Received: 29 November 2021 | Accepted: 3 August 2022

DOI: 10.1111/ejn.15823

RESEARCH REPORT

EJN European Journal of Neuroscience FENS WILEY

Domain-specific and domain-general neural network engagement during human–robot interactions

Ann Hogenhuis[1] | **Ruud Hortensius**[2]

[1]Liberal Arts and Sciences, Utrecht University, Utrecht, The Netherlands

[2]Department of Psychology, Utrecht University, Utrecht, The Netherlands

Correspondence
Ruud Hortensius, Department of Psychology, Utrecht University, Heidelberglaan 1, 3584 CS Utrecht, The Netherlands.
Email: r.hortensius@uu.nl

Abstract

To what extent do domain-general *and* domain-specific neural network engagement generalize across interactions with human and artificial agents? In this exploratory study, we analysed a publicly available functional MRI (fMRI) data set ($n = 22$) to probe the similarities and dissimilarities in neural architecture while participants conversed with another person or a robot. Incorporating trial-by-trial dynamics of the interactions, listening and speaking, we used whole-brain, region-of-interest and functional connectivity analyses to test response profiles within and across social or non-social, domain-specific and domain-general networks, that is, the person perception, theory-of-mind, object-specific, language and multiple-demand networks. Listening to a robot compared to a human resulted in higher activation in the language network, especially in areas associated with listening comprehension, and in the person perception network. No differences in activity of the theory-of-mind network were found. Results from the functional connectivity analysis showed no difference between interactions with a human or robot in within- and between-network connectivity. Together, these results suggest that although largely similar regions are activated when speaking to a human and to a robot, activity profiles during listening point to a dissociation at a lower level or perceptual level, but not higher order cognitive level.

KEYWORDS

human–robot interaction, social cognition, social interaction, two-person neuroscience

1 | INTRODUCTION

An intricate collection of brain networks supports interactions between people. Although some of these networks show distinct response profiles dedicated to specific tasks, for example, understanding hidden mental states, other networks are domain-general and are active during a wide variety of tasks. Together activity in these networks influences the quality and outcome of an interaction (Feng et al., 2021; Redcay & Schilbach, 2019), for example, the level of affiliation and degree of trust. In recent years, studies have begun to ask if these networks extent to other social agents and support engagements with robots (Hortensius & Cross, 2018; Wykowska, 2021). Although most studies have been focussing on activity during the perception of robots in a small number of

regions, mostly within the social domain, a careful assessment of the functional architecture during interactions with robots is warranted.

Moving away from passive perception of individuals presented on a screen, recent studies mapped the functional neural architecture during interactions between individuals (Hari et al., 2015; Schilbach et al., 2013). These studies show subtle difference in neural activation between screen-based and embodied interactions, suggesting that screen-based interactions only capture some aspects of human social behaviour. In parallel, a shift from the sole focus on domain-specific neurocognitive processes supporting social interactions to a focus that includes domain-general neurocognitive processes has been taking place (Barrett & Satpute, 2013; Lockwood et al., 2020; Ramsey & Ward, 2020; Spunt & Adolphs, 2017). According to these perspectives, social interaction can be viewed as being built on specialized, for example, theory of mind, and general neurocognitive processes, for example, control mechanisms. This notion has entered the field of interactions with artificial agents (Cross & Ramsey, 2021), probing the question if and how interactions with artificial agents that can or cannot be seen as social, such as robots, are supported by social and non-social, domain-specific and domain-general networks.

Studies on the perception and interaction with artificial agents, such as social robots, suggest that activity profiles across brain regions can be divided across two levels of neurocognitive processes (Agnieszka et al., 2016; Hortensius et al., 2018; Hortensius & Cross, 2018). Most of these regions that show a similar response profile during human–robot interaction (HRI) as during interactions with other people map onto networks that are related to our perception of other agents, such as the person perception network (Hortensius et al., 2018). Regions that show attenuated activity are associated with higher order neurocognitive processes, mainly theory of mind (Hortensius & Cross, 2018). For example, Chaminade et al. (2010) showed increased activity in the fusiform face area (FFA) and decreased activity in the temporoparietal junction (TPJ) during the perception of emotions displayed by a robot compared to emotions expressed by a human. Similarly, Rauchbauer et al. (2019) showed that during interactions with an embodied social robot, activity in the TPJ decreased compared to interactions with another person.

These past studies focussed on only a few brain regions and networks mostly associated with social processes, thereby only providing a glimpse into the neurocognition of HRIs. This focus on the person perception and theory-of-mind networks potentially biases the assessment of the similarities and differences in neurocognitive processes between interactions with human and artificial agents (Cross & Ramsey, 2021; Henschel et al., 2020). Indeed, engagement with robots, from passive observations of emotions and actions to ongoing interactions, consistently activates object-specific regions in the brain (Henschel et al., 2020), thereby suggesting that a wider net needs to be cast in order to understand the underlying mechanism of these new forms of communication. Similar to the potential overlap between robots and humans in terms of neurocognitive profile, the potential overlap between robots and objects should be considered (Cross & Ramsey, 2021; Henschel et al., 2020). Recently, Cross and Ramsey (2021) called for the inclusion of domain-general cognition, beyond human-based and self-oriented social cognition. In this perspective, a combination of domain-specific and domain-general cognition, including their associated neural networks, should be investigated to probe the neurocognition of HRI. HRI could rely on domain-general cognition such as memory and semantics, beyond or instead of specialized social cognitive processes, such as theory of mind. Indeed, the behavioural and neural mechanisms supporting understanding the mind of a robot do not completely overlap with those supporting the understanding of human minds (Hortensius et al., 2021).

Here, we close this gap and systematically test social and non-social, domain-specific and domain-general neural network engagement during HRIs using whole-brain, region-of-interest (ROI) and functional connectivity analyses. Moving beyond passive observation, we consider the ongoing and natural dynamics during social interactions with an embodied robot (Henschel et al., 2020) and particularly focus on networks related to both perception, for example, person and object perception networks, and cognition, for example, the theory-of-mind and multiple-demand networks. This exploratory network approach allows for a more data-driven and complete assessment potentially uncovering hidden patterns, by considering networks that are related to social and non-social processes, as well as domain-specific and domain-general processes, which might play a role during social interaction with robots (Cross & Ramsey, 2021; Henschel et al., 2020).

2 | MATERIALS AND METHODS

2.1 | Data statement

The publicly available data set from Rauchbauer et al. (2019) was used for analyses and extracted from OpenNeuro (Poldrack & Gorgolewski, 2017; ds001740, https://

openneuro.org/datasets/ds001740/versions/2.2.0). This data set is part of a multimodal corpus collected during the conversation with a human and robotic agent, consisting of behavioural, physiological and functional MRI (fMRI) data. As the authors report the details of this corpus elsewhere (Rauchbauer et al., 2019, 2020), we describe the relevant details related to the acquisition of fMRI data and the processing of the conversational data (https://hdl.handle.net/11403/convers/v2), and the data processing and analyses performed in the current study.

2.2 | Participants

Twenty-five participants completed the experiment. All had normal or corrected-to-normal vision and no history of psychiatric or neurological disorders. Participants received information prior to the study but remained naïve to the goal of the study, provided written informed consent and were reimbursed upon completion of the study. Following Rauchbauer et al. (2019), three participants were excluded from the final analysis. One participant was excluded for not following the task instructions correctly, whereas two other participants were excluded due to technical issues during data acquisition. The final sample comprised 15 females and 7 males, between 18 and 49 years old.

