UC San Diego
UC San Diego Previously Published Works

Title

Large scale validation of an early-age eye-tracking biomarker of an autism spectrum disorder subtype.

Permalink

https://escholarship.org/uc/item/5bf6m102

Journal

Scientific Reports, 12(1)

Authors

Wen, Teresa
Cheng, Amanda
Andreason, Charlene
et al.

Publication Date

2022-03-11

DOI

10.1038/s41598-022-08102-6

Copyright Information

Peer reviewed

scientific reports

OPEN

Large scale validation of an early-age eye-tracking biomarker of an autism spectrum disorder subtype

Teresa H. Wen[1✉], Amanda Cheng[1], Charlene Andreason[1], Javad Zahiri[1], Yaqiong Xiao[1], Ronghui Xu[3], Bokan Bao[1,2], Eric Courchesne[1], Cynthia Carter Barnes[1], Steven J. Arias[1] & Karen Pierce[1✉]

Few clinically validated biomarkers of ASD exist which can rapidly, accurately, and objectively identify autism during the first years of life and be used to support optimized treatment outcomes and advances in precision medicine. As such, the goal of the present study was to leverage both simple and computationally-advanced approaches to validate an eye-tracking measure of social attention preference, the GeoPref Test, among 1,863 ASD, delayed, or typical toddlers (12–48 months) referred from the community or general population via a primary care universal screening program. Toddlers participated in diagnostic and psychometric evaluations and the GeoPref Test: a 1-min movie containing side-by-side dynamic social and geometric images. Following testing, diagnosis was denoted as ASD, ASD features, LD, GDD, Other, typical sibling of ASD proband, or typical. Relative to other diagnostic groups, ASD toddlers exhibited the highest levels of visual attention towards geometric images and those with especially high fixation levels exhibited poor clinical profiles. Using the 69% fixation threshold, the GeoPref Test had 98% specificity, 17% sensitivity, 81% PPV, and 65% NPV. Sensitivity increased to 33% when saccades were included, with comparable validity across sex, ethnicity, or race. The GeoPref Test was also highly reliable up to 24 months following the initial test. Finally, fixation levels among twins concordant for ASD were significantly correlated, indicating that GeoPref Test performance may be genetically driven. As the GeoPref Test yields few false positives (~ 2%) and is equally valid across demographic categories, the current findings highlight the ability of the GeoPref Test to rapidly and accurately detect autism before the 2nd birthday in a subset of children and serve as a biomarker for a unique ASD subtype in clinical trials.

Autism spectrum disorder (ASD) begins during prenatal life[1,2], yet most children do not receive a diagnosis and start treatment until 3–4 years later[3,4]. Although genetic, neural, metabolomic, and molecular systems are adversely impacted in ASD[2,5,6], it is nevertheless detected and diagnosed using clinical judgement.

There has been a recent surge in research designed to discover biologically-based markers of ASD which can increase the pace of diagnosis, remove the requirement for highly-trained professionals, provide prognostic information, guide treatment plans, or be used as outcome measures in clinical trials[7]. Currently, only two ASD biomarkers are being considered for the FDA Biomarker Qualification Program[8,9]. However, they were established at "late" ages in children and therefore may not be generalizable to toddlers and infants, for whom biomarkers are of greatest utility. Moreover, these biomarkers only identify a subset of ASD children, indicating that additional biomarkers for other ASD subtypes are needed.

Dramatically reduced attention to social information is a key feature of ASD noted since its discovery in 1943[10]. Unsurprisingly, considerable effort has been leveraged to understand and quantify social visual attention abnormalities, most recently using eye-tracking[11–29]. Despite varying stimuli and participant age, a meta-analysis

[1]Autism Center of Excellence, Department of Neurosciences, University of California, San Diego, 8110 La Jolla Shores Dr., La Jolla, CA 92037, USA. [2]Department of Bioinformatics and Systems Biology, University of California, San Diego, La Jolla, CA, USA. [3]Herbert Wertheim School of Public Health and Department of Mathematics, University of California, San Diego, La Jolla, CA, USA. ✉email: thwen@health.ucsd.edu; kpierce@health.ucsd.edu

of 38 published eye-tracking studies indicated that subjects with ASD attend definitively less to social stimuli compared to typically developing individuals[30].

Our previous eye-tracking work using a novel preferential looking paradigm known as the 'GeoPref Test' identified a subgroup of ASD toddlers with heightened visual attention towards geometric relative to social images[16,25]. The effect was robust with > 85% test–retest reliability, 86–100% positive predictive value (PPV), and 98% specificity, although sensitivity was consistently lower at ~ 20%[16,25]. ASD toddlers with a strong non-social preference also had higher symptom severity, worse language and cognitive ability[16], weak functional connectivity between social-visual brain networks[31], and worse school-age outcomes compared to social-preferring toddlers[20]. Because this work was conducted within the context of population-based screening[32,33], one strength was the inclusion of non-ASD contrast groups which mimic the natural constituency found during routine pediatric practice. This is of considerable importance for the development of biomarkers with real-world utility. Weaknesses of the initial studies, however, include small sample sizes (i.e., 110 and 333), and a lack of rigorous methodology to examine predictive validity. We also failed to combine metrics to improve predictive accuracy, focusing instead on fixation levels alone. While other groups have utilized similar paradigms[13–15,26,27,34–36] or a near identical version of the GeoPref Test[25] and report similar results[19,37–48], validation statistics are often unreported.

In addition to validation, establishing impactful biomarkers requires understanding for whom, and at what ages, the marker best applies. In fields outside of autism it is well known that biomarker efficacy differs by race, ethnicity, age, and sex[49,50]. With minor exception, eye-tracking studies within the autism field often have relatively few subjects[11,22,29,51], yet sample sizes in the thousands are necessary for resolving demographic effects and establishing medical biomarkers[52–54]. Moreover, there is a need to examine associations between biomarkers and clinical profiles as a pathway towards individualized medicine. Finally, ASD biomarkers which are tuned for high specificity/low false positive rate are necessary for circumventing the financial burden and familial stress associated with false positive results.

There is also evidence for a genetic component of visual social preference[55–58], but few studies have explored this using eye-tracking. We addressed this previously by comparing GeoPref Test performance in sibling pairs and found that among sibling pairs concordant for ASD, fixation to geometric images was strongly correlated[16]. In a different study, the time non-ASD monozygotic twins spent looking at eyes or mouth in an eye-tracking task was highly intercorrelated[59]. Such findings underscore the potential genetic basis of social visual attention patterns.

Given the necessity for well-developed, clinically relevant ASD biomarkers both for diagnostic purposes and use in clinical trials, the goal of the present study was to comprehensively validate an eye-tracking based biomarker in a large, diverse group of toddlers, producing the largest eye-tracking study of ASD and other delays to date. Here, we report ASD classification accuracy using both a simple, scalable approach associated with a single metric—percent fixation—that can be easily leveraged by non-researchers, as well as a more complex approach which utilizes machine learning algorithms to conduct tenfold validation and an independent replication set based on multiple eye-tracking metrics.

Methods

Participants.
Subjects were referred through the community or via a population-based screening method known as *Get SET Early*[32,33]. Following screening at well baby check-ups using the CSBS IT-Checklist[60], toddlers were referred to the University of California, San Diego Autism Center of Excellence for in-depth diagnostic evaluations and eye-tracking, and invited for repeat testing every ~ 12 months until age 3. Toddlers were assessed by licensed Ph.D.-level clinical psychologists blind to eye-tracking results using the Mullen Scales of Early Learning[61], the Autism Diagnostic Observation Schedule[62], and the Vineland Adaptive Behavior Scales[63]. Parents were given diagnostic feedback and toddlers referred for treatment as appropriate.

Of the 1,685 toddlers enrolled in the study, 266 (15.8%) were excluded largely due to compliance (See Supplemental Methods eFigure 1). The remaining 1,419 toddlers (mean age: 24.37 months, range: 12.00–49.11) were separated into diagnostic groups based on most recent diagnoses including ASD, ASD features (ASD-Feat), global developmental delay (GDD), language delay (LD), typically developing (TD), typical toddlers with an ASD sibling (TypSibASD), and Other (Table 1 and Supplemental Methods). This final sample is independent from our previous work[16,25]. To validate the GeoPref Test on the largest sample possible, secondary analyses included 444 toddlers from our previous work resulting in a cumulative sample of 1,863 toddlers. Among this sample are 11 monozygotic twins, 27 dizygotic twins, and 109 sibling pairs. Given the goal of early biomarker discovery, the first (i.e., youngest age) eye-tracking data collection was used in analyses.

Eye-tracking apparatus, stimuli, and procedures.
Eye-gaze data was collected using the Tobii T120 (Tobii, Stockholm, Sweden; www.tobii.com; 60 Hz sampling rate; 1280 × 1024) while toddlers watched 'The GeoPref Test' (62.22 s), which consisted of two rectangular areas of interest (AOIs, 525 × 363 pixels) each containing dynamic geometric (DGI) or social images (DSI; social images used with permission from Gaiam Americas Inc., Copyright 2003, Gaiam Americas, Inc.), identical to stimuli used in our previous work[16,17,25] (Fig. 1A; Supplemental Methods eFigure 2). To control for biases due to spatial location, side of stimulus presentation varied across subjects.

To ensure that only the toddler's gaze was tracked and free from parent influence, standardized instructions were read to parents. A five-point calibration was then performed using animated cartoon ducks with sounds, and data was only used if calibration results, determined via graphical output and verified via screenshots, fell within manufacturer-reported parameters (accuracy, 0.5 degrees[64]). For a subset of toddlers, a flashing star with chime appeared for 6.23 s prior to the start of the experiment to ensure toddlers fixated to the screen.

