Title (50-word maximum): Seeing social: A neural signature for conscious perception of social interactions

Abbreviated title (50-character maximum): A neural signature for conscious social perception

Author names and affiliations, including postal codes:

Rekha S. Varrier[1] and Emily S. Finn[1]

[1]Department of Psychological and Brain Sciences, Dartmouth College, Hanover 03755

Corresponding author email address: Rekha S. Varrier, Rekha.S.Varrier@dartmouth.edu;

Emily S. Finn, emily.s.finn@dartmouth.edu

Number of pages: 65

Number of:

1. Figures – 7
2. Tables – 3

Number of words for:

- Abstract – 250
- Introduction – 647
- Discussion – 1273

Conflict of interest statement

The authors declare no competing financial interests.

Acknowledgments

This project was supported by a NARSAD Young Investigator Award (grant number 28392) from the Brain & Behavior Research Foundation (E.S.F.), a Neukom CompX Faculty Grant from the Neukom Institute for Computational Science at Dartmouth College (E.S.F.), and by grant number R01MH129648 from the National Institute of Mental Health (E.S.F.).

Seeing social: A neural signature for conscious perception of social interactions

Rekha S. Varrier and Emily S. Finn

Abstract (250/250 words)

Social information is some of the most ambiguous content we encounter in our daily lives, yet in experimental contexts, percepts of social interactions—i.e., whether an interaction is present and if so, the nature of that interaction—are often dichotomized as correct or incorrect based on experimenter-assigned labels. Here, we investigated the behavioral and neural correlates of subjective (or "conscious") social perception using data from the Human Connectome Project in which participants (n = 1049; 486 men, 562 women) viewed animations of geometric shapes during fMRI and indicated whether they perceived a social interaction or random motion. Critically, rather than experimenter-assigned labels, we used observers' own reports of "Social" or "Non-social" to classify percepts and characterize brain activity, including leveraging a particularly ambiguous animation perceived as "Social" by some but "Non-social" by others to control for visual input. Behaviorally, observers were biased toward perceiving information as social (versus non-social), and neurally, observer reports (as compared to experimenter labels) explained more variance in activity across much of the brain. Using "Unsure" reports, we identified several regions that responded parametrically to perceived socialness. Neural responses to social versus nonsocial content diverged early in time and in the cortical hierarchy. Lastly, individuals with higher internalizing trait scores showed both a higher response bias towards "Social" and an inverse relationship with activity in default-mode and visual association areas while scanning for social information. Findings underscore the subjective nature of social perception and the importance of using observer reports to study percepts of social interactions.

Significance Statement (113/120 words)

Simple animations involving two or more geometric shapes have been used as a gold standard to understand social cognition and impairments thereof. Yet experimenter-assigned labels of what is social versus non-social are frequently used as a ground truth, despite the fact that percepts of such ambiguous social stimuli are highly subjective. Here, we used behavioral and fMRI data from a large sample of neurotypical individuals to show that participants' responses reveal subtle behavioral biases, help us study neural responses to social content more precisely, and covary with internalizing trait scores. Our findings underscore the subjective nature of social perception and the importance of considering observer reports in studying its behavioral and neural dynamics.

Introduction (647 /650 words)

A remarkable feature of human perception is how quickly and automatically we identify social information in the environment: consider pareidolia (seeing illusory faces in everyday objects; (Palmer & Clifford, 2020) or the cocktail party effect (perceiving self-relevant cues in otherwise unattended information streams; Wood & Cowan, 1995).

In the brain, the superior temporal sulcus (STS) has been classically associated with social cognition. Posterior STS regions are involved in perceiving animacy (Lee et al., 2014) and determining the nature of interactions (Isik et al., 2017), while anterior regions are involved in mentalizing, language and gaze detection (Deen et al., 2015). Social signal detection, however, may begin even earlier in the lateral occipital and infero-temporal regions, where recent work has proposed Gestalt-like perceptual mechanisms (i.e., grouping *social* entities like facing dyads; Walbrin & Koldewyn, 2019; Abassi & Papeo, 2020; Papeo, 2020; Landsiedel et al., 2022). The recently proposed third visual stream (Pitcher & Ungerleider, 2021) posits a specialized pathway

for social information that connects primary visual cortex (V1) to the motion-processing region (V5/MT) and culminates in the STS.

Our tendency to spontaneously perceive social interactions in simple animations of geometric shapes emphasizes the relevance of motion to social perception (Heider & Simmel, 1944; Scholl & Tremoulet, 2000). This phenomenon transcends age and culture (Barrett et al., 2005; Mohammadzadeh et al., 2012), but perhaps not species (Schafroth et al., 2021). Although robust, percepts of these animations do vary across individuals. People with autism are less likely to report social interactions (Abell et al., 2000) and show commensurately lower brain activity in social processing regions (Castelli, 2002; Kana et al., 2015). Even neurotypical individuals differ in their socio-perceptual tendencies (Rasmussen & Jiang, 2019; Li et al., 2020) in ways that covary with traits like loneliness, anxiety, and autism-like phenotypes (Kanai et al., 2012; Powers et al., 2014; Lisøy et al., 2022).

Past work has largely used stimuli handcrafted to be perceived as social or non-social, and relied on these experimenter-assigned labels to contrast behavior and/or brain activity. Consequently, effects could reflect not only responses to social information, but also differences across animations in basic physical properties (e.g., speed), which are rarely systematically controlled. Further, this approach ignores the fact that social perception is inherently subjective, even when labels are based on objective physical properties (Tremoulet & Feldman, 2000; Blakemore et al., 2003; Walbrin et al., 2018), and treats deviations from the intended percept as errors. Observer-based labels have been used in behavioral studies of single-agent biological motion (Davis & Gao, 2004; Johnson & Tassinary, 2005) and fMRI studies with non-social (Hebart et al., 2012) or social (Petrini et al., 2014; Nguyen et al., 2019) stimuli, although observer labels are rarely used to probe what constitutes a “Social” stimulus in the first place. Therefore,

here we eschew the assumption of experimenter labels as ground truth and use observer reports to more decisively isolate brain activity associated with perceiving social interactions.

We used a large dataset (n = 1049 healthy adults) from the Human Connectome Project to investigate the behavioral and neural correlates of subjective (or "conscious") social perception. We leveraged two unique features of this dataset: (1) a single animation that yielded high variability in reported percepts, allowing us to isolate neural responses to conscious social perception while holding visual input constant; and (2) a task design that permitted participant reports of "Unsure" as an intermediate between "Social" and "Non-social", allowing us to identify brain regions whose activity scaled with the extent of perceived social content. Results revealed that people show a slight behavioral bias toward perceiving information as social, and that observer responses explain more variance in activity than experimenter labels in many brain regions. Occipital, temporal and prefrontal regions showed higher responses to social information, and these differences emerged early in time and in the cortical hierarchy. Finally, internalizing traits influenced both behavior and brain activity during social signal detection.

Materials and Methods

We primarily used data from the *social cognition task* of the Human Connectome Project (henceforth referred to as the "HCP study" or "HCP dataset"; Van Essen et al., 2013). The dataset is openly accessible, and consists of a large sample of neurotypical individuals, enabling us to study both the dominant and non-dominant percepts for specific animations. The social task was one of seven cognitive tasks that were run as part of the HCP task battery (Barch et al., 2013). In this task, participants watched ten 20s animations, of which five each were considered generally social and generally non-social (experimenter-assigned labels of Mental and Random,

respectively). At the end of each animation, participants indicated whether they perceived a social interaction by pressing buttons ("Social", "Non-social", "Unsure"). To distinguish experimenter-assigned labels from observer responses, in this paper we use the terms Mental and Random for the former, and "Social", "Unsure" and "Non-social" for the latter. In the HCP dataset, participants also completed trait-level questionnaires, which enable the study of individual differences. Here, we focused on internalizing symptoms, which include anxiety, loneliness, and social withdrawal (details below in section *Correlation between traits, behavior, and neural activity*).

As participants had to wait until the end of each 20s-long animation to make a response, the behavioral data in the HCP does not reveal *when* the perceptual decisions were made, and any differences in decision time are likely to influence the trajectory of brain activity during each trial. Hence, we additionally performed an online study on 100 neurotypical individuals (henceforth referred to as the "online RT experiment") to gain insight into when in the course of the animation-watching decisions might have been made, and how this varied across animations and individuals.

Participants

Data from the HCP social cognition task are publicly available in the online HCP repository (https://db.humanconnectome.org/; for each participant, fMRI data sub-folders: *tfMRI_SOCIAL_RL* and *tfMRI_SOCIAL_LR*; behavioral: **TAB.txt*). Trait scores used to study individual differences were from the restricted category. In the demographic data reported below, the age was obtained from the restricted and gender from the unrestricted category. We obtained complete fMRI data from 1049 individuals for the HCP social cognition task (ages 22-37; 562 female and 486 male). Of these, 823 participants responded on all trials in a reasonable response time (RT > 100ms) and were included in the behavioral data analyses. For the various fMRI analyses, depending on the comparison, participants with incomplete data were excluded. Thus,

for the various fMRI analysis, we had n = 777, 870 and 814 for comparisons involving RANDOM MECH, COAXING–BILLIARD and the ALL animations, respectively (see *fMRI data analysis* sub-section *"Social" vs, "Non-social*" for details). Lastly, for the trait-behavior analysis, we included all participants who had complete behavior and trait data (n = 817), and for both the trait-fMRI and trait-behavior-fMRI analyses, we included participants with behavior, fMRI and trait data (n = 812).

For the online RT experiment that we conducted in July 2021, we recruited 100 neurotypical individuals (ages 18-48, M = 23.2, SE = 0.64). from the United States and United Kingdom via the online platform Prolific (www.prolific.co , Palan & Schitter, 2018). Prior to the experiment, all participants read and acknowledged the virtual consent forms in accordance with the Institutional Review Board of Dartmouth College, Hanover, New Hampshire, USA. Participants with good-quality data (n = 90) were used in preliminary analyses and out of these, n = 83 were used to supplement the COAXING–BILLIARD fMRI data analyses. See sub-section *Data acquisition and pre-processing* for details on the selection criteria.

Stimuli

Stimuli in the HCP study were ten 20-second-long animations chosen from previous studies (Castelli et al., 2000; Wheatley et al., 2007). Longer animations had been trimmed to 20s by the HCP researchers (Barch et al., 2013). The animations were presented in two runs with five animations each (run duration 3min 27s) interleaved with fixation blocks of 15s without jitter. The order of presentation was maintained across all participants (see Table 1). The number of Mental (M) and Random (R) animations were balanced within and between runs (run 1: 2M, 3R; sequence M–R–R–M–R; run 2: 3M, 2R; sequence M–M–R–M–R. For a list of the animations as provided

by the HCP and their properties, see Table 1. Note that in this paper, we drop the suffixes in the filenames ("-A" and "-B") for brevity.

Each animation consisted of two or more shapes in motion ("agents") with or without stationary elements ("props"). Seven of them (3M, 4R) had a large red and a smaller blue triangle as agents, and the remaining three (FISHING, RANDOM MECH, and SCARING) were more diverse in the number, color, and/or form of agents and props.

For the online RT experiment, we presented the same animations used in the HCP study and in the same presentation sequence, with a self-timed break after the fifth stimulus in lieu of the break between the two runs in the HCP study. In the practice phase, we randomly showed either a generally social or non-social animation (that was not one of the 10 animations used in the main task) to each participant. For a social practice example, we used MOCKING–B from the HCP repository, and for a non-social practice example, we created a two-agent animation comparable in appearance to MOCKING–B using a custom app Psyanim (the latter available here: https://github.com/rvarrier/HCP_socialtask_analysis/tree/main/stimuli – the repository will be made public on publication. In the meantime, please get in touch with us for the file).

The differences in physical properties that we noted above amongst the HCP animations could have influenced both behavior and brain activity. Hence, we factored these into our analyses either by comparing the brain activity for "Social" and "Non-social" responses *within* the same animation (i.e., same visual input) or by regressing out physical properties like the optic flow and mean brightness before comparing individual pairs of animations in the analysis comparing timecourses (explained in the sub-section *fMRI Timecourse Analysis* under *fMRI data analysis*). The presence of these visual differences also motivated our decision to perform the online RT experiment to estimate decision times and select a pair of animations with similar decision times

(details in the *fMRI data analysis* section). Lastly, we also included animation as a grouping variable ("random effect") in certain behavioral and fMRI data analyses when pooling data from multiple animations.

Table 1.

Detailed information about the experimental stimuli used in the HCP social cognition Task.

Run number	**Presentation sequence**	**Animation file names (.AVI extension)**	**Experimenter-assigned category**	**Description of agents**
1, *tfMRI_SOCIAL_RL*	1	COAXING-B	Mental	Bigger red triangle, smaller blue triangle.
	2	BILLIARDS-A	Random	
	3	DRIFTING-A	Random	
	4	FISHING	Mental	1 circle with a "fishing pole", 1 oblong-shaped "fish"
	5	RANDOM MECHANICAL	Random	4 circles, 1 triangle, 1 long rectangle of multiple colors
2, *tfMRI_SOCIAL_LR*	1	SCARING	Mental	4 circles: 3 small pink, 1 large blue
	2	SEDUCING-B	Mental	Bigger red triangle, smaller blue triangle.
	3	STAR-A	Random	
	4	SURPRISING-B	Mental	

	5	TENNIS-A	Random	

Experimental design

In the HCP study, participants were given the following instructions about the task: *"You will now watch short clips and decide if the shapes are having a mental interaction or not. For a mental interaction, press the button under your index finger. If you are not sure, press the button under your middle finger. For a random interaction, press the button under your ring finger. After each clip, there will be a response slide. Please respond while that slide is on the screen."* They had three seconds to respond. In our online RT experiment, participants were given similar instructions, but were asked to respond twice to each animation: once *during* the animation as soon as they made a decision (left/right arrows for "Social"/ "Non-social") and a second time immediately *after* each animation within 3 seconds (left/right/down arrows for "Social"/ "Non-social"/ "Unsure" similar to the HCP study).

