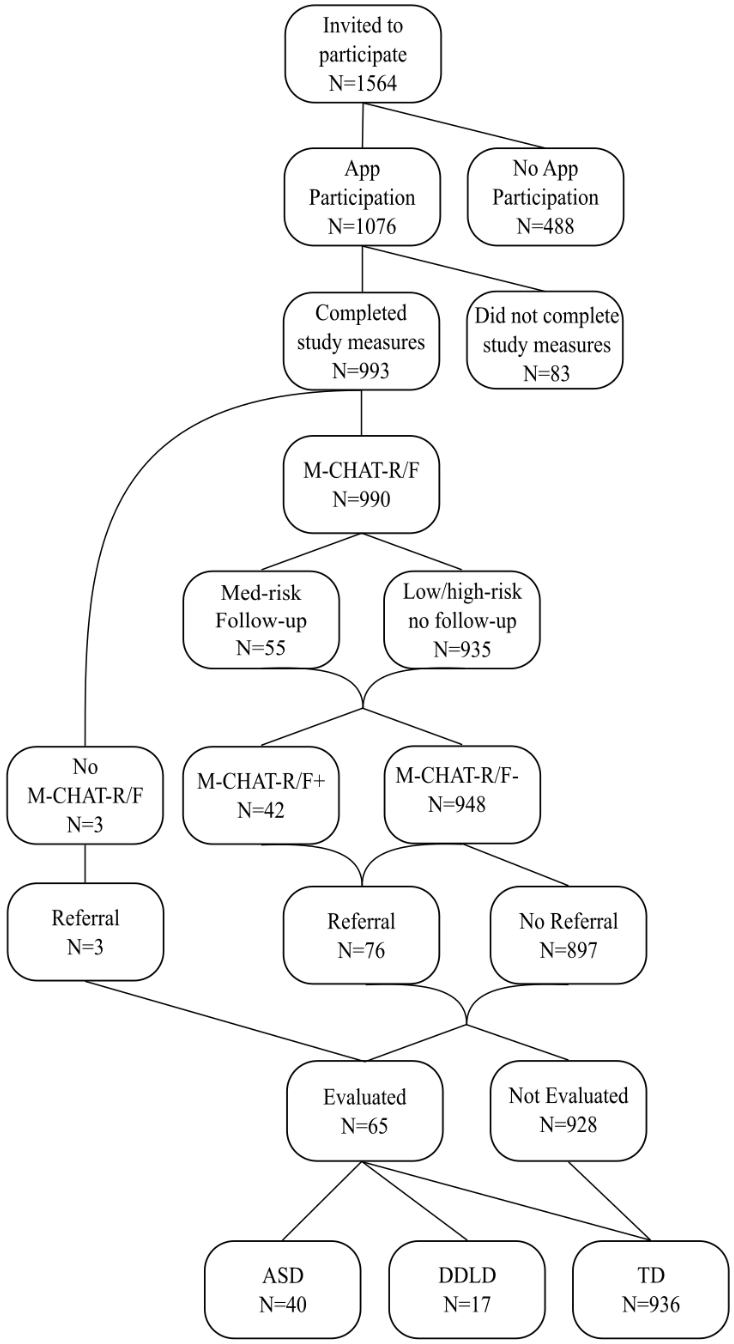
Supplemental Online Content

Chang Z, Di Martino JM, Aiello R, et al. Computational methods to measure patterns of gaze in toddlers with autism spectrum disorder. *JAMA Pediatr*. Published online April 26, 2021. doi:10.1001/jamapediatrics.2021.0530

eFigure 1. Study consort diagram



Demographic and clinical characteristics of subgroups

Demographic characteristics, including age, sex, race, ethnicity, and caregiver education, and clinical characteristics, including M-CHAT-R/F scores, ADOS calibrated severity scores and Mullen Scales of Early Learning scores for children with typical development (TD), autism spectrum disorder (ASD), and developmental delay/language delay (DDLD) are shown in eTable 1.