2.3 | Experimental paradigm

As part of a cover story, participants were invited to discuss a marketing campaign on fruit and vegetables. The task involved a real-life bidirectional conversation with a person or a conversational robot on images of the marketing campaign via a live video feed (Figure 1). Besides gender-matched confederates, participants conversed with a social robot called Furhat (Furhat Robotics, Al Moubayed et al., 2012). Participants were told that as an autonomous conversational agent, this robot had information on the marketing campaign. The Furhat robot is an embodied agent with a human form and contains a semi-transparent mask on which a human face is back projected. To increase the comparability between the two conversational agents, the authors added a wig, glasses and clothes that resembled the gender-matched human confederate. The robot was controlled through a Wizard of Oz set-up by the same confederate as in the human–

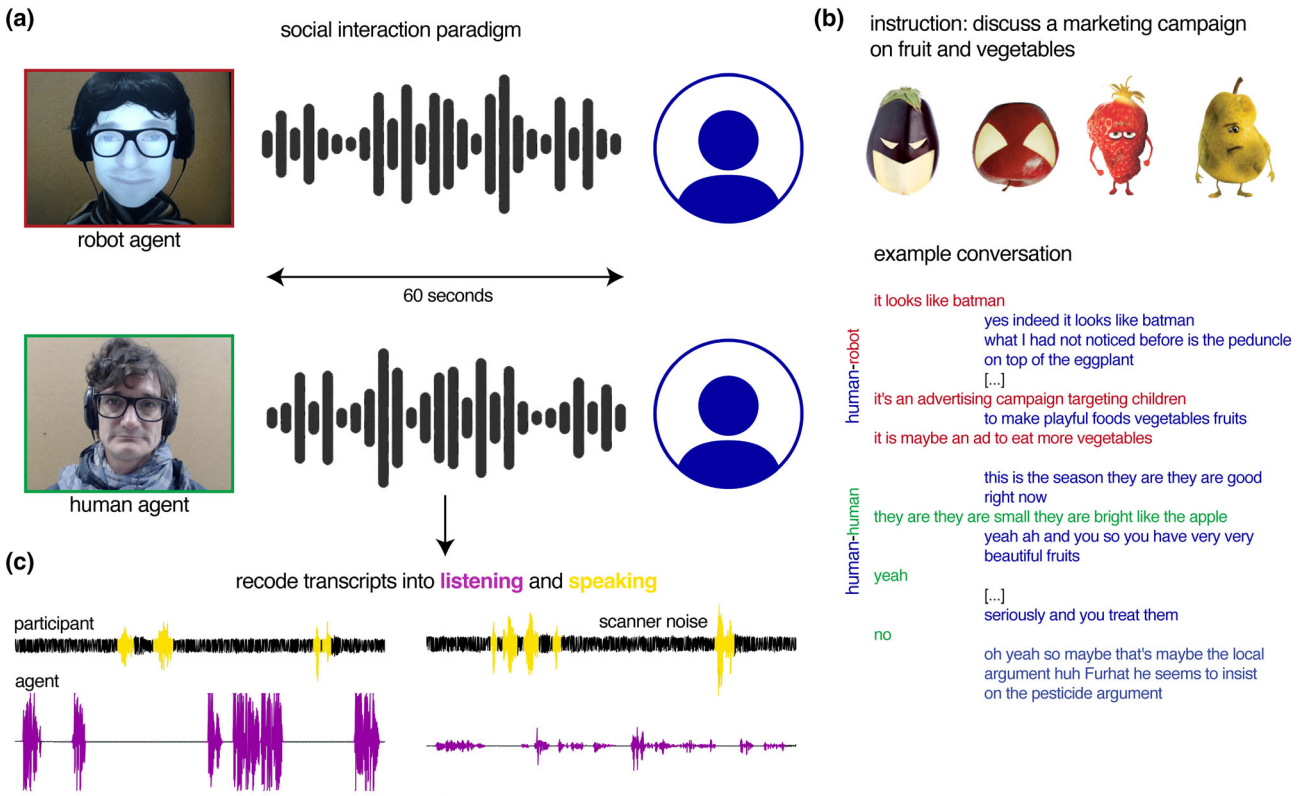

FIGURE 1 Task and procedure. (a) Participants had a real-life bidirectional conversation with a human and Furhat robot and (b) discussed a marketing campaign on fruit and vegetables. We used the transcripts from each 60 s conversation to recode audible speech segments into speaking (when the participants spoke) and listening events (when the agent spoke and the participant listened). (c) Translated example conversations are taken from Rauchbauer et al. (2019).

human conversations. The confederate used a web interface to select the appropriate pre-recorded vocal response. These conversational responses were partly based on previous acquired data on human–human interactions (HHIs) (Chaminade, 2017). About 30 conversational responses were scripted for each image. The order of the conversational agents was not randomized but alternated through each run. Each participant engaged in 12 conservations of 60 s with each conversational agent, resulting in 24 min of recorded conversations in total. As reported in Rauchbauer et al. (2019), all participants were unaware of the Wizard of Oz set-up and believed the cover story.

2.4 | fMRI data acquisition

Functional and structural MRI data were extracted from the data set. Whole-brain MRI data were acquired with a 3 T Siemens Prisma MRI scanner using a 20-channel coil (Siemens Medical, Erlangen, Germany) at the Centre IRM-INT in Marseille, France. Functional images were acquired using an echo-planar imaging sequence (repetition time: 1205 ms; echo time: 30 ms; number of slices [axial, co-planar to anterior/posterior commissure plan] per volume: 54; 2.5 mm isotropic resolution; flip angle: 65°; field of view: $210 \times 210\ mm^2$; matrix size: $84 \times 84\ mm^2$; multiband acquisition factor: 3; and number of volumes per run: 385). A high-resolution structural image was collected for each participant using a GR_IR sequence (repetition time: 2.4 ms; echo time: .00228 ms; .8 mm isotropic resolution; 320 sagittal slices; and field of view: $204.8 \times 256 \times 256$ mm). A field map was collected in the same session (repetition time: 7060 ms; echo time: 59 ms; 2.5 mm isotropic resolution; flip angle: 90°; and field of view: $210 \times 210\ mm^2$).

2.5 | fMRI preprocessing

Data quality was assessed before further preprocessing using imaging quality metrics calculated using MRIQC (Version 0.15.2; Esteban et al., 2017). Signal-to-noise ratio ranged from 1.89 to 2.64 across the data set, whereas mean ± standard deviation (*SD*) framewise displacement (Power et al., 2014) was .23 ± .11. After this initial quality check, further steps were taken in form of preprocessing the raw images. The results included in this manuscript come from preprocessing performed using FMRIPREP 20.2.1 (Esteban et al., 2019; Esteban, Markiewicz, Goncalves, et al., 2020; RRID:SCR_016216), which is based on nipype 1.5.1 (Esteban, Markiewicz, Burns, et al., 2020; RRID:SCR_002502; Gorgolewski et al., 2011).

2.6 | Anatomical data preprocessing

A total of 2 T1-weighted (T1w) images were found within the input BIDS data set. All of them were corrected for intensity non-uniformity with N4BiasFieldCorrection (Tustison et al., 2010), distributed with ANTS 2.3.3 (Avants et al., 2008; RRID:SCR_004757). The T1w reference was then skull stripped with a NIPYPE implementation of the antsBrainExtraction.sh workflow (from ANTS), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid, white matter and grey matter was performed on the brain-extracted T1w using fast (FSL 5.0.9, RRID:SCR_002823; Zhang et al., 2001). A T1w-reference map was computed after registration of 2 T1w images (after intensity non-uniformity correction) using mri_robust_template (FREESURFER 6.0.1; Reuter et al., 2010). Volume-based spatial normalization to two standard spaces (MNI152NLin2009cAsym and MNI152NLin6Asym) was performed through non-linear registration with antsRegistration (ANTS 2.3.3), using brain-extracted versions of both T1w reference and the T1w template. The following templates were selected for spatial normalization: ICBM 152 Nonlinear Asymmetrical template Version 2009c ([Fonov et al., 2009], RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym) and FSL's MNI ICBM 152 non-linear 6th Generation Asymmetric Average Brain Stereotaxic Registration Model ([Evans et al., 2012], RRID:SCR_002823; TemplateFlow ID: MNI152NLin6Asym).

2.7 | Functional data preprocessing

For each of the four BOLD runs found per subject (across all tasks and sessions), the following preprocessing steps were performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of FMRIPREP. A B0-non-uniformity map (or field map) was directly measured with an MRI scheme designed with that purpose (typically, a spiral pulse sequence). The field map was then co-registered to the target echo-planar imaging reference run and converted to a displacements field map (amenable to registration tools such as ANTS) with FSL's fugue and other SDCflows tools. Based on the estimated susceptibility distortion, a corrected echo-planar imaging reference was calculated for a more accurate co-registration with the anatomical reference. The BOLD reference was then co-registered to the T1w reference using flirt (FSL 5.0.9, Jenkinson & Smith, 2001) with the boundary-based registration (Greve & Fischl, 2009) cost function. Co-registration was configured with nine degrees of freedom to account for distortions remaining in the BOLD reference.

