

Neural signatures of social inferences predict the number of real-life social contacts and autism severity

Received: 26 September 2019

Accepted: 12 July 2023

Published online: 20 July 2023

 Check for updates

Anita Tusche^{1,2}✉, Robert P. Spunt¹, Lynn K. Paul¹, Julian M. Tyszka¹ & Ralph Adolphs^{1,3}

We regularly infer other people's thoughts and feelings from observing their actions, but how this ability contributes to successful social behavior and interactions remains unknown. We show that neural activation patterns during social inferences obtained in the laboratory predict the number of social contacts in the real world, as measured by the social network index, in three neurotypical samples (total $n = 126$) and one sample of autistic adults ($n = 23$). We also show that brain patterns during social inference generalize across individuals in these groups. Cross-validated associations between brain activations and social inference localize selectively to the right posterior superior temporal sulcus and were specific for social, but not nonsocial, inference. Activation within this same brain region also predicts autism-like trait scores from questionnaires and autism symptom severity. Thus, neural activations produced while thinking about other people's mental states predict variance in multiple indices of social functioning in the real world.

Although it is a platitude that we are a social species, the number of—and participation in—social relationships differs dramatically across individuals. This variance in people's social ties and network characteristics can impact health and well-being^{1–4}. The possibility that some of the variability in people's social networks might have a neural basis has received considerable attention. A popular evolutionary account, the social brain hypothesis, proposed that the large size of the primate neocortex evolved in order to sustain the demands of increasingly complex social interactions and larger social groups^{5,6}. While this hypothesis is debated^{7–9}, it has motivated a number of neuroimaging studies that attempt to link individual differences in social network characteristics to variance in brain structure, specifically in brain areas implicated in processing social information and behaviors^{10,11}. For example, online social network size on social media has been shown to correlate with variability in gray matter density in the right superior temporal sulcus (STS), a region implicated in social perception and processing biological motion¹² (but see ref. 13). Moreover, the size and complexity of social networks have been

reported to correlate with the volume of the amygdala, a structure associated with processing facial expressions^{12,14,15} (also see ref. 16 for lesion evidence). Evidence for the link between variance in social network size and the brain's gray matter volume spans primate species^{5,17–19}.

A smaller number of neuroimaging studies in humans and non-human primates have started to examine how these findings based on brain structure might also extend to brain function, specifically in brain regions associated with social cognition and social information processing. For instance, initial correlational evidence linked variability in people's social network characteristics to brain activation in the amygdala or posterior superior temporal sulcus (pSTS) in response to social stimuli such as biological motion²⁰ or social working memory^{14,21}. Similarly, measures of functional connectivity between brain regions have been correlated with social network variability, both during social exclusion²² and task-free rest periods^{23–26}, and social network size of active interactions on social media (Twitter)²⁷.

¹Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125, USA. ²Department of Psychology, Queen's University, Kingston, Ontario K7L 3N6, Canada. ³Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA.

✉ e-mail: anita.tusche@gmail.com

Previous associations of variance in brain structure and function with social network characteristics motivate the hypothesis that neural correlates of people’s abilities to process social information could serve as predictive markers for one particular social network metric: the commonly used social network index (SNI)^{4,10,15,26,28}, which captures the number of social contacts. However, prior work leaves open critical limitations with respect to generalizability and specificity and with respect to delineating a possible cognitive process that might explain the association found. First, all prior work has been exclusively based on conventional correlations; no prior study has used a predictive framework with cross-validation, leaving out-of-sample generalizability unclear. Here we address this issue using both cross-validation and two replication samples of neurotypical individuals. Second, no prior studies have further tested generalizability by incorporating participant samples whose social cognition and social behavior are outside the neurotypical range. Here we address this issue by testing a fourth group with autism. By using a predictive model trained on data

from neurotypical individuals and tested on data from the autism group, we provide the most direct test of generalization to populations that are neurodiverse. Third, prior studies have typically been unable to link social network size to a well-isolated, specific cognitive process. Here we used a well-validated neuroimaging task that isolates social inferences, both to define brain regions of interest (ROIs) as the basis for all subsequent analyses, and for the multivariate activation patterns that served as input features for our prediction of individuals’ number of social contacts. To probe the (social) specificity of our findings, we further incorporated a control task using nonsocial (factual) inferences. These features of our approach provide the most compelling and specific evidence to date that social network characteristics like the number of social contacts can be predicted from a specific neural marker: regional patterns of brain activity produced when people think about others’ mental states.

The ability to infer other people’s mental states is a pervasive component of social cognition: we attribute others’ beliefs, intentions,

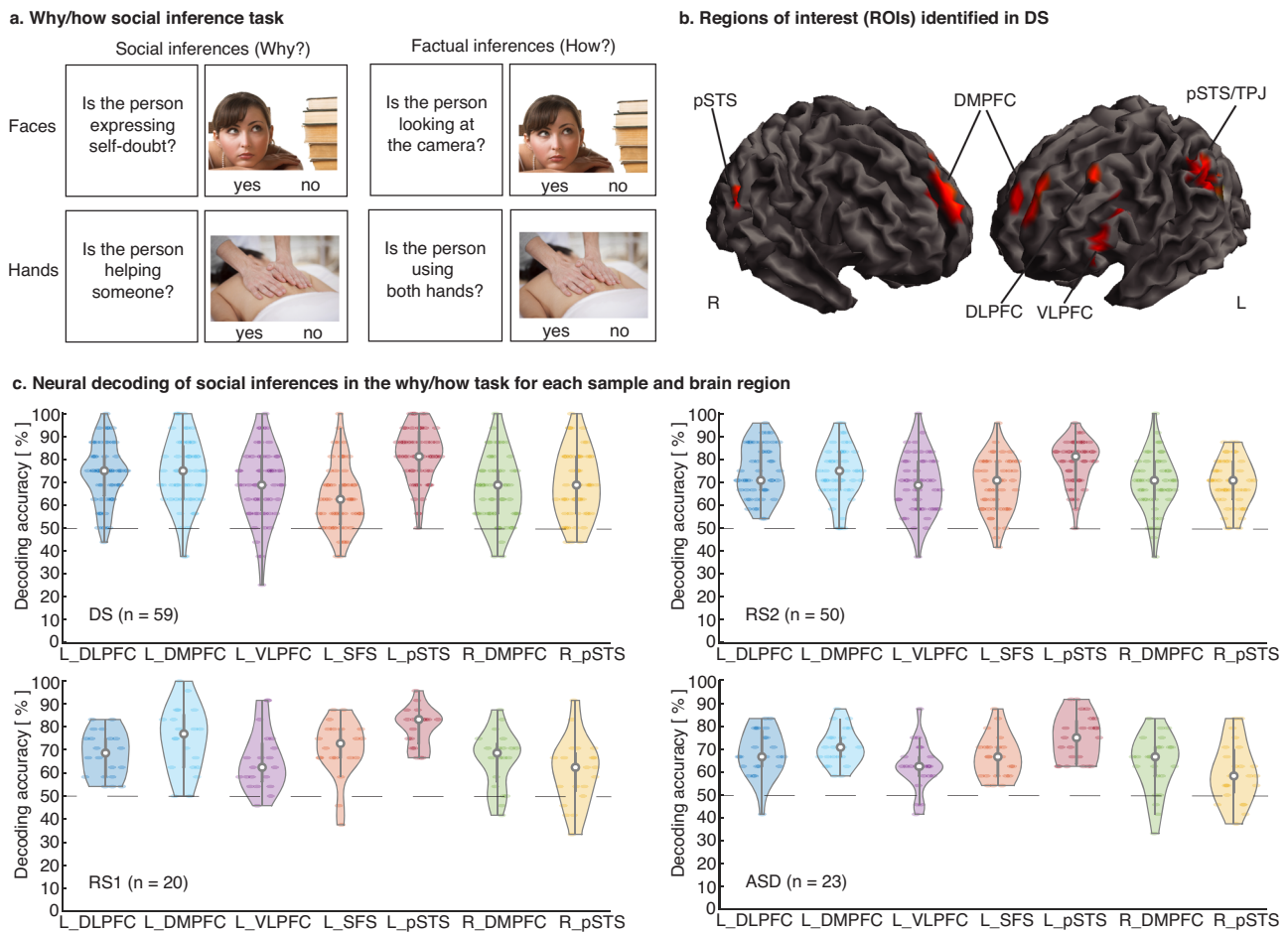


Fig. 1 | Measuring social inference processing in the brain. a Why/how social inference task (fMRI). In each task block, participants respond to yes/no questions that require either social inferences about other people’s internal states (why? high level of inference; left column) or factual inferences (how? low level of inference; right column). The task uses two stimulus sets (emotional facial expressions, top row; intentional hand actions, bottom row)⁴⁸. An additional condition using non-social stimuli to elicit why/how inferences is not displayed (implemented for the replication samples 1 and 2 (RS1, RS2) and the autism spectrum disorder (ASD) group). **b** Brain regions from which social inference could be decoded in the neurotypical discovery sample (DS; why vs. how; $p < 0.05$, FWE whole-brain corrected at the voxel level; simple t test of whole-brain decoding accuracy maps at the group level implemented in SPM12; see Table 2 for details and full names of abbreviated brain areas). These clusters served as regions of interest (ROIs) for the neural prediction of real-world social behavior in all four participant samples.

c Activation patterns in the ROIs (identified in the DS group) also decoded participants’ engagement in social inferences (why vs. how) in RS1, RS2, and ASD groups, verifying the ROIs selected based on results in DS. For each ROI and participant sample, decoding accuracies are shown for each individual (circles within violin plots; n biologically independent samples: DS = 59, RS1 = 20, RS2 = 50, ASD = 23). Dotted lines illustrate the chance level of the prediction (50%). Central white marks of boxplots (gray bars within violin plots) indicate ROI-based median accuracies, which were well above chance in each ROI and sample (all p ’s < 0.05 , permutation tests, FDR-corrected). Edges of boxplots indicate the 25th–75th percentiles and whiskers of boxplots illustrate minima and maxima. Decoding accuracies of social inferences were comparable across ASD and the matched RS1 groups (two-sample t tests, p ’s > 0.52 , FDR corrected). Source data are provided as a Source Data file, for exact p values see Table S8.

emotions, or desires all the time, in the real world, as in films and novels. This social inference ability is sometimes broadly referred to as “theory of mind” or “mentalizing”^{29–31}. The capacity to infer other people’s minds is a fundamental component of social intelligence³², is thought to be required for successful social interactions³³, contributes to prosocial behaviors^{34,35}, and has been linked to variance in social network characteristics^{11,22,36}. Typically, this social ability develops between 3–5 years of age^{37,38}, for a review, see ref. 30, but is notably delayed in children with autism spectrum disorder (ASD)³⁹. Moreover, deficits in the ability to make social inferences have been suggested to contribute to the diagnostic difficulties in communication and social interaction in ASD (for a review, see ref. 39). Further evidence that the ability to make social inferences is related to the capacity for successful social interactions comes from behavioral findings in typically developed adults: variability in people’s ability to make social inferences covaried with the size of their support clique, an index of an individual’s ability to successfully build and maintain a social network⁴⁰.

As reviewed above, there are thus associations between the number of social contacts and neural variability, and between the number of social contacts and social cognition. What is missing is a direct link between all three: can we predict the number of social contacts from the specific brain activations produced while participants engage in a specific social inference task? Here we focused on social inference both because of its plausible link to social relationships and because this cognitive process is known to engage a specific set of brain structures. The brain network includes the pSTS, temporoparietal junction (TPJ), temporal poles, and medial prefrontal cortex (mPFC)^{41–46}. While the specifics can vary depending on the details of the social inference task⁴⁴, this evidence suggests that we can capture social inference processing from activation in a set of well-described brain areas. One of the best-validated tasks for social inference processing in fMRI studies is the why/how task (Fig. 1a). This task has strong psychometric properties and reliably activates the network of brain regions described above, even in single participants^{47,48}.

We used the why/how task together with questionnaire- and observation-based assessments of social behavior to address three specific questions: First, can we predict the real-world number of social contacts based on the multivariate pattern of neural activations to social inferences measured in the laboratory? Second, does this activation pattern during social inferences generalize to predicting other indices of social behavior, such as autism-like traits? Third, does this prediction generalize to people with autism? To address these questions, we applied multivariate prediction techniques^{49,50} to brain activations obtained during the why/how task^{47,48}. Going beyond conventional correlation approaches, we used a predictive framework with cross-validation to quantify whether the pattern of brain activations during social inference is associated with the number of social contacts and other indices of real-world social behavior. Finally, we also tested the specificity of the findings by comparing them to predictions obtained when using an fMRI task with an identical structure but requiring causal inferences about nonsocial events rather than social ones. We tested these questions in four participant samples: a neurotypical DS, two neurotypical replication samples (RS1, RS2), and a sample of high-functioning people with ASD (see Table 1 for details; all statistical evaluations of neural predictions used non-parametric permutation tests).