Data acquisition and pre-processing

HCP social cognition task dataset

Behavioral data: In analyzing the behavioral data, we included only participants who responded to all 10 animations and in whom the response times (RT) were not unrealistically short (i.e., RTs < 100ms were excluded), resulting in n = 823. Note that even if participants had arrived at a decision before the end of the video, we still need to account for the time taken to perceive the appearance of the response screen before responding (Gottsdanker, 1982).

fMRI data: The fMRI data were acquired using a 3T Skyra scanner with 2mm isotropic voxels and a TR of 0.72s (see Barch et al., 2013) for more acquisition details). Each run comprised 274 scan

volumes, and there were two runs per participant. We used minimally preprocessed voxel-wise fMRI data (Glasser et al., 2013), parcellated this into 268 parcels spanning the whole brain as per Shen atlas (Shen et al., 2013) and discarded the first five scan volumes (TRs) within each run to reduce initial artifacts. Next, to make BOLD response magnitudes comparable across participants, we z-scored parcel-wise timecourses in each run. Further, since our analyses were to be performed at the trial-level, we split the run time series into trial-wise timecourses of 40s each – i.e., 20s animations (28 TRs) flanked by 10s fixation periods (14 TRs) on either side (except for the first animation within each run which included only 6 pre-stimulus TRs). Data preprocessed in this manner were used for all fMRI analyses except one (the timecourse analysis, explained later) which required comparing two *individual* animations: COAXING and BILLIARD. For the timecourse analysis, the z-normalization was done at the individual trial-level, to remove differences in mean activity that were due to the order of presentation (since order was not randomized between participants). In both cases, we lastly baseline-corrected each trial timecourse by subtracting the signal magnitude at the trial onset (i.e., the TR immediately before stimulus onset).

Online RT experiment

In the online RT experiment, we excluded trials in which either of the two responses ("during" phase and "after" phase) were missing or where the two responses were not congruent (i.e., participants changed their response on watching the full animation). The latter was done to ensure that the response time from the "during" phase corresponded to the percept reported in the end, to match the HCP task; however, note that the two are not perfectly comparable since in the "during" phase participants did not have a choice to respond "Unsure". Lastly, as a quality check,

participants with fewer than 8 out of 10 good-quality (i.e., congruent) responses were also excluded, giving us 90 participants (n = 34, 33 and 23 with 10/10, 9/10 and 8/10 congruent responses, respectively). Based on decision times, the animation pair COAXING and BILLIARD were used in fMRI analyses to contrast "Social" and "Non-social" perception (see the sub-sections *"Social" vs. "Non-social"* and *fMRI timecourse analysis*). To estimate the decision time here, we used data from 83 of the 90 participants – who responded "Social" to COAXING and "Non-social" to BILLIARD congruently. The remaining 7 participants either missed a response, did not give a congruent response to *both* animations or did not respond to COAXING (BILLIARD) as "Social" ("Non-social").

Behavioral data analysis

Using the behavioral data from the HCP, we performed four analyses to measure whether there was a general bias toward social percepts, or in other words, a shift towards "Social" responses. Our dependent variables were:

(1) Percentages of "Social" and "Non-social" responses within participants; compared using a paired t-test.

(2) Decision criterion (c), the signal detection theory metric quantified as $\frac{-(Z(Hit\ rate)+Z(False\ alarm\ rate))}{2}$ (Stanislaw & Todorov, 1999), where *Hit rate* and *False alarm rate* were computed for each participant as fractions of "Social" responses for animations labelled by the experimenters as Mental and Random, respectively. Note that we do not compute other signal detection theory metrics like *d'* and bias. In this case, *d'* would be a measure of conformity to the experimenter labels (which is not of interest given our theoretical framework), and bias would be largely redundant with c, which already

quantifies the relative magnitude of "Social" compared to "Non-social" responses. Further, *c* is preferable to bias because it is more independent from *d'* (Banks, 1970).

(3) Response time (RT) differences between "Social" vs. "Non-social" trials. We compared RTs using both a non-parametric paired (Wilcoxon signed-rank) test and a more controlled linear mixed effects (LME) analyses to further account for the differences between individual animations. The LME model (LMEM) was of the form: $log(RT) = f(response;\ random\ intercepts: participant, animation)$. The factor *response* was categorical with two levels: "Non-social" (coded as the base level) and "Social", and analysis was performed using the Python package *pymer4* (Jolly, 2018). We used the logarithm of the RT in seconds to bring the residuals of the LMEM closer to a normal distribution (which is an assumption for LMEMs).

(4) Percentage of "Unsure" responses for the two animation labels (Mental, Random). These were compared using a logistic regression model: $uncertainty = f(stimLabel;\ random\ intercepts: participant, animation)$ where the factor *stimLabel* was categorical [Mental, Random], and the dependent variable *uncertainty* had a value of "1" for "Unsure" response trials and "0" otherwise. Keeping Random (0) as the baseline in the analysis, positive (negative) regression coefficients for *stimLabel* would indicate a lower (higher) uncertainty in categorizing Random trials.

fMRI data analysis

GLM-based regression: Our primary approach to fMRI data analysis was a general linear model (GLM) based on animation onset and offset. We computed the regression coefficients for each animation separately for the majority of analyses. For each animation, we fitted each parcel's activity timecourse to a "slope" regressor (line steadily increasing from 0 to 1 from stimulus onset

to offset [duration = 20 s], padded by zeros before and after) that was convolved by the Glover hemodynamic response function (HRF; Glover, 1999). (Preliminary analyses indicated that a steadily increasing slope regressor captured more variance in the BOLD data than a traditional boxcar regressor.) This renders one slope regression coefficient (β) per parcel, participant, and trial (animation). We also performed a separate GLM analysis across all animations (details in the section below). For this analysis, we used a *run*-level regressor and estimated coefficients for each parcel, participant, and *run*. Similar to the slope regressors used at the trial level, regressor values increased (decreased) steadily *during* an animation labelled "Social" ("Non-social"), and were 0 at all other timepoints (including "Unsure" responses); thus, the run-level regression coefficient here summarizes a *contrast* between "Social" and "Non-social". For each participant, we then averaged these coefficients across the two runs.

"Social" vs. "Non-social": To identify brain regions showing a consistent and generalizable difference between "Social" and "Non-social" responses, we compared the regression coefficients between "Social" and "Non-social" percepts in three analyses: (1) controlled for visual input, (2) controlled for decision times and (3) across all animations (Table 2). For analyses with individual animations, we included all participants who gave a valid response to the animation(s) in that analysis, resulting in slightly different numbers of participants in each analysis (for an overview, see sub-section *Participants*). Each analysis is described in detail below:

(1) **Controlled for visual input**: We selected the most ambiguous animation, namely RANDOM MECH, since it had the relatively most balanced "Social" and "Non-social" response groups. We excluded participants who gave an "Unsure" response to this stimulus (leaving $n = 777$) and then split regression coefficients based on observer responses ("Social": $n = 107$, "Non-social": $n = 670$, see Figure 1a), and compared them with two-sample t-tests assuming unequal

variances. While the individual groups of responders are not balanced, the actual number of individuals who responded "Social" to RANDOM MECH is still higher than traditional studies that have studied social perception with animations.

(2) **Controlled for difficulty/ambiguity (COAXING vs. BILLIARD)**: We chose two animations which were most comparable in their "difficulty" or ambiguity, as proxied by two measures: (1) the relative proportions of dominant and non-dominant responses, and (2) the time taken to arrive at a response. We used McNemar's statistic (McNemar, 1947) to compare the relative proportions of dominant ("Social" and "Non-social" for COAXING and BILLARD, respectively) and non-dominant responses ("Non-social" and "Social" for COAXING and BILLARD, respectively) in the HCP dataset (dominant: COAXING $n = 886$, BILLIARD $n = 876$; non-dominant: COAXING $n = 6$, BILLIARD $n = 16$; continuity correction performed) as well as the online RT experiment (dominant: COAXING $n = 84$, BILLIARD $n = 83$; non-dominant: COAXING $n = 0$, BILLIARD $n = 1$; exact binomial distribution used due to the extremely small sample sizes in the non-dominant group), and found that the proportions were not significantly different in either case (details in the *Results* sub-section on decision time).Response times were based on the data we obtained from the online RT experiment, where the decision time to report "Social" to COAXING (*median* = 3.45s, *SE* = 0.27s) and "Non-social" to BILLIARD (*median* = 3.7s, *SE* = 0.25s) were the closest and did not significantly differ (see Figure 2c and the *Results* sub-section on decision time). Hence, we compared regression coefficients for each of these two animations within participants using a paired t-test. Note that we excluded participants who gave an uncertain or non-dominant response for one or both animations (i.e., who responded to COAXING as

"Non-social" or "Unsure" or BILLIARDS as "Social" or "Unsure"), giving us n = 870 for this analysis.

(3) **Across all animations (ALL):** We also performed a more general comparison between brain activity associated with "Social" vs. "Non-social" responses by identifying regions that showed a mean run-level regression coefficient that was different from 0 per a one-sample t-test (for details on how the run-wise regressor was estimated, see sub-section *GLM-based regression* above). To minimize biases due to missed responses, we used only participants who had given all 10 responses and had complete fMRI data from both runs (n = 814).

Lastly, we identified brain regions that were significant in all three of the above comparisons and showed changes in the same direction (either "Social" > "Non-social" in all three comparisons or vice versa) at a corrected threshold (false-discovery rate [FDR] $q < .05$, corrected for 268 comparisons [parcels]). We henceforth refer to this procedure as the "intersection analysis" and the resultant parcels as "robust social perception regions".

Comparison between using observer responses and experimenter-assigned labels

To evaluate whether observer responses actually explain more variance in the fMRI data than experimenter-assigned labels, we also ran two LMEMs for each brain parcel. The dependent variable was the trial-level regression coefficient (β) with run-normalized BOLD data (see *Data acquisition and pre-processing* for details), and both models included subject ID and animation as random effects. The fixed effect in the first model was the observer responses ($\beta = f(response;\ random\ intercepts: participant, animation)$; *response* = {"Social", "Non-social"}), and in the second model, it was the experimenter-assigned labels ($\beta = f(stimLabel;\ random\ intercepts: participant, animation)$; *stimLabel* = {Mental,

Random}). These models are referred to in subscripts as *Obs* and *Exp*, respectively. We then assessed which label type explained more variance in the data by taking a difference between the Akaike information criteria (AIC) for each model. Lower AICs indicate better model fits, so $AIC_{Obs} < AIC_{Exp}$ indicates that the response-based model is better and vice-versa. We identified parcels for which models differed in their AIC fits by at least 10, which corresponds to a relative likelihood of 99.32% for the model with the better fit (Wagenmakers & Farrell, 2004).

Table 2.

Details of the three "Social" vs. "Non-social" comparisons (based on observer reports) performed as part of the GLM analysis.

Analysis Nr.	**"Social" responder group**	**"Non-social" responder group**	**Rationale for the analysis**	**Statistical comparisons for fMRI**	**ID for the analysis used in figures and text**
1	RANDOM MECHANICAL (Within-animation, between-participant)		• Controls for low-level input • Most ambiguous animation	Two-sample t-test	RANDOM MECH
2	COAXING	BILLIARD	• Likely similar decision times *during* animation (as estimated from online RT experiment)	Paired t-test	COAXING-BILLIARD
	(Between-animation, within-participant)				

3	All "Social"	All "Non-social"	• Maximizes power by comparing all "Social" responses with all "Non-social" responses within participants	One-sample t-test after averaging run-wise estimates	ALL ANIMATIONS
	Coded 1 ("Social") and -1 ("Non-social"), in run-wise GLMs.				

335
336

337 ***"Social" vs. "Unsure" vs. "Non-social"***: We also leveraged the "Unsure" responses to identify
338 brain regions that responded parametrically to level of perceived socialness. We predicted that the
339 neural response in such regions during animations ultimately marked "Unsure" would be
340 intermediate to that of "Social" and "Non-social" responses. However, intermediate does not
341 necessarily mean halfway, and hence we performed conjunction analyses – i.e., we identified brain
342 regions showing "Social" > "Unsure" and "Unsure" > "Non-social" (or vice-versa) and took the
343 intersection of these. We performed this analysis across all the animations using an LMEM of the
344 form: $beta = \mathrm{f}(\mathrm{response}, \mathrm{RI{:}\ participant})$ which was performed separately for "Social" vs.
345 "Unsure" (LMEM 1) and "Unsure" vs. "Non-social" (LMEM 2). In each LMEM, *response* was a
346 categorical variable that had the values "Social" and "Unsure" in LMEM 1 (baseline "Unsure"),
347 and "Unsure" and "Non-social" in LMEM 2 (baseline "Non-social"). Thus, a positive LMEM
348 estimate for *response* would indicate a higher neural response corresponding to a higher perceived
349 socialness. From this, we identified parcels which showed the same directionality for LMEM 1
350 and 2 at the multiple comparison-corrected threshold, and which were also in the set of robust
351 social perception regions from the GLM analysis described above.