Data was processed using Tobii Studio (Tobii Fixation Filter, velocity threshold: 35 ms/window). Total fixation duration, fixation count within each AOI, and fixation duration within each AOI were exported and analyzed

	Mean (SD)						
	ASD (N = 725)	ASD-Feat (N = 103)	GDD (N = 128)	LD (N = 198)	Other (N = 162)	TD (N = 487)	TypSibASD (N = 60)
M/F	563/162	85/18	98/30	149/49	113/49	295/192	28/32
Age, months	26.40 (8.25)	23.84 (9.17)	26.38 (9.81)	20.78 (7.44)	23.15 (9.31)	23.32 (9.17)	21.86 (8.79)
Range	[12–49]	[11–44]	[12–46]	[10–48]	[11–48]	[11–48]	[12–44]
Ethnicity (%)							
Hispanic or Latino	32.69%	29.13%	51.56%	42.93%	23.46%	19.71%	23.33%
Non-Hispanic or Latino	56.00%	63.11%	42.97%	50.00%	69.14%	71.25%	65.00%
Unknown/not reported	11.31%	7.77%	5.47%	7.07%	7.41%	9.03%	11.67%
Race (%)							
Caucasian	52.14%	66.99%	54.69%	53.03%	65.43%	67.15%	66.67%
American Indian/ Alaska Native	0.83%	0.97%	0.78%	1.52%	1.23%	0.41%	0.00%
Asian	13.10%	5.83%	3.91%	8.08%	4.94%	8.01%	3.33%
Black/African American	2.48%	1.94%	1.56%	3.54%	1.85%	2.46%	3.33%
Pacific Islander/Native Hawaiian	0.83%	2.91%	2.34%	1.01%	1.23%	0.82%	1.67%
More than one race	9.79%	8.74%	10.94%	5.56%	13.58%	8.83%	8.33%
Not reported/ unknown/other	20.83%	12.62%	25.78%	27.27%	11.73%	12.32%	16.67%
Mullen (t score)							
Visual reception	38.00 (13.04)	49.17 (12.31)	37.31 (11.82)	49.01 (10.13)	52.35 (11.87)	56.16 (9.68)	57.34 (9.34)
Fine motor	37.59 (16.23)	47.62 (12.05)	37.68 (12.39)	49.50 (10.31)	49.85 (11.84)	54.21 (9.14)	54.71 (8.86)
Receptive language	28.47 (12.78)	44.09 (13.06)	34.60 (10.79)	39.78 (11.86)	48.08 (13.15)	52.12 (10.78)	51.68 (10.42)
Expressive language	27.80 (13.03)	40.97 (12.17)	29.35 (11.23)	32.79 (9.59)	44.67 (12.04)	48.93 (10.72)	52.68 (10.44)
Early learning composite	69.01 (19.39)	91.66 (18.37)	71.19 (18.05)	86.20 (14.65)	96.83 (20.22)	105.83 (13.73)	108.63 (13.46)
Vineland (standard score)							
Communication	77.20 (13.90)	91.33 (12.41)	81.62 (13.32)	86.45 (9.66)	94.23 (12.98)	99.56 (11.38)	98.60 (15.28)
Daily living	84.65 (12.61)	92.38 (11.53)	87.59 (14.63)	95.27 (10.54)	95.23 (13.02)	99.67 (11.21)	100.67 (10.00)
Socialization	84.23 (11.97)	95.16 (9.72)	90.57 (12.92)	96.12 (10.89)	98.15 (11.73)	102.62 (9.44)	103.48 (8.82)
Motor skills	90.76 (10.67)	97.40 (10.63)	86.75 (16.67)	96.49 (8.81)	96.62 (11.95)	99.26 (9.63)	100.57 (8.80)
Adaptive behavior composite	81.08 (11.12)	92.50 (10.40)	84.50 (12.94)	91.37 (10.56)	95.07 (11.79)	99.99 (9.86)	101.10 (8.97)
ADOS (module T or 1)							
ADOS RRB score	4.36 (2.16)	2.22 (1.71)	1.39 (1.51)	0.96 (1.20)	1.13 (1.35)	0.52 (0.92)	0.57 (1.01)
ADOS Sa/CoSo tot score	13.24 (4.78)	6.94 (4.73)	4.72 (3.65)	3.82 (2.86)	4.06 (3.57)	2.87 (2.59)	2.05 (1.87)
ADOS total score	17.60 (5.87)	9.16 (5.37)	6.10 (4.20)	4.77 (3.18)	5.20 (4.16)	3.39 (2.96)	2.62 (2.24)

Table 1. Summary of clinical characteristics for 1,863 toddlers who completed The GeoPref Test. Values are presented as means (standard deviations are noted in parentheses unless otherwise noted, and age range provided in brackets). ASD: Autism Spectrum Disorders, ASD-Feat: ASD Features, GDD: Global Developmental Delay, LD: Language Delay, TD: Typically Developing, TypSibASD: Typical Sibling of subject with ASD; ADOS: Autism Diagnostic Observation Scale, SA/CoSo Score: Social Affect/Communication Social Score, RRB Score: Restricted and Repetitive Behavior Score.

offline. Percent fixation duration/AOI was computed by dividing the total fixation duration within an AOI by the fixation duration across the entire video. N-1 total fixations/total fixation duration was used to calculate saccades/sec within each AOI.

Defining ASD subgroups. For comparability with our earlier work and to tune the test towards a low false positive rate, 69% fixation to either DGI or DSI was used to distinguish between ASD toddlers who strongly preferred to visually fixate on geometric images (ASD_Geo) or social images (ASD_Soc)[16,25]. Given that experience-dependent mechanisms may differentially impact toddlers with distinct attention patterns, stratifying ASD toddlers in this way may have relevance for understanding long term outcomes among subtypes identified through eye-tracking. ASD toddlers who lacked a strong preference for either stimulus (i.e. they fixated on DGI for 32–68% of the time) were categorized as ASD Middle responders (ASD_Mid).
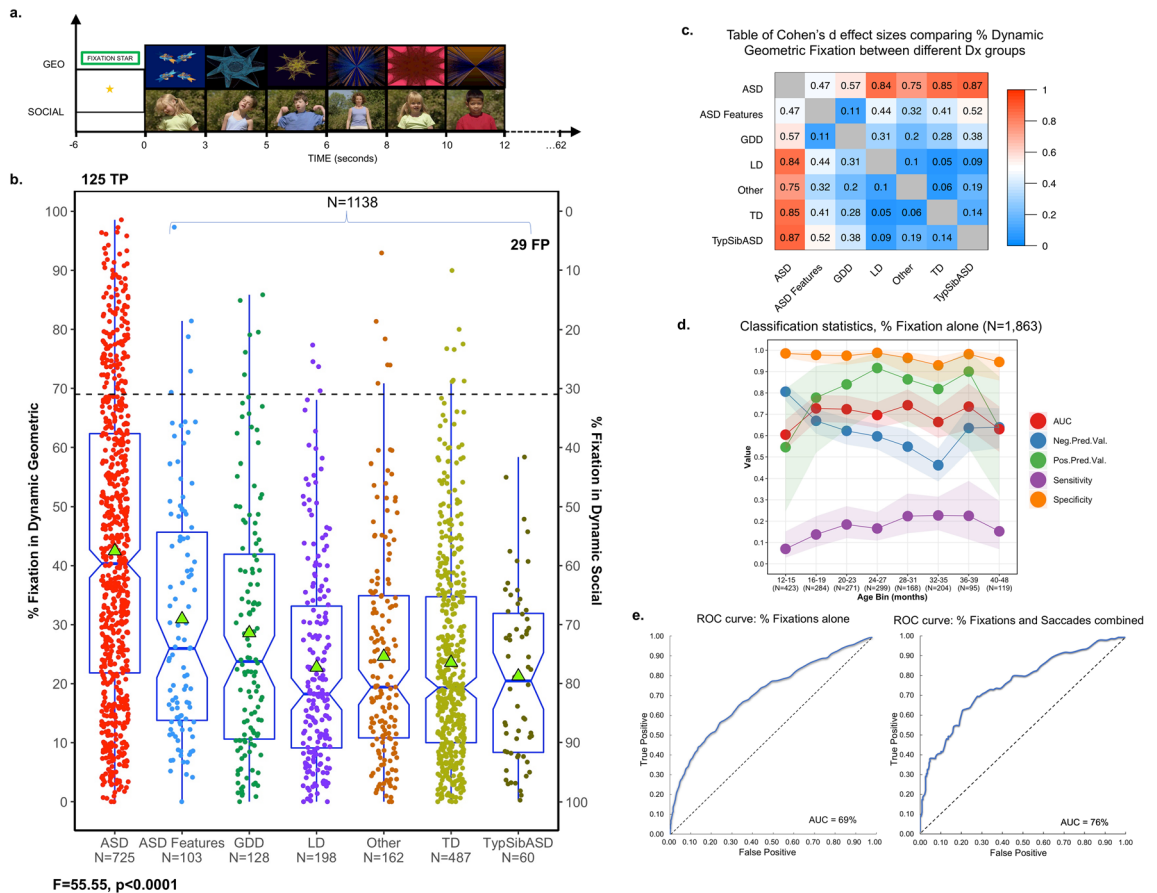
Figure 1. Validation of visual attention patterns among a combined sample of 1,863 toddlers of varying developmental types. (**a**) Sample images from the GeoPref eye-tracking test across the first 12 s. Social images courtesy of Gaiam Americas Inc., Copyright 2003, Gaiam Americas, Inc. To standardize initial fixations across toddlers, testing was preceded by a fixation star which flashed for 6.23 s. The entire length of the test was 62.23 s. (**b**) In the current study, 444 toddlers from a previous analysis[16] were combined with a new, independent sample of 1,419 toddlers in order to increase power and provide the largest dataset possible for subsequent analyses. As such, scatterplot demonstrates percent fixation to dynamic geometric images (DGI) or dynamic social images (DSI) for the combined sample of 1,863 toddlers. Percent fixation was defined as fixation duration within geometric or social areas of interest divided by total fixation duration to the entire video. Dashed lines depict threshold for 69% fixation. Green triangles indicate average percent fixation. Notched boxplots show median, range, and first/third quartiles. Notch ranges indicate 95% confidence intervals around the median. TP: true positive, FP: false positive. F statistic pictured was obtained from a one-way ANOVA conducted to compare percent DGI fixation across diagnostic groups. (**c**) Effect size matrix of pairwise comparisons of average percent fixation to DGI for toddlers of varying developmental types. Warmer colors indicate larger effect sizes and cooler colors indicate smaller effect sizes. (**d**) Validation statistics for classification of toddlers as ASD vs. non-ASD using a 69% fixation threshold after grouping toddlers into 4-month age bins. Toddlers > 40 months were binned together given the small sample size of toddlers beyond this age. (**e**) ROC curves showing classification performance (ASD vs. non-ASD) among 20% of toddlers which make up an independent, hold-out test using left, percent DGI fixation alone (threshold: 61.3%), or right, multiple eye-tracking metrics (percent fixation, threshold: 61.3% and saccades/second, threshold: 2.29 saccades/second). ASD: Autism Spectrum Disorders, ASD-Feat: ASD Features, GDD: Global Developmental Delay, LD: Language Delay, TD: Typically Developing, TypSibASD: Typical Sibling of subject with ASD, AUC: Area under the curve, Neg. Pred. Val.: Negative Predictive Value, Pos. Pred. Val.: Positive Predictive Value.

