Supplementary Information

Table of Contents

1. Supplementary Methods: Participants

1.a. Recruitment

Children were recruited with support from grants from the Simons Foundation and from the National Institute of Mental Health (R01 MH083727). Additional support was provided by the Marcus Foundation, the Whitehead Foundation, and the Georgia Research Alliance. Data collection occurred in the Autism Program of the Yale Child Study Center, New Haven, CT, and at the Marcus Autism Center, Children's Healthcare of Atlanta and Emory University School of Medicine, Atlanta, GA. The research protocol was approved as non-significant risk by Human Investigations Committees at the Yale University School of Medicine and at the Emory University School of Medicine, and the data collected were used for research purposes only with no relationship to clinical care. Families were free to withdraw from the study at any time.

The sample at High Risk for ASD (HR-ASD) was enrolled by identifying mothers who had a child with ASD and were also pregnant with another child. These mothers were recruited into the study via direct contacts with the following sources: through prior clinical evaluations of young children with ASD at the Yale Child Study Center, New Haven, CT, U.S.A.; via collaborations with local parent support groups; via large paediatric practices; and via general advertisement. The Low-Risk, Typical Development expected sample (LR-TDx) was created by recruiting pregnant mothers via collaborations with Yale – New Haven Hospital OB-GYN and Paediatrics departments (before or at the time of the child's birth, or subsequently via well-baby clinics), as well as by advertisements on pertinent websites and via direct mailing. For the external validation sample (main text, **Figure 3**), additional recruitment of LR-TDx children was

conducted at the Marcus Autism Center, Atlanta, GA, U.S.A., and via contact with Atlanta paediatric practices.

1.b. *Inclusionary and Exclusionary Criteria.*

Participating mothers and infants were included in this study after the mothers provided written informed consent for themselves, permission for their infants, as well as permission for their older children and/or permission to provide information on older children.

1.b.i. *High Risk (HR-ASD) sample*: High-risk status was ascertained by having an older sibling who met criteria for an ASD based on expert clinical diagnosis and at least one of two diagnostic instruments: the *Autism Diagnostic Interview – Revised* [1] and/or the *Autism Diagnostic Observation Schedule* [2]. Potential participants were excluded from the study if they displayed any one of the following: gestational age below 36 weeks; hearing loss or visual impairment determined at birth; non-febrile seizures; known medical conditions associated with autistic-like behaviours (e.g., Fragile X Syndrome, Tuberous Sclerosis); or any other identified genetic disorder. Infants requiring tube feeding or ventilation post-discharge were also excluded from participation. Ascertainment of exclusionary and inclusionary criteria included review of the mother's clinical file, completion of a medical history questionnaire, interview with caregivers, and a clinical examination of the baby by a paediatrician. Additional exclusionary criteria related to the child's ability to complete experimental procedures were applied once the child began to participate; these criteria are detailed below in Clinical Assessment Procedures, section 1.c.

1.b.ii *Low Risk (LR-TDx) Sample*: Children were enrolled in the Low Risk (LR-TDx) sample if there was no family history of ASD in first, second, or third degree relatives; nor developmental delays in first degree relatives; nor pre- or perinatal complications. As above, additional exclusionary criteria related to the child's ability to complete experimental procedures were applied once the child began to participate in procedures; these criteria are detailed below in Clinical Assessment Procedures, section 1.c.

1.c. *Clinical Assessment Procedures*

1.c.i. *Medical, Developmental, and Genetic Assessment of Child*. An attending paediatrician at Yale – New Haven Hospital completed a thorough baby check-up at 15 months, and well-child exams were reviewed for each child from birth throughout the first 2 years of life. The exam was used to assess the presence of identifiable medical conditions that might impact the child's development and to rule out sensorimotor difficulties that could compromise the child's participation in the experimental procedures. Otoacoustic emissions testing was used to assess hearing. The child's medical history was further reviewed regarding the course of birth and delivery. Attention was given to the course of delivery, presence of risk factors such as foetal brachycardia, low Apgar scores, newborn perinatal course, evidence of trauma, presence of dysmorphic features, skin findings, presence of seizures, primitive reflexes, motor abnormalities and asymmetries, as well as to prenatal exposure to harmful substances (e.g., valproic acid).

In addition, all infants at risk for ASD, contingent on parental consent, underwent paediatric and genetics assessments at 24 months, utilizing a template consistent with the format

and content of the exam used by the Autism Genetic Resource Exchange (AGRE) [3,4]. The evaluation ruled out known genetic and developmental syndromes that might be confused with autism. In addition to eliciting medical history and constructing a three-generation pedigree, the Genetics Counsellor sought to obtain all available relevant medical records and test results on the child participant. Upon parental consent, a blood sample was also obtained for genetic analyses.

Gestational age at birth was not significantly different between groups, $t_{34} = 0.08$, $P = 0.938$, with ASD mean(SD) = 38.7(1.2) weeks and TD = 38.7(1.7) weeks[5].

1.c.ii. *Direct Behavioural Assessment of Child*. *The Mullen Scales of Early Learning* [7] were administered at ages 6, 12, 18, 24, and 36 months to obtain standardized measures of cognitive functioning. *The Autism Diagnostic Observation Schedule*, modules T/1 and 2 (ADOS; [2,7]) were administered at ages 12, 18, 24 and 36 months (typically ADOS 1, or the Toddler Version, at ages 12, 18, and 24 months, and ADOS 1 *and* 2 at the age of 36 months).

1.c.iii. *Basic Visual Function*. This study measured how infants and toddlers watch social stimuli, and how their patterns of preferential looking might relate to level of social functioning, autistic symptomatology, and diagnostic outcome. As noted above, gestational age was not significantly different between groups, an important consideration in light of recent findings that the development of binocular vision is experience-dependent, and varies in relation to postnatal experience [8]. As a prerequisite for participation, prior to presentation of experimental stimuli, we tested each child's ability to shift and stabilize gaze. We included this procedure as a basic control against obvious symptoms of conditions affecting eye movement (e.g., conditions—such as nystagmus, Duane syndrome, or strabismus—that could adversely impact a child's ability to visually fixate video scenes of social content of the type used in this study). Children were shown a series of animated targets on an otherwise blank screen, and the elicited

behaviours (saccading to the target and maintaining fixation) were measured with eye-tracking equipment as a minimal check of eye movement function. Children passed the screen if they were able to saccade to the target and maintain stable foveation, defined as less than 5º/sec of drift in visual fixation [9]. The screen was conducted at each longitudinal visit, and with one exception (described below), all children passed.

