

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

John N. Constantino^{1,2,3}, Stefanie Kennon-McGill¹, Claire Weichselbaum¹, Natasha Marrus^{1,3}, Alyzeh Haider¹, Anne L. Glowinski¹, Scott Gillespie⁴, Cheryl Klaiman^{5,6}, Ami Klin^{5,6,7} & Warren Jones^{5,6,7}

Long before infants reach, crawl or walk, they explore the world by looking: they look to learn and to engage¹, giving preferential attention to social stimuli, including faces², face-like stimuli³ and biological motion⁴. This capacity—social visual engagement—shapes typical infant development from birth⁵ and is pathognomically impaired in children affected by autism⁶. 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 information⁷. In a series of eye-tracking experiments conducted with 338 toddlers, including 166 epidemiologically ascertained twins (enrolled by representative sampling from the general population), 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 characteristics that are the most highly heritable, preferential attention to eye and mouth regions of the face, are also those that are differentially decreased in children with autism ($\chi^2=64.03$, $P<0.0001$). These results implicate social visual engagement as a neurodevelopmental endophenotype not only for autism, but also for population-wide variation in social-information seeking⁸. In addition, these results reveal a means of human biological niche construction, with

phenotypic differences emerging from the interaction of individual genotypes with early life experience⁷.

Despite evidence that autism is among the most highly heritable neuropsychiatric conditions⁹, with a majority of genetic risk attributable to common (polygenic) factors^{10,11}, its neurobiological mechanisms remain unknown¹². Autism is instead defined behaviourally, by atypical trajectories of social development⁶ that can result in profound impairments in social-communicative function¹³ and poor inclusion into wider societies that are often less than tolerant¹⁴. Atypical social visual engagement is observable within the first six months in infants that are later diagnosed with autism⁶ and continues 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 and peers in social interaction (Extended Data Figure 1a and Methods). We examined pairwise concordance in social visual engagement as a function of zygosity, collecting eye-tracking data from 82 monozygotic twins (41 pairs), 84 dizygotic twins (42 pairs) and 84 non-siblings (42 randomized pairs). We established measurement stability over 15 months, and assessed measurement utility as an endophenotype for social disability, testing 88 toddlers with

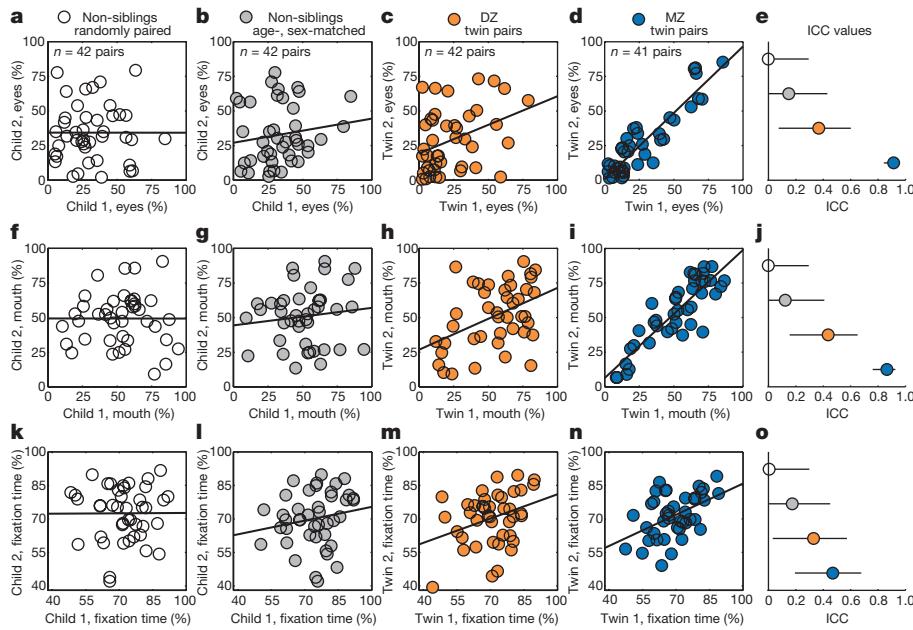


Figure 1 | Monozygotic twins exhibit high twin-twin concordance for eye- and mouth-looking, significantly greater than dizygotic twins or age- and sex-matched non-siblings. **a–d**, Paired measurements of eye-looking in non-siblings paired randomly in 10,000 resamplings without replacement (**a**), age- and sex-matched non-siblings (**b**), dizygotic (DZ) twins (**c**) and monozygotic (MZ) twins (**d**). **e**, Intraclass correlation coefficients (ICCs) 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). Data plotted in **a**, **f**, **k** are representative, selected to match the mean ICC of all 10,000 re-samplings.

¹Department of Psychiatry, Washington University, St Louis, Missouri 63110, USA. ²Department of Pediatrics, Washington University, St Louis, Missouri 63110, USA. ³Intellectual and Developmental Disabilities Research Center, Washington University, St Louis, Missouri 63110, USA. ⁴Pediatric Biostatistics Core, Emory University School of Medicine, Atlanta, Georgia 30307, USA. ⁵Marcus Autism Center, Children's Healthcare of Atlanta, Atlanta, Georgia 30329, USA. ⁶Division of Autism & Related Disabilities, Department of Pediatrics, Emory University School of Medicine, Atlanta, Georgia 30329, USA. ⁷Center for Translational Social Neuroscience, Emory University, Atlanta, Georgia 30329, USA.

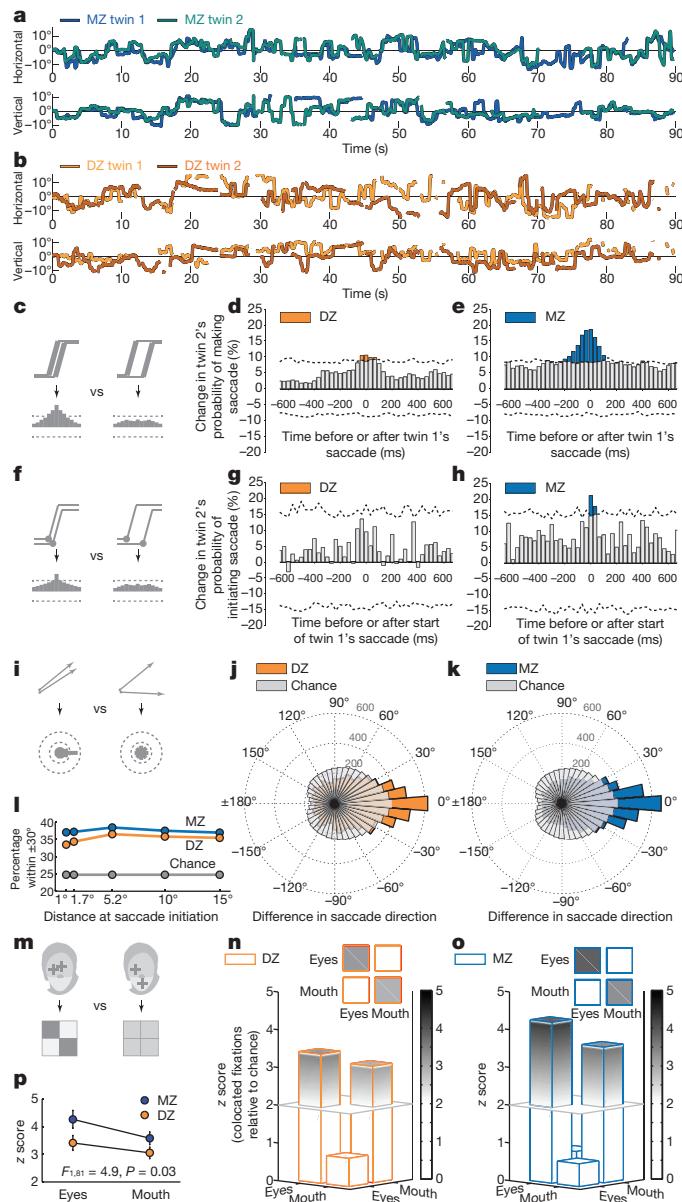


Figure 2 | Monozygotic twins exhibit greater probability of shifting their eyes at the same moments, in the same directions, and onto the same semantic content when viewing scenes of social interaction.

a, b, Example eye position data for monozygotic (**a**) and dizygotic (**b**) twins. Gaps in plots reflect blinks or off-screen fixations. **c**, Schematic peristimulus time histograms (PSTHs) showing probability of co-occurring saccades. Left, if saccades co-occur, the saccade probability of twin 2 increases with saccades of twin 1. Right, if saccades do not co-occur, the probability remains unchanged. **d, e**, There is a small increase in probability of time-locked saccades for dizygotic twins (**d**) compared to a large increase for monozygotic twins (**e**). **f**, Schematic PSTHs showing the probability of time-locked saccade initiation (as opposed to **c–e**, which measure co-occurrence of entire saccades). **g, h**, No significant change for dizygotic twins, but significant time-locking of saccade initiation was seen in monozygotic twins. **d, e, g, h**, Dotted lines show 95% confidence intervals for change expected by chance (that is, no time-locking), measured by permutation testing. **i**, Schematics showing the probability of saccades shifting in the same or different directions (angular difference, twin 1 – twin 2). **j, k**, Polar histograms measuring the distribution of differences in saccade directions for dizygotic and monozygotic twins, in relation to the upper bound (95% confidence interval) of results expected by chance, measured by permutation testing. **l**, Across all comparisons, monozygotic twins shift saccades in more similar subsequent directions than dizygotic twins. **m**, Schematics showing probability of fixating on the same semantic content at the same moment. Left, collocated, co-occurring eye and mouth fixations. Right, non-collocated, non-co-occurring fixations. **n, o**, Collocated, co-occurring fixations for dizygotic and monozygotic 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**); data are shown as mean \pm s.e.m. from individual variation.

in typical development, 18–24 months of age (mean \pm s.d. = 21.3 ± 4.3) months), coinciding with large shifts in language, cognition, and adaptive behaviour²⁰, and affording wide variation in our trait of interest⁶ (see Methods).

