NeuroImage 235 (2021) 118049

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Representation of social content in dorsomedial prefrontal cortex underlies individual differences in agreeableness trait

Sandra Arbula [a,*], Elisabetta Pisanu [a], Raffaella I. Rumiati [a,b]

[a] Neuroscience Area, International School for Advanced Studies (SISSA), via Bonomea 265, Trieste 34136, Italy
[b] Scuola superiore di studi avanzati Sapienza (SSAS), Rome, Italy

ARTICLE INFO

Keywords:
Personality
Agreeableness
Social cognition
Dorsomedial prefrontal cortex
Representational similarity analysis

ABSTRACT

Personality traits reflect key aspects of individual variability in different psychological domains. Understanding the mechanisms that give rise to these differences requires an exhaustive investigation of the behaviors associated with such traits, and their underlying neural sources. Here we investigated the mechanisms underlying agreeableness, one of the five major dimensions of personality, which has been linked mainly to socio-cognitive functions. In particular, we examined whether individual differences in the neural representations of social information are related to differences in agreeableness of individuals. To this end, we adopted a multivariate representational similarity approach that captured within single individuals the activation pattern similarity of social and non-social content, and tested its relation to the agreeableness trait in a hypothesis-driven manner. The main result confirmed our prediction: processing social and non-social content led to similar patterns of activation in individuals with low agreeableness, while in more agreeable individuals these patterns were more dissimilar. Critically, this association between agreeableness and encoding similarity of social and random content was significant only in the dorsomedial prefrontal cortex, a brain region consistently involved during attributions of mental states. The present finding reveals the link between neural mechanisms underlying social information processing and agreeableness, a personality trait highly related to socio-cognitive abilities, thereby providing a step forward in characterizing its neural determinants. Furthermore, it emphasizes the advantage of multivariate pattern analysis approaches in capturing and understanding the neural sources of individual variations.

1. Introduction

Every human being is unique, and one important part of this uniqueness is determined by personality. Understanding the neural sources of inter-individual variability in personality is the major goal of personality neuroscience (Yarkoni, 2015). Differences across personality traits are reflected in different motivational, emotional, cognitive and behavioral responses to particular stimuli. According to many theories, these responses are relatively stable, but manifest only in certain contexts (Corr et al., 2013; DeYoung, 2010; Gray, 1982; Tellegen and Waller, 1981). Yet, numerous studies identifying the neural correlates of different personality traits have focused on spontaneous, resting state brain activity (e.g., Cai et al., 2020; Dubois et al., 2018; Kuper et al., 2019; Mulders et al., 2018; Nostro et al., 2018), or structural brain features (e.g., Avinun et al., 2020; Lewis et al., 2018; Omura et al., 2005; Owens et al., 2019; Riccelli et al., 2017; Taki et al., 2013), in which indices of personality traits were rarely clearly evident, and often diverged between studies. Another important limitation is that these types of associations are relatively uninformative about the mechanisms that contribute to distinct trait-specific behaviors.

An alternative approach holds that differences in personality affect specific cognitive mechanisms and, in turn, modulate the neural correlates observed during task-based neuroimaging. Such an association has been hypothesized between the trait agreeableness - one of the five broad dimensions of personality within the five-factor model (Costa and McCrae, 1992; John et al., 2008) - and mentalization processes (Allen et al., 2017; Nettle and Liddle, 2008). Agreeable individuals have more empathic, altruistic and cooperative tendencies, which require the ability to understand others' mental states, emotions and intentions (Allen and DeYoung, 2016). This hypothesis has been tested both indirectly, by associating agreeableness with prosocial behavior and empathy (Graziano et al., 2007; Habashi et al., 2016; Penner et al., 1995), and directly, by showing its impact on the performance in a mentalizing task (Nettle and Liddle, 2008). More recently, two studies have taken a step further by investigating the neural correlates of ability and testing whether those correlates are related to agreeableness (Allen et al., 2017; Udochi et al., 2020). Both studies considered the default network as the neural substrate of mentalization, and found

