Infant viewing of social scenes is under genetic control and atypical in autism

John N. Constantino[1,2,3], **Stefanie Kennon-McGill**[1], **Claire Weichselbaum**[1], **Natasha Marrus**[1,3], **Alyzeh Haider**[1], **Anne L. Glowinski**[1], **Scott Gillespie**[4], **Cheryl Klaiman**[5,6], **Ami Klin**[5,6,7], and **Warren Jones**[5,6,7]

[1]Department of Psychiatry, Washington University, St. Louis, MO

[2]Department of Pediatrics, Washington University, St. Louis, MO

[3]Intellectual and Developmental Disabilities Research Center, Washington University, St. Louis, MO

[4]Pediatric Biostatistics Core, Emory University School of Medicine, Atlanta, GA

[5]Marcus Autism Center, Children's Healthcare of Atlanta, Atlanta, GA

[6]Division of Autism & Related Disabilities, Department of Pediatrics, Emory University School of Medicine, Atlanta, GA

[7]Center for Translational Social Neuroscience, Emory University, Atlanta, GA

Abstract

Long before infants reach, crawl, or walk, they explore the world by looking: they look to learn and to engage[1], giving preferential attention to social stimuli including faces[2], face-like stimuli[3], and biological motion[4]. This capacity—social visual engagement—shapes typical infant development from birth[5] and is pathognomonically impaired in children affected by autism[6]. Here we show that variation in viewing of social scenes—including levels of preferential attention and the timing, direction, and targeting of individual eye movements—is strongly influenced by genetic factors, with effects directly traceable to the active seeking of social information[7]. In a series of eye-tracking experiments conducted with 338 toddlers—including 166 epidemiologically-ascertained twins, 88 non-twins with autism, and 84 singleton controls—we

find high monozygotic twin-twin concordance (0.91) and relatively low dizygotic concordance (0.35). Moreover, the measures that are most highly heritable, preferential attention to eye and mouth regions of the face, are also those that are differentially diminished in children with autism (X^2=64.03, P<0.0001). These results—which implicate social visual engagement as a neurodevelopmental endophenotype—not only for autism, but for population-wide variation in social-information-seeking[8]—reveal a means of human biological niche construction, with phenotypic differences emerging from the interaction of individual genotypes with early life experience[7].

Despite evidence that autism is among the most highly heritable of neuropsychiatric conditions[9], with a majority of genetic risk attributable to common (polygenic) factors[10,11], its neurobiological mechanisms remain unknown[12]. Autism is instead defined behaviorally, by atypical trajectories of social development[6] that can result in profound impairments in social-communicative function[13] and poor inclusion in wider societies that are often less-than-tolerant[14]. Atypical social visual engagement is observable within the first 6 months in infants later diagnosed with autism[6] and carries forward through later life[15,16].

In the present study, as an entry to understanding the genetic structure of factors affecting normative social development—factors that may be influenced by common genetic variation in the population at-large and are disrupted, at the extreme, in autism—we examined patterns of concordance in how children visually engage with (look at) caregivers[6] and peers[15] in social interaction (Extended Data Figure 1a; Methods). We examined pairwise concordance in social visual engagement as a function of zygosity, collecting eye-tracking data from 82 monozygotic twins (MZ, 41 pairs), 84 dizygotic twins (DZ, 42 pairs), and 84 non-siblings (42 randomized pairs). We established measurement stability over 15 months, and assessed the measures as an endophenotype for social disability, testing 88 toddlers with autism spectrum disorder (ASD) in comparison and replication cohorts (N=43, N=45).

In experiment 1, we measured macro-level indices of social visual engagement, calculating percentages of time spent looking at eye and mouth regions (Extended Data Figure 1b,c). In experiment 2, we measured micro-level indices, testing for concordance—on timescales of tens of milliseconds—in timing of eye movements, in direction of eye movements, and in the collocation of contemporaneous visual fixations. We also tested whether observed concordance could be partitioned into variation reflecting either stimulus response[17] (responding to specific features of the exact stimulus presented) or goal-directed action[18] (individual differences in seeking social information).

Trait heritability was estimated according to the classic twin design, with epidemiologically-ascertained samples of MZ and DZ twins, together with a sample of non-siblings. Non-siblings had no familial biological relationship to one another, lived apart, and were compared with twins in two ways: individually matched on sex and age (mean(SD) age difference: 0.99(0.27) days) and randomly matched in 10,000 re-samplings without replacement[19]. We restricted analyses to like-sex twin pairs (inclusion of opposite-sex DZ pairs yielded either no change or accentuation of MZ-DZ differences). For age at time of testing, we selected a dynamic period in typical development, 18–24 months of age

(mean(SD))=21.3(4.3) months), coinciding with large shifts in language, cognition, and adaptive behavior[20], and affording wide variation in our trait of interest[6] (see Methods).

We confirmed that groups did not differ significantly in age at time of testing ($F_{(2,247)}$=2.3, p=0.10; MZvs.DZ, t_{164}=1.59, p=0.11); participant demographics (Extended Data Table 1); calibration accuracy, oculomotor function (Extended Data Figure 2); or percentage of time spent in eye-looking, mouth-looking, or attention to task (Extended Data Figure 1d-f).

For concordance in eye- and mouth-looking (Figure 1d, 1i), MZ intraclass correlations (ICCs, case 2,1[21]) were remarkably high: 0.91 for eyes (95% CI: 0.85–0.95) and 0.86 for mouth (95% CI: 0.76–0.92). This contrasted markedly with DZ correlations: eyes, 0.35(0.07–0.59) and mouth, 0.44(0.16–0.65) (Figure 1c, 1h). In non-siblings, correlations did not differ significantly from zero: either when age- and sex-matched (Figure 1b, 1g) or when randomly-matched (Figure 1a, 1f, Extended Data Table 2a). In all groups, within-subject stability (test-retest reliability) was consistently high, indicating 'trait-like' stability (Extended Data Figure 3). These results are consistent with broad heritability of 0.86–0.90 for eye- and mouth-looking[22].

When seen for follow-up 15 months after initial testing, at 36.8(1.7) months (mean(SD)), MZ twins again demonstrated pairwise concordance in eye-looking of 0.93(0.75–0.98), while DZ concordance was 0.25(0.00–0.60) (Extended Data Figure 4a-4l and Table 2b, see Methods), indicating strong preservation of genetic influence on social visual engagement over development (Extended Data Figure 4m-4n). Moreover, longitudinal *within-subject stability*—from 21 until 36 months—was high in both groups: equal to ~0.70 (Extended Data Figure 5a-5e).

To test the specificity of these measures to social engagement, we compared concordance in eye-looking with concordance of two additional indices: time spent looking at non-social content (inanimate objects/background) and time spent attending to task (maintaining stable onscreen fixation[23]). In MZ twins, eye-looking was significantly more concordant than non-social object-looking—eyes, 0.91(0.85–0.95) vs. object, 0.66(0.46–0.80)—and more concordant than time spent attending to task, 0.46(0.19–0.67) (Figure 1n). In contrast, in DZ twins, eye-looking (0.35(0.07–0.59)) was not more concordant than either object-looking, 0.09(0.0–0.38), or attention to task, 0.34(0.05–0.58) (Figure 1c,h,m). Similarly, in age- and sex-matched non-siblings, all ICC estimates overlapped (Figure 1b,g,l). While heritable effects of domain general visual attention are likely observable in other contexts, these analyses indicate effects that are differentially social (Figure 1e,j,o).

In the next experiment, we measured moment-by-moment, micro-level concordance (Figures 2a,b). Macro-level concordance observed in the first experiment does not guarantee micro-level concordance; similarly, micro-level concordance could be present but present too weakly to yield global similarities. Comparison of the two creates an opportunity to test how genetic variation might influence social visual engagement across varying phenomenological timescales.

We first analyzed concordance in timing of eye movements (Figure 2c). Data in Supplementary Videos provide an immediately-appreciable sense of moment-to-moment

MZ concordance—weakened substantially in DZ twins—when viewing social scenes. Number and rate of eye movements did not differ significantly by group (~1944 fixations/child: mean(SD) rates of 1.66(0.59) fixations/sec DZ and 1.66(0.49) fixations/sec MZ; all $t<0.65$, $p>0.20$). However, MZ twins demonstrated greater probability of moving their eyes at the same times: for each movement by twin 1, within 350 milliseconds, there was an 18.6% increase in twin 2's probability of *also* making an eye movement (Figure 2d,e). More surprisingly, when analyses were restricted to moments of motor *initiation* of a saccade (Figure 2f), we observed a 21.1% increase in probability of time-locked eye movements: within +/−16.7 msec, MZ twins, but not DZ twins, initiated saccades at the same moments (Figure 2g,h). These results suggest that MZ toddlers, freely viewing naturalistic social stimuli, may synchronize not only the timing of overt eye movements, but also the activity of neuronal ensembles commonly associated with those movements: activity connecting areas of cortex to brainstem and cranial nerves[23,24], ultimately resulting in time-locked shifts of gaze[24].

We next tested for concordance in *direction* of eye movements (Figure 2i). We mined the eye movement data to identify contemporaneous collocated fixations: instances when both twins fixated on the same approximate locations at the same moments. In such cases, twins not only share an approximate fixation location, they share an approximate pattern of retinal irradiance (stimulation of retinal photoreceptors). By identifying these instances, we could then test the probability—given initially shared retinal stimulation—of twins subsequently moving their eyes in the same or different directions. We varied the criteria for collocation from within 1° (*i.e.,* shared stimulation of the rod-free, capillary-free foveola of the retina); to 1.7° (shared stimulation of rod-free fovea); to 5.2° (shared stimulation of whole fovea); to 10° (shared quadrant of visual information); to finally within 15° ('collocated' in only the broadest sense of looking at the same hemi-field of the presentation monitor). Across all comparisons (Figures 2l), MZ twins were more likely than DZ twins to shift saccades in more similar subsequent directions (see Methods).

Next, we compared twins' probability of fixating on the same social content at the same moments (Figure 2m). If twin 1 and twin 2 both looked at the eyes (or mouth) at the same time, this counted as a 'hit' for shared fixation; if twin 1 looked at the eyes when twin 2 looked at the mouth (or vice versa), this counted as a 'miss'. While both groups show more co-occurring, collocated fixations than chance (Figures 2n,o), MZ twins exhibited greater concordance than DZ ($F_{1,81}=4.89$, $p=0.030$; Figure 2p).

In summary, MZ twins exhibit strikingly high concordance in levels of eye-looking; greater probability of shifting their eyes at the same moments in time; greater probability of shifting their eyes in the same subsequent directions; and greater probability of contemporaneously fixating the same semantic content. These high levels of MZ concordance, observed at both macro- and micro-levels, indicate strong biological basis for variation in social visual engagement[7], with a substantial portion of that variation attributable to additive genetic influence. While concordance in micro-level measures was more modest than macro, even modest micro-level concordance marks repeatable shifts in probability: repetition of these shifts—recurring as frequently as every 400–500msec—suggests a striking means by which small probabilistic differences might amount, developmentally, to large eventual effects.

To further explore potential underlying biological mechanisms, we tested whether observed concordance could be partitioned into variation reflecting either stimulus response[17] or goal-directed action[18]. This distinction is intriguing because it relates to what aspects of social behavior may be more or less phylogenetically-conserved: have evolutionary pressures favored biological systems that rely on specific responses to particular features of external stimuli (in the manner of feature detectors[25]), or have evolutionary pressures favored systems that specialize in particular modes of seeking, internally driven with less direct dependence on the exact stimulus *per se*[26,27]? The related question in autism, when social development is disrupted, is whether to focus research on biological determinants related to processing particular social cues (afferent sensory systems) or to the seeking and adaptive usage of such cues (systems subserving social engagement and reciprocity, and the valuation of social stimuli).

To test this question, we conducted *post-hoc* comparisons capitalizing on two elements of the experimental protocol: because presentation order of video stimuli was randomized (with total duration longer than some toddlers' willingness to sit), each twin saw a separate set of videos, the majority of which were the same (M(SD)=86.4(19.3)%) but some of which were different (13.6(19.3)%), seen by only one among the pair. Moreover, each twin saw two different categories of video, one emphasizing dyadic mutual gaze (Extended Data Fig 1) and the other triadic peer interaction (Extended Data Fig 6).

We conducted three tests. For the first, analyses were restricted to measures made only when both twins watched the *same* videos; the null hypothesis held that concordance would be equal ($ICC_{sameVideos}=ICC_{allVideos}$), the alternative stated that concordance would be greater ($ICC_{sameVideos}>ICC_{allVideos}$). Greater concordance when watching the same videos would evidence stimulus response (more concordant responding given the exact same stimulus). For the second and third tests, analyses were restricted to measures made only when each twin watched *different* videos or *different content categories* of video; the null hypothesis held that concordance would be zero ($ICC_{differentVideos}=0$ and $ICC_{differentContent}=0$), the alternative stated that concordance would be greater than zero ($ICC_{differentVideos}>0$ and $ICC_{differentContent}>0$). Greater-than-zero concordance when each twin watched different videos or different content categories would evidence goal-directed action, less dependent on the exact stimulus (flexibly seeking social information as a form of active niche-picking[7]). For each test, we measured levels of eye-looking and quantified physical image properties of all eyes stimuli[28] (Extended Data Figure 7).

In the first test, we were unable to reject the null: concordance when watching the same videos was no greater than concordance when watching all videos (Figure 3b,g,l). In the second test, however, when each twin watched different videos, MZ and DZ concordance was significantly greater than zero (Figure 3c,h,m). And in the third test, when each twin watched different content categories, MZ concordance was significantly greater than zero but DZ concordance was not (Figure 3d,i,n). These results suggest that the relevant biological mechanisms more likely relate to systems subserving goal-directed seeking and valuation of social information.

Finally, to directly assess the functional significance of these measures, we compared the above results with data from two independent cohorts of toddlers with ASD (Figure 4): one primary comparison sample, N=43, and a second replication sample, N=45 (all consecutive referrals for diagnostic evaluation). In toddlers with ASD, the same measures of social visual engagement that were most highly heritable—eye- and mouth-looking—are markedly reduced (Figure 4a-c), providing a robust index of diagnostic membership (Figure 4d-f, AUCs=0.88(0.84–0.92) and 0.86(0.82–0.91)).

Taken as a whole, these findings lend insight into the means by which phenotypic differences emerge from the interaction between individual genotypes and individually-experienced environments, theorized decades ago as the means by which children 'make their own environments'[7] via developmental successions of reliable and repeated couplings between organism and environment[29]. Similar notions have been advanced in phenotypic studies contrasting the experiential development of children with autism and their typically-developing peers[30], yet never heretofore demonstrated as having directly traceable genetic influence. Inherent to the classic twin design is the fact that interactions between genetic and unmeasured environmental factors will be subsumed under the category of additive genetic influence. Although the twins' individual experiences of dynamic social stimuli were markedly influenced by genetic factors—reflecting a form of gene-environment *correlation* influencing their assimilation of standardized social scenes presented in the laboratory—it is likely that earlier life events already interacted with genetic variation, in development of both typical social visual engagement and autism susceptibility. Elucidating mechanisms by which genes interact with experienced (measured) environments is critical for future identification of preventive-intervention targets. The current findings underscore the notion that social visual engagement constitutes a neurodevelopmental endophenotype, not only for autism but for population-wide variation in goal-directed seeking and valuation of social information.