Results

Behavior

Number of social contacts: the number of social contacts in real life was assessed by the SNI (<http://www.midss.org/content/social-network-index-sni>) and varied considerably across individuals in all four participant samples: a neurotypical discovery sample (DS), two neurotypical replication samples (RS1, RS2), and a sample of high-functioning people with ASD (Table 1). For completeness, Table 1 also

reports scores of network diversity and the number of embedded networks of the SNI. While the current study focused on the number of social contacts, high positive correlations among SNI measures (Table S1) indicate that they might capture related metrics of social network characteristics. Importantly, SNI scores showed sufficient variance across participants in each sample to justify the neural prediction of individual differences (Fig. 2a).

Autism-like traits and autism symptom severity: Table 1 also displays summary scores for three additional measures of social functioning that we investigated in different participant samples: in neurotypical individuals, we measured autism-like traits with the Autism Quotient (AQ, in DS and RS1)⁵¹ and the Social Responsiveness Scale-2 (SRS-2, Adult Form, Self-Report, in RS2)⁵². Both the AQ and SRS-2 have self-report versions used to screen for ASD in adults. We report the SRS-2 for our larger replication sample RS2 as this measure has been suggested to be preferable to the AQ for assessing the Broad Autism Phenotype⁵³. Within the general population, autism-like traits have been proposed to be continuously distributed⁵². Consistent with this notion, the distribution and range of behavioral scores of social functioning (AQ, SRS-2) in all three neurotypical subject samples allowed us to justify neural decoding of individual differences (Fig. 2b). In the ASD group, we measured autism symptom severity from the calibrated severity scores of the ADOS (social affect domain [ADOS SA]; restricted, repetitive behavior domain [ADOS RRB]^{54,55}). We focus on symptom severity in the social affective domain (ADOS SA). As for the SNI scores, ADOS SA scores indicated a sufficient variance for analyses of individual differences in autism severity in this sample (Fig. 2b).

Behavior on the why/how task (fMRI): consistent with previous implementations of the why/how task^{47,48}, participants in all samples performed at (or close to) ceiling (average response accuracy [%]: DS = 94.61 ± 2.42, RS1 = 94.95 ± 2.82, RS2 = 94.80 ± 2.45, ASD = 93.04 ± 4.11; M ± SD; see Table S2 for summary scores on response accuracies, RTs, and d’s separately for each task condition and sample). Accuracy scores in the ASD group were significantly lower compared to those in the two non-matched neurotypical samples (DS: $t(80) = 2.14$, [0.11, 3.03], $p = 0.04$; RS2: $t(71) = 2.28$, [0.22, 3.30], $p = 0.03$; independent sample t tests, two-tailed, uncorrected), but not significantly different from accuracies in the matched RS1 group ($t(41) = 1.75$ [−0.30, 4.12], $p = 0.09$). As expected, accuracy scores for neurotypical community samples were comparable (DS vs. RS1: $t(77) = 0.52$, [−0.96, 1.64], $p = 0.60$; DS vs. RS2: $t(107) = 0.42$, [−0.73, 1.12], $p = 0.68$; RS1 vs. RS2: $t(68) = 0.22$, [−1.20, 1.50], $p = 0.83$). Consistently high response accuracies (Table S2) demonstrate that participants in all four participant samples gave valid task performances.

fMRI results: outline of the analysis approach

Our analysis strategy proceeded in several steps: First, we used an established why/how fMRI task^{47,48} to locate functionally delineated regions of interest (ROIs) from which social inference processing could be decoded in a discovery group of neurotypical individuals (DS, $n = 59$). Second, we examined whether neural activation patterns from these ROIs predict individual differences in people’s number of social contacts in DS. Third, using the same set of ROIs for consistency across all participant groups, we replicated this prediction of individual differences in two community samples of neurotypical individuals (replication samples 1 and 2, RS1 and RS2). Fourth, we ensured the specificity of our findings to social inferences by comparison with neural activations obtained during nonsocial causal inferences. Fifth, we investigated the generalizability of our findings to other measures of peoples’ social functioning outside of the lab: autism-like traits in neurotypical individuals (AQ scores⁵¹ in DS and RS1, and SRS-2⁵² scores in RS2). Finally, we extended our investigation to individuals diagnosed with Autism Spectrum Disorder (ASD). We examined whether activation patterns subserving social inferences in neurotypical

Table 1 | Demographic information and summary scores of social functioning

	Discovery sample (DS)	Replication sample 1 (RS1)	Replication sample 2 (RS2)	Autism spectrum disorder sample (ASD)
Sample size (total sample)	59 (60)	20 (20)	50 (55)	23 (25)
Age (M ± SD)	28.29 ± 5.21 [19,40]	28.45 ± 6.51 [21,46]	33.56 ± 7.28 [21,49]	27.52 ± 8.25 [18,48]
Sex (male/female)	33/26	13/6	31/19	17/6
IQ scores (WASI-II)				
Full-scale IQ	108.14 ± 10.14 [90,132]	111.35 ± 9.20 [97,130]	105.92 ± 7.41 [93,127]	107.70 ± 15.40 [77,133]
Verbal comprehension index	108.91 ± 11 [87,135]	110.50 ± 12.32 [85,137]	107.48 ± 6.08 [97,120]	107.43 ± 16.53 [74,135]
Perceptual reasoning index	105.14 ± 10.94 [83,128]	110.15 ± 7.92 [87,121]	102.58 ± 9.93 [84,128]	107.04 ± 13.02 [74,127]
Social network metrics				
Network size	16.68 ± 10.23 [3,57]	21.35 ± 13.88 [6,57]	13.42 ± 8.79 [0, 37]	13.96 ± 7.60 [4,31]
Network diversity	5.07 ± 1.63 [2,9]	4.94 ± 1.75 [1,9]	4.34 ± 1.66 [0, 8]	4.87 ± 1.79 [2,8]
Embedded networks	1.68 ± 1.59 [0, 7]	3.00 ± 2.29 [0, 8]	1.24 ± 1.15 [0, 4]	1.26 ± 1.14 [0, 4]
Autism quotient scores (AQ)	15.55 ± 2.73 [8,20]	13.35 ± 4.53 [6,22]	-	-
Social responsiveness scale (SRS-2)	-	-	50.44 ± 8.38 [38,67]	-
Symptom severity scores in ASD				
ADOS SA	-	-	-	7.52 ± 1.50 [5,10]
ADOS RRB	-	-	-	6.86 ± 2.51 [0, 10]

Mean ± standard deviation [range], IQ intelligence quotient, WASI-II Wechsler Abbreviated Scales of Intelligence-II, ADOS autism diagnostic observation schedule scores, ADOS SA symptom severity in the social affect domain, ADOS RRB restricted, repetitive behavior domain, AQ Baron-Cohen Autism Quotient. Some entries are derived from incomplete data. Source data are provided as a Source Data file.

individuals would generalize to participant populations that are neurodiverse. Specifically, we tested if they predict individual differences in the number of social contacts and symptom severity in the social domain (ADOS SA⁵⁴) in the ASD group.

Identifying brain regions from which social inference can be decoded in neurotypical adults (why/how fMRI task)

First, we localized activation patterns in the brain from which social inferences can be decoded in the DS group ('why' vs. 'how' inferences in the why/how fMRI task, see Fig. 1a). Using multivariate decoding together with a searchlight approach^{35,49,50,56}, we found that activation patterns in the bilateral dorsomedial prefrontal cortex (DMPFC), bilateral pSTS, left dorsal and ventral lateral prefrontal cortex (DLPFC, VLPFC), and left superior frontal sulcus (SFS) reliably decoded social inferences in the why/how task ($p < 0.05$, FWE corrected at the voxel level for the whole brain as implemented in SPM12). See Table 2 for details and Fig. 1b for an illustration. These brain regions identified in the DS group served as regions of interest (ROIs) for the subsequent cross-validated prediction of the number of social contacts in real life (and other behavioral measures of interest) in all four participant samples. Thus, ROIs were independently defined concerning our key analyses—the neural prediction of individual differences in peoples' social functioning outside of the lab—and different participant samples (RS1, RS2, and ASD), reducing the risk of producing false positive results and circular analysis (i.e., double dipping)⁵⁷.

Complementary univariate analyses further confirmed the involvement of the identified brain regions during social inference processing (Table S6), mirroring findings of previous implementations of the why/how task^{47,48}. Figure S1 illustrates the overlap between univariate and multivariate results. We defined ROIs based on the results of the multivariate decoding analysis as they have been suggested to be more sensitive than the univariate results⁵⁸, and we intended to closely match the multivariate analysis approach for the key prediction of social network scores.

Neural prediction of individual differences in the number of social contacts in DS

Having localized brain areas that reliably decode social inference processing in the DS group (Table 2; Fig. 1b), we proceeded to the key

question of the study: do neural activation patterns obtained during social inferences predict individual differences in the number of social contacts? We tested this hypothesis using an ROI-based cross-participant prediction approach for SNI scores³⁶. Multivoxel activation patterns in the right pSTS reliably predicted differences in the number of social contacts across individuals in the DS ($r = 0.46$ [−0.38, 0.26], $R^2 = 0.21$, $p = 0.004$, permutation test, significant at $p_{\text{adj}} < 0.05$ FDR corrected⁵⁹; for illustration see Fig. 2c). No other ROI yielded predictions above the chance level (Table 3).

Prediction of individual differences in the number of social contacts in two replication samples (RS1 and RS2)

We showed that neural activation patterns for social inferences predict variance in the number of social contacts in a community sample of DS (cross-validated, localized to the right pSTS). To further validate this finding, we aimed to replicate our results in two datasets of neurotypical individuals: a smaller replication sample RS1 ($N = 20$) matching the ASD group (see below) and a larger replication sample RS2 ($N = 50$).

First, as a sanity check, we validated the ROIs. Specifically, we confirmed that neural activation patterns in the same ROIs derived from the DS sample (Table 2) decoded social inferences in the why/how task also in RS1 and RS2. For every ROI, we carried out a decoding analysis of social inferences ('why' vs. 'how') (separately for RS1 and RS2). These analyses yielded decoding accuracies well above chance level (50%) for each ROI and participant sample (all p 's < 0.05 , permutation tests, FDR corrected⁵⁹). For details, see Table 4; for an illustration, see Fig. 1c. The results demonstrate that all ROIs (identified in DS) reliably decode social inferences across all three neurotypical participant samples (and ASD, see below).

Second, we aimed to replicate our key finding from the DS group: neural activation patterns for social inference in the right pSTS predict variance in people's number of social contacts. Matching results in DS, neural response patterns in the right pSTS reliably predicted SNI scores in RS1 ($r = 0.52$ [−0.49, 0.48], $R^2 = 0.27$, $p = 0.036$) and RS2 ($r = 0.30$ [−0.39, 0.28], $R^2 = 0.09$, $p = 0.043$, permutation tests; Fig. 2c). For completeness, supplemental analyses confirmed that predictive information was specific to activation patterns in the pSTS and did not extend to other ROIs listed in Table 2 (all p 's > 0.14 , FDR corrected). In sum, we successfully replicated the key result: neural activation

