Human-robot interaction

The difference in functional connectivity during human-human interaction and human-robot interaction.

Universiteit Utrecht

Author: Ann L.M.P. Hogenhuis[1]

Supervisor: dr. Ruud Hortensius[2]

Second reader: Colin Caret[3]

25-02-2021, Utrecht University

Thesis Artificial Intelligence, Liberal Arts and Sciences, 15 ECTS

Highlights

- No difference in functional connectivity has been observed between human-human and human-robot interaction.
- The right temporoparietal junction (rTPJ) and right fusiform face area (rFFA) are stronger connected during interaction.
- The superior parietal lobule (SPL) is less connected with the rFFA.
- No temporal effects of conversating have been found on the functional connections between the regions of interest.

[1)] *BSc student (nr. 6019293), Liberal Arts and Sciences; Cognitive and Neurobiological Psychology at Utrecht University*

[2)] *Department of Psychology, Utrecht University, Heidelberglaan 1, 3584 CS, Utrecht, The Netherlands*

[3)] *Department of Philosophy and Religion studies, Utrecht University, Janskerkhof 13, Utrecht, 3512 BL, The Netherlands*

Abbreviations

HHI; human-human interaction, HRI; human-robot interaction, FFA; fusiform face area, FG; fusiform gyrus, TPJ; temporoparietal junction, MOG; middle occipital gyrus, MFG; middle frontal gyrus, STS; superior temporal sulcus, VMPFC; ventral medial prefrontal cortex, DMPFC; dorsal medial prefrontal cortex, FC; functional connectivity, FMRI; functional magnetic resonance imaging, BOLD; blood oxygenation level dependent, WOZ; Wizard of Oz, TR; repetition time, TE; echo time, MNI; Montreal Neurological Institute, EPI; echo-planar imaging, GM; general model, CSF; cerebrospinal fluid, GSR; global signal ratio, WM; white matter, ROI; region of interest, DMN; default-mode network

Box 1. Integrating AI with interdisciplinary studies

The emergence of robotic agents in our social environment is accompanied by the co-evolution of a new scientific field in which social- and beta sciences have to blend to satisfy future users. With a background in interdisciplinary studies, the resulting field of 'human-robot interaction' is ideal to integrate these multiple disciplines. Not only is the collaboration important for future inventions, investigating human-robot interaction allows me to personally develop on two tracks; coding and analyzing behavioral data. Curious? Check out GitLab for my artificial development and continue reading to explore the cognitive effect of humans communicating with robots!

Abstract

The developments in artificial intelligence are leading us towards a new scientific frontier of socially engaged robots. This poses new questions regarding the impact of unfamiliar agents that alter our social environment. In order to optimise the communication with these robots, the question that has been answered in the current study is whether the social interaction between humans is distinct from the social interaction with robots. In previous studies the neural embodiment of social features has been examined through the activation of the main social networks, i.e. the person perception- and the theory-of-mind network. Here, a novel approach was performed by extracting the functional connections between the hubs of these networks (fusiform face area; FFA & temporoparietal junction; TPJ). In order to do so, an existing fMRI dataset was analysed. For the experiment the participants (N = 22) alternately conversated with a robot and a person during scanning. Correlation analysis between the extracted time series revealed identical connectivity patterns for human-human and human-robot interaction (HHI & HRI) that are stable over time. Specifically, the connection between the FFA and TPJ was increased, while control regions exhibited decreased functional connectivity with the FFA during conversation. As a result, I propose a novel theory concerning a general interaction network that is connected during communication regardless of the type of conversational agent. To further explore this theory, future research should include a full connectome study to ascertain whether functional connectivity during human-robot interaction remains similar to human-human interaction.

Keywords: human-robot interaction, functional connectivity, social robotics, social interaction, fMRI

Table of contents

1. Introduction

A next frontier in the social life of humans is the interaction with artificial agents such as social robots. This may have many advantages, such as being a cure for loneliness, an educational tool for children with autism or for more practical uses in for example hospitals (Broadbent, 2017; Cross et al., 2019; Dautenhahn, 2007; Wiese et al., 2017). With such benefits it is likely that their introduction in our social environment will continue to grow. This poses many challenges regarding a new type of interaction with other autonomous agents. For example, one of the pitfalls in human-robot interaction (HRI) is that we have a desire to simulate mankind and create an as humanlike robotic exterior as possible. However a realistic appearance, while a robot is still an object, could actually be a deterrent (Mori et al., 2012). The fact that the human façade cannot easily be copied relates to the way in which humans are programmed. To predict behaviour, cognitive capacities like empathic concern and moral decision making are embedded in the brain (Adolphs, 2009; L. Schilbach, 2015; Wykowska et al., 2016). These skills make it easier to predict human behaviour, however if it makes artificial behaviour predictable as well is the question. To implement our social capacities on non-human agents, engeneers in artificial intelligence try to optimise every feature of the robot that may shape our perception. Integrating a neuropsychological viewpoint is therefore crucial to deploy the knowledge of human behaviour that shapes the social experience induced by HRI.

To delineate the neural mechanism of social interaction, the brain should be examined as a functional network in the form of a multi-wired organ instead of a collection of isolated regions (van den Heuvel & Hulshoff Pol, 2010). This approach results in a pattern of connected brain areas, i.e. networks. Based on recent reviews, the most involved brain networks corresponding to social interaction are the person perception network and the theory-of-mind network (Adolphs, 2009; Redcay & Schilbach, 2019; Schurz et al., 2014). The involvement of the person perception network is straightforward, knowing that its function relates to the recognition of comparable agents (Hortensius et al., 2018; Hortensius & Cross, 2018). Secondly, the theory-of-mind network, is associated with the embodiment of the widely reviewed theory that humans are not only able to recognize others, but that we could predict thoughts, intentions and motives as well (Mitchell, 2008; Saxe & Baron-Cohen, 2006; Saxe & Kanwisher, 2003). The theory-of-mind network could be complementary to the person perception network by using the generated information of recognition to infer mental states, resulting

in a more global network of interaction. (Greven et al., 2016; Greven & Ramsey, 2017; Koster-Hale & Saxe, 2013).

However, most theories remain based on communication between humans. For the interaction with a robot instead of a person on the other hand, it has not been determined whether a similar mechanism is applicable (Chaminade et al., 2007; Peelen & Downing, 2007; Redcay & Schilbach, 2019; L. Schilbach, 2015). In the last decades some studies have been performed regarding the communication with robots. For example, Chaminade and colleagues have demonstrated increased activity in the fusiform face area (FFA; key hub person perception network) during HRI and decreased activation in the temporoparietal junction (TPJ; key hub theory-of-mind network) compared to HHI (2010). Suggesting differential engagement of the theory-of-mind and person perception network. However the mechanisms that are primarily related to the communication with robots are debatable (Adolphs, 2009). Clear evidence for the neural mechanism behind HRI is therefore still missing. A possible explanation may rely on the current methods that have been used to analyse HRI. Namely, communication with a robot has been investigated in terms of activity. However, interaction depicts a multifaceted type of behaviour concerning several brain areas. A different way in which these multiple areas could by combined, is by looking at the connections between them. For example, Greven and Ramsey have provided evidence that the person perception network, especially the fusiform gyrus, is effectively connected to hubs of the theory-of-mind network during HHI (2016). Functional connectivity may therefore be a useful method to obtain more information on interaction. Still, the question remains if the results from the study of Greven and Ramsey applies for the interaction with robots as well. With the rise of robotic agents in our environment it is important to know whether humans will react in a similar manner. A promising way to obtain these insights in HRI is therefore to examine the functional connectivity between brain areas that are important for social interaction.

The question that will be answered in this article is to what extent humans communicate in a comparable way with robots as they do with humans. The aim of this study is thus to discover whether the functional connectivity during the interaction with humans is different from the functional connectivity observed during the interaction with robots. In order to do so, two key networks for social cognition, the person perception network and the theory-of-mind network, will be investigated. To narrow this study, the connectivity between the FFA, as part of the person perception network, and

the TPJ, as part of the theory-of-mind network, during HHI in comparison with HRI will be studied over several moments in time. By using functional connectivity as a measure in combination with these brain areas, two hypotheses were constructed. Firstly, it is expected that the functional connectivity between the TPJ and the FFA will be less for HRI than for HHI, because the prediction of robotic behaviour may not be as accurate as it is for human behaviour. Secondly, it is expected that the connection between the TPJ and the FFA will increase over time for HRI because participants may improve the skill to predict the behaviour of the robot, causing to see them as a more humanlike interlocutor by strengthening the connection.

2. Methods

2.1. Database

A publicly available dataset extracted from OpenNeuro (Poldrack & Gorgolewski, 2017) was analysed (Rauchbauer et al. 2019; ID: ds001740). In their paradigm, a multimodal corpus was collected during the conversation with an artificial agent, consisting of eye tracking, physiological and fMRI data. This database was approved in France by the ethics committee 'Comité de Protection des Personnes Sud Méditerrannée I'. The current secondary data analysis was approved by the Ethics Committee of the Faculty of Social and Behavioural Sciences of Utrecht University (protocol number: 20-0127).

2.2. Participants

Twenty-five participants (M_{age} = 28.5 yrs., s.d. = 12.4 yrs.) completed the experiment. They received information and provided written informed consent, but were naïve to the goal of the study. All had normal or corrected to normal vision and no history of psychiatric or neurological disorders. In the original study from Rauchbauer and colleagues, three participants were excluded from the final analysis because one participant did not perform the task correctly, while for other participants the scans revealed technical artefacts (2019). In the present study the same three participants were excluded, two based on excessive movement (> 3 mm) and one was removed due to technical issues during the data acquisition.