Statistical analyses. All analyses were conducted using R and relevant packages (e.g., pROC, ROCR[65–74]). Results are reported using associated 95% confidence intervals (CI) and effect sizes as appropriate.

Visual attention preference stratified by diagnosis and correlation with clinical measures. One-way ANOVA was performed to compare percent fixation towards DGI across diagnostic groups. Follow-up pairwise comparisons of group means were conducted using Tukey's HSD.

Linear regression with percent DGI fixation as the outcome variable and sex, age, and diagnosis as predictor variables was used to examine associations between demographics and DGI fixation levels. Pearson's correlations were conducted to examine relationships between DGI fixation and performance on clinical measures.

Across ASD subgroups, performance on clinical measures and associated subscales was assessed via one-way ANCOVA with age and sex as covariates. Subgroup-level differences were assessed using Tukey's HSD.

Classification accuracy: traditional approaches. To determine whether validation statistics for the current sample were comparable to those observed in our previous work[16,25], and to tune the test in favor of a low false-positive rate, the same 69% fixation threshold was applied here to the full cohort to compute Receiver Operating Characteristic (ROC) curves, an Area Under the Curve (AUC) statistic, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

Validation statistics were also computed after stratifying toddlers based on age bins (4-months), ethnicity, race, and sex.

Saccade profiles within ASD. Our previous work indicated that ASD_{Geo} toddlers exhibited significantly fewer saccades when viewing DGI and significantly more when viewing DSI relative to other ASD toddlers[16,17,25]. To determine whether these patterns were also present in this larger sample, a one-way ANCOVA with age and sex as covariates and a follow up Tukey's HSD test was used.

Classification accuracy: cross validation with independent hold out set. Rigor, generalizability, and diagnostic accuracy of the GeoPref Test were examined using a tenfold cross-validation approach conducted on 80% of the cohort who were randomly selected to be included in the analysis. Ten-fold cross-validation is a widely used method for parameter tuning and threshold selection in statistical learning and classification problems[75]. Average performance statistics for all 10 validations were averaged and reported, and then used to classify the remaining 20% of toddlers which made up an independent, hold-out test set. See Supplemental Methods.

Classification sensitivity enhancement was also examined using tenfold cross-validation following the inclusion of both percent fixation and a second eye-tracking metric, number of saccades per second within DSI, as variables of interest.

Test–retest reliability. Additional eye-tracking sessions (not used in aforementioned analyses) occurred in 535 toddlers (mean age: 19.48 ± 6.18 months) within 1 to 25 + months of their first eye-tracking test. After grouping toddlers based on interval length (immediate: 0–1 month, n = 39; short term: 2–6 months, n = 100; intermediate term: 7–12 months, n = 211; long term: 13–24 months, n = 163; and very long term: > 25 months, n = 22), intraclass correlations and paired t-tests were used to examine differences in percent DGI fixation between tests.

Development of visual social attention preference across the first years of life. To better understand the trajectory of typical versus atypical visual attention patterns, Pearson's correlations comparing age and percent DSI/DGI fixation were conducted for all diagnostic groups. The impact of age on eye-tracking performance was further examined using linear regression with percent fixation on DSI as the dependent variable, and diagnosis, age, and diagnosis x age as predictor variables.

Examination of genetic underpinnings of visual social attention. Intraclass correlations were used to determine the concordance among social visual attention patterns between siblings, twin pairs and a random pairing of 850 unrelated subjects.

Ethical approval statement. This study was approved by the Institutional Review Board at the University of California, San Diego (IRB #181,652, #172,066, #081,722, #041,715, #140,673, #130,352, #110,049, #070,229) and performed in accordance with the UCSD Human Research Protections Program guidelines. Prior to data collection, informed consent was obtained from all subjects and/or their legal guardians for study participation. GeoPref Test social images are copyrighted by Gaiam Americas Inc. and were used with permission.

Results

Visual attention preference stratified by diagnosis and correlation with clinical measures. Similar to our previous study[25], results from the independent sample of 1,419 toddlers indicated significant differences in the amount of time a toddler fixated on DGI based on diagnostic group membership ($F(6,1,412) = 43.74$, $p < 0.0001$). There were no differences in terms of data distribution (Kolmogorov–Smirnov Z Test, $D = 0.042$, $p = n.s.$) or effect sizes between the new independent sample and our prior smaller sample of 444[25] toddlers, and thus both were combined to increase power and provide the largest dataset possible for subsequent analyses (N = 1,863). See Supplemental Results and eFigure 3 for the independent sample data. Within the combined sample of 1,863, toddlers with ASD exhibited the highest percent fixation to DGI compared to all other toddler types (ASD 95% CI [40.67, 44.30] vs. ASD-Feat 95% CI [26.88, 35.03], mean difference: 11.53 ± 4.03, $p < 0.0001$, $d = 0.47$ 95% CI [0.26, 0.68]; ASD vs. GDD 95% CI [24.86, 32.35], mean difference: 13.88 ± 3.46, $p < 0.0001$, $d = 0.57$ 95% CI [0.38, 0.76]; ASD vs. LD 95% CI [20.30, 25.19], mean difference: 19.74 ± 7.44, $p < 0.0001$, $d = 0.84$ 95% CI [0.68, 1.00]; ASD vs. Other 95% CI [21.70, 27.51], mean difference: 17.88 ± 6.15, $p < 0.0001$, $d = 0.75$ 95% CI [0.57, 0.92]; ASD vs. TD 95% CI [22.03, 25.12], mean difference: 18.91 ± 7.54, $p < 0.0001$, $d = 0.85$ 95% CI [0.73, 0.97]; ASD vs. TypSibASD 95% CI [17.50, 25.04], mean difference:

21.21 ± 10.29, $p < 0.0001$, $d = 0.87$ 95% CI [0.61, 1.14]). Preference for DGI was also high for the ASD-Feat group (ASD-Feat vs. TD mean difference: 7.38 ± 3.51, $p < 0.05$, $d = 0.41$ 95% CI [0.20, 0.62]; ASD-Feat vs. LD mean difference: 8.21 ± 3.41, $p < 0.05$, $d = 0.44$ 95% CI [0.20, 0.68]). TD, LD, GDD, TypSibASD, and toddlers categorized as Other exhibited a stronger preference for DSI, and preference strength was comparable between groups. See Fig. 1. There were no differences in percent fixation levels when data was stratified by sex, although small differences were associated with ethnicity and race. See Supplemental Results eFigure 4–6.

Since diagnosis and age were significant predictors of DGI fixation (overall $R^2 = 0.21$, $F(7,1855) = 72.27$, $p < 0.0001$), correlation coefficients were computed for each diagnostic group separately to determine relationships between DGI fixation and clinical symptoms. For toddlers with ASD, percent DGI fixation was significantly correlated with all clinical measures and all associated subscales. In contrast, apart from the visuospatial subscale on the Mullen, there were no significant relationships found between DGI fixation and clinical profiles within the TD or TypSibASD toddlers. See Supplemental Results eTable 1.

Among the 3 ASD subtypes (see Methods), significant differences in symptom severity (ADOS total score $F(2,707) = 37.21$, $p < 0.0001$), cognitive ability (Mullen Early Learning Composite $F(2,692) = 21.04$, $p < 0.0001$), and adaptive behavior (Vineland Adaptive Behavior Composite ($F(2,707) = 20.84$, $p < 0.0001$) were observed, suggesting the possibility of unique underlying biological profiles. The largest differences were between toddlers who strongly preferred geometric (ASD_{Geo}) and those that strongly preferred social (ASD_{Social}) images. See Fig. 2a.

Classification accuracy: traditional approaches.
ROC curves leveraging the full dataset yielded 98% specificity, 17% sensitivity, 81% PPV, and 65% NPV when the 69% fixation threshold was used. See Supplemental Results eTable 2. These values are nearly identical to validation statistics computed in previous work using the same fixation threshold[16]. These results indicate that even among a large population of toddlers of various developmental types, the GeoPref Test accurately distinguishes ASD from non-ASD toddlers, with relatively few false positives (i.e., 2%), which is a key criterion for biomarker tests[76].

Age-binned data show that while specificity remains high throughout development (> 90%), sensitivity is consistently low. Negative predictive value is highest at 12 months and lowest by 32 months, while peak positive predictive value is achieved by 24 months. See Fig. 1d.

