Behavioral/Cognitive

Heterogeneous Encoding Patterns of Dynamic Audiovisual Social Stimuli in Autism Spectrum Disorder (Heterogeneous Encoding Patterns in Autism)

Avery L. Van De Water,[1] Lisa Byrge,[2,3] Daniel P. Kennedy,[4,5,6] Dorit Kliemann,[1,7,8]* and James A. Traer[1,7]*

[1]Department of Psychological and Brain Sciences, The University of Iowa, Iowa City, Iowa 52242, [2]Department of Psychological & Brain Sciences, University of North Florida, Jacksonville, Florida 32224, [3]Biomedical Sciences Program, University of North Florida, Jacksonville, Florida 32224, [4]Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana 47405-7007, [5]Program in Neuroscience, Indiana University, Bloomington, Indiana 47405-7007, [6]Cognitive Science Program, Indiana University, Bloomington, Indiana 47405-7007, [7]Iowa Neuroscience Institute, The University of Iowa, Iowa City, Iowa 52242, and [8]Department of Psychiatry, The University of Iowa, Iowa City, Iowa 52242

The neural basis of social difficulties in autism spectrum disorder (ASD) remains uncertain. Prior studies typically used static stimuli or reduced complex stimuli to only a few dimensions, obscuring real-world complexity. Here, we used videos with social content to engage brain networks and trained encoding models to predict brain activity from high-dimensional audio, visual, and more abstract socially relevant features (e.g., valence). We hypothesized that brain activity in autistic adults would be less accurately predicted and more heterogeneous within and across individuals than in neurotypical (NT) participants, particularly for socially relevant features. Participants [48 ASD (13 female), 90 NT (25 female)] viewed six videos (5–20 min). Brain activity was measured using fMRI. Cross-validated encoding models were assessed on prediction accuracy, variance, and pattern similarity. Brain activity was predicted significantly above chance across brain networks and videos with striking consistency between groups. Auditory and language networks were most accurately predicted. Within-participant predictions were equally variable between groups; however, autistic participants exhibited lower within-group encoding pattern similarity than NT participants and greater across-group than within-group similarity. Analyses using only socially relevant features showed the largest across-group differences in encoding pattern similarity. The novel application of encoding models to multiple videos in ASD suggests greater heterogeneity in neural representations of social information in ASD. Although group-averaged encoding accuracies were comparable across groups, ASD participants showed more heterogeneous regional encoding patterns. These idiosyncratic patterns underscore the need to incorporate individual variation in brain representations when developing translational neuroimaging approaches in autism.

Key words: **autism spectrum disorder; encoding models; heterogeneity; social cognition; video fMRI**

Received Jan. 9, 2026; revised Aug. 14, 2026; accepted Aug. 20, 2026.

Author contributions: D.K. and J.A.T. designed research; A.L.V.D.W., L.B., D.P.K., D.K., and J.A.T. performed research; A.L.V.D.W., L.B., D.P.K., D.K., and J.A.T. contributed unpublished reagents/analytic tools; A.L.V.D.W., L.B., D.K., and J.A.T. analyzed data; A.L.V.D.W., L.B., D.P.K., D.K., and J.A.T. wrote the paper.

As this study was part of a larger project, we thank J. Michael Tyszka (California Institute of Technology) and Hu Cheng (Indiana University) for expertise in MRI data acquisition, scanner troubleshooting, and site alignment; Lynn K. Paul for support in recruitment and behavioral assessments; Tim Armstrong, Steven Lograsso, and Arispa Weigold for help in data acquisition; Paola Galdi for support in denoising code optimization; Jonathan Y. Peters, Jordan Luna, Pasi J. Jouhikainen, Jax D. Skye, and other members of the Social Cognitive Neuroscience Lab for support in manual feature annotation; Umit Keles for feedback on analyses and results; as well as Ralph Adolphs for his major contribution to the larger project and feedback. This work was supported in part through computational resources provided by the University of Iowa, Iowa City, Iowa; the Iowa Initiative for Artificial Intelligence (D.K., J.A.T.); and by Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute. We acknowledge the Indiana University Pervasive Technology Institute (https://pti.iu.edu; Stewart et al., 2017) for providing supercomputing and storage resources that have contributed to the research results reported within this paper. Research reported in this publication was supported by the National Institute of Mental Health of the National Institutes of Health under award numbers R01MH110630 and R01MH136198. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This manuscript is the result of funding in whole or in part by the National Institutes of Health (NIH). It is subject to the NIH Public Access Policy. Through acceptance of this federal funding, NIH has been given a right to make this manuscript publicly available in PubMed Central upon the Official Date of Publication, as defined by NIH.

*D.K. and J.A.T. contributed equally to this work.

The authors declare no competing financial interests.

Correspondence should be addressed to Dorit Kliemann at dorit-kliemann@uiowa.edu or James A. Traer at james-traer@uiowa.edu.

This paper contains supplemental material available at: https://doi.org/10.1523/JNEUROSCI.0055-26.2026

https://doi.org/10.1523/JNEUROSCI.0055-26.2026

Copyright © 2026 the authors

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Significance Statement

Individuals with autism spectrum disorder (ASD) experience difficulties in social communication and interaction, yet the underlying neural mechanisms remain unclear. Here, we present a novel application of encoding models to neuroimaging in ASD. Neural representations of dynamic and complex social information were more heterogeneous across autistic as compared with neurotypical individuals while encoding model prediction accuracy was similar for population averages across groups. Effects were robust across six videos and spatial scales. Variability in neural representation across individuals, rather than systemic population differences, may thus be associated with atypical brain function in complex social information processing in autism. The findings highlight the importance of accounting for individual variability when studying the neurobiology of autism and developing translational neuroimaging approaches.

Introduction

Social difficulties are core to autism spectrum disorder (ASD). To date, many findings on the underlying neural basis do not consistently replicate across studies, tasks, and samples (Kliemann et al., 2018; King et al., 2019). Prior research has favored highly controlled stimuli (e.g., static images, isolated from context), which may have limited stimulus complexity and thus ecological validity. This may obscure essential aspects of social difficulties in autism (Dufour et al., 2013; Kana et al., 2015). More "realistic" designs, such as social interaction games, may be especially suited to study social difficulty in ASD (Redcay et al., 2013; Banker et al., 2025). For example, reduced social attention was magnified by complex multisensory load in naturalistic environments (Haskins et al., 2022).

Passive video viewing has been increasingly used to approximate real-world cognition (Hasson et al., 2004; Finn and Bandettini, 2021) and has shown promise in populations where dynamic information processing may be altered, such as schizophrenia (Yang et al., 2020) and depression (Gruskin et al., 2020). It seems particularly suited for ASD, because it requires integration of richly structured information unfolding over time, a mechanism suggested to be altered by several theories, for example, atypical sensory integration (Robertson and Baron-Cohen, 2017), predictive coding (Sinha et al., 2014), and central coherence (Happe and Frith, 2006). Particularly relevant for ASD, videos can improve compliance, reduce motion, and increase engagement (Vanderwal et al., 2015; Eickhoff et al., 2020).

Encoding models provide a powerful framework to assess relationships between brain activity and complex high-dimensional stimulus features (Huth et al., 2012; Dupre la Tour et al., 2022; Im et al., 2025). Prior work has favored voxel-level approaches in few participants ($n < 10$) with deep data and repeated stimulus presentations. Regarding sample size, shorter videos in larger samples are more feasible to study atypical samples (Nakai et al., 2024). Regarding spatial resolution, parcel-level network approaches have revealed interindividual differences, brain–behavior predictions (Finn et al., 2015; Dubois et al., 2018), and atypical sensory integration and social cognition systems in ASD (Uddin et al., 2014; Washington et al., 2014). Candidate networks integrate cross-modal perceptual information (e.g., posterior multimodal network, PMM; Ji et al., 2019) with higher-level cognitive and social representations (Andrews-Hanna et al., 2014; Braga and Buckner, 2017) and networks previously implicated in ASD pathophysiology (e.g., default mode, DMN; Kennedy et al., 2006; Padmanabhan et al., 2017).

Heterogeneity of autistic symptoms (Geschwind and Levitt, 2007; Dinstein et al., 2011) may indicate heterogeneous neural patterns on the intra- (Dinstein et al., 2011; Shama et al., 2025) and interindividual level (Hahamy et al., 2015; Hong et al., 2023), potentially contributing to inconsistent group-level findings (Kliemann et al., 2018; Rodgaard et al., 2019; He et al., 2020). Greater interindividual neural variability during video viewing in ASD may reflect more idiosyncratic representations of dynamic information (Hasson et al., 2009; Byrge et al., 2022; Turner et al., 2025). It is therefore critical to explicitly investigate individual-level variability with encoding models, even if the derived metrics represent only indirect proxies for underlying brain activity.

This study sought to compare how encoding models predict cortical activity using visual, auditory, and more abstract socially relevant features (e.g., valence) across autistic and neurotypical participants. We assessed the following: (1) prediction accuracy (how well features predicted brain activity); (2) intraindividual variance (how variable predictions were across test-train splits); (3) average neurotypical encoding prediction pattern similarity; and (4) within- versus across-group pairwise comparisons. We hypothesized that autistic participants would show less accurate and more heterogeneous (within- and across-groups and individuals) predictions, especially when focusing on features and networks relevant for social cognition. We tested the hypotheses across spatial granularities, including whole brain, data-driven parcel subsets (top 10%), and hypothesis-driven networks (DMN, PMM). The present work advances the neuroscientific study of ASD by assessing how complex features predict neural representations and how individual encoding patterns indicate neural heterogeneity in autistic social cognition.

Materials and Methods

Participants

Data were previously collected from autistic and neurotypical adults across two sites (Indiana University, IU; California Institute of Technology, Caltech; Byrge et al., 2022). Following data quality exclusions (e.g., technical issues, data quality assessments, head motion, $n = 10$), the full sample (prior to matching per video, see sample matching section and Table S1) included 90 NT participants (25 females, mean age of 26.8 years) and 48 autistic participants (14 females, mean age of 25.1 years; Table 1). The neurotypical dataset included predominantly males (NT = 72.2%; ASD = 70.8%) to reflect the male-to-female ratio observed in ASD populations at the time of study design (~75% male; Loomes et al. (2017).

Participants were included if they were not colorblind, were fluent in English, were able to participate in an MRI, had normal or corrected vision of at least 20/40, and had normal hearing. Neurotypical participants were further included if they had no history of psychiatric disorder or neurological incident or disease and not taking medication for mental illness. Participants in the autistic group had a prior clinical diagnosis of

Table 1. Behavioral and demographic characteristics

Variable	Group *n* (ASD/NT)	ASD mean (SD) [range]	NT mean (SD) [range]	$p[t(df)]$ or $p[U]$ or $p[\chi^2(df)]$	Effect size
Sex	48/90	34 M, 14 F	65 M, 25 F	$1.00[\chi^2(1) = 0.00]$	$V = 0.00$
Age	48/90	25.1 (5.54) [18.0–44.0]	26.79 (7.82) [18.0–54.0]	$0.384[U = 1{,}965.5]$	$r_{rb} = -0.09$
FSIQ	47/84	113.81 (13.38) [84.0–142.0]	110.0 (10.16) [87.0–136.0]	$0.094[t_{(129)} = 1.70]$	$d = 0.33$
Handedness	48/90	39 R, 8 L, 1 A	79 R, 9 L, 2 A	$0.525[\chi^2(2) = 1.29]$	$V = 0.10$
Location	48/90	16 Cal, 32 IU	29 Cal, 61 IU	$1.00[\chi^2(1) = 0.00]$	$V = 0.00$
AQ	48/89	27.12 (8.67)[10.0–48.0]	16.09 (6.57) [4.0–40.0]	$7.25 \times 10^{-12}[U = 3{,}654.0]$	$r_{rb} = -0.71$
MASC	39/85	29.38 (8.38) [7.0–42.0]	35.71 (3.67) [27.0–46.0]	$4.55 \times 10^{-5}[t_{(122)} = -4.52]$	$d = -1.13$
EQ	46/90	31.35 (15.83) [9.0–90.0]	46.53 (13.51) [6.0–75.0]	$1.79 \times 10^{-8}[U = 845.5]$	$r_{rb} = -0.59$
VIQ	47/84	115.36 (17.16) [74.0–160.0]	111.87 (11.96) [90.0–144.0]	$2.20 \times 10^{-1}[t_{(129)} = 1.24]$	$d = 0.25$
Familiarity	44/81	4.02 (3.1) [1.0–9.0]	4.27 (2.97) [1.0–9.0]	$7.23 \times 10^{-1}[U = 1{,}714.0]$	$r_{rb} = -0.04$

Behavioral and demographic characteristics for participants with autism spectrum disorder (ASD, $n = 48$) and neurotypical participants (NT, $n = 90$); see Figure S1. FSIQ, full scale intelligence quotient; VIQ, verbal intelligence quotient; AQ, Autism Spectrum Quotient score; MASC, Movie for the Assessment of Social Cognition; EQ, Empathy Quotient; familiarity, how familiar participants were with each video; n, sample size; SD, standard deviation; χ^2, chi-square test statistic; *U*, Mann–Whitney *U* test statistic; *t*, Student's *t* test statistic; *p*, probability value; df, degrees of freedom; V, Cramer's V; r_{rb}, rank-biserial correlation; *d*, Cohen's d; F, female; M, male; IU, Indiana University; Cal, California Institute of Technology; R, right-handed; L, left-handed; A, ambidextrous.

ASD, Asperger's syndrome, or Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS) and medication was not an exclusion criterion. To validate ASD diagnoses at both research sites in autistic participants, the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2; Lord et al. (2012), Module 4) was administered by research reliable ADOS-administrators. Inclusion required participants to score within one point of the diagnostic threshold according to the revised algorithm by Hus and Lord (2014), specifically a combined Social Affect and Restricted and Repetitive Behaviors score of ≥7. For autistic participants enrolled at Caltech, diagnoses were also confirmed by expert clinical evaluation.

Exclusion criteria included a full-scale IQ (FSIQ) below 80, a history of moderate or severe head trauma, epilepsy, schizophrenia, bipolar disorder, and current use of hallucinogenic or antipsychotic substances. Intellectual functioning was evaluated using the Wechsler Abbreviated Scale of Intelligence (Wechsler, 2011), used to estimate full scale IQ (FSIQ) and verbal IQ (VIQ) reported here. The Empathy Quotient (EQ; Baron-Cohen and Wheelwright, 2004) was collected to assess empathy with a 60-item self-report. Social cognitive ability was further assessed with the Movie for the Assessment of Social Cognition (MASC; Dziobek et al., 2006). The MASC characterizes the ability to reason about others' mental states and has shown sensitivity to difficulty in autistic populations (Lahera et al., 2014; Muller et al., 2016; Martinez et al., 2017; Schaller and Rauh, 2017), without ceiling effects in neurotypical populations (Dziobek et al., 2006).

This study was approved by the institutional review boards of Indiana University and the California Institute of Technology, and all participants gave written, informed consent. Data were subject-wise excluded for technical (e.g., projector malfunction) as well as data quality reasons (e.g., registration quality, artifacts, head motion, sleep) that are described in the MRI processing section below and in further detail in Byrge et al. (2022; Table 2).

fMRI data

Acquisition

MRI images were acquired using Siemens 3 T Magnetom Prisma Fit scanners (Siemens Medical Solutions) equipped with 64-channel head receiver arrays. Identical acquisition sequences were used across sites. Scanner software versions included VE11B (IU) and VE11C (Caltech and final five IU scans). Functional whole-brain images were acquired using T2*-weighted multiband echo planar imaging (EPI) sequences (voxel size, 2.5 mm^3; TR, 720 ms; TE, 30 ms; flip angle, 50°; 60 interleaved slices; multiband acceleration factor, 6; multiband EPI sequence version R16, CMRR, University of Minnesota). Spin-echo EPI images (TR, 4.39 s; TR, 37.2 ms; flip angle, 90°) were acquired in opposite phase-encoding directions as field maps for EPI distortion correction. For anatomical reference, high-resolution images were acquired using a multi-echo MPRAGE sequence (voxel size, 0.9 mm^3; TR, 2.55 s; TEs, 1.63 ms, 3.45 ms, 5.27 ms, 7.09 ms; TI, 1,150 ms). As described in Byrge et al. (2022), a trigger box upgrade during the final phase of data collection at the Indiana University (IU) site caused intermittent technical issues in 35 scans of the overall sample, resulting in occasional missed triggers and delayed video starts. These scans were empirically identified and corrected through realignment procedures (Byrge et al. (2022).