Head-motion parameters with respect to the BOLD reference (transformation matrices and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL 5.0.9; Jenkinson et al., 2002). BOLD runs were slice-time corrected using 3dTshift from AFNI 20160207 (Cox & Hyde, 1997; RRID:SCR_005927). The BOLD time series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. These resampled BOLD time series will be referred to as preprocessed BOLD in original space, or just preprocessed BOLD. The BOLD time series were resampled into standard space, generating a preprocessed BOLD run in MNI152NLin2009cAsym space. First, a reference volume and its skull-stripped version were generated using a custom methodology of FMRIPREP. Automatic removal of motion artefacts using independent component analysis (ICA-AROMA; Pruim et al., 2015) was performed on the preprocessed BOLD on MNI space time series after removal of non-steady state volumes and spatial smoothing with an isotropic, Gaussian kernel of 6 mm full width at half maximum. Corresponding 'non-aggressively' denoised runs were produced after such smoothing. Additionally, the 'aggressive' noise-regressors were collected and placed in the corresponding confounds file. Several confounding time series were calculated based on the preprocessed BOLD: framewise displacement, DVARS and three region-wise global signals. Framewise displacement was computed using two formulations following Power (absolute sum of relative motions; Power et al., 2014) and Jenkinson (relative root-mean-square displacement between affines; Jenkinson et al., 2002). Framewise displacement and DVARS were calculated for each functional run, both using their implementations in NIPYPE (following the definitions by Power et al., 2014). The three global signals are extracted within the cerebrospinal fluid, white-matter and whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (CompCor; Behzadi et al., 2007). Principal components were estimated after high-pass filtering the preprocessed BOLD time series (using a discrete cosine filter with 128 s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components were then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (cerebrospinal fluid, white matter and combined cerebrospinal fluid + white matter) are generated in anatomical space. The implementation differed from that of Behzadi et al., in that instead of eroding the masks by two pixels on BOLD space, the aCompCor masks have subtracted a mask of pixels that likely contain a volume fraction of grey matter. This mask was obtained by thresholding the corresponding partial volume map at .05, and it ensures that components are not extracted from voxels containing a minimal fraction of grey matter. Finally, these masks were resampled into BOLD space and binarized by thresholding it at .99 (as in the original implementation). Components were calculated separately within the white-matter and cerebrospinal fluid masks. For each CompCor decomposition, the '*k*' components with the largest singular values were retained, such that the retained components' time series are sufficient to explain 50% of variance across the nuisance mask (cerebrospinal fluid, white matter, combined or temporal). The remaining components were dropped from consideration. The head-motion estimates calculated in the correction step were placed within the corresponding confounds file. The confound time series derived from head-motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al., 2013). Frames that exceeded a threshold of .5 mm framewise displacement or 1.5 standardized DVARS were annotated as motion outliers. All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e., head-motion transform matrices, susceptibility distortion correction when available and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTS), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Non-gridded (surface) resamplings were performed using mri_vol2surf (FREESURFER).

Many internal operations of FMRIPREP use NILEARN (Abraham et al., 2014, RRID:SCR_001362), principally within the BOLD-processing workflow. For more details of the pipeline, see the section corresponding to workflows in FMRIPREP's documentation.

The preprocessed BOLD images were used for the first-level and second-level analyses, whereas functional correlation analysis was performed on the ICA-AROMA non-aggressively denoised images. Final nuisance regression was conducted on the ICA-AROMA non-aggressively denoised images using the DENOISER toolbox (Tambini & Gorgolewski, 2020). Besides nuisance signal removal from white-matter, cerebrospinal fluid and global signal sources, the data were high-pass filtered (.01). To validate the functional correlation analysis, final nuisance regression was performed with and without global signal regression (Liu et al., 2017).

2.8 | fMRI data analysis

First-level and second-level analyses were carried out using SPM12 (Wellcome Trust Centre for Neuroimaging, London) in MATLAB 2018a and 2021a (MathWorks, Natick, MA, USA). Besides the original events from Rauchbauer et al. (2019), events for HHI and HRI and the presentation of the images (baseline), we created new events based on the transcriptions of the conversations (https://hdl.handle.net/11403/convers/v2). These events correspond to when the participant spoke and when they listened to the agent (human or robot). One transcript was missing for one run for a participant. Only audible speech segments were used, laughter, inaudible speech segments or scanner noise were not coded in this analysis. More listening events were coded during the HHI, mean ± *SD* number of events: 164.23 ± 20.01, compared to HRI, 104.96 ± 22.22, whereas a comparable number of speaking events were coded for the HHI, 147.45 ± 32.71, and HRI, 155.77 ± 50.56. The duration of each listening and speaking event was shorter for HRI, 1.22 s ± 1.04 and 1.37 ± 1.09, respectively, compared to the HHI, 1.66 s ± 1.52 and 1.54 s ± 1.29. Besides these events, predictors of no interest were included (framewise displacement and six head-motion parameters) and a subset of the anatomical CompCor confounds (i.e., white-matter and cerebrospinal fluid decompositions). Images were masked with a grey-matter mask (threshold: .8). To capture the neural dynamics of interactions with a robot at a trial-by-trial and network level, we ran (1) whole-brain analysis, (2) ROI and (3) functional connectivity analyses.

Besides an initial analysis that successfully replicated the whole-brain analysis of Rauchbauer et al. (2019) using the following simple contrasts, interaction versus baseline and HHI versus HRI (as well as the reversed contrast; Figure S1), we focussed on the dynamics of the conversation in the main whole-brain analysis. We calculated the following contrasts using the new coded events, listening versus speaking for the interactions combined as well as for the HHI and HRI separately, and HHI versus HRI for speaking and listening events separately. All contrast images were smoothed using a 5 mm smoothing kernel. For the second-level analyses, one-sample *t* tests were used for each data set (initial single voxel threshold: $p_{\text{uncorrected}} < .001$, $k = 10$, with an average grey-matter mask applied, cluster-level threshold: $p_{\text{FWE-corrected}} \leq .05$). Labelling of regions was based on the Anatomy Toolbox in SPM (Eickhoff et al., 2005).

Closer inspection of the recording of the interaction led us to believe that the ease of understanding the other agent could potentially influence the results. For example, if the robot is easier to understand, increased activity in brain regions associated with listening comprehension might reflect this confound. To control for differences in voice quality or more general signal to noise between the interactions, we used harmonic-to-noise ratio as a predictor in a control analysis. The mean harmonic-to-noise ratio was calculated for each interaction using PARSELMOUTH-PRAAT (Boersma & Weenink, 2021; Jadoul et al., 2018). Directly contrasting the HRI with the HHI showed that the harmonic-to-noise ratio voice was higher in the HRI compared to the HHI, $t(21) = -3.48$, $p = .002$, Cohen's $d = -.74$, 95% confidence interval [−1.21, −.26]. The mean harmonic-to-noise ratio was centred and scaled and used as a parametric predictor and added to the extended whole-brain analysis that incorporated the type as well as the dynamics of the interaction.