We additionally examined classification statistics after stratification using demographic factors and found that the GeoPref Test performs similarly across sex, ethnic and racial groups. See Supplemental Results eTable 3.

Saccade profiles within ASD.
In our previous work, we demonstrated that ASD_{Geo} toddlers exhibited significantly fewer saccades/second when viewing DGI, but greater saccade rates when viewing non-preferred social images, in contrast to ASD_{Soc} toddlers who had near-typical saccade patterns[16]. This was replicated in the current large sample when saccades/second was examined within DGI ($F(8,1847) = 9.65$, $p < 0.0001$) or DSI ($F(8,1847) = 33.24$, $p < 0.0001$) images. All planned comparisons between ASD_{Geo} and other diagnostic groups were significant, with the largest effect sizes found between ASD_{Geo} and TD toddlers. See Fig. 2b-e.

Classification accuracy: cross validation with independent hold out set.
To enhance the rigor of the GeoPref Test, we next performed tenfold cross validation to determine appropriate DGI fixation thresholds for computing validation statistics. Using this method, the ideal fixation threshold was 61.3%, which yields 95% specificity, 23% sensitivity, 76% PPV, 66% NPV, and 67% accuracy. When this same threshold was applied to an independent, hold-out test set, the GeoPref Test had 96% specificity, 33% sensitivity, 84% PPV, 69% NPV, and 71% accuracy.

Despite high specificity using the 69% fixation threshold[16,25] or the 61% fixation threshold obtained from tenfold cross validation, sensitivity remained low. However, combining eye-tracking measures, including saccades per second within DSI (optimal threshold from tenfold cross validation was 2.29 saccades/sec), and percent DGI fixation (61.3% fixation threshold), enhanced GeoPref Test sensitivity to 33.3%, with little impact on specificity (95.2%), PPV (81.4%) or NPV (71.2%) for the independent, hold-out test set. See Fig. 1e.

Test–retest reliability.
A key component of biomarker validation research is characterizing the stability of test performance across repeated measurements. Five hundred and thirty-five toddlers participated in repeat eye tracking. High levels of reliability were observed for more immediate, 0–1 month retests (intraclass correlation coefficient = 0.76, $p < 0.0001$; paired samples t-test $t(75.32) = -0.72$, $p > 0.05$). Longer interval retests were still reliable, although correlation strength was reduced. This result is expected as the GeoPref Test was created for assessing ASD during the toddler period, during which age-related visual attention changes are expected. See Supplemental Results eTable 4.

Developmental trajectory of visual social attention preference across the first years of life.
Examination of the correlation between social and non-social fixation levels and age across all diagnostic groups and ASD subtypes revealed an interesting trend: both ASD toddlers with the social profile (ASD_{Soc}) and other non-ASD toddlers significantly decreased their interest in social images with age, alongside a concomitant increase in interest in non-social images. In contrast, toddlers with the geometric profile (ASD_{Geo}) as well as those that fell into the middle category (ASD_{Mid}) showed no age-related changes. See Fig. 3. Follow-up linear regression analyses and pairwise comparisons of beta coefficients confirm this result (overall fit: $F(17,1845) = 151.5$, $R^2 = 0.58$; ASD_{Geo} vs. TD $p < 0.001$, ASD_{Geo} vs. TypSibASD $p < 0.05$, ASD_{Geo} vs. GDD $p < 0.0001$, ASD_{Geo} vs. LD $p < 0.001$, ASD_{Geo} vs. Other $p < 0.01$, ASD_{Geo} vs. ASD-Feat $p < 0.001$).
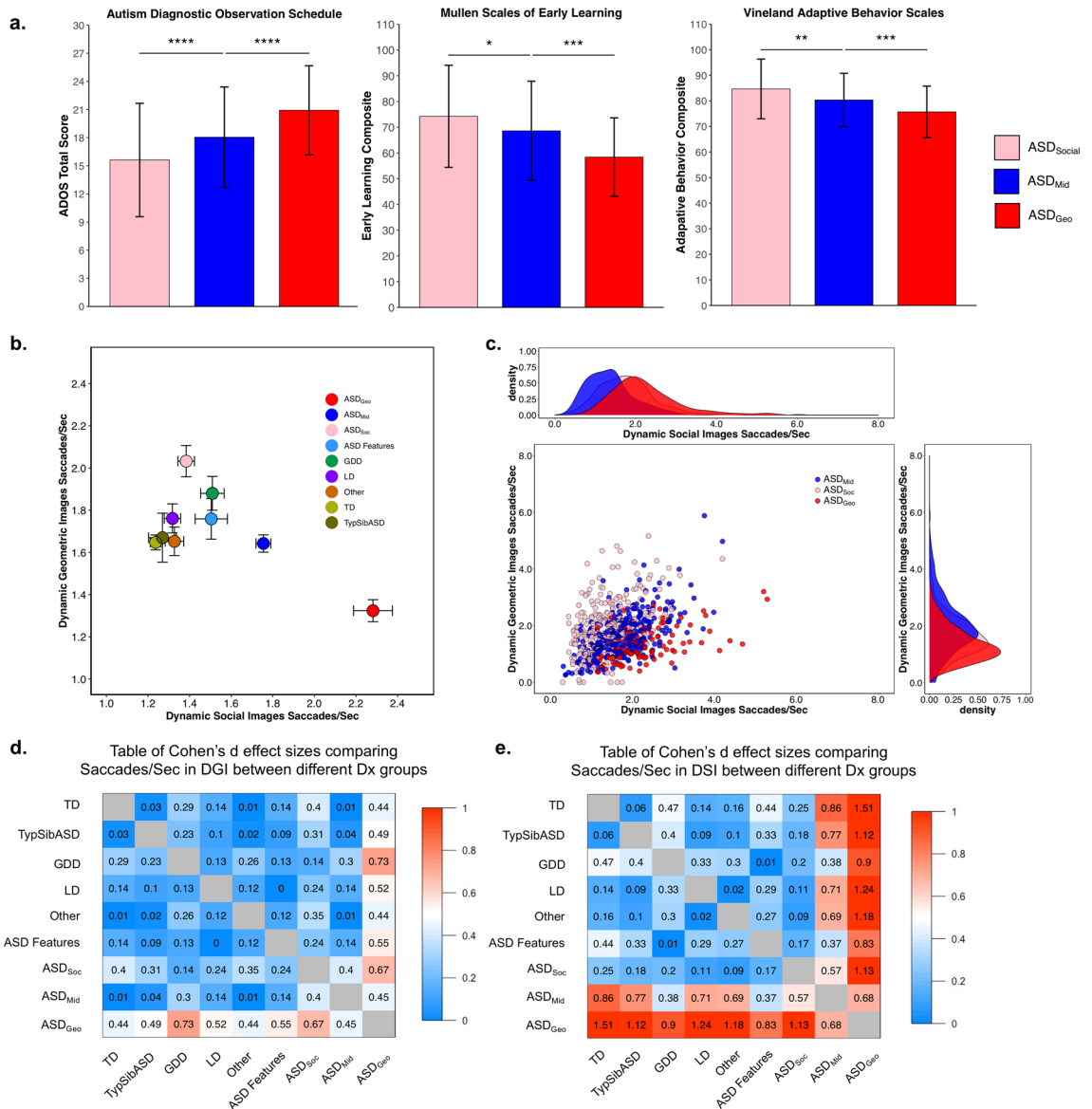
Figure 2. Differences in clinical and saccade profiles across ASD subgroups. (**a**) Bar graphs (mean ± SD) demonstrating clinical scores obtained from ASD toddlers with varying levels of preference for dynamic geometric images (DGI). ASD toddlers with a strong preference for DGI (≥ 69% fixation to DGI) were designated as ASD_{Geo}. ASD toddlers with a strong preference for dynamic social images (DSI) (≥ 69% fixation to DSI) were designated as ASD_{Soc}. Toddlers who lacked a strong preference for either stimulus were designated as ASD_{Mid} toddlers. (**b**) Scatterplot showing average saccades/sec across diagnostic groups (mean ± SEM) to dynamic geometric images (DGI, y-axis) or dynamic social images (DSI, x-axis). (**c**) Scatterplot showing saccades/sec to DGI (y-axis) or DSI (x-axis) for all ASD subgroups. Density plots demonstrate distribution of individual saccade patterns to DGI or DSI. (**d,e**) Cohen's d effect sizes associated with pairwise comparisons of average saccades/sec to DGI or DSI. ASD: Autism Spectrum Disorders, ASD-Feat: ASD Features, GDD: Global Developmental Delay, LD: Language Delay, TD: Typically Developing; TypSibASD: Typical Sibling of subject with ASD.

Sibling and twin correlations highlight the genetic underpinnings of visual social attention in ASD.