Table 2. Number and percentage of variance explained for principal components for visual motion energy, spatial gist, and spectrotemporal audio features after dimensionality reduction included per video

Video	Motion energy	Spatial gist	Spectrotemporal audio
Bang	95 (85%)	9 (95%)	14 (70%)
Partly Cloudy	106 (95%)	20 (95%)	30 (70%)
Office1	89 (75%)	23 (95%)	25 (95%)
Office2	127 (75%)	25 (95%)	27 (70%)
Trailers1	56 (96%)	19 (95%)	17 (70%)
Trailers2	135 (97%)	19 (95%)	13 (70%)

%, percentage.

Stimuli

While in the scanner, participants passively viewed six short videos (each ~5–21 min; Kliemann et al., 2021; Byrge et al., 2022; Turner et al., 2025). The videos varied in length and genre: two 21 min episodes from the TV sitcom *The Office* ("Office1," Season 1 Episode 6, 1,753 timepoints; "Office2," Season 1 Episode 5, 1,749 timepoints; Byrge et al., 2015; Pantelis et al., 2015; Byrge et al., 2022). Additionally, participants watched two 13 min compilations of video trailers spanning genres such as documentaries, sports, and mystery ("Trailers1," 1,121 timepoints; "Trailers2," 1,073 timepoints; Byrge and Kennedy, 2020; Byrge et al., 2022). The selection also included *Bang! You're Dead*, an 8 min black-and-white excerpt from an Alfred Hitchcock thriller (1961, 674 timepoints; "Bang"; Hasson et al., 2008a,b; Hasson et al., 2010; Campbell et al., 2015), and Pixar's *Partly Cloudy*, a 5 min animated short video (442 timepoints; Reher and Sohn, 2009; Richardson et al., 2018; Byrge et al., 2022; Turner et al., 2025). Audio was delivered using Sensimetrics MR-compatible in-ear headphones (Sensimetrics model S14, www.sens.com), which provided high-fidelity audio and some attenuation of scanner noise. Volume was adjusted to a comfortable level for each participant prior to acquisition while running an EPI to account for scanner noise in volume selection.

Because familiarity with the video content could in principle affect neural processing, we asked participants to rate how familiar they were with each video on a scale of 1 (not at all) to 9 (very much) after the scan sessions (Fig. 2, Table 1, Table S3).

Video features

To evaluate visual, auditory, and more abstract (socially relevant) semantic processing, we extracted three types of respective features from the videos. To do so, we first segmented audiovisual videos into 0.72 s intervals, necessary to align with the TR for synchronization between brain data and video features (although, note, some of the features were

computed over multiple intervals to capture different timescales of the stimulus). We standardized all features using *z*-scoring to facilitate comparability across different videos. We used both automated and manual feature extraction methods, enabling a comprehensive analysis of brain responses to audiovisual and more abstract information.

Automated feature extraction. For visual features, we calculated three types of features: (1) We estimated the average changes in hue, saturation, and value between frames using the rgb2hsv function from Scikit-learn (Pedregosa et al., 2011). To do so, we averaged values across the first two frames of each TR and the two preceding frames and then computed frame differences by subtracting the average of preceding frames from the average of the current frames. (2) We estimated motion energy (pymoten; Nunez-Elizalde et al., 2022), analyzing five frames per TR (the first frame of the TR, two preceding frames, and two subsequent frames) and selected the second frame difference (between the first frame of the TR and its preceding frame) for further analysis. (3) We derived low-dimensional scene representations using the gist algorithm (Oliva and Torralba, 2001), retaining principal components that explained 95% of the variance.

For auditory features, we segmented audio waveforms into TR-length bins (Hanning windowed to minimize truncation artifacts) and computed statistics on each segment. For "basic auditory" features we computed statistics of the time series (root-mean-square amplitude, variance, skew, kurtosis) and of the log-power spectrum (mean, variance, skew, kurtosis) for each segment. These features give a very simple approximation of sound power, power fluctuations, and frequency content (e.g., high-vs-low frequency and "broadband noise"-vs-"narrowband tones"; Moore, 2013). Additionally, we extracted spectrotemporal auditory statistics (McDermott and Simoncelli, 2011). Each segment was transformed into a log-power spectrogram with 128 mel-frequency-spaced subbands spanning 0.05–21.05 kHz. Statistics (mean, variance, skew, kurtosis) were computed of each subband power modulation and of each subband modulation power spectra (computed from the subband power modulations with 256-bands). Subband power modulations were cross-correlated across frequency bands, and the 128×128 pairwise correlation values were stored. Collectively, these statistics of subband power modulations and correlations give an indication of sound texture (e.g., "crackling" vs "bubbling" vs "scratching" vs "whooshing" sounds; McDermott and Simoncelli, 2011) for each TR segment. Finally, 40 mel-frequency cepstral coefficients (MFCCs) were extracted for each segment (Imai, 1983). These features have long been used to separate two aspects of sound thought to be of particular use to speech: the rapid sound fluctuations caused by vocal fold vibrations that convey the pitch of a speaker's voice and the slower power modulations induced during speech by movements of the vocal tract (e.g., mouth, tongue). We computed these MFCCs on every segment whether speech was present or not.

Pymoten, gist, and spectrotemporal auditory feature sets contained hundreds or thousands of values for every timepoint. To reduce dimensionality, we transformed each set with a principal component analysis (PCA) and retained components to account for at least 70% of the original feature set variance (Table 2). This reduced the number of features (many of which strongly covaried across the videos and thus provided redundant information) to 9–135 components (across videos). Because the videos themselves had highly variable content, the number of components required to preserve 70% of variance varied dramatically across both feature spaces and videos. The current study was not directly comparing feature spaces (or their respective contributions to signal variance with a banded ridge regression; Dupre la Tour et al., 2022) within or across videos. We thus opted to roughly equate the numbers of components across video and feature spaces (which affects the encoding model fitting and the resulting prediction accuracies), rather than the fraction of variance they contributed. Importantly, these values were constant for all participants (compared in our main analyses) and were chosen prior to hypothesis testing.

Manual feature extraction. Two methods were used to extract additional features manually from each video. First, using an independent sample of 138 participants (mean age, 19.0 years; range, 18–38; 96 female), we obtained continuous ratings of arousal and positive and negative valence using a joystick to indicate changes in real time when watching the video. Data were smoothed (Savitzky–Golay filter, polynomial order 3, 10 s window length) and *z*-scored (following Kirk et al., 2023). Intersubject correlations (ISCs) were computed using a pairwise and a leave-one-out approach to assess variability and consistency across individual ratings (Nastase et al., 2019; arousal: $r = 0.31 \pm 0.14$; positive valence: $r = 0.24 \pm 0.14$; negative valence: $r = 0.24 \pm 0.08$; Table S1).

In the second method for manual feature extraction, features from each video were annotated by six trained annotators using ELAN (v6.4, Max Planck Institute for Psycholinguistics) using one-hot manual annotation to represent the presence or absence of specific features. Two independent raters annotated each feature for each video. We assessed inter-rater reliability (82 and 97% averaged across videos per feature; Table S1) and used averaged consensus ratings for subsequent analyses. Manual annotation included scene transitions (such as zooms and scene changes) and more abstract features, including distance between agents present on the screen (categorized into five levels: direct, near, not so close, distant, and over-the-phone), presence of vocalization (including speech but also nonspeech such as a sigh, gasp, cry, etc.), and background music. Many of the abstract features were selected with the expectation that they were important for social cognition. We annotated five features specifically relevant for social cognition in addition to previously mentioned features (i.e., arousal, positive valence, and negative valence), including theory of mind (reasoning about others' thoughts, intentions, or emotions; Wimmer and Perner, 1983), goal directedness (actions aimed at achieving specific objectives, e.g., carrying a table, dancing; Abell et al., 2000), joint attention (i.e., at least two people focusing their attention on something together, e.g., looking at a computer screen, pointing at a bird in the sky and others look up; Tomasello et al., 2005), physical states (instances reflecting states such as hunger, thirst, or pain; Bruneau et al., 2012), and social touch (i.e., people touching; McGlone et al., 2014). Resulting one-hot encoded features had an average of 17.0% presence across videos (SD = 5%, min = 11.8%, max = 25%), with the highest average feature presence in Partly Cloudy (Table 3). Of one-hot encoded features present in all videos, "Any Vocalizations" had the highest percent presence across videos. Additional features, such as speaker identity, theme songs, and environmental sounds, were annotated by one annotator (Table 3).

MRI processing

Preprocessing

Raw DICOM images were converted to NIfTI format and organized per the Brain Imagining Data Structure (BIDS; Gorgolewski et al., 2016). Preprocessing was conducted using fMRIPrep (v20.0.7; Esteban et al., 2019), incorporating components from ANTs (Avants et al., 2008), FSL (v5.0.9; FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl), and FreeSurfer (v6.0.1; Dale et al., 1999). In short, key steps included bias correction, skull stripping, segmentation, spatial normalization using nonlinear registration to MNI space (2 mm MNI152NLin6Asym), and motion correction. Functional data were corrected for rigid-body motion, distortion using field maps, and coregistered to anatomical reference image. Confound regressors, including head motion parameters, CSF, and WM, were also computed. Slice-time correction was not applied due to the short TR (0.72 s). Detailed processing is outlined below (fMRIPrep boilerplate).

Anatomical data processing. The T1-weighted (T1w) image was corrected for intensity nonuniformity (INU) with N4BiasFieldCorrection (Tustison et al., 2010), distributed with ANTs 2.2.0 (Avants et al., 2008), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white matter (WM), and gray matter (GM) was performed on the brain-extracted T1w using fast (Zhang et al., 2001). Brain surfaces were reconstructed using recon-all (FreeSurfer 6.0.1; Dale et al., 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray matter of Mindboggle (Klein et al., 2017). Volume-based spatial normalization to two standard spaces (MNI152NLin6Asym,

Table 3. Percent presence of manually encoded features per video and feature type

Feature type	Bang	Partly Cloudy	Office1	Office2	Trailers1	Trailers2
Scene cuts	1.7	3.2	5.2	15.2	27	30.1
Theory of mind	16.7	35.7	21.5	31.4	12.6	11.2
Goal directedness	4.1	15.4	10.5	12	10.8	7.6
Joint attention	14.2	8.4	23.4	7.5	5.7	10.8
Physical states		8.4	2.5	7.5	1.9	5.3
Touching	12	8.4	4.1	4.7	17.9	10.6
Background music	31.3	24.4			93	87.6
Conversation	38.3	42.1	45.2	72.3	1.6	11.6
Vocalization	51.5	64.7	79.4	87.4	53.4	57.5
Environmental sounds	14.5	64.7				
Theme music Transitions	18.6	4.8				
Gun cymbal Leitmotif	25					
Happy Leitmotif		22.4				
Sad leitmotif		10.2				
Agent proximity						
Direct	4.7	5.7	0.5	10	2	1.3
Near	25.6	13.8	12.1	40.3	1.5	5.7
Not so Close	3.8	35.3	15.3	21.4		1.7
Distant	6.3	3.8	39.3	8.4	0.4	4.3
Over the phone	6.6		1	4.2		
Individual speakers	1.8–8.6		0.2–48.7	0.4–37.9		
	($n_s = 6$)		($n_s = 12$)	($n_s = 10$)		
Total average	11.8	25	15.3	11.8	18.9	19

Empty cells indicate the absence of specific feature type in respective video. n_s, number of individual speakers.

MNI152Nlin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.2.0), using brain-extracted versions of both T1w reference and the T1w template. The following templates were selected for spatial normalization: FSL's MNI ICBM 152 nonlinear 6th Generation Asymmetric Average Brain Stereotaxic Registration Model (TemplateFlow ID: MNI152NLin6Asym; Evans et al., 2012) and ICBM 152 Nonlinear Asymmetrical template version 2009c (TemplateFlow ID: MNI152NLin2009cAsym; Fonov et al., 2009).

Functional data preprocessing. For each of the available BOLD runs per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. A B0-nonuniformity map (or fieldmap) was estimated based on two (or more) echoplanar imaging (EPI) references with opposing phase-encoding directions, with 3dQwarp (AFNI 20160207; Cox, 1996). Based on the estimated susceptibility distortion, a corrected EPI (echoplanar imaging) reference was calculated for a more accurate coregistration with the anatomical reference. The BOLD reference was then coregistered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration (Greve and Fischl, 2009). Coregistration was configured with six degrees of freedom. Head motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) were estimated before any spatiotemporal filtering using mcflirt (FSL 5.0.9; Jenkinson et al., 2002). The BOLD time series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single, composite transform to correct for head motion and susceptibility distortions. These resampled BOLD time series will be referred to as preprocessed BOLD in original space or just preprocessed BOLD. The BOLD time series were resampled into standard space, generating a preprocessed BOLD run in MNI152NLin6Asym space. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Several confounding time series were calculated based on the preprocessed BOLD: framewise displacement (FD), DVARS, and three region-wise global signals. FD and DVARS are calculated for each functional run, both using their implementations in Nipype (following the definitions by Power et al., 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (CompCor; Behzadi et al., 2007) which was not used in this study and thus removed from this boilerplate description. The head motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al., 2013). All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e., head motion transform matrices, susceptibility distortion correction when available, and coregistrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Nongridded (surface) resamplings were performed using mri_vol2surf (FreeSurfer).

Calculating in-scanner head motion. The head motion parameters were filtered to exclude respiratory frequencies (Power et al., 2019). To quantify in-scanner head motion across scans, we utilized the head motion parameters derived from fMRIPrep to compute filtered framewise displacement traces (FDfilt4) as the sums of the backwards difference of four TRs of motion parameters, consistent with prior established and published procedures (Byrge et al., 2022). Compared with FDs computed by fMRIPrep (Esteban et al., 2019; Power et al., 2019), FDfilt4 is expected to be more effective at separating motion parameters from respiratory fluctuations in multiband acquisitions.

Denoising and spatial smoothing. To minimize non-BOLD-related noise in fMRI data, we employed the rsDenoise tool (https://github.com/adolphslab/rsDenoise; Dubois et al., 2018; Kliemann et al., 2022) using a denoising pipeline closely following the strategy outlined by Finn et al. (2015), including seven consecutive steps: (1) each voxels' signal was *z*-score normalized; (2) using tissue masks, temporal drifts from cerebrospinal fluid (CSF) and white matter (WM) were removed with third-degree Legendre polynomial regressors; (3) CSF and WM mean signals were regressed from gray matter (GM) voxels; (4) rotational and translational realignment parameters and their temporal derivatives were used as explanatory variables in motion regression; (5) signals are low-pass filtered with a Gaussian kernel; and (6) temporal drift from GM signal was removed using a third-degree Legendre polynomial regressor. Cortical data were spatially smoothed with a 6 mm FWHM kernel.

Cortical parcellations

We analyzed the average BOLD time series by dividing the voxel-wise cortical data into 360 parcels, based on changes across various neuroimaging modalities such as cortical architecture, function, connectivity, and topography (Glasser et al., 2016). Following the largely cortico-centric focus of the encoding model fMRI literature to date, we excluded subcortical and cerebellar structures from the current analyses. The 360 cortical parcels were categorized into 12 distinct networks (primary and secondary visual, auditory, language, somatomotor, default mode, posterior and ventral multimodal, cingulo-opercular, dorsal attention, frontoparietal, and orbito-affective), reflecting the functional connectivity and lateralization of the language and motor networks (Ji et al., 2019). Consistent with prior work in ASD (Kliemann et al., 2024; Li et al., 2024; 2025; Czekoova et al., 2025), we focused on a parcel-level approach (instead of voxel-level) to investigate how brain function may differ at the systems level (see d'Ascoli et al., 2025; Schad et al., 2025 for similar parcel-level encoding models). For each parcel, we thus averaged the preprocessed and denoised BOLD time series across voxels. Atypical network organization has been reported in individuals with ASD (Persichetti et al., 2025); therefore, the validity of the parcellation was confirmed for all participants with a homogeneity analysis (Table S2, Fig. S3).

Data exclusion due to MRI quality

To evaluate data quality, MRI images were processed using MRIQC (v0.15.2; Esteban et al., 2017) for initial quality assessment using functional image quality metrics (IQMs): ghost-to-signal ratio (GSR; Giannelli et al., 2010), signal-to-noise ratio (SNR), temporal SNR (TNSR; Kruger and Glover, 2001), DVARS standard deviation (Power et al., 2012), and average full-width at half-maximum smoothness (FWHM avg; Friedman et al., 2008). Outliers for each IQM metric [across the broader project sample also including subjects not included here, see Byrge et al. (2022)], defined as values exceeding the median for that site (IU or Caltech) plus 1.5 times the site-specific interquartile range (IQR), were flagged and manually reviewed by D.K. and L.B. All flagged scans were excluded from further analyses based on a consensus decision. Additionally, all reports generated by fMRIPrep were reviewed by two independent raters from IU or Caltech, who flagged potential issues with anatomical or functional scans and their alignment. These flagged scans were then independently reviewed by D.K. and L.B., who reached a consensus on whether to include or exclude them from the dataset.