* Corresponding author.
E-mail address: saarbul@sissa.it (S. Arbula).

https://doi.org/10.1016/j.neuroimage.2021.118049
Received 18 December 2020; Received in revised form 1 April 2021; Accepted 2 April 2021
Available online 10 April 2021

it to be modulated by individual variations in mentalizing ability, but only Udochi et al. (2020) found that it also predicted individual differences in agreeableness. One key distinction between the two studies is that Allen et al. (2017) defined the default network from resting state functional connectivity, while Udochi et al. (2020) adopted a task-based approach. In particular, in the latter work the authors found that the default network activity during social animation processing predicted both social cognitive ability and agreeableness. While this finding provides new insights on the common neural substrates of agreeableness and mentalization, it still remains unclear how these neural variations underlie behavioral differences. Do they reflect differences in processing social information among individuals with different degrees of agreeableness?

To test this hypothesis we used representational similarity analysis (RSA; Kriegeskorte et al., 2008) that allowed us to examine whether individual differences in the representation of social content are related to differences in agreeableness of individuals. In particular, we compared the patterns of neural activity during social and random (i.e., non-social) animations viewing within each participant in order to measure on an individual basis how similarly the two types of contents are encoded. We predicted that the levels of agreeableness would correlate positively with the dissimilarity in neural representations of social and random content, and that this association will be mainly evidenced in brain regions involved in mentalization processes (i.e., mentalizing network) (Molenberghs et al., 2016; Schurz et al., 2014). Moreover, we hypothesized a possible correlation also with task accuracy, showing greater social-random pattern dissimilarity for higher accuracies. On the other hand, we did not expect any other trait to be related to this pattern dissimilarity.

With respect to univariate approaches that relate individual differences to differences in activation of single voxels, or activation averaged across voxels in restricted brain regions, the representational similarity approach is more sensitive to individual variations since it takes into account distributed patterns of activity, which hold much more information that can be compared within and between individuals (Etzel et al., 2020). On the other hand, it suffers certain limitations in inferring what type of information is encoded in the representational space (Naselaris and Kay, 2015). One way to overcome this issue is to predict neural dissimilarities between experimental conditions according to a psychological model that makes clear inferences about the observed behavior (Carlson et al., 2018; Ritchie et al., 2019). Here we adopted this strategy to investigate the neural sources of agreeableness trait, by basing our predictions of social content encoding on well-established behavioral facets of agreeableness and its link with mentalization ability.

2. Methods

2.1. Participants recruitment and personality assessment

Participants' recruitment involved a prescreening personality assessment that was administered through an online questionnaire, advertised through different recruitment services (Facebook, Sona). Five personality traits (extraversion, agreeableness, conscientiousness, neuroticism and openness to experience) were assessed using the Italian adaptation of the Big Five Inventory (Ubbiali et al., 2013), which comprises 44 items rated on a five-point Likert scale. Fifty-five participants were selected from a larger sample ($N = 143$) based on their personality scores in order to cover uniformly all five traits derived from the five-factor model (Costa and McCrae, 1992; John et al., 2008). In particular, within each trait the score distribution was normal (Kolmogorov-Smirnov *ps*' > 0.15) and at least 5 subjects had scores above ± 1 SD from the Italian population mean (Ubbiali et al., 2013). Six participants were excluded due to excessive motion during scanning, yielding a final sample of 49 participants (19 males) with mean age 23.14 (SD = 4.24, range = 18–34). None of them reported any history of neurologic or psychiatric disorders. All participants gave written informed consent prior to their participation in the study and received 30 euros for their participation. The study was approved by the Regional Ethics Committee of Friuli Venezia Giulia and was conducted according to the guidelines of the Declaration of Helsinki. Behavioral and demographic information are reported in the participants.tsv file on OpenNeuro.

