LETTERS

Two-year-olds with autism orient to non-social contingencies rather than biological motion

Ami Klin[1], David J. Lin[1]†, Phillip Gorrindo[1]†, Gordon Ramsay[1,2] & Warren Jones[1,3]

Typically developing human infants preferentially attend to biological motion within the first days of life[1]. This ability is highly conserved across species[2,3] and is believed to be critical for filial attachment and for detection of predators[4]. The neural underpinnings of biological motion perception are overlapping with brain regions involved in perception of basic social signals such as facial expression and gaze direction[5], and preferential attention to biological motion is seen as a precursor to the capacity for attributing intentions to others[6]. However, in a serendipitous observation[7], we recently found that an infant with autism failed to recognize point-light displays of biological motion, but was instead highly sensitive to the presence of a non-social, physical contingency that occurred within the stimuli by chance. This observation raised the possibility that perception of biological motion may be altered in children with autism from a very early age, with cascading consequences for both social development and the lifelong impairments in social interaction that are a hallmark of autism spectrum disorders[8]. Here we show that two-year-olds with autism fail to orient towards point-light displays of biological motion, and their viewing behaviour when watching these point-light displays can be explained instead as a response to non-social, physical contingencies—physical contingencies that are disregarded by control children. This observation has far-reaching implications for understanding the altered neurodevelopmental trajectory of brain specialization in autism[9].

Preferential attention to biological motion is a fundamental mechanism facilitating adaptive interaction with other living beings. It is present throughout a wide range of species, from humans[10,11] to monkeys[12] to birds[13]. Developmentally, it can be found in newly hatched chicks[14] and in human infants as young as 2 days old[1]. Recognition of biological motion remains intact in a variety of forms, from degraded presentations, through varying states of occlusion, and in cases when information-bearing components are reduced to their most minimal[15,16]. In addition, perception of biological motion can be preserved even when other types of motion perception are impaired, as in individuals with Williams syndrome[17] (a condition noted for visuo-spatial deficits) and in patients suffering from circumscribed brain lesions[18]. Furthermore, biological motion perceived through other sensory modalities—such as when listening to sounds of human motion[19]—evokes activity in the same areas of the brain that are typically responsive to visual presentations.

Collectively, these findings describe a mechanism that is evolutionarily well-conserved, developmentally early-emerging, highly robust in signal detection (withstanding degradation on signalling and receiving sides), and redundantly represented by several sensory modalities. Each of these aspects suggests ready benefits for adaptive interaction with other living beings: following the movements of a conspecific, looking at others to entreat or avoid interaction, learning by imitation, or directing preferential attention to cues that build on biological motion (such as facial expression and gaze direction[5]).

Notably, many of the same behaviours have also been shown as deficits in children with autism: deficits in social interaction, diminished eye contact and reduced looking at others, problems with imitation, deficits in recognizing facial expressions, and difficulties following another's gaze[20]. Autism is a lifelong, highly prevalent, and strongly genetic disorder defined by impairments in social and communicative functioning and by pronounced behavioural rigidities[21]. Although the preponderance of evidence points to prenatal factors instantiated in infancy, knowledge of the first two years of life in autism remains largely limited to retrospective data and indirect observations[20]: because autism is rarely diagnosed before 18 months, relatively little is known about autism during the first two years of development.

In later life, much more is known about the consequences—cognitive, social and behavioural—of having autism. Altered visual scanning, of both faces and social scenes[22,23], as well as altered neural processing of social information, have been documented[24,25]. In school-age children with autism, perception of biological motion is impaired[26], but the manner in which very young children with autism relate to biological motion in early life, during periods critical for brain development and before compensatory coping strategies are established, has not, to our knowledge, been previously studied.

In the current study, we sought to address whether preferential attention to biological motion is altered in children with autism by two years of age, and what other factors might guide the visual attention of children with autism if they do fail to orient towards biological motion.

To answer these questions, we created five sets of point-light animations, counterbalanced for a total of ten. The animations consisted of children's games, such as playing 'peek-a-boo' or 'pat-a-cake', and were created with live actors and motion capture technology (see Supplementary Information). The motion capture sessions included a simultaneous audio recording. The experimental task was a preferential-looking paradigm (Fig. 1a and Supplementary Movie 1): a point-light animation of biological motion was presented on one half of a computer screen, together with the audio soundtrack of the actor's vocalizations. On the other half of the screen, the same animation was presented, but that point-light figure was inverted in orientation (shown upside-down) and played in reverse order (the frames of animated action played from the end of the sequence until its beginning). Only the one (forward) audio soundtrack was presented.

Inverted presentation disrupts perception of biological motion in young children[27], and is processed by different neural circuits in infants as young as 8 months old[28]. Also, by playing the inverted

[1]Yale Child Study Center, Yale University School of Medicine, New Haven, Connecticut 06519-1124, USA. [2]Haskins Laboratories, New Haven, Connecticut 06511, USA. [3]Interdepartmental Neuroscience Program, Yale University School of Medicine, New Haven, Connecticut 06520-8074, USA. †Present addresses: Division of Health Sciences and Technology, Harvard Medical School, Boston, Massachusetts 02115, USA (D.J.L.); Neuroscience Graduate Program at Vanderbilt Kennedy Center for Research on Human Development, Vanderbilt University, Nashville, Tennessee 37203, USA (P.G.).

