

Journal Pre-proof

Altered Theory of Mind Engagement and Neural Alignment in Social Anxiety During Movie Viewing

Saskia B.J. Koch, Margot Mangnus, Robin Devillers, Peter Hagoort, Jana Bašnáková, Arjen Stolk



PII: S2667-1743(26)00034-0

DOI: <https://doi.org/10.1016/j.bpsgos.2026.100721>

Reference: BPSGOS 100721

To appear in: *Biological Psychiatry Global Open Science*

Received Date: 29 August 2025

Revised Date: 16 January 2026

Accepted Date: 3 March 2026

Please cite this article as: Koch S.B.J., Mangnus M., Devillers R., Hagoort P., Bašnáková J. & Stolk A., Altered Theory of Mind Engagement and Neural Alignment in Social Anxiety During Movie Viewing, *Biological Psychiatry Global Open Science* (2026), doi: <https://doi.org/10.1016/j.bpsgos.2026.100721>.

This is a PDF of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability. This version will undergo additional copyediting, typesetting and review before it is published in its final form. As such, this version is no longer the Accepted Manuscript, but it is not yet the definitive Version of Record; we are providing this early version to give early visibility of the article. Please note that Elsevier's sharing policy for the Published Journal Article applies to this version, see: <https://www.elsevier.com/about/policies-and-standards/sharing#4-published-journal-article>. Please also note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2026 Published by Elsevier Inc. on behalf of Society of Biological Psychiatry.

Altered Theory of Mind Engagement and Neural Alignment in Social Anxiety During Movie Viewing

Saskia B. J. Koch^{a,b}, Margot Mangnus^a, Robin Devillers^c, Peter Hagoort^{a,c}, Jana Bašnáková^a & Arjen Stolk^d

^aDonders Institute for Brain, Cognition, and Behaviour, Radboud University Nijmegen, the Netherlands

^bBehavioral Science Institute, Radboud University Nijmegen, the Netherlands

^cMax Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

^dPsychological and Brain Sciences, Dartmouth College, Hanover, NH, USA

Corresponding author

Saskia B. J. Koch, Donders Centre for Cognitive Neuroimaging, P.O. Box 9101, NL-6500 HB, Nijmegen, the Netherlands, Phone: +31(024) 36 10750, E-mail: saskia.koch@donders.ru.nl

Ethical considerations

The study was approved by the local ethics committee (Committee on Research Involving Human Subjects, Arnhem-Nijmegen region, the Netherlands, file number: 2019-6059).

Consent to participate

Participants provided verbal and written consent before their participation.

Declaration of conflicting interest

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding statement

This work was supported by the NWO Gravitation Grant (No. 024.001.006) awarded to the Language in Interaction Consortium. SBJK is funded by a Starting Grant from the European Research Council (ERC, No. 101163381 - *SocialDynamics*).

Data availability statement

Data and analysis scripts are publicly available in the Radboud Data Repository (<https://doi.org/10.34973/qhd6-df16>).

Running title

Theory of Mind and Neural Variability in Social Anxiety.

Keywords

Functional magnetic resonance imaging, pupillometry, intersubject correlation, naturalistic neuroscience, interpretation bias, social anxiety, theory of mind

Abstract

Background: Social anxiety is marked by a persistent fear of social scrutiny, yet its neurocognitive mechanisms remain incompletely understood. We tested whether social anxiety is associated with altered engagement of Theory of Mind (ToM) systems supporting inferences about others' thoughts and emotions, or with broader interpretive tendencies biasing social information processing toward more internally guided responses.

Methods: Functional MRI, heart rate, and pupil diameter were recorded from 43 individuals with elevated social anxiety and 43 low-anxiety controls during viewing of a nonverbal animated film. ToM engagement was assessed via scene-specific neural activation, while broader interpretive processing was indexed using dynamic intersubject correlation (ISC), quantifying the degree to which neural responses were shared versus idiosyncratic across viewers over time.

Results: Socially anxious participants showed reduced activation in the left posterior superior temporal sulcus (pSTS) during ToM-relevant scenes, alongside preserved engagement of the broader ToM network. Dynamic ISC analyses revealed increased alignment in early sensory regions and reduced alignment in higher-order regions outside the ToM network, consistent with more individualized processing at higher levels. These effects occurred without group differences in autonomic arousal or pupil-linked attention, and an exploratory comparison with an autism cohort revealed opposite alignment patterns in overlapping higher-order regions.

Conclusions: Social anxiety is associated with focal alterations in ToM-related processing and broader shifts in movie-driven neural alignment across the cortical hierarchy. Divergent alignment patterns in social anxiety and autism suggest that movie-driven neural alignment may serve as a transdiagnostic marker distinguishing mechanistically distinct forms of atypical social cognition.

Plain Language Summary

Social anxiety is marked by fear of being judged by others, but it remains unclear how it shapes the interpretation of social situations. We measured brain activity, heart rate, and eye responses while people with higher or lower social anxiety watched a short, animated film depicting social interactions. Individuals with elevated social anxiety showed subtle changes in brain activity related to understanding others, along with more individualized brain responses during the movie, even without signs of increased arousal or attention. These findings suggest that social anxiety alters how social information is processed, even in situations that are not overtly threatening.

Introduction

Social anxiety disorder, or social phobia, is characterized by a persistent and excessive fear of being judged or scrutinized in social situations (1,2). Typically emerging in childhood or adolescence and affecting roughly one in eight adults, the condition is associated with substantial impairment in daily functioning (3,4). However, beyond its threat-related features, the neurocognitive mechanisms shaping how social information is processed in social anxiety remain incompletely understood (5). Here, we focus on two candidate socio-cognitive processes implicated in the condition: Theory of Mind (ToM) and broader interpretive biases.

Theory of Mind, or mentalizing, refers to the capacity to infer others' unobservable thoughts and emotions, supporting the interpretation and prediction of social behavior (6,7). In social anxiety, ToM difficulties have been proposed to increase uncertainty about how one is perceived by others, thereby amplifying fears of negative evaluation (8–10). Consistent with this account, socially anxious individuals show reduced performance on ToM tasks such as the Reading the Mind in the Eyes Test and the Movie Assessment of Social Cognition (11–14), alongside altered activity and connectivity in canonical ToM regions including medial prefrontal cortex, precuneus, temporoparietal junction, and superior temporal sulcus (15–18). Other findings, however, point to exaggerated interpretation of others' emotional expressions and social signals (10,19,20), raising questions about whether ToM is underutilized, overengaged, or otherwise dysregulated in social anxiety. Notably, much of this evidence derives from explicitly evaluative tasks that may themselves induce anxiety, leaving it unclear whether altered mentalizing reflects a core feature of the condition or context-dependent effects of evaluation.

A complementary body of work has focused on interpretive biases, defined as systematic tendencies to assign excessive personal significance to social cues (21–24). Although models vary in emphasis, they converge on the idea that internally generated expectations of negative evaluation, often grounded in a negative self-image, disproportionately shape how external input is processed, both online and in retrospect (25–32). Neuroimaging studies support this account, showing altered engagement of precuneus and frontoparietal regions during social feedback (33), as well as increased interindividual variability in neural responses during naturalistic movie viewing (34). Rather than reflecting noise or disengagement, such variability may index more individualized, internally guided processing of shared social input (35,36). However, it remains unclear to what extent these interpretive tendencies operate independently of threat-related arousal or attention (37), and how they relate to ToM computations supporting mental state inference, as few studies have examined these dimensions within a single experimental framework.

To address this gap, we used a movie-viewing paradigm that enables concurrent assessment of scene-specific mentalizing and broader interpretive tendencies under minimally evaluative conditions. Forty-three individuals with elevated social anxiety and 43 low-anxiety controls viewed the animated short *Partly Cloudy* during functional magnetic resonance imaging (fMRI). The film depicts the development of a relationship between a stork and a cloud, conveyed entirely through nonverbal cues. Mental state events known to recruit the canonical ToM network (38–40) are embedded within a continuous stream of social information, enabling event-related analyses of ToM-relevant scenes alongside dynamic intersubject correlation (ISC) as an index of movie-driven neural alignment, reflecting the degree to which neural responses are shared versus individualized across viewers over time (41). Prior work using this paradigm in autism reported preserved ToM-related activation alongside elevated alignment in non-ToM regions across much of the film, suggesting a dissociation between scene-specific mentalizing and broader interpretive processing (42). Here, we apply this framework to examine how these processes manifest and interact in social anxiety.