For the ROI and functional connectivity analyses, group-based ROIs (9 mm sphere) were created for the five networks (Tables S1–S5). Literature-derived coordinates were used for the theory-of-mind network (bilateral TPJ, precuneus [PC], dorsomedial prefrontal cortex [dMPFC], middle medial prefrontal cortex [mMPFC] and ventromedial prefrontal cortex [vMPFC]; Richardson et al., 2018) and person perception network (bilateral FFA, occipital face area [OFA], extrastratiate body area [EBA] and posterior superior temporal sulcus [pSTS]; Julian et al., 2012). Coordinates for object-selective regions were derived from Julian et al. (2012) (bilateral lateral occipital complex [LOC] and superior parietal lobule [SPL]), as well as Henschel et al. (2020) (bilateral fusiform gyrus [FG], superior parietal lobule [SPL], and middle occipital gyrus [MOG]; using independent coordinates from Dubey et al., 2020). For the language (12 regions) and multiple-demand networks (20 regions), we used the data from Diachek et al. (2020; Experiment 1, $n = 383$, https://osf.io/pdtk9/) and drew spheres around the peak voxels of the group maps masked with the network parcels. For the ROI analysis, beta estimates were extracted for each event for each ROI and averaged across networks. These estimates were entered in a 2 (interaction: HHI and HRI) × 2 (dynamics: listening and speaking) repeated-measures analysis of variance (ANOVA) for each network separately. Given the exploratory nature of this study, Bonferroni correction was used to correct for multiple comparison ($p < .05/5$) and η^2_G was calculated as effect size measure.

To delineate the functional connectivity within and across networks, we calculated functional correlation indices during the interaction. Given the short duration of each listening and speaking event during the interaction, we used the entire 60 s conversation, thus speaking and listening combined. For this functional correlation analysis, the averaged *z*-transformed time course across voxels was extracted for each ROI per subject using

NILEARN (Abraham et al., 2014, RRID:SCR_001362). Pearson's correlation coefficients were calculated between these time courses for all possible combinations of ROIs per interaction (HHI and HRI) for each subject. After Fisher z transformation, the average within-network (e.g., every theory-of-mind region to every theory-of-mind region) and between-network (e.g., every theory-of-mind region to every language region) correlations were calculated (Blank et al., 2014; Paunov et al., 2019; Richardson, 2019; Richardson et al., 2018). A paired-sample t test was used to test for differences in functional correlation between HHI and HRI for each within- and between-network combination (Bonferroni correction $p < .05/15$). Lastly, we tested for temporal effects, that is, the repeated experience of these interactions, on functional connectivity by comparing the within- and between-network correlation between the two interactions by separating the four runs using a 2 (interaction: HHI and HRI) × 4 (run: 1–4) repeated-measures ANOVA for each network. ROI and functional connectivity analyses were executed in R Version 4.2.0 (2022) using the AFEX package (v1.1-1; Singmann et al., 2021) with post hoc tests executed using the EMMEANS package (1.7.4-1; Lenth, 2021).

To provide further evidence on the strength of the relationship and evidence for the null and alternative hypothesis, we used Bayesian regression models implemented in the BRMS package (Version 2.21.5; Bürkner, 2017) with STAN (Version 2.21.2; Carpenter et al., 2017) For the ROI analysis, we specified the following linear model: value ~ dynamics.d * interaction.d + (1 | sub), with value representing the beta estimates extracted for each event for each ROI averaged across networks, and dynamics and type of interaction as fixed effects and a random intercept for participants (sub). To allow for comparison with null hypothesis significance testing (http://talklab.psy.gla.ac.uk/tvw/catpred/; e.g., Bara et al., 2021), deviation coding was used with dynamics and type of interaction coded as .5 (listening and HHI) and −.5 (speaking and HRI). For the functional connectivity analyses, the following model was specified: corz ~ run * interaction + (run | sub), with corz representing the Fisher z-transformed Pearson's correlation coefficients between the time courses for all possible combinations of ROIs within or between the network(s) with type of interaction (HHI and HRI) and run (1–4) as fixed effects and a random intercept and slope for participants (run | sub). The hypothesis function was used to specify the slope for the HRI: run + run:interactionhri = 0. The models were fitted with weakly informative prior, normal (0,1) and a Gaussian distribution. Four Markov chains with 4000 iterations (and a warm-up of 2000 iterations) were used. We calculated the posterior distributions with 95% credible intervals for the ROI and functional connectivity analyses. All models converged, with Rhat values below 1.1.

3 | RESULTS

Considering the dynamics of the interaction, the whole-brain analysis revealed patterns of activation in regions associated with language comprehension when contrasting listening to speaking and in regions associated with language production when contrasting speaking to listening regardless of type of interaction (Figure 2a). Although these patterns overlapped between interactions with humans and robots, differences in activity patterns between the type of interactions appeared when directly contrasting the interaction type for listening and speaking separately (Figure 2a and Table 1). Critically, listening to a human compared to a robot resulted in increased activity in the superior parietal lobule and bilateral inferior parietal lobule, middle frontal gyrus, paracingulate gyrus and left middle temporal gyrus. The reversed contrast revealed increased activity in the bilateral Heschl's gyrus, lateral occipital cortex, insular cortex and inferior frontal gyrus. More subtle differences were observed when contrasting speaking to a human with speaking to a robot, with more activation observed in bilateral temporal pole when speaking to human compared to robotic agent. These results cannot be attributed to differences in voice quality, as similar results were obtained when using harmonic-to-noise ratio as a parametric predictor (Figure S2).

The results of the ROI analysis revealed a significant interaction between dynamics and type of interaction in the person perception, $F(1, 21) = 24.75$, $p < .001$, $\eta^2_G = .09$, and language networks, $F(1, 21) = 51.50$, $p < .001$, $\eta^2_G = .10$ (Figure 2b,c and Table S6). In both these networks, listening to a robot led to more activation compared to listening to a human (person perception: $t(21) = -5.289$, $p < .001$; language: $t(21) = -5.022$, $p = .001$), whereas speaking to a human led to more activation compared to speaking to a robot for the language, $t(21) = 3.402$, $p = .003$, but not the person perception network, $t(21) = 1.12$, $p = .28$. Other networks did not show an interaction between dynamics and type of interaction. A main effect of dynamic was found in the theory-of-mind, $F(1, 21) = 101.48$, $p < .001$, $\eta^2_G = .37$, and multiple-demand networks, $F(1, 21) = 18.47$, $p < .001$, $\eta^2_G = .09$, with more activation observed for listening compared to speaking. Although a main effect of type of interaction was found in the object-specific network, increased activation for HRI compared to HHI, $F(1, 21) = 6.69$, $p = .017$, $\eta^2_G = .07$,