Data exclusion due to head motion

Excessive motion was defined as mean FDfilt4 greater than median FDfilt4 plus 1.5 times the IQR of the mean FDfilt4 across all scans. Scans with excessive motion were excluded per site (mean FDfilt4 > 0.4808 for IU, mean FDfilt4 > 0.5625 for Caltech). For the remaining scans, we censored frames with excessive motion (FDfil4 > 3.75 mm), including 10 preceding and 30 subsequent frames. After all described exclusions (including data quality), the number of valid participants varied across videos (e.g., not all participants had data for each video).

Sample matching

The full sample with at least one run of valid fMRI data consisted of 138 participants (NT $n = 90$, ASD $n = 48$; Table 1). Because not all participants had valid data for all videos, we selected subsets of participants for each video in which the subset NT participants were matched to the subset ASD participants on potentially confounding variables (age, sex, FSIQ, and mFDfilt4). To achieve per video matching while maintaining as many autistic participants as possible, we applied an iterative exclusion procedure sequentially removing NT participants (prioritizing those with the largest deviation from the ASD group mean for a variable) until all variables were matched [$p > 0.05$, using an independent samples *t* test (Student, 1908), Mann–Whitney *U* test (Mann and Whitney, 1947), or chi-square tests (Pearson, 1900; McHugh, 2013)], depending on standard test assumptions. Participants with missing data on any of the relevant variables were excluded from matching (FSIQ: ASD $n = 1$, NT $n = 6$; Table S3, Fig. 2, Fig. S1). These "matched samples" did not seek to match the number of ASD and NT participants.

For supplemental analyses, we also selected participant subsets per video that matched the number of ASD and NT participants as well as all potential confounding variables (*n*-matched samples: Bang, $n = 32$; Partly Cloudy, $n = 37$; Office1, $n = 40$; Office2, $n = 35$; Trailers1, $n = 41$; Trailers2, $n = 40$; see Table S5 for details). Patterns of between-group effects in encoding model analyses largely replicated in the *n*-matched sample (Figs. S6–S9, Tables S5–S7).

Experimental design and statistical analyses

Encoding models

For each video, we used linear ridge regressions to predict time-varying BOLD responses within each cortical parcel from time-varying features (Fig. 1; https://github.com/gallantlab/himalaya; custom code implementation https://research-git.uiowa.edu/scnlabp/ens_a). Regressions were fit separately for each participant to account for individual variability in brain responses (LeBel et al., 2023). We performed two sets of analyses: (1) using all visual, auditory, and abstract features (ALL model) and (2) using only abstract features relevant to social processing, which we hypothesized would be more relevant to autistic difficulties (SOC model). Of note, the aim of this study was not to statistically compare the two models (or feature sets therein) directly, but rather to assess the difference between ASD and NT groups viewed through the lens of each encoding model.

Both the fMRI data and the corresponding features were divided into training and testing datasets, with ~20% of the data allocated to the test set. Test time indices were selected in continuous blocks to respect temporal autocorrelation. Specifically, contiguous time blocks with random starts and durations were sampled from uniform distributions. This process was repeated until 20% of the total time indices were labeled as test data. BOLD and feature time series have inherent temporal autocorrelation over a timescale of several seconds. To maximize independence of training and testing sets, we applied a conservative temporal buffer to account for the hemodynamic lag so that any timepoint within 14 s of a test index was excluded from the training set (see Fig. S2 for effects of temporal buffer). If <60% of data remained available for training after removing these "buffer" timepoints, the sampling process was repeated. This random sampling of test-train data was repeated 5,000 times, and for each sample the encoding model was trained and tested as described below.

To tune regularization parameters, the training data (for each test-train split) were further subdivided using fivefold cross-validation, whereby the encoding model was trained on 80% of training data and tested on the held out 20% of training data. This was repeated with five regularization hyperparameters, log-spaced between 10^{-1} and 10^{4}, and the parameter that minimized mean squared error was selected for use when the encoding model was trained on all the training data.

To account for the temporal delay between stimulus presentation and the BOLD response (reflecting the hemodynamic response; Friston et al., 1998), each feature was shifted to include delays of 0, 3, 6, and 9 timepoints. For each parcel, the regression model independently estimated a weight for each feature at each delay. The delay yielding the best prediction was retained for each feature-parcel pair. This allowed the model to flexibly capture parcel-specific and feature-specific temporal dynamics using a finite impulse response (FIR) modeling approach (Goutte et al., 2000), enabling the estimation of unique hemodynamic response functions per feature in each parcel (see also Dupre la Tour et al., 2022).

To evaluate model performance, the trained parcel-wise encoding models predicted BOLD responses in the held-out test data. Prediction accuracy was assessed using Pearson's correlation coefficient (*r*; Pearson, 1895), measuring the linear relationship between predicted and observed BOLD response for each parcel. Each model underwent the full train-and-test procedure, including random test set selection, regularization optimization, and delay selection, repeated over 5,000 iterations per participant. For each parcel, we computed the mean, standard deviation, and variance of prediction accuracy across the 5,000 iterations. The average prediction accuracy per participant represented a proxy for encoding model performance (i.e., how well could brain activity be predicted across all encoding model iterations). Individual-level variance, in turn, represented how variable prediction accuracies were across model

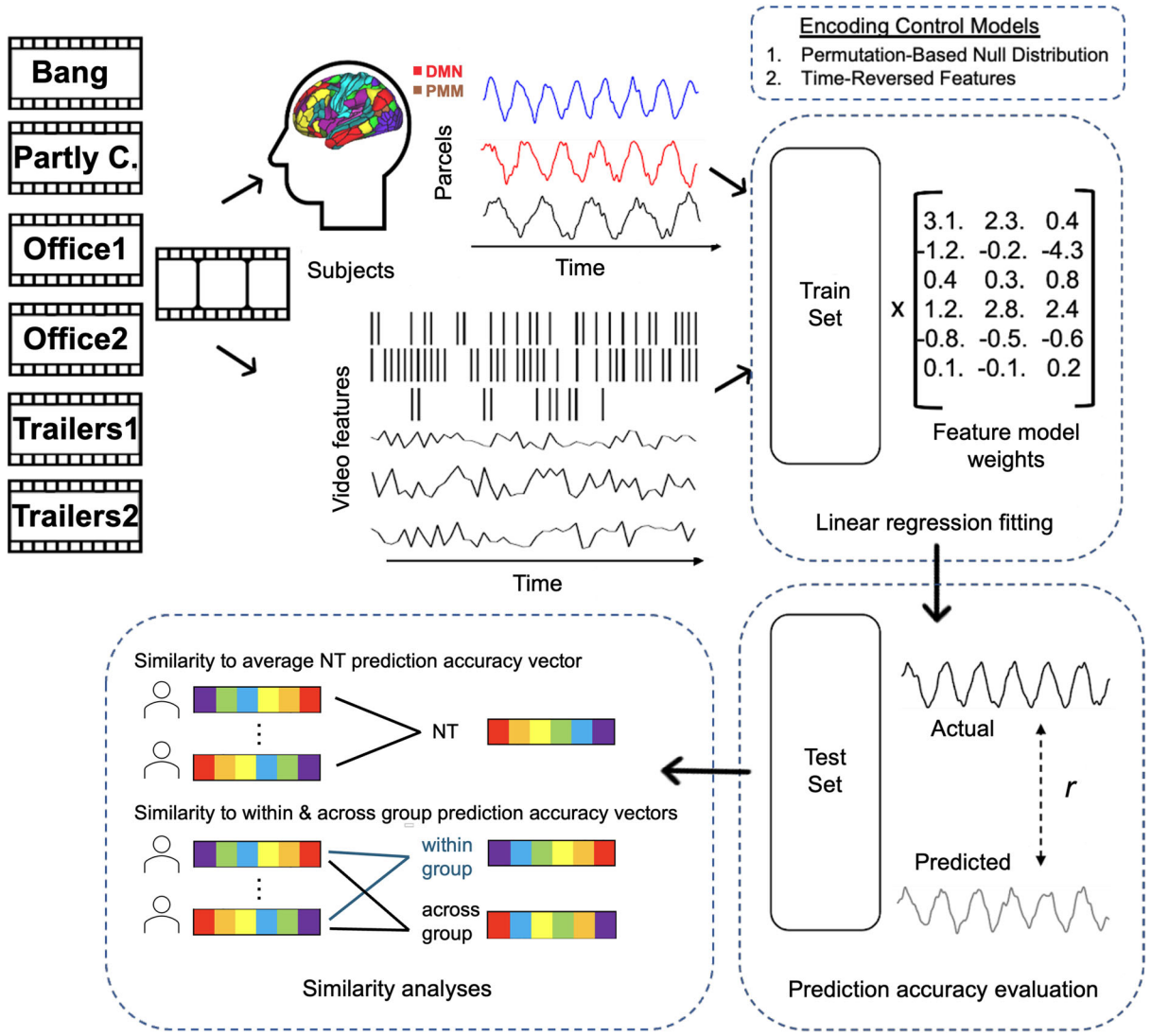


Figure 1. Overview of study design and simplified encoding model approach. Participants watched (and heard) six videos of different genres while in the fMRI scanner. Brain activity (parcel-wise; Glasser et al., 2016) and stimulus features (extracted from videos and audio waveform) were used as training data to fit a linear regression to predict held-out brain activity (test data). Prediction accuracy was evaluated by correlating predicted with measured fMRI data (Pearson's *r*) and compared against two control models (permuted null distribution and time-reversed feature model). Similarity of encoding patterns was compared with (1) the neurotypical average and (2) pairwise within and across groups. DMN, default mode network; PMM, posterior multimodal network; *r*, Pearson's correlation coefficient; NT, neurotypical participants; ASD, autistic participants; Partly C., Partly Cloudy.

iterations (i.e., due to different test-train splits). Thus, variance reflected a proxy for individual level variability in the correlation of brain responses with video features.

Encoding models were calculated for each individual participant (LeBel et al., 2023) and dependent measures were compared with control models, as well as between groups of participants, as outlined in the next section.

Statistical analysis

Evaluating within-group significance via permutation-based null distributions and time-reversed feature control models

To determine whether encoding models predicted brain activity significantly above chance, we used a nonparametric permutation test with 5,000 permutations (Holmes et al., 1996; Nichols and Holmes, 2002). For each parcel, we computed the observed mean correlation between predicted and actual BOLD responses across participants. A null distribution of mean correlations was generated by randomly flipping the sign of each participant's correlation value (±1) for each permutation. Empirical *p* values were calculated as the proportion of permuted means that were greater than or equal to the observed mean. To further address overfitting, we compared the observed encoding model accuracy to a control model that sought to predict brain responses from time-reversed features. Such time-reversed signals preserved the original power spectrum, temporal variation, and autocorrelation of the features, while disrupting temporal alignment with the BOLD signal. In other words, time-reversed features do not allow meaningful predictions of BOLD signal, and predictions made therefrom indicate whether the pipeline is susceptible to overfitting of spurious correlations. This serves as a strong conservative control model. Here, we conducted a one-sided permutation test between the observed and control distributions, testing the hypothesis that the model using original features would outperform the control using time-reversed features. Activity in a parcel was considered significantly predicted only if the observed mean correlation significantly exceeded both the null permutation distribution and the time-reversed feature control model.

Between-group significance assessed via two-sample permutation testing

For parcels where BOLD activity was significantly predicted in at least one of the two groups (i.e., NT and/or ASD) above the permuted null and time-reversed feature control model, group differences were

evaluated using two-sample permutation tests. Group labels were randomly shuffled across participants' correlation values, and the mean group difference was recomputed for each of 5,000 iterations to generate a null distribution. For direct between-group comparisons, we performed a two-sided test, calculating p values as the proportion of random permutations whose mean difference deviated from zero by at least as much as the observed mean difference. This approach tested for group differences in either direction without assuming which group would exhibit higher similarity or prediction accuracy.

Multiple-comparisons correction

For all statistical tests across parcels, we applied false discovery rate (FDR; Benjamini and Hochberg, 1995) correction to account for the number of parcels tested for each participant (i.e., 360) with a significance threshold of $q < 0.05$.

Linear mixed effects analyses

To complement the permutation-based statistical framework that included the control encoding models, we conducted a post hoc secondary analysis using linear mixed effects models (LMEs) using all 360 parcels (regardless of whether a given parcel's activity was predicted significantly above control models). These LMEs were intended to characterize encoding model prediction accuracy across diagnostic groups, cortical regions, and videos while accounting for repeated observations within participants. All models included acquisition site as a fixed effect and participant as a random intercept. Random intercepts for video or cortical region were included where appropriate. Fixed effects were evaluated using Wald z tests, nested models were compared using likelihood-ratio tests, and p values from regional and behavioral analyses were corrected for multiple comparisons using the Benjamini–Hochberg FDR ($p < 0.05$). We first tested whether prediction accuracy differed globally between autistic and neurotypical participants while controlling for acquisition site. To characterize the spatial distribution of prediction accuracy, we next fitted separate parcel- and network-level models, including participant and video random intercepts. These analyses were repeated separately in autistic and neurotypical participants to obtain group-specific regional estimates, as well as across all participants. We also tested whether prediction accuracy differed across videos using a global model that included participant and parcel random intercepts. Video effects were subsequently evaluated separately within each cortical parcel and functional network using analogous mixed effects models with participant-specific random intercepts.

Similarity analyses

To evaluate the extent to which individuals and groups exhibited similar (or dissimilar) patterns of prediction accuracy, we conducted two complementary analyses: (1) similarity to the average NT group response and (2) pairwise similarity within and across groups (Richardson et al., 2018; Turner et al., 2025) for similar approaches applied to functional connectivity. We hypothesized: (1) ASD encoding patterns would be less similar than NT to the NT group average; (2) ASD encoding patterns would be more variable than in the NT group; and (3) the within-group pairwise similarity would be particularly reduced in ASD for encoding patterns in networks relevant for social processing (i.e., DMN and PMM). We assessed encoding pattern similarity across five levels of granularity: (1) across the whole brain at the parcel level (360 parcels; Glasser et al., 2016); (2) at the network level (12 networks; Ji et al., 2019); (3) the top 10% most predictive parcels as defined in the neurotypical group (36 parcels); (4) DMN parcels (77 parcels); and (5) PMM parcels (7 parcels). Evaluating similarity across these five spatial granularities allowed us to probe whether individual differences in encoding pattern similarity were robust across the global brain, the most predictive regions (top 10%), and functional networks relevant for social information processing. Analyses were repeated across both encoding models (i.e., ALL, SOC).

Similarity to the average NT encoding pattern. To assess group-level differences in similarity to the NT pattern, we computed Fisher z-transformed Pearson's correlations between each participant's and NT group average encoding prediction accuracy vector. For NT participants, a leave-one-out (LOO) approach was used to avoid circularity. Autistic group differences in similarity to the average NT prediction accuracy were evaluated using a two-sided Mann–Whitney U test (Mann and Whitney, 1947).

Pairwise within- versus across-group encoding pattern similarity. To examine similarity of encoding accuracy patterns between participants, we first computed Fisher z-transformed pairwise correlations (Pearson's r) of each participant's encoding accuracy vector (e.g., encoding accuracy for all 360 parcels) to every other participant's encoding accuracy vector, within-group, and across-group. For each participant, we then averaged their within-group correlations (e.g., an autistic participant's correlation values to all other autistic participants) and across-group correlations (e.g., an autistic participant's correlation values to each neurotypical participant), creating two scores per participant: (1) average within-group similarity and (2) average across-group similarity. We hypothesized that within-group similarity would be reduced in the autistic as compared with the neurotypical group. We also hypothesized that for each group, the within-group similarity would be higher than across-group similarity. We applied paired, two-sided Mann–Whitney U test for within- versus across-group similarity comparisons and two-sided Wilcoxon signed-rank sum test for between group comparisons (e.g., NT vs ASD within-group similarity).

In an exploratory analysis to assess brain–behavior relationships, a within-group minus across-group pairwise similarity difference metric was calculated for autistic participants for each search space (i.e., whole brain parcel-level, whole brain network-level, 10% of parcels in NT adults with highest prediction accuracy, DMN parcels, and PMM parcels) and tested for correlation with ADOS, AQ, and MASC scores, separately. To complement, we also fitted separate LME models (using within-minus-across-group similarity as the dependent variable, with participant and video included as random intercepts) for AQ, ADOS, and MASC scores at both the parcel and network spatial scales.

Code accessibility

Code for minimal preprocessing of MRI data (https://fmriprep.org/en/20.0.7/), denoising fMRI data (https://github.com/adolphslab/rsDenoise), encoding models (https://github.com/gallantlab/himalaya; custom code implementation and statistical analyses and plotting of results are available https://research-git.uiowa.edu/scnlabp/ens_a).

Data availability

Code for fMRI data, encoding models, and statistical analyses are publicly available in online repositories (https://research-git.uiowa.edu/scnlabp/ens_a).