To further dissociate social-cognitive effects from general arousal or attention, we measured whole-brain activity alongside heart rate and pupil diameter (43–46) and collected an unannounced post-viewing assessment of participants' movie interpretations. Based on prior evidence for negativity bias in social anxiety (32,47), we predicted greater use of negatively valenced emotion language in participants' plot summaries, but no a priori group differences in autonomic arousal or pupil-linked attentional measures given the non-evaluative nature of the paradigm. At the neural level, we expected group differences in one or more core ToM regions during mentalizing events (12), alongside increased interindividual variability, expressed as reduced alignment, in frontoparietal regions previously implicated in social anxiety (34). Finally, motivated by spatially overlapping ISC effects reported in autism using the same paradigm, we conducted a post-hoc cross-study comparison to examine convergent and divergent patterns of neural alignment across the two conditions (48).

Methods and Materials

Participants

Eighty-six adults were recruited via Radboud University's research database, social media platforms, and campus postings. Exclusion criteria included current use of psychotropic medication, severe cognitive impairment, systemic disease, or a history of neurological treatment. Social anxiety was assessed using the 24-item Liebowitz Social Anxiety Scale (LSAS) (49,50), which measures fear and avoidance across common social situations (e.g., eating in public, interacting with strangers) on a 4-point Likert scale. A validated cutoff score of 30 was used to classify participants into high (LSAS \geq 30) or low (LSAS < 30) social anxiety groups, yielding 43 individuals per group (50). Groups were matched on sex, age, and verbal and nonverbal IQ (51,52) (Table 1). Low-anxiety participants reported no current psychiatric diagnoses. MRI data were acquired from all participants ($n = 86$), heart rate recordings from 74 participants, pupillometry from 75 participants, and post-viewing questionnaires from 84 participants. All participants provided written informed consent in accordance with local ethics guidelines (Committee for Human-Subject Research, region Arnhem-Nijmegen, file No. 2019-6059) and received monetary compensation for their participation.

Table 1. Demographic characteristics.

	<i>Group</i>		<i>Group difference</i>		
	<i>High SA (n = 43)</i>	<i>Low SA (n = 43)</i>	<i>Test statistics</i>	<i>p</i>	<i>Bayes Factor</i>
Gender (female)	27 (62.79%)	25 (58.14%)	$\chi^2_{(1)} = 0.05$	$p = .825$	$BF_{01} = 3.53$
Age (years)	26.33 (5.92)	26.05 (5.28)	$t_{(84)} = 0.23$	$p = .818$	$BF_{01} = 4.34$
Verbal IQ (WAIS-III)	123.72 (14.65)	122.18 (16.23)	$t_{(84)} = -1.35$	$p = .182$	$BF_{01} = 2.01$
Nonverbal IQ (RPM) ¹	103.05 (11.07)	101.79 (10.29)	$t_{(83)} = 0.54$	$p = .588$	$BF_{01} = 3.88$
Social Anxiety (LSAS)	51.42 (18.58)	15.23 (7.57)	$t_{(55.56)} = 11.82$	$p < .001$	$BF_{10} > 100$

Values are presented as frequency (%) for categorical variables or mean (SD) for continuous variables. SA, Social Anxiety; WAIS, Wechsler Adult Intelligence Scale; RPM, Raven's Progressive Matrices; LSAS, Liebowitz Social Anxiety Scale; BF_{01} , Bayes Factor in favor of the null hypothesis of no group differences; BF_{10} , Bayes Factor in favor of the alternative hypothesis.

¹Nonverbal IQ data were missing for one low social anxiety participant.

Experimental protocol

Participants viewed the animated short *Partly Cloudy* (5:45 min), previously validated as a ToM localizer (38). The film portrays an evolving relationship between a stork and a cloud across episodes of cooperation, rejection, injury, and reconciliation. Following prior work, three event types were annotated (Figure 1): *Mental* (44 seconds across 4 scenes), *Pain* (26 seconds across 7 scenes), and *Control* (24 seconds across 3 scenes). Mental scenes were selected for content likely to elicit inferences about characters' mental states (e.g., distress, hope, betrayal). Pain scenes depicted physical discomfort (e.g., electric shocks, animal attacks), controlling for affective salience in the absence of mental state inference. Control scenes consisted of visually engaging but socially neutral content (e.g., flying birds, empty skies). These annotations formed the basis for event-related analyses contrasting Mental scenes with Pain and Control scenes to isolate mentalizing-related activity. The film was presented without interruption, enabling dynamic intersubject correlation (ISC) analyses across the full duration of viewing.

To minimize state anxiety, participants were acclimated to the MRI environment using a mock scanner and instructed simply to watch the film while minimizing head motion. Following the scan, participants completed an unannounced written plot summary. Responses were scored by independent raters using a taxonomy distinguishing mental and emotional language from other content-related categories (53). Group differences in mental state references were assessed using independent sample *t*-tests. A quasi-Poisson regression modeled the proportion of negatively valenced emotion words relative to each participant's total emotion-related word count.

<< Insert Figure 1 here >>

Data acquisition

Heart rate was recorded continuously at 5000 Hz using a finger pulse sensor connected to a BrainAmp ExG MR amplifier (BrainVision software). Pupil diameter was measured at 1000 Hz using an EyeLink 1000 Plus eye-tracker. MRI data were acquired on a Siemens 3T scanner with a 32-channel head coil. High-resolution structural images were obtained using a T1-weighted MPRAGE sequence (TR = 2200 ms, TI = 1100 ms, TE = 2.6 ms, flip angle = 11°, voxel size = 0.8 mm isotropic, acceleration factor = 2). Functional images were acquired using a multi-band multi-echo sequence (TR = 1500 ms, TEs = 13.4/34.8/56.2 ms, flip angle = 75°, voxel size = 2.5 mm isotropic, acceleration factor = 2). Head motion did not differ between groups, as assessed by mean framewise displacement ($M_{\text{high SA}} (\pm \text{SD}) = 0.16 (\pm 0.08)$, $M_{\text{low SA}} = 0.16 (\pm 0.07)$, $t_{(84)} = 0.20$, $p = .841$, $\text{BF}_{01} = 4.36$), maximum displacement ($M_{\text{high SA}} = 0.97 (\pm 1.27)$, $M_{\text{low SA}} = 1.15 (\pm 1.13)$, $t_{(84)} = 0.67$, $p = .507$, $\text{BF}_{01} = 3.65$), and total translational ($M_{\text{high SA}} = 118.48$

(± 110.45), $M_{\text{low SA}} = 107.42 (\pm 78.70)$, $t_{(84)} = -0.53$, $p = .594$, $BF_{01} = 3.92$) and rotational movement ($M_{\text{high SA}} = 2.29 (\pm 2.05)$, $M_{\text{low SA}} = 2.37 (\pm 1.91)$, $t_{(84)} = 0.18$, $p = .856$, $BF_{01} = 4.38$).

Heart rate analysis

Scanner artifacts were removed using a deconvolution filter (BrainAmpConverter toolbox), followed by band-pass filtering between 0.2 Hz and 3 Hz to isolate physiologically relevant frequencies (54). Heartbeats were detected using a 600 ms peak detection window, and inter-beat intervals were used to derive continuous heart rate time series. Autonomic arousal was assessed by comparing mean heart rate trajectories across the film and by quantifying movie-driven alignment using dynamic ISC. For each participant, ISC indexed the similarity between their time series and the leave-one-out average of the corresponding group (55), computed using a 30-second sliding window with 100 ms steps. Group differences were evaluated using cluster-based permutation testing applied to independent samples *t*-tests, with 10,000 Monte Carlo permutations and a significance threshold of $p < .05$ (56).

Pupillometry analysis

Blink- and saccade-related artifacts were removed using noise-based detection and adaptive velocity thresholding (57,58), with additional outliers excluded by visual inspection (59). Pupil time series were z-scored and corrected for luminance fluctuations using fifth-order polynomial regression in R (60), with frame-wise brightness values computed according to Rec. 709 standards (61). Event-related pupil responses were analyzed using a 3 (event type: Mental, Pain, Control) \times 2 (group: High SA, Low SA) mixed-design ANOVA on mean pupil diameter. Significant interactions were followed by independent samples *t*-tests with a Bonferroni-corrected significance threshold of $p < .017$ (0.05/3 tests). Dynamic ISC of pupil diameter was computed using the same leave-one-out sliding-window approach as for heart rate (see above).

fMRI analysis

Preprocessing was performed in SPM12. Multi-echo images were combined into a single time series, realigned using second-degree B-spline interpolation, and unwarped using field maps. Functional images were then coregistered to structural images, normalized to MNI space, and spatially smoothed with an 8 mm FWHM Gaussian kernel. First-level general linear models included regressors for Mental, Pain, and Control events, as well as the film's end credits. Motion parameters (including derivatives, squared terms, and cubic terms) and signals from white matter, cerebrospinal fluid, and non-brain voxels were included as nuisance regressors.