To probe whether similar parametric patterns that were seen across all animations also emerge when controlled for visual input, we again leveraged the most ambiguous animation (RANDOM MECH). We plotted the timecourses for a subset of the parcels in which "Unsure" was the closest to the halfway point between "Social" and "Non-social" both in terms of the mean regression coefficient and the magnitude of activity at the end of the stimulus presentation period (20s) for each parcel and response, the rationale being that the signal during the final timepoints of the animation should most closely reflect a participant's ultimately reported percept.

fMRI timecourse analysis*:* To identify the brain regions where the earliest differences in activity between "Social" and "Non-social" percepts emerged, we performed paired t-tests (within participant) for each timepoint (TR) between BOLD responses corresponding to a pair of "Social" and "Non-social" animations (COAXING and BILLIARD, respectively). This pair of animations was chosen because decisions as to whether the animation was social or non-social were likely made at comparable times *while* watching them as explained previously in the analysis sub-section *"Social" vs. "Non-social"*. To ensure that the differences in BOLD activity between COAXING and BILLIARD were not due to differences in basic visual input between the two animations, we performed these comparisons on the residual timecourses obtained after regressing out two low-level visual features, total optic flow and mean brightness. We first estimated these two features for each animation frame using the *pliers* software package (McNamara et al., 2017), then down-sampled the resulting timecourses to match the temporal resolution of the fMRI data (i.e., the TR), z-normalized them and convolved them with an HRF. We then performed a linear regression on each participant's trial timecourse (including 14 TRs flanking the stimulus duration on either end; same procedure as the slope regressors described earlier) to regress out the changes in BOLD activity related to these features. We then used the resultant *residual* timecourses for COAXING

and BILLIARD for the timecourse analysis. We compared these at each timepoint (TR) and for each parcel using paired t-tests (within participant). For each parcel, we thus identified the earliest timepoint at which BOLD activity begins to diverge (i.e., $p < 0.05$). As additional consistency checks, we (1) only performed this analysis in the robust social perception regions from the GLM intersection analysis, and (2) selected a TR *t* as the divergence point only if the difference between "Social" vs. "Non-social" at *t+1* was also significantly different ($p < 0.05$) in the same direction.

Note that this analysis does not factor in the hemodynamic lag. This is because although the HRF *peaks* a few seconds after an event (in our case, the animation onset), the neural responses—and corresponding start of the BOLD response—should have begun nearly instantly upon stimulus presentation (Friston et al., 1994), so here we investigated where these earliest changes could be observed. Further, in using the median decision times from the online RT experiment for COAXING and BILLIARD as the expected decision time for the HCP dataset, we did not factor in the motor response delay (i.e., time taken after a decision has been made to press a button) in the online RT experiment. Hence it is possible that some of the pre-decisional processes closer to the decision time may have in fact been post-decisional. While we cannot exclude this possibility, this was unlikely since motor responses on arriving at a decision are typically quicker than the TR used in the HCP task (0.72s).

We also did not multiple comparison-correct across timepoints in this analysis since the primary goal was to identify the *earliest* differences in activity, and to infer this correctly, false negatives are undesirable. Still, in identifying the earliest timepoints, we only selected a region if the subsequent timepoint was also significant ($p < 0.05$ uncorrected), limiting the odds of a false positive further by 95%.

We also did not perform this analysis within the same animation (RANDOM MECH) and across all animations like in the GLM analysis (sub-section *"Social" vs. "Non-social"*) because of the heterogeneity in decision times for RANDOM MECH and across animations (see Figure 2c). This heterogeneity means that the neural processes at each time point could have been vastly different between individuals and animations, thus clouding any potential interpretations from these comparisons.

Correlations between traits, behavior, and neural activity

Past work has shown that within the neurotypical population, social perception covaries with traits like loneliness, anxiety, psychopathy and autism-like phenotypes (Sacco et al., 2016; Lessard & Juvonen, 2018; Desai et al., 2019; Abassi & Papeo, 2022; Lisøy et al., 2022; Williams & Chakrabarti, 2022). In particular, individuals high on internalizing traits such as loneliness and anxiety are more sensitive to social cues (Gardner et al., 2005), tend to form illusory social connections by anthropomorphizing inanimate objects (Epley et al., 2008; Powers et al., 2014) and show smaller grey matter volumes in a brain region typically associated with social processing, the pSTS (Kanai et al., 2012). Here, we probed whether internalizing traits affect behavior and/or brain activity associated with social perception using the internalizing T-score provided by the HCP (Barch et al., 2013). This score is based on participants' responses to the internalizing dimension questions from the Achenbach Adult Self-Report questionnaire (ASR; Achenbach et al., 2017). Internalizing symptoms refer to symptoms like anxiety, depression, and withdrawal, and are typically contrasted with externalizing behaviors such as rule-breaking and aggression. The ASR was designed to assess behavioral, emotional, and social functioning across a wide spectrum of the population, so it is sensitive to individual differences (i.e., produces a range of scores) even in healthy/subclinical populations. Specifically, we used the participant-level

internalizing T-score (labelled *ASR_Intn_T* in the HCP dataset) which is normalized for age and sex (*M* = 48.72, *STD* =10.75, *range* = 30-97; see Figure 7a-c for the full distribution) and which sums across the three ASR scales *Anxious/Depressed*, *Withdrawn* and *Somatic Complaints*.

To assess whether internalizing score relates to a behavioral bias toward "Social" percepts, we compared participants' internalizing scores with the following behavioral variables: (1) the difference between % of "Social" and % of "Non-social" responses (calculated as percentages to control for missing data; Spearman (rank) correlation r_s); (2) responses to the most ambiguous animation, RANDOM MECH; specifically comparing "Non-social" to "Social" or "Unsure" responders (two-sample t-test); and (3) the percent of "Unsure" responses on Mental versus Random trials (Spearman correlation). We tested the specificity of these correlations by additionally performing correlations with externalizing scores (*ASR_Extn_T* in the HCP dataset) and comparing the strength of the relationships using the CorrelationStats package (https://github.com/psinger/CorrelationStats). For RANDOM MECH, we used the LMEM $traitScore \sim \mathrm{f}(\mathrm{respGroup}, \mathrm{traitType}, \mathrm{respGroup} * \mathrm{traitType\ interaction}, RI\colon \mathrm{participant})$ to study how the internalizing and externalizing trait scores (dependent variable: *traitScore*, factor *traitType* with two levels [internalizing, externalizing]) vary for each response group (factor *respGroup* with two levels: ["Non-social", "Social"/ "Unsure"]), and with *participant* as a random effect.

To quantify if and where internalizing traits relate to brain activity while scanning animations for social information, for each parcel, we performed an LME analysis where the dependent variable was the trial-level slope regression coefficient, the fixed factor was internalizing score and the random factor was animation. This yields brain regions that respond proportionately to internalizing score in that individual across animations and parcels.

Lastly, to check for interactions between social percepts, internalizing symptoms, and neural activity, we tested how much the difference between neural responses to "Social" and "Non-social" trials depended on internalizing symptoms. For this, we fitted regression coefficients from the GLM analysis across all animations (ALL) – which represents the contrast between activity to "Social" vs. "Non-social" responses (one estimate per participant; see sub-sections *GLM-based regression* and "*Social" vs. "Non-social"* above) – to the internalizing symptom scores (also one estimate per participant) in a linear regression.

Code availability

All the code for analyzing data from both the HCP and online RT experiment, as well as the anonymized data from the online RT experiment, will be made available upon publication here: https://github.com/rvarrier/HCP_socialtask_analysis. In the meantime, please get in touch with the authors for these.

Results

In this study, we used behavioral and fMRI data from the Human Connectome Project (HCP) social cognition task to characterize the behavioral and neural processes underlying conscious perception of social interactions. We started by evaluating the behavioral data for any response bias: are people more inclined to declare information "Social" (as opposed to "Non-social")? We next identified brain regions that robustly differentiated between "Social" and "Non-social" percepts, including a subset that showed a parametric response pattern to degrees of perceived socialness, and showed that observer responses explain more variance than experimenter-assigned labels in many regions' activity levels. Next, we used a timepoint-by-timepoint analysis to identify

where and when brain activity begins to diverge according to whether social information is subjectively perceived. Lastly, we studied the relationship between internalizing behavior scores, tendency toward social percepts, and brain activity while scanning for social information. As a reminder, here, we use the terms Mental and Random to refer to experimenter-assigned stimulus labels, and "Social" and "Non-social" to refer to observers' actual reported percepts of those stimuli.

Some animations are more ambiguous than others

First, we examined the degree to which participants' percepts of "Social" versus "Non-social" information agreed with one another as well as the intended stimulus category. In the HCP social cognition task, participants passively watched ten 20-s animations of geometric shapes (Heider-Simmel-like; Castelli et al., 2000), see *Methods* sub-section *Stimuli* for a detailed description of the animations) and then made a behavioral response — "Social", "Non-social" or "Unsure"—to indicate whether they perceived a social interaction in the animation. Five animations were intended to evoke social interactions (experimenter-assigned Mental) and five were not (experimenter-assigned Random). Although on average participants' percepts aligned with experimenter labels, the degree to which animations were perceived as "Social" and "Non-social" varied considerably. This was true in both the HCP behavioral data and the secondary online dataset (online RT experiment) we collected to study the time taken for individuals to arrive at decisions while watching each animation (Figure 1a and 2a). While animations like DRIFTING and BILLIARD were seen almost unanimously as "Non-social", animations like RANDOM MECH and FISHING had a higher percentage of the non-dominant percept as well as "Unsure" responses. This underscores the need to use participants' own percepts to categorize what is or is not "Social" rather than experimenter-assigned labels. In later analyses, we leveraged this

ambiguity by comparing neural activity corresponding to "Social" and "Non-social" responses within the most variably perceived animation (RANDOM MECH), thereby isolating activity associated with a conscious social percept while controlling for visual input.

Responses are biased toward "Social"

Next, we used behaviorally reported percepts to determine whether there was a response bias towards "Social". We hypothesized that evolutionarily, there may be a bias towards perceiving information as social, since the cost of a false positive (i.e., mistakenly thinking someone is trying to engage you in a social interaction) is lesser than that of a false negative (i.e., missing out on social cues that are important for group dynamics, reproduction, and survival). We predicted that this bias would manifest as a higher "Social" response rate, shorter response times for "Social" percepts, and more "Unsure" responses to animations labeled Random by experimenters (because of a reluctance to declare something entirely non-social). Our findings are described below:

(1) 'Social' responses are more frequent: Comparing the frequency of percepts for each participant (limited to trials where participants were sure of their response—i.e., excluding "Unsure" trials) showed that the percentage of "Social" responses was higher (M = 52.9%, SE = 0.29%) than that of "Non-social" responses (M = 47.1%, SE = 0.29%; paired t-test, t = 9.96, $p < 10^{-21}$; Figure 1b).

(2) The response criterion further shows a bias towards "Social": Next, we computed criterion (c), a metric from signal detection theory that quantifies response biases. If the mean c is significantly different from zero, this suggests a bias in responses towards "Social" ($\bar{c} < 0$) or "Non-social" ($\bar{c} > 0$). We found that criterion was significantly negative at the population level (M = –0.05, SE = 0.01; Wilcoxon test, *test statistic* = 26813, $p < 10^{-17}$; Figure 1c), further confirming the response bias towards "Social". In this computation, we used the experimenter-assigned labels

to show that although the experimenters aimed to create a balanced set of five Mental and five Random animations, actual observer reports indicate that individuals ended up perceiving more animations as "Social". Thus, percepts did not fully conform to the expectations of the experimenters.

(3) Responders may have been quicker to declare something as "Social" than "Non-social": Next, to get at a more subconscious measure of perceptual decision-making for social information, we compared response times between "Social" (*median* = 0.87s, *SE* = 0.009s) and "Non-social" (*median* = 0.9s, *SE* = 0.012s) responses (Figure. 1d) and found that "Social" responses were overall faster (Wilcoxon test, *test statistic* = 144885, $p < 10^{-3}$). Since response times could differ by animation due to their heterogeneity, we additionally performed an LME analysis with response ("Social" or "Non-social" [baseline]) as the fixed effect, and both animation and participant as random effects. We observed a trend towards shorter RTs for "Social" responses, but this did not reach significance (*Est.* = –0.04, $p = .1$).

(4) "Unsure" responses were more common for animations intended as Random compared to those intended as Mental: We studied the distribution of "Unsure" responses between animations that were intended to be "Social" (Mental) or "Non-social" (Random) and noted that there was a higher percentage of "Unsure" responses in the animations intended as Random (*M* = 9.4%, *SE* = 0.5%; Figure 1e) compared to those intended as Mental (*M* = 2.7%, *SE* = 0.3%). This indicated that people were more hesitant to label something "Non-social" (as opposed to "Social") when their confidence is low. In other words, they err on the side of false alarms rather than misses; this fits with the idea that misses are likely costlier than false alarms. We formally compared the frequency of "Unsure" responses using logistic regression with Mental (coded 1) and Random (coded 0) label as the fixed effect and participant ID and

animation as random intercepts. Results showed higher uncertainty on Random trials even after accounting for the differences in animations (*Est*. = -1.61, p =.005).

To summarize, the behavioral data overall showed a bias towards "Social" responses based on frequency of each response type, response times and degree of uncertainty.