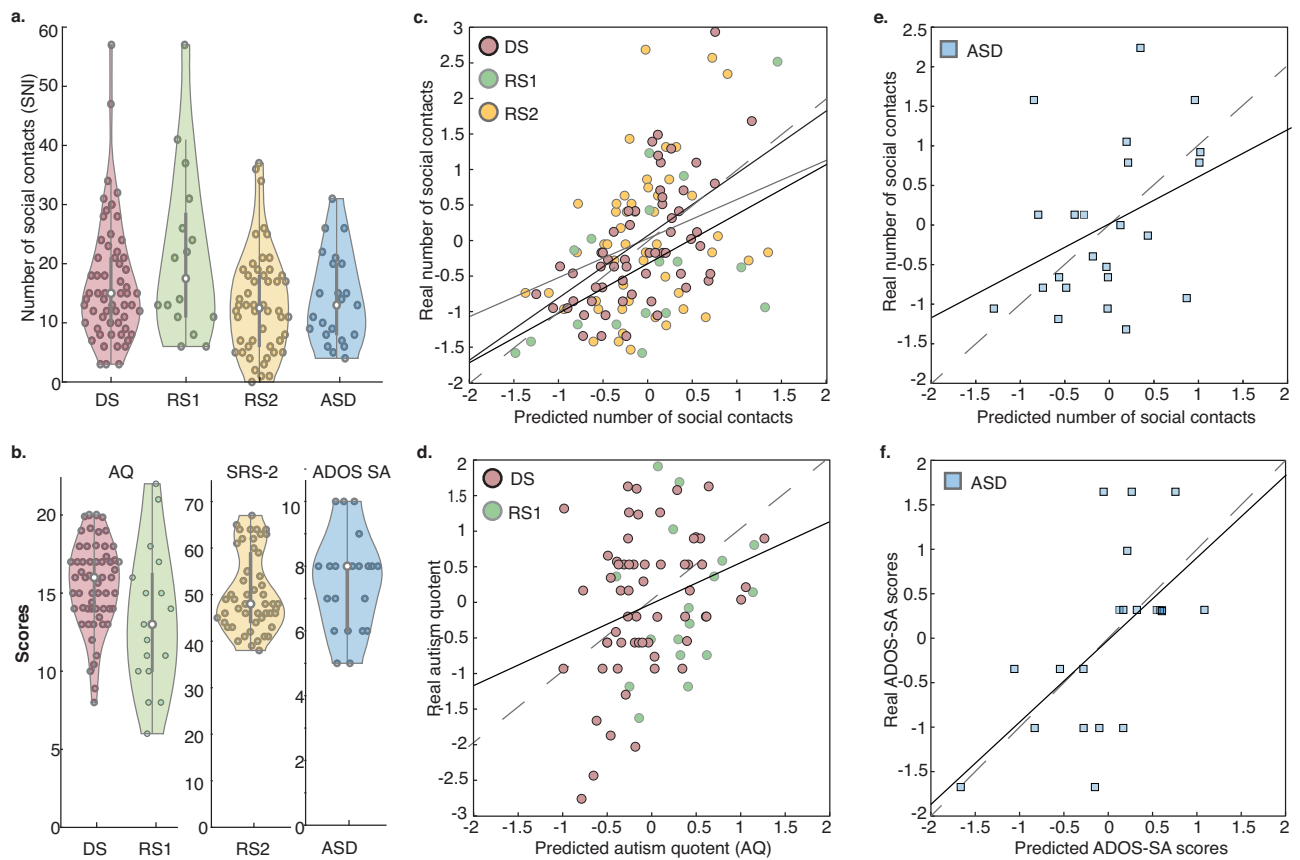


Fig. 2 | Neural prediction of real-world social functioning. **a** Violin plots of the number of real-life social contacts (SNI scores of social network size) in the neurotypical Discovery Sample (DS, $n = 59$), neurotypical Replication Samples 1 and 2 (RS1, $n = 17$; RS2, $n = 50$), and Autism Spectrum Disorder group (ASD, $n = 23$). **b** Violin plots of autism-like traits in neurotypical samples (Autism Quotient, AQ in DS [$n = 59$] and RS1 [$n = 17$]; Social Responsiveness Scale-2 in SRS-2, $n = 50$) and autism symptom severity in the social affective domain in the ASD group (ADOS SA, $n = 21$). Circles within violin plots represent participants; gray bars indicate the 25th–75th percentiles; box plot whiskers illustrate the minima and maxima; white dots indicate the median. **c** Illustration of the match of predicted and actual number of real-life social contacts (SNI, z scored) from brain activation in the right pSTS (DS: $r = 0.46$ [−0.38, 0.26], $R^2 = 0.21$, $p = 0.004$; RS1: $r = 0.52$ [−0.49, 0.48], $R^2 = 0.27$, $p = 0.036$; RS2: $r = 0.30$ [−0.39, 0.28], $R^2 = 0.09$, $p = 0.043$; Pearson correlations, p values derived from non-parametric permutation tests; leave-one-

participant-out-cross-validation). Solid lines indicate trend lines for each sample; dashed lines are for reference and indicate the theoretical perfect correspondence of values on both axes (predicted = real scores). **d** Neural prediction of autism-like trait scores (Autism Quotient, AQ, z scored for comparability across behavioral measures) in the community samples (DS, RS1) from activation patterns in the right pSTS ($r = 0.29$ [−0.34, 0.24], $R^2 = 0.09$, $p = 0.021$, Pearson correlation, permutation test). **e** Neural activation patterns the pSTS from all three neurotypical samples predict the number of social contacts (SNI scores, $r = 0.38$ [−0.35, 0.35], $R^2 = 0.14$, $p = 0.037$) and **f** symptom severity (ADOS SA scores) in the ASD group ($r = 0.60$ [−0.36, 0.36], $R^2 = 0.36$, $p = 0.001$; Pearson correlations, permutation tests; cross-sample prediction: train on data of all neurotypical individuals, test on data of the ASD group). Higher ADOS SA scores indicate more severe symptoms in the social affective domain. Source data are provided as a Source Data file.

Table 2 | Brain regions decoding social inferences in the why/how task in the DS group (whole-brain searchlight decoding)

Brain region	Side	k	t	MNI		
				x	y	z
pSTS (posterior superior temporal sulcus)/ TPJ (temporo-parietal junction)/ SMG (supramarginal gyrus)	L	754	19.09	−46	−66	24
pSTS	R	80	8.93	48	−62	22
DMPFC (dorsomedial prefrontal cortex)	L	430	10.31	−8	60	26
DMPFC	R	161	7.60	4	58	24
SFS (superior frontal sulcus)	L	34	9.68	−30	20	50
DLPFC (dorsal lateral prefrontal cortex)	L	232	8.99	−46	12	36
VLPFC (ventral lateral prefrontal cortex)	L	84	6.42	−50	26	14

Results are reported at a statistical threshold of $p < 0.05$, FWE corrected at the voxel level for the whole brain (cluster threshold of five voxels); only peak activations of clusters are reported; L left hemisphere, R right hemisphere, k cluster size in voxels, MNI Montreal Neurological Institute. Source data available on OSF (<https://doi.org/10.17605/OSF.IO/RNT8S>).

Table 3 | Neural prediction of social network size scores in the DS group

Region of Interest (ROI)	Accuracy [5th, 95th]	R ²	P values
R_pSTS	0.46 * [-0.38, 0.26]	0.21	0.004 *
L_pSTS	-0.03 [-0.29, 0.30]	0.0008	0.524
L_DMPFC	-0.13 [-0.31, 0.39]	0.02	0.709
R_DMPFC	0.16 [-0.34, 0.26]	0.03	0.167
L_SFS	0.16 [-0.38, 0.25]	0.03	0.123
L_DLPFC	-0.11 [-0.33, 0.28]	0.01	0.681
L_VLPFC	-0.11 [-0.36, 0.30]	0.01	0.658

Prediction effect sizes (accuracy scores) are represented by Pearson correlation coefficients of predicted and self-reported numbers of real-life social contacts (supplemental R square values are also reported, error degrees of freedom = 57). Upper and lower boundaries of prediction accuracies achieved by chance (empirical null distributions) are displayed for each ROI [5th and 95th percentile]. Permutation tests assessed the statistical significance of the ROI-wise prediction accuracies (P values). * indicates p values that survive FDR correction ($P_{\text{adj}} = 0.028$)⁵⁹, controlling for the false discovery rate across multiple ROI-wise tests. L/R_left/right hemisphere. ROIs available on OSF (<https://doi.org/10.17605/OSF.IO/RNT8S>). R_pSTS/L_pSTS right and left posterior superior temporal sulcus, L_DMPFC/R_DMPFC right and left dorsomedial prefrontal cortex, L_SFS left superior frontal sulcus, L_DLPFC left dorsal lateral prefrontal cortex, L_VLPFC left ventral lateral prefrontal cortex.

patterns in pSTS during social inferences predict individual differences in people's real-world number of social contacts across three neuro-typical participant samples.

Supplemental analyses tested whether variance in the ability to decode social inferences in the why/how task across participants might be linked to variance in SNI scores. To this end, for each ROI, we correlated participants' ROI-wise decoding accuracies in the why/how task with their SNI scores (separately for each participant sample). We found that participant-specific decoding accuracies for social inferences per se (see Fig. 1c for illustration) were not correlated with their SNI scores (neither in DS, RS1, or RS2; all p 's > 0.31, FDR corrected⁵⁹). In other words, individuals for whom we could more accurately decode social inferences in the fMRI why/how task were not necessarily participants with more social contacts.

Additional post-hoc tests also examined whether conventional univariate measures of brain responses in the why/how task—that focus on estimated neural activity in single voxels rather than multi-voxel activation patterns—reflect individual differences in the number of social contacts (SNI scores). To this end, for every participant, we estimated the average univariate activation during social inferences for each ROI (based on the contrast of [why–how] inferences from participant-specific GLMs, see Methods). Correlations with SNI scores did not yield significant results in the right pSTS or any other ROI (all p 's > 0.31, FDR corrected⁵⁹). These findings suggest that information predictive of social function (like SNI scores) can be obtained from distributed neural activations patterns, but not from proportionally stronger—or weaker—overall univariate activation in these localized brain areas during social inferences.

Domain-specificity of neural predictions of the number of social contacts

Next, we tested whether the ability to predict the number of social contacts is specific to social inference activations in the brain or extends to neural markers obtained during nonsocial inferences. In other words, if participants engage in inferences that are not primarily social in nature, can we still predict variance in individuals' SNI scores from their brain data? To address this question, we conducted a supplemental analysis of data from a nonsocial inference condition in the why/how task we had collected for the RS1 and RS2 groups (the DS group did not perform this control condition, Table 5). In this condition of the why/how task, participants made why/how inferences regarding nonsocial stimuli (e.g., inferring that a rainstorm causes water to pour out of a gutter vs. describing water pouring out of a

Table 4 | ROI-wise decoding of social inference in the why/how task in each participant sample

Regions of interest (ROI)	Decoding accuracy [%]			
	Discovery sample (DS)	Replication sample 1 (RS1)	Replication sample (RS2)	Autism spectrum disorder sample (ASD)
L_pSTS	82 [31,69]	80 [36,65]	79 [36,64]	75 [38,62]
R_pSTS	69 [39,61]	61 [44,56]	70 [40,60]	60 [45,55]
L_DMPFC	74 [35,65]	75 [38,62]	74 [38,62]	71 [40,60]
R_DMPFC	66 [40,60]	66 [42,58]	70 [40,60]	64 [43,57]
L_SFS	63 [42,58]	71 [40,61]	69 [40,59]	66 [42,58]
L_DLPFC	74 [36,64]	68 [41,59]	75 [37,63]	68 [41,59]
L_VLPFC	69 [39,61]	65 [42,58]	69 [40,59]	63 [44,57]

All analyses pertain to the fixed set of regions of interest (ROIs) obtained from the DS group. Thus, ROIs are completely independent of RS1, RS2, and ASD. ROI-wise decoding results for DS are displayed for comparative purposes only. Values represent the average percentage of cross-validated correct classification (decoding accuracy) of social inferences (why vs. how) based on ROI-wise activation patterns in the why/how task in each sample (columns) and ROI (rows). Values in brackets represent the 5th and 95th percentile of the empirical null distributions of decoding accuracies achieved by chance (1000 folds). Neural activation patterns in each ROI and participant sample predicted social inferences well above chance (permutation tests, all p 's < 0.05, FDR corrected), validating the ROI selection across samples. Source data are provided as a Source Data file. R_pSTS/L_pSTS right and left posterior superior temporal sulcus, L_DMPFC/R_DMPFC right and left dorsomedial prefrontal cortex, L_SFS left superior frontal sulcus, L_DLPFC left dorsal lateral prefrontal cortex, L_VLPFC left ventral lateral prefrontal cortex.

gutter)⁴⁸. Importantly, these supplemental analyses did not yield significant predictions of SNI scores for the right pSTS (RS1: $r = -0.45$ [-0.57, 0.47], $R^2 = 0.20$; $p = 0.88$; RS2: $r = -0.17$ [-0.38, 0.28], $R^2 = 0.03$, $p = 0.72$). For completeness, we repeated the analysis for the other ROIs, none of which was significant (all p 's > 0.23 in RS1 and RS2, FDR corrected for each sample). Overall, these findings suggest that neural information predictive of variability in metrics of social network size (SNI scores) is specifically related to making social inferences rather than general causal inferences.

Following up on evidence of domain-specificity, we also aimed to predict behavioral indices that were not explicitly linked to the social domain. More precisely, we repeated our neural decoding analysis with one major difference: we predicted variance in participants' intelligence scores instead of the number of social contacts (Full-Scale IQ assessed in the WASI-II; see Methods) (Table 1). Neural activation patterns in the pSTS obtained during social inferences did not predict variance in IQ scores in DS, RS1, or RS2 (all p 's > 0.62). Complementary tests for any of the other ROIs also yielded no significant predictions of IQ scores (all p 's > 0.50, FDR corrected). This finding adds further specificity and, together with our inability to predict SNI scores from nonsocial inference activations in the brain, argues for discriminative validity: neural responses designed to capture a social cognitive process (social inferences) only predict aspects of social behavior.