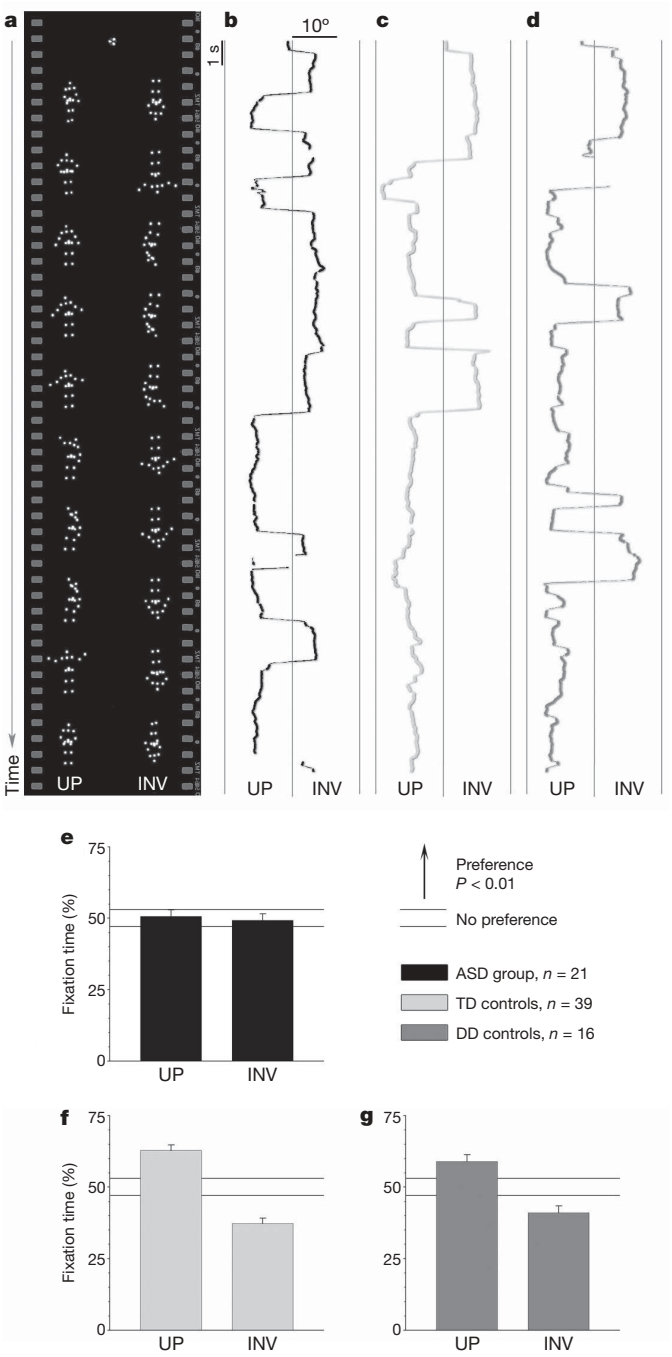

Figure 1 | Two-year-olds with autism show no preferential attention to biological motion, whereas control children show significant preferences. **a**, Example still images from point-light biological motion stimuli, with centring cue at start. Each animation showed an upright (UP) and inverted (INV) figure with accompanying soundtrack matching the actions of the upright figure. The upright figure enacted childhood games. Figures were identical except that the inverted figure was rotated 180° and its movements were played in reverse order. **b**–**d**, Visual scanning data of individual children are plotted as horizontal location by time. Breaks in the data occur for blinks or offscreen fixations. **b**, Visual scanning data from one toddler with autism (ASD), for one animation. **c**, Data from one typically developing toddler (TD). **d**, Data from one developmentally delayed but non-autistic toddler (DD). **e**, For the ASD group, fixation to upright and inverted biological motion occurs at chance levels. **f**, Typically developing toddlers give preferential attention to upright animations. **g**, Developmentally delayed toddlers also give preferential attention to upright animations. Horizontal guidelines denote percentages not significantly different from chance. Error bars are s.e.m.

animation backwards, its relative levels of motion complexity, speed and gestalt coherence were preserved, but its motion was not an exact mirror of the upright. Each animation lasted an average of 30 s. The order of presentation was randomized, and the presentation of the upright figure was counterbalanced to appear on the left and right side of the screen equally often.

Evidence for recognition and preferential attention to biological motion was measured by the child's viewing patterns: increased looking towards the upright figure indicated preferential attention to biological motion[1] and the perceptual matching of human voice with a mental template of human action[8]. Visual scanning was measured with eye-tracking equipment, with data collected at 60 Hz (Fig. 1b–d) (see Methods in Supplementary Information).

With the written, informed consent of their parents or legal guardians, 76 children with a mean chronological age of 2.05 (s.d. = 0.62) participated. These children comprised three groups (see Supplementary Table 1): 21 toddlers with autism spectrum disorders (ASD), 39 typically developing toddlers, and 16 developmentally delayed but non-autistic toddlers. Toddlers with autism were matched to the typically developing toddlers on non-verbal mental age and chronological age, and matched to the developmentally delayed, non-autistic toddlers on verbal mental age and chronological age (see Supplementary Information).

Whereas typically developing toddlers provide normative data, the developmentally delayed but non-autistic children act as controls against developmental confounds, assuring that the findings are specific to autism rather than attributable to delays in cognitive development or language function.

Results are plotted in Fig. 1e–g. When viewing point-light displays of human biological motion, two-year-olds with autism spectrum disorders are random in their looking patterns: 50.7% on the upright figure versus 49.3% on the inverted (Fig. 1e). In contrast, both control groups demonstrated significant preferential attention to the upright animations: 62.7% upright for the typically developing group, and 58.9% upright for the developmentally delayed group (Fig. 1f, g). Comparison across groups was significantly different (by one-way analysis of variance (ANOVA), $F_{2,73} = 7.95$, $P < 0.001$). In pairwise comparisons, looking by the ASD group differed significantly from that of each control group ($P < 0.001$ in comparison with the typically developing group, and $P = 0.0185$ relative to the developmentally delayed group). The two control groups did not differ significantly from one another ($P = 0.27$). All data were normally distributed (all $P > 0.4$, $k < 0.15$, Lilliefors).