To assess ToM engagement, two mixed-design ANOVAs tested effects of event type (Mental > Pain; Mental > Control) and group. A conjunction-null analysis identified regions consistently engaged

across both contrasts. Statistical significance was assessed using cluster-level family-wise error correction ($p_{FWE} < .05$, voxel-wise $p < .001$), and anatomical labels were assigned using the SPM Anatomy Toolbox (62). Bayesian analyses quantified evidence for the absence of group differences in canonical ToM regions by extracting activation estimates from 8-mm spheres centered on conjunction peaks and computing Bayes Factors (BFs) (63). To control for temporal structure, an additional first-level regressor indexing linear progression through the film was included; this did not alter the results, indicating robustness to event timing.

Movie-driven neural alignment was assessed using voxelwise dynamic ISC applied to motion- and noise-corrected fMRI time series (42). ISC was computed using the same leave-one-out sliding-window approach, applied to downsampled data for computational efficiency (7.5-mm spatial; 4.5-s temporal resolution). Group differences were identified using cluster-based permutation testing. For visualization, t -values were summed across time points to generate 3D statistical maps, and spatial overlap with ToM activation was quantified as the proportion of overlapping voxels within each significant cluster.

Results

Post-viewing movie assessment

Following the film, participants summarized the plot in their own words. High- and low-social-anxiety groups did not differ in the proportion of mental state terms used ($M_{\text{high SA}} = 0.053 (\pm 0.057)$, $M_{\text{low SA}} = 0.048 (\pm 0.064)$, $t_{(82)} = 0.47$, $p = .642$, $BF_{01} = 4.00$; Figure 2A,B), nor in overall use of emotion-related language ($M_{\text{high SA}} = 0.042 (\pm 0.057)$, $M_{\text{low SA}} = 0.034 (\pm 0.038)$, $t_{(82)} = 0.73$, $p = .469$, $BF_{01} = 3.49$). Contrary to predictions, however, socially anxious individuals used a lower proportion of negatively valenced emotion words in their summaries ($M_{\text{high SA}} = 0.65 (\pm 0.31)$, $M_{\text{low SA}} = 0.85 (\pm 0.21)$, $t_{(50)} = -2.61$, $p = .012$; Figure 2C), indicating a more positive emotional framing of the narrative.

<< Insert Figure 2 here >>

Heart rate responses

Heart rate fluctuated over the course of the film but followed highly similar trajectories across groups. No group differences were observed in mean heart rate ($M_{\text{high SA}} = 65.45 (\pm 8.69)$, $M_{\text{low SA}} = 64.11 (\pm 9.90)$, $t_{(72)} = 0.62$, $p = .536$, $BF_{01} = 3.52$; Figure S1A), nor in movie-driven alignment of heart rate dynamics as indexed by dynamic and mean ISC ($M_{\text{high SA}} = 0.22 (\pm 0.13)$, $M_{\text{low SA}} = 0.21 (\pm 0.10)$, $t_{(72)} = 0.23$, $p = .81$, $BF_{01} = 4.06$; Figure S1B). These results indicate comparable autonomic arousal across groups throughout the movie.

Pupil responses

Pupil diameter varied robustly as a function of event type but showed highly similar response profiles across groups. Dilation was greatest during *Pain* events, followed by *Mental* and *Control* events ($F_{(2,146)} = 33.93$, $p < .001$, $BF_{10} > 100$; *Pain* > *Mental*: $t_{(74)} = 2.72$, $p = .008$, $d = 0.31$; *Mental* > *Control*: $t_{(74)} = 7.91$, $p < .001$, $d = 0.91$; Figure 2D). No main effects or interactions involving group were observed (group: $F_{(1,73)} = 0.10$, $p = .750$, $BF_{01} = 6.22$; group \times event: $F_{(2,146)} = 0.48$, $p = .619$, $BF_{01} = 6.66$). Dynamic ISC analyses revealed pronounced fluctuations in pupil alignment across viewers but no group differences at any time point or in mean ISC ($M_{\text{high SA}} = 0.58 (\pm 0.24)$, $M_{\text{low SA}} = 0.61 (\pm 0.19)$, $t_{(73)} = 1.09$, $p = .280$, $BF_{01} = 2.51$; Figure S2). Together, these findings indicate comparable event-related and continuous attentional engagement across groups.

Theory of Mind engagement

Whole-brain contrasts revealed robust engagement of the canonical ToM network during *Mental* events relative to both *Pain* and *Control* scenes (Figure 3A, Table S1). Across participants, this network included bilateral temporoparietal junction (rTPJ: $xyz_{\text{MNI}} = [46, -52, 22]$), left TPJ ($[-44, -60, 22]$), precuneus ($[0, -54, 42]$), and medial prefrontal cortex (mPFC; $[6, 50, 32]$), all significant at $p_{\text{FWE}} < .001$.

Within this otherwise preserved network, socially anxious participants showed selectively reduced activation in the left posterior superior temporal sulcus (pSTS) across both mentalizing contrasts (*Mental > Control*: [-44, -62, 18], $t = 4.41$, $p_{FWE} = .003$, $BF_{10} > 100$; *Mental > Pain*: [-50 -64, 14], $t = 4.58$, $p_{FWE} = .009$; $BF_{10} > 100$; Figure 3B). Bayesian analyses provided evidence against group differences in other core ToM regions (rTPJ: $BF_{01} = 4.08$; precuneus: $BF_{01} = 1.22$; mPFC: $BF_{01} = 1.42$), indicating largely preserved ToM engagement outside this focal region.

<< Insert Figure 3 here >>

Movie-driven neural alignment

Dynamic ISC analyses identified three spatiotemporal clusters showing group differences in neural alignment during movie viewing (Figure 4). Two clusters showed reduced alignment in socially anxious participants. Cluster 1 spanned the first two-thirds of the film ($cluster\ stat = 8585$, $p = .003$) and included superior parietal ([24, -52, 74], $t_{max} = 5.54$, $t_{sum} = 105.26$) and bilateral temporal regions (left: [-60, -10, -22], $t_{max} = 4.06$, $t_{sum} = 94.37$; right: [66, -40, -10], $t_{max} = 5.04$, $t_{sum} = 95.64$). Cluster 2 emerged during the final third ($cluster\ stat = 4497$, $p = .02$) and encompassed supplementary motor ([0, 8, 62], $t_{max} = 4.39$, $t_{sum} = 56.54$), medial prefrontal ([-6, 50, 8], $t_{max} = 6.77$, $t_{sum} = 56.01$) and precentral ([-54, 2, 20], $t_{max} = 3.84$, $t_{sum} = 51.21$) regions. In contrast, Cluster 3 showed increased alignment in socially anxious participants across much of the film ($cluster\ stat = -8478$, $p = .002$), involving occipital ([42, -70, 8], $t_{min} = -3.96$, $t_{sum} = -110.16$), inferior frontal ([30, 32, -10], $t_{min} = -4.91$, $t_{sum} = -108.48$), and superior temporal ([48, 2, -10], $t_{min} = -3.81$, $t_{sum} = -97.43$) cortices. Spatial overlap between ISC clusters and the ToM network was limited, with no cluster exceeding 20% overlap (Figure 5). Post-hoc analyses revealed no association between ISC values within these clusters and individual pSTS activation among socially anxious participants (all $p > .20$), indicating that alignment effects were not directly coupled to ToM-related activation.

<< Insert Figure 4 here >>

<< Insert Figure 5 here >>

Comparison with autism

Because Cluster 1 closely overlapped with a cluster previously reported in individuals with Autism Spectrum Condition (ASC) using the same paradigm (42), we conducted an exploratory post-hoc cross-study comparison. The ASC and high-social-anxiety groups were demographically matched and comparable in social anxiety but differed in autism traits (Table S2). Spatial overlap between clusters was substantial (~60%), spanning supramarginal, inferior temporal, anterior midcingulate, and precuneus regions (Figure 6A). Despite this anatomical convergence, alignment patterns diverged:

socially anxious participants showed reduced alignment, whereas individuals with ASC showed increased alignment across much of the film (Figure 6B). These findings suggest distinct modes of social information processing, characterized by more idiosyncratic responses in social anxiety and more stereotyped responses in autism, relative to controls.