2.2. Social cognition task

Participants were presented with short animations representing different shapes that moved randomly or interacted in a socially meaningful way. The animations were originally developed by Castelli et al. (2000) and Wheatley et al. (2007) and shortened to 20 s for the battery of fMRI tasks used in the Human Connectome Project (Barch et al., 2013). During each of the 4 runs 5 different animations were presented interleaved with 15 s long fixation blocks. After each animation, a 3 s long instruction screen appeared prompting the participants to respond by pressing one of the three keys on a response pad positioned under their right hand. They were told to press their ring finger if the shapes were moving randomly, their index finger if the shapes interacted in a socially meaningful way (as if they were taking into consideration each other feelings and thoughts), and the middle finger if they were not sure about the type of interaction. Each run consisted of 2 or 3 videos of each condition. One social and one random video were presented prior to scanning for practice and were not reused during the testing phase.

2.3. MRI acquisition

MRI data were collected on a 3 Tesla whole-body scanner (Achieva Philips) equipped with an 8-channel head coil at the "S. Maria della Misericordia Hospital" in Udine. For each of the four runs of the social cognition task, 101 functional image volumes with 37 contiguous axial slices were collected with a T2*-weighted echo-planar sequence (TR: 2 s, TE: 30 ms, FA: 82°, voxel size: $3 \times 3 \times 3$ mm, acquisition matrix: 80×80). A high-resolution T1-weighted anatomical image was acquired at the beginning of the session (170 sagittal slices, TR/TE: 8.1/3.7 ms, FA: 12°; voxel size: $1 \times 1 \times 1$ mm, acquisition matrix: 240×240). Additionally, to correct for spatial distortion of functional images, a pair of spin echo images with opposite phase encoding directions and same orientation as the functional scans, were acquired at the beginning and in the middle of the whole scanning session. Stimuli were presented using E-Prime 2 (Psychology Software Tools, Inc; Schneider et al., 2012) and delivered through MRI-compatible goggles mounted on the head coil. During the scanning session, participants performed three other tasks in a counterbalanced order, which are not reported in the present study.

2.4. MRI quality assessment

All MRI data were converted from DICOM format into the Brain Imaging Data Structure (BIDS; https://bids.neuroimaging.io/) using the Dcm2Bids tool (https://github.com/cbedetti/Dcm2Bids). Quality of structural and functional data was assessed using the MRI Quality Control tool (MRIQC) (Esteban et al., 2017) and compared to a set of quality metrics from the MRIQC web API (Esteban et al., 2019) using the MRIQCeption tool (https://github.com/elizabethbeard/mriqception). Out of 55 scanned participants, data from 6 were classified as outliers based on the following quality metrics: AFNI's outlier ratio and quality index, intensity changes (DVARS) and frame-wise displacement (FD). A full report of image quality metrics is available on OpenNeuro.

2.5. MRI preprocessing

Results included in this manuscript come from preprocessing performed using fMRIPrep version 1.5.1rc2 (Esteban et al., 2019; RRID:SCR_016216), a Nipype (Gorgolewski et al., 2011, 2017; RRID:SCR_002502) based tool.

Each T1w (T1-weighted) volume was corrected for INU (intensity non-uniformity) using N4BiasFieldCorrection v2.2.0 (Avants et al., 2008; RRID:SCR_004757) and skull-stripped using antsBrainExtraction.sh v2.2.0 (using the OASIS template). Brain surfaces were reconstructed using recon-all from FreeSurfer v6.0.1 (Dale et al., 1999; RRID:SCR_001847), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (Klein et al., 2017; RRID:SCR_002438). Spatial normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c (Fonov et al., 2009; RRID:SCR_008796) was performed through nonlinear registration with the antsRegistration tool of ANTs v2.2.0 (Avants et al., 2008; RRID:SCR_004757), using brain-extracted versions of both T1w volume and template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL v5.0.9; Zhang et al., 2001; RRID:SCR_002823).