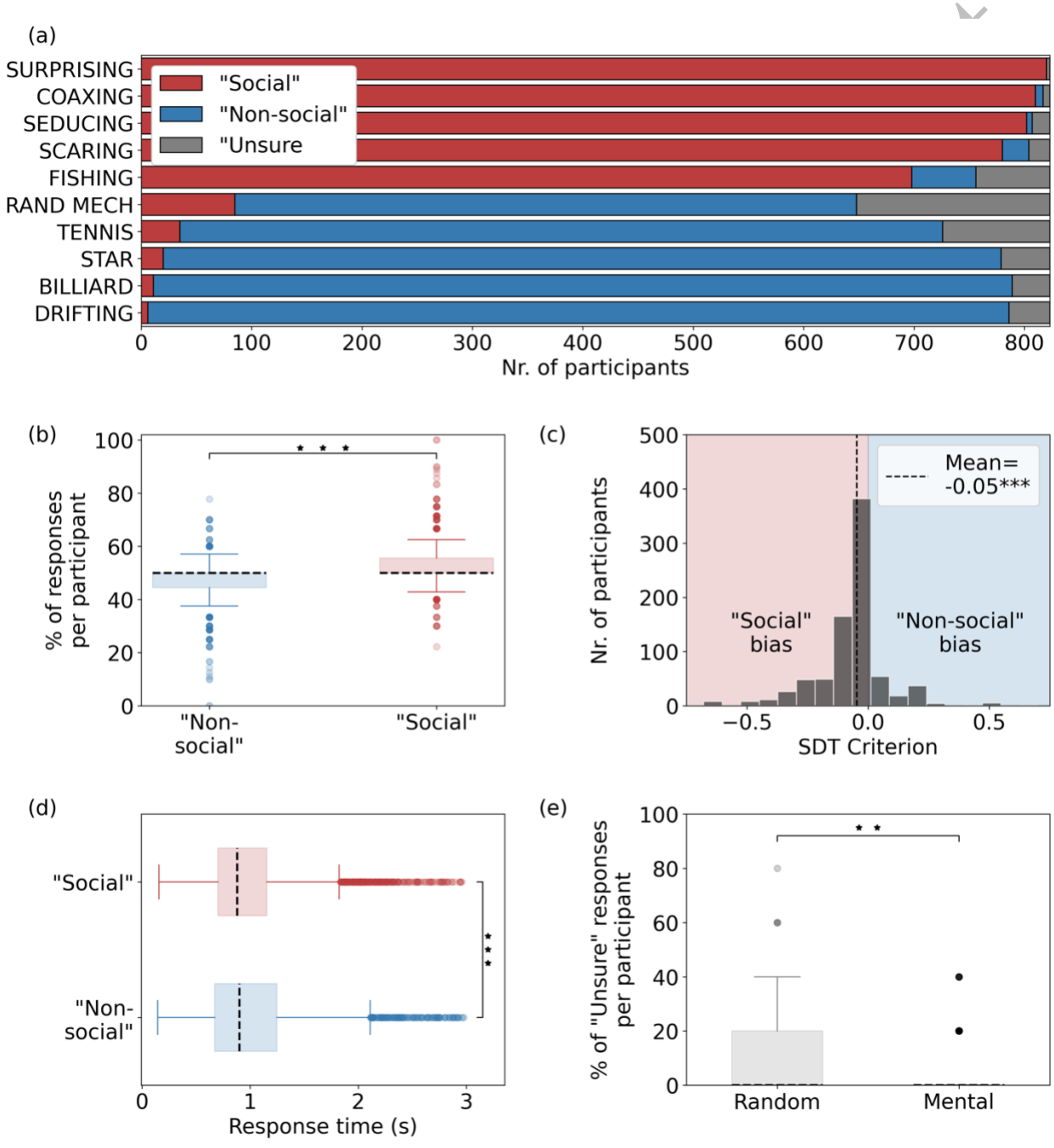


Figure 1. Behavioral data from the HCP participants (n=823) show a bias toward "Social" responses. *(a) Number of responses per type ("Social", "Non-social", "Unsure") for each animation (sorted from most to least "Social"). (b) Percentages of "Social" and "Non-social" responses. "Social" responses were more frequent (*$t = 9.96$, $p < 10^{-21}$*, paired t-test). (c) Signal detection theory metric* criterion c *across participants based on experimenter-assigned labels. Mean criterion was negative (* $\bar{c} = -0.05$*, Wilcoxon signed rank test statistic* = 26813, $p < 10^{-17}$*), indicating a bias toward false alarms (i.e., declaring an animation labeled Random by experimenters as "Social"). (d) Response time for "Social" and "Non-social" responses. "Social" responses tended to be quicker (Wilcoxon signed-rank test statistic* = 144885, $p < 10^{-3}$*). (e) "Unsure" responses for animations labelled Mental and Random by experimenters. There was a higher percent of "Unsure" responses for Random responses (LMEM:* Est. = –2.15, p < .005). ***: p<.001, ***: p < .0001.*

Decision time as to whether an animation is "Social" varies widely between individuals and animations

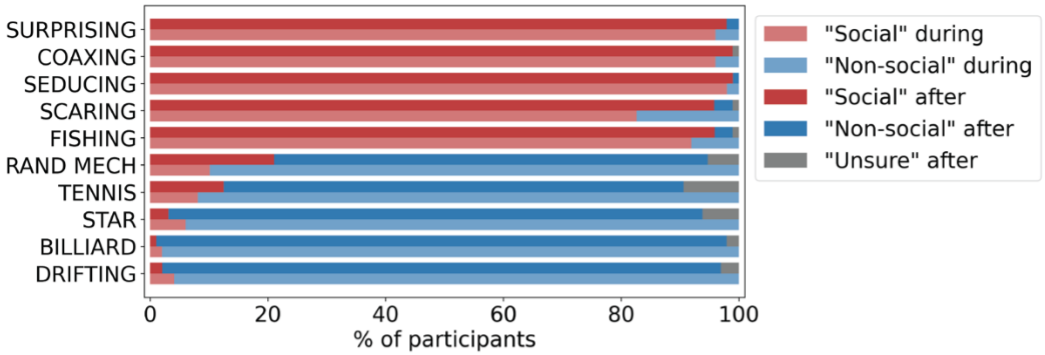
In the HCP study, participants had to wait until the end of each animation (lasting 20 s) to make a behavioral response. However, the decision as to whether an animation was "Social" or "Non-social" was presumably made sometime during passive viewing, although the decision time could have varied widely across animations and participants. This variability, in turn, might influence the timecourse of brain activity (e.g., visual attention for the same animation may be different when a participant makes a decision 2 seconds into the animation vs. 15 seconds into it). Hence, getting information as to when decisions could likely have been made during each animation was critical to modeling and interpreting neuroimaging data. To this end, we performed an independent online behavioral study using the same animations where participants

(final n = 90) were instructed to indicate their percepts as soon as they had arrived at a decision ("during" phase). To compare the results with the HCP study, participants were also instructed to respond at the end of each trial ("after" phase).

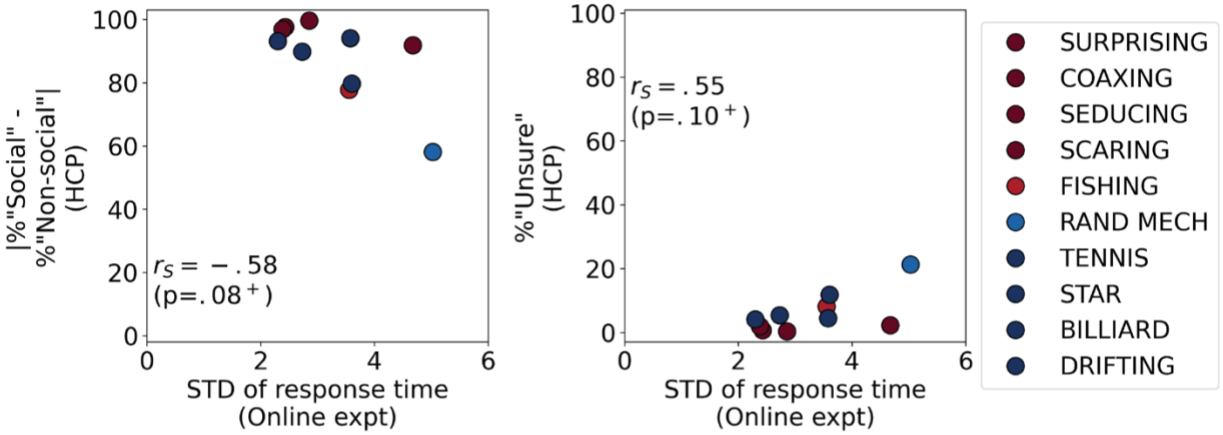
The consensus across participants of which animations were generally "Social" versus "Non-social" in the online sample was comparable to that of the HCP sample (see Figure 2a). As a corollary to this, the animations with high variability in decision times in the online RT experiment also tended to have less consensus across participants in the HCP study – the latter operationalized as (1) the absolute value of the difference between % "Social" and % "Non-social" animations (Figure 2b, left) and (2) higher number of "Unsure" responses (Figure 2b, right). The reaction time data from the "during" phase (Figure 2c) showed that while most responses were made in the earlier half of the 20 second animations, there was a high variability in decision time both within and across animations. This means that the brain activity corresponding to an especially ambiguous animation (e.g., SCARING, RANDOM MECH) could have been vastly different even amongst participants who reported the same percept for these, depending on when each participant made their decision and how this affected their attention before and after the decision. Hence, we identified two animations with the most comparable decision times, namely, COAXING (*median* = 3.45s, *SE* = 0.27s), a predominantly "Social" animation, and BILLIARD (*median* = 3.7s, *SE* = 0.25s), a predominantly "Non-social" animation, whose decision times were not significantly different (Wilcoxon signed-rank test [paired], $t = 1619$, $p = .57$). These animations also did not differ (*McNemar test statistic* = 0, $p = 1$, exact correction done) in their proportion of dominant ("Social" and "Non-social" for COAXING and BILLARD, respectively) and non-dominant responses ("Non-social" and "Social" for COAXING and BILLARD, respectively; see *Methods* sub-section *"Social" vs.*

586 *"Non-social"* for more details). This was the case with the HCP dataset too (*McNemar test*
587 *statistic* = 3.7, p =.055, continuity correction done), although the differences in proportions was
588 close to significance in spite of a large proportion of the participants (n = 870) reporting the
589 dominant percept to COAXING and BILLIARD. Therefore, we used this pair of animations in
590 later analyses as a control for stimulus difficulty/ambiguity.

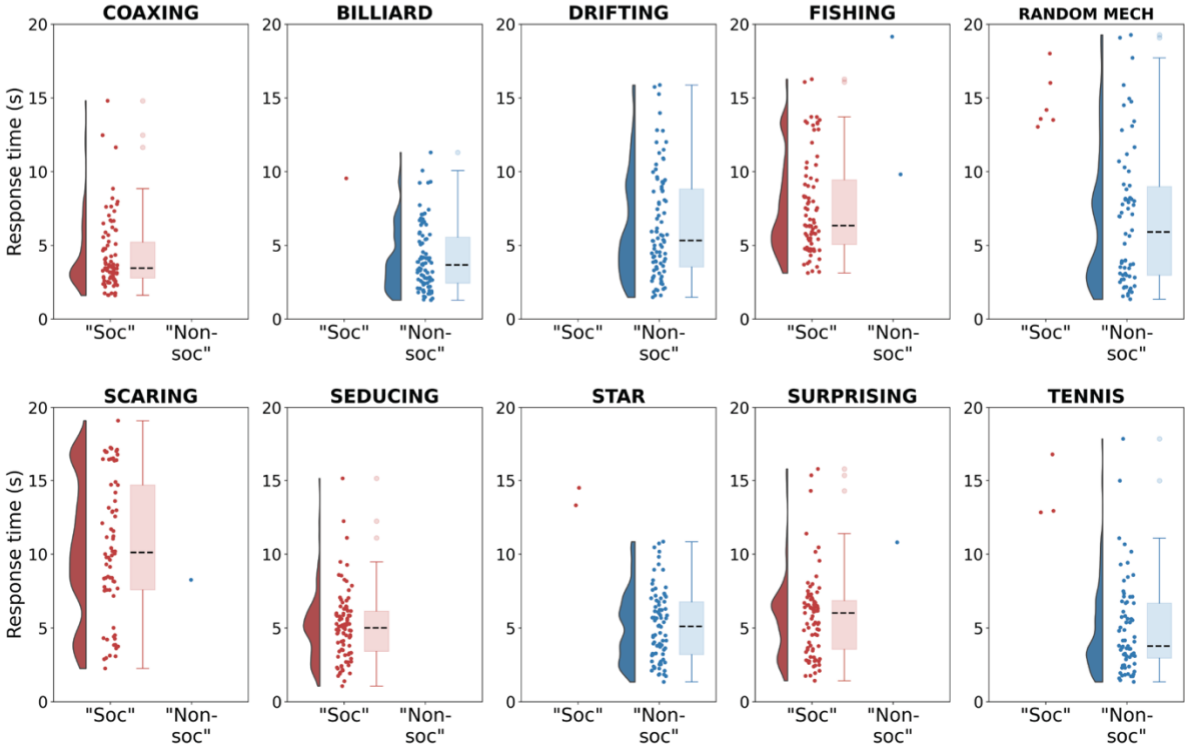
(a) Responses during and after each animation



(b) Uncertainty in decision time during an animation



(c) Response time during each animation



591

Figure 2: Results of the online RT experiment to characterize decision time for each animation. *(a) Number of "Social", 'Non-social" and "Unsure" responses per animation made during (lighter shades) and after (darker shades) each animation. Order of animations on the Y-axis is the same as for the HCP data in Figure 1a. The degree to which animations were reported "Social" is comparable to the HCP behavioral data in Figure 1a. (b) Standard deviation of response time while watching each animation (in seconds; X-axis) vs. two indicators of uncertainty from the HCP behavioral data on the Y-axes (left: absolute difference between number of "Social" and "Non-social" responses, an indicator of how definitive responses for this animation were across participants; right: % of "Unsure" responses). Spearman (rank) correlation shows a trend ($p \leq .1$, marked with '+') for the animations with higher variation in response times in the online RT experiment (X-axes) to also a less definitive response (left) and a higher % of "Unsure" responses (right) in the HCP behavioral data. Dots are colored according to each animations' average perceived socialness (average response from panel (a)). (c) Distribution of response times for "Social" and "Non-social" responses while watching each animation (in seconds). As seen in (b), decision times varied more for some animations than others. Note that COAXING (dominantly "Social") and BILLIARD (dominantly "Non-social") had similar mean decision times that were both relatively early in the animation.*

Much of the brain responds more strongly to what is perceived as social information

In the next set of analyses spanning this and the next two sections, we used the HCP fMRI data to understand where and when the brain distinguishes social from non-social information. For all fMRI analyses, whole-brain data were parcellated into 268 regions covering the cortex, subcortex, and cerebellum using the Shen atlas (Shen et al., 2013) to ease the computational burden of voxel-wise analyses.

In the first fMRI analysis, we focused on the question of "where" by comparing overall neural responsiveness while viewing animations ultimately deemed "Social" versus "Non-social". In addition to regions along the STS which are known to be involved in animacy and interaction perception, we hypothesized that differences might emerge as early as visual regions. We compared "Social" and "Non-social" responses using a general linear model (GLM) approach—again, using the participant's reported percept rather than the experimenter-assigned label as input to the model—in three separate contrasts to ensure results were robust to different confounding factors: 1) within the single most ambiguous animation (RANDOM MECH), which controls for visual input (since all participants saw the same animation, but reported different percepts; across-participants); 2) between two animations with similar decision times (COAXING vs. BILLIARD), to control for the effect of when the decision was likely made on the timecourse of brain activity during passive viewing (within-participants); and 3) across all ten animations, to maximize power and ensure generalizability (within-participants). We then took the intersection of the regions showing a significant difference in all three analyses.