2.3. Experimental paradigm

During the experiment participants were instructed to have a real-life bidirectional conversation with a person and an artificial agent. To maintain naivety on the goal of the experiment there was an alternative rationale for the study. The cover story entailed that participants had to discuss the effectiveness of a marketing campaign about fruit and vegetables. Participants were told that they had to talk freely about the intentions of the campaign so that the company could test whether the message was clear to the public.

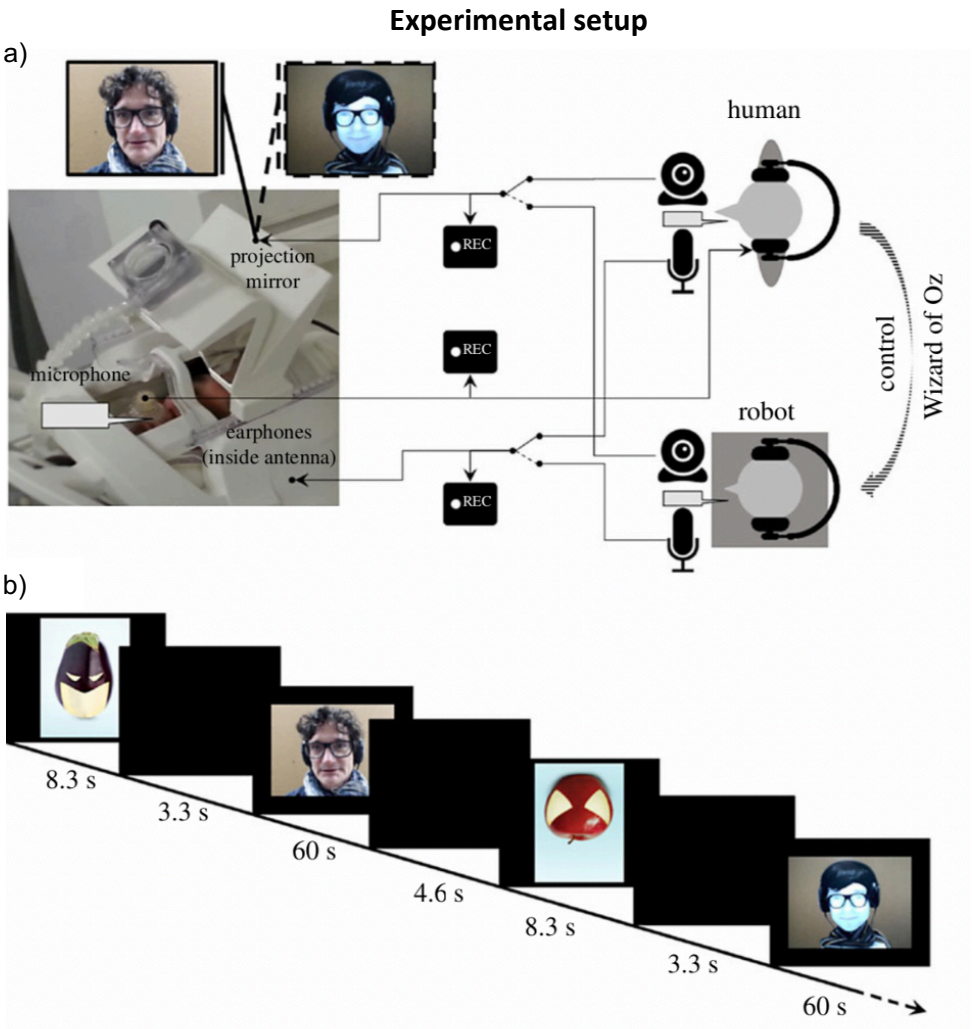

Figure 1. The experimental paradigm demonstrating the fMRI setup. a) The participants have been recorded during the conversation with a human and with a robot, which were projected on a screen using a mirror. The robot was controlled through the Wizard of Oz (WOZ) paradigm by the same person as in the human condition (this figure was extracted from Rauchbauer et al., 2019). b) In the bottom panel the queue of the experimental paradigm is illustrated. First a screen depicting the picture of the cover story was shown for 8.3 sec. followed by a black screen for 3.3 sec. After that the conversation with either a person or the robot began for about 60 sec. The sequence was repeated three times during one run, resulting in twelve repetitions and 24 conversations.

Each participant underwent four repetitions of three conversations with a robot and three conversations with a person. The sequence of the conversations alternated from robot to human. A block consisted of the presentation of an image for 8.3 sec., followed by a black screen for 3.3 sec. and the conversation for 60 sec. (Figure 1). In total, 24 minutes of conversation was recorded, including twelve minutes with an artificial agent and twelve minutes with a person. Importantly, the 'Wizard of Oz' (WOZ) paradigm was used, meaning that the robot was controlled by the same person as in the conversation with a human condition. Thus, possible differences found between the HHI and HRI conditions could be less likely ascribed to a quality difference in conversation per se. The participants were unaware of the WOZ set-up.

2.3.1. Artificial agent

The participants conversated with an artificial agent from Furhat robotics (Al Moubayed et al., 2012). This is a robot in the form of a semi-transparent mask on which a human face is projected (Figure 1). To make the robot more humanlike, the authors added a wig, glasses and clothes. The Furhat robot can be controlled through the WOZ method, i.e. when a researcher would press 'yes these are superheroes', this was said by the robot. These words have to be programmed on forehand, which means that a limited amount of answers can be replied during the conversation. Rauchbauer and colleagues stated that 30 French conversational feedbacks were scripted for each image.

2.4. fMRI data acquisition

All images are obtained through a 3 Tesla MRI scanner (Siemens Medical, Erlangen, Germany), equipped with a 20-channel coil. BOLD sensitive functional images using an EPI sequence were obtained [functional parameters: TR = 1205 ms, TE = 30 ms, 2.5 mm isotropic voxels, 65° flip angle, 54 axial slices, a field of view from 210mm – 210mm, a matrix of 84-84 mm, and multiband acquisition factor 3]. For the anatomical scans, the images were obtained with a GR_IR sequence [structural parameters: TR = 2,4 ms, TE = 0,00228 ms, 0.8 mm isotropic voxels, 320 axial slices, a field of view from 204.8 – 256 – 256 mm (Rauchbauer et al., 2019).

2.5. Preprocessing

First of all, to check the quality of the data and calculate the imaging quality-metrics, MRIQC was used (version 0.15.2; Esteban et al., 2019; Figure S1). After the 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, Markiewicz, et al. (2018); Esteban, Blair, et al. (2018); RRID:SCR_016216), which is based on Nipype 1.5.1 (Gorgolewski et al. (2011); Gorgolewski et al. (2018); RRID:SCR_002502).

2.5.1. Anatomical data preprocessing

A total of 2 T1-weighted (T1w) images were found within the input BIDS dataset. All of them were corrected for intensity non-uniformity (INU) 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 (CSF), white-matter (WM) and grey-matter (GM) was performed on the brain-extracted T1w using fast (FSL 5.0.9, RRID:SCR_002823, Zhang, Brady & Smith, 2001). A T1w-reference map was computed after registration of 2 T1w images (after INU-correction) using mri_robust_template (FreeSurfer 6.0.1, Reuter, Rosas & Fisch, 2010). Volume-based spatial normalization to two standard spaces (MNI152NLin2009cAsym, MNI152NLin6Asym) was performed through nonlinear 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], 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],

2.5.2. 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-nonuniformity 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 EPI (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 EPI (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 & Fisch, 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 6mm FWHM (full-width 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 (FD), DVARS and three region-wise global signals. FD 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). FD 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 CSF, the WM, and the whole-brain masks. Additionally, a set of physiological

regressors were extracted to allow for component-based noise correction (CompCor; Behzadi et al., 2007). Principal components were estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components were then calculated from the top two percent variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differed from that of Behzadi and colleagues, 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 GM. This mask was obtained by thresholding the corresponding partial volume map at 0.05, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks were resampled into BOLD space and binarized by thresholding it at 0.99 (as in the original implementation). Components were calculated separately within the WM and CSF 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 percent of variance across the nuisance mask (CSF, WM, 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 0.5 mm FD or 1.5 standardised 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).

2.6. fMRI data analysis

This study included the replication of the whole-brain analysis results from Rauchbauer et al. (2019). This analysis was conducted on the preprocessed BOLD images obtained from fMRIPrep (not the AROMA denoised images, as they will be used for the functional connectivity analysis) using SPM12

(Wellcome Trust Centre for Neuroimaging, London) in MATLAB 2020a (Mathworks, Natick, MA, USA). To reveal activated areas, a set of three contrasts was used 1) baseline > HRI + HHI, 2) HRI > HHI, 3) HHI > HRI. First-level analysis included a design matrix, containing these contrasts, several noninterest regressors (framewise displacement, head-motion correction, and a subset of the anatomical CompCor confounds, i.e. white matter and CSF decompositions), four times three HRI events and four times three HHI events. Before conducting group-analysis (p = *0.05*; false discovery rate at cluster level) the obtained images were smoothed using a 5mm smoothing kernel.

2.6.1. Functional connectivity extraction

Nuisance regression was conducted on the ICA-AROMA non-aggressively denoised images produced by fMRIPrep. Included parameters consisted of white matter, cerebrospinal fluid and global signal ratio to minimize the influence of motion and physiological artefacts on further processing. For denoising, the Toolbox 'denoiser' was used using Python 3 (Tambini & Gorgolewski, 2020). The obtained files were used for further steps in the functional connectivity analysis.

