Distinct neural mechanisms of social orienting and mentalizing revealed by independent measures of neural and eye movement typicality

Authors: Michal Ramot*, Catherine Walsh, Gabrielle E. Reimann, Alex Martin

Affiliation:

Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892

*Correspondence to: michal.ramot@nih.gov

Abstract

Extensive study of typically developing individuals and those on the autism spectrum has identified a large number of brain regions associated with our ability to navigate the social world. Although it is widely appreciated that this so-called 'social brain' is composed of distinct, interacting systems, these component parts have yet to be clearly elucidated. Here we used measures of eye movement and neural typicality - based on the degree to which subjects deviated from the norm - while typically developing (N = 62) and individuals with autism (N = 36) watched a large battery of movies depicting social interactions. Our findings provide clear evidence for distinct, but overlapping, neural systems underpinning two major components of the 'social brain', social orienting and inferring the mental state of others.

Introduction

Movie viewing involves many complex mental tasks. These include, though certainly are not limited to, applying mechanisms of directed attention to select the most relevant information, and understanding the behavior of the characters and predicting their future actions through mentalizing. In social, dynamic scenes, attentional selection (measured through eye movements) is driven not only by low-level visual features, but is preferentially modulated by social cues. Saliency maps rely on low-level features such as contrast, color and motion to predict fixations [1-3]. Yet the presence of faces in the scene is a better predictor of fixations than saliency maps [4,5], and orientation to faces is further enhanced in the presence of accompanying speech [6]. Other social cues, such as gaze direction, emotion and touch are also better at predicting attentional focus than low-level visual features [7], and information derived from head orientation and body position on top of gaze direction have also been shown to strongly modulate social attention [8,9].

Directors of Hollywood movies are particularly adept at manipulating the focus of our attention, using cinematic techniques to tightly control where viewers' attention is drawn [10-13]. This creates a movie experience which is robustly shared across viewers, with previous studies describing widespread correlations in neural responses across individuals, extending well beyond perceptual regions into social processing areas, among others [14]. This attentional synchrony however, seems to be dominated by transient visual and social cues, and is only very weakly modulated by higher level comprehension of the narrative, or inferences based on the mental states of the characters in the scene, as is demonstrated by studies which manipulated comprehension through temporal shuffling of scenes [15], or manipulation of available context [16,17].

It is particularly interesting to consider participants with Autism Spectrum Disorder (ASD) in the context of this disassociation between higher order comprehension / mentalizing and social orienting. Social deficits and impairments in social processing are among the defining

characteristics of ASD [18]. For adolescents and young adults on the high functioning end of the spectrum however, these manifest most consistently as deficits in complex mentalizing or theory of mind tasks, though these deficits can be subtle and difficult to probe experimentally [19-22]. These difficulties also extend to social orienting. High functioning adolescents and young adults with ASD exhibit aberrant social orienting as manifested by aberrant eye movements to faces and other social stimuli, though, again, differences can be subtle and are usually only apparent when examined with complex stimuli or sensitive metrics [23-25].

Movie viewing is an experimental environment uniquely suited to the study of the social brain. On top of the robust and widespread basis of shared responses to the movie, both behavioral and neural, there is a range of individual variance [26]. Eye movement patterns, reflecting attentional selection, vary across individuals [27], as do all aspects of their social comprehension, from basic understanding to empathy for the characters in the scene [28]. This provides an exceptional opportunity for uncovering links between brain and behavior, while the complexity and depth of the social stimuli make it ideal for picking out subtle differences in high functioning ASD. Previous research has focused on correlating the typicality of neural responses during movie viewing – as measured by the inter-subject correlations (ISC) of the neural response time course to the movie - to behavior related directly to the movie in question, such as memory for specific scenes [29]. Differences in ISC between typically developing (TD) and ASD groups have been examined in only a handful of studies, mostly with very few participants and mixed results [30-32]. Similarly, while many previous studies have used movies to study the behavioral aspects of social orienting [33-35], the search for the neural correlates of social orienting has so far utilized only very simple, mostly static and schematic, social stimuli. Moreover, these studies have focused on probing gaze following, which is only one aspect of social cues [9]. These limitations have led to a partial, fragmented understanding of the social orienting network.

Here we sought to exploit the full capacity of the movie viewing environment by expanding the analysis to include an independent measure of behavior, which allowed us to make inferences

which generalize beyond our specific movie stimuli. We compared measures of typicality for eye movements (measured by distance from the average scan path) while watching movie clips outside the scanner with the typicality of neuronal responses derived from voxel-wise ISC while watching a different movie during fMRI acquisition in a large cohort (62 TD and 36 high functioning participants with ASD). This allowed us to uncover the broad neural underpinnings of the social orienting network, providing a much more detailed and complete delineation of this network. A group comparison of the typicality of neuronal responses in TD and ASD participants revealed a second, distinct network, which in a manner congruent with the above observations does not correlate with the eye movements, but instead corresponds to regions previously implicated in mentalizing and theory of mind. Together, these findings present direct evidence and a comprehensive description of two fundamental components of the social brain.

Results

62 TD participants (24 female) and 36 participants with ASD took part in this study. Of the TD group, 36 were matched to the ASD group, in terms of gender (all male), age, IQ and motion (see Methods for more details). In analyses where the TD group is considered separately the full TD dataset was used, whereas in analyses comparing or combining the two groups, only the matched TD subset was used. All participants completed a behavioral session outside the scanner, in which they watched 24 short (14s) movie clips taken from popular Hollywood movies while their eye movements were being recorded. 3 ASD participants and 2 TD participants did not achieve adequate calibration and were removed from the eye-tracking portion of the analysis. These movie clips were chosen in a separate pilot study from a larger set of 60 movie clips, for eliciting the most consistent viewing patterns across subjects. Immediately following the behavioral session, participants took part in an fMRI scan session without eye tracking, during which they watched a 9.5-minute clip taken from a different movie. All movie clips outside and inside the scanner depicted social scenes, with interactions between at least two characters and were presented with sound (see Methods). Additionally,

informant versions of the Social Responsiveness Scale (SRS) measure were obtained from the parent or guardian for the ASD participants.

Eye movement typicality

Even within these carefully selected movie clips, there was a range of individual eye movements, with some participants having more typical viewing patterns than others. For the TD group, we quantified the typicality of eye movements for each participant and each movie by calculating the Euclidean distance of their eye movements from the mean scan path of all other participants for each frame, averaging across all frames of that movie. This method gave us a distance measure of how similar that participant's viewing pattern was to the average viewing pattern of all other TD participants, per movie. Figure 1a shows an example of the mean scan path of all TD participants for one movie clip, with the eye movements of the most typical and least typical participants plotted in green and red, respectively. Supplementary Movie 1 shows these as fixations overlaid on the movie clip. The average eye movement typicality of each participant was defined as the average distance across all 24 movie clips, with an inverse relationship between the two, so that the smaller the distance measure, the more typical the eye movements. This typicality metric served as a measure of the degree to which individual participants' eye movements differed from the group norm.

For the ASD group, we calculated two measures of typicality for each participant: one by quantifying the distance from the mean of all other participants in the ASD group, and the other by quantifying the distance from the mean of all the matched TD participants. The two measures were nearly identical, with a correlation of $r = 0.99$ across participants between the average typicality when compared to the others in the ASD group, and the average typicality when compared to the TD group (Figure 1b). For individual movies, this correlation ranged between $r = 0.93$ to $r = 0.995$. This is in line with our finding that at the group level, the average scan paths were very tightly coupled across groups – the average ASD scan paths for each movie to the average scan paths of both the matched TD subset, and the full TD group, were

always more similar to each other than the average TD scan path was to the most typical TD participant, and correlations along the horizontal scan path were greater than 0.88 for all movies. An example for the average scan paths for the two groups is shown in Figure 1c. Given the high correlation between the two measures, we decided to henceforth define the eye movement typicality for the ASD group as the Euclidean distance from the mean scan path of the TD group, as this more obviously represents "typical" movie viewing and social orienting at the population level.