The current sample size allows us to compare DGI fixation among a larger group of siblings than our previous work[16]. ASD Concordant pairs of siblings or twins exhibited the highest ICC values compared to ASD NonConcordant pairs, NonASD Concordant pairs, NonASD NonConcordant pairs, and randomly paired subjects (Fig. 4). These findings highlight the sensitivity of the GeoPref Test for detecting a genetically driven subtype of ASD.
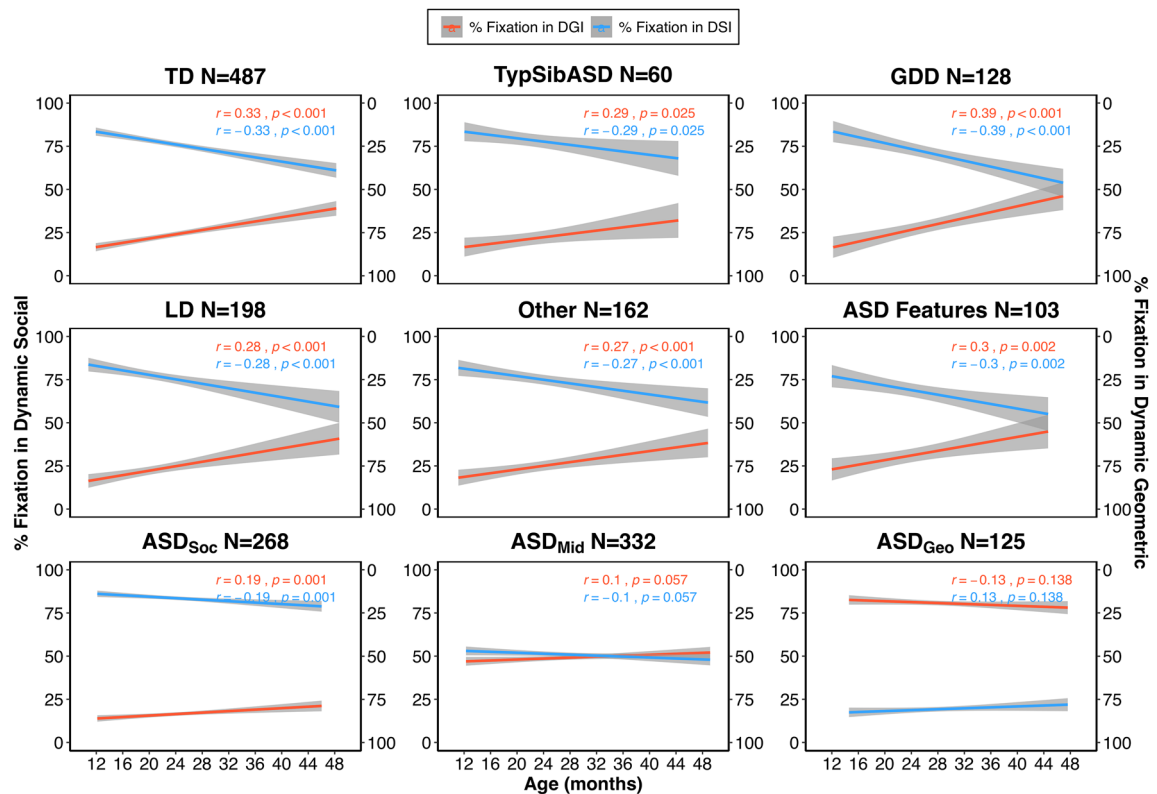
Figure 3. Developmental trajectories of social and non-social (geometric) attention distinguish ASD toddlers from toddlers of other developmental types. Best fit lines depicting developmental trajectories of percent fixation to social images (blue) or non-social images (red) for toddlers of different developmental types and associated Pearson's r and p-values. ASD toddlers were again split into the three ASD subgroups based on visual attention preference for social (ASD_{Soc}) or non-social stimuli (ASD_{Geo}). ASD_{Mid} toddlers did not exhibit a strong preference for social or non-social stimuli. ASD: Autism Spectrum Disorders, ASD-Feat: ASD Features, GDD: Global Developmental Delay, LD: Language Delay, TD: Typically Developing, TypSibASD: Typical Sibling of subject with ASD.

Discussion

ASD begins during pregnancy[2,77–80], and thus it is not surprising that parents become aware of developmental delays within the first months of their child's life. Despite this, age at first diagnosis has remained stable at around ~ 52 months across the past decade[3,4,81]. To the degree that early identification and interventions are beneficial, alternatives to the long diagnostic journey are needed. Biomarkers offer one such alternative.

Using the largest eye-tracking sample to date with 1,863 toddlers who received both eye-tracking and a diagnostic evaluation by licensed psychologists, here we present comprehensive validation of an eye-tracking biomarker of an autism spectrum disorder subtype that quantifies a toddler's attention to non-social images that is highly replicable and reliable. Importantly, participating toddlers were largely first identified through universal screening, underscoring the notion that eye-tracking may be an excellent 2nd tier screen or diagnostic tool.

The GeoPref Test was examined in multiple ways. First, as a simple tool by examining only percent fixation on geometric images, which requires little to no computational sophistication, supporting use by clinicians and researchers alike. Moreover, the user can determine the specificity rate they prefer, as illustrated in the ROC table (See Supplemental Results eTable 2), and select the associated fixation cut off level. For example, in our study, 69% fixation was selected, which results in a 2% false positive rate. Tuning biomarker tests towards a very low false positive rate may be particularly important for disorders of infancy to avoid unnecessary parental stress associated with false positive results. Second, using a more rigorous machine learning, tenfold cross-validation approach with potentially more generalizable results than standard approaches, the present study found 95% specificity and 23% sensitivity. Incorporating an additional measure, saccades per second while viewing social images, increased the sensitivity to 33% while maintaining levels of specificity at 95%. This supports the notion that combining eye-tracking paradigms and/or metrics can bolster classification accuracy[34,82].

Although the GeoPref Test has a low false positive rate and exceptional specificity, sensitivity was modest. Given the considerable heterogeneity inherent in ASD[31,83–86], and the fact that several studies highlight the likelihood of specific subtypes in ASD, this is not surprising[87,88]. Toddlers who demonstrate reduced levels of social visual attention as measured by the GeoPref Test may indeed represent a unique biological subtype. In comparison to toddlers who strongly preferred social stimuli, average symptom severity among ASD_{Geo} toddlers as indexed by the ADOS was 5.6 points higher, while levels of cognition and adaptive behavior were 15.8 and 9.0 points lower, respectively. A unique imaging study of toddlers who received both eye-tracking and brain imaging
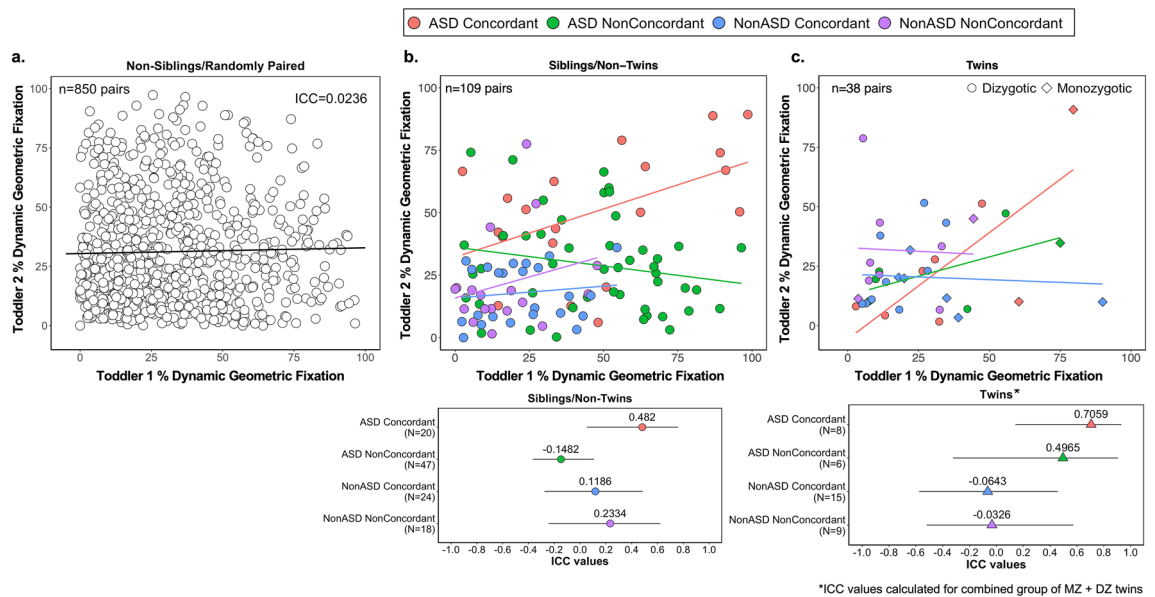
Figure 4. Eye-tracking among siblings and monozygotic and dizygotic twins highlights the genetic basis of the GeoPref Test. Scatterplots comparing percent fixation to dynamic geometric images (DGI) among (**a**) non-sibling/randomly paired toddlers, (**b**) sibling/non-twin toddler pairs, and (**c**) twin toddler pairs. Intraclass correlation coefficients are also shown in (**a**) or plotted, along with the 95% confidence intervals, based on concordance for ASD (**b,c**, bottom). ASD Concordant: both siblings/twins received a diagnosis of ASD. ASD NonConcordant: One sibling/twin received an ASD diagnosis while the other received a non-ASD diagnosis. NonASD Concordant: both siblings/twins received the same diagnosis but were non-ASD. NonASD NonConcordant: Each sibling/twin had a different, non-ASD diagnosis.

revealed that ASD_{Geo} toddlers exhibited unusually low levels of connectivity between areas classically associated with the 'social brain' (e.g., cingulate) and visual cortex[31], and a separate study found strong correlations between social visual attention levels and language/visual cortex connectivity in this subgroup[89]. Collectively, these studies suggest that ASD_{Geo} toddlers are more symptomatic with a unique genetic profile that drives abnormal neural development, particularly as it relates to connectivity with visual cortex. Future studies which incorporate additional eye-tracking paradigms tailored to quantify other features and/or subtypes[12,13,82,90–92] in conjunction with the GeoPref Test will likely capture more of the variance associated with ASD and improve test sensitivity.

Although there is extensive evidence supporting the notion of reduced social attention in ASD[30], it is not possible to definitively conclude whether the present findings reflect a failure of social attention, or a difference in visual preference driven by sensory issues common in ASD[93,94]. For instance, the geometric patterns are per pixel more dynamic than the social videos and involve more color change. The geometric patterns may also be more unpredictable than the social scene. There is evidence that altered sensory processing may be a direct or indirect driver of social attention and/or orienting and higher order social ability among ASD toddlers[95–97], indicating that the GeoPref Test may be a correlative measure of ability in either or both domains. Future characterization of sensory profiles across groups may help to deepen our understanding of drivers of visual attention patterns in ASD in general, and the ASD_{Geo} subtype more specifically.