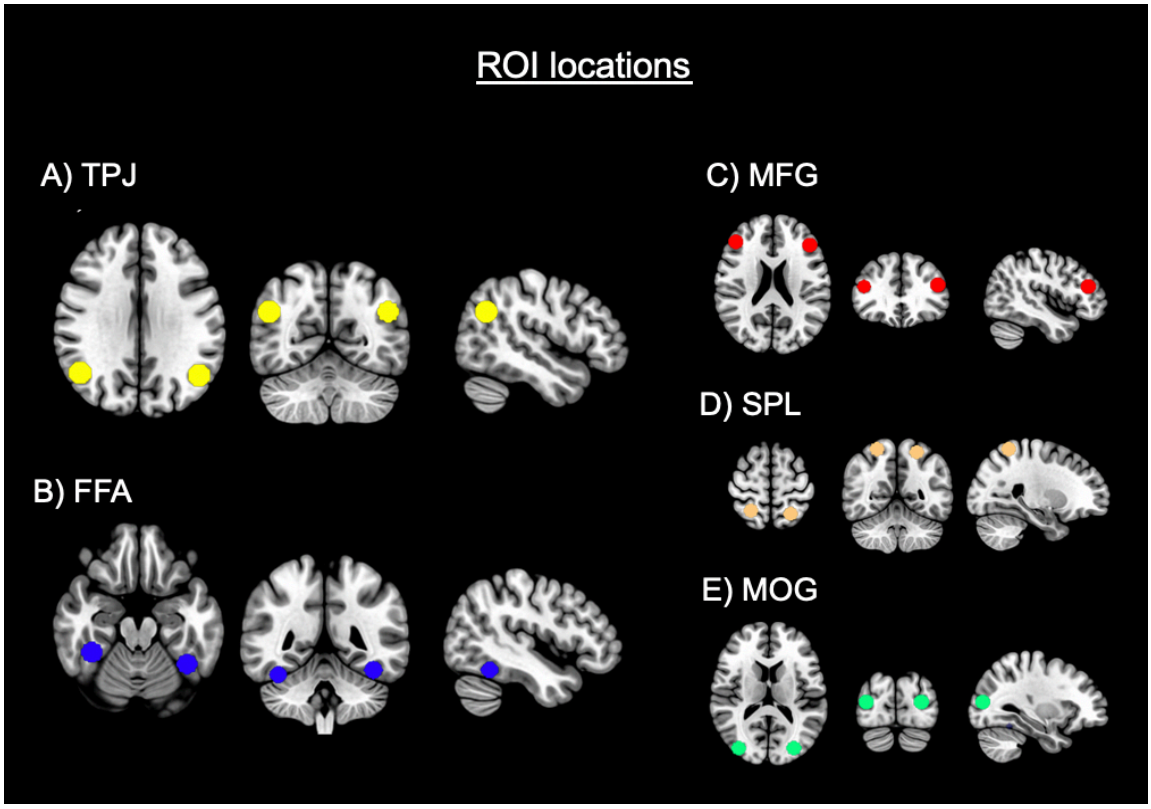

Figure 2. Region of Interest's (ROI's) have been extracted in a 9mm^3 sphere. The included regions are a) temporoparietal junction (TPJ); yellow, b) fusiform face area (FFA); blue, c) middle frontal gyrus (MFG); red, d) superior parietal lobule (SPL); orange, e) middle occipital gyrus (MOG); green. See the supplementary material for the exact coordinates (Table S2).

Based on existing literature, seed-based regions-of-Interest (ROI's) were established, namely the: fusiform face area (FFA), temporoparietal junction (TPJ), middle frontal gyrus (MFG), middle occipital gyrus (MOG) and superior parietal lobule (SPL) (Figure 2). The brain areas on which have been focussed in this analysis were the FFA, which is one of the main hubs in the person perception network, and the TPJ, which is in a central hub of the theory of mind network. However, no consensus has yet been reached on which areas would be of critical importance to HRI (Chaminade et al., 2018; Henschel et al., 2020; Krach et al., 2008; Rauchbauer et al., 2019; Wang & Quadflieg, 2014). Therefore, three control areas were included as well, namely the MFG, MOG and the SPL. The SPL and the MOG are central hubs of the object recognition network and studies have reported increased activity for the perception of robotic agents versus human agents (Henschel et al., 2020). These brain areas were used as control parameters to assess whether HRI activated more object-specific regions, instead of person perception related regions. The MFG is known to be a key hub in a more domain-general network of executive functioning, and will therefore control for a possible general effect of cognitive control (Shenhav et al., 2016). A nine mm^3 sphere was used for each ROI using the MarsBar toolbox (Brett et al., 2002) (for coordinates see Table S2) . For each participant, the overall mean time courses were extracted from these ROI's separately per run.

2.6.2. Statistical analysis

To delineate the strength of the connections between the ROI's the analysis has been divided into two sections, both comparing HHI and HRI. Correlation coefficients were normalised using Fisher's r-to-z-test. These values were tested according to Welch's t-test of unequal means, to discover whether there was a significant effect ($\alpha = 0.05$) between the three contrasts (Baseline > HRI+HHI, HHI>HRI, HRI>HHI. The divisions were as follows: 1) the total mean of all z-transformed time courses from each of the five ROI's (ten in total; right/left) have been correlated using Pearson's correlation test and were depicted in a correlation matrix. 2) Temporal results included all of the four runs separately to ascertain whether functional connectivity changes over time. To control for habituation effects, the relation between the conversations has been analysed over time as well. For this, the mean for each conversation was taken from all four runs per ROI.

3. Results

3.1. Whole-brain analysis

To test whether significant activity is detected in the FFA and TPJ while conversating, the contrasts Baseline>HHI+HRI, HHI>HRI and HRI>HHI were analysed with the FDR-corrected activity maps (p = *0.05)*. Replicating the findings of Rauchbauer and colleagues, similar activation patterns have been found for all contrasts (Figure 3). Compared to baseline, activation during overall interaction is found mainly in the occipital and temporal lobes, including the superior temporal sulcus (STS), and inferior frontal gyrus (IFG). For HRI parietal and frontal lobes are more activated, including the middle frontal gyrus (MFG) and inferior parietal lobule (IPL). Lastly, for HHI temporal regions coding for social interaction (e.g. TPJ & fusiform gyrus; FG) are activated. For a detailed overview, see the article of Rauchbauer et al. 2019. For both HHI and HRI, the TPJ and FFA were activated.

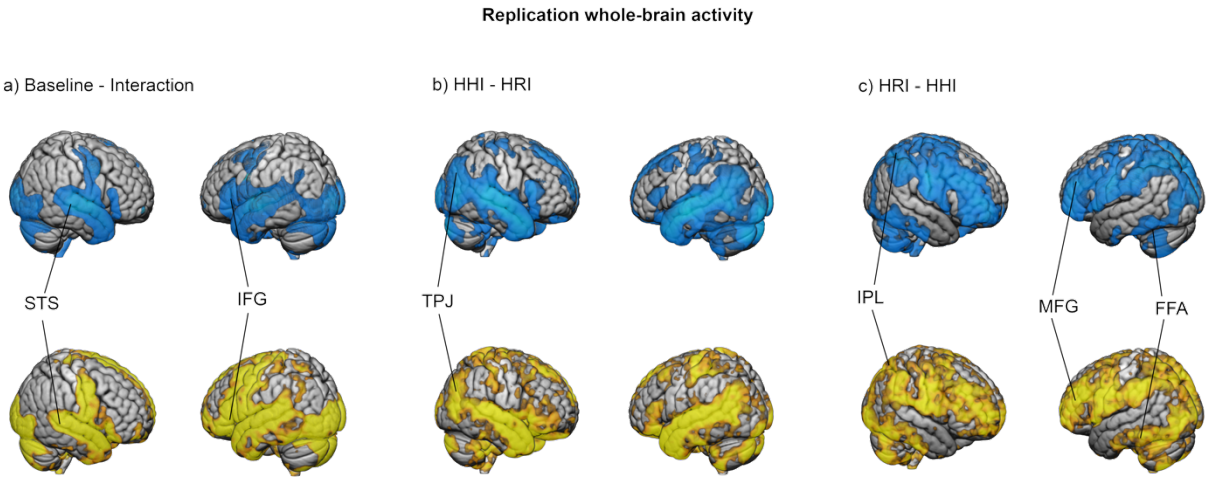

Figure 3. Results from Rauchbauer et al. (2019) have been replicated using the published untresholded contrast maps for the conditions: a) Baseline-Interaction b) human-robot interaction (HRI) - human-human interaction (HHI) c) human-human interaction (HHI) – human-robot interaction (HRI). The results from Rauchbauer et al. are depicted on the rendered MNI brain in the top row in blue (2019). Results from the current study are depicted in yellow. Whole-brain analyses revealed activation in the superior temporal sulcus (STS), inferior frontal gyrus (IFG), temporoparietal junction (TPJ), inferior parietal lobule (IPL), middle frontal gyrus (MFG), fusiform gyrus (FG). Small changes could be ascribed to different preprocessing procedures.

3.2. Functional connectivity analysis FFA-TPJ

For the functional connectivity analysis, I first sought to delineate the connections between the main regions of interest (FFA & TPJ) during the conversation with a robot and with a person. Overall, in contrast to the first hypothesis, no differences have been discovered between the conditions HRI and HHI (Figure 4). That is, no significant effect was obtained for the type of conversational partner, however the functional connectivity interaction did differ from baseline. In line with previous studies on social interaction (Hari et al., 2015; L. Schilbach, 2015), correlation analysis revealed significant (p < 0.001) functional connectivity between the core hubs of the person perception network (FFA) and the theory-of-mind network (TPJ) while conversating. Surprisingly this pattern was found for HRI and HHI, providing new evidence for a shared mechanism. Specifically, the TPJ demonstrates bilateral increased functional connectivity to the right FFA for HHI (rTPJ: r = 0.27, lTPJ: r = 0.16) and HRI (rTPJ:r = 0.26, lTPJ: r = 0.14; Figure 4a and 4b). The connection between the bilateral TPJ and the left FFA exhibits no significant change from baseline (HHI; rTPJ: r = -0.08, lTPJ: r = -0.02 & HRI; rTPJ: r = -0.07, lTPJ: r = -0.02; Figure 4c and 4d).