Results in Fig. 1 are from four of the five types of animations presented. In earlier research[7], a serendipitous observation led us to recognize that one of the animations contained a confounder. Although four animations presented only moving point-lights with an accompanying human voice, one animation included a different sound. The actor in that animation plays pat-a-cake (see Supplementary Movie 2), and the sound of clapping is heard at the same time that two point-lights—the actor's hands—collide. The collision of point-lights and the resulting clapping sound create a causal physical contingency: rather than merely co-occurring (as with the speech sounds and movements in the other animations), the movements of the point-light hands in this case actually cause a noise to occur. In the earlier research we found that a 15-month-old with autism was very sensitive to the occurrence of this clapping, as her preferential looking went from random during other animations to 93.1% upright during the pat-a-cake animation[7].

During the clapping, the causal physical contingency only exists on the upright side: the single audio track plays normally (forward), matching the upright movements, but the action of the inverted figure, playing in reverse, does not move in time to the clapping sounds.

When analysed independently (Fig. 2), the toddlers with ASD showed a significant preference for the upright clapping figure during the pat-a-cake animation, and looking towards the upright figure during this animation was significantly increased relative to other

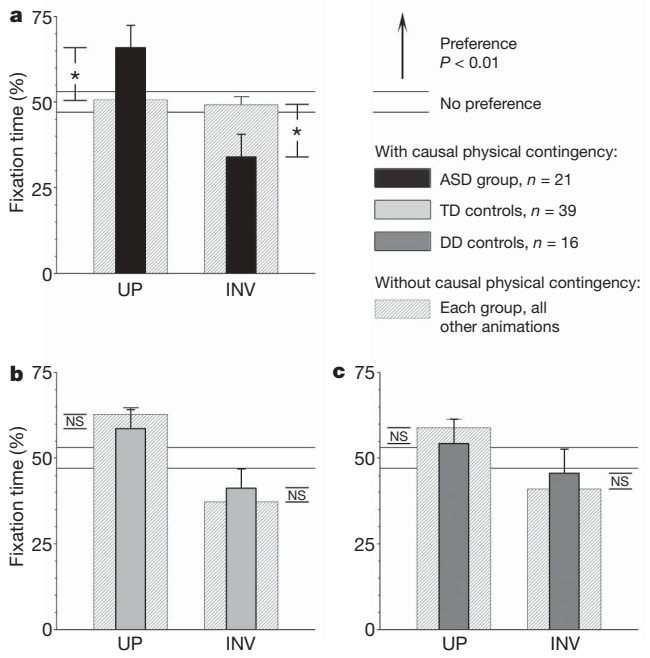

Figure 2 | When the animation contains a physical contingency, two-year-olds with autism do show significant viewing preferences. **a**, During other biological motion animations, ASD toddlers show no preference, but when a physical contingency is present on the upright side, these toddlers show significant preference for the upright figure (different from chance: $P < 0.01$; different from their viewing behaviour to other animations: $P = 0.044$). Whereas other animations presented only moving point-lights and human voice, one type of animation contained an extra cue: as two point-lights, representing the actor's hands, collided, the sound of clapping could be heard (playing 'pat-a-cake'). The collision of point-light 'hands' actually caused a noise (the clap) to occur, localized to the upright (UP) figure and absent from the inverted (INV; the inverted figure's movements were not synchronous with the claps). **b**, Typically developing toddlers show no significant change in preferential viewing. **c**, Developmentally delayed toddlers also show no significant change in preferential viewing. Horizontal guidelines denote percentages not significantly different from chance. Error bars are s.e.m. $*P < 0.05$. See Supplementary Movies for movie data.

animations: 65.9% upright during pat-a-cake versus only 50.7% in the other four animations, $t_{20} = 2.43$, $P = 0.02$. Behaviour of the typically developing and developmentally delayed groups was unchanged: they continued to give preferential attention to the upright figure: 58.6% upright during pat-a-cake versus 62.7% in the other four animations for typically developing ($t_{38} = 0.79$, $P = 0.44$); and 54.4% versus 58.9% for developmentally delayed ($t_{15} = 0.66$, $P = 0.51$). Overall on this animation, results for the three groups did not differ significantly ($F_{2,73} = 0.67$, $P = 0.52$). All data were normally distributed (all $P > 0.36$, $k < 0.15$, Lilliefors).

After this observation, we questioned whether the presence of more subtle synchronies might have had an unanticipated role in the viewing of all animations—that is, whether visual scanning that had appeared random by the toddlers with ASD might actually be related to audiovisual synchronies less obvious than clapping.

To test this, we quantified levels of audiovisual synchrony (AVS) in all animations (Fig. 3). In the pat-a-cake animation, when the point-light hands collide and a clapping sound occurs, an abrupt change in motion coincides with a large change in sound amplitude. We measured AVS in our stimuli to match this case: the synchronous occurrence of change in motion and change in sound[29].

We measured the change in motion by first measuring each point-light's trajectory over time (Fig. 3a). From each point-light's trajectory, we calculated its velocity and then the magnitude of its change in velocity, $|\Delta v|$ (Fig. 3b, c). This served as our measure of change in motion. To measure change in sound, we measured the audio amplitude of the