Predictions of individual differences in AQ and social responsiveness scores

In this set of analyses, we examined whether the pattern of activation during the why/how task might be predictive also of other behavioral indices of interest in the social domain: autism-like traits in the general population (AQ; Social Responsiveness Scales [SRS-2]). Specifically, in two separate tests, we examined whether neural activation patterns in the right pSTS obtained in the why/how task might also predict individual differences in AQ questionnaire scores⁵¹ in DS and RS1 and variance in SRS-2 scores⁵² in RS2. SRS-2 scores were chosen for our second replication sample for two reasons: first, this measure has been suggested to be preferable to the AQ for assessing the Broad Autism Phenotype⁵³. Second, significant predictions for variance in yet another measure of autism-like traits in the general population provide further evidence that neural predictive information generalizes

Table 5 | Details of how the why/how task, fMRI data collection, and image preprocessing varied across participant samples

	Discovery sample (DS)	Replication sample 1 (RS1)	Autism spectrum sample (ASD)	Replication sample 2 (RS2)
Why/how task				
Nonsocial condition	No	Yes	Yes	Yes
Blocks per condition	4	6	6	6
Stimuli per block	8	8	8	9
Total runtime (min)	5:12	16:35	16:35	13:25
Source code	https://github.com/spunt/whyhowlocalizer ¹⁰⁴	https://osf.io/59cbe/	https://osf.io/59cbe/	https://github.com/adolphslab/LOI2 ¹⁰⁵
fMRI data acquisition				
MR scanner	TRIO	TRIO	TRIO	PRISMA
MR EPI sequence	Multi-band	Single-band	Single-band	Multi-band
fMRI data preprocessing				
Source code	House code ¹⁰⁶	House code ¹⁰⁶	House code ¹⁰⁶	fMRIPrep 1.5.3 https://fmriprep.org/en/stable/
ICA (independent component analysis)	No	No	No	Yes

beyond measures of the number of social contacts (or AQ scores). The implemented cross-validated analysis was identical to the one reported above for SNI scores, simply replacing these with AQ scores (collapsed across DS and RS1) or SRS-2 scores in RS2 (Fig. 2b). Neural activation patterns of social inference in the pSTS reliably predicted AQ scores in DS and RS1 ($r = 0.29$ [-0.34, 0.24], $R^2 = 0.09$, $p = 0.021$, permutation test, Fig. 2d) and SRS-2 scores in RS2 ($r = 0.29$ [-0.41, 0.28], $R^2 = 0.09$, $p = 0.045$, permutation test). Notably, AQ and SNI scores were not correlated in DS and RS1 ($r = 0.16$ [-0.06, 0.37], $p = 0.15$). This suggests that predictions of variance in autism-like traits in neurotypical individuals are unlikely to merely reflect covariation of the behavioral scores of interest. Overall, the successful prediction of variance in autism-like traits indicates that predictive neural information in the right pSTS is not specifically related to SNI scores but extends to other measures of human sociability in neurotypical individuals.

Neural processing of social inferences in the why/how task in ASD

This set of analyses aimed to extend the neural prediction of variance in social network characteristics in a sample of high-functioning adults diagnosed with ASD. First, as a sanity check, we confirmed that the neural activation patterns obtained in the why/how task decoded social inferences in the ASD group for each ROI in Table 2 (all p 's < 0.05, FDR-corrected, ROI-wise permutation tests, Table 4, Fig. 1c). This result demonstrates that our ROIs—identified in the neurotypical DS group—also contained information for decoding social inferences in the ASD sample.

Second, we also explored whether the ROIs are equally informative about social inferences in the ASD group as they are in neurotypical individuals. For each ROI, we compared the decoding accuracies of social inferences in the why/how task between the ASD and the matched RS1 groups (see Methods). Decoding accuracies for social inferences did not significantly differ across the ASD and RS1 groups in any ROI (two-sample t tests, all FDR corrected p 's > 0.12). We also compared whole-brain accuracy maps of a searchlight decoding of social inferences in the why/how task (ASD vs. RS1, two-sample t test at the group level as implemented in SPM12). No brain area encoded social inferences differently for the ASD group compared to the RS1 group (even at slightly more liberal statistical thresholds of $p < 0.001$ at the voxel level, $p < 0.05$ FWE corrected at the cluster level, as implemented in SPM12). Together, these results suggest that local neural information content about social inference processing in the why/how

task is not systematically diminished in ASD compared to matched neurotypical individuals in RS1.

Third, supplemental analyses also tested whether ASD and RS1 groups use similar neural activation patterns to encode social inferences in the why/how task. This was an important sanity check for our cross-sample prediction of indices of social functioning (see below). To explicitly test for shared neural representations in the why/how task across both groups, we trained the classifier on the why/how data of one group (ASD) and tested whether we could correctly identify social inferences in the other group (RS1), and vice versa (two-fold cross-validation, cross-sample decoding). For each ROI, we found that response patterns obtained in one group predicted social inference processing in the other group ($p < 0.05$, permutation tests, FDR corrected) (Supplemental Table S7). This finding suggests that—at least part of—the neural representation of social inference generalizes between the ASD and the RS1 groups.

Taken together, the results of these sanity checks suggest that ASD and a matched group of neurotypical individuals (RS1) recruit similar brain regions during social inferences in the why/how task, with matching levels of neural information content (as captured by decoding accuracies), and—at least in part—shared activation patterns. This pre-requisite, in turn, made possible the subsequent investigation outlined below.

Prediction of number of social contacts and autism symptom severity in ASD (SNI and ADOS SA scores)

So far, we have shown that neural activation patterns of social inference processing in the pSTS predict variance in SNI scores and autism-like traits in neurotypical individuals. We next explored whether this predictive information generalized to individuals in the autism group. To this end, we used a cross-sample prediction approach: we trained our model on data from all three neurotypical participant samples (using neural activation patterns in the pSTS as features and SNI scores of 126 individuals as labels) and tested the model on data from the ASD group (using pSTS response patterns and SNI scores of individuals with autism). This approach is the most direct test of generalization of predictive neural information between neurotypical and autism groups. It also addresses potential methodological concerns related to the smaller sample size in the ASD group (i.e., concerns of limited training data in a leave-one-participant-out cross-validation approach in a small participant sample). The cross-sample prediction analysis allowed us to predict the number of social contacts in the ASD group ($r = 0.38$ [-0.35, 0.35], $R^2 = 0.14$, $p = 0.037$, permutation test, Fig. 2e),

suggesting that—at least part of—the predictive information encoded in the pSTS during social inferences generalizes from neurotypical participant groups to the ASD group.

Next, we tested whether the neural information in pSTS generalizes to other behavioral indices of social functioning in the ASD sample, namely symptom severity scores in the social domain (ADOS SA scores). To this end, we repeated the cross-sample prediction analysis with one difference: We replaced the SNI scores in the training data (neurotypical) with participants' autism-like trait scores and the test data (ASD) with autism symptom severity scores. We found that pSTS responses in neurotypical individuals predicted variance in autism symptom severity scores in the ASD sample ($r = 0.60$ [$-0.36, 0.36$], $R^2 = 0.36$, $p = 0.001$, Fig. 2f). Note that ADOS SA scores and SNI scores in the ASD group were not significantly correlated ($r = 0.04$ [$-0.40, 0.46$], $p = 0.86$), indicating that positive predictions of autism symptom severity in the social domain are not merely due to shared variance between ADOS and SNI scores.

Supplementary post hoc tests examined the specificity of these predictions for the social domain. We found that pSTS activation patterns did not predict variance in repetitive behaviors in the ASD group (ADOS RRB) ($r = -0.23$ [$-0.35, 0.38$], $R^2 = 0.05$, $p = 0.83$). Thus, predictive information in the pSTS was specific for variance in symptom severity in the social affective domain (ADOS SA). Next, we used pSTS activation patterns obtained during nonsocial inferences in the why/how task. Neither the prediction for SNI scores ($r = 0.30$ [$-0.34, 0.35$], $p = 0.92$) nor ADOS-SA scores ($r = -0.09$ [$-0.37, 0.37$], $p = 0.63$) yielded significant results here. These results suggest that predictions of SNI scores and autism symptom severity scores in the social affective domain in the ASD group might be specific to neural activation patterns in the social domain (social inferences).

Discussion

We found that patterns of brain responses to social inferences measured in the laboratory predict individual differences in the number of real-life social contacts measured with the SNI. This finding held across four independent participant samples (DS, RS1, RS2, ASD) when using a fixed set of functionally defined brain regions of interest (ROIs), pointing to the robustness and generalizability of our key finding. The number of social relationships measured with the SNI is a metric that has received considerable attention in the literature, in both behavioral and neuroimaging studies^{10–12,14,15,20,22–26,36,60}. Our results directly link peoples' capacity to engage in a specific social cognitive process (thinking about other people's inner mental states) and individual differences in real-world social behavior (yielding varying social network sizes). We found remarkable specificity for a particular neuroanatomical region (the right pSTS). At the same time, our findings generalized across multiple measures of individual differences in social behavior (SNI scores, autism-like traits in neurotypical individuals in community samples (AQ, SRS-2), and autism severity scores (ADOS SA)). We suggest that the patterns of neural activations in the pSTS during social inference reflect an underlying latent variable that comes into play relatively ubiquitously in real-life social behavior. Our findings further argue that variability in the number of social relationships has a partly neural basis. Moreover, the results point to the potential of using process-specific neural markers to characterize individual differences in real-world social behavior.

We predicted the number of social contacts from brain activations during a laboratory task that required inferring other people's mental states (e.g., their feelings, thoughts, or intentions). Such social inferential processing is often referred to as “mentalizing” or “theory of mind”^{29,61,62}, encompassing a rather broad collection of different abilities^{44,63} and recruiting a network of multiple brain regions^{41–43}. Our present findings provide considerable further specificity on how these processes are associated with social behavior. Notably, we found high neuroanatomical specificity: while multiple brain regions decoded

social inferences in the why/how task (Table 2), information that was predictive of variance in the number of social contacts (as well autism-like traits) was exclusively localized to the right pSTS (Table 3). The anatomical specificity of our predictions suggests that different nodes of the brain network recruited during social inferences might implement distinct mental computations, and only the computations in the pSTS generalize to indices of social functioning tested in our study. Our analyses shed light on the nature of the predictive neural computations in the pSTS: brain responses obtained during why/how inferences for nonsocial stimuli (e.g., related to extreme weather phenomena) did not predict the number of social contacts, autism-like traits in neurotypical individuals, or symptom severity in our autism group. This finding suggests that information predictive of the number of social contacts is not coming from domain-general computations of making high-level inferences, but specifically from inferences about other people's mental states. These findings fit emerging evidence for the range of social processes subserved by the STS. In particular, there is an anterior-to-posterior organization that localizes processes ranging, respectively, from language to voices to faces to biological motion to theory-of-mind^{64,65}. The region we found in our study, the very posterior STS, corresponds to the one most specifically implemented in abstract social inferences⁶⁵ (Figure S2a). Finally, neural activation patterns for social inferences in the why/how task did not predict variance in people's cognitive functioning that was not necessarily social in nature, such as intelligence scores or symptom severity regarding restrictive and repetitive behaviors in the ASD group (ADOS RRB scores). This finding further supports the notion of the high specificity of our neural prediction to differences in people's social behavior: neural responses capturing a social cognitive process (social inferences) only predict aspects of social behavior, arguing for discriminant validity. Overall, our results demonstrate the potential of using process-specific, localized neurobiomarkers of cognitive abilities that contribute to social behavior as robust predictors of people's social functioning in the real world.

At the same time, we demonstrated that predictive neural information generalizes to other indices of people's social functioning. Specifically, we could also predict variance in autism-like traits in the general population (AQ and SRS-2 scores) and autistic social symptom severity in the ASD group (ADOS SA scores). This generalizability is especially important given that our primary measure of social network size, the SNI score, is based on self-reports. The ADOS SA scores, by contrast, are based on observer ratings of behavior, arguing that our findings are not merely driven by response biases on questionnaires. The convergence of associations we found, and the fact that they pertain specifically to the activation patterns produced by social inference processes, go well beyond prior findings based on resting-state or structural MRI data. They also raise the important question of what exactly drives our findings. Although we have previously argued for caution in studies that claim to find associations between the SNI and brain structure¹³, variance in brain structure in the right pSTS (namely gray matter density) has previously been linked to variance in the number of friends on online social network services¹². Predictions of indices of social behavior unrelated to metrics of social network size point to the intriguing possibility that social inference activation in the pSTS reflects a latent variable that predicts individual differences in social functioning quite broadly (such as AQ and SRS-2 scores, ADOS SA severity scores, and the number of social contacts). It remains an open question how far this generalizability would extend, e.g., to more sophisticated metrics of social networks such as social network position, characteristics of social ties or social distance^{22,66,67}, or even other social behaviors known to activate similar brain regions of the pSTS such as altruistic behavior^{35,56} (see Figure S2b for illustration of overlap). Multi-trait-multi-method approaches (MTMM)⁶⁸ might be well suited to address these future questions, shedding further light on the underlying latent variable that we propose and on how the cognitive

processes engaged could be causally related to social connectedness and everyday social functioning.