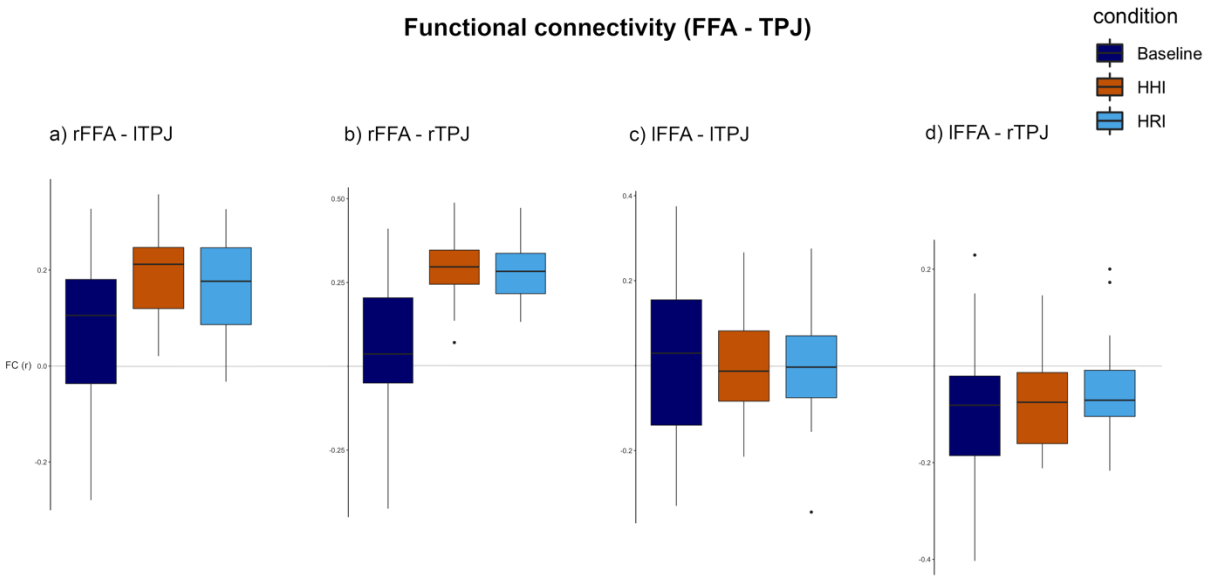

Figure 4. Boxplots (median; 25:75th percentiles, 95% confidence interval) depicting the relation between the contrasts baseline (dark blue), human-human interaction (HHI; orange), human-robot-interaction (HRI; light blue). Each boxplot displays the functional connectivity (FC) between two ROI's: a) left fusiform face area (lFFA) - left temporoparietal junction (lTPJ), b) left fusiform face area (lFFA) - right temporoparietal junction (rTPJ), c) right fusiform face area (rFFA) – left temporoparietal junction (lTPJ), d) right fusiform face area (rFFA) - right temporoparietal junction (rTPJ). The main FC between the rFFA and the bilateral TPJ is significantly different from baseline, while there was no such effect observed for the lFFA. Values can be found in the correlation matrices, depicted in Figure 5.

3.2.1 Changes over time

In contrast to the second hypothesis, no increase in functional connectivity has been demonstrated between the FFA and TPJ for HRI over time ($p > 0.1$). This includes that nor development over time from the three conversations per agent, nor development over the four runs was observed (Figure S3) A similar pattern was observed for HHI. This however does mean that participants did not habituate to the paradigm, verifying the other results. Furthermore, time consistency validates the experiment because different conversations did not influence the overall functional connectivity.

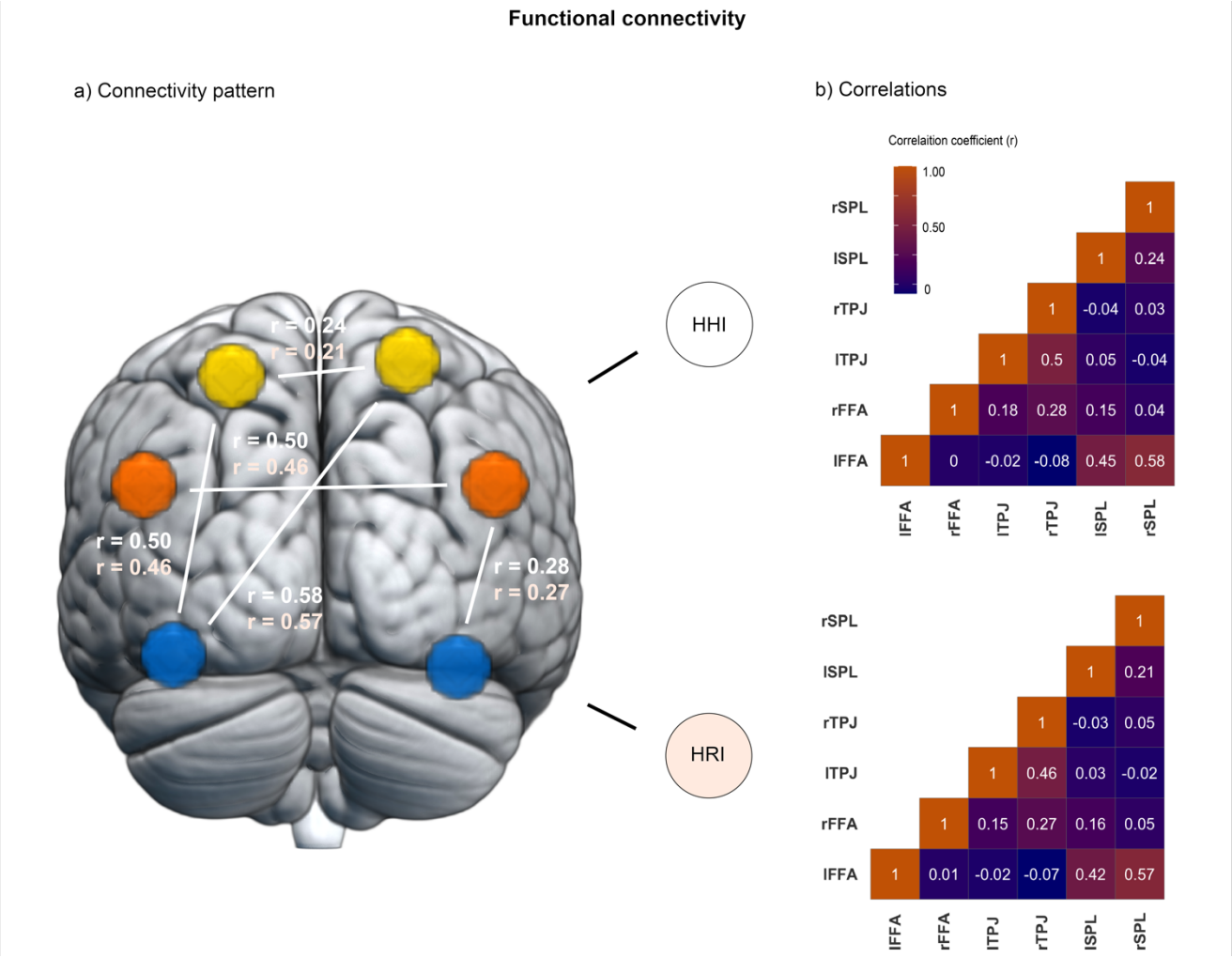

Figure 5. Functional connectivity (FC) expressed as Pearson's correlation coefficient (r) between a) the three main ROI's. Results illustrate a significant correlation (lines) between the average extracted time series per area for the temporoparietal junction (TPJ), fusiform face area (FFA) and the superior parietal lobule (SPL). Correlations in yellow correspond to human-robot interaction (HRI), white to human-human interaction (HHI). b) The main ROI's are depicted in the correlation matrices for HHI and HRI (navy blue; $r = 1.0$, orange; $r = 0$). No significant differences are found between HHI and HRI. For a total overview of the correlation coefficients, including control areas, see supplementary material.

3.3. Object-specific areas and control areas

Lastly, additional results concern the control areas included in this study. Aside of connective variance between the social networks, connections with the object-specific network have altered as well. The connections between the hubs are depicted in Figure 5, all differed from baseline. For HHI and HRI a decrease in functional connectivity is observed ($p < 0.01$) between the bilateral superior parietal lobule (SPL) and the right FFA (HHI; rSPL: $r = 0.05$, lSPL: $r = 0.21$ & HRI; rSPL: $r = 0.07$, lSPL: $r = 0.24$). Even though no hypothesis was included on forehand for the effect between the object-specific and social networks, the result is opposing to prior findings (Henschel et al., 2020), because the SPL is often activated during HRI, in that robots may be perceived as object instead of agents. For all conditions, the left FFA is predominantly connected to the SPL (HHI; lSPL: $r = 0.45$, lSPL: $r = 0.58$ & HRI; rSPL: $r = 0.41$, lSPL: $r = 0.57$).