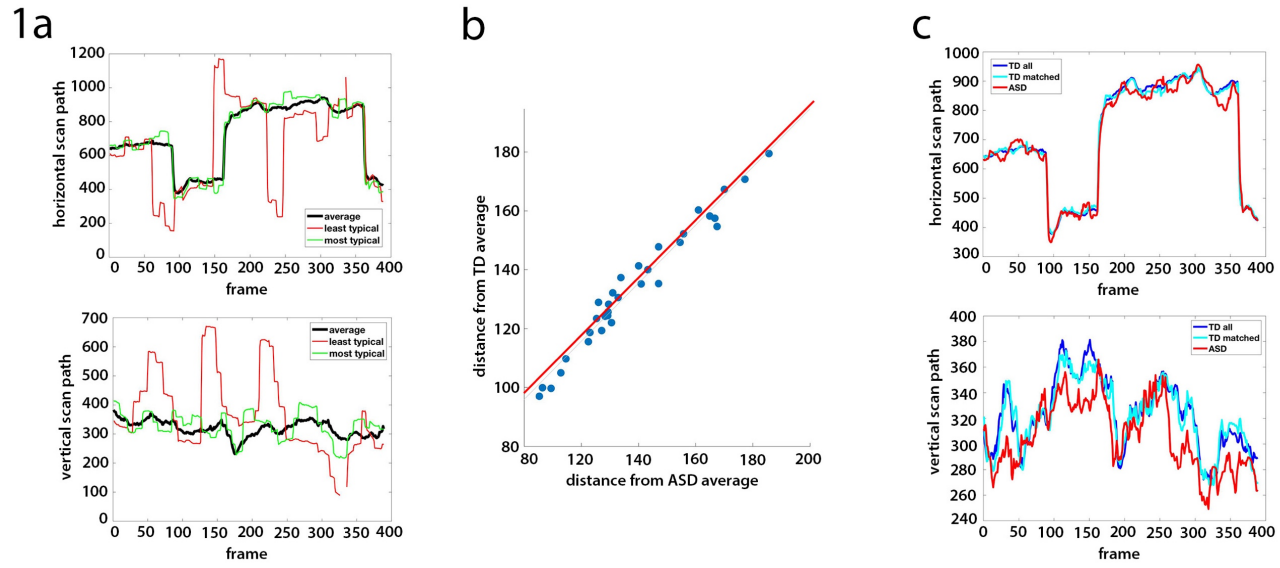

Figure 1: eye movement typicality for TD and ASD participants. Figure 1a shows an example average scan path (black) along the horizontal (top) and vertical (bottom) dimensions, for one sample movie clip. Overlaid are the scan paths for the most typical (green) and least typical (red) TD participants. Figure 1b shows the correlation between the typicality (shown as a distance measure) for each ASD participant (averaged across movies) when compared to the average of all other ASD participants (x-axis), vs. the average typicality when compared to the average of all matched TD participants (y-axis). Each dot denotes one ASD participant. 1c shows the average horizontal (top) and vertical (bottom) scan paths for the same movie as that shown in panel 1a for all TD participants (blue), the matched TD participants only (cyan), and all ASD participants (red).

Stability of eye movement typicality

To test whether this measure of eye movement typicality is a robust and stable individual subject trait, we first divided the 24 movie clips into two sets (odd and even) and calculated the mean typicality for each participant for each set of 12 movies. We next correlated this mean typicality between the two movie sets across participants, separately for the TD and ASD group. Figure 2 shows this correlation between the two sets of data, with the TD participants plotted in blue ($r = 0.78$, $p = 1.9 \times 10^{-13}$ for all TD participants, $r = 0.70$, 2.8×10^{-6} for the matched controls) and the ASD participants plotted in red ($r = 0.73$, $p = 8.5 \times 10^{-7}$). Finally, we repeated 10,000 iterations of this analysis, randomly dividing the movies into two sets each time, and calculated the mean correlation between the two data sets. To gauge the likelihood of randomly getting such correlations between the two halves of the data, we carried out a permutation test, randomly shuffling the subject labels for each iteration. The unshuffled mean correlations across iterations for all TD participants ($r = 0.73$) for the matched subset of TD participants ($r = 0.67$) and for the ASD participants ($r = 0.75$), were all entirely outside the random distribution.

Despite the high correlations between the average scan paths of the two groups, TD participants had significantly more typical eye movements than their ASD counterparts (two sample two tailed t-test of the matched TD vs. ASD participants, $p = 6.57 \times 10^{-6}$). Both groups exhibited a wide range of individual differences in eye movements (note the range of the distance measure across participants, Figure 2), but the variance in typicality of the ASD group was significantly higher than in the matched TD subset (533 and 163, respectively F stat 3.42, significant at $p = 6.57 \times 10^{-4}$), pointing to a wider range of behavior within the ASD group. This variance in behavior within the ASD group was not explained by social behavior though, as measured by the SRS ($r = -0.14$, $p = 0.45$). Together, these results reveal the existence of a typical, “ideal” scan path for these movie clips, which is the same for TD and ASD participants. The difference between the TD and ASD groups is driven by increased variance and increased deviation from the same typical scan path in the ASD group.

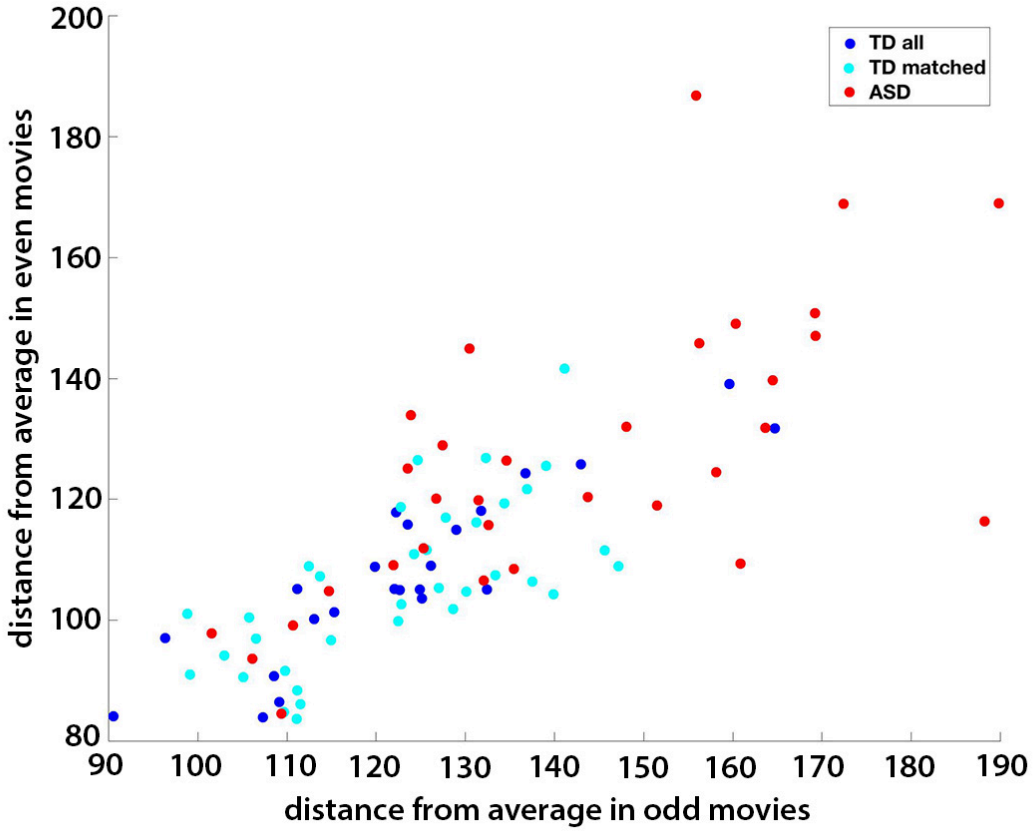

Figure 2: stability of eye movement typicality. Mean typicality (shown as a distance measure) for each of the TD and ASD participants, averaged across all odd movies (x-axis), vs. the mean typicality for each participant averaged across the even movies. Matched TD group participants shown in cyan, the remaining TD participants shown in blue, and ASD participants in red.