We confirmed that groups did not differ significantly in age at the time of testing ($F_{2,247} = 2.3$, $P = 0.10$; monozygotic versus dizygotic twins, $t_{164} = 1.59$, $P = 0.11$); participant demographics (Extended Data Table 1); calibration accuracy, oculomotor function (Extended Data Fig. 2); or percentage of time spent looking at eyes or mouth, or attending to task (Extended Data Fig. 1d–f).

For concordance in eye- and mouth-looking (Fig. 1d, i), monozygotic intraclass correlations (ICC case 2, 1 (ref. 21)) were remarkably high: 0.91 for eyes (95% confidence interval 0.85–0.95) and 0.86 for mouth (95% confidence interval: 0.76–0.92). This contrasted markedly with dizygotic correlations for eyes (0.35 (0.07–0.59)) and mouth (0.44 (0.16–0.65)) (Fig. 1c, h). In non-siblings, correlations did not differ significantly from zero: either when age- and sex-matched (Fig. 1b, g) or when randomly matched (Fig. 1a, f and Extended Data Table 2a). In all groups, within-subject stability (test-retest reliability) was consistently high, indicating ‘trait-like’ stability (Extended Data Fig. 3). These results are consistent with broad heritability of 0.86–0.90 for eye- and mouth-looking²².

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

To test the specificity of these measurements to social engagement, we compared concordance in eye-looking with concordance of two additional indices: time spent looking at non-social content (inanimate objects and/or background) and time spent attending to task (maintaining stable onscreen fixation²³). In monozygotic twins, eye-looking was significantly more concordant than non-social object-looking—eyes,

autism spectrum disorder 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 Fig. 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 response¹⁷ (responding to specific features of the exact stimulus presented) or goal-directed action¹⁸ (individual differences in seeking social information).

Trait heritability was estimated according to the classic twin design, with epidemiologically ascertained cohorts of monozygotic and dizygotic twins, together with a cohort 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 in sex and age (mean \pm s.d. age difference: 0.99 ± 0.27 days) and randomly matched in 10,000 re-samplings without replacement¹⁹. We restricted analyses to same-sex twin pairs (inclusion of opposite-sex dizygotic pairs yielded either no change or accentuation of monozygotic–dizygotic differences). For age at time of testing, we selected a dynamic period

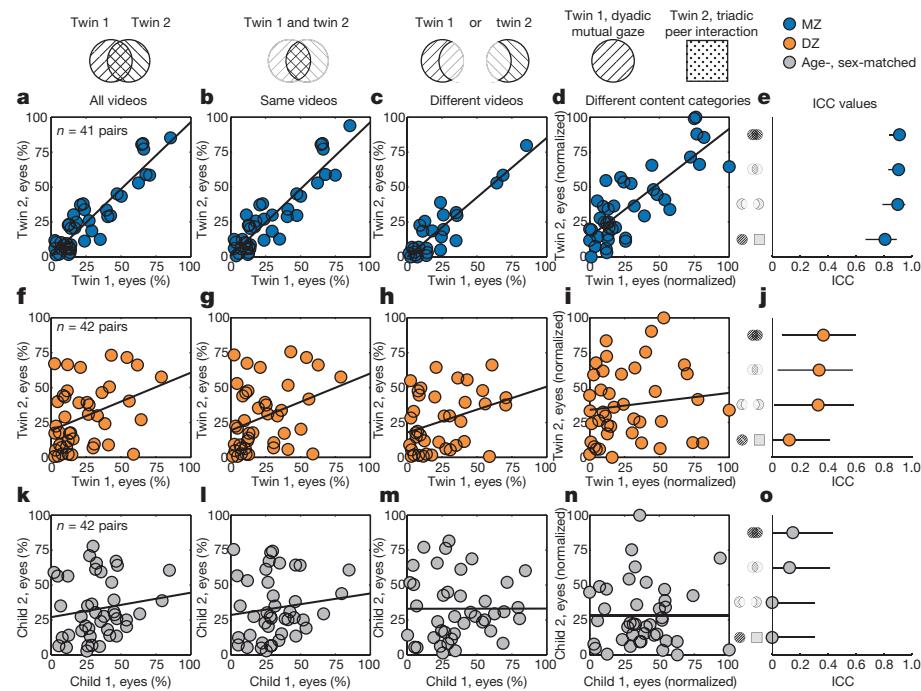


Figure 3 | Monozygotic twins exhibit high twin-twin concordance in eye-looking, whether watching the same or different video stimuli, evidence of active niche-picking in the goal-directed seeking of social information. a–d, Paired measurements of eye-looking in monozygotic twins for all video stimuli presenting dyadic interaction (a), measurements when both twins watched the same dyadic interaction videos (b), measurements collected when each twin watched different dyadic

interaction videos (c), or measurements collected when each twin watched different content categories, showing either dyadic caregiver interaction (twin 1) or triadic peer interaction (twin 2) (d). See Extended Data Fig. 6 for stimuli examples. e, Intraclass correlation coefficients (ICCs) and 95% confidence intervals for a–d. f–j, Measurements in dizygotic twins for the same comparisons in a–e. k–o, Measurements in age- and sex-matched non-siblings for the same comparisons in a–e.

0.91 (0.85–0.95) versus object, 0.66 (0.46–0.80)—and more concordant than time spent attending to task, 0.46 (0.19–0.67) (Fig. 1n). By contrast, in dizygotic 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) (Fig. 1c, h, m). Similarly, in age- and sex-matched non-siblings, all ICC estimates overlapped (Fig. 1b, g, l). While heritable effects of domain general visual attention are likely to be observable in other contexts, these analyses indicate effects that are differentially social (Fig. 1e, j, o).

In the next experiment, we measured moment-by-moment, micro-level concordance (Fig. 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 analysed concordance in timing of eye movements (Fig. 2c). Data in the Supplementary Videos (brief videos of twin eye-tracking data (monozygotic and dizygotic twin gaze data while viewing scenes of dyadic mutual gaze stimuli and triadic peer interaction stimuli)) provide an immediately appreciable sense of moment-to-moment monozygotic concordance—weakened substantially in dizygotic twins—when viewing social scenes. Number and rate of eye movements did not differ significantly by group (approximately 1,944 fixations per child: mean \pm s.d. rates of 1.66 ± 0.59 fixations per second for dizygotic and 1.66 ± 0.49 fixations per second for monozygotic twins; all $t < 0.65$, $P > 0.20$). However, monozygotic twins demonstrated greater probability of moving their eyes at the same times: for each saccadic eye movement by twin 1 (rapid eye movement between fixations), within 350 ms, there was an 18.6% increase in twin 2's probability of also making an eye movement (Fig. 2d, e). More surprisingly, when analyses were restricted to moments of motor initiation of a saccade (Fig. 2f), we observed a 21.1% increase in probability of

time-locked eye movements: within ± 16.7 ms, monozygotic twins, but not dizygotic twins, initiated saccades at the same moments (Fig. 2g, h). These results suggest that monozygotic 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 gaze²⁴.

We next tested for concordance in the direction of eye movements (Fig. 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, but also 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° (that is, 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 within 15° ('collocated' in only the broadest sense of looking at the same hemi-field of the presentation monitor). Across all comparisons (Fig. 2l), monozygotic twins were more likely than dizygotic 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 (Fig. 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 (Fig. 2n, o), monozygotic twins exhibited greater concordance than dizygotic twins ($F_{1,81} = 4.89$, $P = 0.030$; Fig. 2p).

In summary, monozygotic 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 on the same semantic content. These high levels of monozygotic concordance, observed at both macro- and micro-levels, indicate a strong biological basis for variation in social visual engagement⁷, with a substantial portion of that variation attributable to additive genetic influence. While concordance in micro-level characteristics was more modest than in macro-level characteristics, even modest micro-level concordance marks repeatable shifts in probability: repetition of these shifts—recurring as frequently as every 400–500 ms—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 response¹⁷ or goal-directed action¹⁸. This distinction is intriguing because it relates to what aspects of social behaviour may be more or less phylogenetically conserved: have evolutionary pressures favoured biological systems that rely on specific responses to particular features of external stimuli (in the manner of feature detectors²⁵), or have evolutionary pressures favoured 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 a total duration that was longer than some toddlers' willingness to sit), each twin saw a separate set of videos, the majority of which were the same (mean \pm s.d. = $86.4 \pm 19.3\%$) but some of which were different ($13.6 \pm 19.3\%$), and were 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 measurements made only when both twins watched the same videos; the null hypothesis held that concordance would be equal ($ICC_{\text{sameVideos}} = ICC_{\text{allVideos}}$), the alternative stated that concordance would be greater ($ICC_{\text{sameVideos}} > ICC_{\text{allVideos}}$). Greater concordance when watching the same videos would indicate stimulus response (more concordant responding given the exact same stimulus). For the second and third tests, analyses were restricted to measurements made only when each twin watched different videos or videos with different content categories; the null hypothesis held that concordance would be zero ($ICC_{\text{differentVideos}} = 0$ and $ICC_{\text{differentContent}} = 0$), the alternative stated that concordance would be greater than zero ($ICC_{\text{differentVideos}} > 0$ and $ICC_{\text{differentContent}} > 0$). 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-picking⁷). For each test, we measured levels of eye-looking and quantified physical image properties of all eyes stimuli²⁸ (Extended Data Fig. 7).