METHODS

This research was based in the Intellectual and Developmental Disabilities Research Center at Washington University and at the Marcus Autism Center, Children's Healthcare of Atlanta and Emory University School of Medicine. Study protocol was approved by the Washington University School of Medicine Human Research Protection Office (IRB), HRPO #201208010, and by the Emory University Institutional Review Board, IRB00048146. Parents of all participants gave informed consent prior to assessment. Children were shown video scenes of naturalistic caregiver and peer interaction. We measured percentage of visual fixation time to eyes, mouth, body, and object regions (Experiment 1 – macro-level measures of social visual engagement) as well as moment-by-moment variation in timing, direction, and location of eye movements (Experiment 2 – micro-level measures of social visual engagement). Visual scanning was measured with eye-tracking equipment (ISCAN, Inc., Woburn, MA). Analysis of eye movements and coding of fixation data were performed with software written in MATLAB. Data acquisition and processing were performed by experimenters blind to clinical assessment data (zygosity status, diagnosis etc.). Details of participants, experimental procedures, data acquisition, and analysis are provided below.

Participants

This study protocol was approved by the Washington University School of Medicine (WUSM) Human Research Protection Office (IRB), HRPO #201208010, and by the Emory University Institutional Review Board, IRB00048146. The parents of all participants gave informed consent prior to each assessment. A total of 414 children participated (242 twins, 84 non-sibling comparison children, and 88 children diagnosed with autism spectrum disorder (ASD)).

The twin sample was epidemiologically ascertained through the Missouri Family Register (MFR), a birth records registry maintained by the WUSM Department of Psychiatry in collaboration with the State of Missouri as described in detail in Marrus et al.[1]. Age- and sex-matched non-sibling comparison children (N=84 children, N=42 randomly-assigned pairs, age- and sex-matched within each pair) were recruited from the general population via flyers, direct mailings, and advertisement. Children with ASD were consecutive referrals to a diagnostic clinic (Marcus Autism Center), with experimental procedures collected at the time of each child's initial diagnosis (N=88 total from two independent cohorts of 43 and 45). The consenting family member was required to be the legal guardian and primary caregiver and to speak fluent English (given both the English language component of the video stimuli and as English is the sole spoken language in 93.9% of Missouri and 86.7% of Georgia households (quickfacts.census.gov/qfd/states/29000.html)).

Based upon Missouri Family Register data, we were able to contact 330 eligible families of Missouri twins in the specified age range during the calendar years 2011–2013. Of these, 180 enrolled in the Early Reciprocal Social Behavior study (the larger study, described in [31], of which the present experiments were a subcomponent). Of the 180, 121 (242 children) resided in close enough proximity to the St. Louis metropolitan area to be feasibly enrolled in the in-lab eye-tracking component of the study. For these 121 twin pairs, 242 individual eye-tracking data collection sessions were conducted. Sample demographics of the entire epidemiologically-ascertained sample (N=330) are given in Extended Data Table 1. The subset of twins participating in eye-tracking (121 pairs) was well-matched to the entire epidemiologically-ascertained twin group (the total 180 pairs enrolled). There were no group differences in sex, zygosity, race, or ethnicity. There was a difference in level of income, with the eye-tracking subset having slightly higher proportion than the general population of families in the highest wage-earning bracket (X^2=1.55, p=0.21).

Due to the paired nature of planned eye-tracking analyses (requiring complete sets of eye-tracking data from both twins), and due to the restriction of analyses to like-sex DZ twin pairs, the eye-tracking twin sample comprised 166 children (83 twin pairs, 41 MZ and 42 like-sex DZ). Demographics data for these children are given in Extended Data Table 1. Descriptions of eye-tracking data quality control and pairing procedures are given below in sections, 'Data Acquisition & Processing: Quality Control' and 'Data Acquisition & Processing: Pairing Procedures'. Mean age at time of testing in these 166 twins and in the non-sibling control samples (N=84) was 21.3 months (SD=4.26 months). By group, ages and sexes were as follows: non-sibling controls, mean(SD)=20.87(2.77) months, 52.4% male; DZ twins, 22.13(4.89) months, 52.4% male; and MZ twins, 20.94(4.74) months, 58.5% male.

For children with ASD, all eye-tracking data were collected at the time of initial diagnosis. Personnel blind to the diagnostic status of the children performed all aspects of eye-tracking data collection and analysis. Trained clinicians blind to results of all eye-tracking procedures administered all diagnostic measures. Children in each of the two ASD groups met the following inclusionary criteria: (1) they met criteria for Autistic Disorder or ASD on the *Autism Diagnostic Observation Schedule*[32], Module 1; and (2) they received a diagnosis of either Autistic Disorder (32 of 43 children in cohort 1), Pervasive Developmental Disorder-Not Otherwise Specified (11 of 43 children in cohort 1), or Autism Spectrum Disorder (45 of 45 children in Cohort 2) by two experienced clinicians upon independent review of all available clinical data, including standardized testing and video of the diagnostic examination. At time of testing of ASD cohort 1 (2011–2013, as in the twin sample), diagnostic guidelines followed DSM-IV-TR criteria[33]; all children would also meet criteria for ASD per current, DSM-5 criteria[13]. At time of testing of ASD cohort 2 (2015–2016), diagnostic guidelines followed DSM-5 criteria. Mean age at time of testing was 22.8(4.0) for ASD cohort 1 and 25.8(3.4) for cohort 2. Because the ASD cohorts were consecutive clinical referrals, age at time of testing depended on the age of referral. ASD cohorts were older than the epidemiologically-ascertained sample (t_{291}=2.78, P<0.001 and t_{293}=7.36, P<0.001 for cohorts 1 and 2, respectively). However, age was not significantly correlated with eye- or mouth-looking in either ASD cohort (r_{eyes}=0.01, p=0.95, r_{mouth}=-0.08, p=0.61 for cohort 1; r_{eyes}=-0.01, p=0.92, r_{mouth}=-0.12, p=0.41 for cohort 2), and constraining analyses to ASD subsamples age-matched to the twin and non-sibling typically-developing samples did not significantly change the area under the ROC curves in Figure 4 (ASD1$_{age-matched}$ AUC = 0.87(0.83–0.92) and ASD2$_{age-matched}$ AUC = 0.85(0.80–0.90)).

Zygosity Confirmation

Zygosity was determined by the Goldsmith Child Zygosity Questionnaire, which corresponds to DNA marker/blood type determinations of zygosity in 94.8% of cases[34]. The questionnaire was administered during a phone interview with the twins' biological mother or father. Correspondence between the questionnaire-based zygosity determination and genotypic assignment, using DNA acquired by buccal swab, was tested for a randomly selected subset of families (n=24 twin pairs) and in all cases positively confirmed the questionnaire results. In 6 twin pairs, zygosity could not be determined by questionnaire; data from those twins were excluded from the present analyses.

Data acquisition and data processing were performed by experimenters blind to the zygosity status of each twin pair. Measures of eye movement were made directly by videooculography for each child, collected in semi-automated fashion (by an experimenter using automated data collection software), and analyzed in fully automated fashion; aside from final group assignment, there were no components of data collection or analysis adjusted on the basis of zygosity. As a result, eye-tracking-based measures of social visual engagement benefit from the relative absence of rater/observational biases that have elsewhere been cited as a potential confound in twin studies (*e.g.*, [35]).

Experimental Procedures

Twins and non-sibling control participants were tested individually and accompanied at all times by a parent or primary caregiver. Eye-tracking data collection procedures matched those reported in [6] and [36]. Eye-tracking was accomplished by a video-based, dark pupil/corneal reflection technique with hardware and software created by ISCAN, Inc. (Woburn, MA, USA). The system was remotely-mounted within a wall panel beneath the stimuli presentation monitor, concealed from the child's view by an infrared filter.

Children were led into the testing room one at a time while an age-appropriate children's video was playing on the stimuli presentation monitor. Each twin was tested separately while the other was engaged in other assessments. Experimenters remained out of view behind a curtain, while the parent buckled the child into a car seat. The car seat was mounted on a pneumatic lift so that viewing height (aligned vertically to fall within the lower $1/3^{rd}$ of the stimuli presentation monitor) and distance from the monitor (approximately 28–30 inches; 71–76 cm) were standardized for all participants. The stimuli presentation monitor was a 20-in (50.8-cm) computer monitor with refresh rate of 60 Hz. Lights in the room were dimmed in order to direct attention toward the stimuli presentation monitor. Audio was played through a set of concealed speakers. The experimenter was able to observe the child at all times using a live video feed.

A five-point calibration method 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, on an otherwise blank screen, all with accompanying sounds. The calibration routine was followed by verification of calibration in which more calibration targets were presented at any of nine on-screen locations. Throughout the remainder of the testing session, calibration targets were shown between experimental videos to measure possible drift in accuracy. After calibration checks, the system was re-calibrated if excessive drift (>3° of visual angle) in calibration accuracy occurred. Please see *Quality Control* section below for measures of calibration accuracy and concordance thereof.

Stimuli

Following calibration and verification, 27 videos in randomized presentation order were shown to each child. Each video lasted an average of 44.2 sec, for a total viewing time of 19 min 54 sec. Videos comprised two content categories designed to recapitulate naturalistic social experience (as also described in [6] and [15]). The first category of video displayed an adult female actor who spoke directly to the viewer/camera, as shown in Extended Data Figure 1a, Extended Data Figure 7a, and Supplementary Videos 1 and 2, representing what would be experienced in dyadic interaction with a caregiver ('Dyadic Mutual Gaze', 15 videos in total). The actors were filmed in naturalistic settings that emulated the real-world environment of a child's room, with pictures and toys. The other category of videos consisted of children interacting in a daycare setting ('Triadic Peer Interaction', 12 videos total, shown in Extended Data Figure 7i and Supplementary Videos 3 and 4). The adult actors or the parent or legal guardian of the child actors provided written informed consent for filming and for publication of images (in Extended Data Figures 1, 6, and 7, and in

Supplementary Videos 1–4). The two content categories were randomly interleaved during presentation.

Levels of eye- and mouth-looking differ between content categories (see Extended Data Figure 1d-e for 'Dyadic Mutual Gaze' and Extended Data Figure 7d-e for 'Triadic Peer Interaction'). Where cross-category comparisons are made (Figure 3), normalization is required (described below in Data Analysis & Statistics: Macro-Level Indices of Social Visual Engagement). In all other analyses where levels of eye- and mouth-looking constitute the primary comparison, a single video stimuli content category was used (Figures 1, 3, Extended Data Figures 1, 3, 4, & 5, dyadic; Figure 4, Extended Data Figures 6 & 7, triadic). Other analyses (micro-level measures in Figure 2, controls in Extended Data Figure 2) require no normalization and summarize results for all stimuli.

Videos were presented as full-screen audiovisual stimuli; in 32-bit color; at 640×480 pixels in resolution; at 30 frames per sec; with mono-channel audio sampled at 44.1 kHz. Stimuli were sound and luminosity equalized, and were piloted in an independent sample of children before the start of study in order to optimize engagement for typical infant and toddler viewers.

Data Acquisition and Processing

Analysis of eye movements and coding of fixation data were performed with software written in MATLAB (MathWorks). The first phase of analysis was an automated identification of non-fixation data, comprising blinks, saccades and fixations directed away from the presentation screen. Saccades were identified by eye velocity using a threshold of $30°$ per sec[23]. We tested the velocity threshold with the 60-Hz eye-tracking system described above and, separately, with an eye-tracking system collecting data at 500Hz (SensoMotoric Instruments GmbH). In both cases saccades were identified with equivalent reliability as compared with both hand coding of the raw eye-position data and with high-speed video of the child's eyes. Blinks were identified as described in [15]. Off-screen fixations (when a participant looked away from the video) were identified by gaze vectors directed to locations beyond the stimuli presentation monitor.

Eye movements identified as fixations were coded into four regions of interest that were defined within each frame of all video stimuli: eyes, mouth, body (neck, shoulders and contours around eyes and mouth, such as hair) and objects (surrounding inanimate stimuli) (Extended Data Figure 1b,c and Extended Data Figure 6b,c). The regions of interest were hand traced for all frames of the video and stored as binary bitmaps. Automated coding of fixation time to each region of interest then consisted of a numerical comparison of each child's coordinate fixation location data with the bitmapped regions of interest. Extended Data Tables 3c and 3d give the percentage of time spent fixating on each region of interest as well as the corresponding time (in minutes) spent fixating. Average total duration of included video trials per child was M = 18.2 min (SD = 3.1min) for MZ twins and M=17.9 min (SD = 3.4 min) for DZ twins.

Quality Control—In 10 of the 242 twin data collection sessions (4.13%), data could not be collected due to the following: child fussiness, child sleep, and/or temporary equipment

failure. In 16 of 242 twin data collection sessions (6.61%), data were collected but checks of calibration accuracy either could not be performed or, when performed, indicated sufficiently low quality data (*i.e.*, calibration accuracy in excess of +/−3°) that the data should not be used for analyses. Determination of quality was performed at time of data collection, independently from further analyses, and by separate staff from those who conducted primary analyses (in addition, to ensure that exclusion of these data did not introduce bias, analyses were repeated with the 16 sessions of low quality data *included*; although the inclusion introduced additional measurement error, there was no statistically significant change in overall results). In the remaining 216 of 242 sessions (89.26%), data were successfully collected. The 26 sessions with missing values (10 by failure-to-collect and 16 by low-quality-collection) spanned 24 cases with values missing for one twin and 2 cases (1 twin pair) with values missing for both twins.

Average calibration accuracy for all groups was less than 1° of visual angle. Extended Data Figure 2a-c shows total variance in calibration accuracy, and Extended Data Figure 2d-f shows average calibration accuracy. Pairwise concordance in calibration accuracy was measured as both fixation position and distance to address the possibility that systematic saccadic overshoots or undershoots to particular locations might be more concordant in MZ versus DZ twins. Position was measured as horizontal and vertical fixation location relative to the center of the calibration accuracy validation target (in degrees of visual angle) while distance was measured from fixation location to center of the calibration accuracy validation target (also in degrees of visual angle). Concordance in calibration accuracy fixation position was not statistically different from 0 in either MZ, DZ, or non-sibling controls for fixation position (Extended Data Figure 2g-j). Concordance in calibration accuracy fixation position also did not differ significantly between groups, calculated by intraclass correlation coefficient[21,37] (ICC, case (2,1)): ICC_{MZ}=0.07(0.00–0.29), ICC_{DZ}=0.00(0.00–0.20), $ICC_{non-sib}$=0.01(0.00–0.24). Results were also non-significant for distance: ICC_{DZdist}=0.00(0.00–0.22), ICC_{MZdist}=0.15(0.00,0.45), $ICC_{non-sib\_dist}$=0.00(0.00–0.26).

It is theoretically possible that eye movement accuracy could be more concordant in MZ vs. DZ twins (*i.e.*, ballistic muscle movements of the eyes might be incrementally more similar in MZ than DZ). MZ group results, although not approaching statistical significance, exhibit a slight numerical increase in ICC value. However, given the size of this effect in the current experimental testing framework, power analyses indicate that approximately 1600 pairs of MZ twins would be required to reject or confirm its existence (80% power, α=0.05). More importantly, the magnitude of such an effect in the present context, if it existed, would be substantially smaller in size than that of the semantic content regions in our stimuli: stated differently, the relative increase or decrease in concordance in accuracy would, based on current measures, operate in a range of tenths of degrees of visual angle, whereas the size of our semantic target regions is 40–80-fold greater in size (summary of size of regions of interest in the stimuli is given in Extended Data Table 3a). Such an effect could not, in and of itself, account for the large differences in MZ vs. DZ concordance in looking to semantic content regions.