Results

We trained and tested parcel-wise encoding models using features extracted from six videos of different genres to predict brain activity in 360 cortical parcels in a medium sized sample of neurotypical and autistic adults (NT: $n = 90$, ASD: $n = 48$ after exclusions). Extracted features included visual and auditory statistics as well as relatively abstract socially relevant features (manually annotated). The abstract features were chosen to focus on social content (e.g., higher-level socio-cognitive aspects, such as theory of mind, valence, arousal) and spatial/auditory relationships relevant for social cognition (e.g., conversation, proximity between people in a scene).

Matched groups did not differ in how familiar they were with each video (e.g., Bang, $U = 1{,}075.0$, $p = 0.505$, $r_{rb} = -0.08$; Fig. 2, Table S3; see Table S5 for n-matched sample descriptive information) or verbal IQ (e.g., Bang, $t_{(97)} = 1.05$, $p = 0.301$, $d = 0.26$; Fig. 2, Table S3). As expected, the autistic groups showed overall difficulty in social cognitive functioning compared with neurotypical participants, as assessed with the MASC (e.g., Bang, $U = 370.0$, $p = 1.09 \times 10^{-4}$, $r_{rb} = -0.53$). Autistic participants also

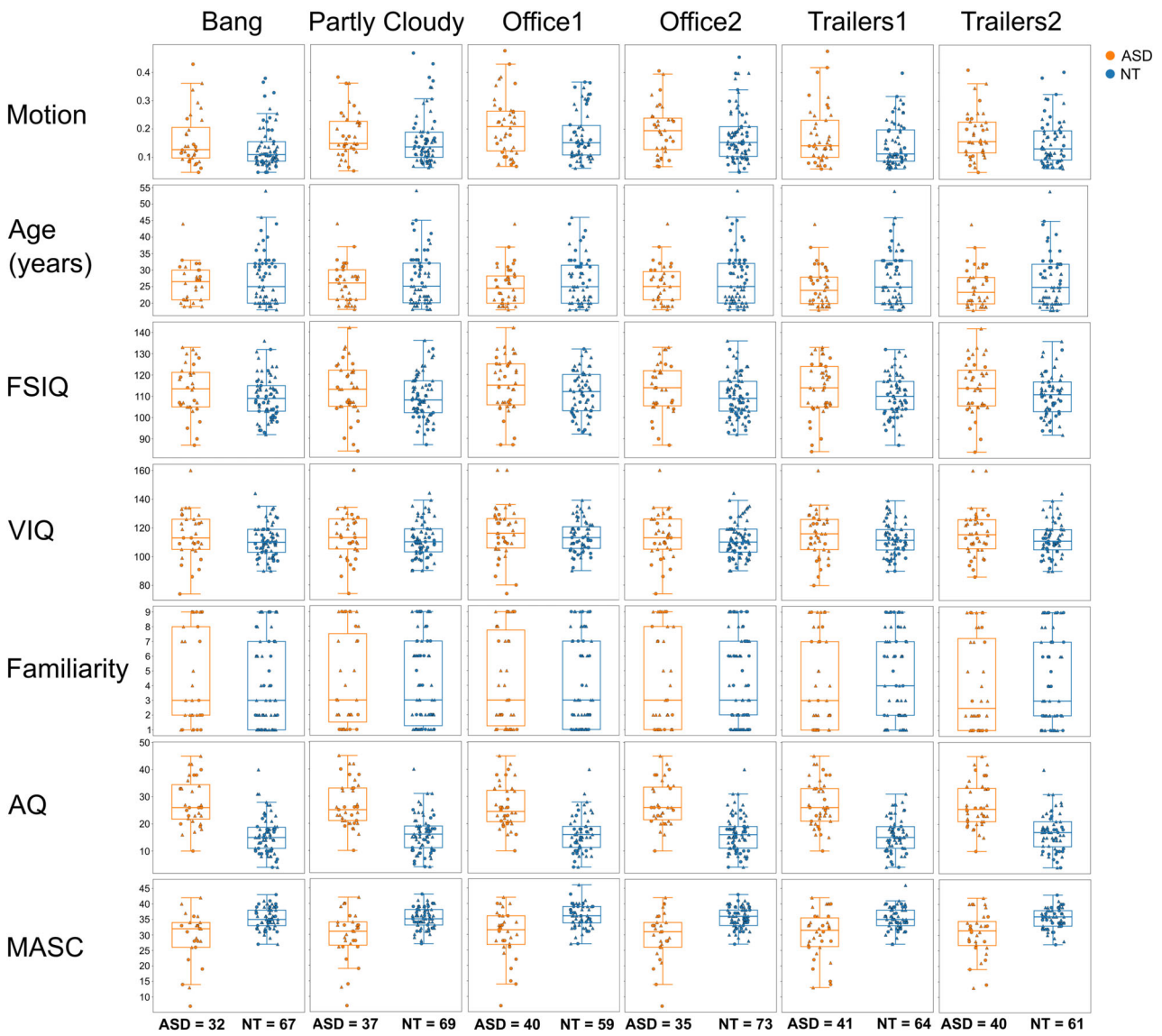


Figure 2. Demographic and socio-cognitive characteristics for each video by group. Boxplots showing (from top to bottom) head motion (mean filtered framewise displacement; FDfilt4), age (in years), intellectual functioning (FSIQ), verbal IQ (VIQ), familiarity with video content, self-reported levels of autistic symptoms (AQ), and mental state inference accuracy assessed with the MASC (Table S3). Individual datapoints are denoted by site (triangle, Caltech; circle, IU). ASD, autistic participants; NT, neurotypical participants; IU, Indiana University; Caltech, California Institute of Technology; MASC, Movie for the Assessment of Social Cognition; familiarity, how familiar participants were with each video; FSIQ, full scale intelligence quotient; VIQ, verbal intelligence quotient; AQ, Autism Spectrum Quotient; Motion, mean FDfilt4.

showed higher levels of autistic symptoms as measured by the AQ (e.g., Bang, $U = 1{,}836.0$, $p = 3.41 \times 10^{-9}$, $r_{rb} = -0.74$) and lower levels of self-reported empathy as measured by the EQ (e.g., Bang, $U = 446.5$, $p = 1.32 \times 10^{-5}$, $r_{rb} = -0.56$), compared with neurotypical participants.

Encoding models predicted brain activity across the whole brain robustly across videos—strikingly similar for autistic and neurotypical samples

First, we assessed the accuracy of how well the encoding models predicted changes in BOLD activity within each group for each video. Here, the NT group served as reference for typical model performance to compare against the hypothesized atypical model performance in the autistic group.

ALL model

Models trained upon all three feature spaces (visual, auditory, abstract) predicted brain activity in held-out data across multiple brain networks in the neurotypical sample significantly greater than chance, and significantly greater than the time-reversed feature control model, in line with prior work. The number of significantly predicted parcels varied per video (Bang, 61%; Partly Cloudy, 25%; Office1, 62%; Office2, 54%; Trailers1, 71%; Trailers2, 64%; Table 4). Activity could be significantly predicted in parcels within all 12 networks, with the exceptions of the ventral multimodal and orbito-affective networks in Bang and Office2. Across all videos, the auditory, language, secondary visual, and posterior multimodal networks were predicted with the highest accuracy and with the largest proportion of parcels per network. We also observed relatively high predictions across primary visual, dorsal attention, default mode, cingulo-opercular, frontoparietal, and somatomotor networks (Fig. 4).

Despite these overall similarities of predictions across the whole brain, idiosyncratic prediction patterns emerged across videos. In other words, specific videos showed different levels

Table 4. Proportion of significantly predicted parcels within each network

		Bang		Partly Cloudy		Office1		Office2		Trailers1		Trailers2	
Network	Group	ALL	SOC	ALL	SOC	ALL	SOC	ALL	SOC	ALL	SOC	ALL	SOC
Whole brain	NT	61	59	25	37	62	76	54	68	71	69	64	68
	ASD	53	54	11	30	54	59	49	57	52	43	51	42
Primary visual	NT	17	0	0	17	100	100	83	83	100	100	100	67
	ASD	17	0	0	17	100	100	83	67	100	100	83	50
Secondary visual	NT	72	61	57	70	96	96	100	100	100	93	98	98
	ASD	59	54	11	44	100	98	100	98	100	87	98	100
Auditory	NT	93	100	100	100	100	93	100	100	100	100	100	87
	ASD	93	93	100	100	87	87	93	93	93	87	93	80
Language	NT	91	91	57	83	96	100	87	100	70	83	91	87
	ASD	74	87	22	43	87	96	91	91	65	78	74	65
DMN	NT	47	47	6	16	58	74	18	45	62	57	61	57
	ASD	26	29	0	22	61	53	22	19	30	30	45	27
PMM	NT	86	43	71	100	100	100	100	100	100	100	100	100
	ASD	71	43	0	86	100	100	100	100	100	86	100	100
Dorsal attention	NT	91	91	17	39	83	78	78	91	96	91	70	91
	ASD	78	91	0	35	61	74	70	91	57	48	65	65
Frontoparietal	NT	60	64	4	18	60	86	22	40	74	68	46	54
	ASD	70	70	4	20	30	40	8	24	36	38	22	14
Somatomotor	NT	38	36	5	10	13	18	44	44	21	21	23	38
	ASD	36	41	5	8	5	8	31	49	26	3	8	18
Cingulo-opercular	NT	63	64	21	29	32	77	61	82	71	73	52	64
	ASD	59	64	16	27	25	46	46	63	43	20	36	18
Orbito-affective	NT	0	0	0	17	33	67	0	33	33	33	17	17
	ASD	0	0	0	0	17	17	17	33	17	0	33	0
Ventral multimodal	NT	0	0	25	75	25	25	0	0	50	25	50	50
	ASD	0	0	0	0	25	50	0	25	25	0	50	0

Results reported for each group (neurotypical, autistic) and each encoding model (ALL, SOC). Encoding prediction accuracy is the average Pearson's correlation coefficient across 5,000 train/test model iterations for each participant. Significant parcels included in averages for each network were significant within their respective group (i.e., NT or ASD) above the time-reversed feature control model and above the permuted null model [p(FDR) < 0.05 for each comparison]. See Figure 4 for average prediction accuracy per network. ALL, all features model; SOC, abstract features model; mean FD, average filtered framewise displacement traces; NT, neurotypical participants; ASD, autistic participants; %, percentage; FDR, false discovery rate correction.

of prediction accuracy for certain parcels (and networks) than other videos. For the auditory and language networks, prediction accuracy was highest for Bang and Trailers1 (both contained prominent soundtracks, as well as speech), while being lowest for Partly Cloudy, the video without speech (Figs. 3, 4). Activity in visual networks was most accurately predicted in Office1 and the two trailer videos (both contained many rapid scene cuts, and high-action sequences; Figs. 3, 4).

The remaining task-positive networks showed relatively low prediction accuracies: the cingulo-opercular network had overall lower prediction accuracies than the aforementioned networks; however, there were still several parcels with prediction accuracies significantly above chance and control models. These were present most prominently in Bang, Trailers1, Trailers2, and Office2, with few significant predictions in other videos. The frontoparietal network also showed relatively lower predictions across a smaller number of parcels above controls across videos, however notably lower than the dorsal attention network, which was prominent for Bang and Trailers1. Activity in posterior multimodal parcels was best predicted in the two trailers. Predictions of parcel-wise activity in the default mode network resembled that of the posterior multimodal network in general, although with notably higher performance in Bang and least in Partly Cloudy. Parcels in the somatomotor network were overall very sparsely predicted above control models. The highest number of significantly predicted parcels in the somatomotor network was found in Office2 and Bang. Trailers1 yielded generally higher prediction than Trailers2 across networks, with the exception of the ventral multimodal network. Similarly, Office1 yielded higher predictions than Office2, with the exception of auditory, language, and somatomotor networks. These differences in encoding prediction accuracy were likely due to specific differences in video content (or perhaps choice of features, which were slightly different for each video; Table 3).

The network distribution of top 10% parcels differed across videos. For Trailers1 the top 10% best predicted parcels contained: 24 parcels from the secondary visual network, 8 auditory, 2 language, and 2 posterior multimodal. For Office1 the top 10% of parcels contained: 14 parcels from the visual network, 10 auditory, 5 default, 4 dorsal attention, 2 language, and 1 posterior multimodal network.

SOC model

Encoding models trained on abstract, socially relevant features significantly predicted brain activity across multiple networks in the neurotypical sample, broadly resembling the pattern observed for the all-features encoding model. The proportion of significantly predicted parcels varied per video (Table 4). Notably, for all videos except Bang and Trailers1, the SOC feature model predicted a numerically higher proportion of significant parcels than the ALL model. Activity of parcels in all 12 networks could be significantly predicted, with the exception of the ventral multimodal and orbito-affective networks for Bang. Compared with the ALL model, the SOC feature model (for Office2) yielded significant predictions in the orbito-affective network, whereas the ALL model had no significant predictions in the network from any video.

Across all videos, the auditory language, posterior multimodal, and secondary visual networks showed the highest prediction accuracy and largest proportion of significant parcels

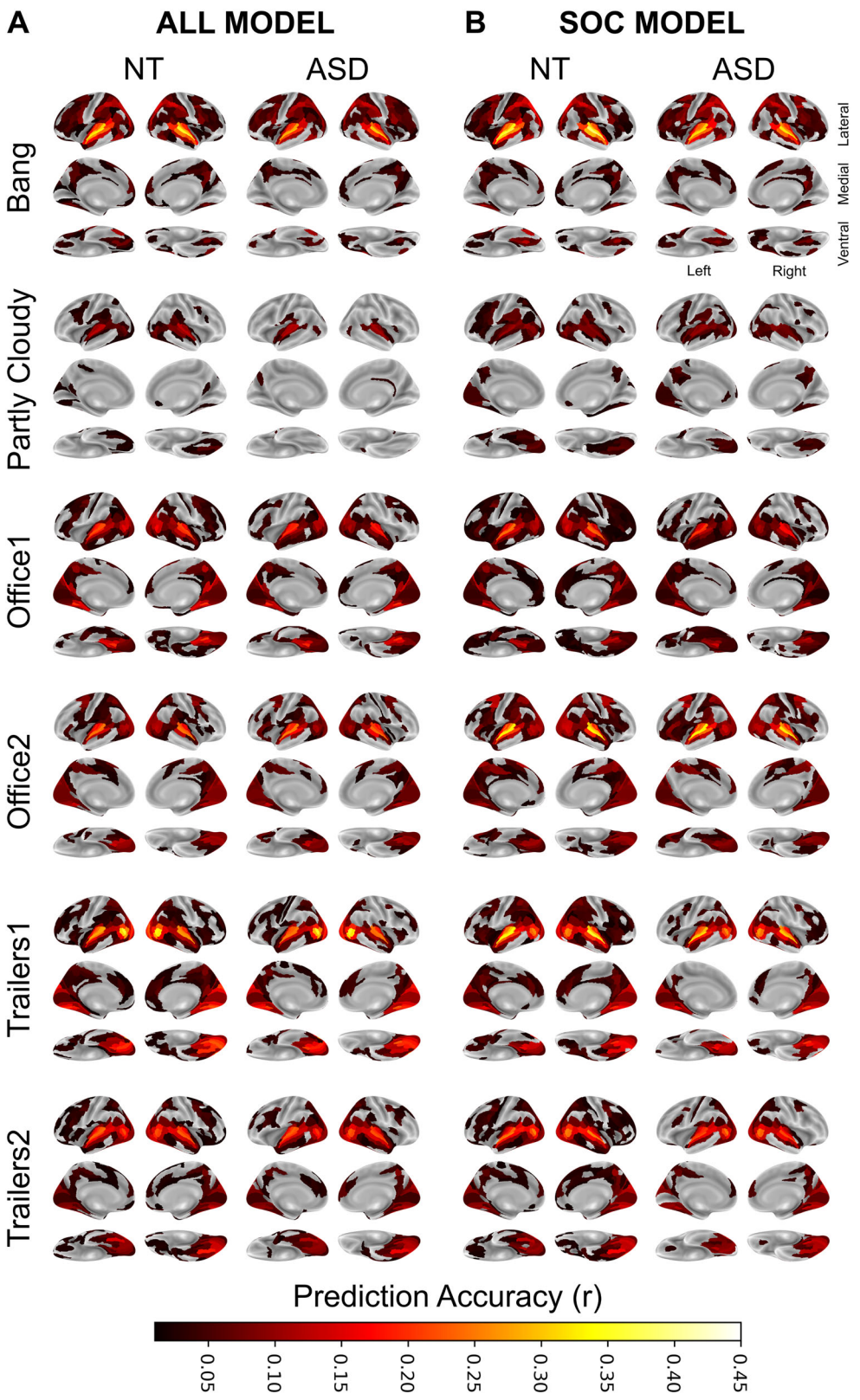


Figure 3. Encoding model prediction accuracy per video, model, and group. The prediction accuracy (Pearson's *r* between predicted and actual brain responses from held-out test data) for the ALL model (***A***) and the SOC model (***B***) averaged across participants in each group (ASD vs NT) per video. Data is presented on the inflated cortical surface (in medial, lateral, ventral views of the fsaverage template) in 360 cortical parcels (Glasser et al., 2016). Only parcels with significant above chance and above the time-reversed feature control models are displayed (pFDR < 0.05). ALL, all features model; SOC, abstract features model; NT, neurotypical participants; ASD, autistic participants; *r*, Pearson correlation coefficient; FDR, false discovery rate correction.