These results confirmed prior research in ASD: while many studies in older children and adults have found differences in how individuals with ASD look at particular aspects of their surrounding environment [10-14], studies of eye movements in autism – that is, studies of the movements of the eyes themselves rather than of the content towards which the eyes are directed – have generally confirmed normal oculomotor function in children with autism in (a) maintaining steady fixation, as well as in velocity, duration, latency and accuracy of saccades [15]; in (b) rates of intrusive saccades [16]; in (c) vestibular-ocular reflex [17]; and in (d) in foveopetal ocular drift [16]. These studies suggest that the mechanics of oculomotor function appear to be generally intact in individuals with autism, and that differences in visual scanning are unlikely to arise from physiological aspects of eye movement [18]; and are instead more likely to arise from the way in which eye movements are deployed to specific content and within specific contexts [10].

In the current study, one child in the HR-ASD sample failed our eye movement screen. That child was identified as having congenital nystagmus and was immediately referred to a pediatric neurologist and ophthalmologist for further evaluation and follow-up care. Although the nystagmus prevented collection of point-of-gaze data, this child remained in the study and was followed until 36 months. We collected sample recordings of his eye movements (i.e. without point-of-gaze calibration) at each visit. At 24 months, and confirmed at 36 months, he was found to have no clinical ASD diagnosis.

1.c.iv. *Parent Interviews and Inventories*. A comprehensive questionnaire and inventory was administered covering aspects of prenatal and perinatal history, general health history, and treatment and intervention history (if any). Items pertaining to the prenatal and perinatal history of the baby were obtained at the 1st week and 3 month visits. Items pertaining to the overall health history of the baby were obtained at the 6, 12, 18, 24, and 36 month visits. Items pertaining to intervention history (if any) were obtained at 12, 18, 24, and 36 months. *The Vineland Adaptive Behaviour Scales – II* [19] were administered at 12, 18, 24, and 36 months to obtain standardized measures of adaptive function in the domains of communication, daily living skills, socialization, and motor skills. The *Autism Diagnostic Interview – Revised* (ADI-R) [1] was administered to the parents by a trained and experienced interviewer with established reliability with the training site at the age of 36 months.

1.c.v. *Reliability*: All diagnostic measures were administered by trained clinicians who were blind to experimental procedures and results. In addition, supervising experienced clinicians, all with post-doctoral expertise in the clinical assessment of children with ASD and related developmental disorders, observed all diagnostic procedures and co-coded diagnostic instruments for reliability checks for every 5th assessment throughout the protocol. Procedures were videotaped and archived for subsequent re-scoring, checking, and correction of possible drift during study duration.

1.d. *Participants: Assignment of Group Membership & Best Estimate Diagnostic Procedures*

Group membership of "ASD" or "non-ASD" for the N=59 HR-ASD children was carried out at the age of 24 months and was then confirmed at 36 months. No changes in group

membership were observed between 24 and 36 months. As noted, one of the N=51 LR-TDx children was flagged by research staff as a child with ASD-like concerns at 12 months of age, and confirmed with ASD outcome at 24 and 36 months. All diagnostic measures were administered by trained clinicians who were blind to experimental procedures and results. Parents were informed that clinicians were blind to participants' risk status, with a request to refrain from any discussion of the older sibling (clinical questions and concerns regarding the older sibling or the child in question were addressed by clinicians not involved in the experimental or diagnostic ascertainment protocol of the project). As noted above (Section 1.c.v.), procedures involving direct contact with children and families that required reliability maintenance were videotaped and archived for subsequent re-scoring, checking, and correction of possible drift during study duration. At least two supervising clinicians independently assigned overall clinical diagnosis on the basis of a review of all available data (ADI-R and ADOS results and protocols, videotaped or direct observation of ADOS, cognitive and communication assessments, history, and any other clinically-relevant data). Disagreements were discussed after data were entered for calculation of inter-rater reliability in order to obtain consensual clinician-assigned diagnosis (see [20,21] for additional details on best-estimate diagnostic procedures). A third experienced clinician reviewed all of the materials for the N=11 male children with ASD included in the main study, and also for the N=2 males with ASD included in the external validation sample. Diagnostic ascertainment at the age of 36 months was completed in the same fashion.

Assignment of group membership of "ASD" or "non-ASD" was carried out at 24 months and then ascertained at 36 months with the involvement of at least one experienced clinician not involved in the 24-month diagnostic procedures. A best estimate diagnostic procedure was

chosen as the gold standard for group membership (this choice was made in light of findings that indicate that experienced clinicians' judgment of children at the age of 24 months is a better predictor of later diagnosis than cut-off scores on the ADOS [20-23]. While ADOS scores for individual children may vary during the first 2 to 3 years of life, best-estimate clinician-assigned diagnosis shows much more stability, and, in our group, approaches 100% [21]. This is likely the result of the much broader frame of reference that is adopted during a best estimate diagnostic process, which includes the ADOS but also extends to other areas, specifically covering the following: historical developmental data; stability of traits in speech-language and communication symptoms (including communicative intent, voice and intonational quality of speech); results and profiles of standardized assessments and observations of speech-language-communication; and adequate weighting of low-frequency but highly-specific stereotypic behaviours (including repetitive behaviours, unusual attachments, and exceptionally restricted interests). Spurts in development and intensive intervention targeting speech-language and communication skills between 24 and 36 months of age can also impact the stability of specific scores, while the broader frame of reference taken by experienced clinicians will account for these factors.

For the analyses focused on phenotypic heterogeneity among the High Risk siblings, we divided the high-risk male infants who were not diagnosed with ASD at outcome into (1) those for whom there was never any clinical concern and whose typical development was ascertained at 24 and 36 months (HR-ASD_No-Dx), N=18; and (2) those for whom there were clinical concerns documented at any one of the clinical assessments. These concerns represented transient or subthreshold symptoms that did not meet criteria for ASD at 24- or 36-month evaluations. Because there are no consensual criteria for the diagnostic assignment of this

subthreshold category [24], also called "Broader Autism Phenotype" (BAP) [25], we followed currently adopted conventions as defined above, and as ascertained through the best-estimate diagnostic procedure [26]. N=10 male infant siblings met these criteria (HR-ASD_BAP).

Group membership of "TD" for the LR-TDx children was assigned at 24 months if there were no concerns of ASD and if children's developmental assessment scores on the *Mullen* did not show either two scores falling 1 SD below the mean or one score falling 1.5 SDs below the mean. At 33 months, the entire LR-TDx group also completed a *Vineland* in order to ascertain maintenance of TD status; any case for whom there was any developmental concern was then invited to complete a full clinical characterization protocol at the age of 36 months. All 25 males from the LR-TDx cohort were confirmed to have typical outcome.