In total, 70 parcels showed "Social" > "Non-social" activity (FDR $q <$ .05, black contours in Figure 3 and Table 3) consistently across all three comparisons. (No parcel showed "Non-social" > "Social" across analyses, though there were some results in this direction in the uncorrected analyses [Figure 3b-c]). Of these, 66 parcels showed positive (i.e., above baseline) activations for both the "Social" ($\beta_{\text{"Social"}} > 0$) and "Non-social" ($\beta_{\text{"Non-Social"}} > 0$) responses for both RANDOM MECH and COAXING–BILLIARD, suggesting that on the whole, much of the brain showed higher *activation* and not lower *deactivation* to "Social" compared to "Non-social". These parcels spanned the occipitotemporal and prefrontal cortex, the cerebellum, and some sub-cortical regions (details in Table 3). We henceforth refer to this set of 70 parcels as the "robust social perception

regions" since they show both specific (after controlling for visual input and decision time) and generalizable activation associated with the subjective experience of a social percept.

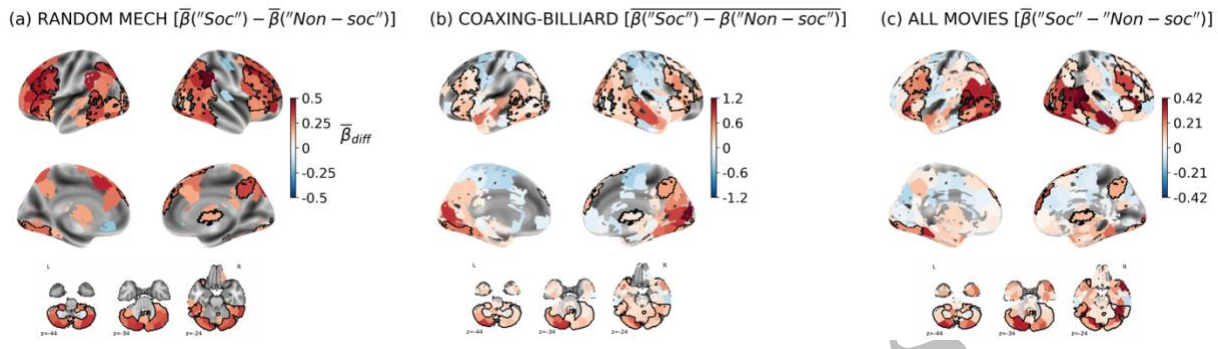


Figure 3: Identifying regions showing differential activity between "Social" and "Non-social" percepts. *Mean differences between GLM regression coefficients (β) for (a) RANDOM MECH (mean (RANDOM MECH "Social") – mean (RANDOM MECH "Non-social"), (b) COAXING–BILLIARD (mean(COAXING "Social" – BILLIARD "Non-social")) and (c) ALL (estimated from run-level regressors, see Methods). Colored regions are significant at an uncorrected threshold ($p < 0.05$) in each of the three analyses, while black contours in a-c show the robust social perception regions significant after correction for multiple comparisons (FDR $q < .05$) in all three analyses. Note: Colorbar ranges are different between the three subplots, since each was estimated separately using different analyses, and hence the values should not be directly compared.*

Table 3.

List of the robust social perception regions (parcels) identified from the intersection analysis (cf. black contours in Figure 3), together with their MNI coordinates.

Lobe	Shen parcels (MNI [x, y, z] coordinates)
Occipital	• Bilateral lateral occipital cortex ([32, –61,49], [48, –62,35], [41, –75, 28], [30, –83, 20], [45, –74,3], [–28, –62,40], [–32, –87,13], [–43, –70, –14], [–48, –67,1], [–36, –84, –4]). • Bilateral Occipital pole ([31, –92 –11], [–22, –97, –10]). • Bilateral occipital fusiform gyrus ([37, –69, –17], [18, –83, –11], [–26, –63, –12], [–15, –84, –13]).
Temporo–occipital	• Bilateral middle and inferior temporal gyrus ([49, –58, 14], [59, –44,9], [47, –60, –15], [55, –56, +5], [61, –43, –18], [42, –46, –23], [–58, –48,5], [–60, –50, –14], [–47, –40, –24]). • Left temporo–occipital fusiform cortex ([–43, –52, –17]). • Left posterior superior temporal gyrus ([–58, –46,6]).
Parietal	• Right angular gyrus ([49, –58, 14]). • Right precuneus ([6, –57, 38]).
Frontal	• Bilateral middle and inferior frontal gyrus ([41,15,48], [54,25,1], [40,18,29], [40,18,29], [–46,28,27], [–39,17,47], [–53,18,11]). • Bilateral lateral precentral gyrus ([40,4,34], [–46,8,29]). • Bilateral frontal operculum ([37,21,6], [–32,22,6]). • Bilateral frontal pole ([29,51,19], [9,53,24], [24,31,36], [48,36,15], [–10,56,30]). • Right superior frontal gyrus ([15,37,49], [24,31,36], [14,6,65], [25,12,49]). • Bilateral anterior insula ([37, 21, 6], [–32,20, –16], [–32,22,6]). • Left frontal orbital cortex ([–46,28, –7], [–32,20, –16].
Sub–cortical	• Bilateral cerebellum (large parts of it) ([32, –78,–40], [39,–75, –30], [23, –36, –43], [7,–68,–38], [46,–46,–43], [16,–47,–53], [12,–84,–35], [20,–73,–50], [37,–57,–33], [23,–72,–29], [42,–64,–49], [7,–54,–34], [30,–36,–31], [–6,–66,–38], [–21,–70,–49], [–40,–75,–29], [–30,–80,–40], [–43,–64,–46], [–10,–82,–32], [–46,–47,–43], [–26,–70,–31], [–24,–38,–44]). • Right thalamus ([6, –10, 5]).

657

658 **Subjective percepts better explain brain activity in robust social perception regions**

659 Next, we tested whether observer response-based labels (“Social”, “Non-social”) explained more

660 variance in the neural data than experimenter-assigned labels (Mental, Random) by comparing

661 models based on each label type. Overall, across all 268 parcels, response labels better explained

brain activity ($AIC_{Obs-Exp}$: $M = –2.23$, $SE = 0.47$; one-sample t-test: $t = –4.77$, $p < 10^{-5}$). Of these, 44 parcels were better fit by the observer-based model ($AIC_{Obs-Exp} > 10$) and 11 parcels were better fit by the experimenter-based model ($AIC_{Obs-Exp} < –10$). Observer-based labels better fit the neural activity in bilateral occipito-temporal regions, left pre-frontal cortex and the cerebellum (parcels colored pink in Figure 4) — several of them overlapping with the robust social perception regions identified in the GLM analysis above (Figure 4) — whereas experimenter-assigned labels better fit the neural activity in the right temporal cortex (parcels colored green in Figure 4). Overall, this suggests that activity in the robust social perception regions reflect the conscious perception of social information rather than merely incoming visual input.

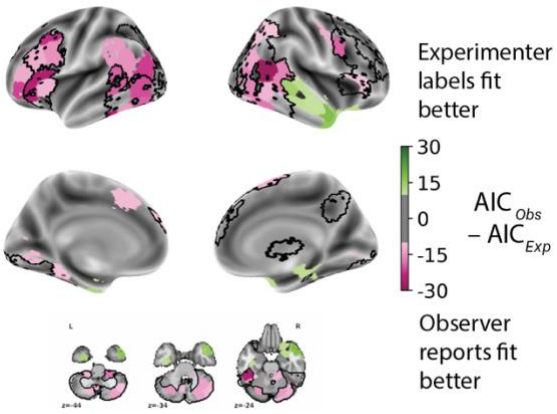


***Figure 4**: **Models based on observer reports, compared to experimenter labels, better explain brain activity in robust social perception regions.** Comparison between two linear mixed-effects models fitting trial-level GLM regression coefficients (β) as a function of either 1) observer responses ("Social" vs. "Non-social") or 2) experimenter labels (Mental vs. Random). Parcels are colored by the difference in Akaike information criterion (AIC) between the two models and were thresholded at 10 (i.e., parcels with |AIC| < 10 are not plotted): parcels colored pink (AIC*

< -10) indicates a better fit for the observer response-based model and green (AIC > 10) indicates a better fit for the experimenter label-based model. Both models included participant and animation as random effects. Black contours correspond to the robust social perception regions identified in Figure 3, and these largely overlap with parcels that show better fits for the observer-based models (pink).

Some brain regions show parametric responses to degree of perceived socialness

The previous analysis identified social information-processing regions that robustly showed a higher response to information ultimately reported as "Social". By leveraging "Unsure" responses as an intermediate level of perceived socialness between "Social" and "Non-social", we further probed the neural correlates of conscious social perception—i.e., an "Unsure" response would indicate that some evidence for a social interaction was detected, but not enough to be fully confident in a "Social" response.

Specifically, we identified regions showing parametric responses, i.e., $\beta_{"Social"} > \beta_{"Unsure"} > \beta_{"Non\text{-}social"}$ (condition *S* > *U* > *NS*) or $\beta_{"Social"} < \beta_{"Unsure"} < \beta_{"Non\text{-}social"}$ (condition *S* < *U* < *NS*) using conjunction analyses across all animations (see *Methods* for details). We further limited this analysis to the robust social perception regions (cf. black contours in Figure 3).

Several parcels showed a consistent S > U > NS response pattern (Figure 5a-c). These were located in posterior and inferior parts of the temporal cortex including parts of the motion-processing region V5/MT (with more parcels in the right hemisphere), middle and inferior frontal gyrus, precuneus, right thalamus and postero-lateral parts of the cerebellum. Other regions that showed a differential response to "Social" compared to "Non-social" but do not show up here, such as the superior temporal and occipital regions, posterior parietal regions and superior frontal regions, likely have more dichotomous responses to any amount of social content ([Social",

"Unsure"] > "Non-social") or only a high level of evidence in favor of "Social" (Social" > ["Unsure", "Non-social"]).

To verify that similar parametric patterns emerge when controlling for visual input, we plotted the timecourses for each response type for a subset of the parcels showing parametric responses pattern to the most ambiguous animation (RANDOM MECH, Figure 5d; see *Methods* for how these parcels were chosen). Visualizing these timecourses confirmed that these regions show parametric neural responses to degrees of reported socialness, albeit with large errorbars for the smaller groups ("Social" and "Unsure").

Thus, it appears that many regions, predominantly in temporal, occipital, and subcortical areas, show a graded response to degree of social information. This result further underscores how using observer-based labels can increase sensitivity and specificity in linking brain activity to conscious experience.

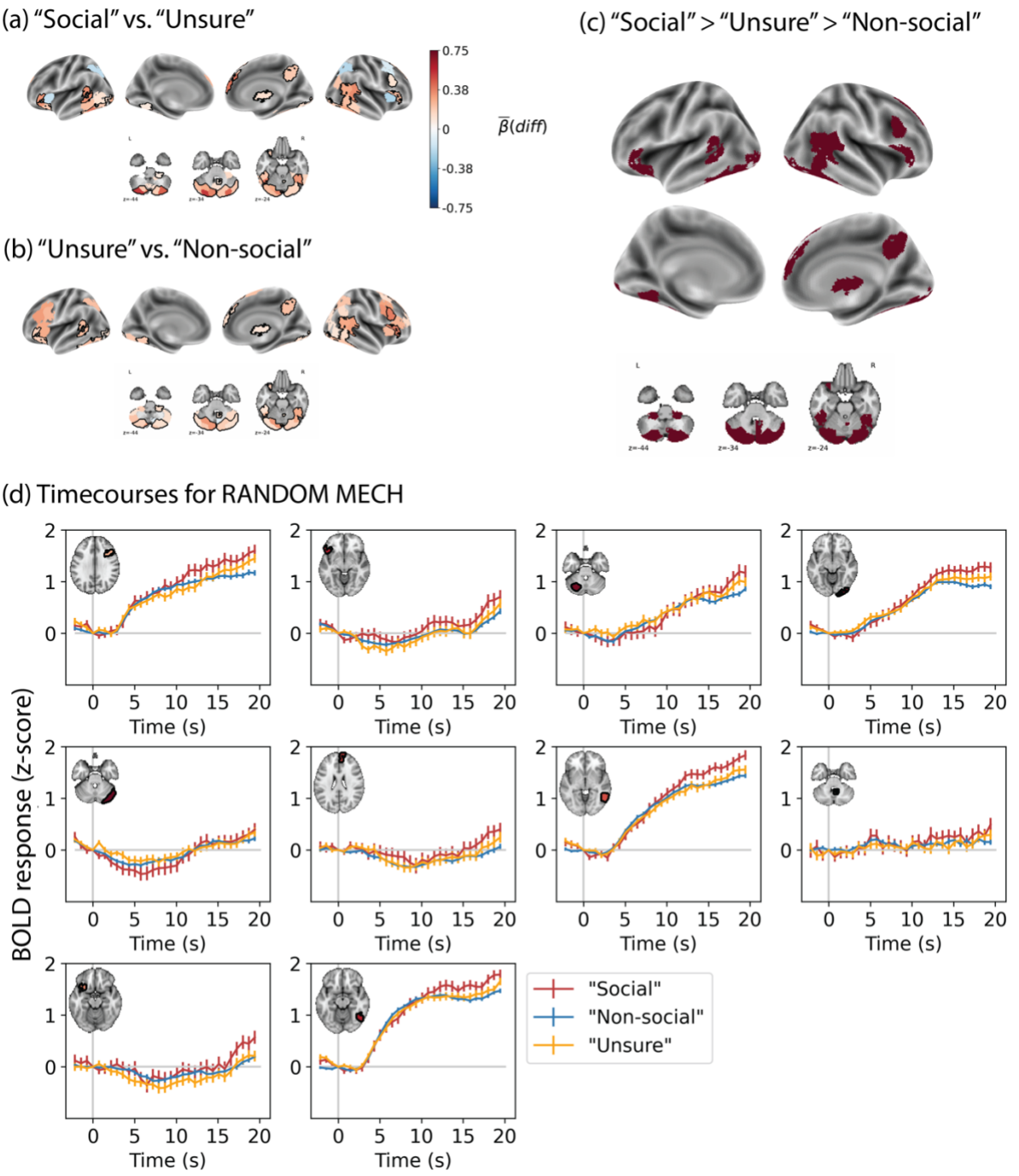


Figure 5. Brain regions showing parametric responses to social content. *(a-b) Colored parcels show mean differences (FDR q < .05) in slope regression coefficients ("Social"-"Unsure" and "Unsure"-"Non-social") within the robust social perception regions (cf. black contours in Figure 3). Black contours and the regions colored dark red in (c) highlight the 35 parcels that showed a*

graded response to perceived socialness ("Social" > "Unsure" > "Non-social" or vice-versa; in other words, the intersection of (a) and (b)) . (d) Timecourses for the most ambiguous animation (RANDOM MECH) in a subset of 10 of the parcels from (a), confirming that activity associated with "Unsure" percepts is intermediate to "Non-social" percepts even when controlling for visual input.