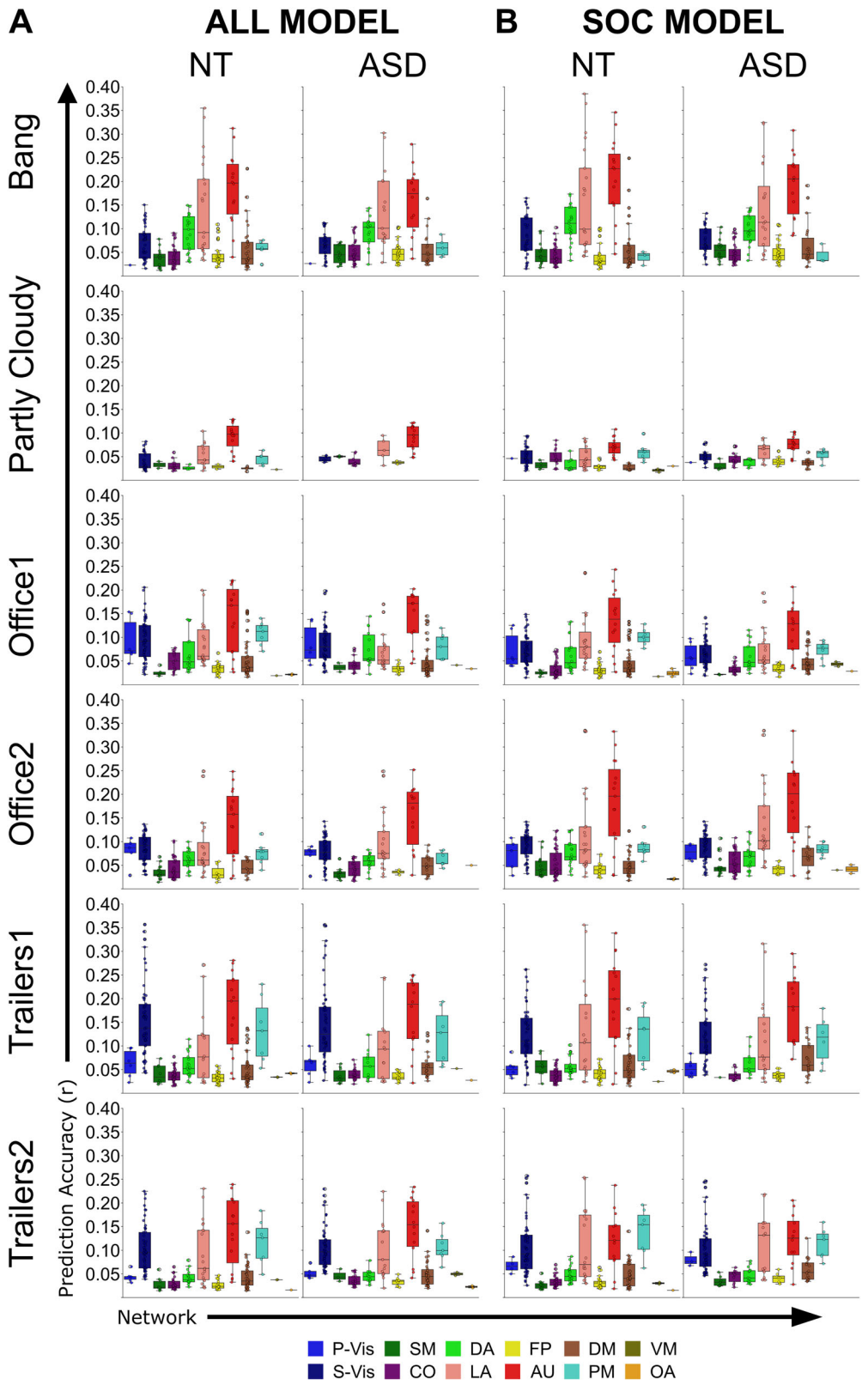


Figure 4. Encoding model prediction accuracy for each network per video, model, and group. The prediction accuracy (Pearson's *r* between predicted and actual brain responses from held-out test data) averaged across participants for each video for the ALL model (***A***) and the SOC model (***B***) per group (ASD vs NT). Bars are color coded by network. Boxplots represent variation across significant parcels above chance and above the time-reversed feature control models (pFDR < 0.05). See Table 4 for proportion of significant parcels per network. ALL, all features model; SOC, abstract features model; NT, neurotypical; ASD, autistic; *r*, Pearson's correlation coefficient; FDR, false discovery rate correction; P-Vis, primary visual; S-Vis, secondary visual; SM, somatomotor; CO, cingulo-opercular; DA, dorsal attention; LA, language; FP, frontoparietal; AU, auditory; DM, default mode; PM, posterior multimodal; VM, ventral multimodal; OA, orbito-affective.

(Fig. 4, Table 4). Relatively high predictions were observed across dorsal attention, primary visual, default mode, cingulo-opercular, somatomotor, and frontoparietal networks, largely mirroring ALL model results.

Idiosyncratic patterns emerged across videos (Figs. 3, 4). Prediction accuracy in auditory and language networks was highest for Bang and Trailers1 and lowest for Partly Cloudy, consistent with the ALL model. Primary and secondary visual network activity was best predicted in the Office videos and trailers, again largely matching ALL model patterns. Within hypothesized networks, prediction accuracy in the posterior multimodal and default mode networks was descriptively higher for Office2 compared with the ALL model, whereas Bang showed reduced posterior multimodal prediction performance. In contrast, default mode predictions were greater for Office2 and Partly Cloudy, with Office1 showing reduced prediction accuracy but an increased proportion of significant DMN parcels.

The remaining task-positive networks showed relatively low prediction accuracy, similar to the ALL model. The cingulo-opercular and frontoparietal networks exhibited sparse but significant predictions across videos (Fig. 4, Table 4), with dorsal attention showing comparatively stronger effects, particularly for Bang and Trailers1. Parcels in the somatomotor network were sparsely predicted, with the highest proportions in Office2, Trailers2, and Bang. Trailers1 generally yielded higher prediction accuracy compared with Trailers2, and Office2 had greater prediction accuracy than Office1 across most networks, patterns that diverged from the ALL model and suggest sensitivity to video content and feature selection.

Summary of ALL and SOC models

Across encoding models (i.e., ALL, SOC), we find the highest prediction accuracy in the auditory and language networks, with Bang having the highest average prediction accuracy across videos, and Partly Cloudy with the lowest. Among hypothesized networks, we consistently see the PMM with the highest average prediction accuracy, with video trailers having the highest average prediction accuracy.

Similar prediction accuracy across the cortex in neurotypical and autistic samples

One of our main hypotheses tested whether encoding models predicted brain activity in autistic participants less accurately than in neurotypicals. We tested this hypothesis, with the main statistical permutation-based framework, across the whole brain only in parcels that were significantly predicted above both chance and the control model in at least one of the two samples. For both the ALL and SOC model, there were no significant group differences in how well the model features predicted brain activity (all $p > 0.05$). Both groups showed strikingly similar levels of predictions in parcels across the whole brain (Figs. 3, 4). Among parcels significantly predicted in at least one group, and searching across all encoding models (i.e., ALL, SOC feature encoding models), we find no group differences in prediction accuracy.

Complementary LME-based analysis (using all 360 parcels, accounting for acquisition site and repeated observations) confirmed the absence of a significant group effect ($\beta = -0.0023$, SE = 0.0031, $z = -0.73$, $p = 0.465$, 95% CI [−0.0084, 0.0038]). Across all participants, prediction accuracy differed significantly across videos ($\chi^2(5) = 7{,}237.78$, $p < 0.001$; Fig. S4) and cortical parcels (likelihood-ratio test: $\chi^2(359) = 69{,}608.21$, $p < 0.001$; Fig. 5). In addition to global prediction accuracy differences across videos, parcel- and network-level prediction accuracy similarly demonstrated pronounced spatial variation across the cortex (Fig. S5).

Similar variance in individual encoding model prediction accuracy across groups

To test whether the autistic group would show greater variance in how accurate encoding models predicted brain activity over the 5,000 train/test iterations in each participant, we compared individual variance scores per video and parcel between groups. Against our hypothesis, we did not observe increased variance in autistic participants' individual prediction accuracies for any of the six videos, regardless of whether models were trained only on abstract features (SOC) or on all features (ALL; all $p > 0.05$).

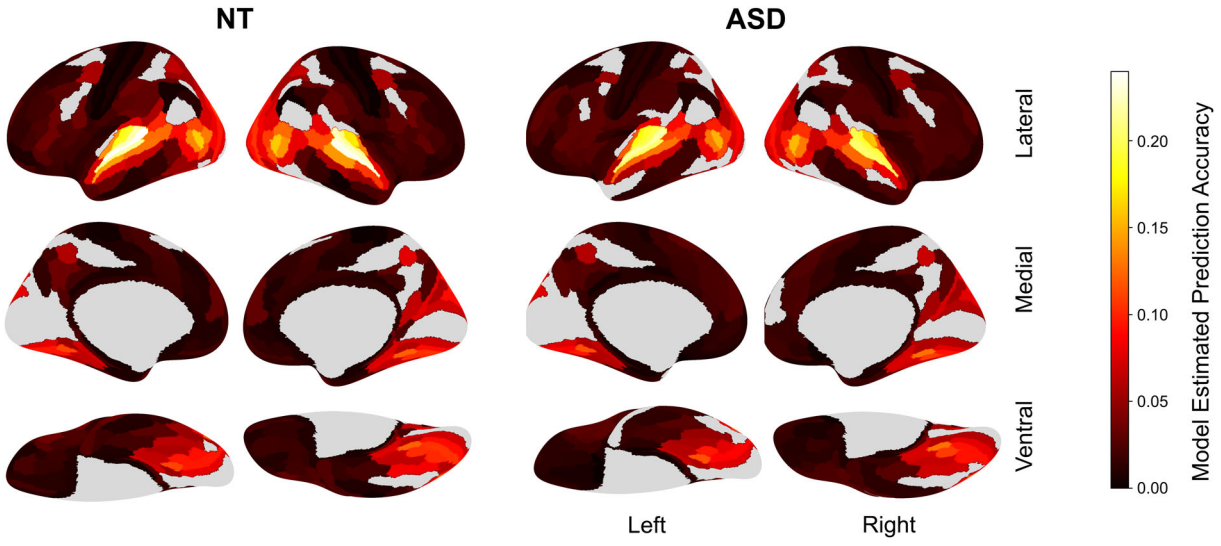


Figure 5. LME model prediction accuracy combined across videos. Linear mixed effects model estimates of encoding model prediction accuracy plotted for each cortical parcel combined across videos for NT and ASD participants. Values represent fitted prediction accuracy after accounting for acquisition site, with participant included as a random intercept. Prediction accuracy was highest in auditory, posterior multimodal, secondary visual, and language networks and showed highly similar spatial distributions between groups, consistent with the absence of a significant overall group effect. Gray parcels did not survive FDR correction (pFDR > 0.05). NT, neurotypical participants; ASD, autistic participants.

Reduced similarity to the average neurotypical encoding pattern in ASD

To test the hypothesis of reduced similarity to typical encoding patterns in ASD, we compared patterns of prediction accuracy across the brain. The pattern for each individual was compared with the NT group average pattern. This analysis was repeated at different levels of spatial granularity (i.e., whole brain parcel-level vector, 360×1; whole brain network-level vector, 12×1; 10% of parcels with highest prediction accuracy in NT adults, 36×1; DMN parcels, 77×1; and PMM parcels, 7×1). For all models, videos, and granularities, the neurotypical group was quantitatively more similar to the NT average than the autistic group (Table S4). While this trend was in line with our hypotheses, not all comparisons reached statistical significance.

In the ALL model, differences between groups (i.e., in similarity to the NT average) were significant for Bang and Trailers1 across all cortical parcels in the whole brain. Here, Bang showed the largest between group difference (NT, $r = 0.78$; ASD, $r = 0.64$; $U = 678$, $r_{rb} = -0.296$, $p = 0.003$), which remained significant when only considering the pattern for the top 10% best predicted parcels (NT, $r = 0.79$; ASD, $r = 0.62$; $U = 798$, $r_{rb} = -0.206$, $p = 0.041$; Table S4, Fig. 6). In contrast, spatial distribution of prediction accuracy was not significantly different for the two hypothesized networks (DMN, PMM) in Bang. For these networks, patterns across predicted parcels differed between groups only for Office1 (DMN: NT, $r = 0.55$; ASD, $r = 0.43$; $U = 890$, $r_{rb} = -0.208$, $p = 0.039$) and Office2 (PMM: NT, $r = 0.47$; ASD, $r = 0.23$; $U = 958$, $r_{rb} = -0.202$, $p = 0.036$). Similarity of network averages did not significantly differ between groups for any video. In other words, when using all features, similarity to the NT average network-level prediction accuracy pattern was consistent between groups.

In the SOC model, again the autistic group was less similar to the NT average prediction accuracies among all parcels in Bang (NT, $r = 0.81$; ASD, $r = 0.68$; $U = 642$, $r_{rb} = -0.323$, $p = 0.001$) and Trailers1 (NT, $r = 0.73$; ASD, $r = 0.65$; $U = 890$, $r_{rb} = -0.271$, $p = 0.006$) and additionally Office1 (NT, $r = 0.59$; ASD, $r = 0.48$; $U = 821$, $r_{rb} = -0.257$, $p = 0.011$), compared with the neurotypical group. For Trailers1 and Office1, similarity also differed significantly between groups when only considering the top 10% predicted parcels (Trailers1: NT, $r = 0.64$; ASD, $r = 0.48$; $U = 924$, $r_{rb} = -0.249$, $p = 0.011$; Office1: NT, $r = 0.55$; ASD, $r = 0.41$; U = 807, $r_{rb} = -0.267$, $p = 0.008$).

Significant differences in the prediction accuracy pattern were further present when considering network averages. Office1 also revealed group differences in pattern similarity to the NT average when considering only parcels in the default mode network (NT, $r = 0.46$; ASD, $r = 0.32$; $U = 854$, $r_{rb} = -0.234$, $p = 0.020$). Similarity to the NT average in parcels of the PMM was significantly increased for NT participants for Office2 and Bang (Office2: NT, $r = 0.38$; ASD, $r = 0.10$; U = 891, $r_{rb} = -0.244$, $p = 0.011$; Bang: NT, $r = 0.39$; ASD, $r = 0.12$; $U = 706$, $r_{rb} = -0.275$, $p = 0.006$; Table S4, Fig. 6).

Atypical pairwise encoding pattern similarity in ASD

To further assess heterogeneity of encoding prediction patterns within and across groups, we calculated pairwise similarity between each participant and (1) all participants in their respective group (i.e., within-group similarity) as well as (2) all participants in the other group (i.e., across-group similarity). We then tested the hypotheses whether within-group similarity would be greater than across-group similarity in each group (ASD-within vs ASD-across; NT-within vs NT-across) and whether within-group similarity would be greater in the neurotypical as compared with the autistic group (NT-within > ASD-within). Again, we conducted similarity analyses at five levels of spatial granularities (i.e., whole brain parcel-level, whole brain network-level, top 10% of parcels with highest prediction accuracy defined in NT participants, DMN parcels, and PMM parcels).