1.e. *ASD, HR-non-ASD, and TD Clinical Characterization Data: Group Comparisons*

Clinical characterization data for the outcome comparisons between the N=11 ASD and N=28 HR-non-ASD male children, and between the N=11 ASD and the N=25 TD male children are provided here. As noted in the main text, from the original N=59 HR-ASD children, N=12 converted to a diagnosis of ASD at 24 months, confirmed again at 36 months: 10 males and 2 females. Because of the small number of females, they were excluded from current data analyses. Of the remaining N=47 children in the HR-ASD group, N=28 were males and N=19 were females. One male child from N=51 LR-TDx group showed concerning behaviour at 12 months and converted to a diagnosis of ASD by 24 months (and again confirmed at 36 months); that child was, therefore, included in the ASD group (N=11 in total). For comparison's sake, we also conducted analyses with that child excluded (described below in Section 2.c.v). As noted

above, the typically-developing status of the remaining N=50 LR-TDx children was assessed at 24 and was then confirmed again at 33 (and, if necessary, 36) months. Of these, N=25 were males and N=25 were females. The male TD children's data provided the normative benchmarks for the typical growth charts of social visual attention used in data analyses.

As diagnostic group membership was first assigned at 24 months, we provide here diagnostic (ADOS) and developmental (*Mullen* and *Vineland*) summaries at that age for the ASD group (N=11), all males, and for the group of HR-non-ASD (N=28), all males, from the HR-ASD risk-based cohort. Data comparisons are provided below.

	ASD Group [1]	**HR-non-ASD Group** [1]	t_{37} values	*p* values
N	11	28		
ADOS-SA [2]	**7.55** (4.46)	**3.93** (2.59)	3.169	0.003
ADOS-RRB [3]	**3.91** (1.7)	**1.96** (1.31)	3.817	<0.001
ADOS-Total [4,5]	**11.45** (5.06)	**5.89** (2.92)	4.306	<0.001
Mullen, NV AE [6]	**23.36** (6.20)	**25.46** (4.59)	-1.163	0.252
Mullen, RL AE [7]	**22.45** (7.59)	**24.50** (6.66)	-0.829	0.412
Mullen, ELV AE [8]	**22.18** (7.56)	**26.75** (6.26)	-1.932	0.061
Vineland, CommAE [9]	**19.73** (5.85)	**25.14** (5.68)	-2.657	0.012
Vineland, SocAE [10]	**16.18** (3.63)	**19.00** (2.19)	-2.978	0.005

[1] ASD Group = Autism Spectrum Disorders; HR-non-ASD = Non-Autism Spectrum Disorder Outcome from High Risk Group
[2] ADOS-SA: *Autism Diagnostic Observation Schedule*, Social Affect Cluster; higher scores on the ADOS indicate more severe autistic symptomatology;
[3] ADOS-RRB: *Autism Diagnostic Observation Schedule*, Restricted and Repetitive Behaviours Cluster; higher scores on the ADOS indicate more severe autistic symptomatology;
[4] ADOS-Total: *Autism Diagnostic Observation Schedule*, Sum of ADOS-SA + ADOS-RRB; higher scores on the ADOS indicate more severe autistic symptomatology.
[5] Autism Spectrum Cut-off = 8;
[6] Mullen, NV: *Mullen Early Scales of Learning*, Visual Reception (Nonverbal Function) Age Equivalent;
[7] Mullen, RL: *Mullen Early Scales of Learning*, Receptive Language Age Equivalent;
[8] Mullen, EL: *Mullen Early Scales of Learning*, Expressive Language Age Equivalent;
[9] Vineland, CommAE: *Vineland Adaptive Behaviour Scales*, Communication Domain Age Equivalent (in months);
[10] Vineland, SocAE: *Vineland Adaptive Behaviour Scales*, Socialization Domain Age Equivalent (in months).

The ASD and the HR-non-ASD groups differed significantly in levels of autistic symptomatology; as expected, the ASD group displayed higher levels of symptoms in the Social Affect (ADOS-SA) and Restricted & Repetitive Behaviours (ADOS-RRB) clusters as well as in the ADOS Total (ADOS-Total) scores. At 24 months, the ASD group had a mean ADOS-Total score of 11.45, exceeding by close to 3.5 points the ASD cut-off score of 8. The ASD and HR-non-ASD groups also differed significantly on levels of adaptive behaviours in the

Communication and Socialization domains, with the ASD group displaying significantly lower levels of abilities in these areas. The ASD and HR-non-ASD groups exhibited comparable levels of nonverbal and verbal function (although levels of Expressive Language skills trended in the direction of higher levels obtained for the HR-non-ASD group). These results confirm that the ASD group displayed higher levels of autistic symptomatology and lower levels of social and communicative adaptive skills while being comparable on levels of cognitive and receptive language skills. These results are consistent with expected differences between ASD and HR-non-ASD groups from an initial HR-ASD cohort (e.g., [26]).

For regression analyses (**Extended Data Figure 1**), we used 24-month outcome data to maximize comparability with previous cross-sectional and longitudinal work [11,20,21]. Correlation between decline in eye fixation and ADOS scores at either 24 or 36 months was not significantly different, $z = 0.86$, $P = 0.390$ (Fisher's r to z transform). Although we selected the ADOS-SA as our primary outcome measure, decline in eye fixation was also significantly associated with ADOS Total score, $r = -0.731$ [-0.23 - -0.93, 95% CI], $P = 0.01$. We also compared measures between the time of diagnostic membership assignment at 24 months and diagnostic confirmation at 36 months: a comparison of ADOS scores at each of the 2 testing times for the N=11 ASD children revealed no significant differences in group measures of ADOS-SA (mean = 7.55, SD = 4.45, and mean = 7.27, SD = 3.63, respectively, for 24 and 36 months); ADOS-RRB (mean = 3.91, SD = 1.70, and mean = 3.27, SD = 1.48, respectively, for 24 and 36 months); and ADOS-Total (mean = 11.45, SD = 5.06, and mean = 10.55, SD = 4.20, respectively, for 24 and 36 months). The comparison was between ADOS module 1 at 24 and at 36 months in order to ensure comparability of scores across the 2 ages. Although all but 3 children with ASD were able to complete the ADOS-2 (i.e., met the expressive language criteria for a meaningful

administration of this module), they were scored on both ADOS-1 and ADOS-2 because, as a rule, scores on a higher module are typically higher (i.e., indicating greater disability) because demands upon the child are higher [23].

For the comparison between the N=11 ASD and the N=25 TD children, at 6 months of age, there were no between-group significant differences in either nonverbal/cognitive or (pre)verbal/language skills [6], measured as age-equivalent scores in months, with Mean (SD) for TD = 5.8 (0.6) vs. ASD = 5.6 (0.9) for nonverbal/cognitive skills; and TD = 5.1 (0.6) vs. ASD = 4.7 (1.1) for receptive/expressive (pre)verbal/language skills. At 12 and 24 months, the two groups did not differ in nonverbal/cognitive skills ($P = 0.118$ and $P = 0.136$, respectively), but did differ in receptive/expressive language skills (consistent with the expected communication deficit in toddlers with autism), with means of 10.5 (2.0) for TD vs. 8.8 (2.2) for ASD at 12 months ($P = 0.032$); and 27.4 (4.4) for TD vs. 22.3 (7.3) for ASD at 24 months ($P = 0.036$).