Processing of social versus non-social information diverges early in time and in the cortical hierarchy

The previous analyses showed that several regions spanning the whole brain are more responsive to information that is ultimately reported as social (versus non-social). However, given that these analyses modeled the entire 20 s animations, any differences, especially in early visual regions, could reflect (1) the accumulation of evidence that *led* to the perception of an animation as "Social", (2) the *consequence* of having perceived an animation as "Social" (i.e.., top-down attention effects on sensory regions), or (3) a combination of both. To gain a better understanding of the dynamics of evidence accumulation leading to a "Social" percept, we compared BOLD activity at each timepoint (TR) after stimulus onset to determine the timepoint of earliest divergence between "Social" and "Non-social" percepts.

To ensure that the differences observed at each timepoint are comparable in terms of the underlying cognitive processes (i.e., evidence accumulation versus decision-making versus post-decisional processes), we performed this analysis on the animation pair which had the most comparable decision times in our auxiliary behavioral experiment, namely COAXING and BILLIARD. Decision times for these animations were both early and close in time (as explained in *Materials and Methods* and the *Results* section for the online RT experiment, also see Figure 2c). These animations were similar visually with the same two triangular agents on the screen

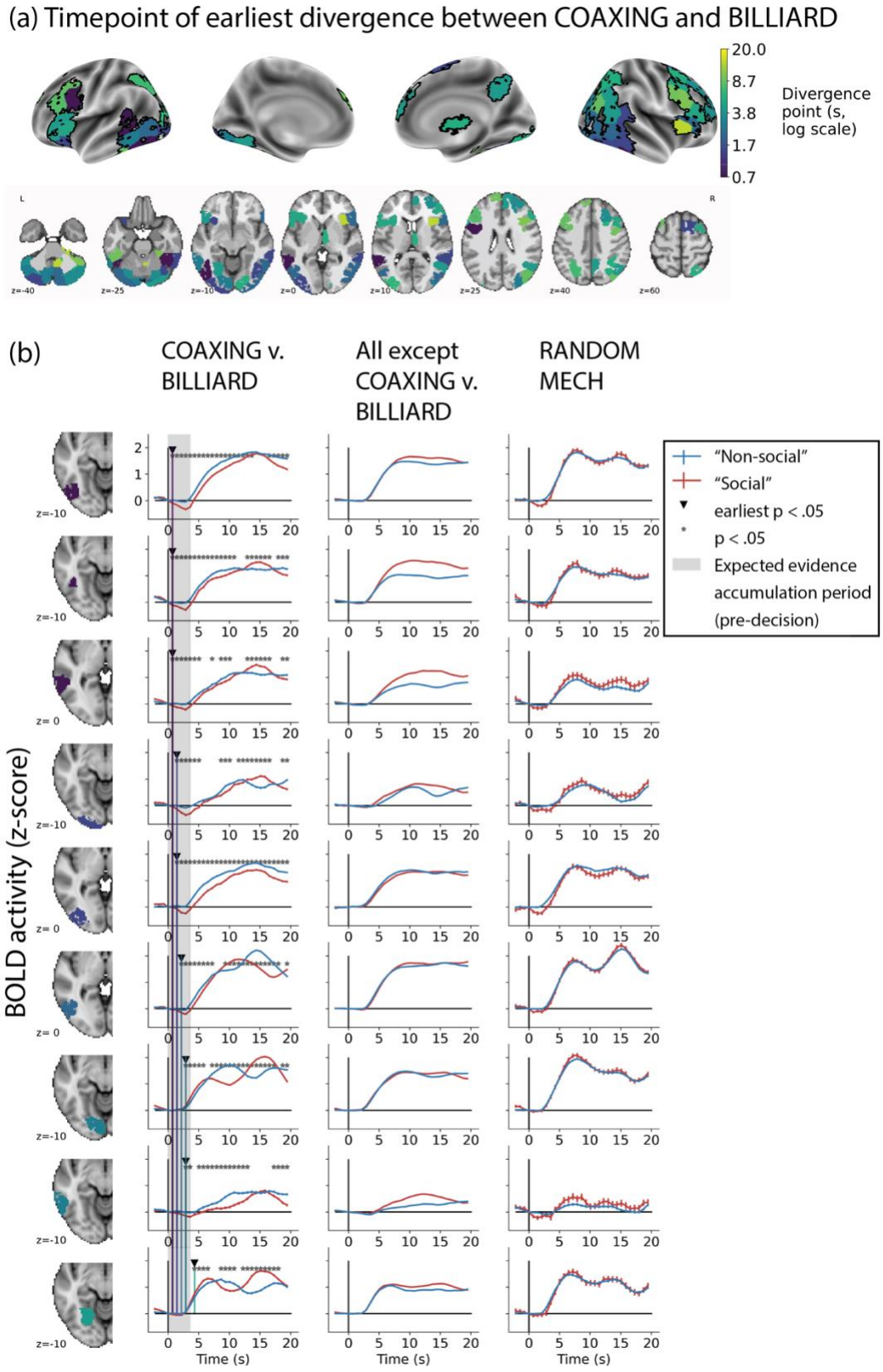
(see Table 1); nevertheless, they did vary in their temporal dynamics and some low-level visual features. To minimize the effect of these on the BOLD activity, we regressed two low-level visual features (total optic flow and mean brightness) from the BOLD responses of each animation and participant, and compared the residual COAXING and BILLIARD timecourses at each TR. To guard against spurious fluctuations early in the animations, we again limited our analysis to the robust social perception regions (cf. Figure 3, black contours).

In many regions, differences in brain activity between "Social" and "Non-social" percepts emerged early, i.e., in TRs 1-3 after stimulus onset (Figure 6a). These early differences were seen in both hemispheres, in posterior regions such as the fusiform gyrus, lateral occipital cortex, pSTS and posterior parts of the cerebellum as well as in frontal areas such as the lateral precentral gyrus, posterior parts of the middle and inferior frontal gyrus (MFG, IFG), the orbitofrontal cortex (OFC) in the left hemisphere, and the IFG and supplementary motor area (SMA) in the right hemisphere. Later TRs, which are more likely to reflect post-decisional activity, showed divergences in the bilateral inferior and superior frontal regions, the right precuneus, bilateral intraparietal sulcus (IPS) and bilateral posterior cerebellum.

To visualize the earliest differences in the posterior regions and to understand how generalizable these dynamics are, we plotted (Figure 6b) the residual BOLD timecourses for COAXING–BILLIARD (left column, our main analysis) alongside the averaged "Social" and "Non-social" timecourses across all the other animations (All except COAXING–BILLIARD; middle column) and within the most ambiguous animation (RANDOM MECH; right column). The two latter analyses are not as well suited to pinpointing *when* differences emerged because decision times were likely more variable across individuals and animations (per our online RT experiment), thus making timecourses noisier and less comparable. Despite this, we see similar

765 relative trends in these posterior regions (each row) as to when and how they distinguish between
766 "Social" and "Non-social" reports. Responses emerged much later for the "All except
767 COAXING–BILLIARD" condition in line with the later and more variable decision times for
768 most animations; see Figure 2c). When comparing within the same animation (RANDOM
769 MECH), we see trends emerging early on, although the magnitudes are smaller and the error for
770 the "Social" responder group are large, possible because of the smaller group size (n = 107)
771 compared to the majority percept of "Non-social" (n = 670). Note that the latter two timecourses
772 are plotted only for visual examination and that we did not perform statistical analyses here. Also
773 note that the directionality of the difference between "Social" and "Non-social" should not be
774 strongly interpreted especially in the case of COAXING–BILLIARD, since despite our attempts
775 to normalize activity at the trial level (see *Methods*), order effects (COAXING was always the
776 first stimulus in the first run, immediately followed by BILLIARD) and/or the shape of the
777 hemodynamic response (i.e., presence of initial dip) could have affected the BOLD response
778 between trials.

779 To summarize, while watching an animation that was eventually reported as "Social",
780 differences in brain activity emerged early across much of the brain, involving both ventral
781 visual processing regions and occipito-temporal regions involved in action and animacy
782 detection as well as social cognition. The early reactivity in these regions is in line with the
783 recently suggested "third visual pathway" that projects directly from early visual cortex to the
784 superior temporal sulcus and is specialized for social perception (Pitcher & Ungerleider, 2021).


(a) Timepoint of earliest divergence between COAXING and BILLIARD
20.0
8.7
3.8
1.7
0.7
Divergence point (s, log scale)
L
R
z=-40
z=-25
z=-10
z=0
z=10
z=25
z=40
z=60
(b)
COAXING v. BILLIARD
All except COAXING v. BILLIARD
RANDOM MECH
"Non-social"
"Social"
earliest p < .05
p < .05
Expected evidence accumulation period (pre-decision)
BOLD activity (z-score)
Time (s)

785

Figure 6. Timecourse analysis showing when and where differences between "Social" and "Non-social" percepts emerge. *(a) Brain map of the earliest timepoint at which brain activity diverges between "Social" and "Non-social" responses for the COAXING and BILLIARD animations, respectively (within-participant analysis). Analysis was limited to the robust social perception regions (cf. Figure 3, black contours), and BOLD signal timecourses were residualized with respect to the visual features of brightness and optic flow to minimize the effects of any differences in low-level sensory information between the two animations. Colors show how early (purple-blue) or late (yellow-green) activity diverged. (b) BOLD signal timecourses in the left posterior regions illustrating how "Social" and "Non-social" activity diverge in the pre-decisional period for COAXING and BILLIARD. Rows: regions are sorted by the earliest divergence TR and then from posterior to anterior. Columns: left, timecourses for the two animations matched for approximate decision time, COAXING ("Social") and BILLIARD ("Non-social"). Middle and right, timecourses from the same regions from two supporting analyses: across all animations except COAXING–BILLIARD ("Social" vs. "Non-social" response trials), middle; and for the most ambiguous animation, RANDOM MECH ("Social" vs. "Non-social" responders), right.*

Individual differences in behavior and brain activity while viewing animations covary with internalizing symptoms

Lastly, we explored whether individual differences in behavioral and neural responses to social animations covaried with trait-level measures. Specifically, we focused on internalizing symptoms from the Achenbach Adult Self-Report Scale, because past work has shown that certain internalizing traits (e.g., loneliness, anxiety) are associated with a stronger tendency to perceive

808 visual cues as socially salient. We hypothesized that individuals with higher internalizing scores
809 would show stronger behavioral and neural reactivity to potentially social information.
810 Using the behavioral data, we tested whether the response bias towards "Social" (cf.
811 Figure 1a) was even stronger for individuals higher on internalizing symptoms. Indeed, there was
812 a positive relationship between the bias toward "Social" responses and internalizing score
813 (Spearman rank correlation r_s =.10, p =.003, Figure 7a). We tested the specificity of this
814 relationship by contrasting it to the correlation with externalizing trait scores, which index more
815 "acting out" behaviors like rule-breaking and aggression and have not been linked to social
816 perception tendencies (though note that internalizing and externalizing symptoms were correlated:
817 r_s = .51, $p < 10^{-55}$). The correlation with externalizing symptoms was not significant (r_s = .06, p =
818 .096), although the two correlations were not significantly different (t = 1.3, p = .094).
819 Furthermore, individuals with higher internalizing scores were more likely to give a "Social" or
820 "Unsure" (as opposed to "Non-social") response to the most ambiguous animation, RANDOM
821 MECH ("Social" or "Unsure", M = 49.3, SE = 0.69; "Non-social", M = 47.7, SE = 0.45; unpaired
822 t-test, t = 2.05, p = .04, Figure 7b). Mean externalizing symptoms were also higher for the "Social"
823 or "Unsure" group (M = 49.3, SE = 0.57) compared to the "Non-social" group (M = 47.9, SE =
824 0.38). While the difference in externalizing symptoms was smaller (unpaired t-test, t = 1.95, p
825 =.051), it was not significantly different from the internalizing symptoms (interaction between
826 response and score type [internalizing vs. externalizing]: p = .63).
827 Lastly, individuals with higher internalizing scores were also more likely to give an "Unsure"
828 response to animations intended as Random (Spearman r_s = .098, p = .005), but not to animations
829 intended as Mental (r_s = –.024, p = .49), indicating a preference for false alarms over misses when
830 it comes to detecting social information (difference between correlations: t = 2.47, p = .007; Figure