To ensure best practices for consistent data collection, each eye-tracking session was also qualitatively rated at time of collection by in-lab staff on a scale from 0–5, using a scoring

system in which staff were trained and checked for reliability. The score was based on quality of the eye image throughout the session, amount of measurement error during each calibration check, and perceived degree of overall child engagement during testing. The ratings were used to ensure best practices for consistent data collection. Sessions in which data could not be collected (the 10 of 242 mentioned above) were given scores of 0; poor quality sessions (the 16 of 242 noted above) were given scores of 1. As noted, analyses were repeated with and without the 16 sessions of low quality data included, to no change in results.

For each experimental trial (each video stimulus), we used a minimum-valid-data criterion of fixation time greater than or equal to 20% of total trial duration. The criterion was established on the basis of prior analyses of an independent sample of eye-tracking data (207 children, aged 16.5–30 months; with threshold identified in that sample by analyzing the full set of fixation percentages as two separable components of a finite mixture model).

In the present data set for MZ and DZ twins, 4,232 video trials were presented; application of the exclusion criterion excluded 4.44% of collected trials (188 videos), leaving 4,044 included video trials. For number of video trials included, excluded, or presented, there were no significant differences between MZ and DZ twins ($t_{164} = 1.25$, p=0.21; $t_{164} = 0.17$, p=0.87; $t_{164} = 1.31$, p=0.19, for included, excluded, and presented, respectively; data were tested with both a two-sample t test as well as a Wilcoxon signed rank test / Mann-Whitney U test with comparable results [no significant differences in any analysis]). We set no threshold for minimum number of trials sufficient for inclusion of a child's data in final analyses; if usable data were collected, with trials fixated at a level greater than or equal to the minimum-valid criterion described above, the child's data were included.

Of 27 possible video trials, the mean number of included trials for MZ twins was 24.8 (3.7) and for DZ twins was 24.0 (4.7) (data given as mean(SD)). The mode number of included trials per child for both MZ and DZ groups was 27; likewise, the median number of included video trials for children in both groups was 26, while the minimum number of included trials collected for any single participant was 8 for one MZ twin and 4 for one DZ twin.

Pairing of Participant Data—Due to the paired nature of planned analyses, only twin pairs having complete eye-tracking data sets from *both* twins could be analyzed. Of the 121 total twin pairs enrolled in the eye-tracking study, 96 pairs (192 children) had complete data sets. Twenty-five twin pairs had missing data (24 pairs missing one twin's data, 1 pair missing both). In 10 of the 25 twin pairs with incomplete data, an additional eye-tracking session was scheduled and conducted, re-testing both twins; we repeated all analyses with and without these data (*i.e.*, either constraining analyses to the first attempted testing session [constrained to the 96 twin pairs succeeding in collection on first visit] or including results from the next data collection session in which data were successfully collected for both twins); in either case, twin pair data were always collected on the same day for each twin. With and without these data included, there was no statistically significant change in overall results. Of the 15 twin pairs with insufficient data, the proportion of twins included/excluded did not differ between MZ and DZ twins: 5/15 (33.3%) were MZ (2f:3m) and 5/15 (33.3%)

were like-sex DZ (1f:4m); the remaining 5/15 were opposite-sex DZ and, like all opposite-sex DZ, were not a part of the analyses.

For analyses of pairwise concordance in eye-tracking measures, the final set of children with successfully collected, paired data consisted of N=41 MZ twin pairs (82 children), N=42 DZ like-sex twin pairs (84 children), and N=42 age- and sex-matched non-sibling comparison children (84 children). Analyses were also conducted for the DZ opposite sex twin pairs (N=17 pair, 34 children); inclusion of these children *increased* the MZ – DZ differences in all cases (*i.e.*, DZ concordance was reduced by inclusion of opposite sex pairs); in light of those results, consistent with other published studies, and in order to be conservative in our estimates of concordance and heritability, we constrained present analyses to DZ like-sex twins.

For age- and sex-matched non-sibling comparison, paired children had no biological relationship to one another; were matched on sex; and were matched to within 1 day of mean chronological age (average difference in age: mean(SD) = 0.99 (0.27) days; average absolute difference in age: mean(SD) = 4.8 (0.22) days). Our rationale for including this age- and sex-matched, non-sibling control group was specifically to include an overt comparison for effects of age and sex on measures of social visual engagement. Our previous work using this same experimental paradigm[6] shows evidence that these behaviors, as with many others that emerge in early development (*e.g.*, walking and talking), undergo progressive changes that are very sensitive to a child's developmental stage and may be sensitive to sex effects (*e.g.*, precociousness in female verbal abilities[38]). Because it is therefore possible that genetically unrelated individuals could show a degree of concordance based on similarities in developmental stage or sex, we included this comparison sample.

Finally, we also included a comparison with fully randomized pairings of the non-sibling controls: these analyses were conducted irrespective of age and sex, across 10,000 randomized pairings without replacement, to calculate concordance estimates and confidence intervals across all permutations[19]. The weak—but non-zero—ICC values in age- and sex-matched non-sibling controls (main text, Figure 1b, 1g, 1l) do appear to indicate that a portion of concordance may be due to developmental effects independent of direct biological familial relationship. We are under-powered to confirm or reject such an effect; however, the graded pattern of results across all 4 groups—from fully randomized pairings (Figure 1a, 1f, 1k), to age- and sex-matched non-sib controls, to DZ twins, to MZ twins—suggests that such an effect may exist.

Data Analysis and Statistics

As noted above in the Data Acquisition and Processing section, eye movements identified as fixations were coded into four regions of interest defined within each frame of all video stimuli: eyes, mouth, body (neck, shoulders and contours around eyes and mouth, such as hair) and objects (surrounding inanimate stimuli) (Extended Data Figure 1b,c). Supplementary Videos 1–4 show examples of coded data.

Code Availability—Analysis of eye movements and coding of fixation data were performed with software written in MATLAB (MathWorks) either via the commandline or in scripts, available upon request.

Macro-Level Indices of Social Visual Engagement—In Experiment 1, we measured macro-level indices of social visual engagement, calculating proportion of time spent looking at eyes, mouth, and body regions. Percentage of total time spent attending to video stimuli (Extended Data Figure 1f), as well as time spent fixating specifically on eyes (Extended Data Figure 1d), mouth (Extended Data Figure 1e), body, and nonsocial object regions was calculated. The twin-twin and paired non-sibling concordance plots for proportion of time looking at each region were constructed (main text Figure 1), and intraclass coefficients calculated. Because negative values of the intraclass correlation coefficient (ICC) only arise when estimates of the variance components are negative or zero —which is mathematically possible but not theoretically meaningful[39,40]—all reported ICC values fall within the range [0,1].

As observed in previous work[6], the 18–24-month developmental period in which testing was conducted corresponds with large changes in typical infant eye- and mouth-looking, with amount of mouth-looking in typical infants rising to a peak value at approximately 18 months of age (when single word vocabulary is also rapidly increasing). Given floor and ceiling effects noted in the distributions of eye- and mouth-looking, respectively, analyses of concordance were also repeated with non-parametric measures, with no appreciable difference in results: we compared correlations for MZ and DZ twin pairs using both Spearman's rank correlation and intraclass correlation coefficient (ICC). Non-parametric results were as follows: 1) eyes: MZ, $\rho = 0.843$ ($P<0.001$) and DZ, $\rho = 0.333$ ($p=0.031$); and mouth: MZ, $\rho = 0.822$ ($P<0.001$) and DZ, $\rho = 0.405$ ($p=0.008$).

Likewise, macro-level measures of eye- and mouth-looking differ, as expected, by video content category (Extended Data Figure 1d-f and Extended Data Figure 6d-f). For this reason, analysis of concordance in levels of looking across different content categories, as undertaken in Figure 3, parts d, i, and n, requires normalization (measures of Pearson correlation would, of course, by unaffected by such differences, but measures of agreement and consistency, as is the case for the intraclass correlation coefficients, are affected by differences in scale). To analyze measures of concordance on a common scale, data were normalized by linear transformation as follows: for each set of measured levels of eye-looking in dyadic mutual gaze stimuli and triadic peer interaction stimuli (the X and Y axes of Figure 3 parts d,i,n), the minimum value was identified and the range was calculated; the minimum value was subtracted from each individual value and then each value was multiplied by the range, resulting in values scaled from 0 to 100 (comparable results were found by using a Z score transformation, but because the data were not normally distributed, we used this non-parametric alternative).

In addition, as described in the main text, to test the specificity of the measures to *social* engagement, we compared concordance in eye- and mouth-looking with concordance of time spent looking at nonsocial content (inanimate object and background regions), and time spent attending to task (maintaining stable onscreen fixation with less than 5°/sec of eye

movement[23]). Interestingly, in MZ twins, eye-looking (ICC: 0.91, 95% CI: 0.85–0.95) was significantly more concordant than nonsocial object-looking (ICC: 0.66, 95% CI: 0.46–0.80) and more concordant than time spent maintaining steady fixation (0.46 (0.19–0.67)); mouth-looking, by contrast, (mouth ICC: 0.86, 95% CI: (0.76–0.92) was more concordant in MZ twins than time spent maintaining steady fixation but was *not* more concordant than time spent looking at nonsocial content. This difference is consistent with other studies emphasizing the distinct evolutionary and functional role of the eyes in social interaction[41].

Measures of Trait-like Stability—To measure the extent of trait-like stability of these behaviors, we measured within-subject stability / test-retest reliability across both short and long timescales. For short timescales, results are plotted in Extended Data Figure 3. Within-subject stability is strong, irrespective of group membership, and within-subject stability results present a striking contrast to the twin-twin concordance results which vary by degree of genetic relatedness (plotted below each respective panel for comparison). These measures reflect trait-like stability for any given individual during *single-day* testing sessions, quantified by intraclass correlation coefficients with a 2-way random effects model, ICC(2,1)). Another related measure, not plotted, is that of *inter*-individual variation—the reliability of measured differences between any individuals A and B (*i.e.,* the stability with which the measured trait is higher/lower in individual A than individual B, individual C, etc., given a series of repeated measures). In that case, the observed ICC values are on the order of 0.9 for each group, quantified in that case by a fixed rather than random effects model, ICC(3,k). Both measures are strong evidence that the levels of looking across individuals are highly reliable.

Regarding the question of stability over longer timescales, we invited back as many participants as possible for follow-up at the age of 36 months. We were able to collect and analyze data for N=22 MZ twins (11 pairs, age at Time 1, mean(SD) = 21.1(2.6) months, age at Time 2, 36.9(2.6) months) and for N=44 DZ twins (22 pairs, age at Time 1, mean(SD) = 22.1(2.5) months, age at Time 2, 36.8(1.0) months; ages for combined groups, Time 1 = 21.7(2.6), Time 2 = 36.8(1.7).

We analyzed these data in three ways: (1) twin-twin concordance of measures at Time 2 alone (Extended Data Figure 4a-l); (2) within-subject stability from Time 1 until Time 2 (Extended Data Figure 5a-e); and (3) twin-twin concordance from Time 1 until Time 2 (Extended Data Figure 5f-j).

For the first comparison (twin-twin concordance of measures at Time 2 alone), results show robust MZ twin-twin concordance at Time 2 alone relative to diminished DZ twin-twin concordance (similar to results observed at Time 1 alone). Results are plotted in Extended Data Figure 4a-l and given in Extended Data Table 2b.

For the second comparison (within-subject stability *from* Time 1 *until* Time 2), results show comparable within-subject stability over time for both groups, irrespective of zygosity. Results are plotted in the top row of Extended Data Figure 5, parts a-e. These results indicate that within-subject stability of eye-looking from 21 until 36 months is very high in both groups: 0.72 for MZ twins (95% CI: 0.44–0.87) and 0.69 for DZ twins (95% CI: 0.50–0.82).

Also, as expected for within-subject stability, the two groups (with 95% confidence intervals that fully overlap mean estimates for both groups) do not differ significantly in this regard.

Finally, for the third comparison (*twin-twin concordance* from Time 1 until Time 2, bottom row of Extended Data Figure 5, parts f-j and Extended Data Table 2c), the results differ starkly as a function of zygosity: concordance of twin 1's eye-looking at 21 months with Twin 2's eye-looking at 36 months for MZ twins is 0.70 (0.40–0.86), whereas for DZ twins the twin-twin concordance is 0.22 (0.00–0.49); for mouth looking, the difference is 0.73 (0.45–0.88) for MZ and 0.07 (0.00–0.36) for DZ.

Taken as a whole, these analyses of within-subject stability and twin-twin concordance strongly support the notion that social visual engagement exhibits heritable trait-like characteristics during this period of early childhood: there is substantial within-subject stability across all participant groups in marked contrast to differences in twin-twin concordance varying by zygosity; MZ twin-twin concordance is preserved over 15 months of time and substantially contrasts with DZ twin-twin correlations at both 21 months and 36 months; and within-subject stability is extremely strong when examined on both short and long timescales.

Physical Image Properties of Eye Regions—To address the question of whether observed concordance could be partitioned into variation reflecting stimulus response[17] (responding to specific features of the exact stimulus presented) or goal-directed action[18,26] (individual differences in the seeking of social information, able to be dissociated from an exact stimulus), we measured concordance in eye-looking across varying conditions in which twins watched either the same or different stimuli (as described in the main text and presented in Figure 3).

To quantify differences in the physical image properties of stimuli seen by each twin, we analyzed image property profiles of regions demarcated as eyes across all frames of all videos presented. Specifically, we analyzed the lightness and color (color opponency in red-green and blue-yellow following the CIE 1976 model [42]), contrast (RMS, root-mean-squared contrast), orientation gradients (sum of local maxima of image intensity gradient), and amount of motion (sum of change in image intensity) present within all eye regions in all videos[28,43]. Image property profiles for representative videos are plotted in Extended Data Figure 7. Variation in stimulus image properties can be seen in the histograms themselves (parts c-h and m-r) as well as in the statistical comparisons of video image property profiles (parts i-j and s-t, compared by two-sample Kolmogorov-Smirnov tests). These data underscore the notion that "eyes" are a semantic content category rather than a singular stimulus image property[28], a notion consistent with research distinguishing stimulus-driven or "bottom-up"[43–47] processes in visual saliency from those that are goal-directed or "top-down"[26,27,48–51].