Finally, for the analyses focused on phenotypic heterogeneity among the High Risk infant siblings, the HR-ASD_BAP and the HR-ASD_No-Dx groups differed significantly in levels of autistic symptomatology; as expected, the HR-ASD_BAP group displayed higher levels of symptoms in the ADOS Total (ADOS-Total) scores relative to the HR-ASD_No-Dx, with means of 9.7 (3.1) and 4.8 (2.4), respectively ($t(26) = 4.65$, $P < 0.001$). These results are consistent with expected differences between these groups (e.g., [26]).

1.f. *Necessary Trade-offs in Selection of Experimental Design*

1.f.i. *Sample Size*. Sample size was selected according to power calculations based on earlier work in toddlers with ASD[11], in which the effect size of ASD vs. TD between-group

differences in eye fixation was equal to $d = 1.56$ (Cohen's d). We expected greater variability in typical infant development and thus reduced our expectations of effect sizes and increased our sample allocation ratio. In order to observe cross-sectional results with "large" expected effect sizes (0.8 or greater), with standard α-level of 0.05, power equal to 0.8, and with sample allocation ratio equal to 4 (increased from 2.4 TD:ASD in [11]), we calculated that samples of N=12 children with ASD and N=50 TD children would be required. With expected sex ratios in ASD leading to at least 9 boys with ASD and 25 TD children, this sample size was also adequately powered to detect sex-specific differences (requiring slightly larger expected effect sizes, at $d = 1.0$, but still smaller than those observed in [11]). With an expected 20% conversion rate (conversion from HR-ASD to ASD, consistent with other studies of infant siblings[26-29]), confirmation of N=12 ASD children at outcome was expected to require an initial cohort of 60 HR-ASD infants.

In line with these estimates, we conducted analyses on the first consecutively enrolled cohort of N=59 HR-ASD siblings of children with ASD, and N=51 low-risk (LR-TDx) children. Analyses were conducted when 12 HR-ASD infants received confirmatory ASD diagnoses at 36 months. The total sample size of N=110 compares well with other prominent studies of infants at-risk for ASD (e.g.,[27]). We also increased statistical power by repeated sampling: these 110 children completed more than 960 visits and more than 7,893 trials of experimental eye-tracking procedures. Sex ratios in the ASD outcome group were approximately as expected (10:2 male to female), and, in addition, one LR-TDx infant also received an ASD diagnosis at 36 months. In total, N=11 male ASD children were confirmed at 36 months. Although this sample is still relatively small in absolute terms, effect sizes for between-group comparisons of longitudinal eyes, body, and object fixation (plotted in **Figure 2** in **Main Text**), with Cohen's d values

ranging from 1.18 to 1.26, indicate adequate power for detection of differences. In addition, this sample is the largest yet collected with data at the age of 2 months for children later diagnosed with ASD. Some other prominent studies[28,30] have included data collection in the first year of life, but more often only at 6 and/or 12 months of age. The current study complements those approaches by using a smaller overall sample size but more frequent sampling, with more than 400 successful data collection sessions completed by the time these infants reached the age of 6 months.

1.f.ii. *Female Infants with ASD*. We did not include data from females in the main analyses because we remain conservative in our appraisal of what conclusions could or should be drawn on the basis of 2 female participants. Given the almost 5:1 male:female ratio in autism spectrum disorders [31], the sample size problem for studying females with autism—especially when utilizing a risk-based infant siblings strategy—is daunting but not unexpected. In the future, we expect to remedy this problem and will ultimately chart the developmental profiles of female infants who are subsequently diagnosed with ASD; however, more infants will still need to be seen before a sample of sufficient size is amassed.

2. Supplementary Methods: Experimental Procedures

2.a. *Considerations in Selection of Experimental Design*

Our goal in this study was to test the extent to which performance-based measures of evolutionarily highly-conserved [32-46], and developmentally early-emerging [38,47-52] mechanisms of typical social adaptation may be disrupted in autism spectrum disorders (ASD), at a point prior to

the manifestation of overt symptoms of social disability. For this purpose, we adopted a design marked by 3 main features.

2.a.i. *Focus on Foundational Social Skills*. We chose to focus on disruptions of foundational social skills that are already online in typical babies from the first days and weeks of life [33]. This choice was made because several research groups have conducted studies of the natural course of autistic symptoms using observational and experimental methods, with no clear indicators of overt ASD symptomatology in the first year of life [53]. We chose to focus instead on whether normative mechanisms of social development were intact or disrupted in infants with ASD, and how those mechanisms were shaped in TD infants during early development. This approach follows the idea that between genetic liability and behavioural symptoms lies the disruption of highly conserved, normative mechanisms of socialization; the disruption of these mechanisms is not a symptom in and of itself, but rather a divergence in developmental course that will later give rise to symptoms [54-56].

2.a.ii. *Dense, Prospective Sampling of Early Infant Behavior*. The dizzying pace of social and communicative development in the first year of life [57], together with the corresponding brain specialization in that same timeframe, suggests that measures of infancy must keep pace with the accomplishments of infancy. To that end, in order to quantify atypical deviations from normative developmental trajectories, we opted for a high-density sampling design, with data collection occurring 5 times before the age of 6 months, a total of 7 times by the age of 12 months, and 10 times by the age of 24 months.

2.a.iii. *Longitudinal Growth Charts of Preferential Visual Attention to Conspecifics*. The intensive data collection allowed us to model, with sufficient statistical power in repeated measurements, "growth charts" of normative social visual attention with the hypothesis that

deviations therefrom would indicate a marker of unfolding ASD [54]. Like many other phenomena in nature [58], babies' deployment of social visual attention is highly variable; at single cross-sectional time points, or even at 2 or 3 time points, that variability will drastically weaken statistical power to detect meaningful developmental changes. However, dense repeated sampling can shed light on robust predictability of maturational patterns.

We addressed variability in individual data by using Functional Data Analysis (FDA) to generate growth curves, as FDA explicitly models statistical variation in both time scale and amplitude [59-68]. This approach greatly improved detection of common features in trajectory shape and individual deviations (in magnitude and timing) relative to normative data. We also repeated all analyses with traditional growth curve analysis using hierarchical linear modeling (HLM) [69].