Neural typicality

To assess the typicality of the neural responses for each participant during movie viewing, we analyzed the fMRI data acquired while participants were watching a different 9.5-minute movie inside the scanner. We calculated the correlation of the time course of each voxel to the average time course of that voxel for all the other participants, giving us a measure of how typical (i.e. similar to the average) the neural responses to the movie were for that participant, per voxel. The map in Figure 3 shows the average typicality of each voxel, defined as the mean

typicality for that voxel across all TD participants. High typicality values indicate a high level of correlations across individuals in the activity of that voxel during the movie. High levels of these inter-subject correlations (ISC) while watching an engaging movie spanned large areas of the cortex, including but not limited to sensory regions, spreading into many regions of association cortex, though sensory regions tended to be the most highly correlated between subjects. This is in accordance with several previous studies of ISC during movie viewing [14,36,37]. Note that due to scanning parameter constraints, the field of view did not cover the entire brain, with some areas primarily in motor cortex missing coverage. Additionally, some voxels were removed from the analysis for failing to meet minimal signal to noise requirements (see Methods).

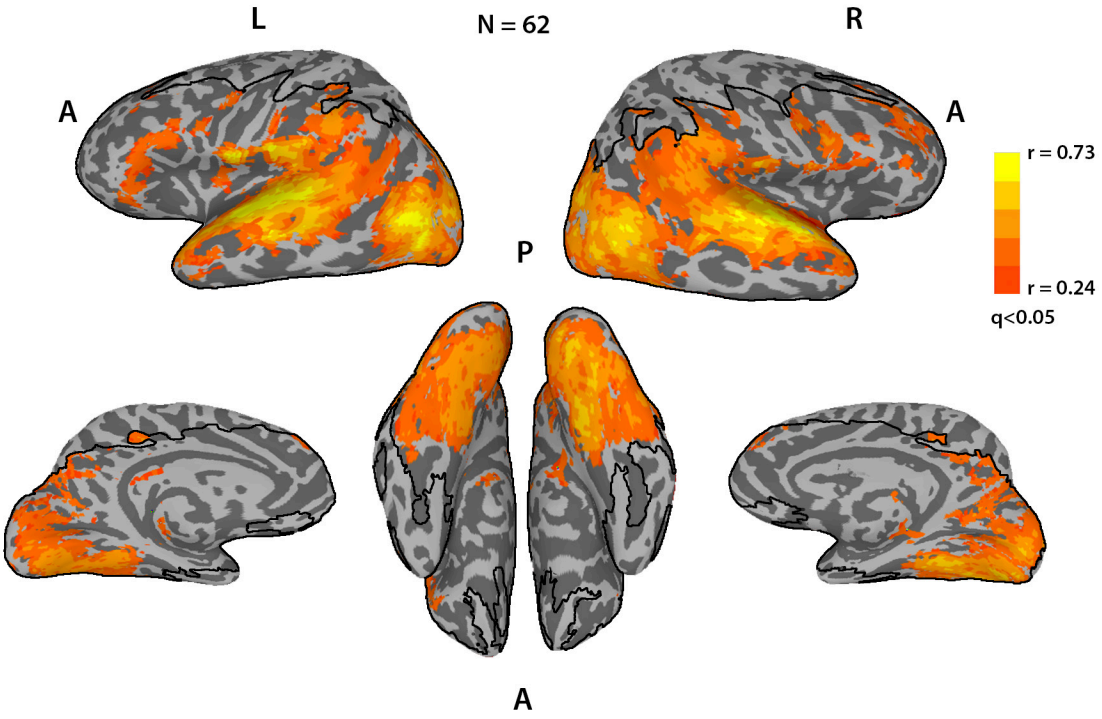

Figure 3: average neural typicality. Average neural typicality (ISC) across the entire brain for the TD subjects, thresholded at q<0.05, FDR corrected. Black lines delineate the field of view, voxels outside this boundary were not imaged or were removed from the analysis for poor temporal signal to noise ratio (tSNR).

We carried out the same analysis for the ASD group, and as with the eye movements, calculated neural typicality for each participant for each voxel to both the average time course

of that voxel for all other ASD participants, and to the average time course of all the matched TD participants. As with the eye movements, these measures were very tightly correlated (mean correlation across participants was r = 0.92, calculated across voxels for each participant, and then averaged across participants), so we will henceforth define the neural typicality of the ASD group in relation to the TD average, for similar reasons as above.

Correlation between eye movement typicality and neural typicality

In order to search for the neural underpinnings of social orienting, we sought to combine these two independent measures of brain and behavior by conducting a whole brain search for voxels in which there was a correlation across participants of the typicality of the neural responses to the movie and the typicality of the eye movement patterns during the short movie clips shown outside the scanner. Note that the eye movement typicality measure as we have defined it is in fact a distance measure, meaning that the greater it is, the less typical the eye movements. Voxels in which there was a significant anti-correlation between the two measures are those in which the more typical (similar to the average) the neural response to the movie, the more typical the eye movements of that participant were to the short movies outside the scanner.

Figure 4a shows the results of this analysis for the TD group, revealing multiple regions associated with social and language processes, including superior temporal sulcus (STS), inferior frontal gyrus (IFG), anterior insula, posterior and anterior cingulate cortex (PCC and ACC respectively), medial prefrontal cortex (MPFC) and subcortically the hippocampus, putamen, and caudate, bilaterally, for which eye movement typicality is strongly correlated with neural typicality in response to a movie (corrected for multiple comparisons through a permutation-based cluster size correction, $p<0.05$).

We next tested whether this same network also underlies social orienting in participants with ASD, i.e. will there be a significant correlation between eye movement typicality and neural typicality within this network for the ASD group. We defined a mask of the voxels which were

found to be significant in the TD analysis and averaged the neural typicality within this mask for each of the ASD participants. We then correlated this average neural typicality value with the eye movement typicality value across all the ASD participants, and found a significant correlation ($r = -0.42$, $p = 0.01$). Supplementary Figure 1 shows a scatter plot with these data, overlaid with a similar analysis for the entire TD group ($r = -0.56$, $p = 4.2 \times 10^{-6}$) and the matched TD group ($r = -0.59$, $p = 1.6 \times 10^{-4}$). The high correlations for the TD groups are expected, as it is this correlation which was used to define the network. These data are shown together only to put the ASD data in context.

Since the same network seemed to underlie social orienting in both TD and ASD participants, we carried out an additional analysis on the combined group of the ASD participants with the matched TD participants. Figure 4b shows the overlay of this analysis with the analyses of just the matched TD participants (blue) or just the ASD participant (red). Combining the groups (the ASD group with the matched TD group) gave more widespread correlations with the eye movements, and these overlapped with the TD only and ASD only analyses substantially, with only 14% of the voxels in those two analyses not contained within the combined group analysis.