In the first test, we were unable to reject the null hypothesis: concordance when watching the same videos was no greater than concordance when watching all videos (Fig. 3b, g, l). In the second test, however, when each twin watched different videos, monozygotic and dizygotic concordance was significantly greater than zero (Fig. 3c, h, m). And in the third test, when each twin watched different content categories, monozygotic concordance was significantly greater than zero but dizygotic concordance was not (Fig. 3d, i, n). These results suggest that the relevant biological mechanisms are more likely to

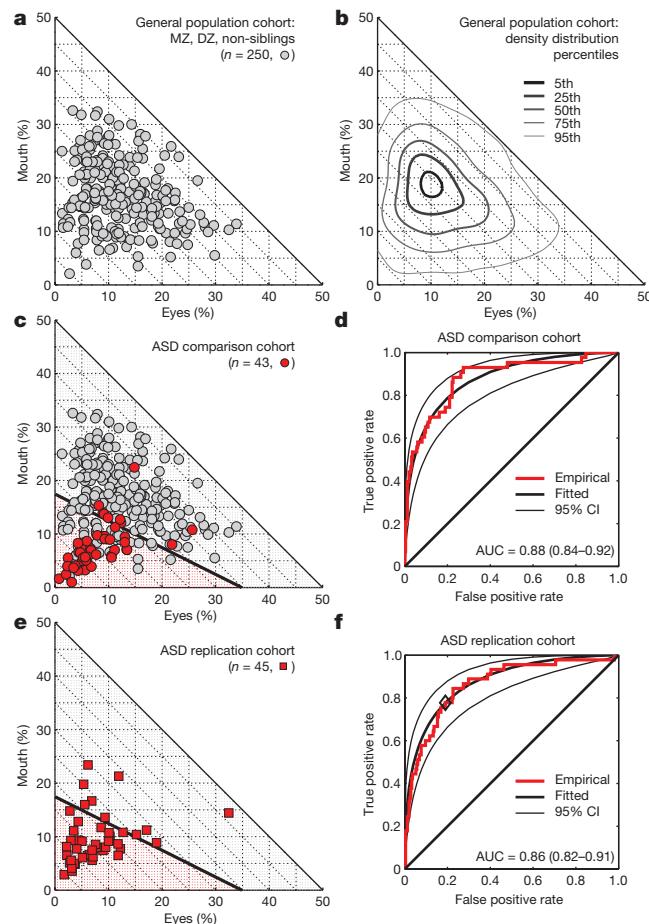


Figure 4 | Comparison of social visual engagement in epidemiologically ascertained toddlers from the general population relative to two cohorts of toddlers diagnosed with autism spectrum disorder.

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 autism spectrum disorder (ASD). The 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. CI, confidence interval. e, Data from the replication cohort of 45 toddlers with ASD. Classification boundary as in c. f, Classification based on individual levels of social visual engagement. As in d, 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.

relate to systems subserving goal-directed seeking and valuation of social information.

Finally, to directly assess the functional significance of these measurements, we compared the above results with data from two independent cohorts of toddlers with autism spectrum disorder (Fig. 4): one primary comparison cohort, $n = 43$, and a second replication cohort, $n = 45$ (all consecutive referrals for diagnostic evaluation). In toddlers with autism spectrum disorder, the same measurements of social visual engagement that were most highly heritable—eye- and mouth-looking—are markedly reduced (Fig. 4a–c), providing a robust index of diagnostic membership (Fig. 4d–f; area under the curve (AUC) = 0.88 (95% confidence interval 0.84–0.92) for the primary comparison cohort and AUC = 0.86 (0.82–0.91) for the replication cohort).

Taken together, 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 environment²⁹. Similar notions have been advanced in phenotypic studies contrasting the experiential development of children with autism and their typically developing peers³⁰, yet, to our knowledge, this has never before been 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 also for population-wide variation in goal-directed seeking and valuation of social information.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 September 2016; accepted 25 May 2017.

Published online 12 July 2017.

1. Gibson, E. J. Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annu. Rev. Psychol.* **39**, 1–41 (1988).
2. Valenza, E., Simion, F., Cassia, V. M. & Umiltà, C. Face preference at birth. *J. Exp. Psychol. Hum. Percept. Perform.* **22**, 892–903 (1996).
3. Goren, C. C., Sarty, M. & Wu, P. Y. K. Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics* **56**, 544–549 (1975).
4. Simion, F., Regolin, L. & Buff, H. A predisposition for biological motion in the newborn baby. *Proc. Natl Acad. Sci. USA* **105**, 809–813 (2008).
5. Simion, F., Leo, I., Turati, C., Valenza, E. & Dalla Barba, B. How face specialization emerges in the first months of life. *Prog. Brain Res.* **164**, 169–185 (2007).
6. Jones, W. & Klin, A. Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature* **504**, 427–431 (2013).
7. Scarr, S. & McCartney, K. How people make their own environments: a theory of genotype > environment effects. *Child Dev.* **54**, 424–435 (1983).
8. Constantino, J. N. & Charman, T. Diagnosis of autism spectrum disorder: reconciling the syndrome, its diverse origins, and variation in expression. *Lancet Neurol.* **15**, 279–291 (2016).
9. Constantino, J. N. et al. Autism recurrence in half siblings: strong support for genetic mechanisms of transmission in ASD. *Mol. Psychiat.* **18**, 137–138 (2013).
10. Gaugler, T. et al. Most genetic risk for autism resides with common variation. *Nat. Genet.* **46**, 881–885 (2014).
11. Robinson, E. B. 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* **68**, 1113–1121 (2011).
12. Geschwind, D. H. & State, M. W. Gene hunting in autism spectrum disorder: on the path to precision medicine. *Lancet Neurol.* **14**, 1109–1120 (2015).
13. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* 5th edn (American Psychiatric Association, 2013).
14. Magiati, I., Tay, X. W. & 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.* **34**, 73–86 (2014).
15. Shultz, S., Klin, A. & Jones, W. Inhibition of eye blinking reveals subjective perceptions of stimulus salience. *Proc. Natl Acad. Sci. USA* **108**, 21270–21275 (2011).

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* **59**, 809–816 (2002).
17. Pearce, J. M. & Bouton, M. E. Theories of associative learning in animals. *Annu. Rev. Psychol.* **52**, 111–139 (2001).
18. Dickinson, A. Actions and habits: the development of behavioural autonomy. *Phil. Trans. R. Soc. Lond. B* **308**, 67–78 (1985).
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 Vol. 2 Cognitive Processes* 7th edn (John Wiley & Sons, 2015).
21. McGraw, K. O. & Wong, S. P. Forming inferences about some intraclass correlations coefficients. *Psychol. Methods* **1**, 30–46 (1996).
22. Jacquard, A. Heritability: one word, three concepts. *Biometrics* **39**, 465–477 (1983).
23. Leigh, R. J. & Zee, D. S. *The Neurology of Eye Movements* (Oxford Univ. Press, 2006).
24. Schall, J. D. & Thompson, K. G. Neural selection and control of visually guided eye movements. *Annu. Rev. Neurosci.* **22**, 241–259 (1999).
25. Marr, D. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (W. H. Freeman, 1982).
26. Treue, S. Visual attention: the where, what, how and why of saliency. *Curr. Opin. Neurobiol.* **13**, 428–432 (2003).
27. Hopfinger, J. B., Buonocore, M. H. & Mangun, G. R. The neural mechanisms of top-down attentional control. *Nat. Neurosci.* **3**, 284–291 (2000).
28. Wang, S. et al. Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking. *Neuron* **88**, 604–616 (2015).
29. Oyama, S. *Evolution’s Eye: A Systems View of the Biology–Culture Divide* (Duke Univ. Press, 2000).
30. Klin, A., Jones, W., Schultz, R. & Volkmar, F. The enactive mind, or from actions to cognition: lessons from autism. *Phil. Trans. R. Soc. Lond. B* **358**, 345–360 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the families and children for their participation. Research was supported by grants from the National Institute of Child Health & Human Development, HD068479 (J.N.C.) and U54 HD087011 (Intellectual and Developmental Disabilities Research Center at Washington University, J.N.C., principal investigator); and by the National Institute of Mental Health, MH100019 (N.M.) and MH100029 (A.K., W.J.). Additional support was provided by the Marcus Foundation, the Whitehead Foundation and the Georgia Research Alliance. Epidemiologic ascertainment of twins was made possible by the Missouri Family Register, a joint program of Washington University and the Missouri Department of Health and Senior Services; authorization to access was approved by the MO DHSS Institutional Review Board (S. Ayers, Chair) under auspices of the project entitled Early Quantitative Characterization of Reciprocal Social Behavior. We thank E. Mortenson, S. Sant, T. Gray, Y. Zhang, L. Campbell, L. Malik, A. Khan and E. McGarry for data collection and analysis; A. C. Heath and A. Agrawal for discussions of data analysis and statistics; C. Gunter for helpful comments on the manuscript; C. Drain and D. Hopper for project coordination and data collection; D. Jovanovic and R. Todorovic for contributions to twin family ascertainment; M. Panther for administrative support; and S. Kovar, J. Paredes, and M. Ly for designing and building the eye-tracking laboratory.

Author Contributions J.N.C., A.L.G., A.K. and W.J. developed the initial idea and study design. J.N.C. and W.J. had full access to all data and take responsibility for data integrity and accuracy of analyses. J.N.C. supervised participant characterization. W.J. supervised technology development, data acquisition and analysis. S.K.-M., C.W., N.M. and A.H. collected data, ensured quality control at Washington University, conducted sub-analyses and participated in manuscript writing and revision. S.G., C.K. and W.J. performed data processing at Emory, ensured quality control across sites and participated in manuscript revision. W.J., A.K. and J.N.C. interpreted data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to W.J. (warren.jones@emory.edu) or J.N.C. (constantino@wustl.edu).