Another key question about generalizability is whether our findings would extend to different participant samples, specifically ones not necessarily representative of the discovery sample, either in terms of predictor variables (i.e., the cognitive process of interest or its neural signature in the brain), the dependent variable (i.e., markers of real-world social functioning and connectedness), or both. We explicitly tested the generalizability of our findings in a sample of high-functioning individuals with autism. Marked deficits in social interaction and specifically in understanding other people's mental states are a hallmark of ASD³⁹. Yet, we found that brain responses of social inference measured in the lab predict the number of social contacts, and metrics of autism symptomatology in the social domain (ADOS SA). Notably, the successful predictive model was trained on data from neurotypical individuals. This approach is the most direct test of generalization of predictive neural information between neurotypical and ASD groups. Our results point toward a generic component of information encoded in activation patterns pSTS that is shared across neurotypical and the ASD sample. Overall, the findings suggest that neural predictions of social behavior extend to individuals in non-representative samples with atypical behavior in the social domain.

Evidence of successful cross-sample predictions of social functioning in ASD has important practical implications: limited sample sizes in special interest groups or clinical populations are common, highlighting the broad appeal of using neurotypical groups—that are often easier to recruit and test—to train predictive models of interest. Moreover, on a conceptual level, our results support the idea that social inference abilities are continuously distributed across neurotypical individuals drawn from a community sample and neurodiverse individuals with ASD. Instead of a categorically defined condition, ASD can be viewed as one end of a continuous distribution of core competencies and/or deficits that occur in nature⁶⁹ (but see ref. 70). Our findings across neurotypical and ASD groups support an understanding of the core feature of problems with social interactions of ASD as a quantitative trait rather than a categorically defined condition⁶⁹, with implications for diagnostic classification. Our findings also suggest that neural activations to social inference could be incorporated into developing biomarkers for ASD, offering avenues for diagnosis and mechanistically targeted treatments⁷¹. Biomarkers for autism are scarce, partly due to the heterogeneity in phenotypes and difficulties in identifying “biological tests” to predict treatment responses⁷². Neural measures of atypical cognitive processes, such as the ones we feature in this study, hold the promise of predictive biomarkers sensitive to quantitative variations in core features of ASD.

While reliable across four participant samples, some limitations and open questions need highlighting. For instance, our results remain limited by small sample sizes, reducing our power to detect possible effects in this study. Thus, the lack of prediction from ROIs other than the right pSTS may have resulted partly from small sample sizes. Future replications in much larger samples that test hundreds or thousands of individuals will be necessary to establish the likelihood of any negative findings further. Likewise, while our results show a link between social inferences and social network characteristics, they can't speak to variance in other social behaviors (e.g., normative choices⁷³ or expressions of prosocial phenotypes^{74–76}).

Another important open question concerns the trajectories of neural prediction of social network characteristics over time. ASD is a neurodevelopmental disorder that arises early in life, raising the question of how the predictive information of our identified neural markers changes across the lifespan. Do these markers characterize social deficits when they are first displayed? Can they predict symptom severity later in life or identify subgroups of young children likely to respond to treatments targeting socio-cognitive processing? Likewise, in neurotypical individuals in the general population, social network

characteristics change over time. Longitudinal studies will be necessary to assess the predictive value of functional neural markers across time. It will be interesting in this respect also to investigate the effects of training people's social skills^{77,78} to see how the relationship between neural markers and social networks can be modified.

These considerations raise the perhaps most challenging question about interpreting our findings: their causal relation. Do social inference skills, as indexed in pSTS activation patterns, cause real-world social behavior that influences social network size? Or do people with larger social networks and more social contacts in daily lives develop better social inference skills? In general, it is surprising that there should be any substantial relation at all since social networks are influenced by many factors in life that appear independent of social inference skills (moving to a new location, time available to socialize, type of job, etc.). The only clear evidence for a causal relationship currently comes from studies in Macaques where social membership to groups of variable size could be experimentally assigned⁷⁹. Living in larger social groups *caused* changes in brain structures as measured in increased gray matter volume and changes in brain function in non-human primates⁷⁹. Positive correlations between a monkey's social rank within a group and neuroanatomical markers support the notion that social structure is associated with neural measures. However, these results leave unclear the precise cognitive processes mediating between the two. This evidence would argue that the vagaries of life circumstances cause membership in a particular social group, and this, in turn, causes changes in the neural substrates of social cognition. More in-depth characterization of people's life variables in a longitudinal design could shed further light on these intriguing questions.

It is implausible that the pattern of brain activation and SNI scores we observed in our study are directly causally related. After all, while we acquired neuroimaging data, participants were neither filling out SNI questionnaires nor otherwise instructed to think about their friends or social networks. So what could explain the reliable association that we found? An alternative possibility is that both brain activation patterns and the number of real-world social relationships measured by SNI scores are caused by a shared latent variable: the social cognitive processes engaged during our activation task. Considering the task conditions that elicit the raw data that constituted the predictive multivoxel activation patterns, we propose that it is precisely the intended social inference process that constitutes the latent variable in question. Participants are thinking about other people's mental states and the intentions responsible for the actions depicted in the stimuli, and it is such mental state attribution that is causally related to establishing larger numbers of social relationships (as indexed with the SNI).

This hypothesis is supported by the relationship we found with SRS, AQ, and ADOS severity scores, particularly by the fact that AQ scores are not themselves correlated with SNI scores. Instead, these two metrics of social functioning (autism-like traits as captured in AQ and SNI) are independently associated with brain activation patterns. Our interpretation is that social inference must itself be comprised of multiple cognitive sub-processes, distinct sets of which are causally related to different consequences in the social world: autism-like traits and SNI. Although it is known that social inference is not monolithic^{44,63}, this still leaves open considerable detail to be investigated. The anatomical specificity we found will be particularly informative in further dissecting this issue in future studies. For instance, our prior work has shown that social inference activation can be correlated with social curiosity or attributional complexity skill—but in this instance, the effect was found only in the dmPFC, not in the pSTS⁴⁸. Our findings emphasize the pSTS as one node in the social cognition network whose variable activation patterns are a promising marker for social functioning in the real world.

Method

Participants

The study reports data from four participant samples: the discovery sample (DS, $n = 60$) and two replication samples (RS1: $n = 20$; RS2: $n = 55$) of neurotypical volunteers from the Los Angeles metropolitan area and the group of high-functioning individuals diagnosed with autism spectrum disorder (ASD, $n = 25$) (see Table 1). Individuals in the RS1 group were age-, sex- and IQ-matched to individuals in the ASD group to allow for a matched comparison of neural activation of social inferences (why/how task) across groups. All participants in the ASD group had a prior clinical diagnosis of autism spectrum disorder, which was confirmed by revised algorithm scores on the Autism Diagnostic Observation Schedule (ADOS), Module 4. All participants (DS, RS1, RS2, ASD) participated in a version of the why/how fMRI task (Table 5) and a separate behavioral session in exchange for financial compensation (\$20/h). All participants were right-handed, had a normal or corrected-to-normal vision, spoke English fluently, and had IQ scores in the normal range (as assessed using the Wechsler Abbreviated Scales of Intelligence-II; FSIQ, Table 1). Sex at birth was assessed using self-reports (see Table 1; source data are also reported in a de-identified form on OSF, <https://doi.org/10.17605/OSF.IO/RNT8S>). Sex differences were not hypothesized for the current research question. We excluded data from one participant in the DS group and one participant in the RS2 group due to poor performance in the why/how fMRI task (no responses to > 70% of trials). We excluded data from two individuals in the ASD group due to artifacts in the fMRI data. We excluded data from one participant in RS2 due to an SNI score of 106 that exceeded three standard deviations above the group average (Table 1) and another three individuals in the RS2 group due to excessive motion during the why/how task. Motion outliers at the individual level were identified and excluded based on the approach described in ref. 80,81, which uses low-pass filtering to control false rejections associated with respiration and pseudo-motion signals present in short repetition time multi-band EPI sequences. The six rigid body motion parameters estimated during pre-preprocessing were low-pass filtered temporally using a fifth-order Butterworth filter with a critical frequency of 0.2 Hz. Following filtering, the framewise displacement (FD) time series were calculated⁸² for each participant. Participants with unusually high motion were identified from the set of individual 50th and 95th percentile frame-wise displacement (FD) values using the DBSCAN clustering algorithm⁸³ implemented by scikit-learn⁸⁴. All participants provided written informed consent according to a protocol approved by the Institutional Review Board of the California Institute of Technology (#12-0343).

Behavioral indices of social functioning

Social network characteristics (number of social contacts, SNI): We measured participants' social network characteristics using the SNI^{4,15,26}. The questionnaire assesses how many people participants see or talk to regularly in 12 types of social relationships, including family, friends, workmates, neighbors, religious groups, etc. (for a complete list, see <http://www.midss.org/content/social-network-index-sni>). Our core measure of interest, the estimated number of social contacts, refers to the total number of people with whom the respondent has regular contact at least once every two weeks, regardless of the type of social role involved. While not the focus of our study, the SNI also yields two additional scores: network diversity (number of unique social roles in which the participant regularly interacts with at least one other person) and number of embedded networks (number of different network domains in which a respondent is active).

Autism-like traits in neurotypical individuals in DS, RS1, and RS2 (AQ and SRS-2): Within the general population, autism-like traits have been proposed to be continuously distributed⁵². We measured variance in autism-like traits using the AQ⁸⁵ (Autism Spectrum Quotient; in DS, RS1) and Social Responsiveness Scale-2⁵² (SRS-2; in RS2). Both the

AQ and SRS-2 have self-report versions used to screen for ASD in adults.

The Autism Spectrum Quotient (AQ) is a self-report survey used to screen for autism spectrum traits in adults with normal-range intelligence (https://www.autismresearchcentre.com/arc_tests). It is comprised of 50 items scored using a 4-point Likert scale to indicate how strongly an individual agrees or disagrees with each statement (e.g., "I find it difficult to work out people's intentions"; "I find myself drawn more strongly to people than to things"). Scores at or above 32 are a strong indicator that autism symptoms are present. Note, however, that AQ scores are not designed for diagnostic purposes.

The Social Responsiveness Scale-2 (SRS-2), Adult Form, Self-Report is a 65-item questionnaire that assesses social deficits associated with autism spectrum disorder. It is sensitive to mild forms of impairment in individuals who do not meet the criteria for an autism diagnosis, as well as to variance in the severity of deficits among individuals with ASD. We have collected the SRS-2 for our second neurotypical replication sample (RS2) as the SRS-2 has been suggested to be preferable to the AQ for assessing the Broad Autism Phenotype⁵³. It also allowed assessing if our results could be replicated using yet another behavioral measure of variance in atypical social processing associated with autism.

Symptom severity scores in ASD (ADOS SA): ADOS SA scores served as a quantitative, continuous measure of autistic symptom severity in the social affective domain in the ASD group. The Autism Diagnostic Observation Schedule, 2nd Edition (ADOS-2⁸⁶), is a standardized observational diagnostic measure. ADOS-2 Module 4, designed for verbally fluent adolescents and adults, was administered to all participants with a prior diagnosis of ASD or clinically significant social skills deficits (i.e., social skills deficits that required clinical treatment in the PEERS program at UCLA; <https://www.semel.ucla.edu/peers/young-adults>). Following the ADOS-2 protocol, a clinically trained examiner engaged each participant in a series of standardized discussions and tasks and rated specific aspects of participant's communication and social behavior on a scale from 0 (no abnormality) to 3 (severe abnormality). The revised diagnostic algorithm⁵⁵ was used to score the presence of autistic behaviors in the Social Affect (SA) domain and Restrictive and Repetitive Behaviors (RRB). The ADOS SA score provides a classification of autism spectrum disorder or non-spectrum, with a sensitivity of 89% and specificity of 72.2%. The algorithm scores were then converted to a calibrated measure of severity, which provides a continuous, quantitative index of social-communication (SA) and restricted and repetitive behaviors (RRB) extending beyond the diagnostic categorization of autism⁵⁴. Given our focus on indices of social functioning, the analyses focused on ADOS SA scores.