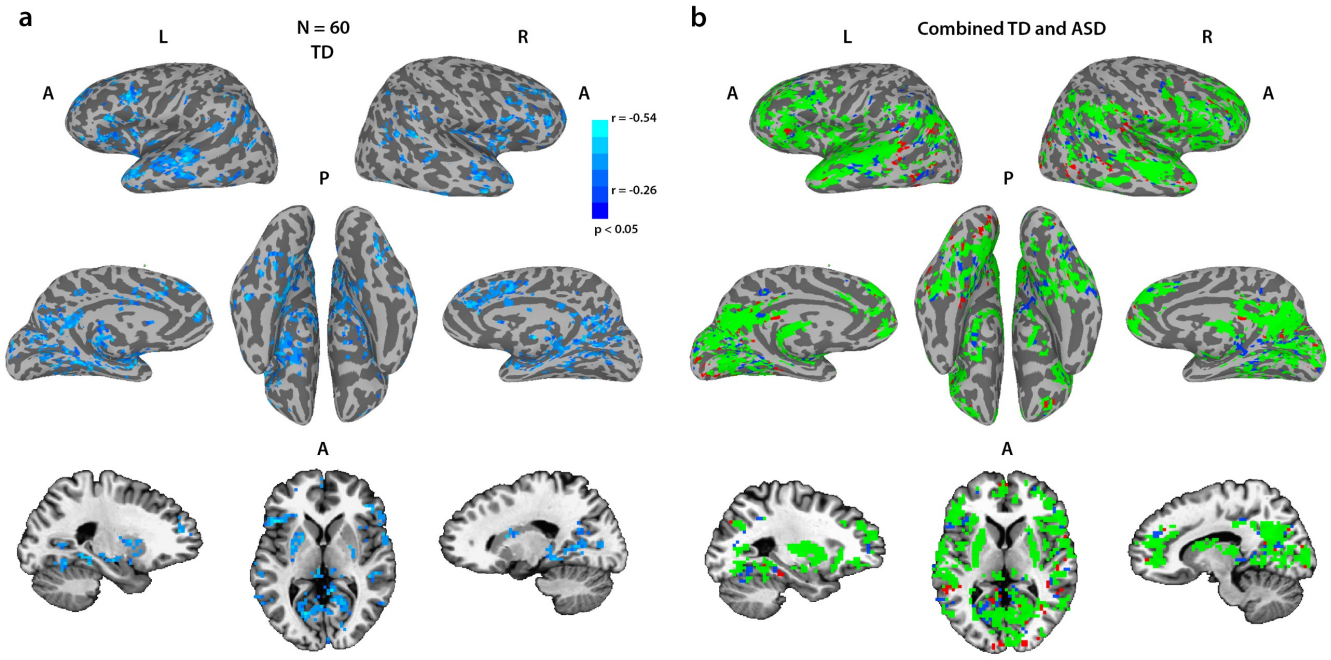

Figure 4: correlations between neural typicality and eye movement typicality. (a) Correlations between eye movement typicality and neural typicality for the full TD group at a corrected threshold of $p<0.05$, corrected through cluster size permutation testing. Panel b shows all voxels whose neural typicality was significantly correlated with eye movement typicality, for the combined group of the ASD participants and their matched controls (green, N=69), with voxels that were only correlated for the corresponding analysis of the matched TD group (N = 36) in blue, and voxels that were only correlated for the corresponding analysis of the ASD group (N = 33) in red. Note that the combined analysis is more widespread than either analysis separately.

Group differences in neural typicality

The eye movement analysis captured many similarities between the two groups, and yet there are clearly differences in social processing between them. We hypothesized that social difficulties in the ASD group would translate to less typical processing in the relevant social brain regions and would therefore be reflected in reduced neural typicality in those areas. To test this, we carried out a group t-test on the neural typicality measure of the matched TD participants and the ASD group, and the result is displayed in Figure 5. Significant differences between the TD and ASD groups were found mainly in the right temporal parietal junction (TPJ) and middle temporal gyrus, and bilaterally in posterior STS, IFG, anterior insula, ACC, PCC, putamen and caudate.

Surprisingly, many of the voxels in the social orienting network identified by the eye movement typicality analysis, did not show significant differences in neural typicality as we would have predicted, considering the difficulty in social orienting that is also a hallmark of ASD. To test whether this was a result of a lack of sensitivity of the neural typicality measure, we examined the neural typicality group difference when averaging the neural typicality across the entire social orienting network, as defined by the TD group alone, and also as defined by the combined matched TD and ASD group (see Figure 4). We found that even though there were no significant group differences in many of the individual voxels, the average neural typicality across the entire network was indeed significantly greater for the TD group both when using the network definition derived from the TD group alone ($p = 3.5 \times 10^{-4}$, two-tailed two sample t-

test using the matched TD group), and when using the network definition derived from the combined group $p = 8.8x10^{-5}$, two-tailed two sample t-test using the matched TD group). At the network level therefore, the social orienting network showed significantly greater neural typicality for the TD group compared with the ASD group.

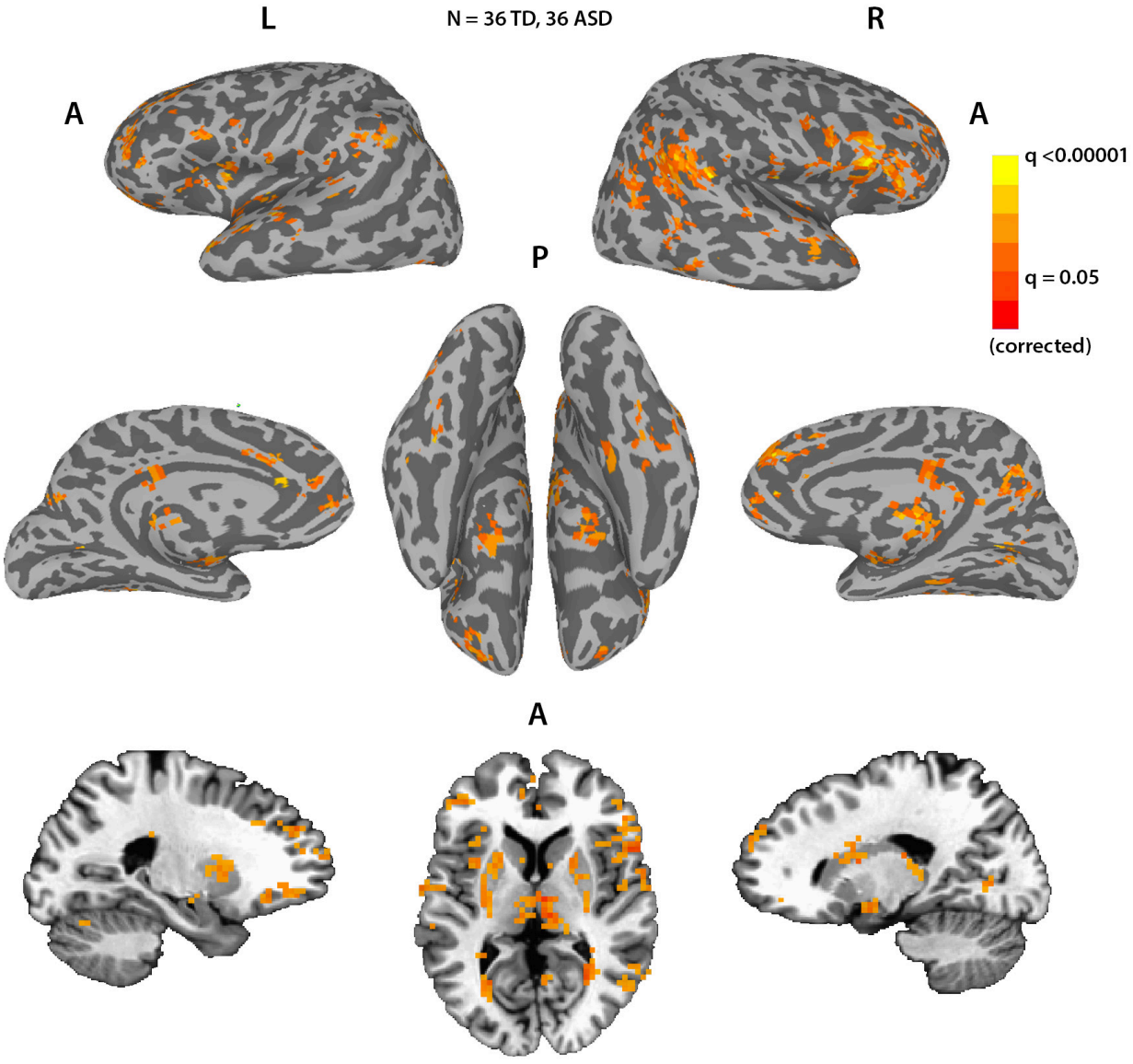

Figure 5: group differences in neural typicality. Voxels that had significantly greater neural typicality for the matched TD group than the ASD group. No voxels showed significant differences in the other direction. FDR corrected.