Although our test was specifically tuned for the 12–48-month age range, an important consideration is whether biomarkers have comparable efficacy across target ages. Here we found good psychometric properties between 12 and 39 months, with decline in accuracy starting around 40 months, which may be attributed to age-related changes in social preference in typically developing infants[98–104]. This process, likely driven by frontal cortex synapse proliferation followed by selective pruning across the first years of life[105], affords the child greater curiosity, control, and preference for novelty as they age[106,107]. Indeed, TD and non-ASD delayed toddlers demonstrated a simultaneous reduction in social preference and increase in geometric preference (i.e., "novelty") across age. While ASD_{Soc} toddlers showed a profile almost identical to typically developing toddlers, ASD_{Geo} toddlers did not. To the degree that eye tracking performance has external validity and can serve as a proxy for real-world social engagement, results from the GeoPref Test may be useful as a prognostic metric. Indeed, one study noted that toddlers with ASD who preferred social images had better symptom profiles at school age than those that preferred geometric images[20].

The psychometric properties of the GeoPref Test were also comparable across demographic categories, which may have been bolstered by the fact that it is a visual-only (i.e., no sound) test, potentially circumventing biases associated with language or culture[108–111]. Although the gap between first age of diagnosis and treatment access is narrowing between Caucasian and non-Caucasian children[3], racial and ethnic inequities persist[3,112]. Females are also more likely to be diagnosed at older ages[108,109,113]. Such findings underscore the need for the development and implementation of culture-free, objective tools which support equal access to early diagnosis and treatment[114].

The high intercorrelation of geometric fixation levels between ASD siblings and twins in our study suggests that eye-tracking biomarkers may be essential for early identification of genetic forms of ASD. A recent study noted a high intercorrelation in visual social attention towards the mouth and eyes among non-ASD twins, but not in unrelated children[59]. These results suggest that eye-tracking measures of social visual attention can be driven by ASD-related genetic variance or by genes which drive social attention. These genes may not be mutually exclusive. For instance, infants homozygous for the CD38 risk allele exhibited less gaze to happy eyes compared to infants heterozygous or homozygous for the non-risk allele[56]. Relatedly, CD38 risk allele expression is associated with higher ASD symptom severity[115]. Future work examining genetic profiles which drive ASD subtypes will help clarify the impact of ASD and other genetic factors on eye-tracking performance.

The large sample size in the current study, and the strong psychometric properties across age, sex, race and ethnic groups generates confidence that the GeoPref eye-tracking Test has value as both a clinical and research tool. Moreover, our sample included multiple non-ASD contrast groups (e.g., LD, GDD) which mimics natural pediatric practice. Importantly, performance on clinical measures in the current study correlated with geometric fixation levels, suggesting that the GeoPref Test can directly index a core ASD phenotype, which is relevant for measuring clinical improvements in clinical trials. Eye-tracking metrics outlined in the current study and others like it may be key in ensuring that severely impacted toddlers receive early diagnosis and treatment access, and promote biotherapeutic and behavioral treatment development, which can contribute to better outcomes and quality of life.

Data availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

References

1. Bonnet-Brilhault, F. *et al.* Autism is a prenatal disorder: Evidence from late gestation brain overgrowth. *Autism Res.* **11**, 1635–1642. https://doi.org/10.1002/aur.2036 (2018).
2. Courchesne, E., Gazestani, V. H. & Lewis, N. E. Prenatal origins of ASD: The when, what, and how of ASD development. *Trends Neurosci.* **43**, 326–342. https://doi.org/10.1016/j.tins.2020.03.005 (2020).
3. Maenner, M. J. *et al.* Prevalence of autism spectrum disorder among children aged 8 Years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016. *MMWR Surveill. Summ.* **69**, 1–12. https://doi.org/10.15585/mmwr.ss6904a1 (2020).
4. Baio, J. *et al.* Prevalence of autism spectrum disorder among children aged 8 years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. *MMWR Surveill. Summ.* **67**, 1–23. https://doi.org/10.15585/mmwr.ss6706a1 (2018).
5. Courchesne, E. *et al.* The ASD living biology: from cell proliferation to clinical phenotype. *Mol. Psychiatry* **24**, 88–107. https://doi.org/10.1038/s41380-018-0056-y (2019).
6. De La Torre-Ubieta, L., Won, H., Stein, J. L. & Geschwind, D. H. Advancing the understanding of autism disease mechanisms through genetics. *Nat. Med.* **22**, 345–361. https://doi.org/10.1038/nm.4071 (2016).
7. Frazier, T. W. *et al.* Evidence-based use of scalable biomarkers to increase diagnostic efficiency and decrease the lifetime costs of autism. *Autism Res.* https://doi.org/10.1002/aur.2498 (2021).
8. McPartland, J. C. *et al.* The autism biomarkers consortium for clinical trials (ABC-CT): Scientific context, study design, and progress toward biomarker qualification. *Front. Integr. Neurosci.* **14**, 1–7. https://doi.org/10.3389/fnint.2020.00016 (2020).
9. Mcpartland, J., Dawson, G., Webb, S. J., Panagiotides, H. & Carver, L. J. Event-related brain potentials reveal anomalies in temporal processing of faces in autism spectrum disorder. *J. Child Psychol. Psychiatry* **45**, 1235–1245. https://doi.org/10.1111/j.1469-7610.2004.00318.x (2004).
10. Kanner, L. Autistic disturbances of affective contact. *Nervous Child* **2**, 217–250 (1943).
11. Klin, A., Jones, W., Schultz, R., Volkmar, F. & Cohen, D. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Arch. Gen. Psychiatry* **59**, 809–816. https://doi.org/10.1001/archpsyc.59.9.809 (2002).
12. Sasson, N. J., Turner-Brown, L. M., Holtzclaw, T. N., Lam, K. S. L. & Bodfish, J. W. Children with autism demonstrate circumscribed attention during passive viewing of complex social and nonsocial picture arrays. *Autism Res.* **1**, 31–42. https://doi.org/10.1002/aur.4 (2008).
13. Sasson, N. J. & Touchstone, E. W. Visual attention to competing social and object images by preschool children with autism spectrum disorder. *J. Autism Dev. Disord.* **44**, 584–592. https://doi.org/10.1007/s10803-013-1910-z (2014).
14. Chevallier, C. *et al.* Measuring social attention and motivation in autism spectrum disorder using eye-tracking: Stimulus type matters. *Autism Res.* **8**, 620–628. https://doi.org/10.1002/aur.1479 (2015).
15. Unruh, K. E. *et al.* Social orienting and attention is influenced by the presence of competing nonsocial information in adolescents with autism. *Front. Neurosci.* **10**, 1–12. https://doi.org/10.3389/fnins.2016.00586 (2016).
16. Pierce, K. *et al.* Eye tracking reveals abnormal visual preference for geometric images as an early biomarker of an autism spectrum disorder subtype associated with increased symptom severity. *Biol. Psychiatry* **79**, 657–666. https://doi.org/10.1016/j.biopsych.2015.03.032 (2016).
17. Moore, A. *et al.* The geometric preference subtype in ASD: Identifying a consistent, early-emerging phenomenon through eye tracking. *Mol. Autism* **9**, 1–13. https://doi.org/10.1186/s13229-018-0202-z (2018).
18. Bovery, M. D. M. J., Dawson, G., Hashemi, J. & Sapiro, G. A scalable off-the-shelf framework for measuring patterns of attention in young children and its application in Autism Spectrum Disorder. *IEEE Trans. Affect. Comput.* https://doi.org/10.1109/TAFFC.2018.2890610 (2018).
19. Hong, M. P. *et al.* Differentiating social preference and social anxiety phenotypes in fragile X syndrome using an eye gaze analysis: A pilot study. *J. Neurodev. Disord.* **11**, 1–10. https://doi.org/10.1186/s11689-019-9262-4 (2019).
20. Bacon, E. C. *et al.* Identifying prognostic markers in autism spectrum disorder using eye tracking. *Autism* **24**, 658–669. https://doi.org/10.1177/1362361319878578 (2020).
21. Kaliukhovich, D. A. *et al.* Social attention to activities in children and adults with autism spectrum disorder: effects of context and age. *Mol. Autism* **11**, 1–14. https://doi.org/10.1186/s13229-020-00388-5 (2020).