Analyses in Figure 3 and Extended Data Figure 7 show that concordance in eye-looking is strongly preserved in MZ twins despite watching different stimuli: in MZ twins, the extent to which twin 1 looks at the eyes in Dyadic Mutual Gaze videos (examples can be found in Supplementary Videos 1 and 2) is highly concordant with the extent to which twin 2 looks at

the eyes in scenes of unscripted peer interaction (Triadic Peer Interaction videos; examples can be found in Supplementary Videos 3 and 4): ICC = 0.81 (95% CI: 0.67–0.89). This effect persists despite the fact that eyes found in the triadic peer interaction videos differ substantially in lightness, color, contrast, orientation gradients, and motion. These eyes are ½ to ¼ the size of eyes found in the dyadic mutual gaze videos (see Extended Data Table 3); they do not engage the viewer in mutual gaze, and they are instead frequently encountered in partial occlusion or profile and frequently present multiple onscreen targets (for each of multiple onscreen characters) rather than a single eye region. Notably, when DZ twins are presented with these different content categories, concordance in their levels of eye-looking no longer differ significantly from 0: ICC = 0.12 (95% CI: 0.00–0.41). These analyses are not meant to suggest that concordance in social visual engagement is stimulus-*independent*, necessarily, there are consistencies across the stimuli presented in the current study and there are limits to the extent of reasonable differences in stimuli that comparisons of the current type would allow (*e.g.*, there would be no expectation that measures of social visual engagement should remain consistent across entirely non-social stimuli). Instead, we take the present analyses as indication that what is heritable does not appear to be a response to a particular physical feature *per se* (*i.e.*, response to a single feature found within a highly uncertain visual world); rather, the evidence indicates that what is conserved is an adaptive action: behavioral seeking (in goal-directed fashion) to engage with relevant social stimuli in the environment[26,27,52] (seeking to engage with stimuli that can exist in a variety of different forms and features). This notion is consistent with basic evolutionary theory (aligning with survival impulses that drive adaptive action), particularly for primate species seeking to survive in highly social environments[53].

Micro-Level Indices—In MZ and DZ twins, we collected 322,672 fixational eye movements (DZ: 161,963; MZ: 160,709; ~1944 fixations per child), occurring at a rate of 1.66 fixations per second (DZ mean(SD) = 1.66(0.59) fixation/sec; MZ mean(SD) = 1.66(0.49) fixation/sec), each lasting an average of 514 milliseconds (DZ = 523(188) msec; MZ = 505(245) msec). As a function of zygosity, there were no significant between-group differences in fixation count, frequency, or duration (tested by 2-sample *t* test, all $p > 0.594$, all $t_{(164)} < 0.534$). Saccadic amplitude data are given in Extended Data Figure 2k-m. Summary statistics regarding saccadic eye movements are limited to instances in which saccades begin and end with within-range, measurable fixations. (In cases in which saccadic eye movements either originate from or result in offscreen/out-of-range fixation locations, or cases in which saccades co-occur with blinks, accurate measurements of saccade amplitude, duration, and velocity are not available and were thus excluded.) We analyzed 133,582 saccadic eye movements (DZ: 68,262; MZ: 65,320; ~804 saccades analyzed per child), with no significant between-group difference in quantity as a function of zygosity: $t_{(164)} = 0.649$, $p = 0.517$.

Timing of Eye Movements—In Experiment 2, we measured concordance in the timing of individual eye movements, testing whether probability of making a saccade was significantly modulated as a function of zygosity. Specifically, we analyzed the time series eye movement data in terms of timing of saccades and timing of saccade initiation using peristimulus (or "peri-event") time histograms (PSTHs, [54]).

Following methods detailed in [15], PSTHs were constructed by aligning each twin pair's individual time series eye movement data to the start of each video stimulus, and by then computing counts of co-occurring saccades in 33.3 msec bins in a surrounding 1333.3 msec window. Bin counts were computed for each twin pair and then averaged across all pairs to obtain group means (plotted in main test Figures 2d,e and 2g,h).

To test whether observed changes in saccade probability differed from those expected by chance, we used permutation testing[19,55]. In each of 1000 iterations, the binary time-series saccade data for each twin (0 = not saccading, 1 = saccading) were permuted by circular shifting[56], following the equation:

$$s_{j,c}(t) = s_j(t - r_j, \text{modulo} T)$$

written as

$$s_{j,c}(t) = s(\langle t - r_j \rangle_T)$$

which, for $r_j >= 0$, equals

$$s_{j,c}(t) = \begin{cases} s_j[t - r_j], & r_j < t \leq T \\ s_j[T - r_j + t], & 0 \leq t \leq r_j \end{cases}$$

where s_j is the measured saccade time-series data for each participant, j; $s_{j,c}$ is the circular-shifted saccade time-series data for the same participant, j; t is a time point in the time series defined over the interval $0 \quad t \quad T$; T is the total duration of the stimulus (in the present case, the duration of a video shown to participants); and r_j is the size of the random circular shift, in the same units of time as t, for each participant, j. Size of the random circular shift for each participant was drawn independently from a random number generator with uniform distribution with possible values ranging from $-T$ to T.

PSTHs were then computed on each of those permuted data sets. By this method, durations of saccades and inter-saccade intervals were preserved for each individual, but the timing of each saccade was made random in relation to the actual timing of the other twin's saccades. The mean instantaneous probability of making a saccade, during each bin, across all 1000 PSTHs from permuted data, quantified the results one would observe if saccade probability were random between twins. If, on the other hand, the timing of one twin's saccades was synchronized with his or her twin sibling, and not random (*i.e.*, if when twin 1 made a saccade, twin 2 exhibited a greater probability of making a saccade), one would expect to see significant deviations from the permuted data distribution. The 2.5th and 97.5th percentiles of instantaneous saccade probability across all PSTHs from permuted data served as a $p = 0.05$ confidence level against which to compare saccade rates in the actual data (two-tailed comparison).

Taken as a whole, this approach enabled the comparison of actual patterns of saccading to randomized, chance patterns, and also allowed us to test the null hypothesis that DZ or MZ

twins demonstrated no greater than chance levels of time-locking of eye movements. Results in main text Figures 2c-h show significant time-locking in MZ twins, to within +/−16.67 milliseconds of saccade initiation. This level of concordance suggests an impressive set of related biological implications. Specifically, this degree of time-locking of eye movements would not be possible without time-locked contractions of rectus and oblique extraocular muscles[23]. Cranial nerves III, IV, and VI supply these muscles[23], whose afferent connections are in turn supplied by the reticular formation in brainstem[57,58]. Synapsing directly upon the reticular formation are projections from the frontal eye fields[59]. With so few synapses separating frontal eye fields from the extraocular muscles[24], spontaneous time-locking of eye movements suggests the likely presence of some, even modest, degree of time-locked neural activity in stages prior to motor movement initiation. Given the present behavioral results, it is intriguing to speculate on the extent of possible concordance in activity of neural systems that play a role in saccadic eye movements[24] (frontal eye fields, supplementary eye fields, parietal eye fields, Area 22, DLPFC).

Direction of Eye Movements—In Experiment 2, we measured twin-twin concordance in direction of eye movements. Saccade direction was computed as an angle (θ), in degrees. Difference in saccade direction was measured as the difference, in degrees, between the angles of twin 1 and twin 2's saccades: θ_{twin1} - θ_{twin2}. Polar histograms of twin-twin differences are plotted in main text Figures 2j,k. As noted in the main text, the analysis began by identifying instances of data in which both twins fixated on the same approximate locations at the same moments in time. Necessarily, these analyses involved selection of thresholds (*i.e.*, analytic definitions of what would constitute the "same" approximate location as well as the "same" moment in time). To assure that any observed differences were not merely the result of selecting one threshold versus another, we conducted analyses across varying thresholds of contemporaneous timing (temporal windows of 66.7msec, 133msec, 250msec, 500msec) and degree of collocation (retinal eccentricities of 1°, 1.7°, 5.2°, 10°, 15°). Main text Figures 2j & k plot results for saccades starting from fixations collocated within 5.2° (at least partially overlapping foveas) and co-occurring within 500msec or less. Main text Figure 2l plots results across varying degrees of collocation, also co-occurring within 500msec or less. Across all comparisons of varying retinal eccentricities and temporal windows, MZ twins were more likely than DZ twins to shift saccades in more similar subsequent directions.

As in the preceding analyses of timing of eye movements, we compared observed differences in twin-twin saccade direction to results expected by chance by means of permutation testing. For permuted analyses, within each twin pair, twin-twin pairings of saccades starting at common locations were randomly shuffled in 1000 iterations, computing the angular difference across all randomly paired saccades in all iterations. The polar histogram data plotted as gray bars in Figures 2j & k shows the upper 95th percentile of differences expected by chance alone across all 1000 permutations (with the upper 95th percentile serving as a $p = 0.05$ confidence level against which to compare the actual observed differences). By comparison, the 50th percentile of permuted data would have less skew. The 95th percentile established the upper limit of similarity in saccade direction expected by chance. Skew seen in the chance distribution (with the histogram shifted

towards more versus less similar saccade directions), is likely due to the nature of the video content and the effect of that content on probable saccade direction (*i.e.*, video stimuli presented content that was, in general, centrally-framed; as a result, saccades are more probably made in specific directions towards or away from that content). While both MZ and DZ twins show an increase in the probability of moving their eyes in a shared direction, also summarized in Figure 2l, MZ twins exhibit greater probability of shifting saccades in more similar subsequent directions.

Collocation of Contemporaneous Fixations—Finally, we measured concordance in the collocation of contemporaneous visual fixations with respect to semantic content regions (eyes and mouth). We compared the twins' probability of fixating on each of these regions at the same moments by creating 2x2 contingency tables of co-occurring fixations (main text Figure 2m). When twin 1 and twin 2 looked at the eyes (or mouth) at the same moments, this counted as a 'hit' for shared fixation; if twin 1 looked at the eyes when twin 2 looked at the mouth (or vice versa), this counted as a 'miss'. The counts of collocated fixations to eyes or mouth thus depend on the exact timing of when these fixations occurred; to compare the observed counts to those expected by chance, we again used permutation testing (permuting the observed sequences of fixations by circular shifting in each of 1000 iterations). Observed counts were normalized relative to the mean and standard deviation of permuted data (yielding counts of collocated fixations as *Z* scores). As noted in the main text, both groups show more co-occurring, collocated fixations on eye and mouth than expected by chance (main text Figures 2n,o), but MZ twins exhibit greater concordance than DZ twins ($F_{1,81} =$ 4.89, *p*=0.030; main text Figure 2p). In addition, the relative difference between hits and misses (difference in *Z* scores for eyes-eyes or mouth-mouth versus eyes-mouth or mouth-eyes, seen in the relative heights of plots in main text Figures 2n,o) is greater for MZ than DZ twins: MZ twins are both *more* likely to look at the eyes or mouth at the same moments in time, as well as relatively *less* likely to split their looking between different regions.

Power Calculations—For determining sample size in the present study, power calculations were based on assumptions from the existing literature on the longitudinal course and genetic structure of reciprocal social behavior[60–63]. Analyses indicated that twin pairs samples of 40 or greater would provide 80% power to detect correlations of approximately *r*=0.38 (approximately half the magnitude of MZ correlations observed in [31]). Actual statistical power to detect concordance between two measurements depends not only on the true genetic correlation (*r*) between them, but on their marginal heritabilities (H^2): When $H^2 = 50\%$, power is above 80% when $r >= 0.27$; and when $H^2 = 20\%$ (lower than anticipated from the existing literature), power is above 80% when $r >= 0.55$ (alpha = 0.001). Given the size of the observed MZ and DZ concordance effects, measurement estimates of our achieved power (1-β error probability) for MZ eye-looking was ~1; in DZ twins, achieved power for eye-looking was 0.77. Future work will follow-up in larger samples.

Additionally, in our final experiment, we tested two further hypotheses (described in the main text). For the first, the null hypothesis stated that concordance when watching the same videos would be equal to the value observed for watching all videos (H_0: $ICC_{same}=ICC_{all}$);

the alternative stated that concordance would be greater (H_1: $ICC_{same} > ICC_{all}$). Given the extremely high intraclass correlation coefficients already observed for MZ twins (eyes, 0.91), we were aware that we would not be sufficiently powered to detect a significant increase in concordance greater than this value for the MZ sample; however, we were adequately powered to detect significant increases, should they be observed, in the DZ and non-sibling samples. Likewise, in the second and third tests—comparing concordance when each twin watched *different* videos and when each twin watched different content categories of videos—the null stated that concordance would be zero (H_0: $ICC_{different} = 0$), the alternative stated that concordance would be greater than zero (H_1: $ICC_{different} > 0$). Here, as in the main set of analyses, we had >80% power to detect correlations of >=0.38.

Extended Data
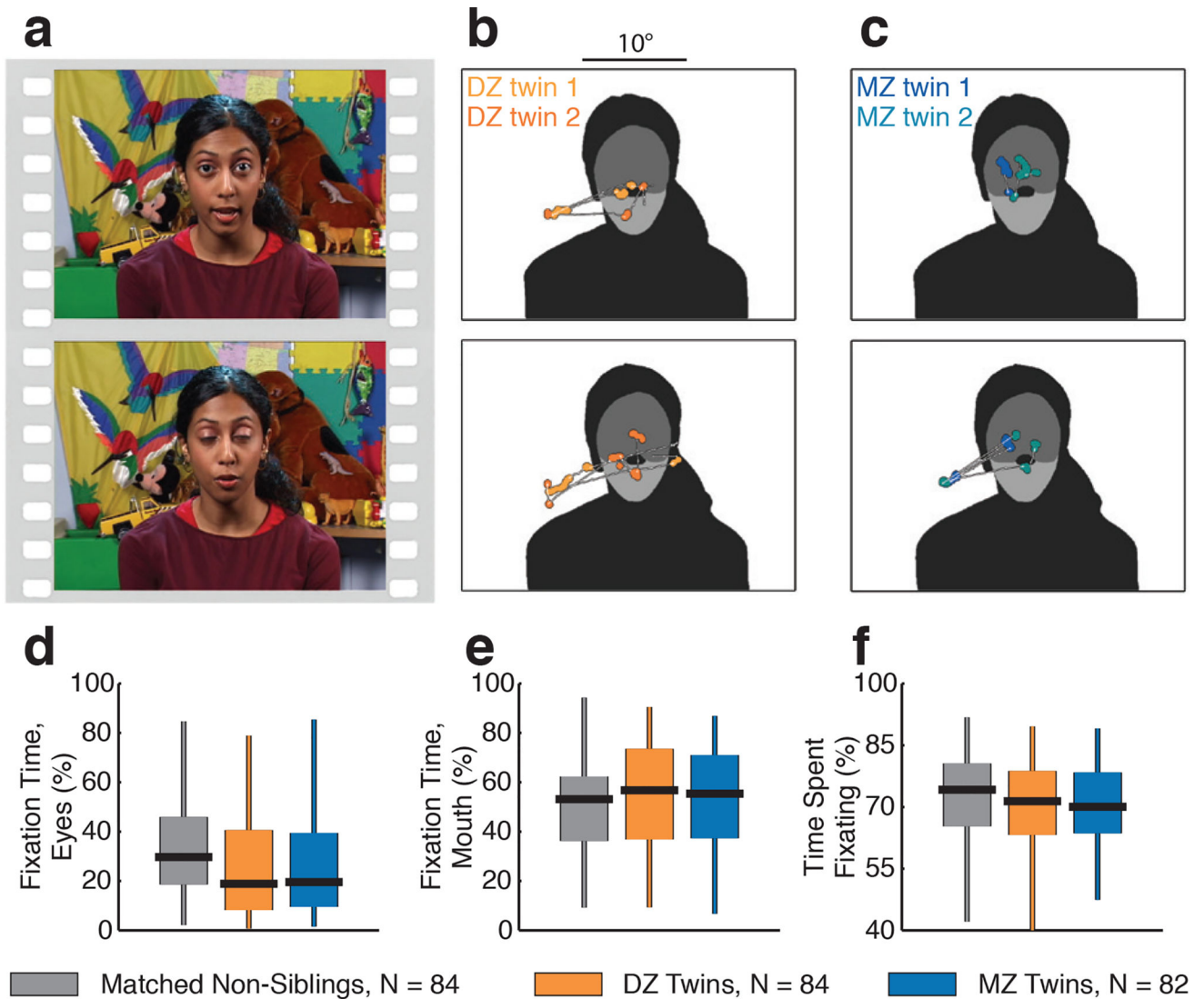
Extended Data Figure 1. Measuring genetic structure of social visual engagement in 250 paired toddlers: dizygotic twins (N=84, 42 pair), monozygotic twins (N=82, 41 pair), and non-sibling comparison children (N=84, randomized to 42 pairs)