Functional brain markers of social inference: why/how task (fMRI)

To assess functional neural markers of social inference processing, participants in the DS performed an established why/how social inference task^{47,48} during fMRI data acquisition (see Table 5 for the hyperlink to source code). The task uses a 2 (inference: why, how) × 2 (target stimuli: faces, hands) factorial design implemented in pseudo-randomized blocks (4 blocks per condition). Each task block consists of a series of 8 images that display social targets in the form of emotional facial expressions (faces) or intentional hand actions (hands) (Fig. 1a). Each image requires a speeded judgment (yes/no), entailing either social inferences about others' internal states ("why?"; high level of inference, e.g., "Is the person helping someone?") or factual inferences about the observed behavior ("how?"; low level of inference, e.g., "Is the person looking to their side?"). Table S3 presents a complete list of block-specific questions. Each image is presented twice, once per inference condition. A brief verbal cue (e.g., "helping?") presented prior to every target image after the first minimizes working memory

demands. Participants had a maximum of 1750 ms to respond before the presentation of the next image. Once an answer was given, the task immediately advanced. Block onsets were fixed across participants. The order and onsets of trials were optimized to maximize the efficiency of separately estimating the contrast of interest [why > how] for each of the two target categories. In addition, the order of why and how blocks were counterbalanced within each target category. The order of trials was achieved by using open-source fMRI design optimization software (<https://zenodo.org/record/58616#.W5mujWaUnbk>) to generate one million pseudo-random designs and for each summing the efficiencies of the contrasts of interest. The most efficient design for each task was retained and used for all participants. Stimulus presentation and response recording used the Psychophysics Toolbox PTB 3.0.13–3.0.16⁸⁷. An LCD projector was used to present the task on a screen at the rear of the scanner bore that was visible to participants through a mirror positioned on the head coil. Participants were given a button box and made their responses using their right-hand index and middle fingers.

Performance in the why/how task was assessed as follows: for each participant, we computed measures of mean percent accuracy, *d'* and response time (RT) for the four conditions (inference [why, how] × targets [faces, hands]). We omitted rare trials with no response (mean = 0.46%, max = 3.91%, across participants). To address negative skewness, accuracy scores were subjected to a Box–Cox transformation before undergoing statistical testing⁸⁸. Mean RTs were computed after excluding values greater than three SDs from the mean. Table S2 reports group-level descriptive statistics.

The version of the why/how social inference task performed by participants in the RS1, RS2, and ASD groups differed in several ways from the one introduced above. These differences were motivated by the specific questions being investigated in the larger study for which it was designed. First, and most importantly, in addition to social target stimuli (hands, faces; Fig. 1a), modified task versions featured a third stimulus category showing the effects of nonsocial processes (e.g., scenes showing the consequences of extreme weather) (for details, see ref. 48). Matching the social stimulus conditions (faces, hands), each nonsocial stimulus was presented twice, requiring either a high level of inference (“why?”) or a low-level inference based on facts displayed in the picture (“how?”). Note that data from this nonsocial control condition was outside of the focus of the current study and was only analyzed for post hoc analyses that examined the specificity of the neural prediction of social functioning (e.g., number of social contacts) for making inferences about other people’s inner mental states. Second, task versions differed regarding the number of trials per block and the number of blocks (Table 5). Third, block-specific questions differed slightly (Table S3). Fourth, there were minor differences in the timing of the trial elements within each block and with the average stimulus onset asynchrony⁴⁸ (see Table 5 for further details and hyperlink to source code).

Functional image acquisition (fMRI)

All imaging data were acquired at the Caltech Brain Imaging Center. Imaging data of DS, RS1, and ASD were collected using a Siemens Trio 3.0 Tesla MRI scanner outfitted with a 32-channel phased-array head coil. For the DS, we acquired 304 whole-brain T2*-weighted echoplanar image volumes (EPis; voxel resolution = $2.5 \times 2.5 \times 2.5$ mm³, 56 slices, TR = 1000 ms, TE = 30 ms, flip angle = 60°, FOV = 200 mm, interleaved acquisition order, multi-band acceleration factor = 4) for the why/how social inference task. For the RS1 and ASD groups, whole-brain T2*-weighted EPI volumes for the why/how task were acquired with the following MR protocol: voxel resolution = $3 \times 3 \times 3$ mm³, 47 slices, TR = 2500 ms, TE = 30 ms, flip angle = 85°, FOV = 192 mm, ascending acquisition order. For RS2, imaging data were collected at Siemens 3.0 Tesla MAGNETOM Prisma.Fit MRI scanner outfitted with a 32-channel phased-array head coil. For RS2, we acquired 1080 whole-brain T2*-

weighted EPI volumes with the following parameters: TR = 700 ms, TE = 30 ms, 60 contiguous oblique transverse slices, slice pitch 20 degrees, 2.5 mm isotropic voxel size, multi-band acceleration = 6, interleaved slice acquisition order, flip angle = 53 degrees. A pair of spin echo EPI volumes (TR = 5500 ms, TE = 48 ms, and multi-band acceleration = 1) with opposing phase encoding polarity were acquired for geometric distortion correction with identical geometry and EPI echo spacing to the T2*-weighted EPI volumes. Participants’ in-scan head motion was minimal (max translation = 2.78 mm, max rotation = 1.88°), and movement parameters were comparable for RS1 and ASDs (no significant between-group difference in frame-to-frame displacement or frame-to-frame rotations as measures of participant motion most likely to cause artifactual BOLD signal correlations; unpaired *t* tests, *p*’s > 0.17, uncorrected). For all participants, we also acquired a high-resolution anatomical T1-weighted image (1 mm isotropic) and field maps used to estimate and correct for inhomogeneity-induced image distortion.

Image preprocessing (fMRI data analysis)

Images in three participant samples (DS, RS1, ASD) were processed primarily using Statistical Parametric Mapping (SPM12, <http://www.fil.ion.ucl.ac.uk/spm>) together with house code (see Table 5 for the hyperlinks). Prior to statistical analysis, functional imaging data were subjected to the following preprocessing steps: (1) the first four EPI volumes were discarded to account for T1-equilibration effects; (2) slice-timing correction was applied; (3) the realign and unwarp procedure was used to perform distortion correction and concurrent motion correction; (4) the participants’ T1 structural volume was co-registered to the mean of the corrected EPI volumes; (5) the group-wise DARTEL registration method included in SPM12⁸⁹ was used to normalize the T1 structural volume to a common group-specific space, with subsequent affine registration to Montreal Neurological Institute (MNI) space; (6) all EPI volumes were normalized to MNI space using the deformation flow fields generated in the previous step, which simultaneously re-sampled volumes to 2 mm isotropic, (7) and smoothed using a Gaussian kernel of 6 mm isotropic, full width at half maximum (FWHM).

Images in RS2 were obtained after a major update of the MR scanner and were preprocessed using the open-source fMRIPrep 1.5.3 analysis pipeline⁹⁰, which utilizes a combination of tools from well-known software packages (see Table 5 for hyperlink). fMRIPrep is designed to provide an easily accessible, standardized, state-of-the-art fMRI data preprocessing pipeline that is robust (e.g., to variations in scan acquisition protocols) and requires minimal user input. An extensive description of the analysis steps is provided in the boilerplate that fMRIPrep generates for every participant processed (Methods S1).

GLM: block-wise estimates of social (“why?”) and factual (“how?”) inferences in the why/how fMRI task

For each participant, a general linear model (GLM) estimated regressors of interest for each block of the why/how task. Blocks were defined by the onset of the first target image and the offset of the final image of the block⁹¹. The GLM included as covariates of no interest the six motion parameters estimated from image realignment and a predictor for every time point where the in-brain frame-wise signal change (calculated as the root mean square derivative, or DVARS) exceeded 2.5 SDs of the mean DVARS across the time series or where frame-wise displacement exceeded 0.5 mm of translation or 0.5° of rotation⁹². Note that participant-wise GLMs in RS2 only included as covariates of no interest the six motion parameters estimated from image realignment due to differences in the preprocessing pipeline in this sample (Table 5, Methods S1). The hemodynamic response was modeled using the canonical (double-gamma) response function and a 1/100 Hz high-pass cutoff filter to eliminate low-frequency drifts in data. GLMs were

estimated using the SPM12 RobustWLS toolbox, which implements the robust weighted least-squares estimation algorithm⁹³. Estimated responses for the block-wise regressors of interest (i.e., ‘why’ and ‘how’ inference task blocks) were used as inputs for the multivariate classification of ‘why’ vs. ‘how’ inferences to identify (DS) or confirm (RS1, RS2, ASD) brain regions that decode social inferences in the why/how task (described below). Moreover, for each participant, we estimated one contrast image based on regressors of interest in conditions that included social stimuli ([why_{face-blocks} AND why_{hand-blocks}]–[how_{face-blocks} AND how_{hand-blocks}]). These participant-specific contrast images were used as inputs for the multivariate regression analyses to predict variance in participants’ number of social contacts (SNI scores) (and other behavioral indices of interest such as AQ, SRS-2, or ADOS SA scores). For each participant, we also estimated one matching contrast image for the nonsocial control condition [why_{nonsocial}–how_{nonsocial}]. Note that this contrast was only estimated for participant samples for which the nonsocial task condition was available (Table 5). These participant-wise contrast images were used as input for the post-hoc multivariate regression analysis that probed the social specificity of our neural predictions.

Identifying ROIs of social inference processing in the why/how task in DS

In the first step, we identified brain regions that are reliably engaged during social inference processing in neurotypical individuals in the DS group. More precisely, this initial analysis aimed to identify multi-voxel activation patterns in the brain that decode social (versus factual) inferences in the why/how task (i.e., why versus how task blocks). To this end, we used a linear support vector machine classifier (libSVM, <http://www.csie.ntu.edu.tw/~cjlin/libsvm>) in combination with a standard whole-brain searchlight approach^{49,94,95} implemented in MATLAB 2021a). This approach has been used in various domains of cognitive and social neuroscience^{95–97}. This data-driven approach does not depend on a priori assumptions about informative brain regions and ensures unbiased information mapping throughout the whole brain.

For each participant, we defined a sphere with a radius of four voxels around a given voxel v_i of the measured brain volume^{56,98–100}. We extracted the parameter estimates of the N voxels within this sphere from the block-wise regressors of interest from the GLM (see above). The resulting N -dimensional pattern vectors were created separately for each block and condition of the why/how task (inferences [why, how] \times target [hands, faces]). We then trained a linear kernel support vector machine classifier to distinguish neural pattern vectors associated with either social or factual inferences (why versus how; fixed cost parameter $c=1$). Training data consisted of all but one neural pattern vector of a particular inference condition obtained for the participant (i.e., leave-two-blocks-out cross-validation; left-out-data: neural pattern vectors for one why block and one how block). The resulting model provided the basis for the prediction of the inference condition of the two left-out task blocks (test data) solely based on their neural pattern vectors. This procedure was repeated several times, always using pattern vectors of a different task block as test data. Predictive information was defined as the average percentage of correct classification of the inference condition and was assigned to the central voxel of the searchlight sphere. This approach was repeated for every voxel of the measured brain volume, yielding a three-dimensional decoding accuracy map for every participant. Participant-specific accuracy maps were then used in a random effect group analysis for DS (single t-test as implemented in SPM12) and tested against chance level (50% in this binary classification) ($p < 0.05$, FWE corrected at the voxel level for the whole brain volume as implemented in SPM12). To identify response patterns that decoded engagement in social inferences independent of the social stimulus triggering this process, we implemented the decoding analysis separately for each target condition of the why/how task (i.e., faces and

hands, Fig. 1a, Table S4 and Table S5). This also allowed us to verify that faces and intentional hand actions elicited common inference processes in the brain¹⁰¹. At the group level, we then identified brain regions that decoded social inference across both target conditions, using the implicit masking function in SPM12 ($p < 0.05$, FWE corrected at the voxel level for the whole brain, cluster threshold of five voxels). The resulting clusters were defined as regions of interest (ROIs) for the multivariate prediction of individual differences in social network characteristics (see below) for all four participant samples. Focusing on a fixed set of brain areas identified in DS minimizes the risk of circular analysis (i.e., double dipping)⁵⁷ and producing false positive results in the remaining three participant samples (RS1, RS2, ASD).