2.b. *Data Acquisition*

2.b.i. *Equipment.* Eye-tracking in both infant and toddler labs (as described in Methods section) was accomplished by a video-based, dark pupil/corneal reflection technique with hardware and software created by ISCAN, Inc. (Woburn, MA, USA). The systems employ remotely mounted eye-tracking cameras with data collected at a rate of 60 Hz. We benchmarked the systems against another eye-tracker collecting data at 500 Hz (SensoMotoric Instruments GmbH, Teltow, Germany), in both infants and in toddlers, to ensure that the 60 Hz frequency was sufficient for reliably identifying on- and offset of saccades at a threshold velocity of 30° per second [9].

2.b.ii. *Calibration*. A five-point calibration scheme was used, presenting spinning and/or flashing points of light as well as cartoon animations, ranging in size from 1° to 1.5° of visual

angle, all with accompanying sounds. For the infants, calibration stimuli began as large targets (>= 10° in horizontal and vertical dimensions) which then shrank (via animation) to their final size of 1° to 1.5° of visual angle. The calibration routine was followed by verification of calibration in which more animations were presented at five on-screen locations. Throughout the remainder of the testing session, animated targets (as used in the calibration process) were shown between experimental videos to measure drift in calibration accuracy. In this way, accuracy of the eye-tracking data was verified before beginning experimental trials and was then repeatedly checked between video segments as the testing continued. In the case that drift exceeded 3°, data collection was stopped and the child was recalibrated before further videos were presented.

In this manner, we included data if the verification procedure indicated fixation locations no further than 3° from target center; in the majority of cases, as seen in **Extended Data Figure 8**, the accuracy was well within this limit. **Extended Data Figure 8** includes "worst case" fixations, because it includes fixations that initiated a halt of data collection and recalibration of the child; these measures are included to show the full range of accuracy testing.

We set the minimum allowable drift at 3.0° because the average eye region in our videos subtended 8.0° by 6.9° of participants' visual angle. By setting the minimum allowable drift to 3.0°, we assured that population variance in calibration accuracy would fall within 6.0°. The actual accuracy, as shown in the kernel density estimates in Extended Data Figure 9 is better than the worst case of a uniform distribution across a 6.0° region. As shown in the figure, the probability distribution of fixation locations relative to target was heavily weighted within the central 1-2° and the minimum discriminable ROI is smaller than the size of the target ROIs in all months. Even the mouth ROI, which subtends only 5.7° in the vertical direction, is discriminable with accuracy well above chance thresholds.

2.c. *Data Analysis*

2.c.i. *Performance of Task.* Given the young ages at which data were collected, and as a control for between-group differences in attention to task and completion of procedures, we tested for differences in duration of data collected per child (TD = 71.25(27.66) min, ASD = 64.16(30.77) min, $t_{34} = 0.685$, $P = 0.498$); and for differences in the distribution of ages at which successful data collection occurred ($k = .0759$, $P = 0.9556$; 2-sample Kolmogorov-Smirnov). Trials in which a child failed to fixate on the presentation screen for a minimum of 20% total trial duration were excluded from analyses. We tested for between-group differences in percentage of time spent saccading, blinking, or looking off-screen. And given our interest in results for the first 6 months alone as well as for the entire 24-month trajectory, we performed these analyses for both time periods.

As seen in **Extended Data Figure 3**, between months 2 and 6, there were no significant between-group differences in overall fixation time (**Extended Data Figure 3a-c**) (no main effect of diagnosis, $F_{(1,21.652)} = 0.958$, $P = 0.339$, nor interaction of diagnosis by age, $F_{(1,20.026)} = 0.880$, $P = 0.359$, by hierarchical linear modeling (HLM), model described below in Section 2.c.iii); nor in percentage of viewing time spent saccading (**Extended Data Figure 3d-f**) (no main effect of diagnosis, $F_{(1,27.189)} = 0.250$, $P = 0.621$, nor interaction of diagnosis by age, $F_{(1,26.430)} = 0.561$, $P = 0.460$, by hierarchical linear modeling). During the entire period of data collection (months 2, 3, 4, 5, 6, 9, 12, 15, 18, and 24), non-fixation data (saccades + blinks + off-screen fixations) were not significantly different between groups, with no main effect of diagnosis ($F_{(1,234.012)} = 2.701$, $P = 0.102$), and with no interaction of diagnosis by month ($F_{(1,1776.615)} = 3.447$, $P = 0.064$). In the

latter analysis, a trend level difference was observed, driven by increased off-screen fixation at month 24 in the ASD group.

2.c.ii. *Accuracy of Calibration.* Calibration accuracy was measured as the distance between a child's fixation location and the centre of the target location (for each target presented). Average calibration accuracy was less than 0.5 degrees of visual angle in the majority of all months (**Extended Data Figure 8a**), and in every month, the average calibration accuracy was less than 1.5 degrees of visual angle.

Calibration accuracy was not significantly different between groups cross-sectionally, at any data collection session (all $P > 0.15$, $t < 1.44$; mean $P = 0.428$; with comparisons conducted as independent samples *t* tests, at each month of data collection, *without* correction for multiple comparisons, so as to reduce the possibility of Type II error and be conservative in identifying between-group differences), nor longitudinally, as either a main effect of diagnosis ($F_{1,2968.336} = 0.202$, $P = 0.65$) or as an interaction of diagnosis by time ($F_{1,130.551} = 0.027$, $P = 0.87$). Longitudinal analyses of calibration accuracy were conducted by hierarchical linear modeling, and the relationship between calibration accuracy and age was modeled as an inverse function. The intercept and B terms were modeled as fixed effects but were allowed to vary by group. Degrees of freedom were calculated by the Satterthwaite method (equal variances not assumed).

2.c.iii. *Longitudinal Data Analyses*. As described in the Methods section, to examine the longitudinal development of social visual attention, both for individual participants and across both ASD and TD groups, we used Functional Data Analysis (FDA) [59] and Principal Analysis by Conditional Expectation (PACE) [60,61,67,68] (main text, **Figure 1D** and **1E**, for examples of individual results, **Figure 2** for group results, and **Extended Data Figure 7**). Although we focused on FDA/PACE in order to overcome limitations inherent to cross-sectional analyses, as

well as some potential limitations of traditional growth curve analyses, we repeated all our analyses using hierarchical linear modeling (**Extended Data Figures 2 - 6,** and **Extended Data Table 1a**).

Following the convention of Ramsay and Silverman[59], we also plotted the correlation surface functions for fixation data in each group; these are continuous estimates of the month-to-month correlations in looking patterns (i.e., a measure of the correlation between fixation at month 2 with fixation at month 3, with fixation at month 4, etc.) (due to space limitations in the Extended Data section, figures available upon request). For eye fixation, in infants later diagnosed with ASD, negative correlations are observed when comparing earliest months to later months, indicating transition from high to low eye fixation; a positive correlation surface emerges by months 5 and 6, indicating thereafter that levels of eye fixation remain low or decline further. In TD children, the correlation surface remains generally high and positive, with surface depressions coinciding with periods of behavioral transition (e.g., between 4-6 months as eye fixation increases, and again between 12-18 months as eye fixation declines to accommodate increasing mouth fixation). The between-group differences in these surfaces indicate differences in underlying developmental processes.