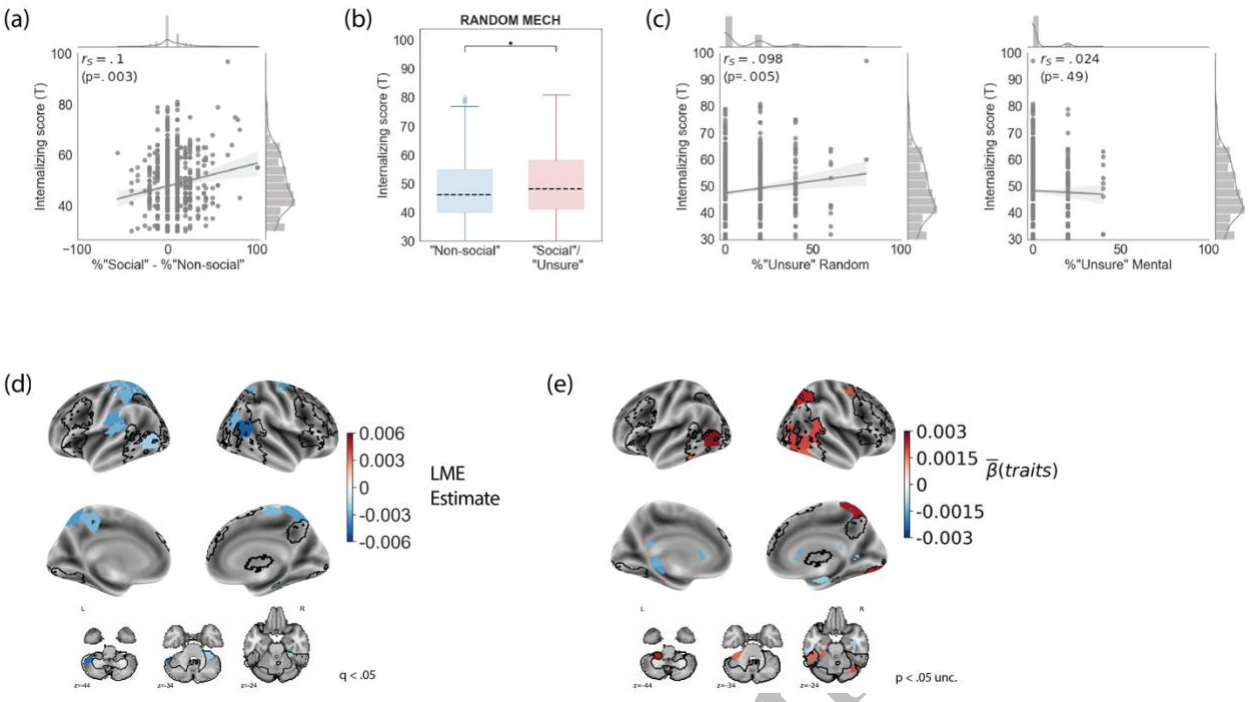
831 7c). Percent "Unsure" responses did not correlate with externalizing symptoms for either Random
832 (r_s = .048, p = .17) or Mental (r_s = .01, p = .75) animations. Together, these analyses support a link
833 between internalizing symptoms and a greater tendency to perceive information as social, perhaps
834 driven by a homeostatic drive to seek social connections.

835 To understand whether overall neural activity while watching animations and scanning for
836 social information also covaried with internalizing symptoms, we related trial-wise brain activity
837 estimates to internalizing symptom scores in an LMEM (fixed effect: internalizing score; random
838 effect: animation). In a whole-brain analysis, 18 parcels showed a significant relationship (FDR q
839 < .05, Figure 7d) between internalizing score and neural responsiveness. In all of these, the LME
840 estimates were negative—i.e., as internalizing scores increased, brain activity decreased—
841 although all 18 parcels showed above-baseline activity as evidenced by the positive regression
842 coefficients ($\bar{\beta} > 0$ for all parcels). Thus, while individuals with higher internalizing scores showed
843 positive activity in these regions when scanning animations for social information, the magnitude
844 of this activity was lower than in individuals with lower internalizing scores. These relationships
845 were seen in the right angular gyrus, the bilateral superior parietal lobule, left supramarginal gyrus,
846 regions along the dorsal midline, and anterior parts of the cerebellum (colored blue in Figure 7d).

847 Interestingly, the lateral occipital parcels from the set of robust social perception regions
848 (shown as black contours in Figure 7d) were not as prominent here, showing only a partial overlap
849 (5 parcels) with the parcels showing trait effects. In the overlapping parcels, which comprised
850 bilateral occipito-temporal parcels and the cerebellum, individuals high on internalizing traits
851 showed overall less reactivity in many brain regions while scanning the environment for social
852 interactions. To reconcile this decrease in neural reactivity (Figure 7d) with the observed increase
853 in behavioral sensitivity (Figure 7a-c), one interpretation is that these individuals have a lower

threshold for the amount of neural activity required to declare something "Social". Yet another interpretation—based on the decrease in neural activity with internalizing symptoms in all trait-sensitive parcels and the observation that 72% of the trait-sensitive regions (13 parcels) lay outside the previously identified robust social perception regions—is that this reflects a general decrease in neural responsiveness with more internalizing symptoms.

Finally, we probed whether the difference in neural reactivity to information ultimately perceived as "Social" showed any trait dependence. No parcel showed a significant relationship between internalizing scores and subject-level "Social" – "Non-social" beta estimates at the corrected threshold across all 268 parcels (FDR $q < .05$). At the *uncorrected* threshold ($p < .05$) however, we found several parcels (most notably in the right occipito-temporal parcels; see Figure 7e) showing a positive relationship such that individuals with higher internalizing symptoms showed a relatively higher responsiveness in this parcel to information eventually declared "Social". Together with the overall effect of traits on brain activity, these results may suggest that individuals exhibiting higher internalizing symptoms show lower brain activity when scanning for social content, but that the magnitude of this dip may be lower when viewing social content. However, this potential relationship should be further tested in datasets with more—and ideally more ambiguous—stimuli, to allow for more variation in both behavioral and neural responses.



871

872 ***Figure 7: Relationship between internalizing trait scores, behavior, and brain activity.*** *(a)*

873 *Response bias (% difference between 'Social" and "Non-social" responses per participant)*

874 *correlates positively with internalizing symptom score (Spearman correlation coefficient* r_s *= .1,*

875 *p = .003). (b) Internalizing scores among individuals who reported some degree of socialness*

876 *("Social" or "Unsure" responses) to the most ambiguous animation, RANDOM MECH, were*

877 *higher than those for individuals who reported this animation "Non-social". * indicates p < .05.*

878 *(c) Internalizing score correlates positively with the percent of "Unsure" responses per*

879 *participant for the generally non-social animations (Random; left;* r_s *= .098, p =.005) but not for*

880 *the generally social* animations *(Mental; right;* r_s *= –.024, p = .49). These correlation*

881 *magnitudes were significantly different (t = 2.47, p = .007). (d) Relationship between neural*

882 *responsiveness during the task and internalizing scores. Colored parcels showed a significant*

883 *positive (red) or negative (blue) relationship (FDR q < 0.05) with internalizing score. The robust*

884 *social perception regions from the GLM analysis (cf. Figure 3) are shown in black. All regions*

885 *show a negative relationship between activation magnitude and internalizing symptoms and*

there is only a partial overlap with the robust social perception regions. (e) Parcels in which the difference between activity to "Social" and "Non-social" percepts (cf. Figure 3c) may be modulated by internalizing symptoms (p <.05, unc). Red: at higher internalizing symptoms, "Social" trials show a relatively higher response than "Non-social"; blue: at higher internalizing symptoms, "Social" trials show a relatively lower response than "Non-social".

Discussion (1273/1500 words)

In this study, we investigated behavioral and neural signatures of social signal detection using a large dataset of neurotypical young adults. Behavioral responses showed a subtle but consistent bias towards perceiving information as social (as opposed to non-social), which manifested as a higher number of "Social" responses and a hesitation to report information as "Non-social". We then used observers' own responses to label fMRI data and found that widespread patterns of brain activity differentiate conscious social percepts, even when controlling for visual input (RANDOM MECH) and decision time (COAXING–BILLIARD). Overall, observer responses explained more variance in activity than experimenter-assigned labels. Several regions also showed parametric responses to degrees of perceived socialness ("Social" > "Unsure" > "Non-social" responses). Further, brain activity for information ultimately deemed "Social" diverged from "Non-social" early both in time and the cortical hierarchy. Lastly, we found that a trait-level measure of internalizing symptoms (e.g., loneliness, anxiety) could explain some of the variability in percepts and brain activity.

Humans are an "obligate social" species predisposed towards social interactions (Epley et al., 2007; Rutherford & Kuhlmeier, 2013), and socially relevant content is processed more efficiently (Rothkirch et al., 2015; Papeo et al., 2017). The response bias towards "Social" in the

current study, and its covariation with internalizing symptoms, support the idea of a homeostatic drive to seek social connection (Tomova et al., 2020). This is in line with previous studies reporting that lonely people tend to form illusory social connections (Epley et al., 2008), overattribute animacy to faces (Powers et al., 2014) and have greater attention and memory for social cues (Gardner et al., 2005). It is also possible that the response bias observed here could have been partially due to the task structure, including contextual effects, instructions, or the very presence of multiple agents, which can induce expectations for social content (Piejka et al., 2022). However, it is unlikely that these factors fully account for our results. Despite their fixed order, trials were pseudorandomized with respect to experimenter-assigned labels, meaning that trial order should not have induced a systematic bias in responses. Furthermore, *all* animations (i.e., not just Mental ones) contained at least two agents, and the Mental and Random animations were largely matched in terms of the number and type of agents (Table 1).

Animated shapes have been used extensively in fMRI studies to characterize brain activity involved in social perception. Past work has converged on a canonical set of regions including bilateral pSTS, lateral occipital cortex (LOC), angular gyrus, superior parietal lobule and medial prefrontal cortex (e.g., Castelli et al., 2000; Tavares et al., 2008; Osaka et al., 2012). However, nearly all of this work has used stimuli generated by experimenters to be seen as obviously social or obviously non-social, then characterized participants' "accuracy" with respect to these labels. There are two major limitations to this approach: one, differences in neural responses to social content may be confounded by differences in low-level visual features (e.g., higher speed for animations labelled social); and two, the idea that experimenter labels represent the "ground truth" is likely unrealistic given that real-world social scenarios are frequently ambiguous, and interpretations vary across individuals. Here, we extended the social

perception literature in an important way: we eschewed experimenter-assigned labels and characterized brain activity according to participants' *own reported percepts*, which allowed us to identify regions that are sensitive to the subjective (or "conscious") perception of a social interaction over and above sensory inputs. Several of these regions also showed parametric responses to degrees of perceived socialness, suggesting an even tighter link between activity in these regions and conscious perceptual experiences.

Because the animations used here and in and past work are typically relatively long in duration (20-40 s), another open question is to what extent the observed brain activity reflects distinct cognitive processes over the course of each trial. For example, early in animations, participants are likely accumulating evidence in favor of each alternative (i.e., social or non-social) until a decision is reached. Following this, post-decisional processes likely come into play, which could include maintaining the decision in working memory and monitoring for any counterevidence. Animations that have been deemed "Social" may enjoy higher levels of attention and engagement through the remainder of the trial, which could partially explain current and past observations (Tavares et al., 2008) of stronger neural responses to social content. Although the heterogeneity in decision times across individuals and animations (cf. Figure 2c) makes it challenging to disambiguate these processes, by leveraging stimuli with comparable decision times, we showed that several occipito-temporal regions start responding differently to information ultimately perceived as "Social" even before participants had likely arrived at a decision, suggesting that this activity may reflect pre-decisional evidence accumulation. This is further supported by the results of our parametric analysis, as well as recent EEG work showing a temporal hierarchy in action perception from encoding visual to social features (Dima et al., 2022). Early differences also emerged in the pSTS—an area critical

to the third visual stream hypothesis (Pitcher & Ungerleider, 2021)—and lateral parts of the precentral gyrus and the supplementary motor area. Recent electrophysiological studies (Isik et al., 2020; Dima et al., 2022) have shown that neural responses to social interactions occur at 300ms and that socio-affective features best predict neural responses at around 418ms — i.e., on a timescale congruent with top-down processing. Thus, early activity in frontal regions may reflect feedback mechanisms that direct attentional resources in sensory cortex to prioritize processing social information.

In our trait-dependent analyses, we found that brain activity during the task was lower for individuals with high internalizing scores in regions including parts of the default mode network (angular gyrus and precuneus) and some of our previously identified robust social perception regions (occipito-temporal, frontal, cerebellar). While this may reflect a decrease in neural reactivity while scanning the environment specifically for social information, we cannot rule out that it may also reflect change in neural activity to any task in these individuals (Piani et al., 2022). The apparent discrepancy whereby individuals with higher internalizing scores show *increased* behavioral sensitivity but *decreased* neural activity to potentially social content could indicate that these individuals have a lower neural threshold for declaring something "Social". However, we found only weak evidence for interactions between internalizing scores, neural responses, and reported percepts. Future work might return to this question using more ambiguous stimuli that evoke more variability in both neural and behavioral responses across people.

One limitation of this dataset is that the stimulus set consisted of only ten animations that were not counterbalanced in order across participants, nor controlled in terms of their visual features. These animations were also not optimal to study ambiguous perception since all *did*

have dominant percepts, although the large sample size still allowed us to leverage non-dominant percepts to separate conscious social percepts from sensory input (i.e., RANDOM MECH analysis). Future studies should replicate and extend these results using stimuli that are better controlled and also more ambiguous (i.e., evoke more balanced responses).

Another limitation is that participants were limited to three discrete response options, when perceptual certainty may have varied even within each response type. Furthermore, even for the same animation, individuals who perceive a social interaction do not necessarily perceive the same *type of* social interaction, and different interpretations could have muddied group-level effects. Future experiments can overcome this limitation by using richer behavioral characterizations of percepts.

In summary, we describe behavioral and neural processes that underlie how people arrive at conscious percepts of social information. Together, our results compel a more nuanced view of social perception in which socialness "is in the eye of the beholder".