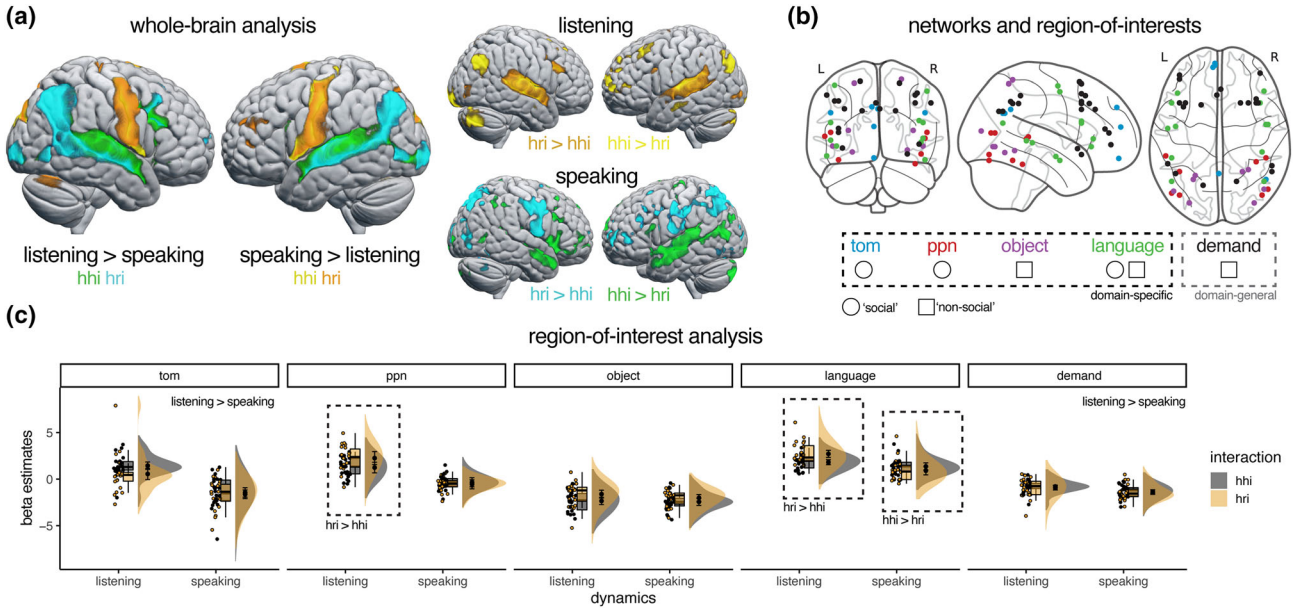

FIGURE 2 Neural network engagement during interactions with a human and robotic agent. (a) Whole-brain analysis showed robust engagement of regions associated with language comprehension and language production during listening and speaking respectively for both interactions with humans and robots. Directly contrasting listening to a robot with listening to a human revealed more activation in the bilateral Heschl's gyrus, lateral occipital cortex, insular cortex and inferior frontal gyrus, whereas the reverse contrast showed more activation in the superior parietal lobule and bilateral inferior parietal lobule, middle frontal gyrus, paracingulate gyrus and left middle temporal gyrus. Only subtle differences were observed when contrasting speaking to a robot with speaking to a human or vice versa. (b) Group-based regions of interest for the theory-of-mind (tom), person perception (ppn), object-specific (object), language and multiple-demand networks (demand) were used in the region-of-interest and functional connectivity analyses, mapping social and non-social, domain-specific and domain-general networks. (c) Listening to a robot led to more activity compared to listening to a human in the person perception and language networks. Regardless of interaction, listening compared to speaking led to more activity in the theory-of-mind and multiple-demand networks. For visual purposes activation maps are shown with an uncorrected threshold of $p < .001$ ($k = 10$) in (a). Rain cloud plots with errors bars reflecting 95% confidence intervals are used in (c) (Allen et al., 2021).

this effect did not survive correction for multiple comparison. Bayesian estimation provided similar results, with a bias towards increased activity for listening compared to speaking for the theory-of-mind, estimated posterior coefficient: 1.99, 95% credibility interval [1.34–2.65], and multiple-demand networks, .44 [.08–.80] (Table S7 and Figure S3). Similarly, activity in the person perception, −.96 [−1.60 to −.32], and language networks, −1.14 [−1.71 to −.58], was modulated by both dynamics and type of interaction. For all other comparisons, 0 was included in the 95% credibility interval.

No differences in within-network functional connectivity were found between interactions with a robot and interactions with a human (Figure 3a,b and Table S8). Increased functional connectivity during HHI compared to HRI was observed between the language and theory-of-mind, $t(21) = 2.65$, $p = .015$, language and multiple-demand, $t(21) = 2.18$, $p = .040$, person perception and multiple-demand, $t(21) = 2.86$, $p = .0094$, and object-specific and multiple-demand networks, $t(21) = 3.16$, $p = .0047$ (Figure 3c). Rerunning the pre-process and analysis pipeline without global signal regression revealed only increased functional connectivity during HHI compared to HRI between the language and theory-of-mind networks, $t(21) = 2.19$, $p = .04$ (Table S8). However, all functional connectivity results did not survive correction for multiple comparison. These results were corroborated by Bayesian analysis, with no robust within- and between-network functional connectivity differences observed between HHI compared to HRI, including the language and theory-of-mind, −.02 [−.06 to .01], person perception and multiple-demand, −.02 [−.05 to .00], and object-specific and multiple-demand networks, −.03 [−.06 to .00] (Table S8). Within- or between-network connectivity did not increase or decrease in the course of the HHI or HRI. Across analyses, no temporal effects on functional connectivity were found across or between interaction type (Figure S3 and Tables S9 and S10).

TABLE 1 Regions associated with the type and dynamics of the interaction

Anatomical region	Cytoarchitectonic location	MNI coordinates			*t* value	Cluster size	p_{FWE}
		x	*y*	*z*			
Listening: HHI > HRI							
Cerebellum right crus II		14	−88	−41	6.10	343	<.001
Cerebellum left crus II		−14	−88	−41	6.16	236	<.001
Superior parietal lobule	7A	−9	−58	56	5.95	203	<.001
Inferior parietal lobule	PGp	−36	−80	42	6.71	162	<.001
Inferior parietal lobule	PGp	46	−80	32	7.41	134	.001
Middle frontal gyrus		−36	25	52	7.32	110	.003
Right caudate		21	5	24	5.47	86	.011
Paracingulate gyrus	p32	−9	50	−4	5.99	84	.013
Middle temporal gyrus	te5	−69	−10	−18	5.81	78	.019
Listening: HRI > HHI							
Planum temporale/Heschl's gyrus	TE 1.1	41	−28	12	12.22	2423	<.001
Planum temporale/Heschl's gyrus	TE 1	−46	−18	2	9.84	2100	<.001
Lateral occipital cortex	FG2	−42	−78	−14	7.21	900	<.001
Lateral occipital cortex	FG2	46	−62	−16	6.88	438	<.001
Intracalcarine cortex	hOc1 (V1)	8	−72	9	5.23	194	<.001
Inferior frontal gyrus	45	54	22	24	7.70	168	<.001
Insular cortex	Id7	−32	22	−1	7.96	159	<.001
Insular cortex	Id7	31	25	−1	7.43	102	.004
Inferior frontal gyrus	44	−39	15	24	5.39	77	.021
Speaking: HHI > HRI							
Temporal pole	TE 3	−54	15	−11	7.74	694	<.001
Temporal pole	TE 5	58	5	−18	7.69	451	<.001
Cerebellum left crus I		−44	−80	−38	6.65	130	.001
Speaking: HRI > HHI							
Cerebellum right crus I		36	−75	−28	5.79	117	.002
Occipital fusiform gyrus	hOc4v (V4(v))	−26	−82	−8	4.66	64	.048

Note: Results from the whole-brain analysis for the different contrasts. Only clusters with a cluster $p_{FWE\text{-corrected}} \leq .05$ ($p_{uncorrected} < .001$, $k = 10$, with an average grey-matter mask applied) are reported. MNI coordinates and *t* value of the peak voxels are reported.

4 | DISCUSSION

In the present exploratory study, we aimed to map the similarities and differences in domain-specific and domain-general neural network engagement during embodied and recursive interactions with a human or robotic agent. Employing whole-brain, ROI and functional connectivity analyses, we mapped functional activity and connectivity across networks associated with domain-general and domain-specific, as well as social and non-social cognitive processes. Whole-brain and ROI analyses suggest that activity in the language network, especially regions associated with listening comprehension, and the person perception network, differentiated between HHI and HRI. No differences in the theory-of-mind network nor robust differences in within- or between-network connectivity between the interactions were observed when considering the dynamics of the interactions. Overall, these results suggest that interactions with an artificial entity such as a social robot might lead to only subtle differences in response profiles of neural networks at a perceptual but not cognitive level.

In the quest for understanding the neurocognitive foundation of social interaction with human and artificial agents, these results help advance our understanding in several ways. The present results could implicate a dissociation between the neurocognitive circuitry supporting different social interactions. This dissociation may shift