Two distinct networks

So far we have used two separate and very different analyses to identify two networks – the social orienting network identified through correlations of neural typicality with eye movement typicality (Figure 4), and a second network, defined by group differences in neural typicality between the TD and ASD groups (Figure 5). For simplicity, we will refer here to this second network as network 2. To examine the overlap between these two networks, we created a conjunction map of voxels belonging to just one of these analyses or to both, using a threshold of $p<0.01$ and correcting for multiple comparisons through cluster size permutation testing (Supplementary Figure 2). While there are some areas of overlap in STS, IFG, anterior insula, ACC, PCC and putamen, other regions belong only to the social orienting network (anterior STS bilaterally, left pSTS and IFG, and most of ACC and PCC bilaterally), or are not correlated with eye movement typicality but show group differences in neural typicality, making them a part of network 2 (right TPJ, middle temporal gyrus, amygdala). To test whether the non-overlapping parts of these networks are functionally distinct, we examined whether the average correlation of each voxel to all the other voxels within the network was significantly greater than its average correlation to all the voxels in the other network. There was a significant difference for the social orienting network, with correlations within the social orienting network significantly greater than correlations between the social orienting network and the network 2 for both the TD and the ASD groups (paired two-tail t-test, $p < 7.4 \times 10^{-14}$ for both groups). Correlations within network 2 were significantly more correlated within than across networks for the TD participants ($p = 2.9 \times 10^{-8}$), but there was no significant difference for the ASD group ($p = 0.61$). Despite this lack of differentiation for the within vs. across correlations within network 2 for the ASD participants, the subset of voxels showing significantly greater within than across network correlations in the TD participants overlapped almost entirely with the subset of voxels in the ASD group which were similarly more correlated within rather than across network, for both networks (99% overlap for voxels in the social orienting network, and 97% overlap for voxels in network 2). Figure 6 portrays the non-overlapping subset of the social orienting network and network 2 identified in Figures 4 and 5, further limited to the voxels which show consistently

greater within than between network correlations in both the TD and the ASD groups. There is an apparent laterality bias difference between the two networks, with the network 2 biased towards the right hemisphere (78% of voxels fall in the right hemisphere), while the social orienting network has a (weaker) left hemisphere bias, with 60% of voxels on the left.

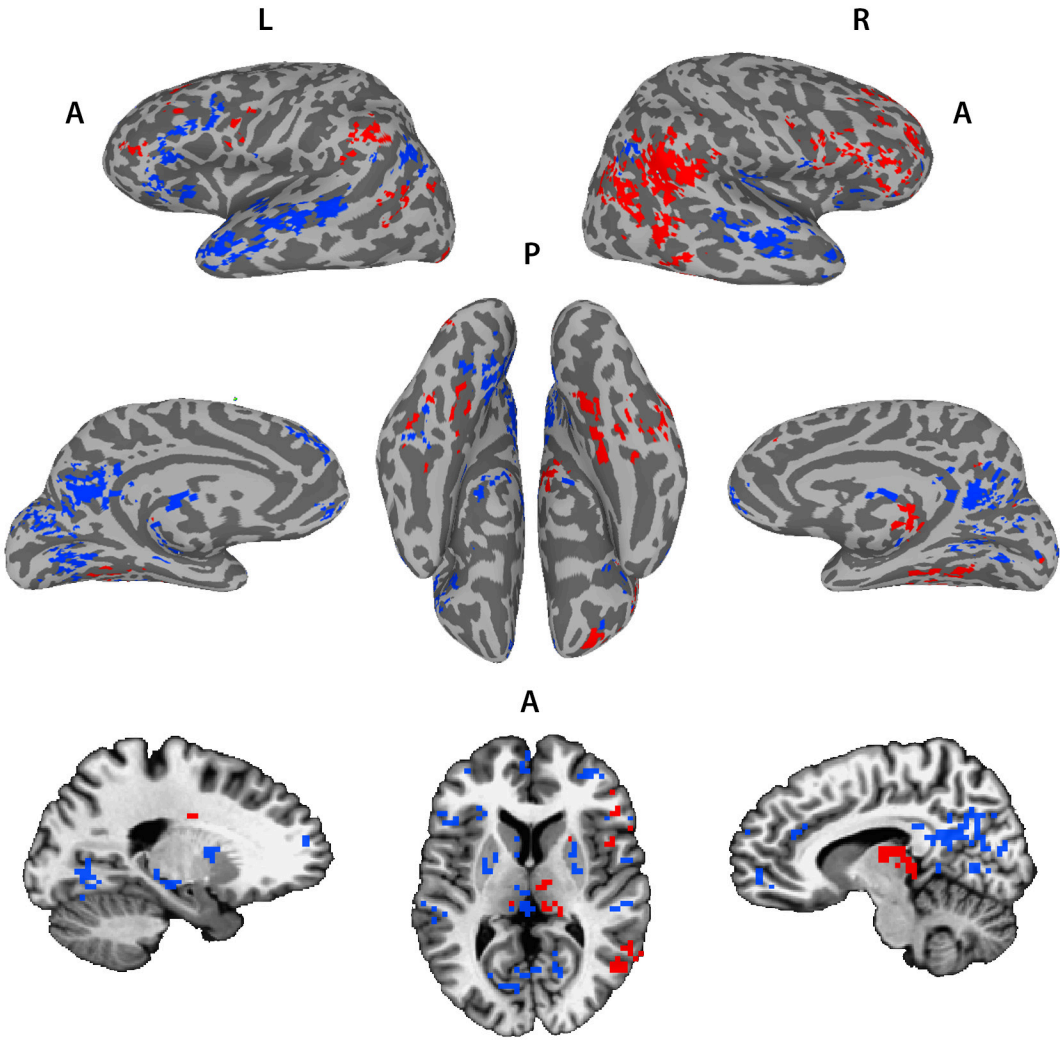

Figure 6: two distinct networks derived from neural typicality group differences and eye movement correlations. Overlay of the voxels with significant correlations between neural typicality and eye movement typicality of the combined matched TD and ASD groups (N = 36 TD + 33 ASD, blue), and voxels showing significant group differences in neural typicality between the matched TD and ASD groups (N1 = 36 TD, N2 = 36 ASD, red).Threshold set at $p<0.01$.

To further examine the distinction between the networks, we tested whether the neural typicality within these two networks shown above was differentially correlated with the SRS in the ASD group. This was indeed the case, with the neural typicality averaged across network 2 correlated to the SRS at r = -0.35, p = 0.037, whereas the neural typicality averaged across the social orienting network correlated to the SRS at r = -0.15, p = 0.38. To test whether the correlation of the neural typicality difference network was significantly greater we used the Steiger z-score test for the difference between two dependent correlations with one variable in common, and the result was significant with a one tailed test (z = -1.9, p = 0.028). Note that these correlations are negative, with a higher (more impaired) SRS score correlating with lower neural typicality, as expected. The regions which overlapped between the social orienting network and network 2 were not significantly correlated to the SRS (r = -0.22, p = 0.21), but also not significantly different from the correlation of the non-overlapping subset of network 2 to the SRS (z = 1.2, one-tailed test p = 0.11).

Discussion

Movie viewing evokes both shared neural responses, and shared behavior, in the form of eye movements orchestrated by carefully constructed visual, auditory and social cues. Yet there is individual variation in both behavior and neural responses. We used the correlation between these two measures across participants to search for the network responsible for social orienting. We also examined differences in neural typicality between a TD group, and a socially impaired ASD group. Based on previous studies which have demonstrated a decoupling of high order comprehension / mentalizing during movie viewing from eye movement patterns [15], our prediction was that differences in neural typicality between a TD group and an ASD group, which has difficulties in both social orienting and mentalizing, would reveal a separate mentalizing network, which would not overlap with the social orienting network.

At the network level, there was significantly lower neural typicality in the ASD group within the social orienting network. This was expected, as on top of previous research [23-25], our own data

described here shows reduced typicality of eye movements in the ASD group compared with the TD group (Figure 2), and this should be reflected in the typicality of neural processing. However, the more robust effects, which could be seen at the single voxel level (Figure 5) centered on regions which seemed to constitute a separate, distinct network, whose correspondence to the mentalizing network will be discussed below.