<< Insert Figure 6 here >>

Journal Pre-proof

Discussion

By leveraging a movie paradigm that reliably engages the ToM network during well-defined scenes, this study provides converging neural evidence relevant to two socio-cognitive accounts of social anxiety: altered mentalizing and broader interpretive biases. Socially anxious individuals showed reduced activation in the left posterior superior temporal sulcus (pSTS), a core subregion of the ToM network, despite otherwise preserved engagement of canonical mentalizing regions. These focal ToM-related differences co-occurred with more widespread alterations in movie-driven neural alignment outside the ToM network, revealed by intersubject correlation analyses and observed in the absence of group differences in physiological arousal or attentional engagement. Together, these findings support an integrative neurocognitive account in which social anxiety is characterized by both circumscribed alterations in ToM-related computations and broader shifts in the neural organization and interpretation of social information, even under minimally evaluative conditions.

Our results extend prior behavioral evidence of reduced performance on emotion recognition and mental state attribution tasks in social anxiety (12) by identifying corresponding neural differences during implicit movie viewing. Low-anxiety participants showed selective recruitment of the left pSTS during scenes requiring inferences about characters' thoughts and emotions, whereas socially anxious individuals exhibited attenuated scene-specific modulation despite preserved engagement of the broader ToM network. This pattern points to a focal alteration in mentalizing rather than a global impairment and demonstrates that ToM-related differences in social anxiety can emerge even in the absence of explicit task demands or evaluative pressure (10).

The pSTS plays a central role in integrating perceptual social cues with internal models of others' intentions (64,65). Through its connectivity with motion- and face-sensitive regions, including area MT and the fusiform face area, it supports the interpretation of dynamic social signals such as gaze, biological motion, and facial expressions (66–72). The left pSTS, in particular, has been implicated in decoding others' emotional states, with variation in its structure and function associated with differences in social perception across contexts, including loneliness and social anxiety (73–76). Reduced engagement of this region during movie viewing may therefore reflect differences in how socially salient affective information is integrated during ongoing perception (77), consistent with group differences in emotional framing observed in post-viewing descriptions.

Beyond these focal ToM effects, dynamic ISC analyses revealed region-specific differences in neural alignment indicative of broader alterations in social information processing in social anxiety. Increased alignment in early sensory regions, including occipital and superior temporal cortex, suggests more

stimulus-locked and uniform perceptual processing. In contrast, reduced alignment in frontoparietal and temporal association cortices indicates greater divergence at higher levels of processing, consistent with more individualized, internally guided interpretations of social content (35,36,78,79). Framed within hierarchical models of cortical organization (80), this pattern points to a redistribution of social information processing across the cortical hierarchy, with increasingly idiosyncratic integration at higher-order association levels. Crucially, these effects occurred despite comparable heart rate and pupil dynamics across groups, arguing against differences in arousal or overall attentional engagement as their primary source. Instead, the observed redistribution may reflect an integrative bias that, in contexts involving expectations of negative evaluation, could influence how social information is weighted toward threat-relevant interpretations.

An additional insight emerged from the substantial spatial overlap between one alignment cluster identified here and a network previously reported in autistic individuals using the same paradigm (42). Despite this anatomical convergence, alignment effects diverged in direction: reduced in social anxiety and increased in autism. This dissociation indicates that similar neural systems can support qualitatively distinct modes of social information processing across conditions. Whereas increased alignment in autism may reflect more stereotyped or stimulus-driven processing, reduced alignment in social anxiety may index more idiosyncratic, internally guided interpretations. Notably, the two groups were comparable in social anxiety severity, suggesting that similar behavioral phenotypes can arise from distinct neural mechanisms. Whereas social anxiety in autism may emerge secondarily from chronic social difficulties (48), social anxiety disorder may be more closely linked to early-emerging interpretive tendencies that prioritize evaluative appraisal over perceptual evidence (81,82). Together, these findings highlight movie-driven neural alignment as a transdiagnostic marker of how social information is processed, rather than a simple index of anxiety severity.

Contrary to predictions, socially anxious participants used fewer negatively valenced emotion words in their post-viewing summaries. This effect was not accompanied by differences in overall mental state or emotion language, suggesting a selective shift in emotional framing rather than reduced engagement with the narrative. Several explanations are possible. Because the assessment was offline and explicit, socially anxious individuals may have engaged in impression management or self-monitoring, downplaying negative content (83,84). Alternatively, the film's positive resolution may have disproportionately shaped retrospective evaluations. It is also possible that negativity biases in social anxiety are more pronounced in self-referential or interactive contexts than in evaluations of fictional social content (33,47,85,86). Future studies incorporating real-time affective or experiential measures could help adjudicate between these possibilities.

We found no direct relationship between pSTS activation during mentalizing events and broader patterns of movie-driven neural alignment, suggesting that these measures index partially independent dimensions of socio-cognitive processing. Although some theoretical models posit a trade-off between internal monitoring and mentalizing resources (25,26), our data do not support a simple antagonistic relationship. Instead, this dissociation is consistent with an integrative neurocognitive account in which social anxiety reflects multiple, not obligatorily coupled vulnerabilities. In this framework, elevated social anxiety may arise from biased interpretive processing in some individuals, altered weighting of mental state inferences in others, or the co-occurrence of both within the same individual (10). These neurocognitive contributions may operate at different levels of the processing hierarchy and be differentially expressed depending on contextual and evaluative demands.

Several limitations warrant consideration. First, although the movie-viewing paradigm enabled the study of implicit mentalizing during continuous social perception, emotional and mental state content were partially correlated. While the Pain condition controlled for affective salience, it did not eliminate emotional mental state processing, such that the Mental condition did not isolate emotion-free mental state inference. Second, behavioral assessment was limited to an unannounced post-viewing summary, capturing retrospective interpretation rather than online processing. As a result, the present data cannot directly link moment-to-moment neural dynamics to unfolding interpretive processes during the narrative (87). Finally, participants were grouped using validated self-report measures indexing continuous variation in social anxiety rather than categorical diagnoses, and comorbid conditions were not systematically assessed. While this dimensional approach enhances sensitivity to individual differences in symptom severity (5,88), it limits conclusions about diagnostic specificity and clinical boundaries.

Conclusions

This study identifies converging neurocognitive signatures of social anxiety during implicit movie viewing, linking focal alterations in ToM-related processing with broader shifts in movie-driven neural alignment across the cortical hierarchy. Reduced alignment in higher-order association regions, observed independently of arousal or attentional differences, suggests more individualized modes of social information processing in social anxiety, while opposite patterns in autism underscore the transdiagnostic potential of movie-driven neural alignment.

Acknowledgements

The authors would like to thank all participants for their participation in this study. This work was supported by the NWO Gravitation Grant (No. 024.001.006) awarded to the Language in Interaction Consortium. SBJK is funded by a Starting Grant from the European Research Council (ERC, No. 101163381 - *SocialDynamics*).

Disclosures

The authors report no biomedical financial interests or potential conflicts of interest.

Supplement Description:

Figures S1-S2 and Tables S1-S2

Journal Pre-proof

References

1. Marks IM, Gelder MG. Different ages of onset in varieties of phobia. *Am J Psychiatry*. 1966;123(2):218–21.
2. American Psychiatric Association. *DSM-III: Diagnostic and Statistical Manual of Mental Disorders, 3rd Edition*. The American Psychiatric Association; 1980. 494 p.
3. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005 June;62(6):593–602.
4. Lim MH, Rodebaugh TL, Zyphur MJ, Gleeson JFM. Loneliness over time: The crucial role of social anxiety. *J Abnorm Psychol*. 2016 July;125(5):620–30.
5. Akiki TJ, Jubeir J, Bertrand C, Tozzi L, Williams LM. Neural circuit basis of pathological anxiety. *Nat Rev Neurosci*. 2025 Jan;26(1):5–22.
6. Premack D, Woodruff G. Does the Chimpanzee Have a Theory of Mind? *Behav Brain Sci*. 1978;1(4):515–26.
7. Wimmer H, Perner J. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*. 1983;13(1):103–28.
8. Catalino LI, Furr RM, Bellis FA. A multilevel analysis of the self-presentation theory of social anxiety: Contextualized, dispositional, and interactive perspectives. *J Res Personal*. 2012 Aug 1;46(4):361–73.
9. Leary MR, Kowalski RM. The self-presentation model of social phobia. In: *Social phobia: Diagnosis, assessment, and treatment*. New York, NY, US: The Guilford Press; 1995. p. 94–112.
10. Hezel DM, McNally RJ. Theory of Mind Impairments in Social Anxiety Disorder. *Behav Ther*. 2014 July;45(4):530–40.
11. Alvi T, Kouros CD, Lee J, Fulford D, Tabak BA. Social anxiety is negatively associated with theory of mind and empathic accuracy. *J Abnorm Psychol*. 2020 Jan;129(1):108–13.
12. Baez S, Tangarife MA, Davila-Mejia G, Trujillo-Güiza M, Forero DA. Performance in emotion recognition and theory of mind tasks in social anxiety and generalized anxiety disorders: a systematic review and meta-analysis. *Front Psychiatry*. 2023 May 19;14:1192683.
13. Baron-Cohen S, Jolliffe T, Mortimore C, Robertson M. Another advanced test of theory of mind: evidence from very high functioning adults with autism or asperger syndrome. *J Child Psychol Psychiatry*. 1997 Oct;38(7):813–22.
14. Dziobek I, Fleck S, Kalbe E, Rogers K, Hassenstab J, Brand M, et al. Introducing MASC: a movie for the assessment of social cognition. *J Autism Dev Disord*. 2006 July;36(5):623–36.
15. Sripada CS, Angstadt M, Banks S, Nathan PJ, Liberzon I, Luan Phan K. Functional neuroimaging of mentalizing during the trust game in social anxiety disorder. *Neuroreport*. 2009 July 15;20(11):984–9.

16. Cui Q, Vanman EJ, Long Z, Pang Y, Chen Y, Wang Y, et al. Social anxiety disorder exhibit impaired networks involved in self and theory of mind processing. *Soc Cogn Affect Neurosci*. 2017 Aug 1;12(8):1284–95.
17. Tei S, Kauppi JP, Jankowski KF, Fujino J, Monti RP, Tohka J, et al. Brain and behavioral alterations in subjects with social anxiety dominated by empathic embarrassment. *Proc Natl Acad Sci*. 2020 Feb 25;117(8):4385–91.
18. Schurz M, Radua J, Aichhorn M, Richlan F, Perner J. Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neurosci Biobehav Rev*. 2014 May 1;42:9–34.
19. Nikolić M, van der Storm L, Colonnese C, Brummelman E, Kan KJ, Bögels S. Are Socially Anxious Children Poor or Advanced Mindreaders? *Child Dev*. 2019 July;90(4):1424–41.
20. Washburn D, Wilson G, Roes M, Rnic K, Harkness KL. Theory of mind in social anxiety disorder, depression, and comorbid conditions. *J Anxiety Disord*. 2016 Jan;37:71–7.
21. Chen J, Short M, Kemps E. Interpretation bias in social anxiety: A systematic review and meta-analysis. *J Affect Disord*. 2020 Nov 1;276:1119–30.
22. Mathews A, MacLeod C. Cognitive vulnerability to emotional disorders. *Annu Rev Clin Psychol*. 2005;1:167–95.
23. Constans JI, Penn DL, Ihen GH, Hope DA. Interpretive biases for ambiguous stimuli in social anxiety. *Behav Res Ther*. 1999 July;37(7):643–51.
24. Wong QJJ, Rapee RM. The aetiology and maintenance of social anxiety disorder: A synthesis of complimentary theoretical models and formulation of a new integrated model. *J Affect Disord*. 2016 Oct;203:84–100.
25. Rapee RM, Heimberg RG. A cognitive-behavioral model of anxiety in social phobia. *Behav Res Ther*. 1997 Aug;35(8):741–56.
26. Clark DM, Wells A. A cognitive model of social phobia. In: *Social phobia: Diagnosis, assessment, and treatment*. New York, NY, US: The Guilford Press; 1995. p. 69–93.
27. Hofmann SG. Cognitive Factors that Maintain Social Anxiety Disorder: a Comprehensive Model and its Treatment Implications. *Cogn Behav Ther*. 2007;36(4):193–209.
28. Moscovitch DA. What Is the Core Fear in Social Phobia? A New Model to Facilitate Individualized Case Conceptualization and Treatment. *Cogn Behav Pract*. 2009 May 1;16(2):123–34.
29. Amir N, Foa EB, Coles ME. Automatic activation and strategic avoidance of threat-relevant information in social phobia. *J Abnorm Psychol*. 1998 May;107(2):285–90.
30. Adamis AM, Walske S, Olatunji BO. Attention mechanisms of social anxiety in daily life: Unique effects of negative self-focused attention on post-event processing. *Behav Res Ther*. 2025 Aug;191:104759.
31. Rapee RM, Lim L. Discrepancy between self- and observer ratings of performance in social phobics. *J Abnorm Psychol*. 1992 Nov;101(4):728–31.
32. Stopa L, Clark DM. Cognitive processes in social phobia. *Behav Res Ther*. 1993 Mar;31(3):255–67.

33. Koban L, Andrews-Hanna JR, Ives L, Wager TD, Arch JJ. Brain mediators of biased social learning of self-perception in social anxiety disorder. *Transl Psychiatry*. 2023 Sept 2;13(1):292.
34. Camacho MC, Balsler DH, Furtado EJ, Rogers CE, Schwarzlose RF, Sylvester CM, et al. Higher Intersubject Variability in Neural Response to Narrative Social Stimuli Among Youth With Higher Social Anxiety. *J Am Acad Child Adolesc Psychiatry* [Internet]. 2023 Dec 7; Available from: <https://www.sciencedirect.com/science/article/pii/S0890856723022347>
35. Finn ES, Corlett PR, Chen G, Bandettini PA, Constable RT. Trait paranoia shapes inter-subject synchrony in brain activity during an ambiguous social narrative. *Nat Commun*. 2018 May 23;9(1):2043.
36. van Baar JM, Halpern DJ, FeldmanHall O. Intolerance of uncertainty modulates brain-to-brain synchrony during politically polarized perception. *Proc Natl Acad Sci*. 2021 May 18;118(20):e2022491118.
37. Henderson HA, Pine DS, Fox NA. Behavioral inhibition and developmental risk: a dual-processing perspective. *Neuropsychopharmacol Off Publ Am Coll Neuropsychopharmacol*. 2015 Jan;40(1):207–24.
38. Jacoby N, Bruneau E, Koster-Hale J, Saxe R. Localizing Pain Matrix and Theory of Mind networks with both verbal and non-verbal stimuli. *NeuroImage*. 2016 Feb 1;126:39–48.
39. Paunov AM, Blank IA, Fedorenko E. Functionally distinct language and Theory of Mind networks are synchronized at rest and during language comprehension. *J Neurophysiol*. 2019 Apr 1;121(4):1244–65.
40. Richardson H, Lisandrelli G, Riobueno-Naylor A, Saxe R. Development of the social brain from age three to twelve years. *Nat Commun*. 2018 Mar 12;9(1):1027.
41. Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R. Intersubject synchronization of cortical activity during natural vision. *Science*. 2004 Mar 12;303(5664):1634–40.
42. Mangnus M, Koch SBJ, Cai K, Romaneli MG, Hagoort P, Bašnáková J, et al. Preserved Spontaneous Mentalizing Amid Reduced Intersubject Variability in Autism During a Movie Narrative. *Biol Psychiatry Cogn Neurosci Neuroimaging* [Internet]. 2024 Oct 26 [cited 2025 June 26];0(0). Available from: [https://www.biologicalpsychiatrycnri.org/article/S2451-9022\(24\)00306-9/fulltext](https://www.biologicalpsychiatrycnri.org/article/S2451-9022(24)00306-9/fulltext)
43. Wascher CAF. Heart rate as a measure of emotional arousal in evolutionary biology. *Philos Trans R Soc B Biol Sci*. 2021 Aug 16;376(1831):20200479.
44. Beatty J. Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources. 1982;17.
45. Burlingham CS, Mirbagheri S, Heeger DJ. A unified model of the task-evoked pupil response. *Sci Adv*. 2022 Apr 22;8(16):eabi9979.
46. Bradley MM, Miccoli L, Escrig MA, Lang PJ. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*. 2008 July;45(4):602–7.
47. Hirsch C, Meynen T, Clark D. Negative self-imagery in social anxiety contaminates social interactions. *Memory*. 2004 July;12(4):496–506.