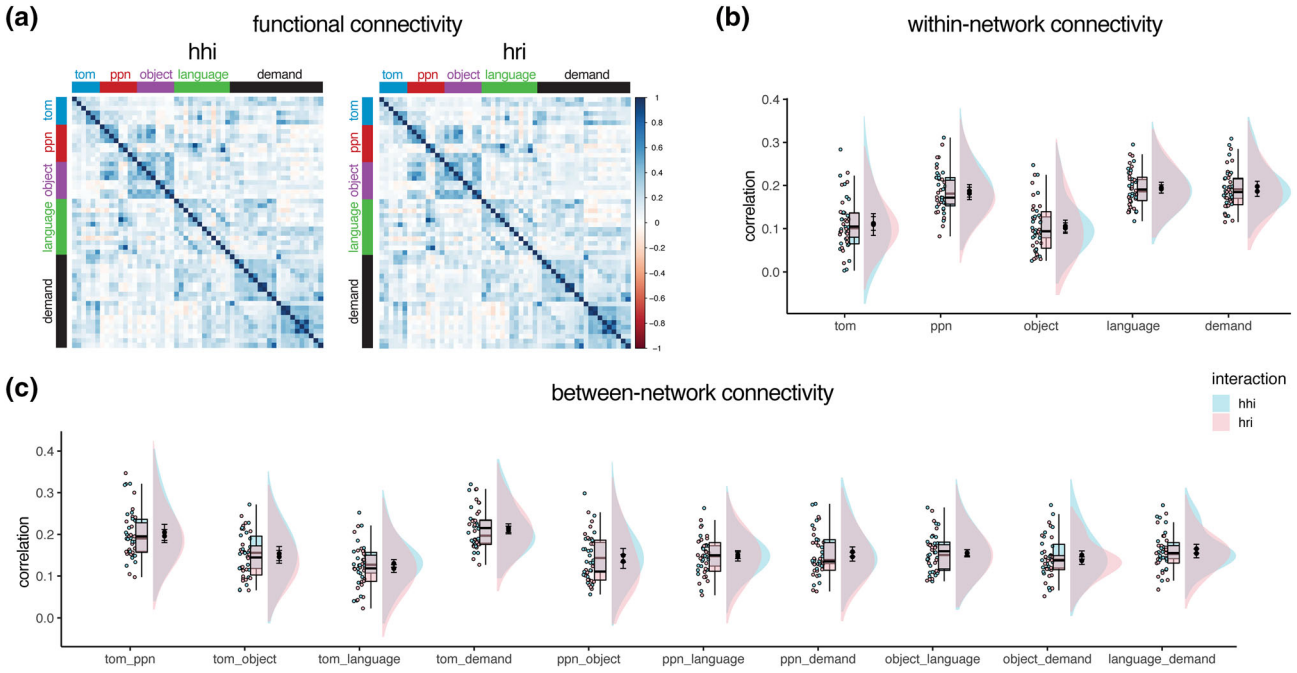

FIGURE 3 Functional network connectivity during interactions with a human and robotic agent. (a) Functional connectivity for interactions with a human (hhi) and robot (hri). (b, c) No differences were found in within-network and between-network connectivity between interactions with a human and robot. Raw correlations are shown in (a), whereas *z*-transformed correlations are shown in (b) and (c). Rain cloud plots with errors bars reflecting 95% confidence intervals are used in (b) and (c) (Allen et al., 2021).

from the level of social and non-social networks to the level of low-level perceptual and higher order neurocognitive processes and networks. Although no differences in activity or connectivity profiles of the theory-of-mind network were found, both the person perception and language networks showed increased activation when listening to a robot compared to listening to a human. This suggests that only during distinct aspects of an interaction (cf. listening), low-level perceptual, as reflected by person perception and language network engagement, but not higher order neurocognitive processes, as reflected by theory-of-mind network engagement, might be affected by the artificial nature of the interacting agent.

Going beyond the mere focus on social brain regions and social cognitive processes, the question arises what role non-social, for example, object-specific regions, play during HRIs (Henschel et al., 2020). This contributes to the debate if robots should be framed as social agents or can be viewed as an object or unique category supported by different neurocognitive processes (Cross & Ramsey, 2021; Prescott, 2017). A recent observation is that engagements with robots consistently activate object-specific regions (Henschel et al., 2020). The fusiform gyrus, middle occipital gyrus, and the inferior parietal lobule are activated during perception of actions (Cross et al., 2012) and emotions expressed by a robot (Cross et al., 2019). In line with these findings, we observed that regions in the object-specific network show higher activation for HRI compared to HHI. Although our results are exploratory in nature, they may point to the involvement of networks beyond social networks.

A broader question pertaining to not only interactions with artificial agents but also with other individuals is to what extent domain-specific and domain-general networks support these interactions. Our results suggest that no clear differences in activation and connectivity profiles of the multiple-demand network are visible when contrasting interactions with other individuals with interactions with artificial agents. All in all, differences between interactions only appear for some domain-specific networks. Ramsey and Ward (2020) put forward a hybrid model for information processing during social interactions that captures the interplay between domain-specific and domain-general networks and provide a first view on the present results. In their hybrid model, both person representation supported by domain-specific processes and domain-general control process play key roles during social interaction. Person representation is informed by social cognitive processes and associated networks, with three possible representational levels: perceptual (e.g., face perception), cognitive (cf. theory of mind), and valence and affect. Across these so-called person-feature maps, information is integrated. For

instance, body perception is supported by processing in not only the person perception network but also the theory-of-mind network (Ramsey, 2018). Besides person representation, control processes play a crucial role in social interaction as well (Ramsey & Ward, 2020). These processes guide ongoing interaction by means of integrated priority maps. The latter maps receive information from of exogenous (person-feature maps across the three levels) and endogenous cues (e.g., goals, memory and an individual's current affective state).

Integration of information occurs both within and between the person representation and control processes in the form of biased competition. Biases at one level (e.g., perceptual person-feature map) influence a subsequent level (e.g., cognitive person-feature map) (Ramsey & Ward, 2020). Although we found subtle differences in functional activation in domain-specific networks, we did not find any differences in functional connectivity within and between domain-specific and domain-general networks. Integrating these findings with the hybrid model of Ramsey and Ward (2020) suggests that person-feature maps at the level of perception, but not cognition, might be impacted by the artificial nature of the agent. Integration and processing of information within and across the person representation and control systems might also be intact during HRI. Future research should replicate and extend our findings and formally test if indeed only subtle perceptual differences between interactions with human and robotic agents are observed.

Mapping activity across a series of networks with different cognitive functionalities, we look beyond individual regions or a few domain-specific social networks (Hortensius et al., 2018; Hortensius & Cross, 2018; Wiese et al., 2017; Wykowska, 2021). Using this network approach, we replicate and extend the original study by Rauchbauer et al. (2019). Although they reported decreased activation in the temporoparietal junction and medial prefrontal cortex during interactions with a robot, we provide a nuance to this observation by considering the dynamics of the interaction. Critically, when considering the dynamics of the interactions, listening and speaking, we do not find decreased activation in these and other regions of the theory-of-mind network. Similar to the observations by Rauchbauer et al. (2019), previous studies mostly reported attenuated activation for the theory-of-mind network when people engage with artificial agents such as social robots (Hortensius & Cross, 2018). The contrast between screen-based, restricted and one-off interactions (previous studies) and embodied and recursive interactions that consider the dynamics of these interactions (current study) likely explains this difference in findings (e.g., Schilbach, 2014). Future studies should explore what the functional significance is of increased or decreased activation in specific networks. Of course, the labelling of networks as social or non-social or domain-general or domain-specific has some level of arbitrarity to it. To fully understand the neurocognitive representation of social cognition during interactions with artificial agents requires more advanced approaches beyond analysing activity patterns, such as representational similarity and multivoxel pattern analyses, together with a network approach (Henschel et al., 2020).

Exposure and beliefs to robots drive how these artificial entities are perceived and interacted with (Agnieszka et al., 2016; Hortensius & Cross, 2018). This experience dependency shapes the neurocognitive response during engagements with robots (Cross et al., 2016; Gowen et al., 2016; Klapper et al., 2014; Özdem et al., 2017; Wykowska et al., 2014). Similarly, the hybrid model of information processing is experience dependent in nature, where informational cues, like beliefs and expectations, influence the ongoing interaction (Ramsey & Ward, 2020). To test similar effects of experience in the current study, we looked into temporal effects on functional connectivity. No temporal effects on within- or between-network connectivity were observed; that is, the connectivity patterns remained stable across the different interactions. However, the current time window is relatively short (less than an hour). Employing a between-session approach, a recent study did not report a change in activity in the pain matrix, a collection of brain regions responsive to observations of another individual's distress or pain, after a 5 day interaction period with a robot (Cross et al., 2019). Using longer time windows (multiple sessions across days/weeks) would allow to answer questions on the experience dependency of the neurocognitive architecture supporting interactions with artificial agents. Future studies can test how external (e.g., the duration of the interaction) and internal (e.g., an individual's expectation of the robot) factors influence the engagement of domain-specific and domain-general, social and non-social networks during interactions with social robots.