eTable 1: Demographic and clinical characteristics

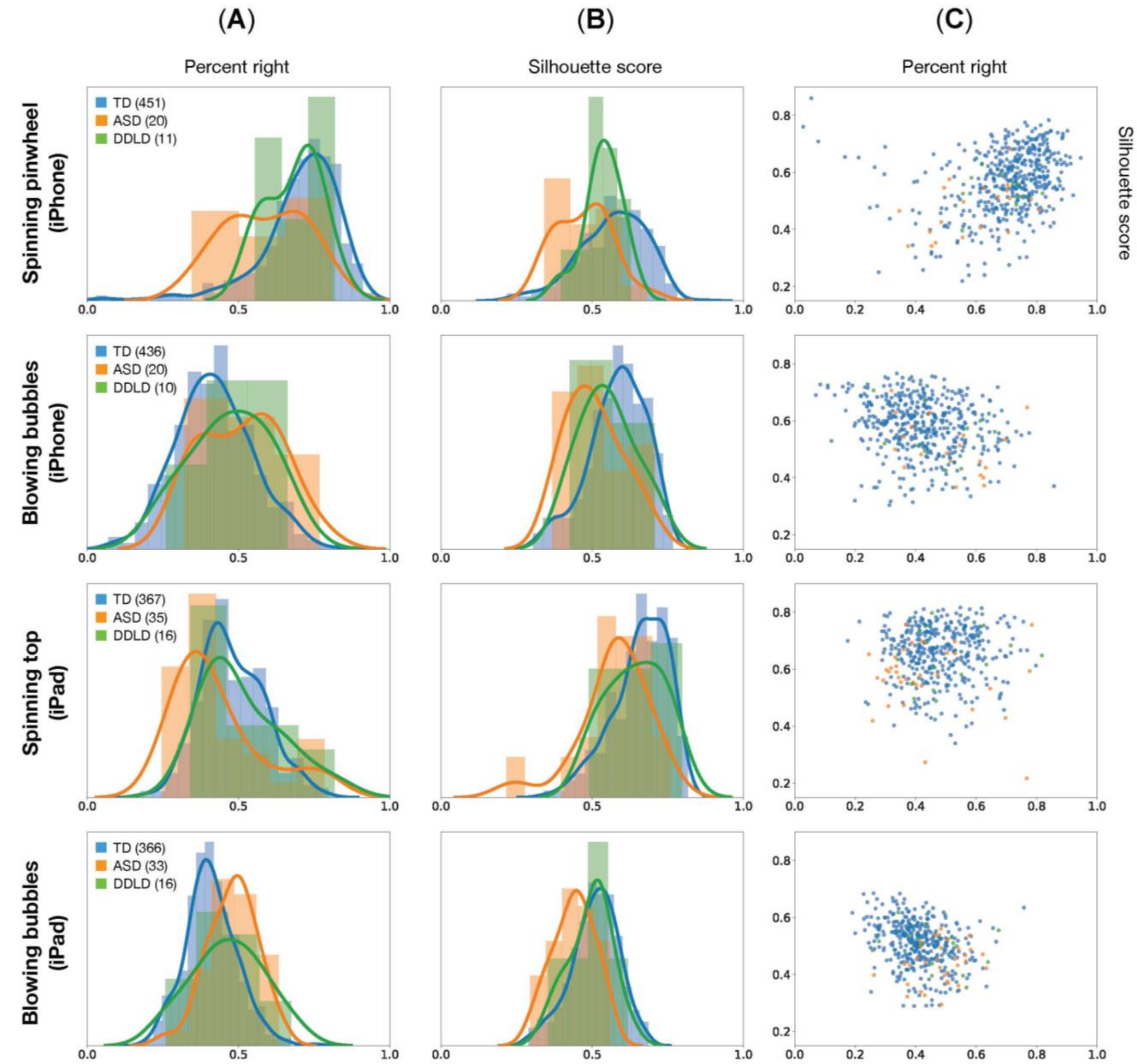
	TD N=936	ASD N=40	DDLD N=17
Age in months, Mean (SD)	20.9 (3.3) [a]	24.2 (4.6) [a]	22.3 (3.5)
Sex			
Female	479 (51.2%) [a,b]	9 (77.5%) [a]	3 (17.6%) [b]
Male	457 (48.8%) [a,b]	31 (22.5%) [a]	14 (82.4%) [b]
Race			
American Indian/Alaskan Native	27 (2.9%)	3 (7.5%)	0 (0.0%)
Asian	36 (3.9%)	1 (2.5%)	0 (0.0%)
Black or African American	152 (16.2%)	6 (15.0%)	6 (35.3%)
Native Hawaiian or Other Pacific Islander	0 (0.0%)	0 (0.0%)	0 (0.0%)
White/Caucasian	568 (60.7%)	19 (47.5%)	7 (41.2%)
More Than One Race	95 (10.1%)	7 (17.5%)	1 (5.9%)
Other	52 (5.6%)	4 (10.0%)	2 (11.8 %)
Unknown/Not Reported	6 (0.6%)	0 (0.0%)	1 (5.9 %)
Ethnicity			
Hispanic/Latino	151 (16.1%) [a]	12 (30.0%) [a]	5 (29.4%)
Not Hispanic/Latino	780 (83.3%) [a]	28 (70.0%) [a]	12 (70.6%)
Unknown/Not Reported	5 (0.5%) [a]	0 (0.0%) [a]	0 (0.0%)
Highest Level of Education			