Confirming ROIs of social inference processing in the why/how task in RS1, RS2, and ASD

For each ROI, we tested if neural activation patterns decoded social inferences in the why/how task in RS1, RS2, and ASD, respectively. The sanity check used the following analysis steps: for every ROI, we separately carried out a decoding analysis of social inference (why versus how) for data of the why/how task (separately for individuals in RS1, RS2, and ASD). This analysis conceptually matched the searchlight decoding analysis implemented in DS with one major difference: neural activation patterns were extracted from a particular ROI identified in DS (not a spherical searchlight cluster). Non-parametric permutation tests assessed the statistical significance of average ROI-wise decoding accuracies for each participant sample. Actual decoding accuracies for the group were compared to accuracies of an empirical null distribution (realized by randomly permuting the pairing of participants’ neural pattern vectors and the binary label coding for the inference condition, fixed permutation order across participants, 1000 permutations). Only predictions above the 95th percentile of null distributions were considered statistically significant¹⁰². For each participant sample, we used the `fdr_hr` function in MATLAB R2021a to apply FDR correction⁵⁹ across the family of ROI-wise statistical tests. Permutation tests were not used in the whole brain searchlight analysis in DS (see above) due to the computational costs associated with the high number of searchlights (~250,000) compared to the seven ROIs.

Neural prediction of individual differences in the number of social contacts

Having identified multi-voxel activation patterns that reliably decoded social inference processing, we proceeded to the key question of the study: can we use functional neural markers of social inferences to predict individual differences in social network characteristics? To address this question, we ran a support vector regression (SVR) analysis of an individual’s SNI scores (labels) and the participant’s ROI-based brain responses in the why/how task (features).

First, we describe the cross-participant prediction in neurotypical groups. For each ROI (Table 2), we performed the following analysis steps: for every participant, we extracted parameter estimates for all voxels in an ROI (Fig. 1b) from the contrast image of individuals’ GLM ([why_{face-blocks} AND why_{hand-blocks}]–[how_{face-blocks} AND how_{hand-blocks}]). The resulting neural pattern vectors (one per participant) were used as input features for the prediction, and the participant’s number of social contacts (SNI scores) served as labels (z-scored for all individuals within a participant sample). To implement the multi-voxel SVR, we used a linear v-SVR in LIBSVM (a popular library for support vector machines that has gained wide popularity in machine learning and many other areas) in MATLAB R2021a. We used a fixed cost parameter $c=1$ and a leave-one-participant-out approach. For example, for the discovery sample DS, this approach yielded a 59-fold cross-validation: training was based on data from 58 participants (training data). We then tested if the model could predict the SNI score of the remaining participant solely based on this participant’s ROI-specific neural activation pattern (test data). ROI-specific prediction accuracies reflect correlations of the

observed and predicted social network score across participants of each participant sample^{35,56,99}. Permutation tests assessed the statistical significance of the prediction by comparisons to the empirical null distribution estimated for this ROI and sample (realized by randomly permuting the pairing of participants' neural pattern vectors and behavioral SNI scores 1000 times). Only predictions above the 95th percentile of null distributions¹⁰² that survived FDR correction across the seven ROIs⁵⁹ were considered statistically significant for participants in DS. FDR correction was implemented using the `fdr_hr` function in MATLAB R2021a. These analysis steps were repeated for data from the RS1 and RS2 groups, as well as for other indices of social functioning such as AQ, SRS-2, and WASI-II IQ scores. ROIs were identical across all four participant samples to explicitly test the generalizability of our findings in DS to data from other participant samples.

Next, we tested whether predictive neural information generalizes from neurotypical participant samples to individuals in the ASD group. To this end, we used a cross-sample approach to predict SNI scores (number of social contacts) and symptom severity in the ASD group. The analysis approach matched the cross-participant prediction described above with one significant difference: we trained our model on data from all three neurotypical participant samples (using participant-wise activation pattern vectors from the right pSTS of 126 individuals as features and their SNI scores as labels) and tested the model on data of the ASD group (using pSTS-wise activation pattern vectors of 23 individuals as features and their SNI scores as labels). Both training and test data were standardized (separately for each sample) using the z-score function in MATLAB R2021a to account for differences in the feature space across samples (e.g., due to differences in the scanner, acquisition protocol, and preprocessing of fMRI data, Table 5). Likewise, behavioral scores were standardized (z-scored) to account for differences in behavioral measures across samples (e.g., scales of autism-like measures in neurotypical samples and ADOS scores in the ASD group, see below). This cross-sample prediction approach is the most direct test of generalization of predictive neural information between neurotypical and autism groups. It also addresses potential methodological concerns related to the smaller sample size in the ASD group (i.e., concerns of limited training data in an alternative leave-one-participant-out cross-validation analysis approach in the ASD group). Permutation tests assessed the statistical significance of predictions. These analysis steps were repeated for the neural prediction of autism symptom severity in ASD (ADOS SA scores).

Statistics and reproducibility

The DS, replication sample 1 (RS1), and autism sample (ASD) were based on existing data (DS = 60, RS1 = 20, ASD = 25, see Table 1), taking advantage of their study design (why/how task: block design, see Table 5) and sample sizes. We collected a separate replication sample (RS2, $n = 55$, see Table 1) of approximately the same size as DS to replicate the effects found in the DS group. No statistical method was used to predetermine sample size. Details on data exclusion criteria are provided in the Participants section. All participant samples completed a version of the why/how fMRI task and a battery of behavior measures to characterize social functioning. No randomization was used. No blinding was applied. The statistical analyses of the data are described in the context of their respective analysis and research question in the Methods (see above). All statistical tests are 2-tailed unless stated otherwise. The study contained multiple replications of the original findings in the DS, including two neurotypical replication samples (RS1, RS2) and one autism sample (ASD). For the prediction in the smaller participant sample of the ASD group, we used a cross-sample prediction approach: we trained our model on data from all neurotypical subject samples and tested the model on data from the ASD group. This approach addresses potential methodological concerns related to the smaller sample size in the ASD group.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The brain data (fMRI) for all neurotypical data generated in this study are available here: Conte Social Inference and Context collection at https://nda.nih.gov/edit_collection.html?id=2643; fMRI data for the autism sample as can be found on the Open Science Framework (OSF, <https://doi.org/10.17605/OSF.IO/RNT8S>). Source data are provided in a de-identified form with this paper and on OSF (<https://doi.org/10.17605/OSF.IO/RNT8S>) Source data are provided with this paper.

Code availability

Behavioral and neural data were analyzed using MATLAB R2021a. fMRI data were analyzed using SPM12 (<https://www.fil.ion.ucl.ac.uk/spm/>); see hyperlinks in the manuscript for code regarding preprocessing and first-level GLMs; code underlying the MVPA analysis is openly available here¹⁰³ or upon request.

References

- Cacioppo, J. T. & Cacioppo, S. Social relationships and health: the toxic effects of perceived social isolation. *Soc. Personal. Psychol. Compass* **8**, 58–72 (2014).
- Smith, K. P. & Christakis, N. A. Social networks and health. *Annu. Rev. Sociol.* **34**, 405–429 (2008).
- Barnes, L. L., De Leon, C. M., Wilson, R. S., Bienias, J. L. & Evans, D. A. Social resources and cognitive decline in a population of older African Americans and whites. *Neurology* **63**, 2322–2326 (2004).
- Cohen, S., Doyle, W. J., Skoner, D. P., Rabin, B. S. & Gwaltney, J. M. Jr. Social ties and susceptibility to the common cold. *JAMA* **277**, 1940–1944 (1997).
- Dunbar, R. I. The social brain hypothesis. *Evolut. Anthropol.* **6**, 178–190 (1998).
- Dunbar, R. I. Neocortex size as a constraint on group size in primates. *J. Hum. Evol.* **22**, 469–493 (1992).
- DeCasien, A. R., Williams, S. A. & Higham, J. P. Primate brain size is predicted by diet but not sociality. *Nat. Ecol. Evol.* **1**, 0112 (2017).
- Rosati, A. G. Foraging Cognition: Reviving the Ecological Intelligence Hypothesis. *Trends Cogn. Sci.* **21**, 691–702 (2017).
- Dunbar, R. & Shultz, S. Why are there so many explanations for primate brain evolution? *Philos. Trans. R. Soc. B: Biol. Sci.* **372**, 20160244 (2017).
- Spagna, A. et al. Gray matter volume of the anterior insular cortex and social networking. *J. Comp. Neurol.* **526**, 1183–1194 (2018).
- Kwak, S., Joo, W.-t., Youm, Y. & Chey, J. Social brain volume is associated with in-degree social network size among older adults. *Proc. R. Soc. B: Biol. Sci.* **285**, 20172708 (2018).
- Kanai, R., Bahrami, B., Roylance, R. & Rees, G. Online social network size is reflected in human brain structure. *Proc. Biol. Sci.* **279**, 1327–1334 (2012).
- Lin, C. et al. No strong evidence that social network index is associated with gray matter volume from a data-driven investigation. *Cortex* **125**, 307–317 (2020).
- Von Der Heide, R., Vyas, G. & Olson, I. R. The social network-network: size is predicted by brain structure and function in the amygdala and paralimbic regions. *Soc. Cogn. Affect. Neurosci.* **9**, 1962–1972 (2014).
- Bickart, K. C., Wright, C. I., Dautoff, R. J., Dickerson, B. C. & Barrett, L. F. Amygdala volume and social network size in humans. *Nat. Neurosci.* **14**, 163–164 (2011).
- Beadle, J. N. et al. Amygdala but not hippocampal damage associated with smaller social network size. *Neuropsychologia* **174**, 108311 (2022).