5 | CONCLUSION

In this exploratory study, we tested neural network engagement during HRIs using whole-brain, ROI and functional connectivity analyses. Going beyond selective brain regions, we analysed the activity and connectivity patterns across networks that can be viewed as social or non-social, and domain-specific or domain-general and that form the building blocks of social interaction. Together, our results point to a dissociation between interactions with human and robotic agents at the perceptual, but not cognitive level.

ACKNOWLEDGEMENTS

We thank Birgit Rauchbauer and Thierry Chaminade for sharing their data and providing additional information on the data set and Kohinoor Darda and Phil McAleer for helpful suggestions during data analyses. We thank the support from the Human-centered Artificial Intelligence focus area at Utrecht University.

CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS

Ann Hogenhuis: Conceptualization; data curation; formal analysis; investigation; methodology; project administration; resources; software; validation; visualization; writing—original draft preparation; writing—review and editing. **Ruud Hortensius:** Conceptualization; data curation; formal analysis; investigation; methodology; project administration; resources; software; supervision; validation; visualization; writing—original draft preparation; writing—review and editing.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/ejn.15823.

DATA AVAILABILITY STATEMENT

The publicly available neuroimaging data set can be accessed via OpenNeuro (https://openneuro.org/datasets/ds001740/versions/2.2.0), whereas the conversational data can be accessed via the Open Resources and TOols for LANGuage (https://hdl.handle.net/11403/convers/v2). Data associated with the secondary data analysis are publicly available on the OSF (https://osf.io/dby4j/), whereas the code can be found on the accompanying GitLab page (https://gitlab.com/human-plus/fchri).

ORCID

Ruud Hortensius https://orcid.org/0000-0002-5279-6202

REFERENCES

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*, 14. https://doi.org/10.3389/fninf.2014.00014

Agnieszka, W., Thierry, C., & Gordon, C. (2016). Embodied artificial agents for understanding human social cognition. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *371*(1693), 20150375. https://doi.org/10.1098/rstb.2015.0375

Al Moubayed, S., Beskow, J., Skantze, G., & Granström, B. (2012). Furhat: A back-projected human-like robot head for multiparty human-machine interaction. In A. Esposito, A. M. Esposito, A. Vinciarelli, R. Hoffmann, & V. C. Müller (Eds.), *Cognitive behavioural systems*. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7403 LNCS. (pp. 114–130). Springer. https://doi.org/10.1007/978-3-642-34584-5_9

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., van Langen, J., & Kievit, R. A. (2021). Raincloud plots: A multi-platform tool for robust data visualization [version 2; peer review: 2 approved]. *Wellcome Open Research*, *4*, 63. https://doi.org/10.12688/wellcomeopenres.15191.2

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26–41. https://doi.org/10.1016/j.media.2007.06.004

Bara, I., Darda, K. M., Kurz, A. S., & Ramsey, R. (2021). Functional specificity and neural integration in the aesthetic appreciation of artworks with implied motion. *European Journal of Neuroscience*, *54*(9), 7231–7259. https://doi.org/10.1111/ejn.15479

Barrett, L. F., & Satpute, A. (2013). Large-scale brain networks in affective and social neuroscience: Towards an integrative functional architecture of the brain. *Current Opinion in Neurobiology*, *23*(3), 361–372. https://doi.org/10.1016/j.conb.2012.12.012

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Blank, I., Kanwisher, N., & Fedorenko, E. (2014). A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. *Journal of Neurophysiology*, *112*(5), 1105–1118. https://doi.org/10.1152/jn.00884.2013

Boersma, P., & Weenink, D. (2021). Praat: Doing phonetics by computer [computer program]. Version 6.1.44.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(i01), 1. https://EconPapers.repec.org/RePEc:jss:jstsof:v:080:i01

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*, 1–32. https://doi.org/10.18637/jss.v076.i01

Chaminade, T. (2017). An experimental approach to study the physiology of natural social interactions. *Interaction Studies*, *18*(2), 254–275. https://doi.org/10.1075/is.18.2.06gry

Chaminade, T., Zecca, M., Blakemore, S.-J., Takanishi, A., Frith, C. D., Micera, S., Dario, P., Rizzolatti, G., Gallese, V., & Umiltà, M. A. (2010). Brain response to a humanoid robot in areas implicated in the perception of human emotional gestures. *PLoS ONE*, *5*(7), e11577. https://doi.org/10.1371/journal.pone.0011577

Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine*, *10*(4–5), 171–178. https://doi.org/10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L

Cross, E. S., Liepelt, R., Hamilton, A. F. D. C., Parkinson, J., Ramsey, R., Stadler, W., & Prinz, W. (2012). Robotic movement preferentially engages the action observation network. *Human Brain Mapping*, *33*(9), 2238–2254. https://doi.org/10.1002/hbm.21361

Cross, E. S., & Ramsey, R. (2021). Mind meets machine: Towards a cognitive science of human–machine interactions. *Trends in Cognitive Sciences*, *25*(3), 200–212. https://doi.org/10.1016/j.tics.2020.11.009

Cross, E. S., Richard, R., Roman, L., Wolfgang, P., & Hamilton, A. F. D. C. (2016). The shaping of social perception by stimulus and knowledge cues to human animacy. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *371*(1686), 20150075. https://doi.org/10.1098/rstb.2015.0075

Cross, E. S., Riddoch, K. A., Pratts, J., Titone, S., Chaudhury, B., & Hortensius, R. (2019). A neurocognitive investigation of the impact of socializing with a robot on empathy for pain. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *374*(1771), 20180034. https://doi.org/10.1098/rstb.2018.0034

Diachek, E., Blank, I., Siegelman, M., Affourtit, J., & Fedorenko, E. (2020). The domain-general multiple demand (MD) network does not support core aspects of language comprehension: A large-scale fMRI investigation. *The Journal of Neuroscience*, *40*(23), 4536–4550. https://doi.org/10.1523/JNEUROSCI.2036-19.2020

Dubey, I., Georgescu, A. L., Hommelsen, M., Vogeley, K., Ropar, D., & Hamilton, A. F. D. C. (2020). Distinct neural correlates of social and object reward seeking motivation. *European Journal of Neuroscience*, *52*(9), 4214–4229. https://doi.org/10.1111/ejn.14888

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, *25*(4), 1325–1335. https://doi.org/10.1016/j.neuroimage.2004.12.034

Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., & Gorgolewski, K. J. (2017). MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS ONE*, *12*(9), e0184661. https://doi.org/10.1371/journal.pone.0184661

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Esteban, O., Markiewicz, C. J., Burns, C., Goncalves, M., Jarecka, D., Ziegler, E., Berleant, S., Ellis, D. G., Pinsard, B., Madison, C., Waskom, M., Notter, M. P., Clark, D., Manhães-Savio, A., Clark, D., Jordan, K., Dayan, M., Halchenko, Y. O., Loney, F., … Ghosh, S. (2020). nipy/nipype: 1.5.1 [computer software]. Zenodo. https://doi.org/10.5281/zenodo.4035081

Esteban, O., Markiewicz, C. J., Goncalves, M., DuPre, E., Kent, J. D., Salo, T., Ciric, R., Pinsard, B., Blair, R. W., Poldrack, R. A., & Gorgolewski, K. J. (2020). fMRIPrep: A robust preprocessing pipeline for functional MRI. Zenodo. https://doi.org/10.5281/zenodo.4252786

Evans, A. C., Janke, A. L., Collins, D. L., & Baillet, S. (2012). Brain templates and atlases. *NeuroImage*, *62*(2), 911–922. https://doi.org/10.1016/j.neuroimage.2012.01.024

Feng, C., Eickhoff, S. B., Li, T., Wang, L., Becker, B., Camilleri, J. A., Hétu, S., & Luo, Y. (2021). Common brain networks underlying human social interactions: Evidence from large-scale neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews*, *126*, 289–303. https://doi.org/10.1016/j.neubiorev.2021.03.025

Fonov, V., Evans, A., McKinstry, R., Almli, C., & Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, *47*, S102. https://doi.org/10.1016/S1053-8119(09)70884-5

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Frontiers in Neuroinformatics*, *5*, 13. https://doi.org/10.3389/fninf.2011.00013