a, Example still images from dyadic mutual gaze video stimuli. **b,** Data from two typically-developing 18-month-old dizygotic (DZ) twins. **c,** Data from two typically-developing 18-month-old monozygotic (MZ) twins. (b) and (c) plot two seconds of eye-tracking data, corresponding to each image in (a) (the image onscreen at midpoint of two-second data sample). Data are overlaid on each image's corresponding regions of interest, shaded to indicate eyes, mouth, body, and object regions. Saccades are plotted as thin white lines with white dots; fixation data are plotted as larger colored dots. **d-f,** Fixation time summaries for each comparison group for percentage of total fixation time on eyes region (d), percentage of total fixation time on mouth region (e), and percentage of total time spent fixating (f). Boxplots span full range of data collected, with vertical lines extending from minimum to maximum values, boxes spanning the 25th to 75th percentiles, and horizontal black lines marking medians.
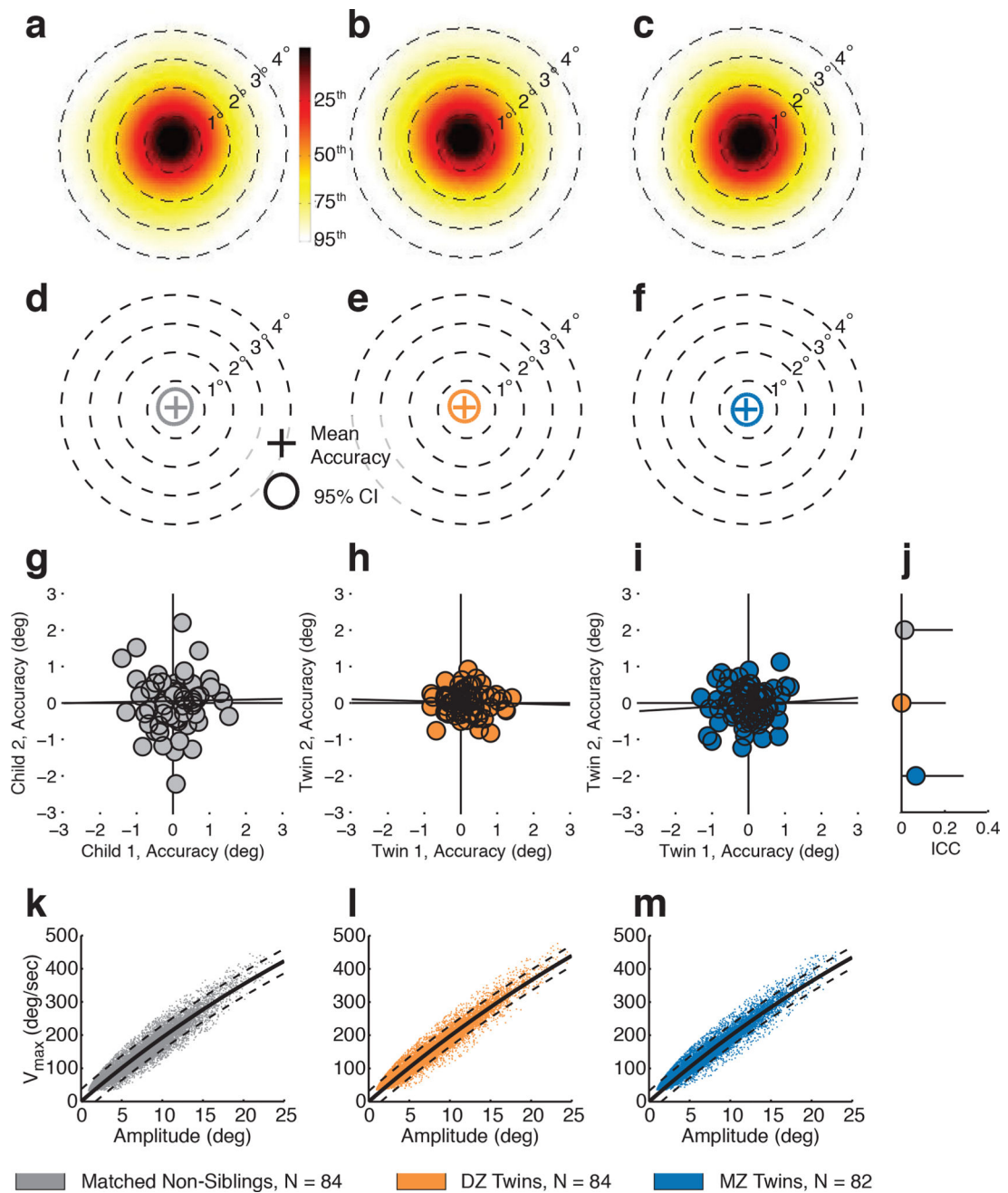
Extended Data Figure 2. Between-group controls for calibration accuracy and oculomotor function

To test for group-wise differences unrelated to subsequent paired comparisons in the main study experiments we measured calibration accuracy and oculomotor function. **a-c,** Total variance in calibration accuracy for age- and sex-matched non-sibling controls (a), DZ twins (b), and MZ twins (c). Plots show kernel density estimates of the distribution of measured fixation locations relative to calibration accuracy verification targets. **d-f,** Average calibration accuracy for non-sibling controls (d), DZ twins (e), and MZ twins (f). Crosses mark the location of mean calibration accuracy, while annuli mark 95% confidence intervals (CI). **g-i,** Concordance in calibration accuracy measures for non-sibling controls (g), DZ

twins (h), and MZ twins (i). Measures in (g-i) are average accuracy per child across all accuracy verification trials. **j,** Intraclass correlation coefficients (ICC, plotted with 95% confidence intervals). **k-m,** Oculomotor relationship between maximum saccade velocity (Vmax) and amplitude for non-sibling controls (k), DZ twins (l), and MZ twins (m).
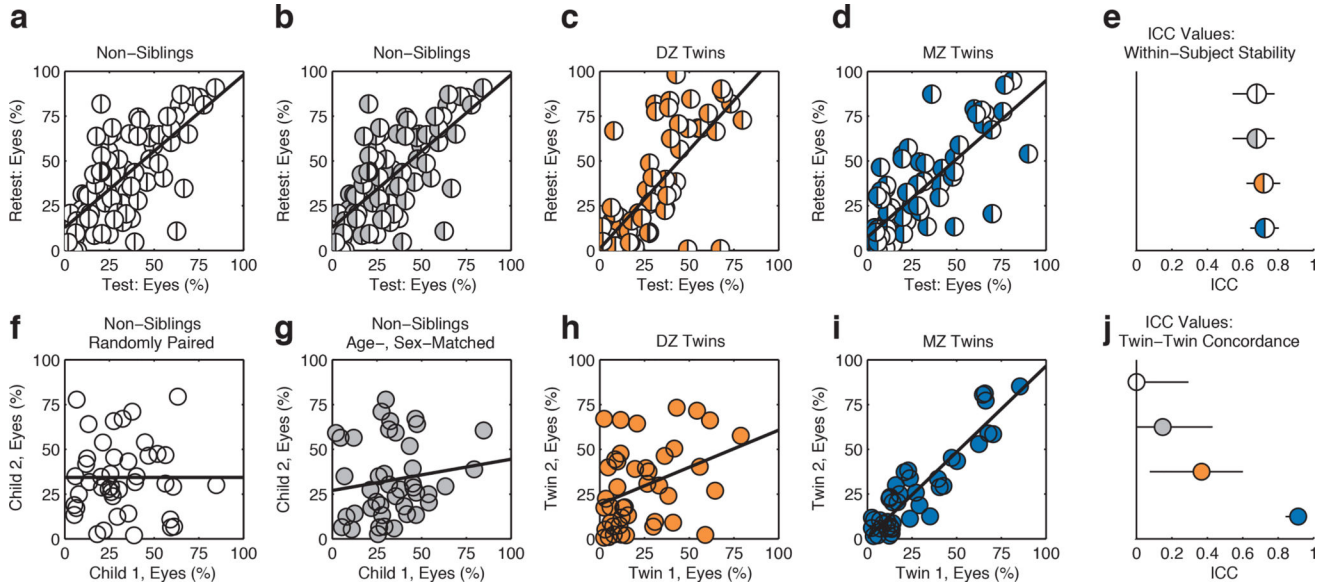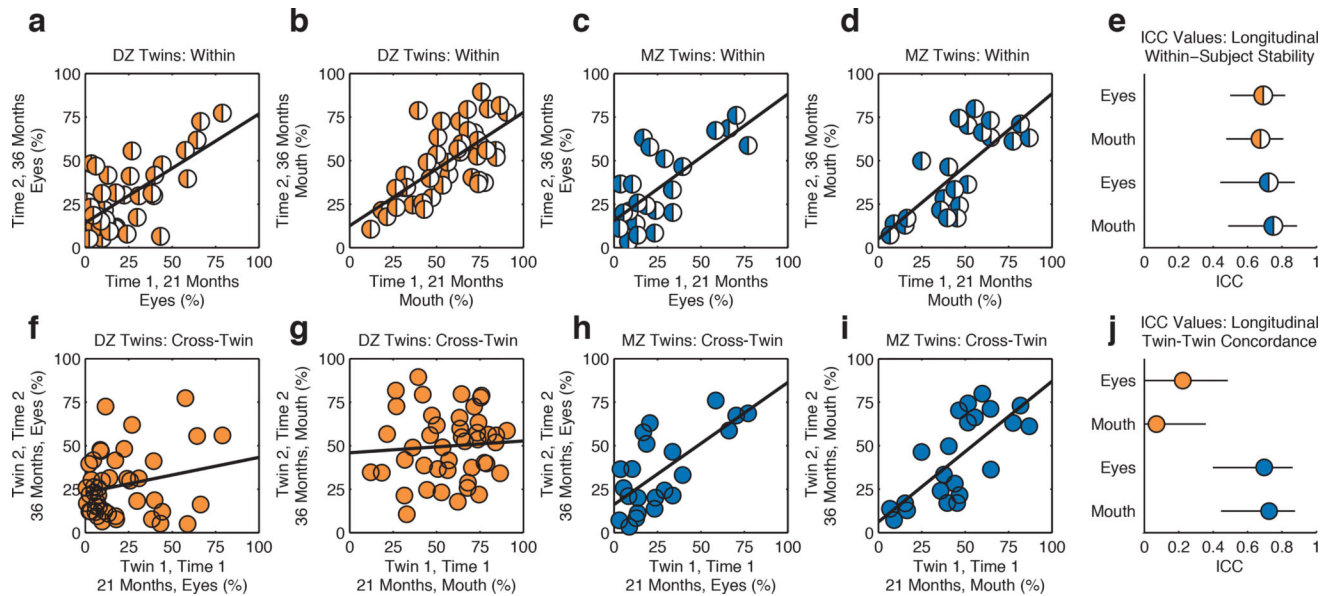

Extended Data Figure 3. Within-subject stability versus between-subject concordance
For heritable traits, one expects to observe substantial within-subject stability contrasting with marked differences, varying by zygosity, in between-subject (twin-twin) concordance. **a-d,** Within-subject stability of observed levels of eye-looking for non-siblings (a, b), DZ twins (c), and MZ twins (d). (Scatter plots in (a) and (b) are repeated for comparison with plots (f) and (g).) **e,** Group-wise summary of within-subject stability (test-retest reliability) of measures of eye-looking quantified by intraclass correlation coefficient (ICC) with 2-way random effects model (ICC(2,1)). Error bars are 95% confidence intervals. Note that estimates assuming fixed rather than random effects of testing (ICC(3,k), not plotted) yield ICC values greater than 0.9 for each group, evidence that the measures of inter-individual variation—the difference between individuals—are also highly reliable. **f-i,** Plots repeated from main text Figure 1a-e, showing paired measures of eye-looking in randomly-paired non-siblings (f), in age- and sex-matched non-siblings (g), in DZ twins (h), and in MZ twins (i). **j,** Intraclass correlation coefficients and 95% confidence intervals for twin-twin concordance in eye-looking.
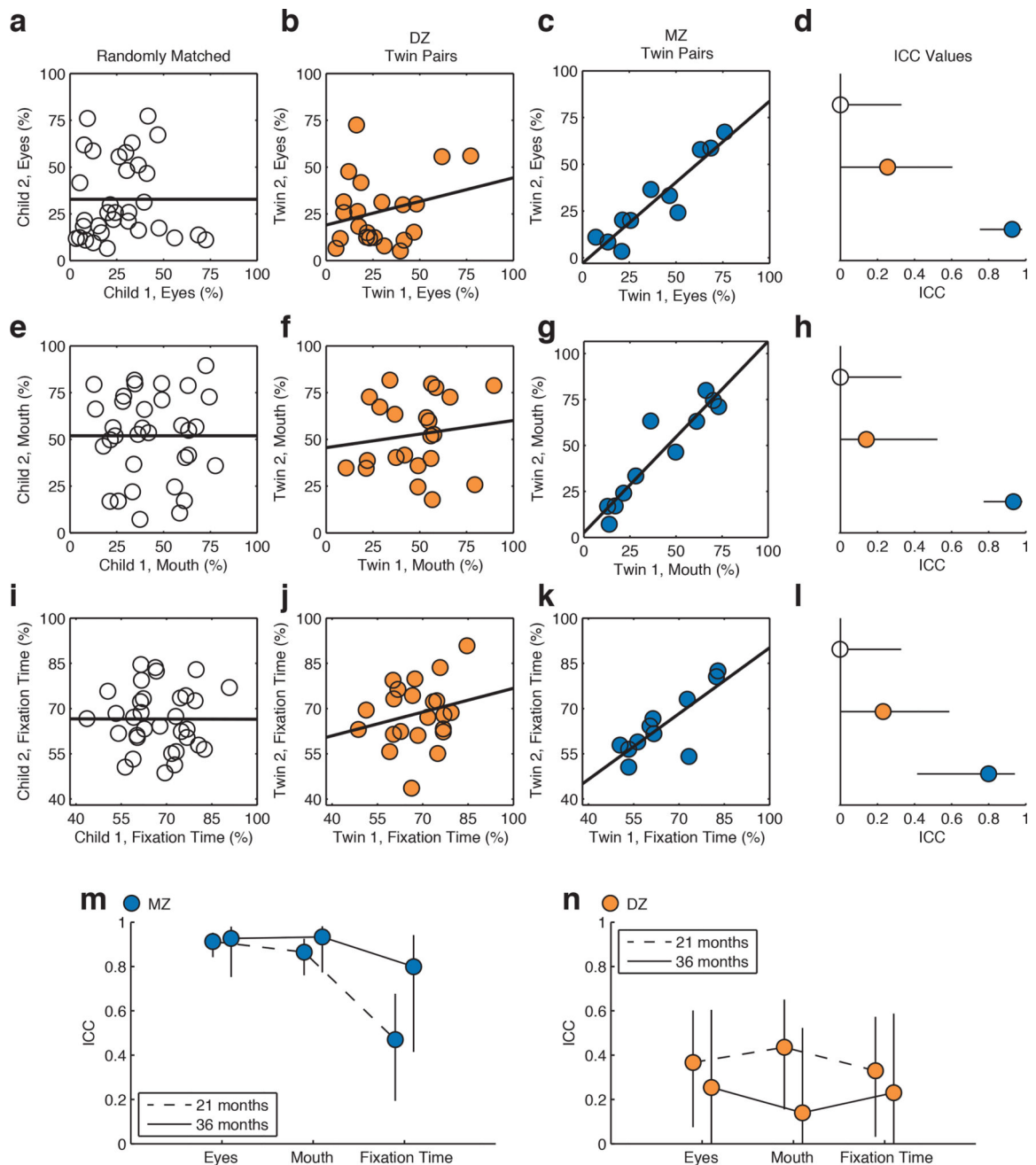
Extended Data Figure 4. Monozygotic (MZ) twins maintain high twin-twin concordance, significantly greater than that observed in dizygotic (DZ) twins, when tested again at 36 months **a-c,** Paired measures of eye-looking in randomly-assigned pairs (a), in DZ twins (b), and in MZ twins (c). **d,** Intraclass correlation coefficients and 95% confidence intervals across groups for eye-looking. **e-h,** Paired measures of concordance in mouth-looking. **i-l,** Paired measures of concordance in percentage of time spent attending to task (maintaining stable onscreen fixation). In all plots, randomly-matched controls in white, DZ twins in orange, and MZ twins in blue. Error estimates are 95% confidence intervals. **m-n,** Summary of MZ (m) and DZ (n) results at initial time of testing (21 months, summary data from Figure 1 in

main text) relative to results at time of longitudinal follow-up (36 months, summary from d, h, & l above). MZ twins exhibit marginally, though not significantly, increased concordance values when tested again at 36 months; in contrast, DZ twins exhibit marginally, though not significantly, decreased concordance values. Plotted data in (a), (e), and (i) are a representative case of random pairing, selected to match the mean ICC value of all 10,000 re-samplings.