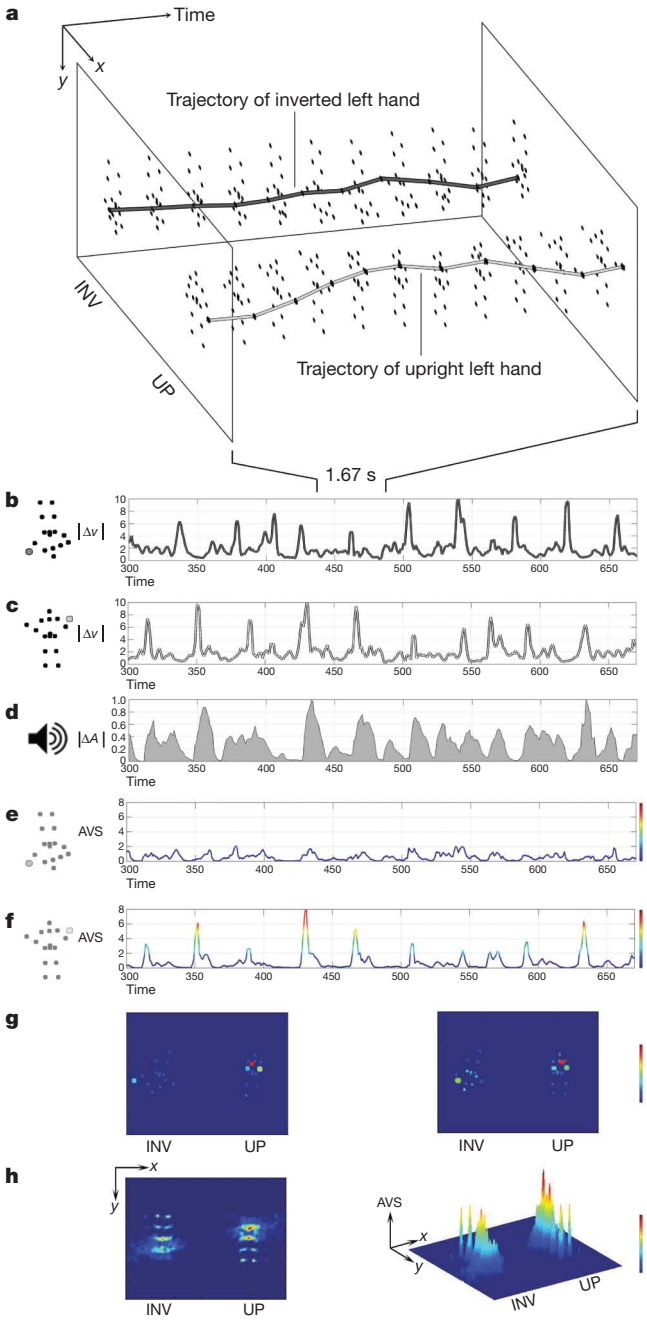

Figure 3 | Quantification of audiovisual synchrony. **a**, We measured spatial trajectories (x-y location over time) of all point-lights throughout each biological motion animation. Example trajectories are for inverted (INV) left hand and for upright (UP) left hand. **b**, Magnitude of change in velocity of inverted left hand, $|\Delta v|$. **c**, Magnitude of change in velocity of upright left hand. **d**, Magnitude of change in short-term amplitude envelope of audio soundtrack, $|\Delta A|$. **e**, AVS of inverted left hand, obtained as a pointwise product of **b** and **d**. **f**, AVS of upright left hand, obtained as a pointwise product of **c** and **d**. **g**, Two still frames from pat-a-cake animation. Colour scale values range from low or no synchrony (dark blue) to maximum synchrony (red). Note that some point-lights are very synchronous (the hands, shown here during claps), whereas others are hardly synchronous (for example, the feet). **h**, Summation of AVS over the duration of an entire animation. Oblique view shows that although there is more AVS on the upright side, the inverted side also contains synchrony: by chance alignment (reverse motion signal aligned with forward audio signal), some change in movement of inverted point-lights can occur synchronously with the change in audio. If preferential viewing in our stimuli were related to the level of AVS, then the relative levels of synchrony on the upright versus inverted side will provide predictions of expected viewing behaviour.

soundtrack (its short-term amplitude envelope) and then calculated its rate of change, the magnitude of ΔA, $|\Delta A|$ (Fig. 3d). The level of AVS of each point-light was then calculated as the product of change in velocity and change in sound amplitude (Fig. 3e, f). This measure of AVS was computed for all point-lights on both the upright and inverted sides (see Supplementary Movie 3 and Supplementary Information).

By then summing the AVS signals of all point-lights over time, we generated cumulative maps of AVS for each animation (Fig. 3h). From these maps, we calculated the difference between maximum AVS on the upright side and maximum synchrony on the inverted side (as a percentage difference to normalize across animations).

Across different animations, this measure of upright versus inverted synchrony then acted as a prediction of which side of the animation would be preferentially attended—if the viewing patterns of children were related to attention to AVS. The relationship between synchrony and preferential viewing was tested by regression (Fig. 4). For the ASD group, preferential looking was significantly and strongly correlated with level of AVS ($R^2 = 0.90$ and $P = 0.01$; Fig. 4a). In the typically developing and the developmentally delayed groups, there was no significant correlation between viewing and AVS ($R^2 = 0.29$ and 0.17, respectively; Fig. 4b, c). Correlation coefficients for the three groups were significantly different from one another ($\chi^2 = 7.24$, $P < 0.05$)[30], with the r value of the ASD group differing from that of the typically developing group ($z = 2.41$, $P < 0.05$) as well as the developmentally delayed group ($z = -2.25$, $P < 0.05$). The two control groups did not differ significantly. The pat-a-cake animation had the greatest upright AVS. When we withheld that animation and re-analysed, the correlation between preferential viewing and AVS remained significant for the ASD group ($R^2 = 0.95$ and $P = 0.018$), but was still not significant for the other groups ($R^2 = 0.04$ for typically developing, and $R^2 = 0.001$ for developmentally delayed).

The results from this *post hoc* quantification of AVS and preferential viewing indicated that the viewing patterns of toddlers with autism—random relative to social content—showed instead a marked reliance on AVS. This one measure accounted for 90% of the autism group's variance in preferential viewing. In contrast, the looking patterns of typically developing and of developmentally delayed, non-autistic children showed no relationship with the levels of AVS. The control children gave preferential attention to biological motion, disregarding AVS in favour of more socially relevant signals.