22. Jones, W., Carr, K. & Klin, A. Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Arch. Gen. Psychiatry* **65**, 946–954. https://doi.org/10.1001/archpsyc.65.8.946 (2008).
23. Klin, A., Lin, D. J., Gorrindo, P., Ramsay, G. & Jones, W. Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature* **459**, 257–261. https://doi.org/10.1038/nature07868 (2009).
24. Shic, F., Bradshaw, J., Klin, A., Scassellati, B. & Chawarska, K. Limited activity monitoring in toddlers with autism spectrum disorder. *Brain Res.* **1380**, 246–254. https://doi.org/10.1016/j.brainres.2010.11.074 (2011).
25. Pierce, K., Conant, D., Hazin, R., Stoner, R. & Desmond, J. Preference for geometric patterns early in life as a risk factor for autism. *Arch. Gen. Psychiatry* **68**, 101–109. https://doi.org/10.1001/archgenpsychiatry.2010.113 (2011).
26. Elison, J. T., Sasson, N. J., Turner-Brown, L. M., Dichter, G. S. & Bodfish, J. W. Age trends in visual exploration of social and nonsocial information in children with autism. *Res. Autism Spectr. Disord.* **6**, 842–851. https://doi.org/10.1016/j.rasd.2011.11.005 (2012).
27. Chawarska, K., MacAri, S. & Shic, F. Context modulates attention to social scenes in toddlers with autism. *J. Child Psychol. Psychiatry* **53**, 903–913. https://doi.org/10.1111/j.1469-7610.2012.02538.x (2012).
28. Chawarska, K., Macari, S. & Shic, F. Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders. *Biol. Psychiatry* **74**, 195–203. https://doi.org/10.1016/j.biopsych.2012.11.022 (2013).
29. Jones, W. & Klin, A. Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature* **504**, 427–431. https://doi.org/10.1038/nature12715 (2013).
30. Chita-Tegmark, M. Social attention in ASD: A review and meta-analysis of eye-tracking studies. *Res. Dev. Disabil.* **48**, 79–93. https://doi.org/10.1016/j.ridd.2015.10.011 (2016).
31. Lombardo, M. V. *et al.* Default mode-visual network hypoconnectivity in an autism subtype with pronounced social visual engagement difficulties. *Elife* **8**, 1–22. https://doi.org/10.7554/eLife.47427 (2019).
32. Pierce, K. et al. Get SET early to identify and treatment refer Autism Spectrum Disorder at 1 year and discover factors that influence early diagnosis. J. Pediatr. (2021).
33. Pierce, K. *et al.* Detecting, studying, and treating autism early: The one-year well-baby check-up approach. *J. Pediatr.* **159**, 458-465.e456. https://doi.org/10.1016/j.jpeds.2011.02.036 (2011).
34. Chang, Z. *et al.* Computational methods to measure patterns of gaze in toddlers with autism spectrum disorder. *JAMA Pediatr.* https://doi.org/10.1001/jamapediatrics.2021.0530 (2021).
35. Vacas, J., Antolí, A., Sánchez-Raya, A., Pérez-Dueñas, C. & Cuadrado, F. Visual preference for social vs. non-social images in young children with autism spectrum disorders. An eye tracking study. *PLoS ONE* **16**, e0252795. https://doi.org/10.1371/journal.pone.0252795 (2021).
36. Buffle, P., Cavadini, T., Posada, A. & Gentaz, E. A study on visual preference for social stimuli in typical Ecuadorian preschoolers as a contribution to the identification of autism risk factors. *Sci. Rep.* https://doi.org/10.1038/s41598-021-87888-3 (2021).
37. Fujioka, T. *et al.* Developmental changes in attention to social information from childhood to adolescence in autism spectrum disorders: A comparative study. *Mol. Autism* **11**, 1–17. https://doi.org/10.1186/s13229-020-00321-w (2020).
38. Vargas-Cuentas, N. I. *et al.* Developing an eye-tracking algorithm as a potential tool for early diagnosis of autism spectrum disorder in children. *PLoS ONE* **12**, 1–13. https://doi.org/10.1371/journal.pone.0188826 (2017).
39. Jensen, K. *et al.* Autism detection in children by combined use of gaze preference and the M-CHAT-R in a resource-scarce setting. *J. Autism Dev. Disord.* **51**, 994–1006. https://doi.org/10.1007/s10803-021-04878-0 (2021).
40. Harrison, A. J. & Slane, M. M. Examining how types of object distractors distinctly compete for facial attention in Autism Spectrum Disorder using eye tracking. *J. Autism Dev. Disord.* **50**, 924–934. https://doi.org/10.1007/s10803-019-04315-3 (2020).
41. Wang, X., Chen, L., Liu, P., Polk, R. J. & Feng, T. Orientation to and processing of social stimuli under normal and competitive conditions in children with autism spectrum disorder. *Res. Autism Spectr. Disord.* **78**, 101614. https://doi.org/10.1016/j.rasd.2020.101614 (2020).
42. Shaffer, R. C. *et al.* Brief report: Diminished gaze preference for dynamic social interaction scenes in youth with autism spectrum disorders. *J. Autism Dev. Disord.* **47**, 506–513. https://doi.org/10.1007/s10803-016-2975-2 (2017).
43. Bradshaw, J. *et al.* The use of eye tracking as a biomarker of treatment outcome in a pilot randomized clinical trial for young children with autism. *Autism Res.* **12**, 779–793. https://doi.org/10.1002/aur.2093 (2019).
44. Shi, L. *et al.* Different visual preference patterns in response to simple and complex dynamic social stimuli in preschool-aged children with autism spectrum disorders. *PLoS ONE* **10**, 1–16. https://doi.org/10.1371/journal.pone.0122280 (2015).
45. Imafuku, M., Kawai, M., Niwa, F., Shinya, Y. & Myowa, M. Longitudinal assessment of social attention in preterm and term infants: Its relation to social communication and language outcome. *Infancy* **26**, 617–634. https://doi.org/10.1111/infa.12402 (2021).
46. Del Valle Rubido, M. *et al.* Exploring social biomarkers in high-functioning adults with autism and Asperger's versus healthy controls: A cross-sectional analysis. *J. Autism Dev. Disord.* **50**, 4412–4430. https://doi.org/10.1007/s10803-020-04493-5 (2020).
47. Hong, M. P. *et al.* Eye gaze and pupillary response in Angelman syndrome. *Res. Dev. Disabil.* **68**, 88–94. https://doi.org/10.1016/j.ridd.2017.06.011 (2017).
48. Le, J. *et al.* Oxytocin biases eye-gaze to dynamic and static social images and the eyes of fearful faces: Associations with trait autism. *Transl. Psychiatr.* https://doi.org/10.1038/s41398-020-0830-x (2020).
49. Theppeang, K. *et al.* Gender and race/ethnicity differences in lead dose biomarkers. *Am. J. Public Health* **98**, 1248–1255. https://doi.org/10.2105/AJPH.2007.118505 (2008).
50. Lew, J. *et al.* Sex-based differences in cardiometabolic biomarkers. *Circulation* **135**, 544–555. https://doi.org/10.1161/CIRCULATIONAHA.116.023005 (2017).
51. Billeci, L. *et al.* Disentangling the initiation from the response in joint attention: An eye-tracking study in toddlers with autism spectrum disorders. *Transl. Psychiatry* **6**, e808–e808. https://doi.org/10.1038/tp.2016.75 (2016).
52. Mattsson, N. CSF biomarkers and incipient alzheimer disease in patients with mild cognitive impairment. *JAMA* **302**, 385. https://doi.org/10.1001/jama.2009.1064 (2009).
53. Brueffer, C. *et al.* Clinical value of RNA sequencing–based classifiers for prediction of the five conventional breast cancer biomarkers: A report from the population-based multicenter Sweden Cancerome Analysis Network—Breast Initiative. *JCO Precis. Oncol.* https://doi.org/10.1200/po.17.00135 (2018).
54. Cui, Y. *et al.* Evaluation of prostate cancer antigen 3 for detecting prostate cancer: A systematic review and meta-analysis. *Sci. Rep.* **6**, 25776. https://doi.org/10.1038/srep25776 (2016).
55. Key, A. P., Jones, D. & Dykens, E. M. Social and emotional processing in Prader-Willi syndrome: Genetic subtype differences. *J. Neurodev. Disord.* **5**, 1–12. https://doi.org/10.1186/1866-1955-5-7 (2013).
56. Krol, K. M., Monakhov, M., Lai, P. S., Ebstein, R. P. & Grossmann, T. Genetic variation in CD38 and breastfeeding experience interact to impact infants' attention to social eye cues. *Proc. Natl. Acad. Sci. USA* **112**, E5434–E5442. https://doi.org/10.1073/pnas.1506352112 (2015).
57. Debladis, J. *et al.* Face processing and exploration of social signals in Prader–Willi syndrome: a genetic signature. *Orphanet J Rare Dis* **3**, 6–8 (2019).
58. Wang, L. *et al.* Heritability of reflexive social attention triggered by eye gaze and walking direction: Common and unique genetic underpinnings. *Psychol. Med.* **50**, 475–483. https://doi.org/10.1017/S003329171900031X (2020).