Reviewer Information *Nature* thanks R. Adolphs, J. P. McCleery, C. Nelson and the other anonymous reviewer(s) for their contribution to the peer review of this work.

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 before assessment. Children were shown video scenes of naturalistic caregiver and peer interaction. We measured the percentage of visual fixation time to eyes, mouth, body, and object regions (experiment 1—macro-level measurements of social visual engagement) as well as moment-to-moment variation in timing, direction, and location of eye movements (experiment 2—micro-level measurements of social visual engagement). Visual scanning was measured with eye-tracking equipment (ISCAN, Inc.). 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 and related variables). 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 before each assessment. A total of 414 children participated (242 twins, 84 non-sibling comparison children, and 88 children diagnosed with ASD).

The twin cohort 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 ref. 31. 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 through 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 (<https://www.census.gov/quickfacts/>)).

On the basis of the 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 Behaviour study (the larger study, described in ref. 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-laboratory eye-tracking component of the study. For these 121 twin pairs, 242 individual eye-tracking data collection sessions were conducted. Cohort demographics of the entire epidemiologically ascertained cohort ($n = 330$) 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 (all 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 a slightly higher proportion than the general population of families in the highest wage-earning bracket ($\chi^2 = 1.55$, $P = 0.21$).

Owing to the paired nature of planned eye-tracking analyses (requiring complete sets of eye-tracking data from both twins), and the restriction of analyses to same-sex dizygotic twin pairs, the eye-tracking twin cohort comprised 166 children (83 twin pairs, 41 monozygotic and 42 same-sex dizygotic twins). Demographics data for these children are given in Extended Data Table 1. Descriptions of eye-tracking data quality control and pairing procedures are can be found in 'Quality control' and 'Pairing of participant data'. Mean age at time of testing in these 166 twins and in the non-sibling control cohorts ($n = 84$) was 21.3 months (s.d. = 4.26 months). By group, ages and sexes were as follows: non-sibling controls, mean \pm s.d. = 20.87 ± 2.77 months, 52.4% male; dizygotic twins, 22.13 ± 4.89 months, 52.4% male; and monozygotic 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 measurements. Children in each of the two ASD groups met the following inclusionary criteria: (1) criteria for autistic disorder or ASD on the Autism Diagnostic Observation Schedule³², Module 1; and (2) a diagnosis of either autistic disorder (32 out of 43 children in cohort 1), pervasive developmental disorder-not otherwise specified (11 out of 43 children in cohort 1), or ASD (45 out of 45 children in Cohort 2) by two experienced clinicians upon independent review of all available clinical data,

including standardized testing and a video of the diagnostic examination. At the time of testing of ASD cohort 1 (2011–2013, as in the twin cohort), diagnostic guidelines followed DSM-IV-TR criteria³³; all children would also meet criteria for ASD per current, DSM-5 criteria¹³. At the 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 months for ASD cohort 1 and 25.8 ± 3.4 months 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 cohort ($t_{291} = 2.78$, $P < 0.001$ 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 ($r_{\text{eyes}} = 0.01$, $P = 0.95$, $r_{\text{mouth}} = -0.08$, $P = 0.61$ for cohort 1; $r_{\text{eyes}} = -0.01$, $P = 0.92$, $r_{\text{mouth}} = -0.12$, $P = 0.41$ for cohort 2), and constraining analyses to ASD subcohorts age-matched to the twin and non-sibling typically developing cohorts did not significantly change the area under the ROC curves in Fig. 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 cases³⁴. The questionnaire was carried out during a phone interview with the biological mother or father of the twins. 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 six 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. Measurements of eye movement were made directly by video-oculography for each child, collected in a semi-automated fashion (by an experimenter using automated data collection software), and analysed 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 measurements of social visual engagement benefit from the relative absence of rater/observational biases that have elsewhere been cited as a potential confounding factor in twin studies (for example in ref. 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 refs 6, 36. Eye-tracking was accomplished by a video-based, dark pupil/corneal reflection technique with hardware and software created by ISCAN, Inc. The system was remotely mounted within a wall panel beneath the stimuli presentation monitor, concealed from the view of the child by an infrared filter.

Children were led into the testing room one at a time while an age-appropriate video for children 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 one-third 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-inch (50.8-cm) computer monitor with a refresh rate of 60 Hz. Lights in the room were dimmed in order to direct the toddler's attention towards 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 ($>3^\circ$ of visual angle) in calibration accuracy occurred. Please see 'Quality control' for measurements of calibration accuracy and concordance thereof.

Stimuli. Following calibration and verification, 27 videos, in a randomized presentation order, were shown to each child. Each video lasted an average of 44.2 s, for a total viewing time of 19 min 54 s. Videos comprised two content categories designed to recapitulate naturalistic social experience (as also described in refs 6, 15). The first category of video displayed an adult female actor who spoke directly to the viewer/camera, as shown in Extended Data Fig. 1a, 7a and Supplementary Videos 1, 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 childcare setting ('triadic peer interaction', 12 videos total, shown in Extended Data Fig. 7*i* and Supplementary Videos 3, 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 Figs 1, 6, 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 Fig. 1*d, e* for dyadic mutual gaze and Extended Data Fig. 7*d, e* for triadic peer interaction). Where cross-category comparisons are made (Fig. 3), normalization is required (described in '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 (Figs 1, 3 and Extended Data Figs 1, 3, 4, 5, dyadic; Fig. 4 and Extended Data Figs 6, 7, triadic). Other analyses (micro-level measurements in Fig. 2, controls in Extended Data Fig. 2) require no normalization and summarize results for all stimuli.

Videos were presented as full-screen audiovisual stimuli; in 32-bit colour; 640 × 480 pixels in resolution; at 30 frames s⁻¹; with mono-channel audio sampled at 44.1 kHz. Stimuli were sound and luminosity equalized, and were piloted in an independent cohort of children before the start of the 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° s⁻¹ (ref. 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 500 Hz (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 ref. 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 Figs 1*b*, *c*, 6*b*, *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 Table 3*c, d* gives 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 mean ± s.d. = 18.2 min (3.1 min) for monozygotic twins and 17.9 min ± 3.4 min for dizygotic twins.

Quality control. For 10 out of the 242 twin data collection sessions (4.13%), data could not be collected owing to the following: child fussiness, child sleep, and/or temporary equipment failure. In 16 out 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 (that is, 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 out 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.

The average calibration accuracy for all groups was less than 1° of visual angle. Extended Data Figure 2*a–c* shows the total variance in calibration accuracy, and Extended Data Figure 2*d–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 monozygotic versus dizygotic twins. The position was measured as horizontal and vertical fixation location relative to the centre of the calibration accuracy validation target (in degrees of visual angle), whereas the distance was measured from fixation location to the centre 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 monozygotic (MZ), dizygotic (DZ), or non-sibling controls for

fixation position (Extended Data Fig. 2*g–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 monozygotic compared to dizygotic twins (that is, ballistic muscle movements of the eyes might be incrementally more similar in monozygotic than dizygotic twins). Group results for monozygotic twins, 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 1,600 pairs of monozygotic twins would be required to reject or confirm its existence (80% power, $\alpha = 0.05$). 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 measurements, 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 3*a*). Such an effect could not, in and of itself, account for the large differences in concordance for monozygotic versus dizygotic twins 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-laboratory 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 out of 242 mentioned above) were given scores of 0; poor quality sessions (the 16 out 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, which did not change the 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 cohort of eye-tracking data (207 children, aged 16.5–30 months; with a threshold that was identified in that cohort by analysing the full set of fixation percentages as two separable components of a finite mixture model).

In the present dataset for monozygotic and dizygotic 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 monozygotic and dizygotic 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 rank sum test/Mann–Whitney *U*-test with comparable results (no significant differences in any analysis)). We set no threshold for the 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 monozygotic twins was 24.8 ± 3.7 and for dizygotic twins was 24.0 ± 4.7 (data given as mean ± s.d.). The mode number of included trials per child for both monozygotic and dizygotic 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 monozygotic twin and 4 for one dizygotic twin.

Pairing of participant data. Owing to the paired nature of planned analyses, only twin pairs having complete eye-tracking datasets from both twins could be analysed. Of the 121 total twin pairs enrolled in the eye-tracking study, 96 pairs (192 children) had complete datasets. Twenty-five twin pairs had missing data (24 pairs missing one twin's data, 1 pair missing both). In 10 out 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 (that is, 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 monozygotic and dizygotic

twins: 5 out 15 (33.3%) were monozygotic (2f:3m) and 5 out of 15 (33.3%) were same-sex dizygotic (1f:4m); the remaining 5 out of 15 were opposite-sex dizygotic twins and, like all opposite-sex dizygotic twins, were not a part of the analyses.

For analyses of pairwise concordance in eye-tracking measurements, the final set of children with successfully collected, paired data consisted of $n=41$ monozygotic twin pairs (82 children), $n=42$ dizygotic same-sex twin pairs (84 children), and $n=42$ age- and sex-matched non-sibling comparison children (84 children). Analyses were also conducted for the dizygotic opposite-sex twin pairs ($n=17$ pair, 34 children); inclusion of these children increased the monozygotic–dizygotic differences in all cases (that is, dizygotic 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 dizygotic same-sex twins.

For the 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 one day of mean chronological age (average difference in age: mean \pm s.d. = 0.99 ± 0.27 days; average absolute difference in age: mean \pm s.d. = 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 measurements of social visual engagement. Our previous work using this same experimental paradigm⁶ shows evidence that these behaviours, as with many others that emerge in early development (for example, walking and talking), undergo progressive changes that are very sensitive to a child's developmental stage and may be sensitive to sex effects (for example, precociousness in female verbal abilities³⁸). 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 cohort.