17. Dunbar, R. I. & Shultz, S. Understanding primate brain evolution. *J. Philos. Trans. R. Soc. Lond. B: Biol. Sci.* **362**, 649–658 (2007).
18. Pérez-Barbería, F. J., Shultz, S. & Dunbar, R. I. J. Evidence for coevolution of sociality and relative brain size in three orders of mammals. *Evolution* **61**, 2811–2821 (2007).
19. Testard, C. et al. Social connections predict brain structure in a multidimensional free-ranging primate society. *Sci. Adv.* **8**, eabl5794 (2022).
20. Dziura, S. L. & Thompson, J. C. Social-network complexity in humans is associated with the neural response to social information. *Psychol. Sci.* **25**, 2095–2101 (2014).
21. Krol, S. A., Meyer, M. L., Lieberman, M. D. & Bartz, J. A. Social working memory predicts social network size in humans. *Adapt. Hum. Behav. Physiol.* **4**, 387–399 (2018).
22. Schmälzle, R. et al. Brain connectivity dynamics during social interaction reflect social network structure. *Proc. Natl Acad. Sci.* **114**, 5153–5158 (2017).
23. Noonan, M., Mars, R., Sallet, J., Dunbar, R. & Fellows, L. The structural and functional brain networks that support human social networks. *Behav. Brain Res.* **355**, 12–23 (2018).
24. Joo, W.-t., Kwak, S., Youm, Y. & Chey, J. Brain functional connectivity difference in the complete network of an entire village: the role of social network size and embeddedness. *Sci. Rep.* **7**, 4465 (2017).
25. Zou, L.-q. et al. What does the nose know? Olfactory function predicts social network size in human. *Sci. Rep.* **6**, 25026 (2016).
26. Bickart, K. C., Hollenbeck, M. C., Barrett, L. F. & Dickerson, B. C. Intrinsic amygdala-cortical functional connectivity predicts social network size in humans. *J. Neurosci.* **32**, 14729–14741 (2012).
27. Mori, K. & Haruno, M. Resting functional connectivity of the left inferior frontal gyrus with the dorsomedial prefrontal cortex and temporoparietal junction reflects the social network size for active interactions. *Hum. Brain Mapp.* **43**, 2869–2879 (2022).
28. Blumen, H. M. & Verghese, J. Gray matter volume covariance network associated with social networks in older adults. *Soc. Neurosci.* **14**, 559–570 (2019).
29. Frith, C. D. & Frith, U. The neural basis of mentalizing. *Neuron* **50**, 531–534 (2006).
30. Frith, U. & Frith, C. D. Development and neurophysiology of mentalizing. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **358**, 459–473 (2003).
31. Premack, D. & Woodruff, G. J. B., sciences, b. Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* **1**, 515–526 (1978).
32. Kihlstrom, J. F. & Cantor, N. Social intelligence. *Handb. Intell.* **2**, 359–379 (2000).
33. Singer, T. & Tusche, A. *In: Neuroeconomics* (2nd Edition) 513–532 (Elsevier, 2014).
34. Tusche, A. & Bas, L. M. Neurocomputational models of altruistic decision-making and social motives: Advances, pitfalls, and future directions. *WIREs Cogn. Sci.* **12**, e1571 (2021).
35. Tusche, A., Bockler, A., Kanske, P., Trautwein, F. M. & Singer, T. Decoding the charitable brain: empathy, perspective taking, and attention shifts differentially predict altruistic giving. *J. Neurosci.* **36**, 4719–4732 (2016).
36. Lewis, P. A., Rezaie, R., Brown, R., Roberts, N. & Dunbar, R. I. Ventromedial prefrontal volume predicts understanding of others and social network size. *Neuroimage* **57**, 1624–1629 (2011).
37. Clements, W. A. & Perner, J. J. C. D. Implicit understanding of belief. *Cogn. Dev.* **9**, 377–395 (1994).
38. Perner, J., Leekam, S. R. & Wimmer, H. J. B. Three-year-olds' difficulty with false belief: the case for a conceptual deficit. *Br. J. Dev. Psychol.* **5**, 125–137 (1987).
39. Frith, U. Mind blindness and the brain in autism. *Neuron* **32**, 969–979 (2001).
40. Stiller, J. & Dunbar, R. I. M. Perspective-taking and memory capacity predict social network size. *Soc. Netw.* **29**, 93–104 (2007).
41. Van Overwalle, F. Social cognition and the brain: a meta-analysis. *Hum. Brain Mapp.* **30**, 829–858 (2009).
42. Mar, R. A. The neural bases of social cognition and story comprehension. *Annu. Rev. Psychol.* **62**, 103–134 (2011).
43. Bzdok, D. et al. Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. *Brain Struct. Funct.* **217**, 783–796 (2012).
44. 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.* **42**, 9–34 (2014).
45. Amodio, D. M. & Frith, C. D. Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* **7**, 268–277 (2006).
46. Saxe, R., Carey, S. & Kanwisher, N. J. A. R. P. Understanding other minds: linking developmental psychology and functional neuroimaging. *Annu. Rev. Psychol.* **55**, 87–124 (2004).
47. Spunt, R. P. & Adolphs, R. Validating the why/how contrast for functional mri studies of theory of mind. *NeuroImage* **0**, 301–311 (2014).
48. Spunt, R. P. & Adolphs, R. Folk explanations of behavior: a specialized use of a domain-general mechanism. *Psychol. Sci.* **26**, 724–736 (2015).
49. Kriegeskorte, N., Goebel, R. & Bandettini, P. Information-based functional brain mapping. *Proc. Natl Acad. Sci. USA* **103**, 3863–3868 (2006).
50. Haynes, J. D. & Rees, G. Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.* **7**, 523–534 (2006).
51. Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J. & Clubley, E. The autism-spectrum quotient (AQ): evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J. Autism Dev. Disord.* **31**, 5–17 (2001).
52. Constantino, J. N. *In: Autism spectrum disorders* (eds. D. G. Amaral, G. Dawson, & D. H. Geschwind) 510–520 (Oxford University Press, Inc., 2011).
53. Ingersoll, B., Hopwood, C. J., Wainer, A. & Donnellan, M. B. A comparison of three self-report measures of the broader autism phenotype in a non-clinical sample. *J. Autism Dev. Disord.* **41**, 1646–1657 (2011).
54. Hus, V., Gotham, K. & Lord, C. Standardizing ADOS domain scores: separating severity of social affect and restricted and repetitive behaviors. *J. Autism Dev. Disord.* **44**, 2400–2412 (2014).
55. Hus, V. & Lord, C. The autism diagnostic observation schedule, module 4: revised algorithm and standardized severity scores. *J. Autism Dev. Disord.* **44**, 1996–2012 (2014).
56. Tusche, A. & Hutcherson, C. A. Cognitive regulation alters social and dietary choice by changing attribute representations in domain-general and domain-specific brain circuits. *eLife* **7**, e31185 (2018).
57. Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. & Baker, C. I. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat. Neurosci.* **12**, 535–540 (2009).
58. Kriegeskorte, N. & Bandettini, P. Analyzing for information, not activation, to exploit high-resolution fMRI. *Neuroimage* **38**, 649–662 (2007).
59. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate – a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
60. Hampton, W. H., Unger, A., Von Der Heide, R. J. & Olson, I. R. Neural connections foster social connections: a diffusion-weighted imaging study of social networks. *Soc. Cogn. Affect. Neurosci.* **11**, 721–727 (2016).

61. Kliemann, D. & Adolphs, R. The social neuroscience of mentalizing: challenges and recommendations. *Curr. Opin. Psychol.* **24**, 1–6 (2018).
62. Wimmer, H. & Perner, J. Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* **13**, 103–128 (1983).
63. Schaafsma, S. M., Pfaff, D. W., Spunt, R. P. & Adolphs, R. Deconstructing and reconstructing theory of mind. *Trends Cogn. Sci.* **19**, 65–72 (2015).
64. Beauchamp, M. S. The social mysteries of the superior temporal sulcus. *Trends Cogn. Sci.* **19**, 489–490 (2015).
65. Deen, B., Koldewyn, K., Kanwisher, N. & Saxe, R. Functional organization of social perception and cognition in the superior temporal sulcus. *Cereb. Cortex* **25**, 4596–4609 (2015).
66. Parkinson, C., Kleinbaum, A. M. & Wheatley, T. Spontaneous neural encoding of social network position. *Nat. Hum. Behav.* **1**, UNSP 0072 <https://doi.org/10.1038/s41562-017-0072> (2017).
67. Peer, M., Hayman, M., Tamir, B. & Arzy, S. Brain coding of social network structure. *J. Neurosci.* **41**, 4897–4909 (2021).
68. Campbell, D. T. & Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.* **56**, 81 (1959).
69. Constantino, J. N. The quantitative nature of autistic social impairment. *Pediatr. Res.* **69**, 55R–62R (2011).
70. Frazier, T. W. et al. Autism spectrum disorders as a qualitatively distinct category from typical behavior in a large, clinically ascertained sample. *Assessment* **17**, 308–320 (2010).
71. Pelphrey, K. Charting a course for autism biomarkers. *Biol. Psychiatry* **82**, 155–156 (2017).
72. Marquand, A. & Ecker, C. In: *Autism Imaging and Devices*. 111–136 (CRC Press, 2017).
73. Hutcherson, C. A. & Tuschke, A. Evidence accumulation, not 'self-control', explains dorsolateral prefrontal activation during normative choice. *Elife* **11**, e65661 (2022).
74. Böckler, A., Tuschke, A. & Singer, T. The structure of human prosociality differentiating altruistically motivated, norm motivated, strategically motivated, and self-reported prosocial behavior. *Soc. Psychol. Pers. Sci.* **7**, 530–541 (2016).
75. Böckler, A., Tuschke, A. & Singer, T. The structure of human prosociality revisited: corrigendum and addendum to Böckler, Tuschke & Singer, 2016. *Soc. Psychol. Pers. Sci.* **9**, 754–759 (2018).
76. Campbell, D., Tuschke, A. & Bo O'Connor, B. Imagination and the prosocial personality: mapping the effect of episodic simulation on helping across prosocial traits. *Appl. Cogn. Psychol.* **36**, 685–698 (2022).
77. Böckler, A., Tuschke, A., Schmidt, P. & Singer, T. Distinct mental trainings differentially affect altruistically motivated, norm motivated, and self-reported prosocial behaviour. *Sci. Rep.* **8**, 13560–13560 (2018).
78. Singer, T. et al. (ed.) *The ReSource Project: background, design, samples, and measurements* https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwisuPCPkJaAAxUkjlkEHTdoDNsQFnoECBQQAQ&url=https%3A%2F%2Fwww.researchgate.net%2Fprofile%2FBethany-Kok%2Fpublication%2F306270103_The_ReSource_Project_Background_design_samples_and_measurements%2Flinks%2F5d4941fba6fdcc370a7fbec9%2FThe-ReSource-Project-Background-design-samples-and-measurements.pdf&usq=A0vVaw23oJSq0CqzOQEJOKrykjCh&opi=89978449 (Leipzig, 2016).
79. Sallet, J. et al. Social network size affects neural circuits in macaques. *Science* **334**, 697–700 (2011).
80. Fair, D. A. et al. Correction of respiratory artifacts in MRI head motion estimates. *Neuroimage* **208**, 116400 (2020).
81. Gratton, C. et al. Removal of high frequency contamination from motion estimates in single-band fMRI saves data without biasing functional connectivity. *Neuroimage* **217**, 116866 (2020).
82. Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L. & Petersen, S. E. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* **59**, 2142–2154 (2012).
83. Schubert, E., Sander, J., Ester, M., Kriegel, H. P. & Xu, X. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. *ACM Trans. Database Syst. (TODS)* **42**, 1–21 (2017).
84. Pedregosa, F. et al. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
85. Baron-Cohen, S., Wheelwright, S., Robinson, J. & Woodbury-Smith, M. The adult Asperger assessment (AAA): a diagnostic method. *J. Autism Dev. Disord.* **35**, 807 (2005).
86. Lord, C., DiLavore, P. C. & Gotham, K. *Autism diagnostic observation schedule*. (Western Psychological Services, 2012).
87. Brainard, D. H. & Vision, S. The psychophysics toolbox. *Spat. Vis.* **10**, 433–436 (1997).
88. Box, G. E. & Cox, D. R. An analysis of transformations. *J. R. Stat. Soc. Ser. B (Methodol.)* **26**, 211–252 (1964).
89. Ashburner, J. A fast diffeomorphic image registration algorithm. *Neuroimage* **38**, 95–113 (2007).
90. Esteban, O. et al. fMRIprep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–116 (2019).
91. Grinband, J., Wager, T. D., Lindquist, M., Ferrera, V. P. & Hirsch, J. Detection of time-varying signals in event-related fMRI designs. *Neuroimage* **43**, 509–520 (2008).
92. Power, J. D. et al. Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* **84**, 320–341 (2014).
93. Diedrichsen, J. & Shadmehr, R. Detecting and adjusting for artifacts in fMRI time series data. *Neuroimage* **27**, 624–634 (2005).
94. Haynes, J. D. et al. Reading hidden intentions in the human brain. *Curr. Biol.* **17**, 323–328 (2007).
95. Heinzle, J. et al. Multivariate decoding of fMRI data. *Neuroforum* **18**, 1–16 (2012).
96. Tuschke, A., Bode, S. & Haynes, J. D. Neural responses to unattended products predict later consumer choices. *J. Neurosci.* **30**, 8024–8031 (2010).
97. Schulreich, S., Tuschke, A., Kanske, P. & Schwabe, L. Altruism under stress: cortisol negatively predicts charitable giving and neural value representations depending on mentalizing capacity. *J. Neurosci.* **42**, 3445–3460 (2022).
98. Wisniewski, D., Reverberi, C., Tuschke, A. & Haynes, J. D. The neural representation of voluntary task-set selection in dynamic environments. *Cereb. Cortex* **25**, 4715–4726 (2015).
99. Kahnt, T., Park, S. Q., Haynes, J. D. & Tobler, P. N. Disentangling neural representations of value and salience in the human brain. *Proc. Natl Acad. Sci. USA* **111**, 5000–5005 (2014).
100. Tuschke, A., Smallwood, J., Bernhardt, B. C. & Singer, T. Classifying the wandering mind: revealing the affective content of thoughts during task-free rest periods. *Neuroimage* **97**, 107–116 (2014).
101. Adolphs, R. & Tuschke, A. From faces to prosocial behavior: cues, tools, and mechanisms. *Curr. Dir. Psychol. Sci.* **26**, 282–287 (2017).
102. Corradi-Dell'Acqua, C., Tuschke, A., Vuilleumier, P. & Singer, T. Cross-modal representations of first-hand and vicarious pain, disgust and fairness in insular and cingulate cortex. *Nat. Commun.* **7**, 10904 (2016).
103. Hebart, M. N., Görgen, K. & Haynes, J.-D. The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Front. Neuroinform.* **8**, 88 (2014).
104. Spunt, B. whyhowlocalizer: First Release (for purposes archiving and making citable on Zenodo) (v1.0.), <https://zenodo.org/record/50244> (2016).

105. Nair, R. adolphslab/LOI2: Archival release June 2023 (Archival), <https://zenodo.org/record/8015371> (2023).
106. Spunt, B. spunt/bspm: Release for archiving purposes, <https://zenodo.org/record/8011875> (2023).

Acknowledgements

This research was supported by funding from NIMH Conte Center P50 MH094258 (R.A.) and the Natural Sciences and Engineering Research Council of Canada RGPIN-2019-04329 (A.T.). We sincerely thank Tim Armstrong for his support during data collection.

Author contributions

A.T., R.P.S. and R.A. authors designed the study. A.T., R.P.S., L.K.P., J.M.T. and R.A. wrote the manuscript, and A.T., R.P.S. and J.M.T. performed data analysis.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-40078-3>.

Correspondence and requests for materials should be addressed to Anita Tusche.

Peer review information *Nature Communications* thanks James Morris, Giorgia Silani and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023