Furthermore, another control area was included; the middle frontal gyrus (MFG). In this area increased functional connectivity is bilaterally found with the TPJ compared to baseline ($\Delta r = 0.03$-0.18, Figure S4). In addition, a reversed but increased connection (negative correlation to positive correlation) between the lMFG and the rTPJ is found (Baseline; $r = -0.04$, HHI & HRI; $r = 0.14$, Figure S5). The functional connectivity between the lMFG and lTPJ is slightly increased as well (HRI; $\Delta r = 0.09$, HHI; $\Delta r = 0.15$). All other regions of interest demonstrated a decrease in functional connectivity with the MFG. The other control area included in this study was the middle occipital gyrus (MOG). This area depicts decreased functional connectivity to the FFA, SPL and MFG, however is increased connected to the TPJ (Figure S4). To conclude, the control areas show decreased functional connectivity with the bilateral FFA and increased functional connectivity with the TPJ.

4. Discussion

The aim of this study was to connect the activated regions associated with human-robot and human-human interaction, and to compare the functional connectivity differences between these regions for both types of interaction. The addressed question was therefore whether social interaction between humans is distinct from the social interaction with artificial agents. While recent work has solely focussed on the activity of the brain during the interaction with robots, no research has been done

regarding the connectivity between the related brain areas (Redcay & Schilbach, 2019). Subsequently, with the use of the existing database presented by Rauchbauer and colleagues (2019), two hypotheses have been tested. First, the main hypothesis predicted differences in functional connectivity between human-robot (HRI) and human-human interaction (HHI) for the fusiform face area (FFA) and the temporoparietal junction (TPJ). Here, no evidence was found for a difference in connectivity between these core hubs of the person perception network and the theory of mind network for HRI versus HHI. Compared to a static state however interaction does induce significantly increased functional connectivity between the FFA and TPJ, suggesting that these areas may be similarly connected during interaction. The second hypothesis predicted an increase in functional connectivity over time between the FFA and TPJ during HRI. Likewise, no effect was found. This provides evidence for a novel theory concerning a common mechanism for social interaction.

4.1. General interaction network

No evidence is found for a division between HRI and HHI in terms of functional connectivity. As a result, the question rises whether the social neural networks that are known for the interaction with humans are similarly connected during interaction with social artificial agents (Henschel et al., 2020). Within the paradigm most regions of interest are similarly connected, insinuating a possible general mechanism for social interaction. In a previous study from Cross and colleagues analogues results have demonstrated that the function of the FFA not only depicts human face recognition, but more general recognition based on agency as well (2016). In accordance with their study, Gobbini and colleagues revealed similar activity in the FFA during the interaction with a robot (2011). These findings indicate that an interaction mechanism may not altogether be devoted to solely to communication between humans.

Within the current study the most significant increase in functional connectivity is found between the FFA and TPJ. Both regions are part of social networks, indicating that while interacting with other individuals these areas are activated to predict and respond to the behaviour of other agents (Carrington & Bailey, 2009). Greven and colleagues suggest that these hubs of the person perception and the theory-of-mind network are connected during the communication between humans and form a feedback circuit, providing each other information about perceptual and intentional traits (2016). In this study the same functional connectivity pattern for HRI is demonstrated, that is, the

connection between the FFA and TPJ seems to have sustained. Consistent functional connectivity during HRI could therefore imply the involvement of a general mechanism for the interaction with humans and robots, that does not differ for the communication between two persons.

Such a general social interaction network has been found in the primate brain, overlapping some of the areas homologues to the human theory-of-mind and mirror neuron networks (Sliwa & Freiwald, 2017). Succeeding the studies on the primate brain, in a recent study from Thompson and colleagues evidence for the development of a social- vs non-social interaction network has been established (2020). The development of an interactive learning network would imply a distinct mechanism specifically for interaction. To what extent this network would be active for non-humanlike cues has not yet been investigated, notwithstanding that in a theoretical review done by Cerulo it appears that non-human agents, such as robots, are treated as autonomous beings (2011). In consistency with these findings, there was no change of functional connectivity over time for all conversations. This stability over time could mean that interaction with robots did not require or did not reach habituation effects. For HHI no habituation effect has been observed either, insinuating that the conversation partner per se does not induce different connectivity patterns related to social interaction (Schilbach et al., 2006) over time between the FFA and the TPJ. Taken together, the fact that no differences are observed between HHI and HRI indicates the existence of a general interaction network.

4.2. Functional connectivity

Most studies have proved clear differences in brain activity between HHI and HRI (Chaminade et al., 2012; Cross et al., 2019; Henschel et al., 2020; Rosenthal-Von Der Pütten et al., 2014). Likewise, the whole-brain analysis in the current study revealed distinct neural correlates for HHI and HRI either. However, activation differences not necessarily induce the same connectivity differences. A robot may therefore be differently processed in separated brain areas, but the connections between the areas coding for interaction could sustain. For example, Rauchbauer and colleagues focussed on a cluster including the middle fusiform gyrus (MFG) during HRI (2019) and found increased activity. While in the current study a decrease in functional connectivity is found between the SPL, FFA and MOG with the MFG. Although these findings appear opposite, i.e. a decrease versus an increase, differential processing of information within these areas does not necessarily affect the connections between the

areas as well (Rogers et al., 2007). A possible reason for the reduction in functional connectivity between the regions of interest and the MFG is that this region is one of the hubs from the resting-state default-mode network (DMN; Passow et al., 2015). Meaning that while in rest, this network becomes more connected. The reduction with the MFG may thus be ascribed to the deactivation of the DMN, that is, participating in a task (interaction). The diverse patterns during HHI and HRI observed in the whole-brain analysis may reflect specific processing within the activated regions. In the current study, nevertheless, no difference is found between the functional connections. Therefore, even though different activation during HRI is evident, there still may be a wider network devoted for interaction that is not embedded in a particular brain region. Functional connectivity is however never based on solely one connection as all regions function in networks (van den Heuvel & Hulshoff Pol, 2010). Therefore, despite the fact that the functional connectivity patterns are comparable in this study during HRI and HHI, connections between undistinguished regions of interest may still differ.

Another compelling discovery entails reduced functional connectivity between the SPL and FFA during interaction. Importantly, in the current study the baseline condition included the eight seconds within the paradigm in which a participant was concentrating on a picture of humanised fruit from the cover story. It is likely that this task induced object-specific areas such as the SPL and the MOG, as well as the FFA to activate. For example, anthropomorphism stimulates the activation within the FFA and the IPL which is spatially close and functionally similar to the SPL (Henschel et al., 2020; Kühn et al., 2014). However, no research has been done on the connectivity between these regions, so as is discussed previously no definite assumptions could be made from these activation studies. The findings from the current study nonetheless do indicate a relation between the SPL and the FFA when concentrating on the anthropomorphised pictures.

4.3. Limitations and future recommendations

There were however a few limitations to this study which should be considered in future research. First of all**,** an article from the same research group is recently published, dividing conversating and listening in separated conditions (Chaminade, 2020). This distinction has not been included in this study, which could have had several implications. Talking to a robot may not be as different as talking to a person, however listening may actually evoke unique connectivity (Ghosh et al., 2008; Sörös et al., 2006). Combining the events of listening and speaking could reflect that interacting is similar

between humans and robots, however one of the mechanisms behind conversation may be independently different. In future research the conditions of talking and listening should therefore be separated.

Secondly, our paradigm consists of a virtual environment, which could influence the results in a manner that it is more difficult to distinguish between the behavioural aspects of a robot and a person via a screen. In line with this hypothesis, Schilbach and colleagues have demonstrated similar activation patterns in the brain during the interaction with a virtual person (2006). The interaction with a robot per screen could therefore have caused similar processing in the brain in the current study. However, results from the whole-brain analysis did illustrate that HRI and HHI activate different brain areas, suggesting that the conditions were indeed observed differently. Therefore, future research should be conducted in a more realistic setting. In addition, a full connectome study should be done without selecting regions-of-interest on forehand to ascertain whether functional connectivity during HRI still remains similar to HHI.

Finally, there may have been bias from the preprocessing procedure, specifically the last step of denoising the preprocessed images. Although this is one of the most important steps for functional connectivity analysis, it is still an ongoing debate until what extent the regression should be done. For this analysis, global signal regression was included as well, because it is still one of the most common procedures. However, in the recent years there has been controversy about the insertion of the global signal ratio (Xu et al., 2018). A possible downside of regressing out the global signal is that a reintroduction of noise arises, which could give more negative correlations. Interpretations of negative functional connectivity should therefore be made with great caution (Chen et al., 2011), i.e. it could be a result of artefacts. In the current study, negative functional connectivity was found between the lFFA and rTPJ, which can possibly be ascribed to a type-II error, that is, a false rejection of the hypothesis. Likewise, reversed functional connectivity (from negative to positive) has been observed between the TPJ and the control regions. Due to the possible bias of negative functional connectivity artefacts, no assumptions have been made for these changes in functional connectivity. To test whether these signal changes may indeed be a result of interaction, the same analysis should be conducted without regressing out the global signal.

5. Conclusion

In this study, a functional connectivity analysis has been conducted to delineate the extent to which humans communicate in a comparable way with robots as they do with humans. At the moment, engineers in artificial intelligence develop advanced humanlike agents that will be deployed in our social environment, however, previous studies demonstrated that these robots are still perceived in a unique manner not comparable to humans (Henschel et al., 2020; Hortensius & Cross, 2018). The altering of the novel social agents in our current context will therefore enquire adaptation from artificial development, as well as from human behaviour. The aim of the current study was to investigate this behavioural effect on humans and, specifically, whether different perception implies that the interaction with robots is different as well. Surprisingly, no such differences have been found in terms of functional connectivity. This may provide the basis for a general interaction network that is connected regardless of the type of conversational agent. Evidence supports this theory for human-human interaction (Greven et al., 2016), here it is demonstrated that this theory may sustain for human-robot interaction. Within the current paradigm, the fusiform face area (FFA) and the temporoparietal junction (TPJ) show increased functional connectivity during the conversations. This indicates stronger wiring between the person perception network and the theory-of-mind network, because the FFA and TPJ are the core hubs of these social networks. To further explore the theory of a general interaction network, more brain regions should be included for a complete overview. A full connectome study is therefore suggested to indicate whether the communication with robots induces similar network wiring to human interaction.