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 permutations¹⁹. The weak—but non-zero—ICC values in age- and sex-matched non-sibling controls (main text and Fig. 1b, g, l) 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 (Fig. 1a, f, k) to age- and sex-matched non-sibling controls to dizygotic twins to monozygotic twins—suggests that such an effect may exist.

Data analysis and statistics. As noted above in 'Data acquisition and processing', 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 Fig. 1b, c). Supplementary Videos 1–4 show examples of coded data.

Macro-level indices of social visual engagement. In experiment 1, we measured macro-level indices of social visual engagement, calculating the proportion of time spent looking at eyes, mouth, and body regions. Percentage of total time spent attending to video stimuli (Extended Data Fig. 1f), as well as time spent fixating specifically on eyes (Extended Data Fig. 1d), mouth (Extended Data Fig. 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 (Fig. 1), and ICCs calculated. Because negative values of the ICCs 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 work⁶, the 18–24-month developmental period in which testing was conducted corresponds with large changes in typical infant eye- and mouth-looking, with the 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 analyses, with no appreciable difference in results: we compared correlations for monozygotic and dizygotic twin pairs using both Spearman's rank correlation and ICC. Non-parametric results were as follows: eyes, monozygotic $\rho=0.843$ ($P<0.001$) and dizygotic, $\rho=0.333$ ($P=0.031$); and mouth, monozygotic, $\rho=0.822$ ($P<0.001$) and dizygotic, $\rho=0.405$ ($P=0.008$).

Likewise, macro-level measurements of eye- and mouth-looking differ, as expected, by video content category (Extended Data Figs 1d–f, 6d–f). For this reason, analysis of concordance in levels of looking across different content categories, as undertaken in Fig. 3d, i, n requires normalization (analysis using Pearson's correlation would, of course, be unaffected by such differences, but analyses of agreement and consistency, as is the case for the ICCs, are affected by differences in scale). To analyse measurements 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 Fig. 3d, 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 that 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 measurements 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° s^{-1} of eye movement²³). Interestingly, in monozygotic twins, eye-looking (ICC (95% confidence interval) = 0.91 (0.85–0.95)) was significantly more concordant than nonsocial object-looking (0.66 (0.46–0.80)) and more concordant than time spent maintaining steady fixation (0.46 (0.19–0.67)). By contrast, mouth-looking (0.86 (0.76–0.92)) was more concordant in monozygotic 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 interaction⁴¹.

Measurements of trait-like stability. To measure the extent of trait-like stability of these behaviours, we measured within-subject stability (test-retest reliability) across both short and long timescales. For short timescales, results are plotted in Extended Data Fig. 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 analyses reflect trait-like stability for any given individual during single-day testing sessions, quantified by ICCs with a two-way random effects model, ICC(2,1). Another related analysis, which is not plotted, is that of inter-individual variation—the reliability of measured differences between any individuals A and B (that is, the stability with which the measured trait is higher or lower in individual A than individual B, individual C, and so on, given a series of repeated measurements). In that case, the observed ICC values are in the order of 0.9 for each group, quantified in that case by a fixed rather than random effects model, ICC(3,k). Both analyses 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 analyse data for $n=22$ monozygotic twins (11 pairs, age at time 1, mean \pm s.d. = 21.1 ± 2.6 months, age at time 2, 36.9 ± 2.6 months) and for $n=44$ dizygotic twins (22 pairs, age at time 1, mean \pm s.d. = 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 analysed these data in three ways: (1) twin–twin concordance of measurements at time 2 alone (Extended Data Fig. 4a–l); (2) within-subject stability from time 1 until time 2 (Extended Data Fig. 5a–e); and (3) twin–twin concordance from time 1 until time 2 (Extended Data Fig. 5f–j).

For the first comparison (twin–twin concordance of measurements at time 2 alone), results show robust monozygotic twin–twin concordance at time 2 alone relative to diminished dizygotic twin–twin concordance (similar to results observed at time 1 alone). Results are plotted in Extended Data Fig. 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 Fig. 5a–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 monozygotic twins (95% confidence interval: 0.44–0.87) and 0.69 for dizygotic twins (95% confidence interval: 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 Fig. 5f–j and Extended Data Table 2c), the results differ strongly 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 monozygotic twins is 0.70 (0.40–0.86), whereas for dizygotic 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 monozygotic and 0.07 (0.00–0.36) for dizygotic.

Taken together, 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; monozygotic twin–twin concordance is preserved over 15 months of time and substantially

contrasts with dizygotic 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 response¹⁷ (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 Fig. 3).

To quantify differences in the physical image properties of stimuli seen by each twin, we analysed image property profiles of regions demarcated as eyes across all frames of all videos presented. Specifically, we analysed the lightness and colour (colour opponency in red–green and blue–yellow following the CIE 1976 model⁴²), contrast (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 Fig. 7. Variation in stimulus image properties can be seen in the histograms themselves (Extended Data Fig. 7c–h, m–r) as well as in the statistical comparisons of video image property profiles (Extended Data Fig. 7i, j, 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 property²⁸, 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 Fig. 3 and Extended Data Fig. 7 show that concordance in eye-looking is strongly preserved in monozygotic twins despite watching different stimuli: in monozygotic twins, the extent to which twin 1 looks at the eyes in dyadic mutual gaze videos (examples can be found in Supplementary Videos 1, 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, 4); ICC = 0.81 (95% confidence interval: 0.67–0.89). This effect persists despite the fact that eyes found in the triadic peer interaction videos differ substantially in lightness, colour, contrast, orientation gradients, and motion. These eyes are half to a quarter 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 dizygotic 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% confidence interval: 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 (for example, there would be no expectation that measurements of social visual engagement should remain consistent across entirely non-social stimuli). Instead, we take the present analyses as an indication that what is heritable does not appear to be a response to a particular physical feature per se (that is, response to a single feature found within a highly uncertain visual world); rather, the evidence indicates that what is conserved is an adaptive action: behavioural 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 environments⁵³.

Micro-level indices. In monozygotic and dizygotic twins, we collected 322,672 fixational eye movements (dizygotic twins, 161,963; monozygotic twins, 160,709; approximately 1,944 fixations per child), occurring at a rate of 1.66 fixations per second (dizygotic mean \pm s.d. = 1.66 ± 0.59 fixations per second; monozygotic mean \pm s.d. = 1.66 ± 0.49 fixations per second), each lasting an average of 514 ms (dizygotic = 523 ± 188 ms; monozygotic = 505 ± 245 ms). As a function of zygosity, there were no significant between-group differences in fixation count, frequency, or duration (tested by two-sample *t*-test, all $P > 0.594$, all $t_{164} < 0.534$). Saccadic amplitude data are given in Extended Data Fig. 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 therefore excluded.) We analysed 133,582 saccadic eye movements (dizygotic: 68,262; monozygotic: 65,320; approximately 804 saccades analysed 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 (micro-level measures), 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 analysed the time series eye movement data in terms of timing of saccades and timing of saccade initiation using peristimulus (or ‘peri-event’) time histograms (PSTHs⁵⁴).

Following methods detailed in ref. 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 ms bins in a surrounding 1,333.3 ms window. Bin counts were computed for each twin pair and then averaged across all pairs to obtain group means (plotted in Fig. 2d, e, g, h).

To test whether observed changes in saccade probability differed from those expected by chance, we used permutation testing^{19,55}. In each of 1,000 iterations, the binary time-series saccade data for each twin (0 = not saccading, 1 = saccading) were permuted by circular shifting⁵⁶, following the equation:

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

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 $0 \leq t \leq T$; 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 datasets. 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 1,000 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 (that is, 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 together, this approach enabled the comparison of actual patterns of saccading to randomized, chance patterns, and also allowed us to test the null hypothesis that dizygotic or monozygotic twins demonstrated no greater than chance levels of time-locking of eye movements. Results in Fig. 2c–h show significant time-locking in monozygotic twins, to within ± 16.67 ms 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 muscles²³. Cranial nerves III, IV, and VI supply these muscles²³, whose afferent connections are in turn supplied by the reticular formation in the brainstem^{57,58}. Synapsing directly upon the reticular formation are projections from the frontal eye fields⁵⁹. With so few synapses separating frontal eye fields from the extraocular muscles²⁴, spontaneous time-locking of eye movements suggests the likely presence of some, even modest, degree of time-locked neural activity in stages before motor movement initiation. Given the present behavioural results, it is intriguing to speculate on the extent of possible concordance in activity of neural systems that play a role in saccadic eye movements²⁴ (frontal eye fields, supplementary eye fields, parietal eye fields, area 22, DLPFC).

Direction of eye movements. In experiment 2 (micro-level measures), we measured twin–twin concordance in the 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: $\theta_{\text{twin1}} - \theta_{\text{twin2}}$. Polar histograms of twin–twin differences are plotted in Fig. 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 (that is, 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 66.7 ms, 133 ms, 250 ms, 500 ms) and degree of collocation (retinal eccentricities of 1°, 1.7°, 5.2°, 10°, 15°). Fig. 2j, k plot results for saccades starting from fixations collocated within 5.2° (at least partially overlapping foveas) and co-occurring within 500 ms or less. Fig. 2l plots results across varying degrees of collocation, also co-occurring within 500 ms or less. Across all comparisons of varying retinal eccentricities and temporal windows, monozygotic twins were more likely than dizygotic 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 1,000 iterations, computing the angular difference across all randomly paired saccades in all iterations. The polar histogram data plotted as grey bars in Fig. 2j, k shows the upper 95th percentile of differences expected by chance alone across all 1,000 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 to be due to the nature of the video content and the effect of that content on probable saccade direction (that is, 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 monozygotic and dizygotic twins show an increase in the probability of moving their eyes in a shared direction, also summarized in Fig. 2l, monozygotic twins exhibit greater probability of shifting saccades in more similar subsequent directions.