References

Abassi, E., & Papeo, L. (2020). The Representation of Two-Body Shapes in the Human Visual Cortex. *Journal of Neuroscience*, *40*(4), 852–863. https://doi.org/10.1523/JNEUROSCI.1378-19.2019

Abassi, E., & Papeo, L. (2022). Behavioral and neural markers of visual configural processing in social scene perception. *NeuroImage*, *260*, 119506. https://doi.org/10.1016/j.neuroimage.2022.119506

Abell, F., Happé, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development*, *15*(1), 1–16. https://doi.org/10.1016/S0885-2014(00)00014-9

Achenbach, T. M., Ivanova, M. Y., & Rescorla, L. A. (2017). Empirically based assessment and taxonomy of psychopathology for ages 1½–90+ years: Developmental, multi-informant, and multicultural findings. *Comprehensive Psychiatry*, *79*, 4–18. https://doi.org/10.1016/j.comppsych.2017.03.006

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin*, *74*(2), 81–99. https://doi.org/10.1037/h0029531

Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., Glasser, M. F., Curtiss, S., Dixit, S., Feldt, C., Nolan, D., Bryant, E., Hartley, T., Footer, O., Bjork, J. M., Poldrack, R., Smith, S., Johansen-Berg, H., Snyder, A. Z., … WU-Minn HCP Consortium. (2013). Function in the human connectome: Task-fMRI and individual differences in behavior. *NeuroImage*, *80*, 169–189. https://doi.org/10.1016/j.neuroimage.2013.05.033

Barrett, H. C., Todd, P. M., Miller, G. F., & Blythe, P. W. (2005). Accurate judgments of intention from motion cues alone: A cross-cultural study. *Evolution and Human Behavior*, *26*(4), 313–331. https://doi.org/10.1016/j.evolhumbehav.2004.08.015

Blakemore, S.-J., Boyer, P., Pachot-Clouard, M., Meltzoff, A., Segebarth, C., & Decety, J. (2003). The detection of contingency and animacy from simple animations in the human brain. *Cerebral Cortex (New York, N.Y.: 1991)*, *13*(8), 837–844. https://doi.org/10.1093/cercor/13.8.837

Castelli, F. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain*, *125*(8), 1839–1849. https://doi.org/10.1093/brain/awf189

Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and Mind: A Functional Imaging Study of Perception and Interpretation of Complex Intentional Movement Patterns. *NeuroImage*, *12*(3), 314–325. https://doi.org/10.1006/nimg.2000.0612

Davis, J. W., & Gao, H. (2004). An expressive three-mode principal components model for gender recognition. *Journal of Vision*, *4*(5), 2. https://doi.org/10.1167/4.5.2

Deen, B., Koldewyn, K., Kanwisher, N., & Saxe, R. (2015). Functional Organization of Social Perception and Cognition in the Superior Temporal Sulcus. *Cerebral Cortex*, *25*(11), 4596–4609. https://doi.org/10.1093/cercor/bhv111

Desai, A., Foss-Feig, J. H., Naples, A. J., Coffman, M., Trevisan, D. A., & McPartland, J. C. (2019). Autistic and alexithymic traits modulate distinct aspects of face perception. *Brain and Cognition*, *137*, 103616. https://doi.org/10.1016/j.bandc.2019.103616

Dima, D. C., Tomita, T. M., Honey, C. J., & Isik, L. (2022). Social-affective features drive human representations of observed actions. *ELife*, *11*, e75027. https://doi.org/10.7554/eLife.75027

Epley, N., Akalis, S., Waytz, A., & Cacioppo, J. T. (2008). Creating Social Connection Through Inferential Reproduction: Loneliness and Perceived Agency in Gadgets, Gods, and Greyhounds. *Psychological Science*, *19*(2), 114–120. https://doi.org/10.1111/j.1467-9280.2008.02056.x

1042 Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of
1043 anthropomorphism. *Psychological Review*, *114*(4), 864–886.
1044 https://doi.org/10.1037/0033-295X.114.4.864
1045 Friston, K. J., Jezzard, P., & Turner, R. (1994). Analysis of functional MRI time-series. *Human*
1046 *Brain Mapping*, *1*(2), 153–171. https://doi.org/10.1002/hbm.460010207
1047 Gardner, W. L., Pickett, C. L., Jefferis, V., & Knowles, M. (2005). On the Outside Looking In:
1048 Loneliness and Social Monitoring. *Personality and Social Psychology Bulletin*, *31*(11),
1049 1549–1560. https://doi.org/10.1177/0146167205277208
1050 Glover, G. H. (1999). Deconvolution of Impulse Response in Event-Related BOLD fMRI1.
1051 *NeuroImage*, *9*(4), 416–429. https://doi.org/10.1006/nimg.1998.0419
1052
1053 Gottsdanker, R. (1982). Age and Simple Reaction Time1. *Journal of Gerontology*, *37*(3), 342–
1054 348. https://doi.org/10.1093/geronj/37.3.342
1055 Hebart, M. N., Donner, T. H., & Haynes, J.-D. (2012). Human visual and parietal cortex encode
1056 visual choices independent of motor plans. *NeuroImage*, *63*(3), 1393–1403.
1057 https://doi.org/10.1016/j.neuroimage.2012.08.027
1058 Heider, F., & Simmel, M. (1944). An Experimental Study of Apparent Behavior. *The American*
1059 *Journal of Psychology*, *57*(2), 243–259. https://doi.org/10.2307/1416950
1060 Isik, L., Koldewyn, K., Beeler, D., & Kanwisher, N. (2017). Perceiving social interactions in the
1061 posterior superior temporal sulcus. *Proceedings of the National Academy of Sciences*,
1062 *114*(43), E9145–E9152. https://doi.org/10.1073/pnas.1714471114
1063 Isik, L., Mynick, A., Pantazis, D., & Kanwisher, N. (2020). The speed of human social
1064 interaction perception. *NeuroImage*, *215*, 116844.
1065 https://doi.org/10.1016/j.neuroimage.2020.116844

Johnson, K. L., & Tassinary, L. G. (2005). Perceiving Sex Directly and Indirectly: Meaning in Motion and Morphology. *Psychological Science*, *16*(11), 890–897. https://doi.org/10.1111/j.1467-9280.2005.01633.x

Jolly, E. (2018). Pymer4: Connecting R and Python for Linear Mixed Modeling. *Journal of Open Source Software*, *3*(31), 862. https://doi.org/10.21105/joss.00862

Kana, R. K., Maximo, J. O., Williams, D. L., Keller, T. A., Schipul, S. E., Cherkassky, V. L., Minshew, N. J., & Just, M. A. (2015). Aberrant functioning of the theory-of-mind network in children and adolescents with autism. *Molecular Autism*, *6*(1), 59. https://doi.org/10.1186/s13229-015-0052-x

Kanai, R., Bahrami, B., Duchaine, B., Janik, A., Banissy, M. J., & Rees, G. (2012). Brain Structure Links Loneliness to Social Perception. *Current Biology*, *22*(20), 1975–1979. https://doi.org/10.1016/j.cub.2012.08.045

Landsiedel, J., Daughters, K., Downing, P. E., & Koldewyn, K. (2022). The role of motion in the neural representation of social interactions in the posterior temporal cortex. *NeuroImage*, 119533. https://doi.org/10.1016/j.neuroimage.2022.119533

Lee, S. M., Gao, T., & McCarthy, G. (2014). Attributing intentions to random motion engages the posterior superior temporal sulcus. *Social Cognitive and Affective Neuroscience*, *9*(1), 81–87. https://doi.org/10.1093/scan/nss110

Lessard, L. M., & Juvonen, J. (2018). Friendless Adolescents: Do Perceptions of Social Threat Account for Their Internalizing Difficulties and Continued Friendlessness? *Journal of Research on Adolescence*, *28*(2), 277–283. https://doi.org/10.1111/jora.12388

Li, G., Chen, Y., Wang, W., Dhingra, I., Zhornitsky, S., Tang, X., & Li, C.-S. R. (2020). Sex Differences in Neural Responses to the Perception of Social Interactions. *Frontiers in*

Human Neuroscience, *14*.
https://www.frontiersin.org/article/10.3389/fnhum.2020.565132

Lisøy, R. S., Biegler, R., Haghish, E. F., Veckenstedt, R., Moritz, S., & Pfuhl, G. (2022). Seeing minds – a signal detection study of agency attribution along the autism-psychosis continuum. *Cognitive Neuropsychiatry*, *0*(0), 1–17. https://doi.org/10.1080/13546805.2022.2075721

McNamara, Q., De La Vega, A., & Yarkoni, T. (2017). Developing a Comprehensive Framework for Multimodal Feature Extraction. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1567–1574. https://doi.org/10.1145/3097983.3098075

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, *12*(2), 153–157. https://doi.org/10.1007/BF02295996

Mohammadzadeh, A., Tehrani-doost, M., & Banaraki, A. K. (2012). Evaluation of ToM (intentionality) in primary school children using movement shape paradigm. *Procedia - Social and Behavioral Sciences*, *32*, 69–73. https://doi.org/10.1016/j.sbspro.2012.01.012

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

Osaka, N., Ikeda, T., & Osaka, M. (2012). Effect of Intentional Bias on Agency Attribution of Animated Motion: An Event-Related fMRI Study. *PLoS ONE*, *7*(11), e49053. https://doi.org/10.1371/journal.pone.0049053

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Palmer, C. J., & Clifford, C. W. G. (2020). Face Pareidolia Recruits Mechanisms for Detecting Human Social Attention. *Psychological Science*, *31*(8), 1001–1012. https://doi.org/10.1177/0956797620924814

Papeo, L. (2020). Twos in human visual perception. *Cortex*, *132*, 473–478. https://doi.org/10.1016/j.cortex.2020.06.005

Papeo, L., Stein, T., & Soto-Faraco, S. (2017). The Two-Body Inversion Effect. *Psychological Science*, *28*(3), 369–379. https://doi.org/10.1177/0956797616685769

Petrini, K., McAleer, P., Neary, C., Gillard, J., & Pollick, F. E. (2014). Experience in judging intent to harm modulates parahippocampal activity: An fMRI study with experienced CCTV operators. *Cortex*, *57*, 74–91. https://doi.org/10.1016/j.cortex.2014.02.026

Piejka, A., Piaskowska, L., & Okruszek, Ł. (2022). Two Means Together? Effects of Response Bias and Sensitivity on Communicative Action Detection. *Journal of Nonverbal Behavior*. https://doi.org/10.1007/s10919-022-00398-2

Pitcher, D., & Ungerleider, L. G. (2021). Evidence for a Third Visual Pathway Specialized for Social Perception. *Trends in Cognitive Sciences*, *25*(2), 100–110. https://doi.org/10.1016/j.tics.2020.11.006

Powers, K. E., Worsham, A. L., Freeman, J. B., Wheatley, T., & Heatherton, T. F. (2014). Social Connection Modulates Perceptions of Animacy. *Psychological Science*, *25*(10), 1943–1948. https://doi.org/10.1177/0956797614547706

Rasmussen, C. E., & Jiang, Y. V. (2019). Judging social interaction in the Heider and Simmel movie. *Quarterly Journal of Experimental Psychology*, *72*(9), 2350–2361. https://doi.org/10.1177/1747021819838764

Rothkirch, M., Madipakkam, A. R., Rehn, E., & Sterzer, P. (2015). Making eye contact without awareness. *Cognition*, *143*, 108–114. https://doi.org/10.1016/j.cognition.2015.06.012

Rutherford, M. D., & Kuhlmeier, V. A. (2013). *Social Perception: Detection and Interpretation of Animacy, Agency, and Intention*. MIT Press.

Sacco, D. F., Merold, S. J., Lui, J. H. L., Lustgraaf, C. J. N., & Barry, C. T. (2016). Social and emotional intelligence moderate the relationship between psychopathy traits and social perception. *Personality and Individual Differences*, *95*, 95–104. https://doi.org/10.1016/j.paid.2016.02.031

Schafroth, J. L., Basile, B. M., Martin, A., & Murray, E. A. (2021). No evidence that monkeys attribute mental states to animated shapes in the Heider–Simmel videos. *Scientific Reports*, *11*(1), 3050. https://doi.org/10.1038/s41598-021-82702-6

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*(8), 299–309. https://doi.org/10.1016/S1364-6613(00)01506-0

Shen, X., Tokoglu, F., Papademetris, X., & Constable, R. T. (2013). Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *NeuroImage*, *82*, 403–415. https://doi.org/10.1016/j.neuroimage.2013.05.081

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Tavares, P., Lawrence, A. D., & Barnard, P. J. (2008). Paying Attention to Social Meaning: An fMRI Study. *Cerebral Cortex*, *18*(8), 1876–1885. https://doi.org/10.1093/cercor/bhm212

Tomova, L., Wang, K. L., Thompson, T., Matthews, G. A., Takahashi, A., Tye, K. M., & Saxe, R. (2020). Acute social isolation evokes midbrain craving responses similar to hunger. *Nature Neuroscience*, *23*(12), 1597–1605. https://doi.org/10.1038/s41593-020-00742-z

Tremoulet, P. D., & Feldman, J. (2000). Perception of Animacy from the Motion of a Single Object. *Perception*, *29*(8), 943–951. https://doi.org/10.1068/p3101

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., & Ugurbil, K. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, *80*, 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192–196. https://doi.org/10.3758/BF03206482

Walbrin, J., Downing, P., & Koldewyn, K. (2018). Neural responses to visually observed social interactions. *Neuropsychologia*, *112*, 31–39. https://doi.org/10.1016/j.neuropsychologia.2018.02.023

Walbrin, J., & Koldewyn, K. (2019). Dyadic interaction processing in the posterior temporal cortex. *Neuroimage*, *198*, 296–302. https://doi.org/10.1016/j.neuroimage.2019.05.027

Williams, E., & Chakrabarti, B. (2022). *The Integration of Head and Body Cues during the Perception of Social Interactions*. PsyArXiv. https://doi.org/10.31234/osf.io/qc3vz

Wood, N., & Cowan, N. (1995). The cocktail party phenomenon revisited: How frequent are attention shifts to one's name in an irrelevant auditory channel? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 255–260. https://doi.org/10.1037/0278-7393.21.1.255

1178

1179