Within- versus across-group similarity

As expected, the neurotypical group showed numerically overall greater pairwise similarity of encoding prediction accuracy patterns to participants of the NT group, as compared with participants from the autistic group. This direction of effect was observed across both encoding models (ALL, SOC), all six videos and different spatial scales (Table S4, Fig. 7). In the ALL model, these within-group comparisons were significant for encoding pattern similarity across all 360 parcels, network averages, and PMM (apart from Office1 for PMM; Table S4). For the top 10% parcels and the DMN network, similarity was significantly greater for Bang (top 10%: NT, $r = 0.45$; ASD, $r = 0.37$; $W = 98$, $r_{rb} = -0.794$, $p = 7.89 \times 10^{-11}$; DMN: NT, $r = 0.26$; ASD, $r = 0.19$; $W = 53$, $r_{rb} = -0.829$, $p = 1.17 \times 10^{-11}$), Office1 (top 10%: NT, $r = 0.16$; ASD, $r = 0.15$; $W = 606$, $r_{rb} = -0.274$, $p = 0.035$; DMN: NT, $r = 0.26$; ASD, $r = 0.20$; $W = 43$, $r_{rb} = -0.827$, $p = 2.08 \times 10^{-10}$), and Trailers1 (top 10%: NT, $r = 0.29$; ASD, $r = 0.27$; $W = 407$, $r_{rb} = -0.529$, $p = 2.30 \times 10^{-5}$; DMN: NT, $r = 0.15$; ASD, $r = 0.12$; $W = 274$, $r_{rb} = -0.64$, $p = 3.01 \times 10^{-7}$) for the ALL model. Significant effects across videos and spatial scales were comparable for the SOC model with minor differences: within- versus across-group similarity was only numerically but not statistically different for Partly Cloudy in the top 10% parcels and for Trailers1 in the PMM. In contrast, Office2 similarity was significantly different in the SOC (but not ALL) model for the DMN (SOC: NT within-group, $r = 0.11$; NT across-group, $r = 0.10$; $W = 891$, $r_{rb} = -0.296$, $p = 0.012$; Table S4).

Strikingly, we observed the opposite direction of effects in the autistic group: the patterns of encoding accuracy observed in autistic participants were (on average) more similar to those of neurotypical participants than to other autistic participants. This effect was observed for most videos and spatial levels (Table S4, Fig. 7) and significant when comparing similarity across all 360 parcels (for all videos except Partly Cloudy) and when comparing prediction accuracy network averages (for all videos). When comparing similarity across the top 10% parcels, DMM, and PMM, only some videos showed significant group differences (top 10%: Bang, Trailers1, Trailers2; DMN: Bang, Office1, Trailers1; PMM: Trailers1, Trailers2). Across-group similarity was significantly increased in Partly Cloudy across all cortical parcels and when comparing network averages (ALL: ASD within-group, $r = 0.31$; ASD across-group, $r = 0.37$; $W = 82$, $r_{rb} = -0.668$, $p < 0.001$; Table S4). Within-group similarity was significantly greater than across-group in Office1 across PMM patterns (ALL: ASD within-group, $r = 0.22$, ASD across-group, $r = 0.16$; $W = 236$, $r_{rb} = -0.370$, $p = 0.019$; Table S4). For the model using only abstract features (SOC), effects were largely comparable to the ALL model with some minor differences (Table S4); significant differences in the same direction across > within were observed in the SOC model, with additional significant effects found (Office1, top 10%: ASD within-group, $r = 0.14$, ASD across-group, $r = 0.20$; $W = 78$, $r_{rb} = -0.706$, $p = 1.16 \times 10^{-6}$; Trailers2, DMN: ASD within-group, $r = 0.10$, ASD across-group, $r = 0.11$; $W = 215$, $r_{rb} = -0.414$, $p = 0.007$).

Within-group similarity compared across groups

Consistent with our hypothesis, we found reduced within-group similarity in the autistic group as compared with the neurotypical

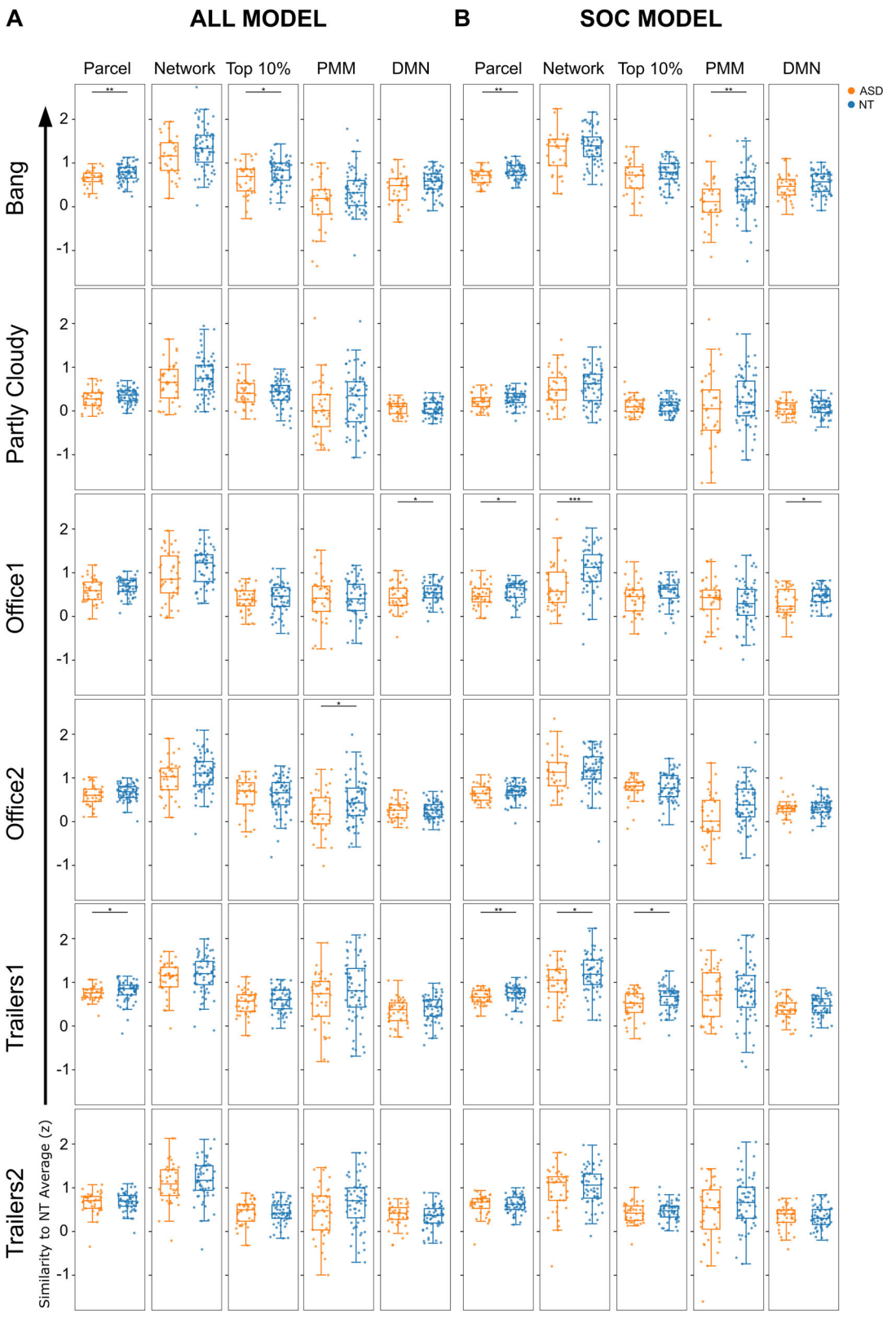


Figure 6. Similarity to the average typical (NT) encoding prediction accuracy per video and group. Typical encoding pattern similarity for all six videos (top to bottom) and different levels of spatial granularity (left to right: whole brain parcels, networks, top 10%, PMM, and DMN) by group (blue, NT; orange, ASD) for the ALL model (***A***), and the SOC model (***B***); see Table 3. ALL, all features model; SOC, abstract features model; NT, neurotypical participants; ASD, autistic participants; *z*, *z*-transformed Pearson's *r* correlation coefficient; top 10% parcels, 10% of parcels with highest prediction accuracy in NT group; DMN, default mode network; PMM, posterior multimodal network; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

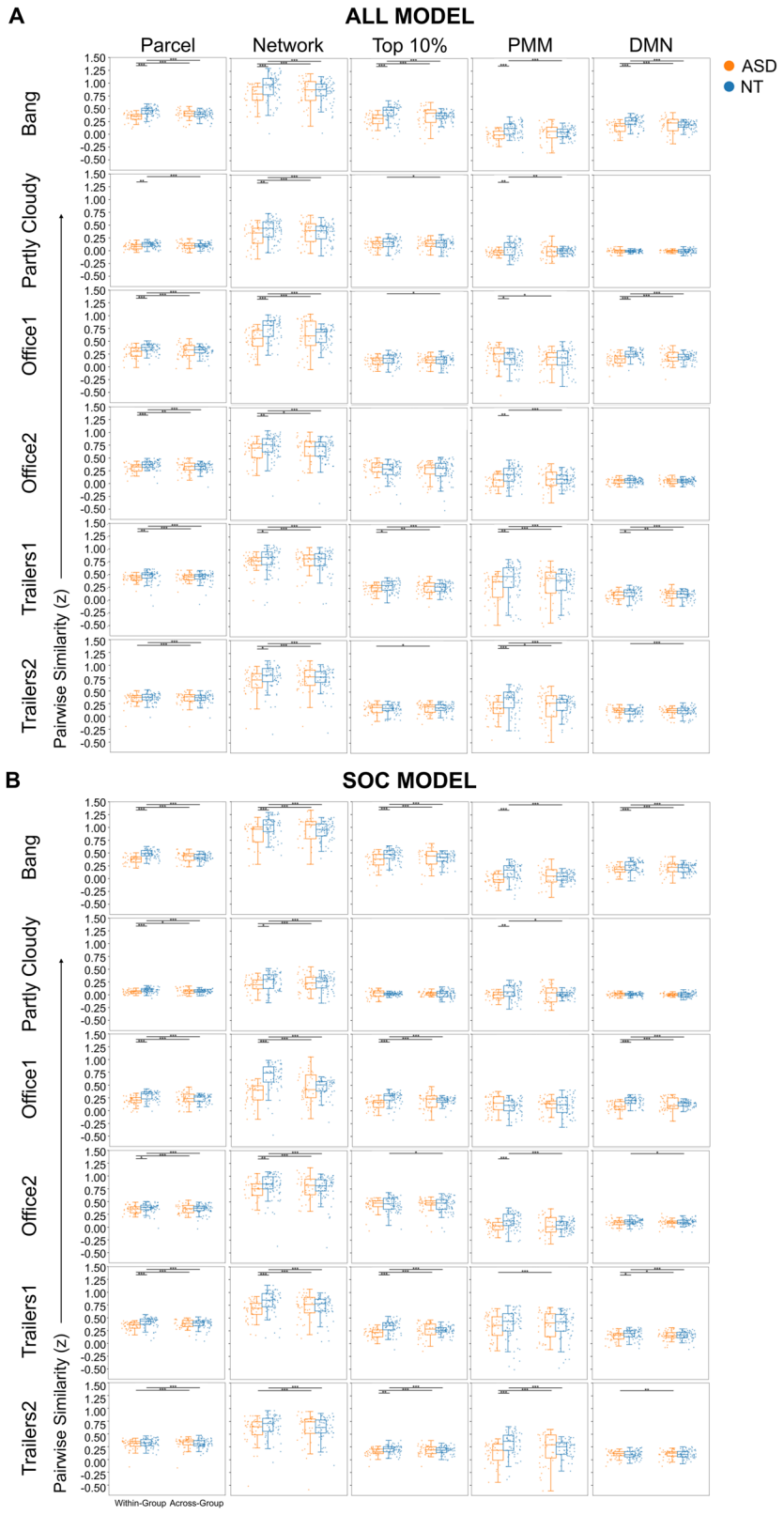


Figure 7. Pairwise similarity of encoding accuracy pattern compared within and between groups per video. Pairwise encoding pattern similarity for all six videos (top to bottom) and different levels of spatial granularity (left to right whole brain parcels, networks, top 10%, PMM, and DMN) by group (blue, NT; orange, ASD) for the ALL model (***A***), and the SOC model (***B***); see Table S4. ALL, all features model; SOC, abstract features model; NT, neurotypical participants; ASD, autistic participants; *z*, *z*-transformed Pearson's *r* correlation coefficient; top 10% Parcels, 10% of parcels with highest prediction accuracy in NT group; DMN, default mode network parcels; PMM, posterior multimodal network parcels; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

group. This effect was relatively robust for most—but not all—videos, with varying effects across different levels of spatial granularities.

For the ALL model, within-group similarity (of network prediction accuracy patterns) was significantly greater in neurotypical than in autistic participants for all videos at the whole brain parcel level and network level, except for Trailers2 (Table S4, Fig. 7). For the top 10% of parcels, we found significant similarity differences only for Bang (NT, $r = 0.45$; ASD, $r = 0.30$; $U = 1{,}717$, $r_{rb} = 0.485$, $p = 1.42 \times 10^{-6}$) and Trailers1 (NT, $r = 0.29$; ASD, $r = 0.24$; $U = 1{,}658$, $r_{rb} = 0.222$, $p = 0.023$). Similarity in the DMN differed between groups also for these two videos (Bang: NT, $r = 0.26$; ASD, $r = 0.14$; $U = 1{,}672$, $r_{rb} = 0.451$, $p = 7.29 \times 10^{-6}$; Trailers1: NT, $r = 0.15$; ASD, $r = 0.11$; $U = 1{,}681$, $r_{rb} = 0.237$, $p = 0.016$; and Office1: NT, $r = 0.26$; ASD, $r = 0.17$; $U = 1{,}807$, $r_{rb} = 0.449$, $p = 7.92 \times 10^{-6}$). For the PMM, within-group similarity differed for all videos between groups; however, for Office1, the within-group similarity was greater in ASD than in NT (ALL: NT within-group, $r = 0.16$; ASD within-group, $r = 0.22$; $U = 903$, $r_{rb} = -0.199$, $p = 0.049$; Table S4).

Again, we observed overall similar between-group results in pattern similarity for the SOC model with few minor differences to the ALL model (Fig. 7, Table S4).

Exploratory brain–behavior analyses

Since the similarity analyses suggested relatively robust group differences between within- versus across-group similarity, we conducted an exploratory brain–behavior analysis. Specifically, we investigated the behavioral relevance of the individual within- minus across-group similarity difference to diagnostic scores (ADOS), levels of autistic symptoms (AQ), and social-cognitive task performance (accuracy in correct mental state inferences measured with the MASC) in the autistic sample.

For Trailers1, we found significant positive correlations between the AQ scores and the similarity difference derived by all parcels, network averages, and for the top 10% of parcels in the ALL model (Fig. 8). This indicates that greater within-group similarity was associated with higher levels of autistic symptoms (as measured by the AQ). In the SOC model, lower within-group similarity among the top 10% of parcels was associated with greater accuracy in the MASC (Trailers1, $r = -0.36$, $p = 0.039$; Fig. 8). For the ADOS, significant associations were observed in Office1 and Office2. In Office1, greater within-group similarity among DMN parcels in the ALL model was associated with higher ADOS scores (Office1, $r = 0.351$, $p = 0.049$). In Office2, greater within-group similarity across all parcels in the SOC model was associated with lower ADOS scores (Office2, $r = -0.432$, $p = 0.022$). In other words, individuals with greater similarity to others within the autistic group (i.e., "more autistic" pattern similarity) exhibited more autistic symptoms (as measured by the ADOS and AQ, see discussion) and greater difficulty in mentalizing.

The complementary LME-based approach further confirmed no robust relations between within- and across-group similarity and AQ, ADOS, or MASC scores combined across videos after FDR correction (all pFDR > 0.05).