Collocation of contemporaneous fixations. 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 2×2 contingency tables of co-occurring fixations (Fig. 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 1,000 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 (Fig. 2n, o), but monozygotic twins exhibit greater concordance than dizygotic twins ($F_{1,81} = 4.89$, $P = 0.030$; Fig. 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 Fig. 2n, o) is greater for monozygotic than dizygotic twins: monozygotic 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 the 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 behaviour^{60–63}. Analyses indicated that twin pair cohorts of 40 or greater would provide 80% power to detect correlations of approximately $r = 0.38$ (approximately half the magnitude of monozygotic correlations observed in ref. 31). Actual statistical power to detect concordance between two measurements depends not only on the true genetic correlation (r) between them, but also on their marginal heritabilities (H^2): When $H^2 = 50\%$, power is above 80% when $r \geq 0.27$; and when $H^2 = 20\%$ (lower than anticipated from the existing literature), power is above 80% when $r \geq 0.55$ ($\alpha = 0.001$). Given the size of the observed monozygotic and dizygotic concordance effects, measurement estimates of our achieved power ($1 - \beta$ error probability) for monozygotic eye-looking was approximately 1; in dizygotic twins, achieved power for eye-looking was 0.77. Future work will follow-up in larger cohorts.

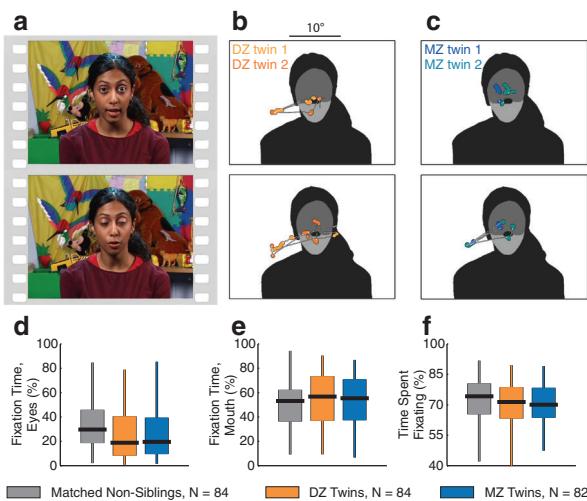
Additionally, in our final experiment, we tested two further hypotheses (described in the main text). For the first, the null hypothesis stated that concordance when watching the same videos would be equal to the value observed for watching all videos ($H_0: \text{ICC}_{\text{same}} = \text{ICC}_{\text{all}}$); the alternative stated that concordance would be greater ($H_1: \text{ICC}_{\text{same}} > \text{ICC}_{\text{all}}$). Given the extremely high ICCs already observed for monozygotic 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 monozygotic cohort; however, we were adequately powered to detect significant increases, should they be observed, in the dizygotic and non-sibling cohorts. 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 hypothesis stated that concordance would be zero ($H_0: \text{ICC}_{\text{different}} = 0$), the alternative stated that concordance would be greater than zero ($H_1: \text{ICC}_{\text{different}} > 0$). Here, as in the main set of analyses, we had $>80\%$ power to detect correlations of ≥ 0.38 .

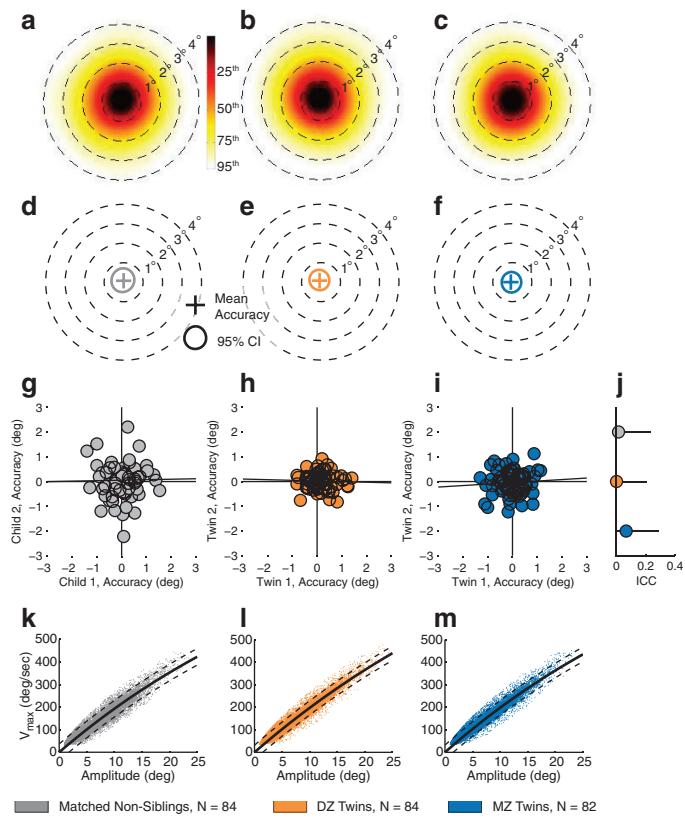
Code availability. Analysis of eye movements and coding of fixation data were performed with software written in MATLAB (MathWorks) either via the commandline or in scripts, available upon request from the corresponding author.

Data availability. The data that support the findings of this study are available from the corresponding author upon reasonable request.

31. Marrus, N. et al. Rapid video-referenced ratings of reciprocal social behavior in toddlers: a twin study. *J. Child Psychol. Psychiatry* **56**, 1338–1346 (2015).
32. Lord, C., Rutter, M., DiLavore, P. C. & Risi, S. *Autism Diagnostic Observation Schedule* (Western Psychological Services, 2002).
33. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* 4th edn, text revision (American Psychiatric Association, 2004).
34. Price, T. S. et al. Infant zygosity can be assigned by parental report questionnaire data. *Twin Res.* **3**, 129–133 (2000).
35. Neale, M. C. & Stevenson, J. Rater bias in the EASI temperament scales: a twin study. *J. Pers. Soc. Psychol.* **56**, 446–455 (1989).
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* **65**, 946–954 (2008).
37. Shrout, P. E. & Fleiss, J. L. Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* **86**, 420–428 (1979).
38. Locke, J. L. *The Child’s Path to Spoken Language* (Harvard Univ. Press, 1993).
39. Fleiss, J. L. & Shrout, P. E. Approximate interval estimation for a certain intraclass correlation coefficient. *Psychometrika* **43**, 259–262 (1978).
40. Giraudau, B. Negative values of the intraclass correlation coefficient are not theoretically possible. *J. Clin. Epidemiol.* **49**, 1205–1206 (1996).
41. Emery, N. J. The eyes have it: the neuroethology, function and evolution of social gaze. *Neurosci. Biobehav. Rev.* **24**, 581–604 (2000).
42. Fairchild, M. D. *Color Appearance Models* (John Wiley & Sons, 2005).
43. Itti, L. & Koch, C. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.* **40**, 1489–1506 (2000).
44. Koch, C. & Ullman, S. Shifts in selective visual attention: towards the underlying neural circuitry. *Hum. Neurobiol.* **4**, 219–227 (1985).
45. Itti, L. & Koch, C. Computational modelling of visual attention. *Nat. Rev. Neurosci.* **2**, 194–203 (2001).
46. Parkhurst, D., Law, K. & Niebur, E. Modeling the role of salience in the allocation of overt visual attention. *Vision Res.* **42**, 107–123 (2002).
47. Wolfe, J. M. & Horowitz, T. S. What attributes guide the deployment of visual attention and how do they do it? *Nat. Rev. Neurosci.* **5**, 495–501 (2004).
48. Tsotsos, J. K. et al. Modeling visual attention via selective tuning. *Artif. Intell.* **78**, 507–545 (1995).
49. Yantis, S. & Egeth, H. E. On the distinction between visual salience and stimulus-driven attentional capture. *J. Exp. Psychol. Hum. Percept. Perform.* **25**, 661–676 (1999).
50. Mazer, J. A. & Gallant, J. L. Goal-related activity in V4 during free viewing visual search: Evidence for a ventral stream visual salience map. *Neuron* **40**, 1241–1250 (2003).
51. Henderson, J. M., Brockmole, J. R., Castelhano, M. S. & Mack, M. in *Eye Movements: A Window on Mind and Brain* Ch. 25 (eds van Gompel, R. P. G., Fischer, M. H., Murray, W. S. & Hill, R. L.) 537–562 (Elsevier, 2007).
52. Lettvin, J. Y., Maturana, H. R., McCulloch, W. S. & Pitts, W. H. What the frog’s eye tells the frog’s brain. *Proc. Inst. Radio Engr.* **47**, 1940–1951 (1959).
53. Ghazanfar, A. A. & Santos, L. R. Primate brains in the wild: the sensory bases for social interactions. *Nat. Rev. Neurosci.* **5**, 603–616 (2004).
54. Moore, G. P., Perkel, D. H. & Segundo, J. P. Statistical analysis and functional interpretation of neuronal spike data. *Annu. Rev. Physiol.* **28**, 493–522 (1966).
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.* **329**, 151–160 (1985).
58. Hanes, D. P. & Wurtz, R. H. Interaction of the frontal eye field and superior colliculus for saccade generation. *J. Neurophysiol.* **85**, 804–815 (2001).
59. Bruce, C. J. & Goldberg, M. E. Physiology of the frontal eye fields. *Trends Neurosci.* **7**, 436–441 (1984).
60. Constantino, J. N. & Todd, R. D. Genetic structure of reciprocal social behavior. *Am. J. Psychiatry* **157**, 2043–2045 (2000).
61. Constantino, J. N. & Todd, R. D. Autistic traits in the general population: a twin study. *Arch. Gen. Psychiatry* **60**, 524–530 (2003).
62. Constantino, J. N. & Todd, R. D. Intergenerational transmission of subthreshold autistic traits in the general population. *Biol. Psychiatry* **57**, 655–660 (2005).
63. Constantino, J. N. et al. Developmental course of autistic social impairment in males. *Dev. Psychopathol.* **21**, 127–138 (2009).