Without High School Diploma	37 (4.0%) [a,b]	4 (10.0%) [a]	3 (17.6%) [b]
High School Diploma or Equivalent	63 (6.7%) [a,b]	5 (12.5%) [a]	6 (35.3%) [b]
Some College Education	102 (10.9%) [a,b]	9 (22.5%) [a]	1 (5.9%) [b]
4-Year College Degree or More	690 (73.7%) [a,b]	22 (55.0%) [a]	6 (35.3%) [b]
Unknown/Not Reported	44 (4.7%) [a,b]	0 (0.0%) [a]	1 (5.9%) [b]
M-CHAT-R/F			
Positive	1 (0.1%) [a,b]	31 (77.5%) [a,c]	10 (58.8%) [b,c]
Negative	935 (99.9%) [a,b]	7 (17.5%) [a,c]	6 (35.3%) [b,c]
Missing	0 (0.0%) [a,b]	2 (5.0%) [a,c]	1 (5.9%) [b,c]
Clinical variables	Mean (SD)		
	ASD		**DDLD**
ADOS-2 Toddler Module			
Calibrated Severity Score	7.6 (1.7) [c]		3.9 (1.6) [c]
Mullen Scales of Early Learning			
Early Learning Composite Score	63.2 (9.9) [c]		74.1 (15.7) [c]
Expressive Language T-Score	28.0 (7.3) [c]		36.1 (11.2) [c]
Receptive Language T-Score	22.9 (4.8) [c]		32.1 (14.1) [c]
Fine Motor T-Score	34.0 (10.4) [c]		38.4 (5.9) [c]
Visual Reception T-Score	33.2 (10.7)		37.4 (12.3)