Eye movements have been used extensively to study social orienting [23,38,39], though to our knowledge this is the first time that such a measure of typicality has been used in this way. The idea that this would be an informative measure is based on the notion that certain stimuli, especially those created explicitly to draw attention in very specific ways such as Hollywood movies, would have a "typical", or in a sense ideal scan path. Indeed, our movie clips were pre-selected in a pilot study (on a separate group of participants) for the consistency of the eye movement patterns they elicited. Despite this being an obvious oversimplification which disregards details and dynamics, the striking similarity between the mean scan path of each of our movie clips not only within TD participants but also between TD and ASD participants as is demonstrated in Figure 1, indicates that such a stereotypical scan path for these clips exists.

Having established that there is a basis for considering typicality, we next investigated whether eye movement typicality would prove to be a stable individual trait, which is not specific to a particular stimulus or context (at least within the broader framework of social movie viewing). This would be a prerequisite for using it as a marker for general social orienting abilities. The highly significant correlations of eye movement typicality across different movie clips (Figure 2) pointed to this being a robust measure of social orienting at the level of the individual. This gave us an independent behavioral measure, with which to search for the neural correlates of social orienting.

Similar to the eye movements, the concept of neural typicality, more often referred to in the literature as inter-subject correlations (ISC), is based on the premise that neural responses across individuals will be similar when driven by a shared stimulus, in regions whose processing

is stimulus related [36]. Figure 3 replicates the results of several studies, which have consistently shown significantly shared neural responses across these same cortical regions during movie viewing [14,36,37].

Evidence that these correlations between subjects indicate shared processing, or a shared experience relevant to the function of the correlated region, can be found for example in studies that have demonstrated that for participants viewing a movie, or listening to a narrative, the similarity of their interpretation of the story predicts the degree of neural similarity in regions involved in narrative interpretation, such as the fronto-parietal network and the default mode network [40]. Likewise, participants recalling the same events have been shown to have more neural similarity than participants recalling different events [41]. A whole brain search for voxels whose neural typicality is strongly coupled to eye movement typicality across subjects was therefore a natural next step to revealing the neural substrate of social orienting.

The regions which exhibited a correlation between the two measures (Figure 4) were very stable between the TD and ASD groups (Supplementary Figure 1). Monkey electrophysiology, as well as lesion studies in both monkeys and humans, have all pointed to an important role for STS in gaze perception and social orienting [42-44]. Human neuroimaging studies have suggested that networks involved in gaze perception extend well beyond STS, including also ACC, MPFC, and hippocampus, see [9] for a review. Recent studies have further expanded the networks involved in social orienting, with one study finding increased activation in STS, IFG and the putamen when comparing gaze cues to non-social symbolic cues [45]. This study further found an interaction effect between groups (TD vs. ASD) and cue type (social vs. symbolic). These human neuroimaging studies are limited however by the very simplistic and specific nature of their stimuli (mostly cartoon gaze cues), which is perhaps why we were able to uncover a broader network linked to social orienting, which in our case is derived from a host of social cues which can be extracted from the naturalistic and dynamic movies. Please note however that we do not have full coverage of the brain, and some areas which might also be involved in social

orienting, most notably frontal and supplementary eye fields, are missing from our analysis (Figure 3).

The eye movement analysis was intended to capture measures of social orienting, but it did not directly address the differences between the TD and ASD participants. To directly test for this, we examined the group differences in neural typicality between our TD and ASD participants (Figure 5). There have been very few studies looking at differences in neural typicality or ISC between TD and ASD groups. The first [31] found differences in visual and auditory regions, but was very underpowered (with 12 ASD and 8 TD participants), while another found differences only between a subset of ASD participants [30]. A third study, though only slightly less underpowered (N=13 in each group), used a full length 67 minute movie, which may have compensated for the low number of participants [32]. This study identified very similar regions to those found in our analysis. These regions have been shown in the past to be involved in mentalizing and in theory of mind [22,46-49]. Right TPJ in particular has long been thought to be fundamental for theory of mind [50-52], whereas left TPJ, MPFC, ACC, PCC and IFG, have all been found to activate for various mentalizing and theory of mind tasks (see Schurz et al. [53] for a meta-analysis). In a study reporting differences between TD and ASD groups in a task which required inferring intentionality from eye gaze, the same regions in right TPJ (referred to there as posterior STS) and middle temporal gyrus were identified as activating more in the TD group [54]. Amygdala on the other hand, is consistently found to be crucial for emotional processing [55,56]. The fact that significant neural typicality differences, at least at the individual voxel level, were found only for these regions, suggests that these regions are involved in the processing that is most atypical in ASD, which should correspond to the social tasks with which the ASD have the greatest difficulty.

The double disassociation between the two networks identified through the eye movement and neural typicality group difference analyses and our two behavioral measures (eye movement typicality, SRS), together with the consistency of greater within compared to across network functional connectivity (Figure 6), strongly point to these being two distinct networks.

Considering the nature of the processing that takes place while watching social movies, the nature of the deficits in ASD, and the previous literature on the regions identified by this analysis, we hypothesize that this second network revealed by the neural typicality group difference analysis corresponds to the mentalizing network. We also found substantial overlap between the two networks, most notably in right IFG, MPFC and PCC, and bilaterally in regions of putamen and the caudate (Supplementary Figure 2). As the two functions of social orienting and mentalizing interact constantly, so must the networks that guide them, and it is perhaps in these regions of overlap that this neural interaction takes place.

Methods

Participants

36 males aged 15-30 (mean age = 20.7) who met the DSM-IV criteria for autistic disorder, an autism cut-off score for social symptoms on the Autism Diagnostic Review (ADR) and/or and ASD cut-off score from social and communication symptoms on the Autism Diagnostic Observation Schedule (ADOS), all administered by a trained, research reliable clinician, were recruited for the experiment. In addition, 63 typically developing participants (24 female) aged 15-30 (mean age = 22.05) were recruited. One was excluded from the analysis because of abnormal brain structure. Of the remaining 62, a subset of 36 males were chosen to match the ASD group, based on gender, age, IQ and motion. Mean age for the matched TD group was 20.8 (range 15-28). For the eye movement analyses, three participants from the ASD group and two from the TD group were excluded for failing to achieve adequate calibration on the eye tracker. All participants were right handed, and had normal or corrected to normal vision. IQ scores were measured by the Wechsler Abbreviated Scale of Intelligence, the Wechsler Adult Intelligence Scale-III, or the Wechsler Intelligence Scale for Children-IV. Full scale IQ scores were all > 94, and were matched between the ASD and the TD groups. The experiment was approved by the NIMH Institutional Review Board (protocol 10-M-0027). Written informed consent was

obtained from all participants or their guardians in the case of minors, in which case written assent was also obtained from the participants themselves.

Eye tracking setup

Eye tracking was recorded with the Eyelink 1000 Plus. Participants' heads were stabilized using a chin and forehead rest, and eye gaze calibration was performed at the beginning of the viewing session for each participant. The movie clips were shown in a randomized order, with a brief 6s pause in between successive clips, during which time a gray screen with a fixation cross was presented. The same screen and fixation were also presented before the first movie presentation. Movies were viewed on a digital monitor with a 1920 x 1080 resolution, and eye tracking data were sampled at 1000 Hz.

Eye tracking preprocessing and analysis

Eye movement data was extracted for each movie clip separately. The first and last 500ms were removed, blinks and missing (offscreen) data were ignored, and the data was despiked. Data was then down sampled from 1000 Hz to the frame rate at which the movies were presented (29.97 fps). This was then used to calculate the position of the eye fixation on the screen for each frame, as well as the average position for all (other) participants in the typicality analyses.

Movie stimuli

An initial set of sixty 14s movie clips was tested on an independent pilot set of 12 participants. The movie clips were then analyzed for the consistency of the eye movement patterns they elicited across these participants, and the 24 most consistent movie clips were selected for the eye tracking session outside the scanner. These movie clips were all taken from Hollywood movies – The Blind Side (6 clips), Goonies (4 clips), How To Lose a Guy in Ten Days (4 clips), The Italian Job (5 clips), and The Neverending Story (5 clips). There was ongoing dialogue between

at least two characters in all these scenes, though characters were often not on the screen together. For the 9.5 minute movie clip shown during fMRI acquisition, a scene from the Princess Bride was selected. This scene involved multiple characters on screen simultaneously, as well as continuous dialogue. All movies were presented with audio.