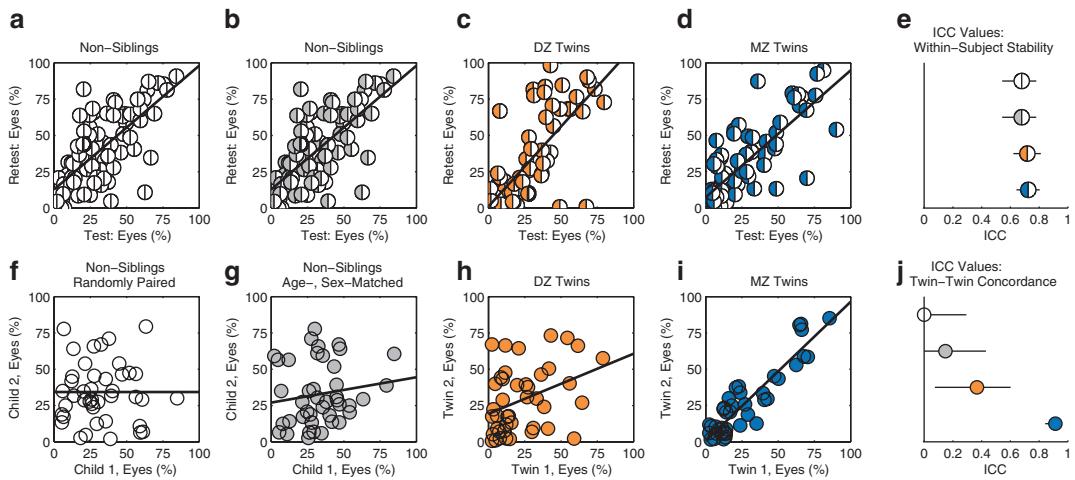


Extended Data Figure 1 | Measuring genetic structure of social visual engagement in 250 paired toddlers. Data consisted of measurements in dizygotic twins ($n = 84$, 42 pairs), monozygotic twins ($n = 82$, 41 pairs), and non-sibling comparison children ($n = 84$, randomized to 42 pairs). **a**, Example still images from 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. Plots (**b**, **c**) show two seconds of eye-tracking data, corresponding to each image in **a** (the onscreen image at midpoint of the two-second data sample). Data are overlaid on the corresponding regions of interest for each image, shaded to indicate eyes (dark grey), mouth (light grey), body (black), and object regions (white). Saccades are plotted as thin white lines with white dots; fixation data are plotted as larger coloured dots. **d–f**, Fixation time summaries for each comparison group for percentage of total fixation time on eye region (**d**), percentage of total fixation time on mouth region (**e**), and percentage of total time spent fixating (**f**). Box plots 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.



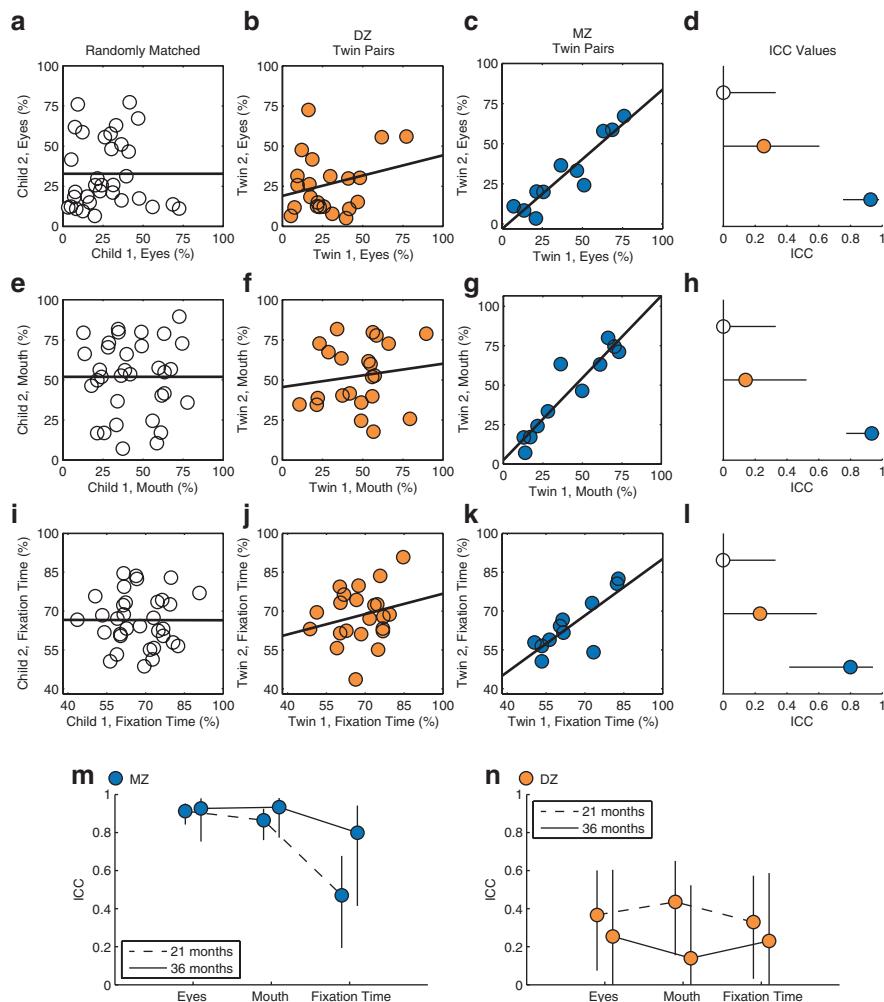
Extended Data Figure 2 | Between-group controls for calibration

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), dizygotic twins (b), and monozygotic 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 (in degrees) for non-sibling controls (d), dizygotic twins (e), and monozygotic twins (f). Crosses mark the location of mean calibration accuracy, while annuli mark 95% confidence intervals (95% CI). **g–i**, Concordance in calibration accuracy measurements for non-sibling controls (g), dizygotic twins (h), and monozygotic twins (i). Measurements in (g–i) are average accuracy per child across all accuracy verification trials. **j**, ICCs, plotted with 95% confidence intervals. **k–m**, Oculomotor relationship between maximum saccade velocity (V_{\max}) and amplitude (in degrees) for non-sibling controls (k), dizygotic twins (l), and monozygotic twins (m).



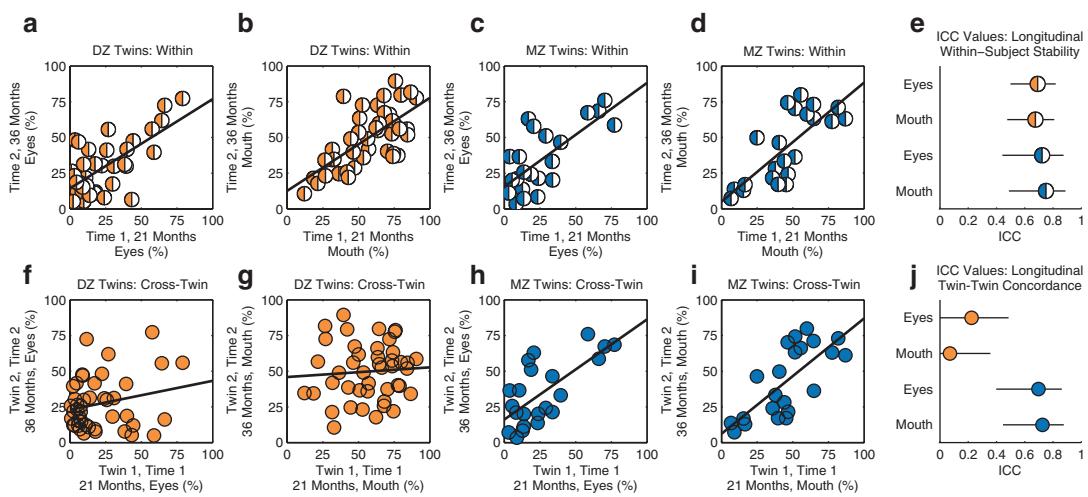
Extended Data Figure 3 | Within-subject stability versus between-subject concordance. For heritable traits, one expects to observe substantial within-subject stability contrasting with marked differences, varying by zygosity, in between-subject (twin-twin) concordance. **a–d**, Within-subject stability of observed levels of eye-looking for non-siblings (**a**, **b**), dizygotic twins (**c**), and monozygotic twins (**d**). Dots are each child's measured level on the test comparison (*x*-axis) versus measured level on the retest comparison (*y*-axis). (*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 measurements of eye-looking quantified by ICC with two-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 analyses of inter-individual variation—the difference between individuals—are also highly reliable. **f–i**, Plots repeated from main text Fig. 1a–e, showing paired measurements of eye-looking in randomly paired non-siblings (**f**), in age- and sex-matched non-siblings (**g**), in dizygotic twins (**h**), and in monozygotic twins (**i**). Dots are measured levels per child, paired so that one child's level of eye-looking is on the *x*-axis versus the paired child on the *y*-axis. **j**, ICCs and 95% confidence intervals for twin-twin concordance in eye-looking.



