two external cameras facing the child and parent, respectively, and one camera mounted on the wall at participants' eye level were positioned in the play space in both testing locations. The external camera angles served as third-person perspectives of the scene for subsequent behavioral coding. The external video and audio streams were recorded simultaneously using GeoVision software (GeoVision), and the files were later synchronized with the eye tracker cameras using a custom Python script.

### Paradigm

Once the eye-tracking equipment was properly set up, the experimenters engaged participants in a calibration procedure to determine each participants' gaze location relative to the scene. The children and their parents were instructed to shift their gaze to different toys in the play area that were indicated with a laser pointer controlled by the experimenter. These fixations were used for later eye tracking calibration. The child-parent dyads then played under two conditions. During the *joint play* condition, children and their parents played together freely with the toys to yield approximately five minutes. The toys were spread randomly in between the dyad members. Parents did not receive any instructions other than to play with their children as they typically would at home (i.e., as naturally as possible). During the *solo play* condition (not analyzed here), the children played with the toys independently while the parents completed a questionnaire in a nearby location, again for approximately five minutes. Our analyses utilized the data collected during the *joint play* condition in order to assess dyadic interactions, as per the goals of the present study.

### Assessments and questionnaires

The child participants with ASD completed a developmentally appropriate Autism Diagnostic Observation Schedule (ADOS-2)[65] as a measure of symptom severity and the Mullen Scales of Early Learning (Mullen) as a measure of verbal and nonverbal abilities.[66] If the child had completed the ADOS or the Mullen within two years of participation in the eye tracking task (ADOS mean (SD): 7.1 months (5.0 months), range = -1.4-14.6 months; Mullen: 7.8 (4.6), range = 0.5-14.6 months), those scores were used for the current study. Given that many of the children performed at a level <0.1 percentile on the Mullen and we were therefore unable to calculate a meaningful *T* score, we used standardized developmental quotients (DQ), or a ratio score of the child's Mullen age equivalency divided by their actual age in months multiplied by 100. For example, a 24mo child who had an age equivalent score of 12m would receive a DQ of 50, whereas a 24mo child who had an age equivalent score of 24m would receive a DQ of 100.Parents also completed the MacArthur-Bates Communicative Development Inventories (MCDI) to determine the parent's perception of the child's vocabulary. This measure was acquired on the same day as the eye tracking data.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Gaze calibration

Following data collection, a trained experimenter calibrated parent and child gaze using Yarbus software (Positive Science, LLC) in order to generate fixation crosshairs to be used for manual coding of gaze behavior. The goal of the calibration procedure was to map the participant's eye movements to the locations on the scene camera video where the participant was looking. The experimenter annotated the first-person video with points indicating the x-y coordinates that corresponded to the eye's location in a given frame of the video. These points were spread over time and over space (i.e., x- and y-coordinates of the video). The procedure of asking parents and children to systematically look at toys before play began (guided by a laser pointer) is instrumental in having moments when it is clearly known to what the participant was looking. Additionally, moments in which a child was examining an object in their hands or was tracking the movement of an object through space were typically used as calibration points. A calibration was considered complete when the indicated calibration points had an intra-calibration correlation of 0.95 or greater in both the x- and y-dimension, and when the experimenter manually confirmed that the crosshair was accurately moving with the eye image. Separate calibrations were conducted for segments of the video when the relationship of the eye movements to the scene may differ, such as having two separate calibrations before and after a bump of the eye tracker or before and after the child moved from sitting to standing. Once points were annotated, a purple crosshair was generated overtop the video (Figures 1B and 1C) that moved with the participant's eye movements to indicate where they were looking at each frame of the video, not just the ones annotated.

### Eye gaze region of interest coding

We first used a custom algorithm to identify short segments of gaze that were most likely directed to the same object (based on minimal eye movements and the participant not blinking). Then, using a custom coding program, highly trained human coders labeled these segments based on the crosshair's location in the dyad member's first-person video of the scene. The coders selected one of 26 values that corresponded to the 26 possible regions of interest (ROIs): one of the 24 toys, the social partner's face, or non-ROI targets that included all other aspects of the play space (e.g., a picture on the wall). Coders could also label a look as missing data if the crosshair landed outside of the scene and the object of fixation could not be readily discerned. Non-look moments, such as saccades and blinks, were not labeled with an ROI. A subset of usable data (10%) for each dyad member was chosen at random and coded by a second coder. Overall raw agreement (mean (SD) = 91.7% (±5.7%)) and Cohen's kappa scores (mean (SD) = 0.88 (±.08) were high, with similarly high values when considering parents and children in each group separately (all agreement > 90%; all Cohen's kappa > .85), suggesting reliable ROI coding.