To test whether AVS could predict looking behaviour in new animations, we designed a follow-up experiment (Fig. 4d) in which we created two new types of animations with increased AVS levels, filling the gap in synchrony signal strength of our original stimuli. We recruited ten additional toddlers with ASD, characterized in the same manner and matched to the original ASD cohort (see Supplementary Information). We used our original results to build a predictive model for expected behaviour, creating weighted binomial prediction intervals around the original regression line[31], with specific predictions for each animation. The probability of both results falling within their respective prediction intervals is equal to the probability of obtaining a value in one interval multiplied by the probability of obtaining a value within the other ($P = [0.1674 - 0.0002] \times [0.0024 - 0]$).

Preferential viewing by this second cohort of toddlers with autism, watching new animations, fit the predictions on the basis of AVS: their viewing on each animation followed the model, a result with a chance likelihood of $P = 0.0004$.

Overall, these results indicate that a skill present in two-day-old, typically developing infants[1], as well as in chronologically, non-verbally, and verbally matched control children (the typically developing and developmentally delayed groups herein), is not functioning properly in children with autism at the age of two.

There are likely to be significant implications of a disruption to such a basic and highly conserved mechanism. One immediate implication of this finding concerns our understanding of another very basic behaviour: how infants with autism look at the faces of other people. We recently found that in comparison with control children, two-year-olds with autism look less at the eyes of others and attend instead to their mouths[24]. The present results indicate an explanation: where on the face is there greatest AVS? These children's sensitivity to synchrony in the present biological motion stimuli is consistent with fixating on the ongoing synchronies between lip motion and speech sound, and the lack of preferential attention towards biological motion is consistent with diminished attention to the eyes and diminished expertise in social action and interaction found in later life.

Developmentally, these results mark an important, early point along an alternative path of neural and behavioural specialization. Although individual and species-specific genetics begin the development of mind and brain, that development over time is shaped critically by experience. For infants with autism, this would suggest that genetic predispositions are probably exacerbated by experiences that are increasingly atypical. By two-years-of-age, the data in this report show that these children are on a substantially different developmental course, having learned already from a world in which the physical contingencies of coincident light and sound are quantifiably more salient than the rich social information imparted by biological motion. Future investigations will benefit from studies, starting still earlier in life, of the developmental unfolding of such selective learning profiles. Exactly which signals are spontaneously attended to and which are missed, and the consequences thereof for structural and

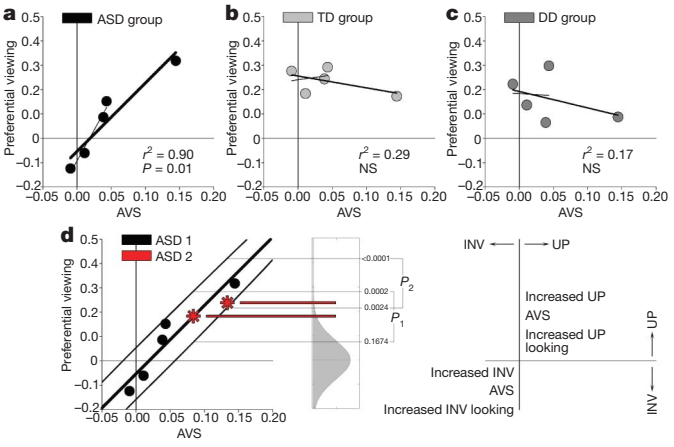

Figure 4 | The level of AVS is highly correlated with preferential viewing in two-year-olds with autism, is uncorrelated with viewing in control children, and can predict ASD viewing patterns in new animations. **a**, Preferential viewing is significantly correlated with AVS in ASD toddlers. Plots pair preferential viewing and AVS. When the animation with greatest upright AVS (pat-a-cake) is withheld from analysis, AVS is still significantly correlated with viewing behaviour in ASD toddlers: $r^2 = 0.95$, $P = 0.018$ (plotted as thin regression line through remaining four data points). **b**, Preferential viewing by typically developing (TD) toddlers is uncorrelated with AVS, across either four or five animations. **c**, Preferential viewing by developmentally delayed (DD) toddlers is also uncorrelated with AVS. **d**, To test whether AVS could predict looking behaviour in new animations, we created two further animation types. The regression from the original data, with weighted binomial prediction intervals, provided a model for expected behaviour. P_1 and P_2 denote prediction intervals for the new animations. Probability of obtaining the results in these intervals is noted to the right of the regression plot. For an independent cohort of toddlers with autism, matched to the original cohort, preferential viewing was predicted on the basis of AVS ($P = 0.0004$). In all plots, the y-axis shows preferential viewing as a difference score: percentage of fixation time to upright (UP) minus percentage of fixation time to inverted (INV). Positive values indicate increased looking at the upright. Similarly, the x-axis shows AVS as synchrony of the upright (as percentage of total synchrony) minus synchrony of the inverted (also as percentage of total). Positive values indicate greater synchrony in the upright figure.

functional brain development, may shed light on the neurobiological anomalies that predispose these altered avenues of learning.

METHODS SUMMARY

Children were recruited through a federally funded STAART Center (Studies to Advance Autism Research and Treatment, NIMH U54-MH66494) based in the Autism Program of the Yale Child Study Center, New Haven. The research protocol was approved by the Human Investigations Committee of the Yale University School of Medicine, and families were free to withdraw from the study at any time. The children were shown counterbalanced presentations of each of five point-light biological motion animations (for a total of ten presentations in the original experiment), and two extra animations (four presentations in the follow-up experiment) (see Fig. 1a and Supplementary Movies 1–3). Preferential viewing in our design was a binary choice, upright versus inverted. Visual scanning was measured with eye-tracking equipment (ISCAN, Inc.). The equipment uses a dark pupil/corneal reflection technique with data collected at the rate of 60 Hz. Analysis of eye movements and coding of preferential fixation data were performed with software written in MATLAB.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 November 2008; accepted 3 February 2009.
Published online 29 March 2009; corrected 14 May 2009 (see full-text HTML version for details).