As a methodological comparison to Functional Data Analysis, we also analysed the data using hierarchical linear modeling [69]. The presence of linear and curvilinear (quadratic and cubic) patterns was assessed for Fixation relative to Age via the following model: $Fixation_{ij} = intercept_j + d_{ij} + B_{1j}\,(Age_{ij}) + B_{2j}\,(Age_{ij})^2 + B_{3j}\,(Age_{ij})^3 + e_{ij}$; where d_{ij} represents the normally distributed random effect modeling within-subject dependence by group; e_{ij} represents the normally distributed residual error; and the B_1, B_2, and B_3 coefficients indicate how fixation levels change with age and by group. Initial evaluation of the data indicated an inverse

relationship between body fixation and age, and was therefore also assessed with the following model: Body Fixation$_{ij}$ = d_i + intercept $_j$ + (B_{1j} /Age$_{ij}$) + e_{ij}. In all cases, the intercept and B terms were modeled as fixed effects but were allowed to vary by group. Degrees of freedom were calculated by the Satterthwaite method (equal variances not assumed). Positively skewed data (eg, body and object fixation trials) were log-transformed; plots show untransformed data. F tests and log-likelihood ratios were used to determine whether a linear, quadratic, cubic, or inverse relationship best described the data.

Growth curves from hierarchical linear modeling (HLM) are plotted in **Extended Data Figure 4, 5, 6**, and the regression parameters for Eyes, Mouth, Body, and Object are given in **Extended Data Table 1a**. Age-related changes in eye fixation were best characterized by a cubic relationship [$F(1,1870.709) = 12.576$, $P < 0.001$, with change in log likelihood (-2LL) indicating significantly improved fit for cubic relative to quadratic, $X^2(2) = 41.14$, $P < 0.01$]. Age-related changes in Mouth Fixation, however, were best characterized by a quadratic relationship [$F(1,1505.768) = 97.592$, $P < 0.001$, with change in log likelihood (-2LL) indicating significantly improved fit for quadratic relative to linear, $X^2(2) = 93.05$, $P < 0.001$, but no improvement for cubic relative to quadratic, $X^2(2) = 2.14$ $P > 0.05$]. Age-related changes in Body Fixation were best characterized by an inverse relationship [$F(1,20.613) = 14.551$, $P = 0.001$, with change in log likelihood (-2LL) indicating significantly improved fit relative to both quadratic, $X^2(2) = 47.298$, $P < 0.001$, and cubic, $X^2(4) = 16.464$ $P < 0.01$, functions]. Finally, age-related changes in Object Fixation were best characterized by a cubic relationship [$F(1,1790.273) = 11.206$, $P = 0.001$, with change in log likelihood (-2LL) indicating significantly improved fit relative to quadratic, $X^2(2) = 23.563$, $P < 0.01$].

Analyses revealed significant main effects of Diagnosis for Eyes, Mouth, Body, and Object Fixation [$F(1,146.416) = 28.82$, $P < 0.001$; $F(1,51.794) = 6.275$, $P = 0.015$; $F(1,24.141) = 5.50$, $P = 0.028$; and $F(1,240.460) = 10.84$, $P < 0.001$; respectively]; as well as significant Diagnosis x Age interactions for Eyes, Mouth; Body, and Object Fixation [$F(1,1870.709) = 12.58$, $P < 0.001$; $F(1,1505.768) = 13.103$, $P < 0.001$; $F(1,20.613) = 4.56$, $P = 0.045$; and $F(1,1790.273) = 11.21$, $P < 0.001$; respectively].

2.c.iv. *Early Looking Behavior Relative to Later Outcomes:* In order to explore the extent to which early looking behaviors related to a spectrum of affectedness, we measured looking behavior, from 2-6 months, in relation to diagnostic outcomes at 36 months (**Figure 3, main text**). To do so, we measured individual levels of eye fixation (**3a**) and rates-of-change (**3b**) in eye fixation. We calculated the mean change in eye fixation between 2 and 6 months, for each child, and created a receiver operating characteristic (ROC) curve to measure the overlap in distributions for affected children (infants who were later diagnosed with ASD) vs. unaffected children (TD infants) on the basis of mean rate-of-change in eye fixation (**3a-3c**) and body fixation (**3g-3i**). Because the ROC curves in **Figure 3c and 3i** are (necessarily) based upon data used to construct the model (and will thus give optimistically biased results), we also conducted an internal validation.

To conduct the internal validation (**3d-3f, 3j-3l**), we used leave-one-out cross-validation (LOOCV), partitioning our data set so that each infant in the cohort was tested as a validation case in relation to the remainder of the data set. For each infant in turn, we removed the diagnostic label (from outcome diagnosis), and then calculated the infant's eye fixation and rate of change in eye fixation through conditional expectation of each outcome possibility (explicitly testing the assumption of each child belonging to either the ASD or the TD groups as ascertained

at outcome). This process yielded two probabilities per child (the likelihoods, given a child's rate-of-change in eye or body fixation, of belonging to either of the outcome groups), and from these probabilities we calculated a single odds ratio. We computed bootstrap 95% confidence intervals for the fitted ROC curve.

2.c.v. *Tests for Influential Observations: Month-Two Data and Individual Children.* As in conventional statistical diagnostics for regression analyses, we conducted a series of tests to assess the impact of observations that might be overly influential on the data as a whole (i.e., outliers or observations with greater than expected leverage). **Extended Data Figure 9** compares longitudinal growth curves when month-two data are included or excluded. Exclusion of the month-two data does not significantly alter the trajectories for eyes, mouth, body, or object fixation (**Extended Data Figure 9a,9b**); nor does it alter the between-group comparisons thereof. We also conducted tests of the influence of month-two data on the relationship between eye fixation and outcome levels of symptom severity within the ASD group (**Extended Data Figure 9c**). When month 2 data are excluded, decline in eye fixation continues to significantly predict future outcome; this relationship reaches trend level significance by 2-9 months (P = 0.097), and is statistically significant thereafter (with r = -0.714 [-0.2 – -0.92, 95% CI], P = 0.014 for 2-12 months). Finally, we tested the influence of month-two data on results for individual children. While confidence intervals for the cross-validated ROC curves increase in size (as expected, in proportion to the reduction in data that arises by excluding month 2), the levels of overlap between-groups remain significantly different from chance, and are not significantly different from the curves calculated when the 2-month data are included (**Extended Data Figure 9d**).