6. References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in neuroinformatics, 8*, 14.

Adolphs, R. (2009). The Social Brain: Neural Basis of Social Knowledge. *Annual Review of Psychology*, *60*(1), 693–716. https://doi.org/10.1146/annurev.psych.60.110707.163514

Al Moubayed, S., Beskow, J., Skantze, G., & Granström, B. (2012). Furhat: A back-projected human-

like robot head for multiparty human-machine interaction. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7403 LNCS*, 114–130. https://doi.org/10.1007/978-3-642-34584-5_9

Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). Presented at the 8th International Conferance on Functional Mapping of the Human Brain. *Japan. Available on CD-ROM in NeuroImage*, *16*. http://www.mrc-cbu.cam.ac.uk/Imaging/marsbar.html

Broadbent, E. (2017). Interactions With Robots: The Truths We Reveal About Ourselves. *Annual Review of Psychology*, *68*(1), 627–652. https://doi.org/10.1146/annurev-psych-010416-043958

Carrington, S. J., & Bailey, A. J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping*, *30*(8), 2313–2335. https://doi.org/10.1002/hbm.20671

Cerulo, K. A. (2011). Social Interaction: Do Non-humans Count? *Sociology Compass*, *5*(9), 775–791. https://doi.org/10.1111/j.1751-9020.2011.00404.x

Chaminade, T., Hodgins, J., & Kawato, M. (2007). Anthropomorphism influences perception of computer-animated characters' actions. *Social Cognitive and Affective Neuroscience*, *2*(3), 206–216. https://doi.org/10.1093/scan/nsm017

Chaminade, T., Rauchbauer, B., Nazarian, B., Bourhis, M., Ochs, M., & Prévot, L. (2018). Investigating the dimensions of conversational agents' social competence using objective neurophysiological measurements. *Proceedings of the 20th International Conference on Multimodal Interaction, ICMI 2018*. https://doi.org/10.1145/3281151.3281162

Chaminade, T., Rosset, D., Da Fonseca, D., Nazarian, B., Lutcher, E., Cheng, G., & Deruelle, C. (2012). How do we think machines think? An fMRI study of alleged competition with an artificial intelligence. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2012.00103

Chen, G., Chen, G., Xie, C., & Li, S. J. (2011). Negative Functional Connectivity and Its Dependence on the Shortest Path Length of Positive Network in the Resting-State Human Brain. *Brain Connectivity*, *1*(3), 195–206. https://doi.org/10.1089/brain.2011.0025

Cross, E. S., Hortensius, R., & Wykowska, A. (2019). From social brains to social robots: Applying neurocognitive insights to human-robot interaction. In *Philosophical Transactions of the Royal Society B: Biological Sciences, 374*(1771). https://doi.org/10.1098/rstb.2018.0024

Cross, E. S., Ramsey, R., Liepelt, R., Prinz, W., & de Hamilton, A. F. 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). https://doi.org/10.1098/rstb.2015.0075

Dautenhahn, K. (2007). Socially intelligent robots: Dimensions of human-robot interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 679–704. https://doi.org/10.1098/rstb.2006.2004

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

Ghosh, S. S., Tourville, J. A., & Guenther, F. H. (2008). A Neuroimaging Study of Premotor Lateralization and Cerebellar Involvement in the Production of Phonemes and Syllables. *Journal of Speech, Language, and Hearing Research*, *51*(5), 1183–1202. https://doi.org/10.1044/1092-4388(2008/07-0119)

Gobbini, M. I., Gentili, C., Ricciardi, E., Bellucci, C., Salvini, P., Laschi, C., Guazzelli, M., & Pietrini, P. (2011). Distinct neural systems involved in agency and animacy detection. *Journal of Cognitive Neuroscience*, *23*(8), 1911–1920. https://doi.org/10.1162/jocn.2010.21574

Greven, I. M., Downing, P. E., & Ramsey, R. (2016). Linking person perception and person knowledge in the human brain. *Social Cognitive and Affective Neuroscience*, *11*(4), 641–651. https://doi.org/10.1093/scan/nsv148

Greven, I. M., & Ramsey, R. (2017). Person perception involves functional integration between the extrastriate body area and temporal pole. *Neuropsychologia*, *96*, 52–60. https://doi.org/10.1016/j.neuropsychologia.2017.01.003

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. In *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

Koster-Hale, J., & Saxe, R. (2013). Theory of Mind: A Neural Prediction Problem. *Neuron, 79*(5), 836–848. https://doi.org/10.1016/j.neuron.2013.08.020

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS ONE*, *3*(7). https://doi.org/10.1371/journal.pone.0002597

Kühn, S., Brick, T. R., Müller, B. C. N., & Gallinat, J. (2014). Is This Car Looking at You? How Anthropomorphism Predicts Fusiform Face Area Activation when Seeing Cars. *PLoS ONE*, *9*(12), e113885. https://doi.org/10.1371/journal.pone.0113885

Mitchell, J. P. (2008). Activity in Right Temporo-Parietal Junction is Not Selective for Theory-of-Mind. *Cerebral Cortex*, *18*(2), 262–271. https://doi.org/10.1093/cercor/bhm051

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley. *IEEE Robotics and Automation Magazine*, *19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Passow, S., Specht, K., Adamsen, T. C., Biermann, M., Brekke, N., Craven, A. R., Ersland, L., Grüner, R., Kleven-Madsen, N., Kvernenes, O., Schwarzlmüller, T., Olesen, R. A., & Hugdahl, K. (2015). Default-mode network functional connectivity is closely related to metabolic activity. *Human Brain Mapping*, *36*(6), 2027–2038. https://doi.org/10.1002/hbm.22753

Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews Neuroscience, 8*(8), 636–648. https://doi.org/10.1038/nrn2195

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

Rauchbauer, B., Nazarian, B., Bourhis, M., Ochs, M., Prévot, L., & Chaminade, T. (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). 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*(8), 495–505. https://doi.org/10.1038/s41583-019-0179-4

Rogers, B. P., Morgan, V. L., Newton, A. T., & Gore, J. C. (2007). Assessing functional connectivity in the human brain by fMRI. *Magnetic Resonance Imaging*, *25*(10), 1347–1357. https://doi.org/10.1016/j.mri.2007.03.007

Rosenthal-Von Der Pütten, A. M., Schulte, F. P., Eimler, S. C., Sobieraj, S., Hoffmann, L., Maderwald, S., Brand, M., & Krämer, N. C. (2014). Investigations on empathy towards humans and robots using fMRI. *Computers in Human Behavior*, *33*, 201–212. https://doi.org/10.1016/j.chb.2014.01.004

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind." *NeuroImage*, *19*(4), 1835–1842. https://doi.org/10.1016/S1053-8119(03)00230-1

Saxe, Rebecca, & Baron-Cohen, S. (2006). The neuroscience of theory of mind. *Social neuroscience, 1*(3–4). https://doi.org/10.1080/17470910601117463

Schilbach, L. (2015). The Neural Correlates of Social Cognition and Social Interaction. *Brain Mapping: An Encyclopedic Reference, 3*, 159–164. https://doi.org/10.1016/B978-0-12-397025-1.00172-X

Schilbach, Leonhard, Wohlschlaeger, A. M., Kraemer, N. C., Newen, A., Shah, N. J., Fink, G. R., & Vogeley, K. (2006). Being with virtual others: Neural correlates of social interaction. *Neuropsychologia*, *44*(5), 718–730. https://doi.org/10.1016/j.neuropsychologia.2005.07.017

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. In *Neuroscience and Biobehavioral Reviews*, *42*, 9–34. Elsevier Ltd. https://doi.org/10.1016/j.neubiorev.2014.01.009

Shenhav, A., Cohen, J. D., & Botvinick, M. M. (2016). *Dorsal anterior cingulate cortex and the value of control*. *19*(10), 4–6. https://doi.org/10.1038/nn.4382

Sliwa, J., & Freiwald, W. A. (2017). Neuroscience: A dedicated network for social interaction processing in the primate brain. *Science*, *356*(6339), 745–749. https://doi.org/10.1126/science.aam6383

Sörös, P., Sokoloff, L. G., Bose, A., McIntosh, A. R., Graham, S. J., & Stuss, D. T. (2006). Clustered functional MRI of overt speech production. *NeuroImage*, *32*(1), 376–387. https://doi.org/10.1016/j.neuroimage.2006.02.046

Thierry, C. (n.d.). *A multimodal corpus of Human-Human and Human-Robot conversations including synchronized behavioral and neurophysiological recordings*. Retrieved February 12, 2021, from

https://hal.archives-ouvertes.fr/hal-02916070

van den Heuvel, M. P., & Hulshoff Pol, H. E. (2010). Exploring the brain network: A review on resting-state fMRI functional connectivity. In *European Neuropsychopharmacology 20*(8), 519–534. Elsevier. https://doi.org/10.1016/j.euroneuro.2010.03.008

Wang, Y., & Quadflieg, S. (2014). In our own image? Emotional and neural processing differences when observing human-human vs human-robot interactions. *Social Cognitive and Affective Neuroscience*, *10*(11), 1515–1524. https://doi.org/10.1093/scan/nsv043

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

Wykowska, A., Chaminade, T., & Cheng, G. (2016). Embodied artificial agents for understanding human social cognition. In *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(693). https://doi.org/10.1098/rstb.2015.0375

Xu, H., Su, J., Qin, J., Li, M., Zeng, L. L., Hu, D., & Shen, H. (2018). Impact of global signal regression on characterizing dynamic functional connectivity and brain states. *NeuroImage*, *173*, 127–145. https://doi.org/10.1016/j.neuroimage.2018.02.036

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.