1. Simion, F., Regolin, L. & Bulf, H. A predisposition for biological motion in the newborn baby. *Proc. Natl Acad. Sci. USA* **105,** 809–813 (2008).
2. Regolin, L., Tommasi, L. & Vallortigara, G. Visual perception of biological motion in newly hatched chicks as revealed by an imprinting procedure. *Anim. Cogn.* **3,** 53–60 (2000).
3. Blake, R. Cats perceive biological motion. *Psychol. Sci.* **4,** 54–57 (1993).
4. Johnson, M. H. Biological motion: a perceptual life detector? *Curr. Biol.* **16,** R376–R377 (2006).
5. Pelphrey, K. A., Morris, J. P., Michelich, C. R., Allison, T. & McCarthy, G. Functional anatomy of biological motion perception in posterior temporal cortex: an fMRI study of eye, mouth and hand movements. *Cereb. Cortex* **15,** 1866–1876 (2005).
6. Frith, C. D. & Frith, U. Interacting minds: a biological basis. *Science* **286,** 1692–1695 (1999).
7. Klin, A. & Jones, W. Altered face scanning and impaired recognition of biological motion in a 15-month-old infant with autism. *Dev. Sci.* **11,** 40–46 (2008).
8. Klin, A., Jones, W., Schultz, R. T. & Volkmar, F. The enactive mind – from actions to cognition: lessons from autism. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* **358,** 345–360 (2003).
9. Johnson, M. H. Functional brain development in humans. *Nature Rev. Neurosci.* **2,** 475–483 (2001).
10. Johansson, G. Visual perception of biological motion and a model for its analysis. *Percept. Psychophys.* **14,** 201–211 (1973).
11. Fox, R. & McDaniel, C. The perception of biological motion by human infants. *Science* **218,** 486–487 (1982).
12. Oram, M. W. & Perrett, D. I. Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *J. Neurophysiol.* **76,** 109–129 (1996).
13. Omori, E. & Watanabe, S. Discrimination of Johansson's stimuli in pigeons. *Int. J. Comp. Psychol.* **9,** 92 (1996).
14. Vallortigara, G., Regolin, L. & Marconato, F. Visually inexperienced chicks exhibit spontaneous preference for biological motion patterns. *PLoS Biol.* **3,** e208 (2005).
15. Thompson, J. C., Clarke, M., Stewart, T. & Puce, A. Configural processing of biological motion in human superior temporal sulcus. *J. Neurosci.* **25,** 9059–9066 (2005).
16. Neri, P., Morrone, M. C. & Burr, D. C. Seeing biological motion. *Nature* **395,** 894–896 (1998).
17. Jordan, H., Reiss, J. E., Hoffman, J. E. & Landau, B. Intact perception of biological motion in the face of profound spatial deficits: Williams syndrome. *Psychol. Sci.* **13,** 162–167 (2002).
18. Jokisch, D., Troje, N. F., Koch, B., Schwarz, M. & Daum, I. Differential involvement of the cerebellum in biological and coherent motion perception. *Eur. J. Neurosci.* **21,** 3439–3446 (2005).
19. Bidet-Caulet, A., Voisin, J., Bertrand, O. & Folumpt, P. Listening to a walking human activates the temporal biological motion area. *Neuroimage* **28,** 132–139 (2005).
20. Chawarska, K., Klin, A. & Volkmar, F. R. *Autism Spectrum Disorders in Infants and Toddlers: Diagnosis, Assessment and Treatment* (Guilford Press, 2008).
21. Volkmar, F. R., Lord, C., Bailey, A., Schultz, R. T. & Klin, A. Autism and pervasive developmental disorders. *J. Child Psychol. Psychiatry* **45,** 135–170 (2004).
22. 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).
23. Jones, W., Carr, K. & Klin, A. Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-olds with autism spectrum disorder. *Arch. Gen. Psychiatry* **65,** 946–954 (2008).
24. Schultz, R. T. Developmental deficits in social perception in autism: the role of the amygdala and fusiform face area. *Int. J. Dev. Neurosci.* **23,** 125–141 (2005).
25. Dalton, K. M. *et al.* Gaze fixation and the neural circuitry of face processing in autism. *Nature Neurosci.* **8,** 519–526 (2005).
26. Blake, R., Turner, L. M., Smoski, M. J., Pozdol, S. L. & Stone, W. L. Visual recognition of biological motion is impaired in children with autism. *Psychol. Sci.* **14,** 151–157 (2003).
27. Pavlova, M. & Sokolov, A. Orientation specificity in biological motion perception. *Percept. Psychophys.* **62,** 889–899 (2000).
28. Reid, V. M., Hoehl, S. & Striano, T. The perception of biological motion by infants: an event-related potential study. *Neurosci. Lett.* **395,** 211–214 (2006).
29. Driver, J. & Spence, C. Multisensory perception: beyond modularity and convergence. *Curr. Biol.* **10,** R731–R735 (2000).
30. Chen, P. Y. & Popovich, P. M. *Correlation: Parametric and Nonparametric Measures* 20–23 (Sage Publications, 2002).
31. Von Collani, E. & Dräger, K. *Binomial Distribution Handbook for Scientists and Engineers* 182–187 (Birkhäuser, 2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by a grant from the National Institutes of Mental Health (U54-MH66494), by Autism Speaks (piloting of methods), and by the Simons Foundation (current methodological developments). We wish to thank the families of the children included in this study for their time and participation. We would also like to thank F. Shic for conceptual discussions and thoughts on methods. We would like to thank L. Dickholtz, A. Kord, L. Else and M. Campisi for their assistance in creating the biological motion stimuli. We also wish to thank K. Carr, A. Blank, A. Bhatt, S. Shultz and K. Knoch for their help in this research project, and our colleagues K. Chawarska, R. Paul and F. Volkmar for conceptual discussions and for their contributions to the clinical characterization of the samples.