48. White SW, Oswald D, Ollendick T, Scahill L. Anxiety in children and adolescents with autism spectrum disorders. *Clin Psychol Rev.* 2009 Apr;29(3):216–29.
49. Liebowitz MR. Social phobia. *Mod Probl Pharmacopsychiatry.* 1987;22:141–73.
50. Oakman J, Van Ameringen M, Mancini C, Farvolden P. A confirmatory factor analysis of a self-report version of the Liebowitz Social Anxiety Scale. *J Clin Psychol.* 2003 Jan;59(1):149–61.
51. Raven J. The Raven Progressive Matrices: A Review of National Norming Studies and Ethnic and Socioeconomic Variation Within the United States. *J Educ Meas.* 1989;26(1):1–16.
52. Wechsler D. *WAIS-III: Administration and Scoring Manual.* Harcourt Brace & Company; 1997. 20 p.
53. Bang J, Burns J, Nadig A. Brief report: Conveying subjective experience in conversation: production of mental state terms and personal narratives in individuals with high functioning autism. *J Autism Dev Disord.* 2013 July;43(7):1732–40.
54. Avram R, Tison GH, Aschbacher K, Kuhar P, Vittinghoff E, Butzner M, et al. Real-world heart rate norms in the Health eHeart study. *NPJ Digit Med.* 2019 June 25;2:58.
55. Nastase SA, Gazzola V, Hasson U, Keysers C. Measuring shared responses across subjects using intersubject correlation. *Soc Cogn Affect Neurosci.* 2019 Aug 7;14(6):667–85.
56. Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods.* 2007 Aug 15;164(1):177–90.
57. Hershman R, Henik A, Cohen N. A novel blink detection method based on pupillometry noise. *Behav Res Methods.* 2018 Feb 1;50(1):107–14.
58. Nyström M, Holmqvist K. An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behav Res Methods.* 2010 Feb 1;42(1):188–204.
59. Mathot S. Pupillometry: Psychology, Physiology, and Function. *J Cogn.* 2018 Feb 21;1(1):16.
60. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw.* 2015;67(1):1–48.
61. ITU I. Parameter values for the HDTV standards for production and international programme exchange. *Recomm ITU-R BT.* 2002;709–5.
62. Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, et al. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage.* 2005 May 1;25(4):1325–35.
63. JASP Team. JASP (Version 0.16.4)[Computer software] [Internet]. 2022. Available from: <https://jasp-stats.org/>
64. Patel GH, Sestieri C, Corbetta M. The evolution of the temporoparietal junction and posterior superior temporal sulcus. *Cortex J Devoted Study Nerv Syst Behav.* 2019 Sept;118:38–50.
65. Stolk A, Noordzij ML, Volman I, Verhagen L, Overeem S, van Elswijk G, et al. Understanding communicative actions: a repetitive TMS study. *Cortex J Devoted Study Nerv Syst Behav.* 2014

- Feb;51:25–34.
66. Dasgupta S, Tyler SC, Wicks J, Srinivasan R, Grossman ED. Network Connectivity of the Right STS in Three Social Perception Localizers. *J Cogn Neurosci*. 2017 Feb;29(2):221–34.
 67. Ferreira RA, Göbel SM, Hymers M, Ellis AW. The neural correlates of semantic richness: Evidence from an fMRI study of word learning. *Brain Lang*. 2015 Apr 1;143:69–80.
 68. Jabbi M, Kohn PD, Nash T, Ianni A, Coutlee C, Holroyd T, et al. Convergent BOLD and Beta-Band Activity in Superior Temporal Sulcus and Frontolimbic Circuitry Underpins Human Emotion Cognition. *Cereb Cortex*. 2015 July 1;25(7):1878–88.
 69. Lahnakoski JM, Glerean E, Salmi J, Jääskeläinen IP, Sams M, Hari R, et al. Naturalistic fMRI Mapping Reveals Superior Temporal Sulcus as the Hub for the Distributed Brain Network for Social Perception. *Front Hum Neurosci* [Internet]. 2012 Aug 13 [cited 2024 July 7];6. Available from: <https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2012.00233/full>
 70. Möttönen R, Calvert GA, Jääskeläinen IP, Matthews PM, Thesen T, Tuomainen J, et al. Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *NeuroImage*. 2006 Apr 1;30(2):563–9.
 71. Shultz S, van den Honert RN, Engell AD, McCarthy G. Stimulus-induced reversal of information flow through a cortical network for animacy perception. *Soc Cogn Affect Neurosci*. 2015 Jan;10(1):129–35.
 72. Allison T, Puce A, McCarthy G. Social perception from visual cues: role of the STS region. *Trends Cogn Sci*. 2000 July;4(7):267–78.
 73. Peelen MV, Atkinson AP, Vuilleumier P. Supramodal Representations of Perceived Emotions in the Human Brain. *J Neurosci*. 2010 July 28;30(30):10127–34.
 74. Samson D, Apperly IA, Chiavarino C, Humphreys GW. Left temporoparietal junction is necessary for representing someone else's belief. *Nat Neurosci*. 2004 May;7(5):499–500.
 75. Kanai R, Bahrami B, Duchaine B, Janik A, Banissy MJ, Rees G. Brain Structure Links Loneliness to Social Perception. *Curr Biol*. 2012 Oct 23;22(20):1975–9.
 76. Bas-Hoogendam JM, van Steenbergen H, Tissier RLM, van der Wee NJA, Westenberg PM. Altered Neurobiological Processing of Unintentional Social Norm Violations: A Multiplex, Multigenerational Functional Magnetic Resonance Imaging Study on Social Anxiety Endophenotypes. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2020 Oct;5(10):981–90.
 77. Davey J, Thompson HE, Hallam G, Karapanagiotidis T, Murphy C, De Caso I, et al. Exploring the role of the posterior middle temporal gyrus in semantic cognition: Integration of anterior temporal lobe with executive processes. *Neuroimage*. 2016 Aug 15;137:165–77.
 78. Lorberbaum JP, Kose S, Johnson MR, Arana GW, Sullivan LK, Hamner MB, et al. Neural correlates of speech anticipatory anxiety in generalized social phobia. *Neuroreport*. 2004 Dec 22;15(18):2701–5.
 79. Jääskeläinen IP, Sams M, Glerean E, Ahveninen J. Movies and narratives as naturalistic stimuli in neuroimaging. *NeuroImage*. 2021 Jan 1;224:117445.

80. Mesulam MM. From sensation to cognition. *Brain*. 1998 June 1;121(6):1013–52.
81. Ni W, Lu H, Wang Q, Song C, Yi L. Vigilance or avoidance: How do autistic traits and social anxiety modulate attention to the eyes? *Front Neurosci* [Internet]. 2023 Jan 11 [cited 2026 Jan 4];16. Available from:
<https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2022.1081769/full>
82. Kleberg JL, Högström J, Nord M, Bölte S, Serlachius E, Falck-Ytter T. Autistic Traits and Symptoms of Social Anxiety are Differentially Related to Attention to Others' Eyes in Social Anxiety Disorder. *J Autism Dev Disord*. 2017 Dec;47(12):3814–21.
83. Prieto-Fidalgo Á, Calvete E. The role of interpretation biases and safety behaviours in social anxiety: an intensive longitudinal study. *Behav Cogn Psychother*. 2024 Jan;52(1):49–64.
84. Wells A, Clark DM, Salkovskis P, Ludgate J, Hackmann A, Gelder M. Social phobia: The role of in-situation safety behaviors in maintaining anxiety and negative beliefs. *Behav Ther*. 1995 Dec 1;26(1):153–61.
85. Ballespí S, Vives J, Sharp C, Tobar A, Barrantes-Vidal N. Hypermentalizing in Social Anxiety: Evidence for a Context-Dependent Relationship. *Front Psychol* [Internet]. 2019 July 9 [cited 2024 July 8];10. Available from:
<https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2019.01501/full>
86. Yoon HJ, Seo EH, Kim JJ, Choo IH. Neural Correlates of Self-referential Processing and Their Clinical Implications in Social Anxiety Disorder. *Clin Psychopharmacol Neurosci Off Sci J Korean Coll Neuropsychopharmacol*. 2019 Feb 28;17(1):12–24.
87. Sonkusare S, Breakspear M, Guo C. Naturalistic Stimuli in Neuroscience: Critically Acclaimed. *Trends Cogn Sci*. 2019 Aug;23(8):699–714.
88. Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med*. 2013 May 14;11:126.