TD: Typical development; ASD: Autism spectrum disorder; DDLD: Developmental delay/language delay
M-CHAT-R/F: Modified Checklist for Autism in Toddlers, Revised with Follow-Up Questions
ADOS-2: Autism Diagnostic Observation Schedule – Second Edition
[a] Significant difference between typical and ASD groups; [b] Significant difference between typical and LD-DD groups; [c] Significant difference between ASD and LD-DD groups (*P*'s < .05)

Descriptive data for children with developmental delay/language delay without ASD

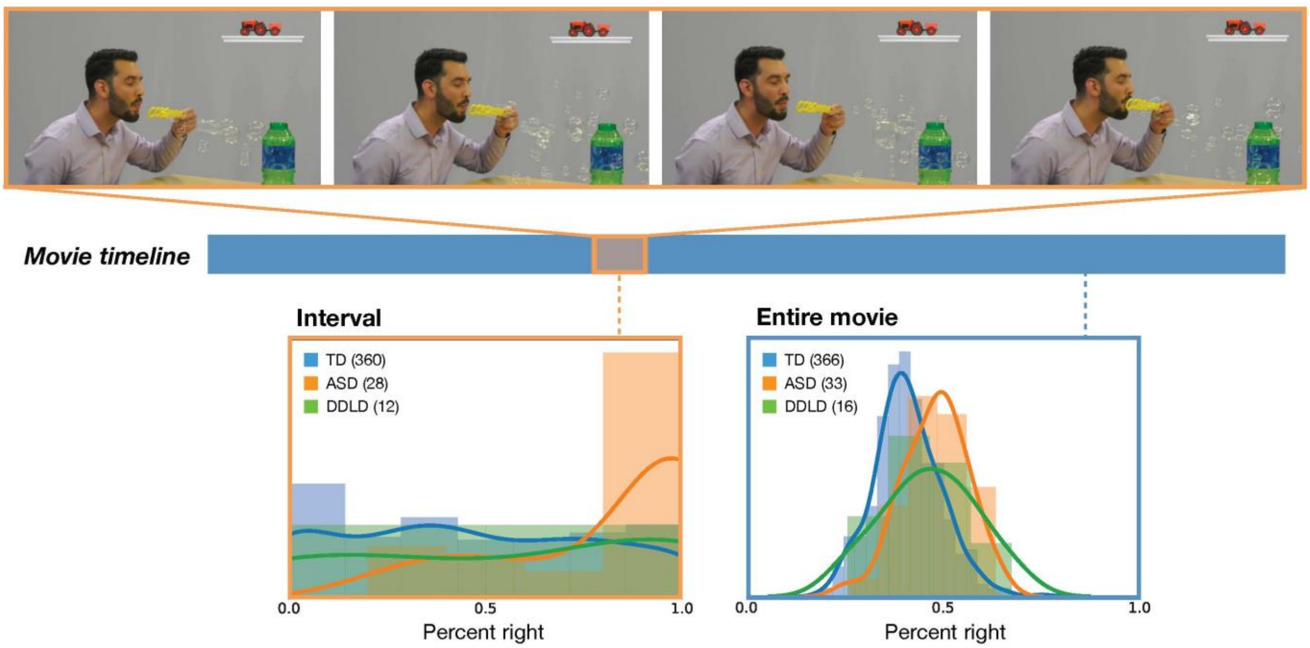
We present and discuss the results for a subgroup of children (N=17) who received a diagnosis of developmental delay and/or language delay without ASD (DDLD). DDLD was defined as having failed the M-CHAT-R/F or having provider or caregiver developmental concerns, were referred for evaluation and having been administered the ADOS-2 and Mullen Scales and determined by a licensed psychologist that the child did not meet DSM-5 criteria for ASD. All children in the DDLD group scored $\geq$ 9 points below the mean on at least one Mullen Early Learning Subscale (1 SD = 10 points).