Extended Data Figure 5. Longitudinal within-subject stability versus longitudinal twin-twin concordance, from 21 until 36 months

DZ and MZ twins both show high levels of longitudinal within-subject stability when tested again 15 months after initial data were collected, but only MZ twins show high levels of longitudinal twin-twin concordance, with twin 1's results at 21 months being highly concordant with twin 2's at 36 months. **a-d,** Within-subject stability of observed levels of eye-looking (a) and mouth-looking (b) for DZ twins, and within-subject stability of eye-looking (c) and mouth-looking (d) for MZ twins. **e,** Summary of longitudinal within-subject stability quantified by intraclass correlation coefficient (ICC) with 2-way random effects model. Error bars are 95% confidence intervals. **f-i,** Longitudinal twin-twin concordance (twin 1 at 21 months paired with twin 2 at 36 months) for eye-looking (f) and mouth-looking (g) in DZ twins, and for eye- (h) and mouth-looking (i) in MZ twins. **j,** Intraclass correlation coefficients and 95% confidence intervals.
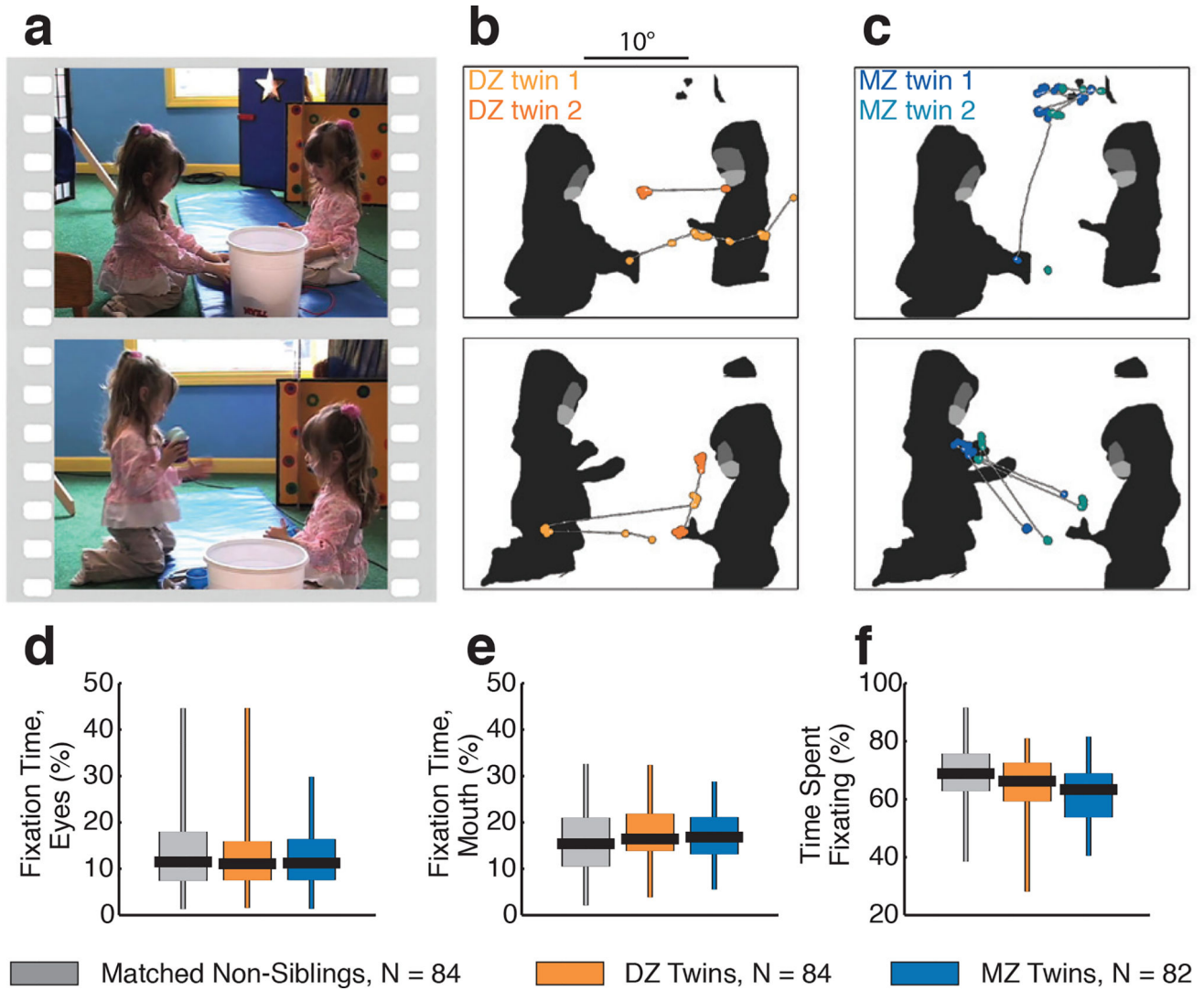
Extended Data Figure 6. Social visual engagement when watching Triadic Peer Interaction stimuli in 250 paired toddlers: dizygotic twins (N=84, 42 pairs), monozygotic twins (N=82, 41 pairs), and non-sibling comparison children (N=84, randomized to 42 pairs)

a, Example still images from triadic peer interaction stimuli. **b,** Data from two typically-developing 18-month-old dizygotic (DZ) twins. **c,** Data from two typically-developing 18-month-old monozygotic (MZ) twins. In (b) and (c), two seconds of eye-tracking data are plotted, corresponding to each image in (a) (the image onscreen at midpoint of the two-second data sample). Data are overlaid on each image's corresponding regions of interest, shaded to indicate eyes, mouth, body, and object regions. Saccades are plotted as thin white lines with white dots; fixation data are plotted as larger colored dots. **d-f,** Fixation time summaries for each comparison group for percentage of total fixation time on eyes region (d), percentage of total fixation time on mouth region (e), and percentage of total time spent fixating (f). Boxplots span full range of data collected, with vertical lines extending from minimum to maximum values, boxes spanning the 25th to 75th percentiles, and horizontal black lines marking medians.

Extended Data Figure 7. Physical image properties that constitute eyes vary significantly from video stimulus to video stimulus in lightness, color, contrast, orientation gradients, and motion

a, Still images sampled from videos depicting dyadic mutual gaze stimuli (an entreating caregiver, engaging the child in mutual gaze and play routines). Still images from 5 of 15 videos are shown (all 15 dyadic mutual gaze videos included in actual analyses). **b,** Eye region demarcated from each still image in (a). Across all demarcated eye regions, across all frames of videos presented, physical image property profiles were analyzed. In the row to the right of each representative still image and corresponding eye region, physical image property profiles, analyzed across all video frames, are given as histograms. **c,** Lightness. **d,**

Red-green color opponency. **e,** Yellow-blue color opponency. **f,** Contrast. **g,** Orientation gradients. **h,** Motion. **i,** For each physical image property analyzed in columns (a-h), row (i) gives corresponding comparison plots across the 5 histograms located in the column directly above. **j,** Statistical comparisons of the measured image property distributions by 2-sample Kolmogorov- Smirnov test. *P* values are corrected for multiple comparisons by the Bonferroni method. For each of the physical image properties analyzed in columns (a-h), row (j) presents the corresponding matrix of statistical comparisons (*i.e.*, the 1st row of colored circles presents comparisons for video 1 vs. 2, video 1 vs. 3, etc.; while the 2nd row presents comparisons for video 2 vs. 3, 2 vs. 4, etc.). **k,** Still images sampled from videos depicting triadic peer interaction stimuli (scenes of children interacting in a daycare setting). Still images from 5 of 12 videos are shown (all 12 triadic peer interaction videos included in actual analyses). **l,** Eye regions demarcated from each still image in (k). **m-t,** All parts of (m-t) are as in (c-j).

Extended Data Table 1
Participant Demographics

	Total Epidemiologically-Ascertained Twins		Eye-Tracking Participants (Twins)		Eye-Tracking Participants (Non-Siblings)	
	%	N	%	N	%	N
Sex						
Male	47.8	172	55.4	92	52.4	44
Female	52.2	188	44.6	74	47.6	40
Zygosity						
Monozygotic	35	126	49.4	82		
Dizygotic	58.3	210	50.6	84		
same sex	36.1	130	50.6	84	N/A	N/A
opposite sex	22.2	80	0	0		
Undetermined	6.7	24	0	0		
Income						
$29,999	19.4	70	15.7	26	6.1	5
$30,000–$59,999	24.4	88	24.1	40	10.6	9
$60,000–$89,999	21.7	78	22.9	38	18.6	16
$90,000	30.6	110	37.3	62	56.6	46
N/A	3.9	14	0	0	8.1	7
Race						
Asian	1.1	4	0	0	4.8	4
Black/African-American	21.1	76	14.5	24	4.8	4
Caucasian	77.8	280	85.5	142	78.5	66
More than one race	0	0	0	0	7.1	6
Unknown / Not reported	0	0	0	0	4.8	4
Ethnicity						
Hispanic	7.8	28	8.4	14	7.3	6
Non-Hispanic	92.2	332	91.6	152	74.4	63

	Total Epidemiologically-Ascertained Twins		Eye-Tracking Participants (Twins)		Eye-Tracking Participants (Non-Siblings)	
	%	**N**	**%**	**N**	**%**	**N**
Unknown / Not reported	0	0	0	0	18.3	15

Extended Data Table 2
Concordance in social visual engagement at 21 months, at 36 months, and from 21 until 36 months

a, Concordance in social visual engagement at 21 months. **b,** Concordance in social visual engagement in subset seen for repeated testing at 36 months (see Methods). **c,** Cross-twin concordance in social visual engagement from 21 (time 1, twin 1) until 36 months (time 2, twin 2). In all cases, results are given as intraclass correlation coefficient (ICC) with 95% confidence intervals in parentheses.

a

	Eyes	Mouth	Body	Object	Time Spent Attending to Task
MZ Twins (N = 41 pairs)	0.91 (0.85 – 0.95) **	0.86 (0.76 – 0.92) **	0.71 (0.52 – 0.83) **	0.66 (0.46 – 0.80) **	0.46 (0.19 – 0.67) *
DZ Twins (N = 42 pairs)	0.35 (0.07 – 0.59) *	0.44 (0.16 – 0.65) *	0.33 (0.04 – 0.57) *	0.09 (0.00 – 0.38)	0.34 (0.05 – 0.58) *
Age-, Sex-Matched Non-Siblings (N = 42 pairs)	0.16 (0.00 – 0.44)	0.13 (0.00 – 0.42)	0.29 (0.00 – 0.55)	0.14 (0.00 – 0.42)	0.14 (0.00 – 0.43)
Randomly-Matched Non-Siblings (N = 42 pairs; 10,000 resamplings)	0.00 (0.00 – 0.29)	0.00 (0.00 – 0.29)	0.00 (0.00 – 0.29)	0.00 (0.00 – 0.30)	0.00 (0.00 – 0.30)

b

	Eyes	Mouth	Body	Object	Time Spent Attending to Task
MZ Twins (N = 11 pairs)	0.93 (0.75 – 0.98) **	0.93 (0.77 – 0.98) **	0.63 (0.08 – 0.88) *	0.95 (0.81 – 0.99) **	0.80 (0.19 – 0.94) **
DZ Twins (N = 22 pairs)	0.25 (0.00 – 0.60)	0.14 (0.00 – 0.52)	0.21 (0.00 – 0.58)	0.00 (0.00 – 0.41)	0.23 (0.00 – 0.59)
Randomly-Matched Pairs (N = 33 pairs)	0.00 (0.00 – 0.33)	0.00 (0.00 – 0.33)	0.00 (0.00 – 0.33)	0.00 (0.00 – 0.33)	0.00 (0.00 – 0.33)

c

	Eyes	Mouth	Body	Object	Time Spent Attending to Task
MZ Twins (N = 11 pairs)	0.70 (0.40 – 0.86) **	0.73 (0.45 – 0.88) **	0.21 (0.00 – 0.58)	0.74 (0.47 – 0.88) **	0.09 (0.00 – 0.49)
DZ Twins	0.22 (0.00 – 0.49)	0.07 (0.00 – 0.36)	0.02 (0.00 – 0.31)	0.07 (0.00 – 0.35)	0.00 (0.00 – 0.30)

c

	Eyes	Mouth	Body	Object	Time Spent Attending to Task
(N = 22 pairs)					

Single asterisk (*) signifies *P* < 0.05 significance value, one-sided comparison relative to 0.

Double asterisk (**) signifies *P* < 0.01 significance value, one-sided comparison relative to 0.