Imagining data collection, MRI parameters and preprocessing

All scans were collected at the Functional Magnetic Resonance Imaging Core Facility on a 32 channel coil GE 3T (GE MR-750 3.0T) magnet and receive-only head coil, with online slice time correction and motion correction. The scans included a 5 minute structural scan (MPRAGE) for anatomical co-registration, with the following parameters: TE = 2.7, Flip Angle = 12, Bandwidth = 244.141, FOV = 30 (256 x 256), Slice Thickness = 1.2, axial slices. EPI scans were collected with the following parameters: TR = 2s, Voxel size 3*3*3, Flip Angle: 60, multi-echo slice acquisition with three echoes, TE1 = 17.5ms, TE2 = 35.3ms, TE3=53.1ms, Matrix = 72x72, slices: 28. 285 TRs were collected for the movie (9 minutes and 30 seconds). All scans used an accelerated acquisition (GE's ASSET) with a factor of 2 in order to prevent gradient overheating.

Post-hoc signal preprocessing was conducted in AFNI (Analysis of Functional Neuro-Images, [57]. The first four EPI volumes from each run were removed to ensure remaining volumes were at magnetization steady state, and remaining large transients were removed through a squashing function (AFNI's 3dDespike). Volumes were slice-time corrected and motion parameters were estimated with rigid body transformations (through AFNI's 3dVolreg function). Volumes were co-registered to the anatomical scan. The data were then entered to a Multi-Echo ICA analysis (ME-ICA), as described in [58], to further remove nuisance signals (e.g., hardware-induced artifacts, residual head motion). Briefly, this procedure utilizes the physical properties of BOLD and non-BOLD fluctuations, namely the fact that whole signal from BOLD sources increases linearly over echo times, signals from non-BOLD sources remain constant across echoes. This allows the removal of non-BOLD fluctuations (noise). The functional and anatomical datasets

were co-registered using AFNI, and then transformed to Talairach space. Voxels with a temporal signal to noise ratio under 40 were removed from further analysis.

Data analysis and statistical tests

All data were analyzed with in-house software written in MATLAB, as well as the AFNI software package. Data on the cortical surface were visualized with SUMA (Surface Mapping, [59]). Two-tail *t* tests were used for all *p* values on correlations, unless otherwise stated. For the maps in Figure 4 which were corrected through a permutation-based cluster size correction, we permuted the subject labels for the eye tracking typicality for 10,000 iterations, and correlated these with the neural typicality. The cluster threshold was defined for each threshold separately ($p<0.05$, $p<0.01$, $p<0.005$, $p<0.001$) as the largest cluster at that threshold at the 95th percentile across all iterations. Voxels were considered significant if they were significant in any of the corrected cluster sizes for the corresponding threshold.

Acknowledgments: We thank Adrian Gilmore, Andrew Persichetti and Stephen Gotts for many helpful conversations and insights, and Kelsey Csumitta for help with recruitment. This work was supported by the Intramural Research Program, National Institute of Mental Health (ZIAMH002920), clinical trials number NCT01031407.

References

1 Itti, L., Koch, C. & Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *Ieee T Pattern Anal* **20**, 1254-1259, doi:Doi 10.1109/34.730558 (1998).

2 Koch, C. & Ullman, S. Shifts in Selective Visual-Attention - Towards the Underlying Neural Circuitry. *Hum Neurobiol* **4**, 219-227 (1985).

3 Itti, L. & Koch, C. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research* **40**, 1489-1506, doi:10.1016/s0042-6989(99)00163-7 (2000).

4 Rider, A. T., Coutrot, A., Pellicano, E., Dakin, S. C. & Mareschal, I. Semantic content outweighs low-level saliency in determining children's and adults' fixation of movies. *J Exp Child Psychol* **166**, 293-309, doi:10.1016/j.jecp.2017.09.002 (2018).

5 Birmingham, E., Bischof, W. F. & Kingstone, A. Saliency does not account for fixations to eyes within social scenes. *Vision research* **49**, 2992-3000, doi:10.1016/j.visres.2009.09.014 (2009).

6 Coutrot, A., Guyader, N., Ionescu, G. & Caplier, A. Influence of soundtrack on eye movements during video exploration. *J Eye Movement Res* **5** (2012).

7 Xu, J., Jiang, M., Wang, S., Kankanhalli, M. S. & Zhao, Q. Predicting human gaze beyond pixels. *J Vision* **14**, doi:Artn 28
10.1167/14.1.28 (2014).

8 Langton, S. R. H., Watt, R. J. & Bruce, V. Do the eyes have it? Cues to the direction of social attention. *Trends in cognitive sciences* **4**, 50-59, doi:Doi 10.1016/S1364-6613(99)01436-9 (2000).

9 Nummenmaa, L. & Calder, A. J. Neural mechanisms of social attention. *Trends in cognitive sciences* **13**, 135-143, doi:10.1016/j.tics.2008.12.006 (2009).

10 Hasson, U. *et al.* Neurocinematics: The Neuroscience of Film. *Projections* **2**, 1-26, doi:10.3167/proj.2008.020102 (2008).

11 Hinde, S. J., Smith, T. J. & Gilchrist, I. D. Does narrative drive dynamic attention to a prolonged stimulus? *Cogn Res* **3**, doi:ARTN 45
10.1186/s41235-018-0140-5 (2018).

12 Smith, T. J. The Attentional Theory of Cinematic Continuity. *Projections* **6**, 1-27, doi:10.3167/proj.2012.060102 (2012).

13 Smith, T. J., Levin, D. & Cutting, J. E. A Window on Reality: Perceiving Edited Moving Images. *Curr Dir Psychol Sci* **21**, 107-113, doi:10.1177/0963721412437407 (2012).

14 Hasson, U., Malach, R. & Heeger, D. J. Reliability of cortical activity during natural stimulation. *Trends in cognitive sciences* **14**, 40-48, doi:10.1016/j.tics.2009.10.011 (2010).

15 Wang, H. X., Freeman, J., Merriam, E. P., Hasson, U. & Heeger, D. J. Temporal eye movement strategies during naturalistic viewing. *J Vision* **12**, doi:Artn 16
10.1167/12.1.16 (2012).

16 Loschky, L. C., Larson, A. M., Magliano, J. P. & Smith, T. J. What Would Jaws Do? The Tyranny of Film and the Relationship between Gaze and Higher-Level Narrative Film Comprehension. *PloS one* **10**, doi:ARTN e0142474
10.1371/journal.pone.0142474 (2015).

17 Hutson, J. P., Smith, T. J., Magliano, J. P. & Loschky, L. C. What is the role of the film viewer? The effects of narrative comprehension and viewing task on gaze control in film. *Cogn Res* **2**, doi:ARTN 46
10.1186/s41235-017-0080-5 (2017).

18 Joseph, R. M., Tager-Flusberg, H. & Lord, C. Cognitive profiles and social-communicative functioning in children with autism spectrum disorder. *J Child Psychol Psyc* **43**, 807-821, doi:Doi 10.1111/1469-7610.00092 (2002).

19 Moran, J. M. *et al.* Impaired theory of mind for moral judgment in high-functioning autism. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 2688-2692, doi:10.1073/pnas.1011734108 (2011).

20 Mathersul, D., McDonald, S. & Rushby, J. A. Understanding advanced theory of mind and empathy in high-functioning adults with autism spectrum disorder. *J Clin Exp Neuropsyc* **35**, 655-668, doi:10.1080/13803395.2013.809700 (2013).

21 Kleinman, J., Marciano, P. L. & Ault, R. L. Advanced theory of mind in high-functioning adults with autism. *Journal of autism and developmental disorders* **31**, 29-36, doi:Doi 10.1023/A:1005657512379 (2001).

22 Jasmin, K. *et al.* Overt social interaction and resting state in young adult males with autism: core and contextual neural features. *Brain : a journal of neurology* **142**, 808-822, doi:10.1093/brain/awz003 (2019).