Author Contributions A.K. and W.J. developed the initial idea and design of the study, interpreted data, wrote the final manuscript, and take full responsibility for the integrity of data and the accuracy of data analysis. A.K. supervised participant characterization. W.J. supervised all technical aspects of experimental procedure, data acquisition and analysis. P.G. contributed to initial development of AVS methods and data analysis. D.J.L., with W.J. and G.R., developed the final AVS methods. G.R., with W.J. and A.K., helped develop new animations for the second experiment. W.J. created the figures. A.K. and W.J. performed the final revision of the manuscript for intellectual content.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.K. (ami.klin@yale.edu) or W.J. (warren.jones@yale.edu).

METHODS

Experimental procedure and setting. Throughout the procedure, toddlers were accompanied by a parent or primary caregiver. To begin the session, the child and caregiver entered the laboratory room while a children's video (for example, Baby Mozart, Elmo) played on a computer monitor. The computer monitor was mounted within a wall panel, and audio was played through a set of concealed speakers. The toddler was seated and buckled into a car seat mounted on a pneumatic lift so that the viewing height (line-of-sight) was standardized for all children. Viewers' eyes were 30 inches (76.2 cm) from the computer monitor, which subtended an approximately $23^\circ \times 30^\circ$ portion of each child's visual field. Lights in the room were dimmed so that only images shown on the computer monitor could be easily seen. The experimenter was concealed from the child's view throughout the testing session but was able to monitor the child at all times by means of an eye-tracking camera and by a second video camera that filmed a full-body image of the child.

After the child was comfortably watching a familiar children's video, the experimenter triggered the presentation of onscreen calibration targets. This was done with software that paused the playing video and presented a calibration target on the otherwise blank screen. A five-point calibration scheme was used, with a variety of small cartoon animations as well as spinning and/or flashing points of light, ranging in size from 0.5° to 1.5° of visual angle, all with accompanying sounds. The calibration routine was followed by a verification of calibration in which more animations were presented at nine on-screen locations.

Throughout the remainder of the testing session, animated targets (as used in the calibration process) were shown between experimental videos to measure drift in calibration. In this way, accuracy of the eye-tracking data was verified before beginning the experimental trials and was then repeatedly checked between video segments as the testing continued. In the case that drift exceeded 3°, data collection was stopped and the child was re-calibrated before further videos were presented.

All aspects of the experimental protocol were performed by personnel 'blinded' to the diagnostic status of child. Most aspects of the data acquisition and all aspects of coding, processing, and data summary are automated so that the separation between the diagnostic characterization protocol and the experimental protocol is assured.

Motion capture stimuli and preferential viewing. Point-light biological motion animations were shown as full-screen audiovisual stimuli on a 20-inch (50.8-cm) computer monitor (refresh rate of 60 Hz non-interlaced). Video frames were 8-bit greyscale images, 640×480 pixels in resolution. The video frame rate of presentation was 30 frames per s. The audio track was a single (mono) channel sampled at 44.1 kHz. The duration of each animation varied with the content of the action, with a mean duration of 30.5 s and a range of 26.4 to 35.5 s. A centring cue lasting 2,800 ms was played immediately before the start of presentation of each animation.

The animations were created with a process called motion capture, in which three-dimensional representations of live performances are recorded in real-time from the movements of actors. The stimuli were created with equipment and support from Animazoo and MetaMotion. Motion is recorded (in three planes of space) directly into computer files by means of an electronic suit worn by the actor. The suit has potentiometers at each joint in the body that track and record movements of the individual wearing the suit. This method enabled us to create a variety of stimuli tailored to young children, featuring routines relevant to childhood experience. As noted, there were five point-light animations portraying an adult's attempts to engage a child. They included the following social approaches: (1) getting the child's attention, (2) playing peek-a-boo ('I can't see you'), (3) playing with a teddy bear, (4) playing pat-a-cake, and (5) enacting a feeding routine.

Preferential viewing in our design was a binary choice, upright versus inverted. To determine viewing preferences that were significantly different from chance, we modelled total viewing time as a binomial distribution. The average viewing time per participant, in the number of frames of video fixated by the toddlers was 5,827 total, 1,165 per animation type. Modelling the binary outcome for this number of frames indicates that results between 47% and 53% should be considered random viewing[31].

Data acquisition and analysis. As noted, visual scanning was measured with eye-tracking equipment (ISCAN, Inc.). The equipment uses a dark pupil/corneal reflection technique with data collected at the rate of 60 Hz (double the frequency of stimuli presentation and of sufficient resolution to identify on- and offset of saccades at a threshold rotational velocity of 30° per s[32]). The eye-tracking camera was mounted remotely, concealed from the child's view behind an infrared filter in a wall panel.

Analysis of eye movements and coding of preferential fixation data were performed with software written in MATLAB. The first phase of analysis was an automated identification of blinks, saccades, and off-screen fixations. Saccades were identified by a velocity threshold. Blinks were identified by eyelid closure (via the rate of change of pupil size and by change in vertical centre-of-pupil data). Off-screen fixations, when a child looked away from the presentation screen, were identified by pupil minus corneal reflection vectors mapping to locations beyond the screen bounds. In the second phase of analysis, eye movements identified as fixations were coded relative to the upright and inverted animations (Fig. 1b–g and Supplementary Movies 1–3).