Extended Data Table 3
Size of experimental stimuli and viewing time summaries

a, Size of Regions-of-Interest, Dyadic Mutual Gaze Stimuli. Data are given as mean (SD) in degrees of visual angle. Object ROIs generally spanned the full horizontal and vertical extent of the background in all video images, excepting cases of some body and hand gestures, as shown in Extended Data Figure 1. The average minimum visual area subtended by any portion of the object ROI is equal to the difference between object and body ROIs. **b,** Size of Regions-of-Interest, Triadic Peer Interaction Stimuli. Data are given as mean (SD) in degrees of visual angle. Eyes and Mouth ROI sizes reflect the average size of a single face within the stimuli. Body ROIs are frequently contiguous between individuals in the stimuli (see Extended Data Figure 6k); measures reflect total body region size. **c,** Total Viewing Time and Time Spent in Fixation, Saccade, Offscreen/Missing, Blink; mean (SD) in minutes. All measures summarized across both Dyadic Mutual Gaze Stimuli and Triadic Peer Interaction Stimuli. Non-sibling controls watched foreshortened subset of video stimuli. **c,** Time Fixating Per Onscreen Region-of-Interest, mean (SD) in minutes. All measures summarized across both Dyadic Mutual Gaze Stimuli and Triadic Peer Interaction Stimuli. Non-sibling controls watched foreshortened subset of video stimuli.

a

	Eyes	Mouth	Body	Object
Horizontal	8.04° (0.46)	7.71° (0.49)	25.11° (2.70)	31.99° (0.05)
Vertical	6.91° (0.44)	5.72° (0.59)	21.71° (0.73)	23.94° (0.49)

b

	Eyes	Mouth	Body	Object
Horizontal	4.64° (2.75)	4.24° (2.54)	20.64° (7.66)	28.00° (6.81)
Vertical	4.06° (2.19)	3.16° (1.54)	20.83° (4.09)	23.62° (1.74)

c

	Fixation	Saccade	Offscreen/Missing	Blink	Total Viewing Time
MZ Twins (N = 42 pairs)	11.43 (2.86)	3.28 (1.07)	3.12 (1.68)	0.36 (0.23)	18.20 (3.12)
DZ Twins (N = 42 pairs)	11.54 (3.24)	3.24 (1.05)	2.82 (1.47)	0.30 (0.19)	17.90 (3.40)
Non-Sibling Controls (N = 42 pairs)	6.93 (3.01)	1.55 (0.71)	1.30 (0.93)	0.25 (0.30)	10.03 (3.81)

d

	Eyes	Mouth	Body	Object
MZ Twins (N = 41 pairs)	2.09 (1.57)	3.71 (1.67)	3.27 (0.83)	2.37 (0.71)
DZ Twins (N = 42 pairs)	2.02 (1.48)	3.85 (1.87)	3.34 (1.00)	2.33 (0.74)
Non-Sibling Controls (N = 42 pairs)	1.92 (1.48)	2.84 (1.87)	1.30 (0.56)	0.87 (0.42)

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Gibson EJ. Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. Annu. Rev. Psychol. 1988; 39:1–41.

2. Valenza E, Simion F, Cassia VM, Umiltà C. Face preference at birth. J. Exp. Psychol. Hum. Percept. Perform. 1996; 22:892. [PubMed: 8756957]

3. Goren CC, Sarty M, Wu PYK. Visual Following and Pattern Discrimination of Face-like Stimuli by Newborn Infants. Pediatrics. 1975; 56:544–549. [PubMed: 1165958]

4. Simion F, Regolin L, Bulf H. A predisposition for biological motion in the newborn baby. Proc. Natl. Acad. Sci. 2008; 105:809–813. [PubMed: 18174333]

5. Simion F, Leo I, Turati C, Valenza E, Dalla Barba B. How face specialization emerges in the first months of life. Progress in Brain Research. 2007; 164:169–185. [PubMed: 17920431]

6. Jones W, Klin A. Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. Nature. 2013; 504:427–31. [PubMed: 24196715]

7. Scarr S, McCartney K. How People Make Their Own Environments: A Theory of Genotype ->Environment Effects. Child Dev. 1983; 54:424–435. [PubMed: 6683622]

8. Constantino JN, Charman T. Diagnosis of autism spectrum disorder: Reconciling the syndrome, its diverse origins, and variation in expression. Lancet Neurol. 2015; 15:279–291. [PubMed: 26497771]

9. Constantino JN, et al. Autism recurrence in half siblings: strong support for genetic mechanisms of transmission in ASD. Mol. Psychiatry. 2013; 18:137–8. [PubMed: 22371046]

10. Gaugler T, et al. Most genetic risk for autism resides with common variation. Nat. Genet. 2014; 46:881–5. [PubMed: 25038753]

11. Robinson EB, et al. Evidence that autistic traits show the same etiology in the general population and at the quantitative extremes (5%, 2.5%, and 1%). Arch. Gen. Psychiatry. 2011; 68:1113–21. [PubMed: 22065527]

12. Geschwind DH, State MW. Gene hunting in autism spectrum disorder: On the path to precision medicine. The Lancet Neurology. 2015; 14:1109–1120. [PubMed: 25891009]

13. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, fifth edition: DSM-5. American Psychiatric Pub; 2013.

14. Magiati I, Tay XW, Howlin P. Cognitive, language, social and behavioural outcomes in adults with autism spectrum disorders: A systematic review of longitudinal follow-up studies in adulthood. Clin. Psychol. Rev. 2014; 34:73–86. [PubMed: 24424351]

15. Shultz S, Klin A, Jones W. Inhibition of eye blinking reveals subjective perceptions of stimulus salience. Proc. Natl. Acad. Sci. U. S. A. 2011; 108:21270–5. [PubMed: 22160686]

16. 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. 2002; 59:809–816. [PubMed: 12215080]

17. Pearce JM, Bouton ME. Theories of associative learning in animals. Annu. Rev. Psychol. 2001; 52:111–139. [PubMed: 11148301]

18. Dickinson A. Actions and Habits: The Development of Behavioural Autonomy. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 1985; 308:67–78.

19. Good, P. Permutation, parametric, and bootstrap tests of hypotheses. Springer; 2000.

20. Liben, L., Müller, U., Lerner, R. Handbook of child psychology and developmental science. Volume 2. Cognitive Processes (Seventh ed.). John Wiley & Sons; 2015.

21. McGraw KO, Wong SP. Forming inferences about some intraclass correlations coefficients. Psychol. Methods. 1996; 1:30–46.

22. Jacquard A. Heritability: One Word, Three Concepts. Biometrics. 1983; 39:465–477. [PubMed: 6626665]

23. Leigh, RJ., Zee, DS. The Neurology of Eye Movements. Oxford University Press; USA: 2006.

24. Schall JD, Thompson KG. Neural Selection and Control of Visually Guided Eye Movements. Annu. Rev. Neurosci. 1999; 22:241–259. [PubMed: 10202539]

25. Marr D. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. New York. 1982; 397doi: 10.2307/2185011

26. Treue S. Visual attention: The where, what, how and why of saliency. Current Opinion in Neurobiology. 2003; 13:428–432. [PubMed: 12965289]

27. Hopfinger JB, Buonocore MH, Mangun GR. The neural mechanisms of top-down attentional control. Nat. Neurosci. 2000; 3:284–291. [PubMed: 10700262]

28. Wang S, et al. Atypical Visual Saliency in Autism Spectrum Disorder Quantified through Model-Based Eye Tracking. Neuron. 2015; 88:604–616. [PubMed: 26593094]

29. Oyama, S. Evolution's eye: A systems view of the biology-culture divide. Duke University Press; 2000.

30. Klin A, Jones W, Schultz R, Volkmar F. The enactive mind, or from actions to cognition: lessons from autism. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 2003; 358:345–60. [PubMed: 12639332]

31. Marrus N, et al. Rapid video-referenced ratings of reciprocal social behavior in toddlers: a twin study. J. Child Psychol. Psychiatry. 2015; 56:1338–46. [PubMed: 25677414]

32. Lord, C., Rutter, M., DiLavore, PC., Risi, S. Autism Diagnostic Observation Schedule. Western Psychological Services; 2002.

33. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, fourth edition, text revision: DSM-IV-TR. American Psychiatric Association; 2004.

34. Price TS, et al. Infant zygosity can be assigned by parental report questionnaire data. Twin Res. 2000; 3:129–33. [PubMed: 11035484]

35. Neale MC, Stevenson J. Rater bias in the EASI temperament scales: a twin study. J. Pers. Soc. Psychol. 1989; 56:446–55. [PubMed: 2926639]

36. 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. 2008; 65:946–54. [PubMed: 18678799]

37. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychol. Bull. 1979; 86:420–428. [PubMed: 18839484]

38. Locke, JL. The child's path to spoken language. Harvard University Press; 1993.

39. Fleiss JL, Shrout PE. Approximate interval estimation for a certain intraclass correlation coefficient. Psychometrika. 1978; 43:259–262.

40. Giraudeau B. Negative values of the intraclass correlation coefficient are not theoretically possible. J. Clin. Epidemiol. 1996; 49:1205–1206. [PubMed: 8827004]

41. Emery NJ. The eyes have it: The neuroethology, function and evolution of social gaze. Neuroscience and Biobehavioral Reviews. 2000; 24:581–604. [PubMed: 10940436]

42. Fairchild, MD. Color Appearance Models. John Wiley & Sons Ltd; 2005.

43. Itti L, Koch C. A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research. 2000; 40:1489–1506. [PubMed: 10788654]

44. Koch C, Ullman S. Shifts in selective visual attention: towards the underlying neural circuitry. Hum. Neurobiol. 1985; 4:219–27. [PubMed: 3836989]

45. Itti L, Itti L, Koch C, Koch C. Computational modelling of visual attention. Nat. Rev. Neurosci. 2001; 2:194–203. [PubMed: 11256080]

46. Parkhurst D, Law K, Niebur E. Modeling the role of salience in the allocation of overt visual attention. Vision Res. 2002; 42:107–123. [PubMed: 11804636]

47. Wolfe JM, Horowitz TS. What attributes guide the deployment of visual attention and how do they do it? Nat. Rev. Neurosci. 2004; 5:495–501. [PubMed: 15152199]

48. Tsotsos JK, et al. Modeling visual attention via selective tuning. Artif. Intell. 1995; 78:507–545.

49. Yantis S, Egeth HE. On the distinction between visual salience and stimulus-driven attentional capture. J. Exp. Psychol. Hum. Percept. Perform. 1999; 25:661–676. [PubMed: 10385983]

50. Mazer JA, Gallant JL. Goal-related activity in V4 during free viewing visual search: Evidence for a ventral stream visual salience map. Neuron. 2003; 40:1241–1250. [PubMed: 14687556]

51. Henderson JM, Brockmole JR, Castelhano MS, Mack M. Visual saliency does not account for eye movements during visual search in real world scenes. Eye Movements A Wind. Mind Brain. 2007; :537–562. DOI: 10.1167/9.3.6

52. Lettvin JY, Maturana HR, McCulloch WS, Pitts WH. What the Frog's Eye Tells the Frogs's Brain. Proc. tha IRE. 1959; 3:1940–1951.

53. Ghazanfar, Aa, Santos, LR. Primate brains in the wild: the sensory bases for social interactions. Nat. Rev. Neurosci. 2004; 5:603–616. [PubMed: 15263891]

54. Moore GP, Perkel DH, Segundo JP. Statistical analysis and functional interpretation of neuronal spike data. Annu. Rev. Physiol. 1966; 28:493–522. [PubMed: 5323104]

55. Manly, B. Randomization, Bootstrap, and Monte Carlo Methods in Biology. Chapman & Hall; 2006.

56. Oppenheim, A., Schafer, R. Digital Signal Processing. Prentice-Hall; 1975.

57. Schnyder H, Reisine H, Hepp K, Henn V. Frontal eye field projection to the paramedian pontine reticular formation traced with wheat germ agglutinin in the monkey. Brain Res. 1985; 329:151–60. [PubMed: 3978438]

58. Hanes DP, Wurtz RH. Interaction of the frontal eye field and superior colliculus for saccade generation. J. Neurophysiol. 2001; 85:804–15. [PubMed: 11160514]

59. Bruce CJ, Goldberg ME. Physiology of the frontal eye fields. Trends Neurosci. 1984; 7:436–441.

60. Constantino JN, Todd RD. Genetic structure of reciprocal social behavior. Am. J. Psychiatry. 2000; 157:2043–2045. [PubMed: 11097975]

61. Constantino JN, Todd RD. Autistic traits in the general population: a twin study. Arch. Gen. Psychiatry. 2003; 60:524–530. [PubMed: 12742874]