We also assessed the impact of the one low-risk infant who received an ASD diagnosis at outcome. Inclusion or exclusion of that child's data did not significantly alter the trajectories for eyes, mouth, body, or object fixation; nor did it alter the clinical relationship to outcome levels of symptom severity; nor the extent of overlap in scores for children with ASD relative to TD outcomes on the basis of their looking patterns in the first 6 months of life (due to space limitations in the Extended Data section, figures available upon request).

We also assessed the impact of the one infant who later received a diagnosis of ASD and who exhibited the steepest decline in early eye fixation (visible in Figure 3A & 3B in the main text). Inclusion or exclusion of that child's data (due to space limitations in the Extended Data section, figures available upon request) did not significantly alter the trajectories for eyes, mouth, body, or object fixation; nor did it alter the clinical relationship to outcome levels of symptom severity; nor the extent of overlap in scores for children with ASD relative to TD outcomes on the basis of their looking patterns in the first 6 months of life.

3. References

1. Rutter, M., Le Couter, A., Lord, C. *ADI-R: Autism Diagnostic Interview-Revised*, Western Psychological Services, Los Angeles, CA (2003).

2. Lord, C., Rutter, M., DiLavore, P. C., Risi, S. *Autism Diagnostic Observation Schedule – WPS (ADOS-WPS)*, Western Psychological Services, Los Angeles, CA (2000).

3. Geschwind, D.H., et al. AGRE Steering Committee. *Am J Hum Genet*. **69**(2):463-6 (2001).

4. Lajonchere, C.J. AGRE Consortium. Changing the landscape of autism research: the autism genetic resource exchange. *Neuron*. **68**(2):187-91 (2010).

5. Abel, K.M., Dalman, C., Svensson, A.C., Susser, E., Dal, H., Idring, S., Webb, R.T., Rai, D., & Magnusson, C. Deviance in fetal growth and risk of autism spectrum disorder. *Am J Psychiatry.* **170**(4):391-8 (2013).

6. Mullen, E. *Mullen Scales of Early Learning. AGS Edition*, American Guidance Service, Inc., Circle Pins, MN (1995).

7. Lord, C., Rutter, M., DiLavore, P., Risi, S. *ADOS Toddler Module: ADOS-T: Manual*, Western Psychological Services, Los Angeles, CA (2008).

8. Jandó, G., et al. Early-onset binocularity in preterm infants reveals experience-depdendent visual development in humans. *PNAS.* **109**(27):11049-52 (2012).

9. Leigh, R.J. & Zee, D.S. *The Neurology of Eye Movements*. 3rd *ed,* Oxford University Press, Oxford, England (1999).

10. 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**(9):809-16 (2002).

11. Jones, W., Carr, K., Klin, A. Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-olds with autism. *Arch Gen Psychiatry*. **65**(8):946-54 (2008).

12. Rice, K., Moriuchi, J.M., Jones, W., Klin, A. Parsing heterogeneity in autism spectrum disorders: visual scanning of dynamic social scenes in school-aged children. *J Am Acad Child Adolesc Psychiatry*. **51**(3):238-48 (2012).

13. Riby, D.M. & Hancock, P.J. Viewing it differently: social scene perception in Williams syndrome and autism. *Neuropsychologia*. **46**(11):2855-60 (2008).

14. Norbury, C.F., et al. Eye-movement patterns are associated with communicative competence in autistic spectrum disorders. *J Child Psychol Psychiatry*. **50**(7):834-42 (2009).

15. Minshew, N.J., Luna, B., Sweeney, J.A. Oculomotor evidence for neocortical systems but not cerebellar dysfunction in autism. *Neurology*. **52**, 917-922 (1999).

16. Nowinski, C.V., Minshew, N.J., Luna ,B., Takarae, Y., Sweeney, J.A. Oculomotor studies of cerebellar function in autism. *Psychiatr Res*. **137**, 11-19 (2005).

17. Goldberg, M.C., Landa, R., Lasker, A., Cooper, L., Zee, D.S. Evidence of normal cerebellar control of the vestibule-ocular reflex (VOR) in children with high-functioning autism. *J Autism Dev Disord*. **30**(6):519-24 (2000).

18. Sweeney, J.A., Takarae, Y., Macmillan, C., Luna, B., Minshew, N.J. Eye movements in neurodevelopmental disorders. *Curr Opin Neurol*. **17**(1):37-42 (2004).

19. Sparrow, S. & Cicchetti, D. *Vineland Adaptive Behavior Scales II. The Interview Edition and Classroom Questionnaire*, American Guidance Service, Circle Pines, MN (2007).

20. Chawarska, K., Klin, A., Paul, R., Volkmar, F.R. Autism spectrum disorders in the second year: stability and change in syndrome expression. *J Child Psychol Psyc*. **48**(2):128-138 (2007).

21. Chawarska, K., Klin, A., Paul, R., Macari, S., Volkmar, F. A prospective study of toddlers with ASD: short-term diagnostic and cognitive outcomes. *J Child Psychol Psychiatry*. **50**(10):1235-1245 (2009).

22. Lord, C., et al. Autism from 2 to 9 years of age. *Arch Gen Psychiatry*. **63**(6):694-701 (2006).

23. Gotham, K., Risi, S., Pickles, A., Lord, C. The Autism Diagnostic Observation Schedule: revised algorithms for improved diagnostic validity. *J Autism Dev Disord.* **37**(4):613-27 (2007).

24. Zwaigenbaum, L., et al. Clinical assessment and management of toddlers with suspected autism spectrum disorder: insights from studies of high-risk infants. *Pediatrics.* **123**(5):1383-1391 (2009).

25. Steer, C.D., Golding, J., Bolton, P.F. Traits contributing to the autistic spectrum. *PLoS One.* 5(9):e12633 (2010).

26. Macari, S.L., Campbell, D., Gengoux, G.W., Saulnier, C.A., Klin, A., Chawarska, K. Predicting developmental status from 12 to 24 months in infants at risk for autism spectrum disorder: a preliminary report. *J Autism Dev Disord*. **42**(12):2636-47 (2012).

27. Tierney, A.L., Gabard-Durnam, L., Vogel-Farley, V., Tager-Flusberg, H., Nelson, C.A. Developmental trajectories of resting EEG power: an endophenotype of autism spectrum disorder. *PLoS One*. **7**(6):e39127 (2012).

28. Ozonoff, S., et al. A prospective study of the emergence of early behavioral signs of autism. *J Am Acad Child Adolesc Psychiatry*. **49**(3):256-66 (2010).

29. Ozonoff, S., et al. Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. *Pediatrics.* **128**(3):e488-95 (2011).

30. Newschaffer, C.J., et al. Infant siblings and the investigation of autism risk factors. *J Neurodev Disord*. **4**(1):7 (2012).