59. Constantino, J. N. *et al.* Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature* **547**, 340–344. https://doi.org/10.1038/nature22999 (2017).
60. Wetherby, A. M. & Prizant, B. M. *Communication and symbolic behavior scales: Developmental profile* (Paul H Brookes Publishing Co., 2002).
61. Mullen, E. M. *Mullen scales of early learning* (AGS, 1995).
62. Lord, C. *et al. Autism diagnostic observation schedule* (Western Psychological Services, 1995).
63. Sparrow, S., Balla, D. & Cicchetti, D. *Vineland Scales of Adaptive Behavior, survey form manual* (American Guidance Service, 1984).
64. Accuracy and precision test method for remote eye trackers. (2011).
65. R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria, 2020).
66. effsize: Efficient Effect Size Computation (2020).
67. Perez-Jaume, S., Skaltsa, K., Pallarès, N. & Carrasco, J. L. ThresholdROC: Optimum threshold estimation tools for continuous diagnostic tests in R. *J. Stat. Softw.* **82**, 1–21. https://doi.org/10.18637/jss.v082.i04 (2017).
68. Robin, X. T. N. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *Bioinformatics* **12**, 77. https://doi.org/10.1186/1471-2105-12-77 (2011).
69. irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84.1. (2019).
70. Wolak, M. E., Fairbairn, D. J. & Paulsen, Y. R. Guidelines for estimating repeatability. *Methods Ecol. Evol.* **3**, 129–137 (2012).
71. Langfelder, P. H. S. Fast R functions for robust correlations and hierarchical clustering. *J. Stat. Softw.* **46**, 1–17 (2012).
72. Langfelder, P. H. S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* https://doi.org/10.1186/1471-2105-9-559 (2008).
73. Wickham, H. F.R, Henry, L., & Müller, K. dplyr: A grammar of data manipulation. R package version 1.0.2. (2020).
74. tidyr: Tidy messy data. R package version 1.1.1.2 (2020).
75. Jung, Y. Multiple predicting K-fold cross-validation for model selection. *Journal of Nonparametric Statistics* **30**, 197–215. https://doi.org/10.1080/10485252.2017.1404598 (2018).
76. Park, J. J., Harari, O., Dron, L., Mills, E. J. & Thorlund, K. Effects of biomarker diagnostic accuracy on biomarker-guided phase 2 trials. *Contemp. Clin. Trials Commun.* **15**, 100396. https://doi.org/10.1016/j.conctc.2019.100396 (2019).
77. Courchesne, E., Yeung-Courchesne, R., Hesselink, J. R. & Jernigan, T. L. Hypoplasia of cerebellar vermal lobules VI and VII in Autism. *N. Engl. J. Med.* **318**, 1349–1354. https://doi.org/10.1056/nejm198805263182102 (1988).
78. Hultman, C. M., Sparén, P. & Cnattingius, S. Perinatal risk factors for infantile autism. *Epidemiology* **13**, 417–423 (2002).
79. Ji, Y., Raghavan, R. & Wang, X. Early life origins of ASD and ADHD. *Oxford Research Encyclopedias, Global Public Health* (2021).
80. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444. https://doi.org/10.1038/s41588-019-0344-8 (2019).
81. Baio, J. Prevalence of Autism Spectrum Disorder among children aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2010. *Surv. Summ.* **63**, 1–21 (2014).
82. Frazier, T. W. *et al.* Development and validation of objective and quantitative eye tracking–based measures of autism risk and symptom levels. *J. Am. Acad. Child Adolesc. Psychiatry* **57**, 858–866. https://doi.org/10.1016/j.jaac.2018.06.023 (2018).
83. Jeste, S. S. & Geschwind, D. H. Disentangling the heterogeneity of autism spectrum disorder through genetic findings. *Nat. Rev. Neurol.* **10**, 74–81. https://doi.org/10.1038/nrneurol.2013.278 (2014).
84. Chawner, S. J. R. A. *et al.* A genetics-first approach to dissecting the heterogeneity of autism: Phenotypic comparison of autism risk copy number variants. *Am. J. Psychiatry* **178**, 77–86. https://doi.org/10.1176/appi.ajp.2020.20010015 (2021).
85. Lombardo, V. M. *et al.* Different functional neural substrates for good and poor language outcome in Autism. *Neuron* **86**, 567–577. https://doi.org/10.1016/j.neuron.2015.03.023 (2015).
86. Kim, S. H., Macari, S., Koller, J. & Chawarska, K. Examining the phenotypic heterogeneity of early autism spectrum disorder: Subtypes and short-term outcomes. *J. Child Psychol. Psychiatry* **57**, 93–102. https://doi.org/10.1111/jcpp.12448 (2016).
87. Chen, H. *et al.* Parsing brain structural heterogeneity in males with autism spectrum disorder reveals distinct clinical subtypes. *Hum. Brain Mapp.* **40**, 628–637. https://doi.org/10.1002/hbm.24400 (2019).
88. Agelink Van Rentergem, J. A., Deserno, M. K. & Geurts, H. M. Validation strategies for subtypes in psychiatry: A systematic review of research on autism spectrum disorder. *Clin. Psychol. Rev.* **87**, 102033. https://doi.org/10.1016/j.cpr.2021.102033 (2021).
89. Xiao, Y. *et al.* Disrupted intrinsic connectivity links to early language and social deficits in toddlers with autism. *bioRxiv.* https://doi.org/10.1101/2021.10.08.463640 (2021).
90. Zhou, P., Ma, W. & Zhan, L. A deficit in using prosodic cues to understand communicative intentions by children with autism spectrum disorders: An eye-tracking study. *First Lang.* **40**, 41–63. https://doi.org/10.1177/0142723719885270 (2020).
91. Frazier, T. W. *et al.* Rapid eye-tracking evaluation oflanguage in children and adolescents referred for assessment of neurodevelopmental disorders. *J. Intellect. Dev. Disabil.* **45**, 222–235. https://doi.org/10.3109/13668250.2019.1698287 (2020).
92. Manyakov, N. V. *et al.* Visual exploration in Autism Spectrum Disorder: Exploring age differences and dynamic features using recurrence quantification analysis. *Autism. Res.* **11**, 1554–1566. https://doi.org/10.1002/aur.2021 (2018).
93. Gunderson, J. *et al.* Social and non-social sensory responsivity in toddlers at high-risk for autism spectrum disorder. *Autism Res.* **14**, 2143–2155. https://doi.org/10.1002/aur.2556 (2021).
94. Baranek, G. T. *et al.* Hyporesponsiveness to social and nonsocial sensory stimuli in children with autism, children with developmental delays, and typically developing children. *Dev. Psychopathol.* **25**, 307–320. https://doi.org/10.1017/s0954579412001071 (2013).
95. Dellapiazza, F. *et al.* Sensory processing related to attention in children with ASD, ADHD, or typical development: results from the ELENA cohort. *Eur. Child Adolesc. Psychiatry* **30**, 283–291. https://doi.org/10.1007/s00787-020-01516-5 (2021).
96. Baranek, G. T. *et al.* Cascading effects of attention disengagement and sensory seeking on social symptoms in a community sample of infants at-risk for a future diagnosis of autism spectrum disorder. *Dev. Cogn. Neurosci.* **29**, 30–40. https://doi.org/10.1016/j.dcn.2017.08.006 (2018).
97. Damiano-Goodwin, C. R. *et al.* Developmental sequelae and neurophysiologic substrates of sensory seeking in infant siblings of children with autism spectrum disorder. *Dev. Cogn. Neurosci.* **29**, 41–53. https://doi.org/10.1016/j.dcn.2017.08.005 (2018).
98. Goren, C. C., Sarty, M. & Wu, P. Y. K. Visual following and pattern discrimination of face like stimuli by newborn infants. *Pediatrics* **56**, 544–549 (1975).
99. Frank, M. C., Vul, E. & Saxe, R. Measuring the development of social attention using free-viewing. *Infancy* **17**, 355–375. https://doi.org/10.1111/j.1532-7078.2011.00086.x (2012).
100. Hunt, J. M. Attentional preference and experience: I. Introduction. *J. Genet. Psychol.* **117**, 131–146. https://doi.org/10.1080/00221325.1970.10533940 (1970).
101. Weizmann, F., Cohen, L. B. & Pratt, R. J. Novelty, familiarity, and the development of infant attention. *Dev. Psychol.* **4**, 149–154. https://doi.org/10.1037/h0030432 (1971).
102. Rose, S. A., Gottfried, A. W., Melloy-Carminar, P. & Bridger, W. H. Familiarity and novelty preferences in infant recognition memory: Implications for information processing. *Dev. Psychol.* **18**, 704–713. https://doi.org/10.1037/0012-1649.18.5.704 (1982).
103. Courchesne, E., Lincoln, A. J., Kilman, B. A. & Galambos, R. Event-related brain potential correlates of the processing of novel visual and auditory information in autism. *J. Autism Dev. Disord.* **15**, 55–76. https://doi.org/10.1007/BF01837899 (1985).

104. Klein, R. P. & Jennings, K. D. Responses to social and inanimate stimuli in early infancy. *J. Genet. Psychol.* **135**, 3–9. https://doi.org/10.1080/00221325.1979.10533411 (1979).
105. Huttenlocher, P. & Dabholkar, A. Regional differences in synaptogenesis in human cerebral cortex. *J. Comp. Neurol.* **387**, 167–178 (1997).
106. Mendel, G. Children's preferences for differing degrees of novelty. *Child Dev.* **36**, 453. https://doi.org/10.2307/1126468 (1965).
107. Morgan, K. & Hayne, H. Age-related changes in visual recognition memory during infancy and early childhood. *Dev. Psychobiol.* **53**, 157–165. https://doi.org/10.1002/dev.20503 (2011).
108. Rutherford, M. *et al.* Gender ratio in a clinical population sample, age of diagnosis and duration of assessment in children and adults with autism spectrum disorder. *Autism* **20**, 628–634. https://doi.org/10.1177/1362361315617879 (2016).
109. Wood-Downie, H. *et al.* Sex/gender differences in camouflaging in children and adolescents with autism. *J. Autism Dev. Disord.* **51**, 1353–1364. https://doi.org/10.1007/s10803-020-04615-z (2021).
110. Fombonne, E. & Zuckerman, K. E. Clinical profiles of black and white children referred for autism diagnosis. *J. Autism Dev. Disord.* https://doi.org/10.1007/s10803-021-05019-3 (2021).
111. Barbaro, J. & Freeman, N. C. Investigating gender differences in the early markers of Autism Spectrum Conditions (ASC) in infants and toddlers. Res. Autism Spectr. Disord. **83** (2021).
112. Yuan, J., Li, M. & Lu, K. Racial/ethnic disparities in the prevalence and trends of Autism Spectrum Disorder in US children and adolescents. *JAMA Netw. Open* **4**, e210771 (2021).
113. Jamison, R., Bishop, S. L., Huerta, M. & Halladay, A. K. The clinician perspective on sex differences in autism spectrum disorders. *Autism* **21**, 772–784. https://doi.org/10.1177/1362361316681481 (2017).
114. Dababnah, S., Shaia, W., Campion, K. & Nichols, H. "'We Had to Keep Pushing'": Caregivers' perspectives on autism screening and referral practices of black children in primary care. *Intellect. Dev. Disabil.* **56**, 321–336 (2018).
115. Lerer, E. *et al.* Low CD38 expression in lymphoblastoid cells and haplotypes are both associated with autism in a family-based study. *Autism Res.* **3**, 293–302. https://doi.org/10.1002/aur.156 (2010).

Acknowledgements

Author contributions

Funding

Competing interests

The GeoPref Test is free for research use but is licensed through UCSD for commercial use and is incorporated into pending patent (K.P. and J.Z.). No other authors have anything to declare.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-08102-6.

Correspondence and requests for materials should be addressed to T.H.W. or K.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.