62. Constantino JN, Todd RD. Intergenerational transmission of subthreshold autistic traits in the general population. Biol. Psychiatry. 2005; 57:655–660. [PubMed: 15780853]
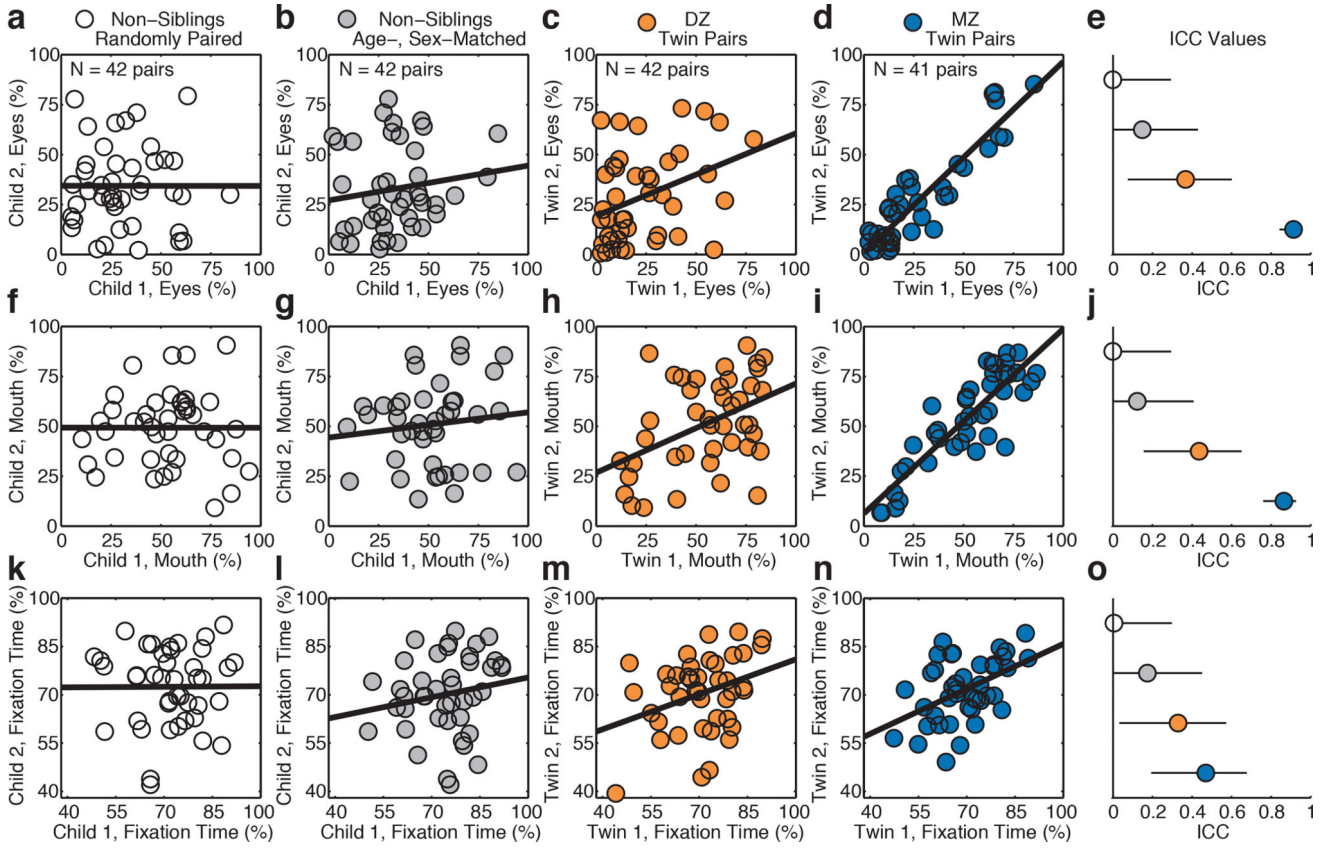
63. Constantino JN, et al. Developmental course of autistic social impairment in males. Dev. Psychopathol. 2009; 21:127–38. [PubMed: 19144226]
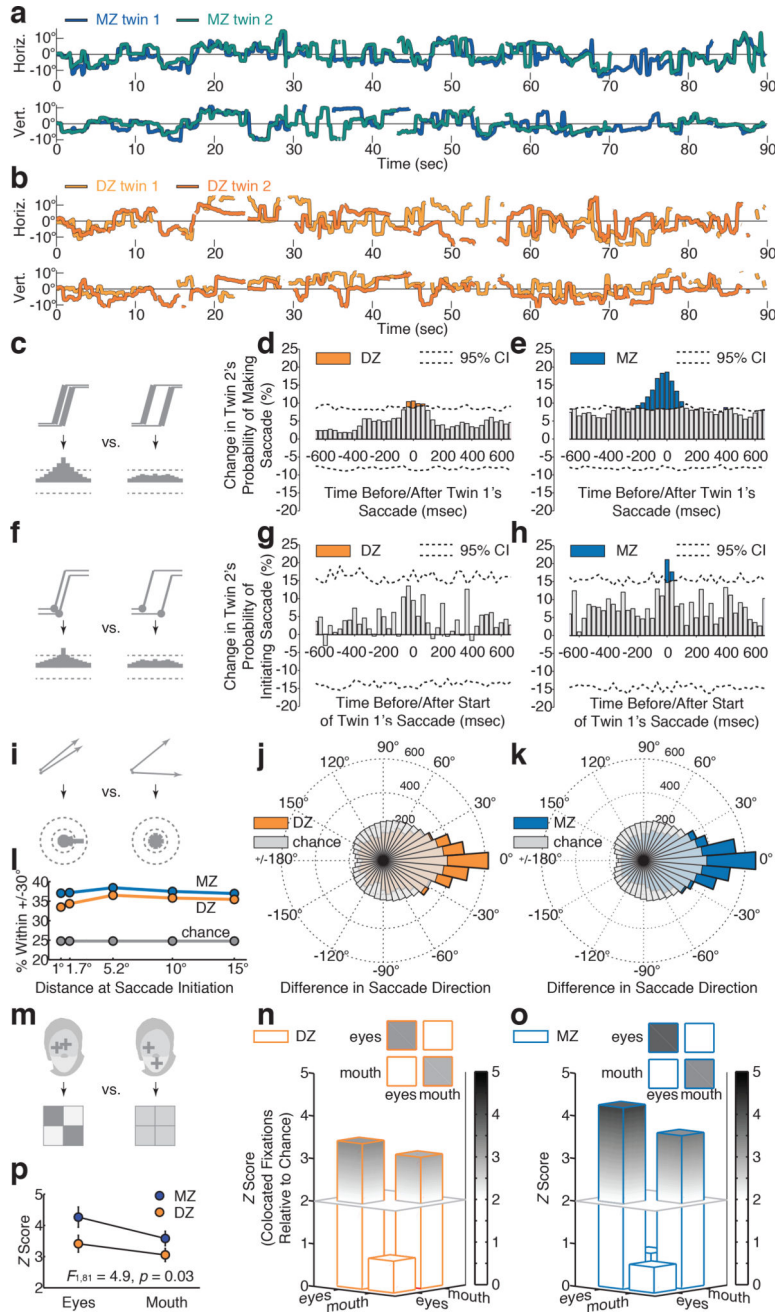
Figure 1. Monozygotic (MZ) twins exhibit high twin-twin concordance for eye- and mouth-looking, significantly greater than dizygotic (DZ) twins or age- and sex-matched non-siblings
Paired measures of eye-looking in, **a,** non-siblings paired randomly in 10,000 re-samplings without replacement; **b,** age- and sex-matched non-siblings; **c,** DZ twins; and **d,** MZ twins. **e,** Intraclass correlation coefficients (ICC) and 95% confidence intervals for eye-looking. **f-j,** Concordance in mouth-looking. **k-o,** Concordance in time spent attending to task (maintaining stable onscreen fixation). Plotted data in (a), (f), (k) are representative, selected to match mean ICC of all 10,000 re-samplings.

Figure 2. When viewing scenes of social interaction, monozygotic (MZ) twins exhibit greater probability of shifting their eyes at the same moments; greater probability of shifting their eyes in the same subsequent directions; and greater probability of fixating the same semantic content at the same moments

a-b, Example eye position data for MZ and DZ twins. Gaps in plots reflect blinks or off-screen fixations. **c,** Schematic peristimulus time histograms (PSTHs) showing probability of co-occurring saccades. At left, if saccades co-occur, twin 2's saccade probability increases with twin 1's saccades; at right, if saccades do not co-occur, probability remains unchanged. Dotted lines in (d-e) and (g-h) show 95% confidence intervals for change expected by chance (*i.e.*, no time-locking), measured by permutation testing. **d-e,** Small DZ vs. large MZ

increase in probability of time-locked saccades. **f,** Schematic PSTHs showing probability of time-locked saccade initiation (as opposed to c-e, measuring co-occurrence of entire saccades). **g-h,** No significant DZ change, but significant MZ time-locking of saccade initiation. **i,** Schematics showing probability of saccades shifting in the same or different direction (angular difference, twin1-twin2). **j-k,** Polar histograms measuring the distribution of differences in saccade directions for DZ and MZ twins, in relation to upper bound (95% CI) of results expected by chance, measured by permutation testing. **l,** Across all comparisons, MZ twins shift saccades in more similar subsequent directions than DZ twins. **m,** Schematics showing probability of fixating on the same semantic content at the same moments. Left: collocated, co-occurring eye and mouth fixations; right: non-collocated, non-co-occurring. **n-o,** Collocated, co-occurring fixations for DZ and MZ twins, plotted as Z scores relative to results expected by chance, measured by permutation testing. **p,** Collocated, co-occurring fixations (diagonals from (n) and (o)); error estimates (s.e.m.) from individual variation.
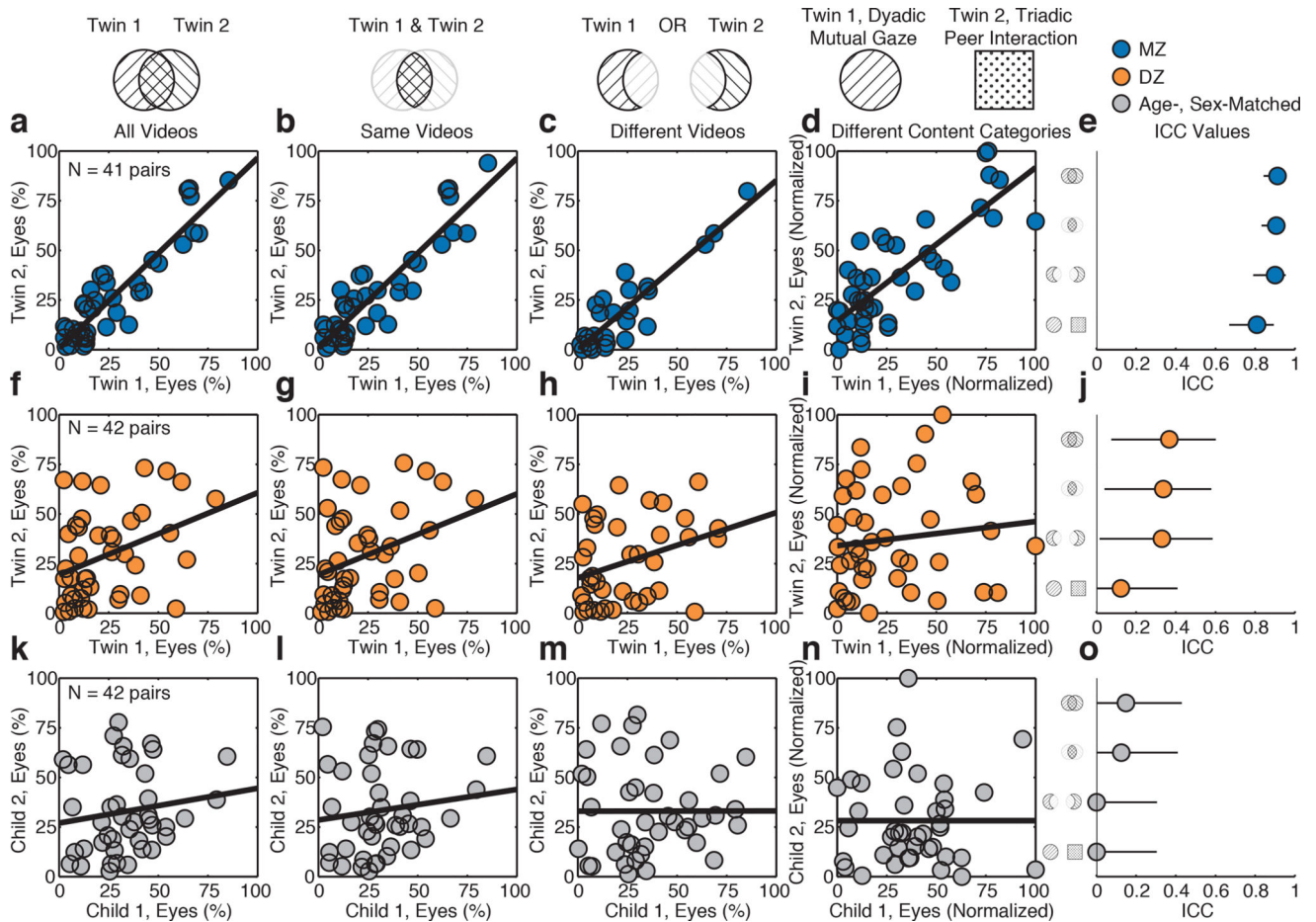
Figure 3. Monozygotic (MZ) twins exhibit high twin-twin concordance in eye-looking, whether watching the same or different video stimuli, evidence of active niche-picking in the goal-directed seeking of social information

a-d, Paired measures of eye-looking in MZ twins for (a) all video stimuli presenting dyadic interaction, (b) measures collected when both twins watched the same dyadic interaction videos, (c) measures collected when each twin watched different dyadic interaction videos, or (d) measures collected when each twin watched different content categories, showing either dyadic caregiver interaction (twin 1) or triadic peer interaction (twin 2). See Extended Data Figure 6 for stimuli examples. **e,** Intraclass correlation coefficients and 95% confidence intervals for a-d. **f-j,** Measures in DZ twins for the same comparisons in a-e. **k-o,** Measures in age- and sex-matched non-siblings for the same comparisons in a-e.
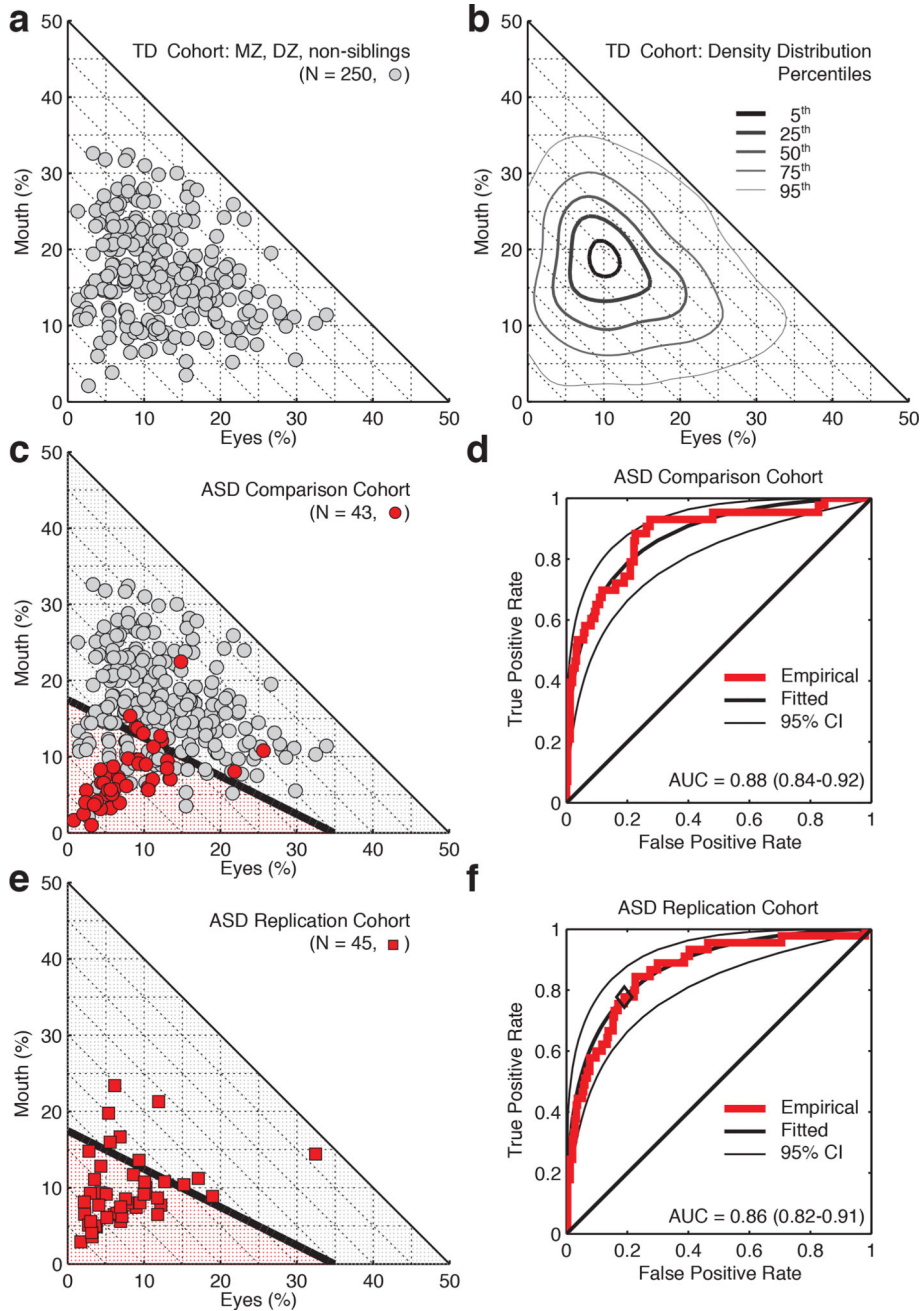
Figure 4. Comparison of social visual engagement in epidemiologically-ascertained toddlers from the general population relative to two cohorts of toddlers diagnosed with autism spectrum disorder (ASD)

a, Raw data marking individual levels of eye- and mouth-looking in 250 epidemiologically-ascertained toddlers watching video scenes of peer interaction. **b,** Population density contours for the data in (a). **c,** Comparison with data from 43 toddlers diagnosed with ASD. Black line marks classification boundary from linear discriminant analysis. **d,** Classification based on individual levels of social visual engagement. Threshold for the receiver operating characteristic (ROC) curve varied by extent of eye- and mouth-looking. **e,** Data from replication cohort of 45 toddlers with ASD. Classification boundary from (c). **f,**

Classification based on individual levels of social visual engagement. As in (e), threshold for the ROC curve varied by extent of eye- and mouth-looking. Black diamond on red empirical ROC marks true positive and false positive rates observed in replication cohort using the optimal threshold identified in (c) and (d).