31. Centers for Disease Control and Prevention. Prevalence of autism spectrum disorders – Autism and Developmental Disabilities Monitoring Network, 14 Sites, United States, 2008. *Surveillance Summaries*. **61**(SS03):1-19 (2012).

32. Ghazanfar, A.A. & Santos, L.R. Primate brains in the wild: the sensory bases for social interactions. *Nat Rev Neurosci*. **5**(8):603-16 (2004).

33. Simion, F., Regolin, L., Bulf, H. A predisposition for biological motion in the newborn baby. *Proc Natl Acad Sci USA*. **105**(2):809-13 (2008).

34. Johnson, M.H. Biological motion: a perceptual life detector? *Curr Biol*. **16**(10):R376-7 (2006).

35. Regolin, L., Tommasi, L., Vallortigara, G. Visual perception of biological motion in newly hatched chicks as revealed by an imprinting procedure. *Anim Cognit*. **3**, 53-60 (2000).

36. Omori, E., & Watanabe, S. Discrimination of Johansson's stimuli in pigeons. *Intl J Comp Psychology*. **9**, 92 (1996).

37. Vallortigara, G., Regolin, L., Marconato, F. Visually inexperienced chicks exhibit spontaneous preference for biological motion patterns. *Plos Biol*. **3**(7):e208 (2005).

38. Rosa Salva, O., Farroni, T., Regolin, L., Vallortigara, G., Johnson, M.H. The evolution of social orienting: evidence from chicks (Gallus gallus) and human newborns. *PLoS One*. **6** (4):e18802 (2011).

39. Ferrari, P.F., Paukner, A., Ionica, C., Suomi, S.J. Reciprocal face-to-face communication between rhesus macaque mothers and their newborn infants. *Curr Biol*. **19**(20):1768-72 (2009).

40. Emery, N.J. The eyes have it: the neuroethology, function and evolution of social gaze. *Neurosci Biobehav Rev*. **24**, 581-604 (2000).

41. Myowa-Yamakoshi, M., Tomonaga, M., Tanaka, M., Matsuzawa, T. Preference for human direct gaze in infant chimpanzees (*Pan troglodytes*). *Cognition*. **89**, 53-64 (2003).

42. Bard, K.A. Primate parenting. In: Bornstein, M., ed. *Handbook of parenting: Vol. 2. Biology and ecology of parenting* (2nd ed., pp. 99-140). Erlbaum, Mahwah, NJ (2002).

43. van Lawick-Goodall, J. The behaviour of free-living chimpanzees of the Gombe Stream Nature Reserve. *Animal Behavior Monographs*. **1**, 161-311 (1968).

44. de Waal, F.X. Darwin's legacy and the study of primate visual communication. *Ann NY Acad Sci*. **1000**, 7-31 (2003).

45. Ehardt, C.L., & Blount, B.G. Mother-infant visual interaction in Japanese macaques. *Dev Psychobiol*. **17**, 391-405 (1984).

46. Kaplan, G., & Rogers, L.R. Patterns of gazing in orangutans (Pongo pygmaues). *Int J Primatol*. **23**, 501-526 (2002).

47. Fox, R., & McDaniel, C. The perception of biological motion by human infants. *Science*. **218**, 486-487 (1982).

48. Simion, F., Regolin, L., Bulf, H. A predisposition for biological motion in the newborn baby. *Proc Natl Acad Sci USA*. **105**(2):809-13 (2008).

49. Bardi, L., Regolin, L., Simion, F. Biological motion preference in humans at birth: role of dynamic and configural properties. *Developmental Sci*. **14**(2):353-9 (2011).

50. Farroni, T., Csibra, G., Simion, F., Johnson, M.H. Eye contact detection in humans from birth. *Proc Natl Acad Sci USA*. **99**(14):9602-9605 (2002).

51. Haith, M.M., Bergman, T., Moore, M.J. Eye contact and face scanning in early infancy. *Science*. **198**(4319):853-855 (1977).

52. Johnson, M.H. Subcortical face processing. *Nat Rev Neurosci*. **6**(10):766-774 (2005).

53. Jones, W. & Klin, A. Heterogeneity and homogeneity across the autism spectrum: the role of development. *J Am Acad Child Adolesc Psychiatry*. **48**(5):471-3 (2009).

54. Rogers, S.J. What are infant siblings teaching us about autism in infancy? *Autism Research*. **2**, 125-137 (2009).

55. Klin, A., Lin, D.J., Gorrindo, P., Ramsay, G., Jones, W. Two-year-olds with autism fail to orient towards human biological motion but attend instead to non-social, physical contingencies. *Nature.* **459**, 257-261 (2009).

56. Klin, A., Jones, W., Schultz, R.T., Volkmar, F.R., Cohen, D.J. The Enactive Mind – from actions to cognition: Lessons from autism. *Philos Trans R Soc Lond B Biol Sci*. **358**, 345-3 (2003).

57. Johnson, M. Functional brain development in humans. *Nat Rev Neurosci.* **2**, 475-483 (2001).

58. Thompson, D. *On growth and form*, Cambridge University Press, Cambridge, UK (1961/1917).

59. Ramsay, J.O. & Silverman, B.W. *Functional Data Analysis (2nd Edition),* Springer – Verlag (2006).

60. Yao, F., Müller, H.G., Wang, J.L. Functional data analysis for sparse longitudinal data. *J Am Stat Assoc*. **100**(470), 577-590 (2005).

61. Tang, R., Müller, H.G. Pairwise curve synchronization for functional data. *Biometrika*. **95**, 875-889 (2008).

62. Yao, F., Müller, H.G., Wang, J.L. Functional linear regression analysis for longitudinal data. *Ann Stat*. **33**, 2873-2903 (2005).

63. Müller, H.G. Functional modeling of longitudinal data. In: Fitzmaurice, G., Davidian, M., Verbeke, G., Molenberghs, G., eds. *Longitudinal Data Analysis (Handbooks of Modern Statistical Methods).* Wiley, New York, NY pp 223-252 (2009).

64. Müller, H.G., Stadtmüller, U. Generalized functional linear models. *Ann Stat.* **33**, 774-805 (2005).

65. Müller, H.G., Stadtmüller, U., Yao, F. Functional variance processes. *J Am Stat Assoc*. **101**, 1007-1018 (2006).

66. Leng, X., Müller, H.G. Classification using functional data analysis for temporal gene expression data. *Bioinformatics*. **22**, 68-76 (2006).

67. Liu B, Müller HG. Estimating derivatives for samples of sparsely observed functions, with application to online auction dynamics. *J Am Stat Assoc.* **104**(486):704-717 (2008).

68. Hall P, Müller HG, Yao F. Estimation of functional derivatives. *Ann Stat*. **37**, 3307-3329 (2009).

69. Singer, J.D., & Willett, J.B. *Applied Longitudinal Data Analysis*, Oxford University Press, New York, NY (2003).