Social preference movies. eFigure 2 shows density plots and histograms for the gaze data for four social preference movies, and corresponding percent right scores (percentage of time looking at the right side of the screen) distributions, silhouette scores distributions, and scatter plots displaying individual percent right scores and silhouette scores. Toddlers with typical development (TD) are shown in blue, those with ASD are shown in orange, and those with DDLD are shown in green. The percent right scores illustrate that, for all four movies (each < 60 seconds in length), the gaze of children with ASD was shifted toward the side of the screen on which the toys were displayed, whereas the opposite was true for the children with typical development. For the percent right scores, for 3 of the 4 movies, the distributions of the DDLD group paralleled the TD group.



eFigure 2. Social versus non-social gaze preference in toddlers with ASD, TD, and DDLD. Gaze data for four movies that depicted a person on one side of the screen playing with toys located on the opposite side of the screen. **(A)** Distribution of percent right scores (percent of time looking at the right side of the screen) for each movie. (**B**) Distribution of silhouette scores for each movie, and (**C**) Scatter plots displaying individual participant percent right (horizontal axis) and silhouette scores (vertical axis). For 'Spinning pinwheel' (iPhone), person is on the right side of screen; for 'Blowing bubbles' (iPhone), person is on the left side; for 'Spinning top' (iPad), person is on the right side; and for 'Blowing bubbles' (iPad), person is on the left side. Toddlers with typical development (TD) are shown in blue, those with ASD are shown in orange, and those with DDLD are shown in green.

Attention during salient moments in a social preference movie. Children's gaze behavior was measured during a salient social segment of the movie during which the person paused expectantly and then enthusiastically blew the bubbles, paralleling a probe used in the ADOS evaluation. eFigure 3 displays the density plots and histograms for the percent right scores reflecting attentional preference for toys (right side) versus the person (left side) for toddlers with TD (blue), ASD (orange), and DDLD (green) during the salient segment of the movie and for the entire movie. Children with ASD more often focused their attention on the toys compared to the person during this segment, whereas children with TD and DDLD showed a distributed pattern of attention to both the toys and the person. When gaze was examined across the entire movie, the children with ASD and DDLD looked more similar, with higher percent right scores (more attention to the toys) compared to the TD group. Thus, whether the children with DDLD appeared more similar to the ASD versus TD group depended on both the movie shown and whether gaze was examined during salient moments or across the entire movie.



eFigure 3. Gaze patterns of toddlers with ASD, TD, and DDLD during salient moments.

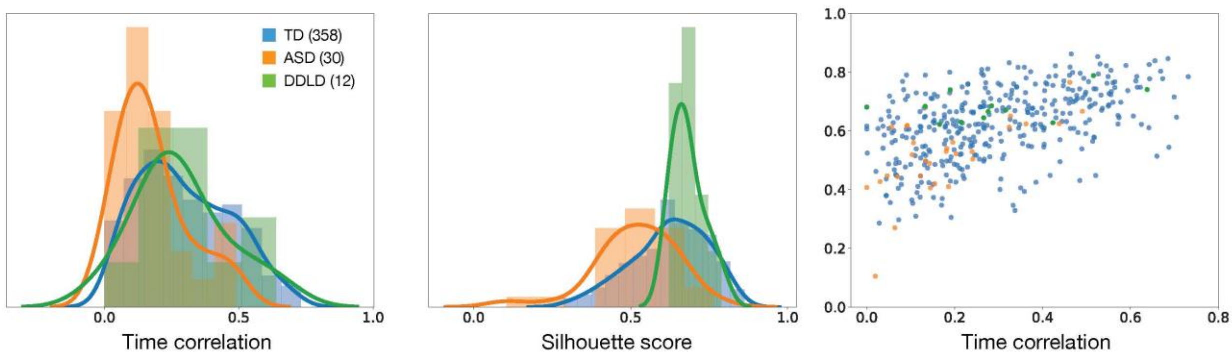
Children's gaze behavior was measured during a segment of the movie during which the person paused expectantly and then enthusiastically blew the bubbles. The graphs display the percent right scores reflecting attentional preference for the toys on the right versus person on the left for toddlers with typical development (blue), ASD (orange), and DDLD (green) during this interval of the movie and for the entire movie.