Figure legends

Figure 1. Experimental stimuli and multimodal recordings. Socially anxious and low-anxiety participants viewed a 6-minute animated film depicting the evolving friendship between a stork and a cloud, while brain activity was continuously recorded alongside heart rate and pupil size. The film, previously validated as a Theory of Mind (ToM) localizer, featured three annotated event types: *Mental* (eliciting inferences about characters' thoughts and emotions), *Pain* (depicting physical discomfort), and *Control* (neutral scenes without foreground characters). Event images were AI-generated (Bing Copilot) to approximate original scenes.

Figure 2. Language use and pupil responses. (A) Word clouds depicting relative frequency of mental state terms in post-viewing summaries for socially anxious and low-anxiety groups. (B) Proportion of mental state terms did not differ between groups. (C) Socially anxious participants used a lower proportion of negatively valenced emotion words. (D) Pupil diameter differed by event type, with no group effects. Violin plots show full distributions; boxplots indicate means (black dots), medians, and interquartile ranges. SA, social anxiety; $**p < .001$, $*p < .05$.

Figure 3. Theory of Mind engagement. (A) Canonical ToM network engaged during Mental scenes (Mental > Control in blue; Mental > Pain in red). (B) Reduced activation in left pSTS in socially anxious participants across both contrasts. (C) No group differences in other ToM regions, with Bayes Factors (BF_{01}) supporting the null hypothesis. Contrast estimates reflect mean activation averaged across contrasts. a.u., arbitrary units; $*p < .05$.

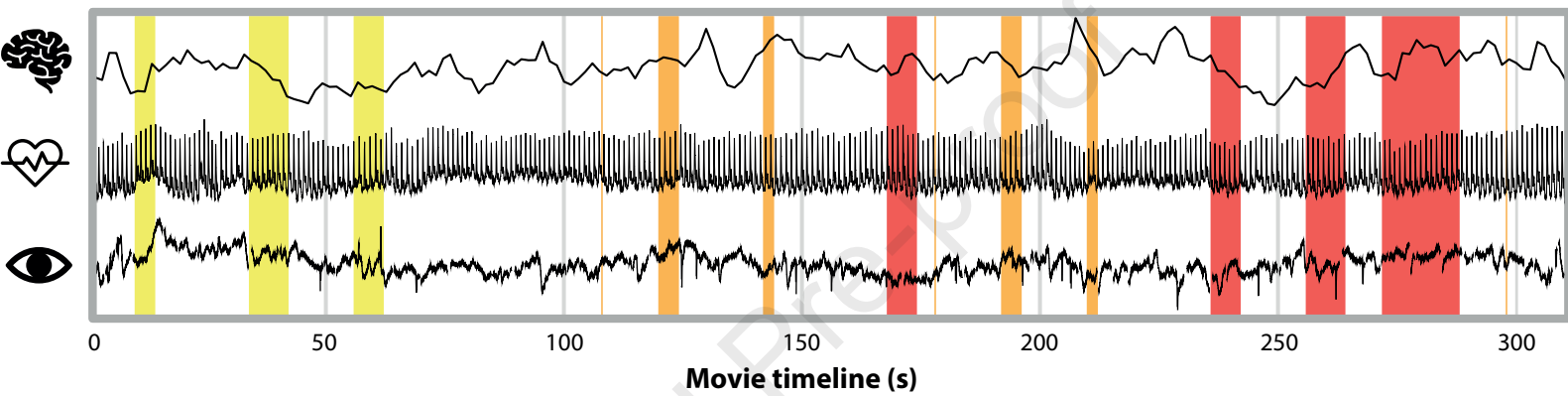
Figure 4. Movie-driven neural alignment. Dynamic ISC identified three spatiotemporal clusters with group differences in neural alignment. Socially anxious participants showed reduced alignment in Clusters 1 and 2 and increased alignment in Cluster 3. Solid lines indicate time intervals with group differences. Pink bands indicate Mental events.

Figure 5. Limited overlap between alignment and ToM activation. ISC clusters showed minimal spatial overlap with ToM regions (outlined), with no cluster exceeding 20%. For visualization, clusters are displayed at a cumulative t -value threshold of 20 or higher. LH, left hemisphere; RH, right hemisphere.

Figure 6. Divergent alignment patterns in social anxiety and autism. (A) Cluster 1 overlapped substantially with a previously reported autism-related cluster. (B) Despite anatomical overlap,

alignment was reduced in social anxiety but increased in autism, relative to controls. ASC, autism spectrum condition.

Journal Pre-proof



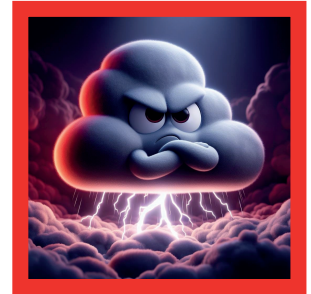
Control events

no specific character-related events



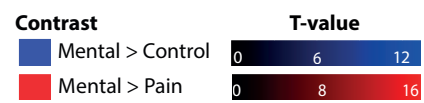
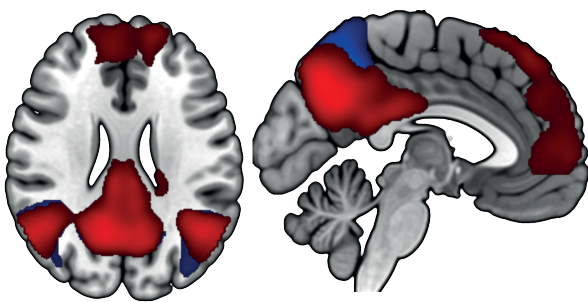
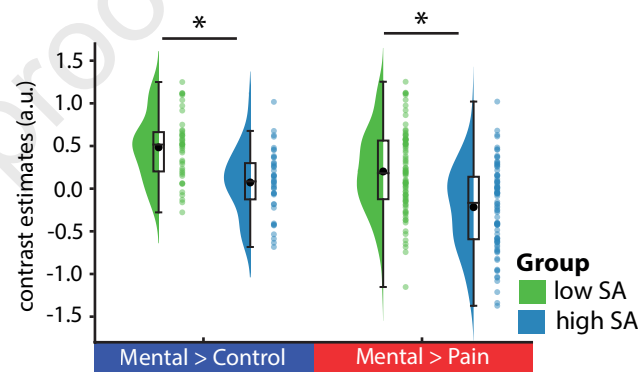
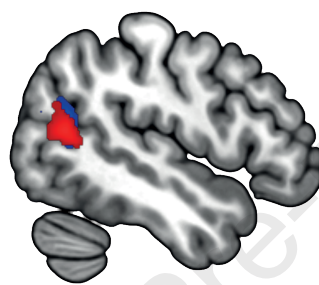
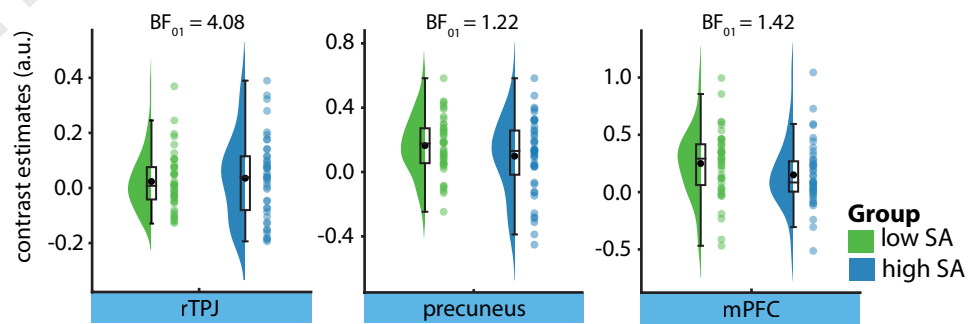
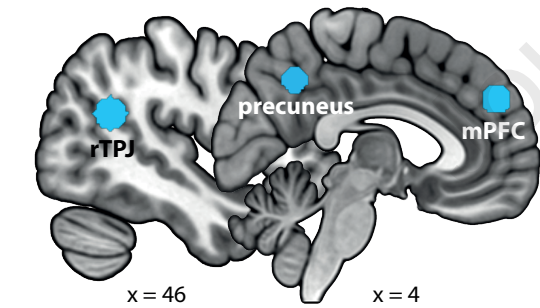
Pain events