Extended Data Figure 4 | Monozygotic twins maintain high twin-twin concordance, which is significantly greater than that observed in dizygotic twins, when tested again at 36 months. **a–c**, Paired measurements of eye-looking in randomly assigned pairs (**a**), in dizygotic twins (**b**), and in monozygotic twins (**c**). **d**, ICCs and 95% confidence intervals across groups for eye-looking. **e–h**, Paired measurements of concordance in mouth-looking. **i–l**, Paired measurements of concordance in percentage of time spent attending to task (maintaining stable onscreen fixation). In all plots, randomly matched controls in white, dizygotic twins in orange, and monozygotic twins in blue. Error estimates are 95%

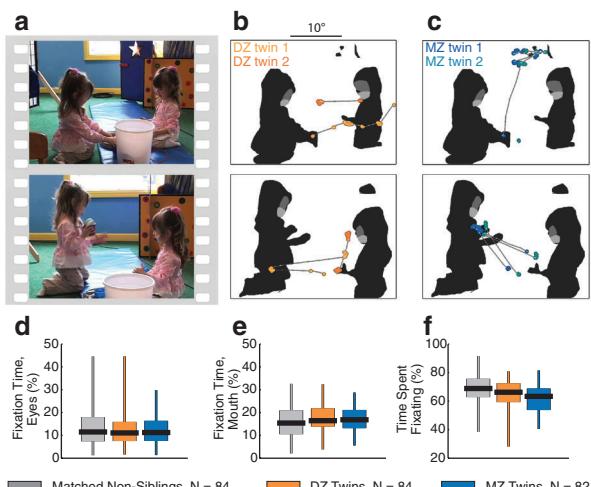
confidence intervals. Dots are individual values for paired children, as in Extended Data Fig. 3f–i. **m, n**, Summary of results for monozygotic (**m**) and dizygotic (**n**) twins at initial time of testing (21 months, summary data from Fig. 1) relative to results at time of longitudinal follow-up (36 months, summary from **d, h, l**). Monozygotic twins exhibit marginally, though not significantly, increased concordance values when tested again at 36 months. By contrast, dizygotic twins exhibit marginally, though not significantly, decreased concordance values. Plotted data in **a, e, and i** are representative random pairings, selected to match the mean ICC value of all 10,000 re-samplings.



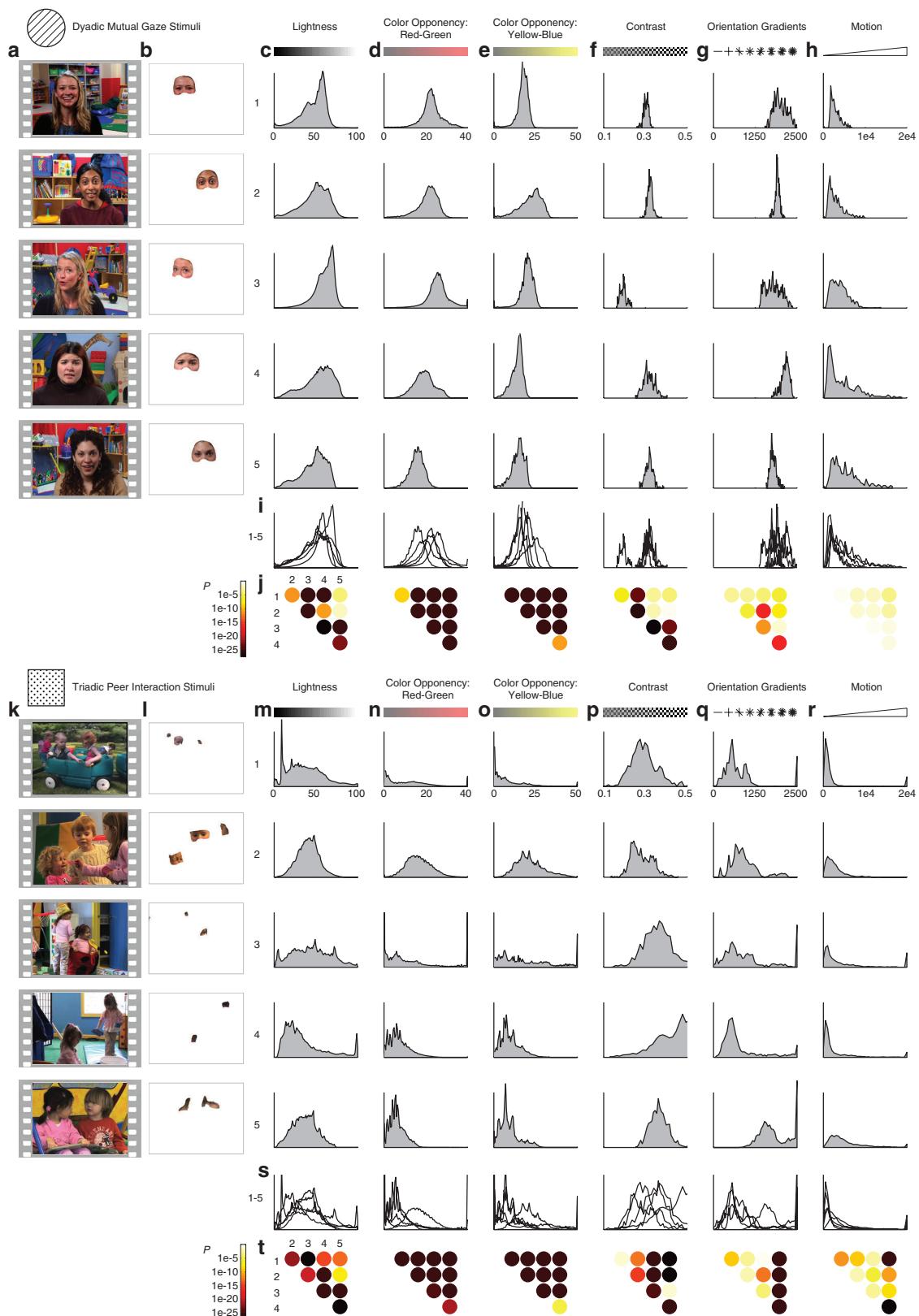
Extended Data Figure 5 | Longitudinal within-subject stability versus longitudinal twin-twin concordance, from 21 until 36 months.

Dizygotic and monozygotic twins both show high levels of longitudinal within-subject stability when tested again 15 months after initial data were collected, but only monozygotic 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 dizygotic

twins, and within-subject stability of eye-looking (**c**) and mouth-looking (**d**) for monozygotic twins. **e**, Summary of longitudinal within-subject stability quantified by ICC with two-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- (**f**) and mouth-looking (**g**) in dizygotic twins, and for eye- (**h**) and mouth-looking (**i**) in monozygotic twins. **j**, ICCs and 95% confidence intervals.



Extended Data Figure 6 | Social visual engagement when watching triadic peer interaction stimuli in 250 paired toddlers. Data consisted of measurements in 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 twins. **c**, Data from two typically developing 18-month-old monozygotic twins. **b**, **c**, Two seconds of eye-tracking data are plotted, corresponding to each image in **a** (the onscreen image 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 coloured 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 | See next page for caption.

Extended Data Figure 7 | Physical image properties that constitute eyes vary significantly from video stimulus to video stimulus in lightness, colour, 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 out 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 analysed. **c–h**, In the rows to the right of each representative still image and corresponding eye region, physical image property profiles, analysed across all video frames, are given as histograms. **c**, Lightness. **d**, Red–green colour opponency. **e**, Yellow–blue colour opponency. **f**, Contrast. **g**, Orientation gradients. **h**, Motion. **i**, For each physical image property analysed in columns (**a–h**),

i gives corresponding comparison plots across the five histograms located in the column directly above. **j**, Statistical comparisons of the measured image property distributions by two-sample Kolmogorov–Smirnov test. *P* values are corrected for multiple comparisons by the Bonferroni method. For each of the physical image properties analysed in columns (**a–h**), **j** presents the corresponding matrix of statistical comparisons (that is, the 1st row of coloured circles presents comparisons for Video 1 versus Video 2, Video 1 versus Video 3, and so on; while the 2nd row presents comparisons for Video 2 versus Video 3, Video 2 versus Video 4, and so on). **k**, Still images sampled from videos depicting triadic peer interaction stimuli (scenes of children interacting in a childcare setting). Still images from 5 out of 12 videos are shown (all 12 triadic peer interaction videos are included in the 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		
<i>same sex</i>	36.1	130	50.6	84	N/A	N/A
<i>opposite sex</i>	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
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

	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 (N = 22 pairs)	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)

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 ICC with 95% confidence intervals in parentheses. *P<0.05, **P<0.01, one-sided comparison relative to 0.

Extended Data Table 3 | Size of experimental stimuli and viewing time summaries**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 = 41 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)

a. Sizes of regions of interest (ROI) for dyadic mutual gaze stimuli. Data are given as mean (s.d.) in degrees of visual angle. Object ROI 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 Fig. 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 the ROIs for triadic peer interaction stimuli. Data are given as mean (s.d.) 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 Fig. 6k); measurements reflect total body region size. **c.** Total viewing time and time spent in fixation, saccade, offscreen/missing and blinks. Data are given as mean (s.d.) in minutes. All measurements are summarized across both dyadic mutual gaze stimuli and triadic peer interaction stimuli. Non-sibling controls watched a foreshortened subset of video stimuli. **d.** Time fixating per onscreen ROI, mean (s.d.) in minutes. All measurements are summarized across both dyadic mutual gaze stimuli and triadic peer interaction stimuli. Non-sibling controls watched a foreshortened subset of video stimuli.