### Coding touch and naming during play

Additional coding was completed for objects that each dyad member's hands came into contact with. Trained coders annotated each frame of the interaction for each hand individually, indicating if each hand was in contact with one of the 24 toys or with nothing. Additionally, parent speech was coded using Audacity. Trained coders found moments of parent speech during the interaction and transcribed parent utterances. An utterance was any segment of speech separated from other speech by 400ms of silence. Utterances that contained a label to one of the 24 toys were considered naming utterances.

### Usable data

Only the moments where the parent and child were alone in the room, with usable eye images, and with the toys already spread in front of them (i.e., excluding moments when the experimenter was completing participant set-up or adjusting a bumped eye camera) were used for data analysis. Note that these segments of data may still have been used to complete an accurate calibration (such as the moments where experimenters shined a laser pointer on different toys to direct participants' gaze). We repeated all analyses while also controlling for usable data length (range = 0.85-6.73 mins) and the findings were unchanged. The reported analyses therefore use the total frames of usable data (only dyad in room, usable eye image, toys available) as the denominator.

### Age

There was no difference in age between children in each group ($t(35) = -0.30$, $p = 0.77$, $d = -0.10$). To ensure that the reported analyses were not due to differences in child age, we also repeated all analyses while regressing age and findings remained unchanged.

### Attention to faces

We first determined the degree to which children and their parents looked at the other's face during the free-flowing interaction. To do so, we calculated the percentage of time spent looking at the social partner's face during the coded interaction for both child and parent in each dyad by dividing the number of frames coded as "face" for a given dyad member by the total frames of usable data for the dyad.

### Mutual face looking

Next, we assessed the extent to which children and their parents mutually looked at each other's face over the course of the interaction. We identified the moments during which the child and parent looked at the other's face at the same time, and we divided the number of instances in which the same frame had been coded as "face" for child and parent by the total frames of usable data for the dyad. Note that gaze to faces was coded separately for parents and children, such that mutual face looking was an unbiased and objective conjunction of these two independent data streams.

### Joint attention

In accordance with the procedure used in previous publications using similar datasets[25,27] we operationalized joint attention (JA) as the child and parent looking at the same object at the same time for at least 500 milliseconds. We allowed for looks elsewhere during the period of JA that were less than 300 milliseconds, which corresponds to the average fixation length of toddlers in similar datasets previously collected by one of the authors (C.Y.). To determine the percentage of time spent in JA during the interaction, we divided the number of frames identified by a custom script as being part of a JA bout by the total number of frames that had been coded for both dyad members. Similar to mutual face gaze, gaze to objects was coded separately for parents and children; thus, moments of JA were defined objectively from the conjunction of these data streams.

### Permutation testing procedures

We employed a permutation testing procedure to determine if rates of mutual face looking and JA were above those that could be expected by chance. To do so, we circularly shifted the parent's eye gaze data stream by a random number of frames relative to their child's gaze over 1000 permutations. We then calculated the percent of frames with mutual face looking or JA for each individual, and we recorded the means for each group. We compared the actual rate of mutual face looking and JA to the permuted means to determine whether intra-dyad coordination exceeded chance.

### Naming during looks with and without joint attention

We divided the child's looks at toys into two categories: those without JA and those with JA. By this definition, the entire child look was considered as a look with JA, not just the segment of the look that overlapped with a parent look at the same toy. We then determined what proportion of looks with and looks without JA overlapped with a parent naming utterance to the toy of the child's interest. A naming utterance was considered as overlapping if the utterance either began during the child's look, ended during the child's look, or spanned the entire duration of the child's look.

### Behaviors preceding joint attention

Finally, we determined the rate at which behavioral cues preceded joint attention. Specifically, we calculated the rate of cues for children to follow into joint attention – look at parent's face, parent touch, and parent naming – and cues for parents to follow into joint attention – look at child's face and child touch. We considered behaviors to precede joint attention if the behavior began before joint attention and if the behavior either extended beyond joint attention, ended during joint attention, or ended within a second before joint attention or any time after. The 1 second window was chosen because past research suggests that joint attention occurs within one second.[28] We also repeated the analyses examining the rate of behaviors preceding joint attention at every 500ms interval from 500ms to 5 seconds and findings were effectively the same.

### Group comparisons

We compared the percentage of time engaged in the various behaviors of interest (attention to faces, mutual face gaze, and JA) across the two groups using independent sample t tests. For certain variables (naming during JA, behaviors preceding JA), we used linear mixed effects models with fixed effects for the variable, fixed effects for group (ASD vs. TD), and random effects for participant. We set alpha at .05 to determine significance. We calculated the effect size (Cohen's *d*) for each result by subtracting the average values for the groups from one another and dividing by the square root of half of the pooled standard deviation.