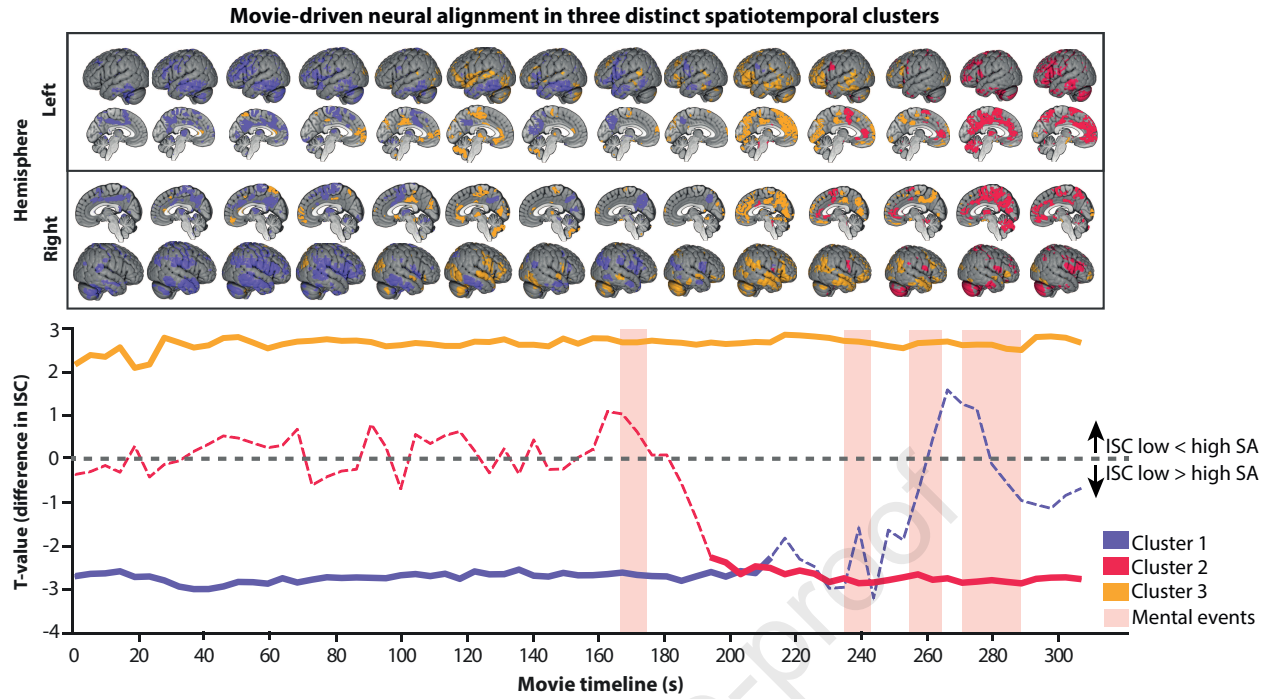
character in the movie is undergoing a physically painful event

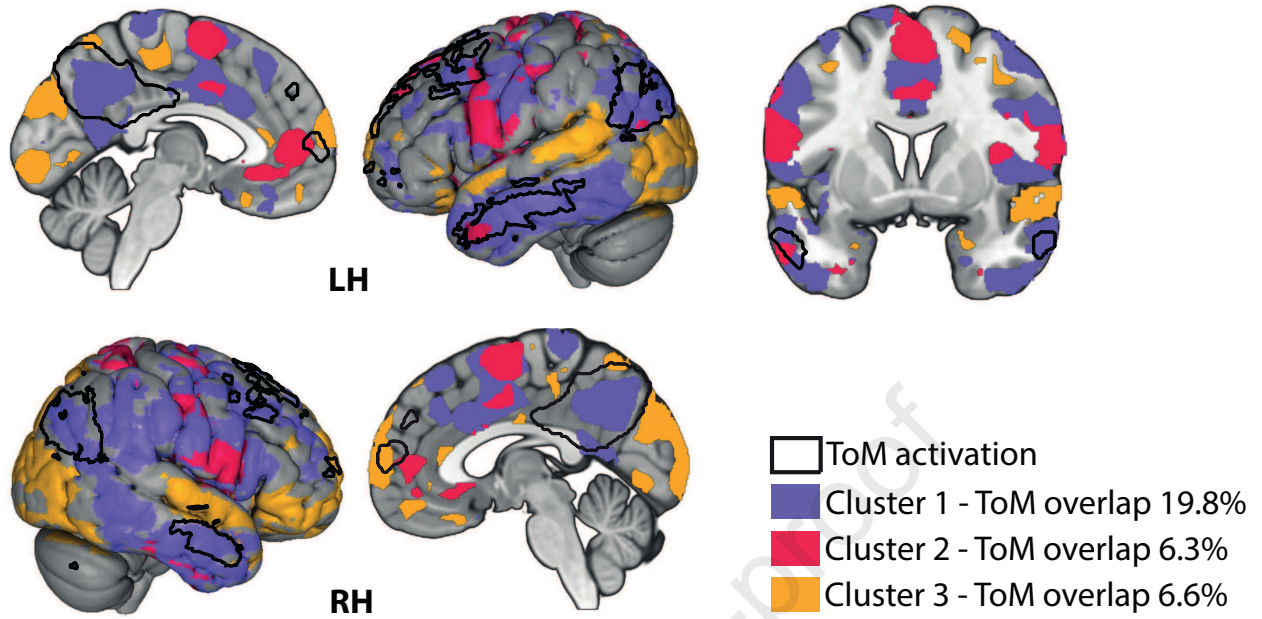


Mental events

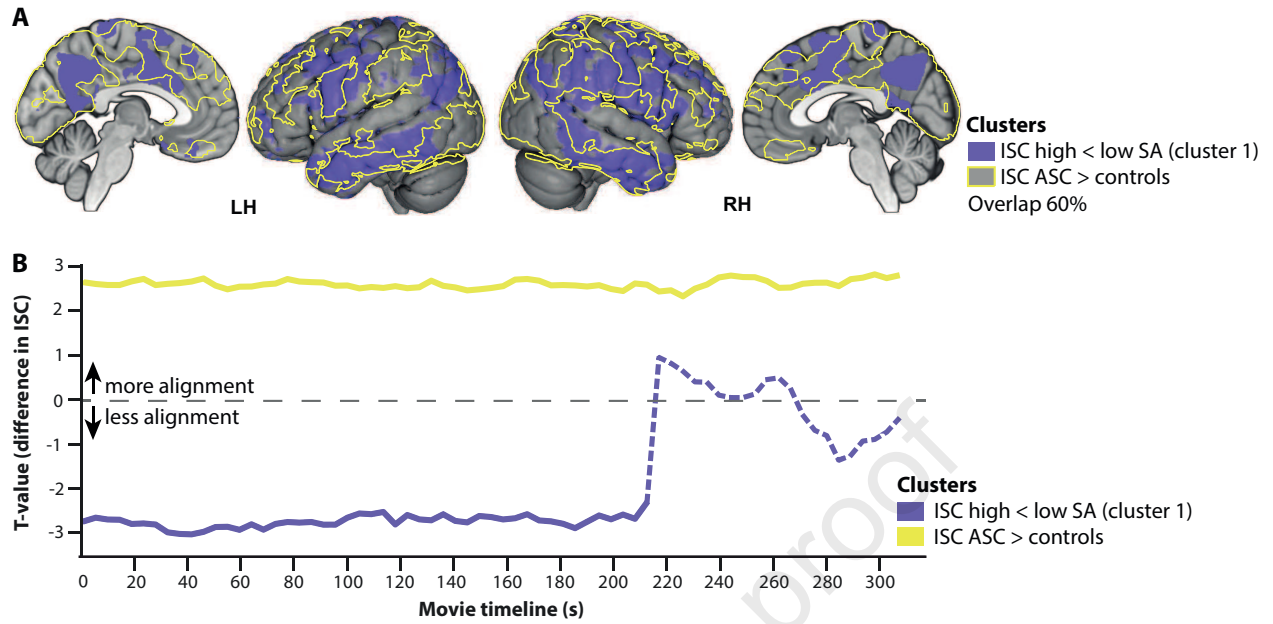
viewer is led to think about the character's thoughts and emotions

A Whole-brain ToM activation**B Reduced pSTS activation in social anxiety****C Activation across ToM regions**



Overlap between alignment and Theory of Mind activation

Divergent alignment patterns in social anxiety and autism



Plain Language Summary

Social anxiety is marked by fear of being judged by others, but it remains unclear how it shapes the interpretation of social situations. We measured brain activity, heart rate, and eye responses while people with higher or lower social anxiety watched a short, animated film depicting social interactions. Individuals with elevated social anxiety showed subtle changes in brain activity related to understanding others, along with more individualized brain responses during the movie, even without signs of increased arousal or attention. These findings suggest that social anxiety alters how social information is processed, even in situations that are not overtly threatening.

Journal Pre-proof