Gowen, E., Bolton, E., & Poliakoff, E. (2016). Believe it or not: Moving non-biological stimuli believed to have human origin can be represented as human movement. *Cognition*, *146*, 431–438. https://doi.org/10.1016/j.cognition.2015.10.010

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Hari, R., Henriksson, L., Malinen, S., & Parkkonen, L. (2015). Centrality of social interaction in human brain function. *Neuron*, *88*(1), 181–193. https://doi.org/10.1016/j.neuron.2015.09.022

Henschel, A., Hortensius, R., & Cross, E. S. (2020). Social cognition in the age of human–robot interaction. *Trends in Neurosciences*, *43*(6), 373–384. https://doi.org/10.1016/j.tins.2020.03.013

Hortensius, R., & Cross, E. S. (2018). From automata to animate beings: The scope and limits of attributing socialness to artificial agents. *Annals of the New York Academy of Sciences*, *1426*(1), 93–110. https://doi.org/10.1111/nyas.13727

Hortensius, R., Hekele, F., & Cross, E. S. (2018). The perception of emotion in artificial agents. *IEEE Transactions on Cognitive and Developmental Systems*, *10*(4), 852–864. https://doi.org/10.1109/TCDS.2018.2826921

Hortensius, R., Kent, M., Darda, K. M., Jastrzab, L., Koldewyn, K., Ramsey, R., & Cross, E. S. (2021). Exploring the relationship between anthropomorphism and theory-of-mind in brain and behaviour. *Human Brain Mapping*, *42*(13), 4224–4241. https://doi.org/10.1002/hbm.25542

Jadoul, Y., Thompson, B., & de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, *71*, 1–15. https://doi.org/10.1016/j.wocn.2018.07.001

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*(2), 143–156. https://doi.org/10.1016/S1361-8415(01)00036-6

Julian, J. B., Fedorenko, E., Webster, J., & Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage*, *60*(4), 2357–2364. https://doi.org/10.1016/j.neuroimage.2012.02.055

Klapper, A., Ramsey, R., Wigboldus, D., & Cross, E. S. (2014). The control of automatic imitation based on bottom–up and top–down cues to animacy: Insights from brain and behavior. *Journal of Cognitive Neuroscience*, *26*(11), 2503–2513. https://doi.org/10.1162/jocn_a_00651

Lanczos, C. (1964). A precision approximation of the gamma function. *Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis*, *1*(1), 86–96.

Lenth, R. V. (2021). emmeans: Estimated marginal means, aka least-squares means (1.7.4-1) [computer software]. https://CRAN.R-project.org/package=emmeans

Liu, T. T., Nalci, A., & Falahpour, M. (2017). The global signal in fMRI: Nuisance or information? *NeuroImage*, *150*, 213–229. https://doi.org/10.1016/j.neuroimage.2017.02.036

Lockwood, P. L., Apps, M. A. J., & Chang, S. W. C. (2020). Is there a 'social' brain? Implementations and algorithms. *Trends in Cognitive Sciences*, *24*(10), 802–813. https://doi.org/10.1016/j.tics.2020.06.011

Özdem, C., Wiese, E., Wykowska, A., Müller, H., Brass, M., & Overwalle, F. V. (2017). Believing androids—fMRI activation in the right temporo-parietal junction is modulated by ascribing intentions to non-human agents. *Social Neuroscience*, *12*(5), 582–593. https://doi.org/10.1080/17470919.2016.1207702

Paunov, A. M., Blank, I. A., & Fedorenko, E. (2019). Functionally distinct language and theory of mind networks are synchronized at rest and during language comprehension. *Journal of Neurophysiology*, *121*(4), 1244–1265. https://doi.org/10.1152/jn.00619.2018

Poldrack, R. A., & Gorgolewski, K. J. (2017). OpenfMRI: Open sharing of task fMRI data. *NeuroImage*, *144*, 259–261. https://doi.org/10.1016/j.neuroimage.2015.05.073

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, *84*, 320–341. https://doi.org/10.1016/j.neuroimage.2013.08.048

Prescott, T. J. (2017). Robots are not just tools. *Connection Science*, *29*(2), 142–149. https://doi.org/10.1080/09540091.2017.1279125

Pruim, R. H. R., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J. K., & Beckmann, C. F. (2015). ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *NeuroImage*, *112*, 267–277. https://doi.org/10.1016/j.neuroimage.2015.02.064

Ramsey, R. (2018). Neural integration in body perception. *Journal of Cognitive Neuroscience*, *30*(10), 1442–1451. https://doi.org/10.1162/jocn_a_01299

Ramsey, R., & Ward, R. (2020). Putting the nonsocial into social neuroscience: A role for domain-general priority maps during social interactions. *Perspectives on Psychological Science*, *15*(4), 1076–1094. https://doi.org/10.1177/1745691620904972

Rauchbauer, B., Hmamouche, Y., Bigi, B., Prevot, L., Ochs, M., & Thierry, C. (2020). Multimodal corpus of bidirectional conversation of human-human and human-robot interaction during fMRI scanning. In *Proceedings of the 12th Language Resources and Evaluation Conference* (pp. 661–668). European Language Resources Association. https://hal.archives-ouvertes.fr/hal-02612820

Rauchbauer, N. B., Morgane, B., Magalie, O., Laurent, P., & Thierry, C. (2019). Brain activity during reciprocal social interaction investigated using conversational robots as control condition. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *374*(1771), 20180033. https://doi.org/10.1098/rstb.2018.0033

Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience*, *20*, 495–505. https://doi.org/10.1038/s41583-019-0179-4

Reuter, M., Rosas, H. D., & Fischl, B. (2010). Highly accurate inverse consistent registration: A robust approach. *NeuroImage*, *53*(4), 1181–1196. https://doi.org/10.1016/j.neuroimage.2010.07.020

Richardson, H. (2019). Development of brain networks for social functions: Confirmatory analyses in a large open source dataset. *Developmental Cognitive Neuroscience*, *37*, 100598. https://doi.org/10.1016/j.dcn.2018.11.002

Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., & Saxe, R. (2018). Development of the social brain from age three to twelve years. *Nature Communications*, *9*(1), 1027. https://doi.org/10.1038/s41467-018-03399-2

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, *64*, 240–256. https://doi.org/10.1016/j.neuroimage.2012.08.052

Schilbach, L. (2014). On the relationship of *online* and *offline* social cognition. *Frontiers in Human Neuroscience*, *8*, 278. https://doi.org/10.3389/fnhum.2014.00278

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience 1. *Behavioral and Brain Sciences*, *36*(4), 393–414. https://doi.org/10.1017/S0140525X12000660

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2021). afex: Analysis of factorial experiments (1.1-1) [computer software]. http://cran.rproject.org/package=afex

Spunt, R. P., & Adolphs, R. (2017). A new look at domain specificity: Insights from social neuroscience. *Nature Reviews Neuroscience*, *18*(9), 559–567. https://doi.org/10.1038/nrn.2017.76

Tambini, A., & Gorgolewski, K. J. (2020). Denoiser: A nuisance regression tool for fMRI BOLD data (1.0.1) [computer software]. Zenodo. https://doi.org/10.5281/zenodo.4033939

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. M. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320. https://doi.org/10.1109/TMI.2010.2046908

Wiese, E., Metta, G., & Wykowska, A. (2017). Robots as intentional agents: Using neuroscientific methods to make robots appear more social. *Frontiers in Psychology*, *8*, 1663. https://doi.org/10.3389/fpsyg.2017.01663

Wykowska, A. (2021). Robots as mirrors of the human mind. *Current Directions in Psychological Science*, *30*(1), 34–40. https://doi.org/10.1177/0963721420978609

Wykowska, A., Wiese, E., Prosser, A., & Müller, H. J. (2014). Beliefs about the minds of others influence how we process sensory information. *PLoS ONE*, *9*(4), e94339. https://doi.org/10.1371/journal.pone.0094339

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57. https://doi.org/10.1109/42.906424

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Hogenhuis, A., & Hortensius, R. (2022). Domain-specific and domain-general neural network engagement during human–robot interactions. *European Journal of Neuroscience*, *56*(10), 5902–5916. https://doi.org/10.1111/ejn.15823