Gaze coordination with speech. One movie was designed to elicit children's gaze patterns when they observed speech of two adults in conversation. The movie was based on previous research showing that 6-11-month-old infants coordinate their gaze with the flow of a conversation they are watching.[1] eFigure 4 shows an illustration of the movie in which the two women were displayed on opposite side of the screen. A distractor object (clock) was located on one side of the screen. The Figure shows the distributions of gaze-speech time correlation and silhouette scores and a scatter plot showing individual data for these two measures. Children with ASD (orange) showed lower correlations between their gaze and the conversational flow, compared to children with TD (blue) and children with DDLD (green), which were similar to each other.

Preliminary conclusions regarding the DDLD findings. These preliminary findings suggest that the gaze patterns to social stimuli of children with DDLD were generally more similar to children with TD than to those with ASD during most but not all of the social preference movies, and during salient moments in a social preference movie, and during the movie designed to assess coordination of gaze with speech. Due to the small sample size of the DDLD group, however, no firm conclusions can be drawn. We are currently collecting larger samples that will allow us to more definitively assess whether the digital assessment tool can reliably distinguish between children with ASD versus DDLD.

(A)

(B)



eFigure 4. Correlation between gaze and speech in children with ASD, TD, and DDLD.

(A) Movie showing two women engaged in a back-and-forth conversation with a distractor object (clock) located in upper right corner. **(B)** Gaze data from the movie displaying the correlations between gaze patterns and the flow of the conversation. Shown are distributions of the time correlation and silhouette scores and a scatter plot jointly showing individual data for these two measures for children with typical development (blue), ASD (orange), and DDLD (green).

Re-analyses of social attention data using independent samples

A subset of children (1 TD, 23 ASD, 11 DDLD) who failed the MCHAT-R/F received a second administration of either the iPad or iPhone app (the version they did not receive in the clinic), which allowed us to include these participants in both the iPad and iPhone analyses. To ensure that this did not influence results, we repeated analyses with completely independent data sets (i.e., a child only contributed data from the first app administration, not both).

iPhone results. Analyses of both the percent right and silhouette scores continued to show significant differences between the TD and ASD groups: ‘Spinning pinwheel,’ $P<.001$ and $r=.51$ for percent right, $P<.001$ and $r=.52$ for the silhouette score; ‘Blowing bubbles,’ $P=.007$ and $r=.35$ for percent right, $P<.001$ and $r=.44$ for the silhouette score.

iPad results. Analyses of both the percent right and silhouette scores continued to show significant differences between the TD and ASD groups: ‘Spinning top,’ $P=.02$ and $r=.33$ for percent right, $P=.008$ and $r=.38$ for the silhouette score; ‘Blowing bubbles,’ $P<.001$ and $r=.53$ for percent right, $P<.001$ and $r=.61$ for the silhouette score.

Performance of model based on combined gaze features by sex, race, and ethnicity

eTable 2 displays the AUCs obtained for models based on the combination of multiple gaze features separately by sex, race, and ethnicity. The AUC values were relatively consistent across groups; however, confidence intervals were larger due to the smaller sample sizes**.**

eTable 2: Model performance for combined features by sex, race, and ethnicity

	iPhone	iPad
All	.88 [.78, .98]	.90 [.82, .97]
Sex		
Female	.88 [.78, .98]	.90 [.73, 1.0]
Male	.87 [.75, .99]	.88 [.79, .98]
Race		
Black or African American	.89 [.59, 1.0]	.71 [.40, 1.0]
White/Caucasian	.88 [.72, 1.0]	.91 [.82, 1.0]
All Other Races	.88 [.75, 1.0]	.92 [.77, 1.0]
Ethnicity		
Not Hispanic/Latino	.87 [.74, .99]	.90 [.82, .98]
Hispanic/Latino	.91 [.75, 1.0]	.84 [.61, 1.0]

eReference

1. Augusti EM, Melinder A, Gredeback G. Look Who's Talking: Pre-Verbal Infants' Perception of Face-to-Face and Back-to-Back Social Interactions. *Frontiers in psychology*. 2010;1:161.