7. References supplemental material

Camilleri, J. A., Müller, V. I., Fox, P., Laird, A. R., Hoffstaedter, F., Kalenscher, T., & Eickhoff, S. B. (2018). Definition and characterization of an extended multiple-demand network. *NeuroImage, 165,* 138–147. https://doi.org/10.1016/j.neuroimage.2017.10.020

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

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

Acknowledgements

I want to thank my supervisor dr. Ruud Hortensius for the excellent guidance during the past months and all the valuable comments he proposed.

Supplementary material

Supplementary material, containing control analysis, graphs and other relevant information are located in the attached document. For the scripts of the analysis please check out GitLab: https://gitlab.com/brasc/fchri

Human-robot interaction, supplementary material

The difference in functional connectivity during human-human interaction and human-robot interaction.

Universiteit Utrecht

Author: Ann L.M.P. Hogenhuis[1]

Supervisor: dr. Ruud Hortensius[2]

Second reader: Colin Caret[3]

[1)] *BSc student (nr. 6019293), Liberal Arts and Sciences; Cognitive and Neurobiological Psychology at Utrecht University*

[2)] *Department of Psychology, Utrecht University, Heidelberglaan 1, 3584 CS, Utrecht, The Netherlands*

[3)] *Department of Philosophy and Religion studies, Utrecht University, Janskerkhof 13, Utrecht, 3512 BL, The Netherlands*

25-02-2021, Utrecht University

Thesis Artificial Intelligence, Liberal Arts and Sciences, 15 ECTS

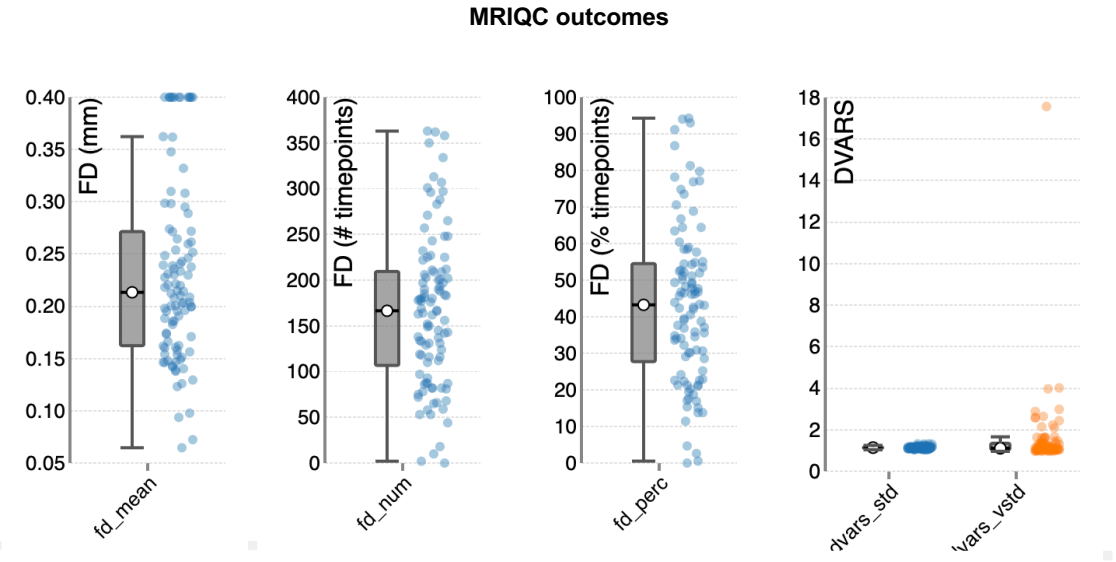

Fig. S1 These plots contain the output from the quality check performed with MRIQC. The included plots illustrate a) framewise displacement, which is an index of head movement. The three plots contain the head movement per volume in millimeters (mm; median = 0.21, Q1 = 0.16, Q3 = 0.27), timepoints (median = 166.50, Q1 = 106.75, Q3 = 209.50) and percentage timepoints (median = 43.25, Q1 = 27.73, Q3 = 54.52). Movement varied from 0.064 – 0.36 mm within the confidence interval (95%). One of the outliers includes participant 19 which has been removed from the results due to overall excessive movement. The remaining outliers have been included in the analysis. b) DVARS, which is the standard deviation of the root mean square from the temporal signal change in 1) std, whole brain signal (median = 1.14, Q1 = 1.10, Q3 = 1.16), 2) vstd, voxel wise signal change (median = 1.11, Q1 = 1.04, Q3 = 1.35). The outlier shown at DVARS_vstd 17.57 has been removed from the analysis. This participant (number 10) encountered technical problems

Table S2 Regions of interest locations for left and right hemispheres

Region of interest			MNI coordinates		
		Hemisphere	**x**	**y**	**z**
Temporoparietal junction (Julian et al., 2012)	(TPJ)	L	-48	-62	30
		R	48	-60	30
Fusiform face area (Julian et al., 2012)	(FFA)	L	-40	-52	-18
		R	38	-42	-22
Middle occipital gyrus (Henschel et al., 2020)	(MOG)	L	-30	-84	12
		R	36	-84	12
Superior parietal lobule (Julien et al., 2012)	(SPL)	L	-24	-56	60
		R	24	-52	64
Middle frontal gyrus (Camilleri et al., 2018)	(MFG)	L	-44	32	22
		R	44	36	20

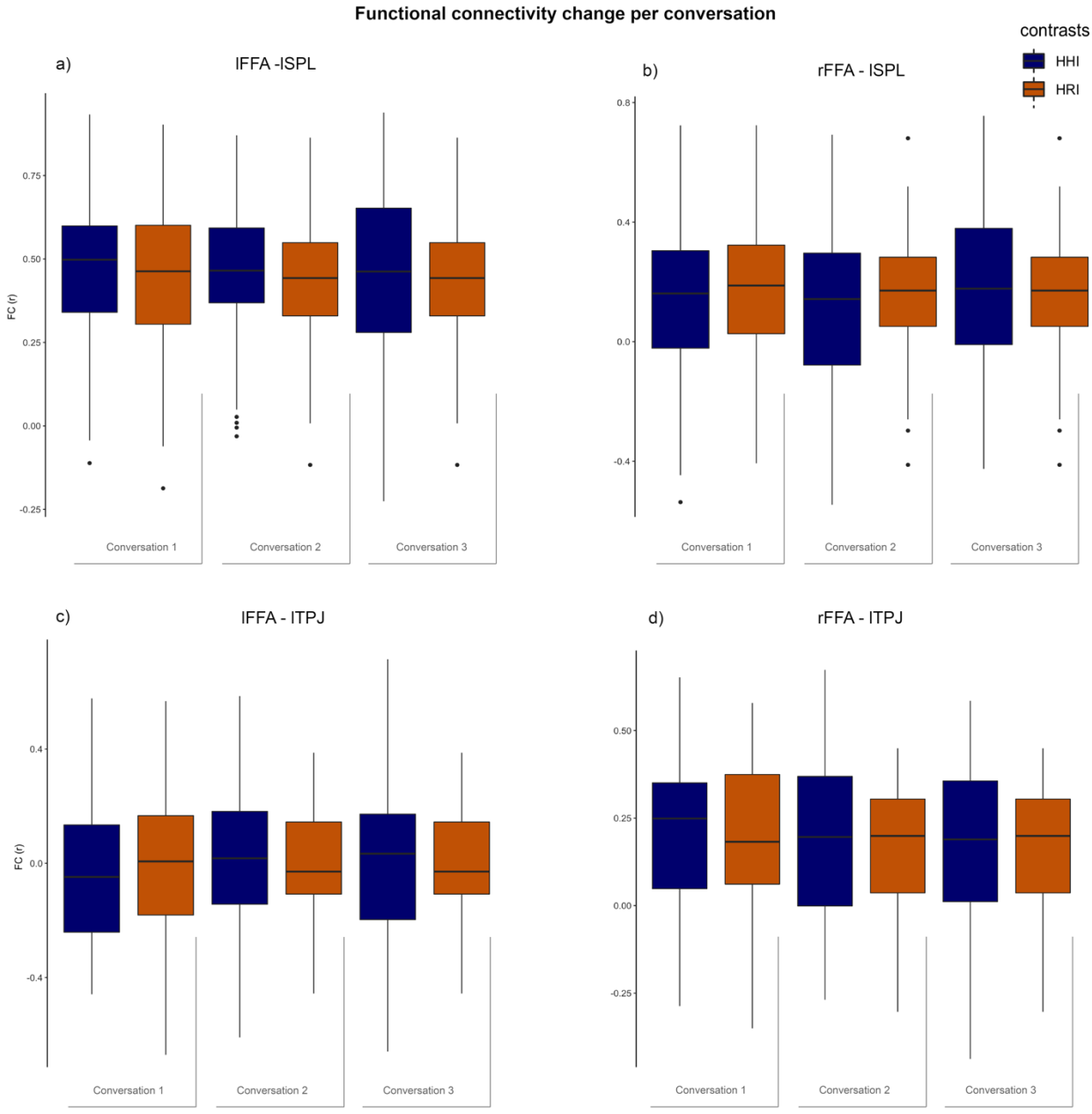

Fig. S3 Examples of stable temporal functional connectivity (y-axis) per conversation (x-axis) between the regions of interest (ROI's). In this figure connections between the main regions of interest are depicted, including a) left fusiform face area (lFFA) – left superior temporal lobule (lSPL), b) right fusiform face area (rFFA) – left superior parietal lobule (lSPL), c) left fusiform face area (lFFA) – left temporoparietal junction (lTPJ), d) right fusiform face area (rFFA) – left temporoparietal junction (lTPJ). No significant changes have been found between the mean of the separated conversations for human-human and human-robot interaction. This applies for every included ROI in this study.