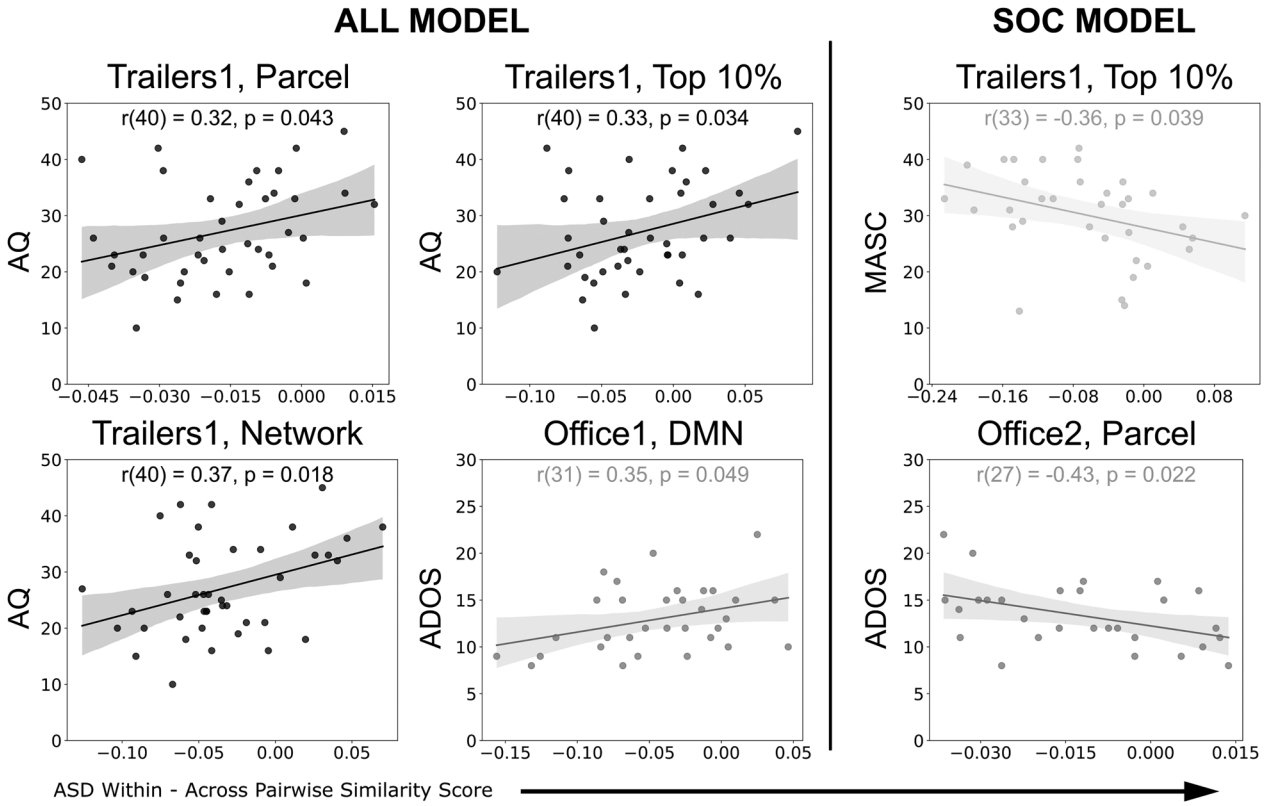


Figure 8. Exploratory brain–behavior relationships in autistic participants. Correlational analysis of the difference in encoding accuracy pattern (within minus between group similarity) with levels of autistic symptoms (AQ, black), diagnostic scores (ADOS, gray), and mental state inference performance (MASC, light gray) in ASD. ALL, all features model; SOC, abstract features modal; AQ, Autism Spectrum Quotient score; ADOS, Autism Diagnostic Observation Schedule; MASC, Movie for the Assessment of Social Cognition; top 10% parcels, 10% of parcels with highest prediction accuracy in NT group; DMN, default mode network parcels; PMM, posterior multimodal network parcels; parcel, all cortical parcels; network, all parcels averaged by network; *r*, Pearson's correlation coefficient; *p*, probability value.

Discussion

This study applied high-dimensional encoding models to neuroimaging data of complex dynamic social information presented in multiple videos of different genres in ASD. Dynamic audio, visual, and abstract social features robustly predicted brain activity in both groups across six videos with consistent cortical network patterns. Contrary to our hypothesis, encoding model prediction accuracies were not reduced, and variance (across different test-train splits) was not increased in ASD across models, brain regions, or networks. However, groups differed in encoding pattern similarity: within-group similarity was decreased in ASD, as was the similarity to the average neurotypical encoding pattern. Encoding (dis)similarity correlated with autistic symptoms and social task performance, however these effects were weak. The results suggest ASD participants showed more heterogeneous regional encoding patterns, even though encoding accuracies and within-subject encoding variability were comparable across groups.

Typical encoding accuracy in ASD

We hypothesized reduced encoding accuracy in ASD, especially with social features (i.e., without lower-level sensory features), assuming these features would induce neural responses underlying social processing. Given the consistency in results across multiple analyses (six videos, two models, and several spatial levels), we consider the present work a robust rejection of our initial hypothesis. Regions in sensory (visual and auditory), as well as language networks, showed the highest and most widespread prediction accuracies, consistent with prior encoding model studies with video and narrative paradigms (Huth et al., 2016; McMahon et al., 2023). Notably, the auditory network was best predicted across all videos, especially in Bang, a socially salient and acoustically complex video. In contrast, activity in auditory and language networks was least accurately predicted in Partly Cloudy, possibly due to the lack of dialogue. Partly Cloudy also showed the lowest overall prediction likely due to the least amount of data (<6 min).

Brain activity was well predicted beyond sensory cortices in both groups, including integrative and associative networks (particularly the PMM, DMN, dorsal attention, and frontoparietal regions). This was against our hypothesis that the PMM (integrating cross-modal information; Jakobs et al., 2012; Xi et al., 2019) and DMN [associated with social-cognitive and self-referential processing (Northoff et al., 2006; Qin and Northoff, 2011) and autistic pathophysiology (Kennedy and Courchesne, 2008; Assaf et al., 2010)] specifically would show the greatest encoding group differences based on their functional relevance for dynamic social information processing.

The lack of group differences in encoding accuracy has several potential explanations. It is unlikely that results were due to overfitting noise. We used a relatively conservative significance assessment and restricted our main analyses to parcels where prediction accuracy exceeded both chance and control model predictions (trained and tested upon time-reversed features). In fact, the lack of group differences can be interpreted as independent sample replication of the encoding patterns, while also highlighting the stability of feature encoding across groups. Another possible explanation for the present results might be the video and/or feature choices. We chose six videos that showed prior relevance for social information processing (Richardson et al., 2018; Byrge et al., 2022) and ASD pathophysiology (Byrge et al., 2015; Pantelis et al., 2015; Kliemann et al., 2021; Turner et al., 2025). Crucially, prior work (Byrge et al., 2022) showed that the videos elicited highly consistent brain responses across two acquisition sites. Including videos of varying genres and content (thriller, animated nonspeech short film, movie trailers with rapidly changing content, socially awkward sitcom) was motivated to maximize the chance of detecting group differences in neural representations. Feature selection emphasized relevance for social cognition and autistic phenotypes. However, selected videos and features represent only a small set of possibilities for the complex social scenes we encounter in daily life. It is possible that other videos and features, such as ratings by autistic rather than neurotypical individuals (Wang et al., 2015), might reveal larger differences across groups. Additionally, we only measured passive perception, whereas explicit tasks demand integration of perception and behavior.

Atypical encoding pattern similarity in autism

This study builds upon previous findings by incorporating high-dimensional encoding models (i.e., using more and different sets of stimulus features) and more diverse stimuli (i.e., across genres). Atypical encoding pattern similarity was most robust across videos and spatial granularities when comparing similarity within and across groups. The reverse pattern of within-group effects in ASD (greater similarity to neurotypical than to autistic participants) may indicate more variable mappings between stimulus features and neural responses, and not simply noise. Rather than a uniform hypoactivation (which should have reduced encoding), atypical spatial encoding patterns may represent internal neural representations of external dynamic events.

The findings of heterogeneous encoding patterns may reflect autistic symptom heterogeneity (Lenroot and Yeung, 2013; Szatmari et al., 2015; Masi et al., 2017; Lombardo et al., 2019; Tang et al., 2023) consistent with prior reports of reduced inter-subject correlation during video watching (Salmi et al., 2013; Byrge et al., 2015; Turner et al., 2025). It is important to note that heterogeneity of encoding patterns (across subjects/groups) and encoding accuracy variability (within-subjects) are only indirect proxies for underlying brain activity and may reflect differences in how well specific features capture individual stimulus–response mappings (Hasson et al., 2009). Group-level metrics (mean, median) may obscure individual variability emphasizing the importance of explicitly examining individual differences in ASD (Byrge et al., 2015; Lombardo et al., 2019). Autistic perception may rely on different neural solutions to the same computational demands, yielding equally accurate but individually variable representations. This view aligns with contemporary theoretical perspectives describing autism in terms of neural heterogeneity (Lombardo et al., 2019).

Exploratory brain–behavior relationships may suggest behavioral relevance of similarity differences

In the autistic sample, we found weak associations of encoding pattern similarity with autistic symptoms (ADOS, AQ) and with social behavior (MASC). Greater within-group similarity among certain cortical parcels was related to higher symptom severity and socio-cognitive functioning, while divergence from these patterns was observed sparsely for the SOC model. Effects were model, stimulus, and sample dependent and overall weak (not surviving multiple-comparison corrections). It is noteworthy that the direction of some brain–behavior associations implies that increasing symptom severity could be related to more homogenous encoding patterns. This in turn implies phenotypic subgroups might be detectable by encoding similarity

(and dis-similarity), though this is beyond the scope of the current work. Our results suggest that such relationships are likely specific to video content, features, sample (including the comparison sample used to derive similarity), and behavioral metrics. Although these interpretations remain speculative, these findings may highlight the (potentially nuanced) relationship between cortical feature representation and behavior relevant to the autistic phenotype, while highlighting the need for replication.

Comparable intraindividual variability in encoding accuracy across groups

Against the hypothesis of increased intraindividual neural variability (Dinstein et al., 2012; Haigh et al., 2015; Heller Murray et al., 2022), we did not observe increased variance in individual encoding models in autism, in line with other recent findings using ISC (Mangnus et al., 2025). While our approach to variability approximation via encoding variance is not directly akin to quantifying variability of neural responses to different trials, it was based on the fundamental assumption that more stable neural responses would result in more consistent feature-brain-response mappings across different timepoints. As each test-train split used different combinations of timepoints, this would translate to less variability in prediction accuracy across splits. Again, we interpret the robustness of these results across all videos and different spatial scales as strong evidence of typical within-subject encoding model variability in the autistic group.

Recommendations for future generalization

Despite two acquisition sites and a relatively large NT sample for an extensive imaging protocol, the ASD sample size remained modest (partly due to the COVID-19 pandemic). Although the observed effects were largely consistent across comparisons of *n*-matched samples (in addition to motion, FSIQ, age, sex; Figs. S6–S9, Tables S5–S7), larger samples may enable the detection of subtle effects or interactions. Observed results may reflect the parcel-level encoding model used. While parcel-level models capture population-level trends, voxel-level models may better capture individual variability (Busch et al., 2024). Future work should compare both approaches to test whether finer-scale models reveal additional individual differences.

The inclusion of temporally resolved stimulus features such as gaze direction, facial affect, or prosody may refine social encoding estimates. Additionally, individualized stimulus features (e.g., emotional or semantic interpretations, participant engagement, across different timescales) could improve model performance and may better reflect how each individual experiences and processes the videos. Eye-tracking may further account for attentional effects and explain individual variation. Concurrent behavioral metrics and higher temporal-resolution modeling (e.g., electroencephalography) could help disambiguate sensory-driven versus social-cognitive encoding components. Extending this approach to longitudinal and developmental designs could test whether encoding patterns are stable across time or responsive to social experience and intervention in autism. Voxel-wise encoding models may be relevant to test spatial hypotheses in specific localized regions of interest, and different regression models (Dupre la Tour et al., 2022) may be used to directly compare models and feature space contributions. Combining encoding models with functional connectivity and representational similarity analyses could further determine whether heterogeneity reflects variability in local tuning or network-level integration.

Conclusions

The application of encoding models to neuroimaging in ASD indicated that both autistic and neurotypical adults recruited similarly distributed cortical systems to represent complex dynamic multimodal features with characteristic and consistent patterns across the cortex. Importantly, this should not be interpreted as direct evidence for intact or impaired "neural encoding" per se, but rather as evidence that stimulus–response relationships captured by these models are similar at the population level. Autism may not be characterized by globally reduced encoding prediction accuracy but by greater interindividual variability in how these stimulus–response mappings are spatially distributed across the cortex.

References

Abell F, Happé F, Frith U (2000) Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. Cogn Dev 15:437–471.

Andrews-Hanna JR, Smallwood J, Spreng RN (2014) The default network and self-generated thought: component processes, dynamic control, and clinical relevance. Ann N Y Acad Sci 1316:29–52.

Assaf M, Jagannathan K, Calhoun VD, Miller L, Stevens MC, Sahl R, O'Boyle JG, Schultz RT, Pearlson GD (2010) Abnormal functional connectivity of default mode sub-networks in autism spectrum disorder patients. Neuroimage 53:247–256.

Avants BB, Epstein CL, Grossman M, Gee JC (2008) Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med Image Anal 12:26–41.

Banker SM, Schafer M, Barkley S, Trayvick J, Chen A, Peters AW, Thinakaran AA, Gu X, Foss-Feig JH, Schiller D (2025) Neural tracking of social navigation in autism spectrum disorder. Biol Psychiatry 100:389–398.

Baron-Cohen S, Wheelwright S (2004) The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. J Autism Dev Disord 34:163–175.

Behzadi Y, Restom K, Liau J, Liu TT (2007) A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage 37:90–101.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol 57:289–300.

Braga RM, Buckner RL (2017) Parallel interdigitated distributed networks within the individual estimated by intrinsic functional connectivity. Neuron 95:457–471 e455.

Bruneau EG, Pluta A, Saxe R (2012) Distinct roles of the "shared pain" and "theory of mind" networks in processing others' emotional suffering. Neuropsychologia 50:219–231.

Busch EL, Rapuano KM, Anderson KM, Rosenberg MD, Watts R, Casey BJ, Haxby JV, Feilong M (2024) Dissociation of reliability, heritability, and predictivity in coarse- and fine-scale functional connectomes during development. J Neurosci 44:e0735232023.

Byrge L, Dubois J, Tyszka JM, Adolphs R, Kennedy DP (2015) Idiosyncratic brain activation patterns are associated with poor social comprehension in autism. J Neurosci 35:5837–5850.

Byrge L, Kliemann D, He Y, Cheng H, Tyszka JM, Adolphs R, Kennedy DP (2022) Video-evoked fMRI BOLD responses are highly consistent across different data acquisition sites. Hum Brain Mapp 43:2972–2991.

Byrge L, Kennedy DP (2020) Accurate prediction of individual subject identity and task, but not autism diagnosis, from functional connectomes. Hum Brain Mapp 41:2249–2262.

Campbell KL, Shafto MA, Wright P, Tsvetanov KA, Geerligs L, Cusack R, Cam CAN, Tyler LK (2015) Idiosyncratic responding during movie-watching predicted by age differences in attentional control. Neurobiol Aging 36:3045–3055.

Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162–173.

Czekoova K, Marecek R, Stanek R, Hartley C, Kessler K, Hlavata P, Oslejskova H, Brazdil M, Shaw DJ (2025) Altered patterns of dynamic functional connectivity underpin reduced expressions of social-emotional reciprocity in autistic adults. Autism Res 18:725–740.

Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9:179–194.
d'Ascoli S, Rapin J, Benchetrit Y, Banville H, King JR (2025) TRIBE: TRImodal brain encoder for whole-brain fMRI response prediction. *arXiv preprint*.
Dinstein I, Pierce K, Eyler L, Solso S, Malach R, Behrmann M, Courchesne E (2011) Disrupted neural synchronization in toddlers with autism. Neuron 70:1218–1225.
Dinstein I, Heeger DJ, Lorenzi L, Minshew NJ, Malach R, Behrmann M (2012) Unreliable evoked responses in autism. Neuron 75:981–991.
Dubois JM, Galdi P, Han Y, Paul LK, Adolphs R (2018) Resting-state functional brain connectivity best predicts the personality dimension of openness to experience. Personal Neurosci 1:e6.
Dufour N, Redcay E, Young L, Mavros PL, Moran JM, Triantafyllou C, Gabrieli JD, Saxe R (2013) Similar brain activation during false belief tasks in a large sample of adults with and without autism. PLoS One 8:e75468.
Dupre la Tour T, Eickenberg M, Nunez-Elizalde AO, Gallant JL (2022) Feature-space selection with banded ridge regression. Neuroimage 264: 119728.
Dziobek I, Fleck S, Kalbe E, Rogers K, Hassenstab J, Brand M, Kessler J, Woike JK, Wolf OT, Convit A (2006) Introducing MASC: a movie for the assessment of social cognition. J Autism Dev Disord 36:623–636.
Eickhoff SB, Milham M, Vanderwal T (2020) Towards clinical applications of movie fMRI. Neuroimage 217:116860.
Esteban O, Birman D, Schaer M, Koyejo OO, Poldrack RA, Gorgolewski KJ (2017) MRIQC: advancing the automatic prediction of image quality in MRI from unseen sites. PLoS One 12:e0184661.
Esteban O, et al. (2019) fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat Methods 16:111–116.
Evans AC, Janke AL, Collins DL, Baillet S (2012) Brain templates and atlases. Neuroimage 62:911–922.
Finn ES, Shen X, Scheinost D, Rosenberg MD, Huang J, Chun MM, Papademetris X, Constable RT (2015) Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. Nat Neurosci 18:1664–1671.
Finn ES, Bandettini PA (2021) Movie-watching outperforms rest for functional connectivity-based prediction of behavior. Neuroimage 235: 117963.
Fonov VS, Evans AC, McKinstry RC, Almli CR, Collins DL (2009) Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. Neuroimage 47:S102.
Friedman L, et al. (2008) Test-retest and between-site reliability in a multicenter fMRI study. Hum Brain Mapp 29:958–972.
Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg MD, Turner R (1998) Event-related fMRI: characterizing differential responses. Neuroimage 7:30–40.
Geschwind DH, Levitt P (2007) Autism spectrum disorders: developmental disconnection syndromes. Curr Opin Neurobiol 17:103–111.
Giannelli M, Diciotti S, Tessa C, Mascalchi M (2010) Characterization of Nyquist ghost in EPI-fMRI acquisition sequences implemented on two clinical 1.5 T MR scanner systems: effect of readout bandwidth and echo spacing. J Appl Clin Med Phys 11:3237.
Glasser MF, et al. (2016) A multi-modal parcellation of human cerebral cortex. Nature 536:171–178.
Gorgolewski KJ, et al. (2016) The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci Data 3:160044.
Goutte C, Nielsen FA, Hansen LK (2000) Modeling the haemodynamic response in fMRI using smooth FIR filters. IEEE Trans Med Imaging 19:1188–1201.
Greve DN, Fischl B (2009) Accurate and robust brain image alignment using boundary-based registration. Neuroimage 48:63–72.
Gruskin DC, Rosenberg MD, Holmes AJ (2020) Relationships between depressive symptoms and brain responses during emotional movie viewing emerge in adolescence. Neuroimage 216:116217.
Hahamy A, Behrmann M, Malach R (2015) The idiosyncratic brain: distortion of spontaneous connectivity patterns in autism spectrum disorder. Nat Neurosci 18:302–309.
Haigh SM, Heeger DJ, Dinstein I, Minshew N, Behrmann M (2015) Cortical variability in the sensory-evoked response in autism. J Autism Dev Disord 45:1176–1190.
Happe F, Frith U (2006) The weak coherence account: detail-focused cognitive style in autism spectrum disorders. J Autism Dev Disord 36: 5–25.
Haskins AJ, Mentch J, Botch TL, Garcia BD, Burrows AL, Robertson CE (2022) Reduced social attention in autism is magnified by perceptual load in naturalistic environments. Autism Res 15:2310–2323.
Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R (2004) Intersubject synchronization of cortical activity during natural vision. Science 303:1634–1640.
Hasson U, Furman O, Clark D, Dudai Y, Davachi L (2008a) Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. Neuron 57:452–462.
Hasson U, Yang E, Vallines I, Heeger DJ, Rubin N (2008b) A hierarchy of temporal receptive windows in human cortex. J Neurosci 28:2539–2550.
Hasson U, Avidan G, Gelbard H, Vallines I, Harel M, Minshew N, Behrmann M (2009) Shared and idiosyncratic cortical activation patterns in autism revealed under continuous real-life viewing conditions. Autism Res 2: 220–231.
Hasson U, Malach R, Heeger DJ (2010) Reliability of cortical activity during natural stimulation. Trends Cogn Sci 14:40–48.
He Y, Byrge L, Kennedy DP (2020) Nonreplication of functional connectivity differences in autism spectrum disorder across multiple sites and denoising strategies. Hum Brain Mapp 41:1334–1350.
Heller Murray ES, Segawa J, Karahanoglu FI, Tocci C, Tourville JA, Nieto-Castanon A, Tager-Flusberg H, Manoach DS, Guenther FH (2022) Increased intra-subject variability of neural activity during speech production in people with autism spectrum disorder. Res Autism Spectr Disord 94:101955.
Holmes AP, Blair RC, Watson JD, Ford I (1996) Nonparametric analysis of statistic images from functional mapping experiments. J Cereb Blood Flow Metab 16:7–22.
Hong J, DaWalt LS, Taylor JL, Haider A, Mailick M (2023) Autism through midlife: trajectories of symptoms, behavioral functioning, and health. J Neurodev Disord 15:36.
Hus V, Lord C (2014) The autism diagnostic observation schedule, module 4: revised algorithm and standardized severity scores. J Autism Dev Disord 44:1996–2012.
Huth AG, Nishimoto S, Vu AT, Gallant JL (2012) A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76:1210–1224.
Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL (2016) Natural speech reveals the semantic maps that tile human cerebral cortex. Nature 532:453–458.
Im EJ, Shirahatti A, Isik L (2025) Early neural development of social interaction perception: evidence from voxel-wise encoding in young children and adults. J Neurosci 45:e2284232024.
Imai S (1983) Cepstral analysis synthesis on the mel frequency scale. In: ICASSP '83. IEEE International Conference on Acoustics, Speech, and Signal Processing, pp 93–96.
Jakobs O, Langner R, Caspers S, Roski C, Cieslik EC, Zilles K, Laird AR, Fox PT, Eickhoff SB (2012) Across-study and within-subject functional connectivity of a right temporo-parietal junction subregion involved in stimulus-context integration. Neuroimage 60:2389–2398.
Jenkinson M, Bannister P, Brady M, Smith S (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17:825–841.
Ji JL, Spronk M, Kulkarni K, Repovs G, Anticevic A, Cole MW (2019) Mapping the human brain's cortical-subcortical functional network organization. Neuroimage 185:35–57.
Kana RK, Maximo JO, Williams DL, Keller TA, Schipul SE, Cherkassky VL, Minshew NJ, Just MA (2015) Aberrant functioning of the theory-of-mind network in children and adolescents with autism. Mol Autism 6:59.
Kennedy DP, Redcay E, Courchesne E (2006) Failing to deactivate: resting functional abnormalities in autism. Proc Natl Acad Sci U S A 103: 8275–8280.
Kennedy DP, Courchesne E (2008) Functional abnormalities of the default network during self- and other-reflection in autism. Soc Cogn Affect Neurosci 3:177–190.
King JB, et al. (2019) Generalizability and reproducibility of functional connectivity in autism. Mol Autism 10:27.
Kirk PA, Holmes AJ, Robinson OJ (2023) Anxiety shapes amygdala-prefrontal dynamics during movie watching. Biol Psychiatry Glob Open Sci 3:409–417.
Klein A, et al. (2017) Mindboggling morphometry of human brains. PLoS Comput Biol 13:e1005350.
Kliemann D, Richardson H, Anzellotti S, Ayyash D, Haskins AJ, Gabrieli JDE, Saxe RR (2018) Cortical responses to dynamic emotional facial

expressions generalize across stimuli, and are sensitive to task-relevance, in adults with and without autism. Cortex 103:24–43.
Kliemann D, Adolphs R, Paul LK, Tyszka JM, Tranel D (2021) Reorganization of the social brain in individuals with only one intact cerebral hemisphere. Brain Sci 11:965.
Kliemann D, et al. (2022) Caltech conte center, a multimodal data resource for exploring social cognition and decision-making. Sci Data 9:138.
Kliemann D, Galdi P, Van De Water AL, Egger B, Jarecka D, Adolphs R, Ghosh SS (2024) Resting-state functional connectivity of the amygdala in autism: a preregistered large-scale study. Am J Psychiatry 181:1076–1085.
Kruger G, Glover GH (2001) Physiological noise in oxygenation-sensitive magnetic resonance imaging. Magn Reson Med 46:631–637.
Lahera G, Boada L, Pousa E, Mirapeix I, Moron-Nozaleda G, Marinas L, Gisbert L, Pamias M, Parellada M (2014) Movie for the assessment of social cognition (MASC): Spanish validation. J Autism Dev Disord 44:1886–1896.
Lanczos C (1964) Evaluation of noisy data. J S Ind Appl Math Ser B Numer Anal 1:76–85.
LeBel A, Wagner L, Jain S, Adhikari-Desai A, Gupta B, Morgenthal A, Tang J, Xu L, Huth AG (2023) A natural language fMRI dataset for voxelwise encoding models. Sci Data 10:555.
Lenroot RK, Yeung PK (2013) Heterogeneity within autism spectrum disorders: what have we learned from neuroimaging studies? Front Hum Neurosci 7:733.
Li L, Zheng Q, Xue Y, Bai M, Mu Y (2024) Coactivation pattern analysis reveals altered whole-brain functional transient dynamics in autism spectrum disorder. Eur Child Adolesc Psychiatry 33:4313–4324.
Li X, Wu S, Yang W, Jiang C, Zhao Z, Wang H, Lin F (2025) Abnormal resting-state brain network dynamics in toddlers with autism spectrum disorder. Eur Child Adolesc Psychiatry 34:2391–2401.
Lombardo MV, Lai MC, Baron-Cohen S (2019) Big data approaches to decomposing heterogeneity across the autism spectrum. Mol Psychiatry 24:1435–1450.
Loomes R, Hull L, Mandy WPL (2017) What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. J Am Acad Child Adolesc Psychiatry 56:466–474.
Lord C, Rutter M, DiLavore PC, Risi S, Gotham K, Bishop S (2012) *ADOS-2: autism diagnostic observation schedule*, Ed 2. Los Angeles, California: Western Psychological Services.
Mangnus M, Koch SBJ, Cai K, Greidanus Romaneli M, Hagoort P, Basnakova J, Stolk A (2025) Preserved spontaneous mentalizing amid reduced inter-subject variability in autism during a movie narrative. Biol Psychiatry Cogn Neurosci Neuroimaging 10:1057–1066.
Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat 18:50–60.
Martinez G, Alexandre C, Mam-Lam-Fook C, Bendjemaa N, Gaillard R, Garel P, Dziobek I, Amado I, Krebs MO (2017) Phenotypic continuum between autism and schizophrenia: evidence from the movie for the assessment of social cognition (MASC). Schizophr Res 185:161–166.
Masi A, DeMayo MM, Glozier N, Guastella AJ (2017) An overview of autism spectrum disorder, heterogeneity and treatment options. Neurosci Bull 33:183–193.
McDermott JH, Simoncelli EP (2011) Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. Neuron 71:926–940.
McGlone F, Wessberg J, Olausson H (2014) Discriminative and affective touch: sensing and feeling. Neuron 82:737–755.
McHugh ML (2013) The chi-square test of independence. Biochem Med (Zagreb) 23:143–149.
McMahon E, Bonner MF, Isik L (2023) Hierarchical organization of social action features along the lateral visual pathway. Curr Biol 33:5035–5047 e5038.
Moore BCJ (2013) *An introduction to the psychology of hearing*, Ed 6. Leiden, The Netherlands: Brill.
Muller N, Baumeister S, Dziobek I, Banaschewski T, Poustka L (2016) Validation of the movie for the assessment of social cognition in adolescents with ASD: fixation duration and pupil dilation as predictors of performance. J Autism Dev Disord 46:2831–2844.
Nakai T, Constant-Varlet C, Prado J (2024) Encoding models for developmental cognitive computational neuroscience: promise, challenges, and potential. Dev Cogn Neurosci 70:101470.
Nastase SA, Gazzola V, Hasson U, Keysers C (2019) Measuring shared responses across subjects using intersubject correlation. Soc Cogn Affect Neurosci 14:667–685.
Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp 15:1–25.
Northoff G, Heinzel A, de Greck M, Bermpohl F, Dobrowolny H, Panksepp J (2006) Self-referential processing in our brain–a meta-analysis of imaging studies on the self. Neuroimage 31:440–457.
Nunez-Elizalde AO, Dupré la Tour T, Visconti di Oleggio Castello M (2022) gallantlab/pymoten: v0.0.4 (Version 0.0.4) [Computer software]. Zenodo.
Oliva A, Torralba A (2001) Modeling the shape of the scene: a holistic representation of the spatial envelope. Int J Comput Vis 42:145–175.
Padmanabhan A, Lynch CJ, Schaer M, Menon V (2017) The default mode network in autism. Biol Psychiatry Cogn Neurosci Neuroimaging 2:476–486.
Pantelis PC, Byrge L, Tyszka JM, Adolphs R, Kennedy DP (2015) A specific hypoactivation of right temporo-parietal junction/posterior superior temporal sulcus in response to socially awkward situations in autism. Soc Cogn Affect Neurosci 10:1348–1356.
Pearson's K (1895) VII. Note on regression and inheritance in the case of two parents. Proc R Soc Lond 58:240–242.
Pearson's K (1900) X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philos Mag Ser 5:157–175.
Pedregosa F, et al. (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830.
Persichetti AS, Shao J, Gotts SJ, Martin A (2025) A functional parcellation of the whole brain in high-functioning individuals with autism spectrum disorder reveals atypical patterns of network organization. Mol Psychiatry 30:1518–1528.
Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE (2012) Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59:2142–2154.
Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE (2014) Methods to detect, characterize, and remove motion artifact in resting state fMRI. Neuroimage 84:320–341.
Power JD, Lynch CJ, Silver BM, Dubin MJ, Martin A, Jones RM (2019) Distinctions among real and apparent respiratory motions in human fMRI data. Neuroimage 201:116041.
Qin P, Northoff G (2011) How is our self related to midline regions and the default-mode network? Neuroimage 57:1221–1233.
Redcay E, Dodell-Feder D, Mavros PL, Kleiner M, Pearrow MJ, Triantafyllou C, Gabrieli JD, Saxe R (2013) Atypical brain activation patterns during a face-to-face joint attention game in adults with autism spectrum disorder. Hum Brain Mapp 34:2511–2523.
Reher K, Sohn P (2009) Partly Cloudy[Motion Picture] (Pixar Animation Studios and Walt Disney Pictures, 2009).
Richardson H, Lisandrelli G, Riobueno-Naylor A, Saxe R (2018) Development of the social brain from age three to twelve years. Nat Commun 9:1027.
Robertson CE, Baron-Cohen S (2017) Sensory perception in autism. Nat Rev Neurosci 18:671–684.
Rodgaard EM, Jensen K, Vergnes JN, Soulieres I, Mottron L (2019) Temporal changes in effect sizes of studies comparing individuals with and without autism: a meta-analysis. JAMA Psychiatry 76:1124–1132.
Salmi J, et al. (2013) The brains of high functioning autistic individuals do not synchronize with those of others. Neuroimage Clin 3:489–497.
Satterthwaite TD, et al. (2013) An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. Neuroimage 64:240–256.
Schad DC, Dixit S, Keck J, Studenyak V, Shpilevoi A, Bicanski A (2025) Vibe: video-input brain encoder for fMRI response modeling. *arXiv preprint*, *arXiv:2507.17958*.
Schaller UM, Rauh R (2017) What difference does it make? Implicit, explicit and complex social cognition in autism spectrum disorders. J Autism Dev Disord 47:961–979.
Shama DM, et al. (2025) Harnessing trial-to-trial variability of EEG spectral characteristics to understand autism. J Autism Dev Disord.
Sinha P, Kjelgaard MM, Gandhi TK, Tsourides K, Cardinaux AL, Pantazis D, Diamond SP, Held RM (2014) Autism as a disorder of prediction. Proc Natl Acad Sci U S A 111:15220–15225.

Stewart CA, Welch V, Plale B, Fox G, Pierce M, Sterling T (2017) Indiana University Pervasive Technology Institute.

Student's (1908) The probable error of a mean. Biometrika 6:1–25.

Szatmari P, et al. (2015) Developmental trajectories of symptom severity and adaptive functioning in an inception cohort of preschool children with autism spectrum disorder. JAMA Psychiatry 72:276–283.

Tang X, et al. (2023) A study of genetic heterogeneity in autism spectrum disorders based on plasma proteomic and metabolomic analysis: multiomics study of autism heterogeneity. MedComm (2020) 4:e380.

Tomasello M, Carpenter M, Call J, Behne T, Moll H (2005) Understanding and sharing intentions: the origins of cultural cognition. Behav Brain Sci 28:675–691; discussion 691–735.

Turner JM, Byrge L, Richardson H, Galdi P, Kennedy DP, Kliemann D (2025) Social inference brain networks in autistic adults during movie-viewing: functional specialization and heterogeneity. Mol Autism 16:42.

Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, Gee JC (2010) N4ITK: improved N3 bias correction. IEEE Trans Med Imaging 29:1310–1320.

Uddin LQ, Kinnison J, Pessoa L, Anderson ML (2014) Beyond the tripartite cognition-emotion-interoception model of the human insular cortex. J Cogn Neurosci 26:16–27.

Vanderwal T, Kelly C, Eilbott J, Mayes LC, Castellanos FX (2015) Inscapes: a movie paradigm to improve compliance in functional magnetic resonance imaging. Neuroimage 122:222–232.

Wang S, Jiang M, Duchesne XM, Laugeson EA, Kennedy DP, Adolphs R, Zhao Q (2015) Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking. Neuron 88:604–616.

Washington SD, et al. (2014) Dysmaturation of the default mode network in autism. Hum Brain Mapp 35:1284–1296.

Wechsler D (2011) *Wechsler abbreviated scale of intelligence, second edition (WASI-II)*. Bloomington, MN: Pearson.

Wimmer H, Perner J (1983) Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition 13:103–128.

Xi Y, Li Q, Gao N, He S, Tang X (2019) Cortical network underlying audio-visual semantic integration and modulation of attention: an fMRI and graph-based study. PLoS One 14:e0221185.

Yang Z, et al. (2020) Individualized psychiatric imaging based on inter-subject neural synchronization in movie watching. Neuroimage 216:116227.

Zhang Y, Brady M, Smith S (2001) Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans Med Imaging 20:45–57.