From within the 304.7 s of total viewing data (9,142 video frames), non-fixation data were not significantly different between the three groups (ANOVA): for all non-fixation data (saccades + blinks + off-screen), ASD = 35.8% (s.d. = 16.4), typically developing = 35.2% (16.1), developmentally delayed = 37.8% (13.2), $F_{2,73} = 0.15$, $P = 0.860$; or separately for saccades, ASD = 15.2% (7.7), typically developing = 13.1% (4.2), developmentally delayed = 15.4% (7.0), $F_{2,73} = 1.3$, $P = 0.277$; for blinks, ASD = 7.4 (7.8), typically developing = 3.9 (5.4), developmentally delayed = 4.7 (5.2), $F_{2,73} = 2.2$, $P = 0.113$; or for off-screen fixations, ASD = 13.3% (12.4), typically developing = 18.3% (13.7), developmentally delayed = 17.8% (12.6), $F_{2,73} = 1.05$, $P = 0.355$.

Quantification of AVS. To quantify AVS, we tracked the locations of the point-lights in our stimuli and compared the change in their motion with the change in the animation's audio soundtrack. Related methods have been described previously[33,34].

We measured the spatial trajectories (x-y location over time) of all point-lights throughout each biological motion animation: 16 point-lights each for the upright and inverted sides of the animation, across five animations (counterbalanced presentations necessarily yielded identical location data, just reversed for left or right presentation). We stored the locations of the point-lights at each frame in each of the animations as a matrix of size $N \times 2 \times 16$, in which the rows (N) signified frames, the columns (2) signified (x, y) coordinate location data, and the Z dimension (16) signified each individual point-light on one side of the animation screen. On the basis of the manner in which the stimuli were created, the location data of the inverted point-light objects were identical to the location data of the upright point-lights except that they were inverted in space and reversed in time.

From each point-light's trajectory, we calculated its velocity over time, and then its change in velocity, $|\Delta v|$. We smoothed the change in velocity data with a moving-average window-size of three samples. This signal, for each of the 32 point-lights in a given animation (16 upright, 16 inverted) provided our measure of change in motion.

To measure the change in audio over time, we measured the audio amplitude of the soundtrack (its short-term amplitude envelope) and then calculated its rate of change, $|\Delta A|$. The short-term amplitude envelope (SAE) of the audio track was computed as the root mean squared (r.m.s.) of a 100-ms square wave moving average of the original audio signal[35]. To normalize this signal for global variance in intensity, we computed two filtered versions of the SAE: one filtered with a moving-average square window of seven samples (local window), and a second with a square window of 35 samples (global window). We then divided the signal filtered at the local window by the signal filtered at the global window. This step is included to normalize for global variance in intensity of a signal while preserving local signal change[36].

Having calculated both change in audio and change in motion for each animation and for all point-lights, we then computed our measure of AVS. By multiplying the change in motion data (each point light, 16 upright, 16 inverted per animation) by the change in audio data (one signal per animation), we generated an audiovisual coincidence matrix for each animation: this gives an AVS value for each point-light at each point in time. High values indicate increased change in motion occurring synchronously with increased change in audio amplitude. Conversely, low values indicate either that one signal was low while the other was high (so the two were not changing synchronously with one another), or that both signals were low (so that with no movement and no change in audio, there was little observable AVS).

To map the AVS values (computed for each point-light) back into the visual space of each presented animation, we overlaid our computed AVS data onto the locations of the point-lights in the original animations (for example, see Supplementary Movie 3). In the movie, colour data are scaled to the maximum AVS value. For each frame of each animation, this generates a matrix of AVS depicting the amount of AVS at each pixel at each frame.

To quantify AVS over the entire duration of an animation, we summed all frames of AVS data, yielding a cumulative map of AVS for each movie. We smoothed these cumulative maps with an averaging filter of size [10 10]. Filter sizes of 6, 8, 10, 12, 16 and 20 all gave similar results. A plot of cumulative AVS is shown in Fig. 3h. AVS level is in arbitrary units (change in motion multiplied by change in audio, summed over all frames), and the maximum value of AVS depends on the number of frames in a given animation (that is, an animation with 1,000 frames is likely to have a larger cumulative signal than one with only 800 frames; to normalize for comparison across animations, we converted to percentages as described later).

To compare AVS on the upright side versus the inverted side, we found the maximum cumulative AVS on each side (for example, 600 on the upright, 400 on the inverted), and converted these values to percentage of total: 600/(600 + 400) = 60%, 400/(600 + 400) = 40%. We then computed a difference ratio of upright to inverted AVS (as plotted in Fig. 4) as the upright percentage minus the inverted percentage: 60% − 40% = 20% (0.2 on the plot in Fig. 4). This generated a normalized score comparable across animations that could be used as a predictor of viewers' looking patterns. We could then test whether or not preferential viewing in our stimuli was related to the level of AVS, as the relative levels of synchrony on the upright versus inverted side provide predictions of expected viewing behaviour.

32. Leigh, R. J. & Zee, D. S. *The Neurology of Eye Movements* 3rd edn 94 (Oxford Univ. Press, 1999).
33. Hershey, J. & Movellan, J. R. in *Advances in Neural Information Processing Systems 12* (eds Solla, S. A., Leen, T. K. & Muller, K. R.) 813–819 (MIT Press, 2000).
34. Bredin, H. & Chollet, G. Measuring audio and visual speech synchrony: methods and applications. *IET International Conference on Visual Information Engineering* 255–260 (2006).
35. Rabiner, L. R. & Schafer, R. W. *Digital Processing of Speech Signals* (Prentice-Hall, 1978).
36. Tchorz, J. & Kollmeier, B. Estimation of the signal-to-noise ratio with amplitude modulation spectrograms. *Speech Commun.* **38,** 1–17 (2002).