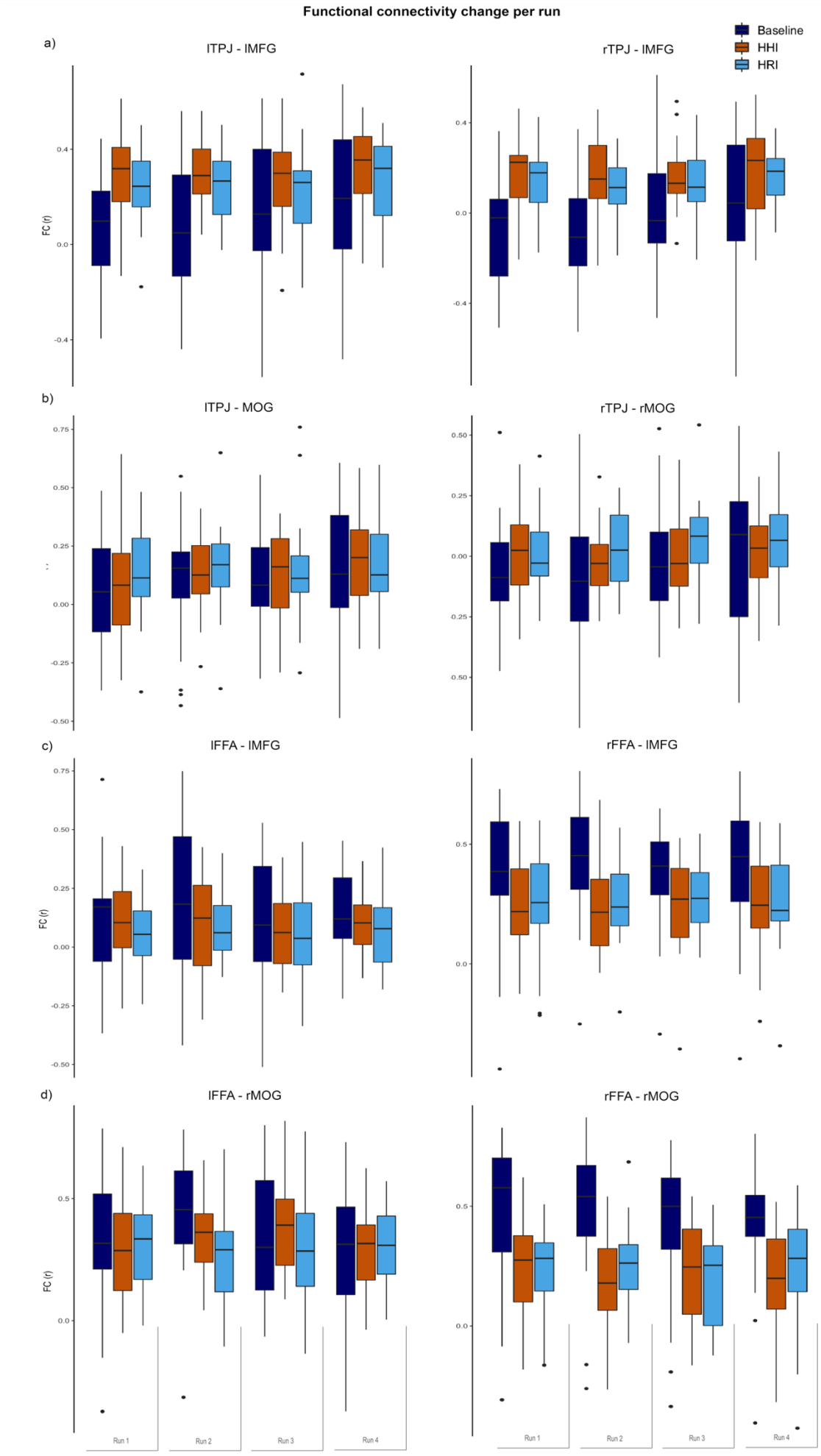
Functional connectivity change per run
Baseline
HHI
HRI
a)
lTPJ - lMFG
rTPJ - lMFG
b)
lTPJ - MOG
rTPJ - rMOG
c)
lFFA - lMFG
rFFA - lMFG
d)
lFFA - rMOG
rFFA - rMOG
FC (r)
Run 1
Run 2
Run 3
Run 4

Fig. S4 In this figure the boxplots illustrate functional connectivity between the control areas and main areas included in the analysis. The left panel demonstrates the left hemisphere, vice versa for the right panel. From upper to lower rows, the connections are: a) ITPJ – IMFG & rTPJ - IMFG b) ITPJ – IMOG & rTPJ - IMOG c) rFFA - rMFG & lFFA - rMFG d) rFFA - rMOG & lFFA – rMOG. Hemispheres for the control areas are picked at random, however the temporal results are bilaterally the same for the control areas. No effect of time was found over multiple runs, suggesting no habituation effect during the interaction with a robot nor with a human.

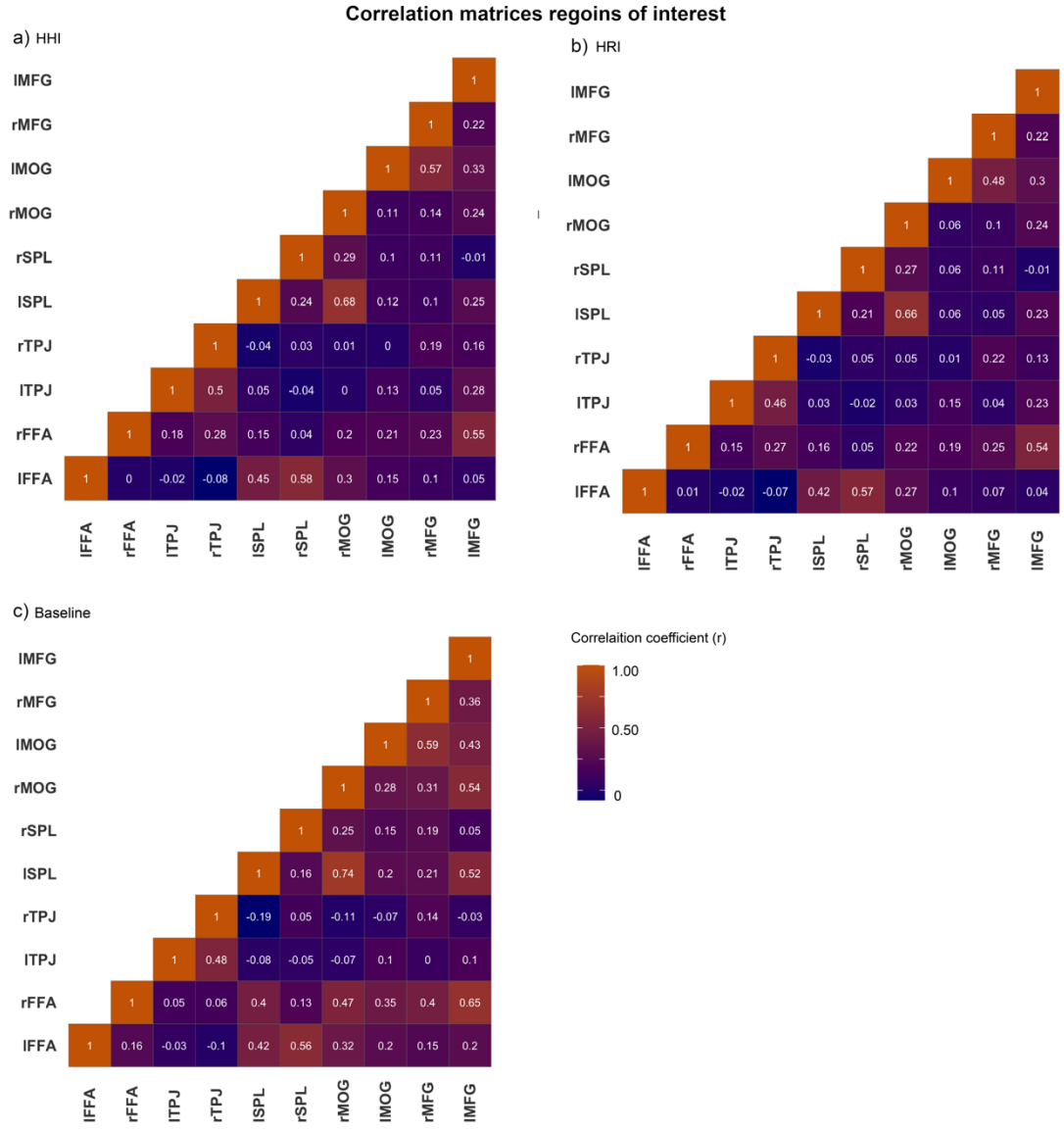

Fig. S5 The depicted plots are identical to Figure 5, however contain control areas and the baseline condition as well (navy blue; r = 1.0, orange; r = 0). Three heatmaps are included, each coding for one contrast (a. HHI, b. HRI, c. Baseline). From left to right the regions of interest are the fusiform face area (FFA), temporoparietal junction (TPJ), superior parietal lobule (SPL), middle occipital gyrus (MOG). Middle frontal gyrus (MFG), all regions are illustrated for left and right hemispheres. No significant differences are found between HHI and HRI.