23 Fletcher-Watson, S., Leekam, S. R., Benson, V., Frank, M. C. & Findlay, J. M. Eye-movements reveal attention to social information in autism spectrum disorder. *Neuropsychologia* **47**, 248-257, doi:10.1016/j.neuropsychologia.2008.07.016 (2009).

24 Harms, M. B., Martin, A. & Wallace, G. L. Facial Emotion Recognition in Autism Spectrum Disorders: A Review of Behavioral and Neuroimaging Studies. *Neuropsychol Rev* **20**, 290-322, doi:10.1007/s11065-010-9138-6 (2010).

25 Snow, J. *et al.* Impaired Visual Scanning and Memory for Faces in High-Functioning Autism Spectrum Disorders: It's Not Just the Eyes. *J Int Neuropsych Soc* **17**, 1021-1029, doi:10.1017/S1355617711000981 (2011).

26 Jang, C. *et al.* Individuality manifests in the dynamic reconfiguration of large-scale brain networks during movie viewing. *Sci Rep-Uk* **7**, doi:ARTN 41414 10.1038/srep41414 (2017).

27 Dorr, M., Vig, E. & Barth, E. Eye movement prediction and variability on natural video data sets. *Vis Cogn* **20**, 495-514, doi:10.1080/13506285.2012.667456 (2012).

28 Eckhardt, B. B., Wood, M. R. & Jacobvitz, R. S. Verbal-Ability and Prior Knowledge - Contributions to Adults Comprehension of Television. *Commun Res* **18**, 636-649, doi:Doi 10.1177/009365091018005004 (1991).

29 Hasson, U., Furman, O., Clark, D., Dudai, Y. & Davachi, L. Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. *Neuron* **57**, 452-462, doi:10.1016/j.neuron.2007.12.009 (2008).

30 Byrge, L., Dubois, J., Tyszka, J. M., Adolphs, R. & Kennedy, D. P. Idiosyncratic Brain Activation Patterns Are Associated with Poor Social Comprehension in Autism. *Journal of Neuroscience* **35**, 5837-5850, doi:10.1523/Jneurosci.5182-14.2015 (2015).

31 Hasson, U. *et al.* Shared and Idiosyncratic Cortical Activation Patterns in Autism Revealed Under Continuous Real-Life Viewing Conditions. *Autism Research* **2**, 220-231, doi:10.1002/aur.89 (2009).

32 Salmi, J. *et al.* The brains of high functioning autistic individuals do not synchronize with those of others. *Neuroimage-Clin* **3**, 489-497, doi:10.1016/j.nicl.2013.10.011 (2013).

33 Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J. & Kingstone, A. Gaze allocation in a dynamic situation: effects of social status and speaking. *Cognition* **117**, 319-331, doi:10.1016/j.cognition.2010.09.003 (2010).

34 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 Psychiat* **59**, 809-816, doi:DOI 10.1001/archpsyc.59.9.809 (2002).

35 Vo, M. L. H., Smith, T. J., Mital, P. K. & Henderson, J. M. Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. *J Vision* **12**, doi:Artn 3 10.1167/12.13.3 (2012).

36 Hasson, U., Nir, Y., Levy, I., Fuhrmann, G. & Malach, R. Intersubject synchronization of cortical activity during natural vision. *Science* **303**, 1634-1640, doi:DOI 10.1126/science.1089506 (2004).

37 Nummenmaa, L. *et al.* Emotions promote social interaction by synchronizing brain activity across individuals. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 9599-9604, doi:10.1073/pnas.1206095109 (2012).

38 Friesen, C. K. & Kingstone, A. The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychon B Rev* **5**, 490-495, doi:Doi 10.3758/Bf03208827 (1998).

39 Chita-Tegmark, M. Social attention in ASD: A review and meta-analysis of eye-tracking studies. *Res Dev Disabil* **48**, 79-93, doi:10.1016/j.ridd.2015.10.011 (2016).

40 Nguyen, M., Vanderwal, T. & Hasson, U. Shared understanding of narratives is correlated with shared neural responses. *NeuroImage* **184**, 161-170, doi:10.1016/j.neuroimage.2018.09.010 (2019).

41 Chen, J. *et al.* Shared memories reveal shared structure in neural activity across individuals. *Nature neuroscience* **20**, 115-125, doi:10.1038/nn.4450 (2017).

42 Perrett, D. I., Hietanen, J. K., Oram, M. W. & Benson, P. J. Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **335**, 23-30, doi:10.1098/rstb.1992.0003 (1992).

43 Heywood, C. A. & Cowey, A. The Role of the Face-Cell Area in the Discrimination and Recognition of Faces by Monkeys. *Philos T Roy Soc B* **335**, 31-38, doi:DOI 10.1098/rstb.1992.0004 (1992).

44 Akiyama, T. *et al.* Gaze but not arrows: a dissociative impairment after right superior temporal gyrus damage. *Neuropsychologia* **44**, 1804-1810, doi:10.1016/j.neuropsychologia.2006.03.007 (2006).

45 Greene, D. J. *et al.* Atypical neural networks for social orienting in autism spectrum disorders. *NeuroImage* **56**, 354-362, doi:10.1016/j.neuroimage.2011.02.031 (2011).

46 Wheatley, T., Milleville, S. C. & Martin, A. Understanding animate agents: distinct roles for the social network and mirror system. *Psychological science* **18**, 469-474, doi:10.1111/j.1467-9280.2007.01923.x (2007).

47 Dufour, N. *et al.* Similar Brain Activation during False Belief Tasks in a Large Sample of Adults with and without Autism. *PloS one* **8**, doi:ARTN e75468 10.1371/journal.pone.0075468 (2013).

48 Schnell, K., Bluschke, S., Konradt, B. & Walter, H. Functional relations of empathy and mentalizing: An fMRI study on the neural basis of cognitive empathy. *NeuroImage* **54**, 1743-1754, doi:10.1016/j.neuroimage.2010.08.024 (2011).

49 Frith, C. D. & Frith, U. The neural basis of mentalizing. *Neuron* **50**, 531-534, doi:10.1016/j.neuron.2006.05.001 (2006).

50 Saxe, R. & Kanwisher, N. People thinking about thinking people - The role of the temporo-parietal junction in "theory of mind". *NeuroImage* **19**, 1835-1842, doi:10.1016/S1053-8119(03)00230-1 (2003).

51 Moriguchi, Y. *et al.* Impaired self-awareness and theory of mind: an fMRI study of mentalizing in alexithymia. *NeuroImage* **32**, 1472-1482, doi:10.1016/j.neuroimage.2006.04.186 (2006).

52 Decety, J. & Lamm, C. The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. *Neuroscientist* **13**, 580-593, doi:10.1177/1073858407304654 (2007).

53 Schurz, M., Radua, J., Aichhorn, M., Richlan, F. & Perner, J. Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neuroscience and biobehavioral reviews* **42**, 9-34, doi:10.1016/j.neubiorev.2014.01.009 (2014).

54 Pelphrey, K. A., Morris, J. P. & McCarthy, G. Neural basis of eye gaze processing deficits in autism. *Brain : a journal of neurology* **128**, 1038-1048, doi:10.1093/brain/awh404 (2005).

55 Adolphs, R., Tranel, D., Damasio, H. & Damasio, A. Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature* **372**, 669-672, doi:10.1038/372669a0 (1994).

56 Adolphs, R., Baron-Cohen, S. & Tranel, D. Impaired recognition of social emotions following amygdala damage. *J Cogn Neurosci* **14**, 1264-1274, doi:10.1162/089892902760807258 (2002).

57 Cox, R. W. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and biomedical research, an international journal* **29**, 162-173 (1996).

58 Kundu, P. *et al.* Integrated strategy for improving functional connectivity mapping using multiecho fMRI. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 16187-16192, doi:10.1073/pnas.1301725110 (2013).

59 Saad, Z. S., Reynolds, R. C., Argall, B., Japee, S. & Cox, R. W. Suma: An interface for surface-based intra- and inter-subject analysis with AFNI. *2004 2ND IEEE INTERNATIONAL SYMPOSIUM ON BIOMEDICAL IMAGING: MACRO TO NANO, VOLS 1 and 2*, 1510-1513 (2004).