Functional data was slice time corrected using 3dTshift from AFNI v16.2.07 (Cox, 1996; RRID:SCR_005927) and motion corrected using mcflirt (FSL v5.0.9; Jenkinson et al., 2002). Distortion correction was performed using an implementation of the TOPUP technique (Andersson et al., 2003) using 3dQwarp (AFNI v16.2.07; Cox, 1996). This was followed by co-registration to the corresponding T1w using boundary-based registration (Greve and Fischl, 2009) with six degrees of freedom, using bbregister (FreeSurfer v6.0.1). Motion correcting transformations, field distortion correcting warp, BOLD-to-T1w transformation and T1w-to-template (MNI) warp were concatenated and applied in a single step using antsApplyTransforms (ANTs v2.2.0) using Lanczos interpolation.

Physiological noise regressors were extracted applying CompCor (Behzadi et al., 2007). Principal components were estimated for the anatomical CompCor variants (aCompCor). A mask to exclude signal with cortical origin was obtained by eroding the brain mask, ensuring it only contained subcortical structures. Six aCompCor components were then calculated within the intersection of the subcortical mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run. Frame-wise displacement (FD) and DVARS (Power et al., 2014) are calculated for each functional run using their implementations in Nipype.

Functional data were masked using the brain mask obtained from fMRIPrep and 14 fMRIPrep derived confounds (six motion parameters, FD, standardized DVARS and six aCompCor) were removed on a voxel-wise level using the Denoiser tool (https://github.com/arielletambini/denoiser). As a final step, functional data were spatially smoothed using a Gaussian kernel of 6 mm full-width at half-maximum.

2.6. Behavioral data analyses

We collected accuracy and reaction times (RTs) data. RTs were filtered for errors and outliers above 3 standard deviations from the subjects mean for each video type condition. One subject was excluded from the behavioral analysis because we did not collect all of his responses due to a response pad issue. Anticipated responses (RTs < 200 ms) were absent. RTs were assessed with repeated measures ANCOVA with type of video as within-subjects factor and five trait scores as continuous predictors. Since accuracy data were not normally distributed (Shapiro-Wilk test $p < .05$), a paired Wilcoxon signed rank test was used to assess accuracy differences between the two types of videos, and Spearman rank correlation was used to assess the accuracy correlations with the five trait measures, separately for each type of video.

2.7. fMRI data analyses

2.7.1. First level GLM analysis

First level GLM analysis was performed using FSL FEAT (www.fmrib.ox.ac.uk/fsl). A GLM model was built for each participant and each run with the two experimental conditions (social and random) as regressors of interest and their temporal derivatives as regressors of no interest. Erroneous trials and those with "not sure" responses were not excluded from the model. The regressors were time locked to the onset and duration of the video and convolved with a double-gamma hemodynamic response function. FILM pre-whitening was used to correct for autocorrelation and low-frequency drifts were removed using a high-pass filter with a 100 s cutoff. All following analyses were performed on subject-level social and random beta maps, averaged over the four runs (fixed effects).

2.7.2. Group level GLM analysis

To localize mean group effects, whole-brain group level analysis was performed with mixed effect (FLAME 1), as implemented in FSL, for the social - random contrast. Statistical map was assessed with a cluster-based threshold of $Z > 3.1$ and corrected at $p = 0.05$ (family-wise error correction).

2.7.3. Representational similarity analysis

We conducted a within subject representational similarity analysis (RSA) to test if the participants' level of agreeableness will be related to the dissimilarity between the neural patterns of activity during social and random content processing. As anticipated in the introduction, we expected to observe a positive correlation between the two measures in brain regions belonging to the mentalizing network (Molenberghs et al., 2016; Schurz et al., 2014), indicating that subjects low in agreeableness will have less distinct neural representations of social and random content, and vice versa. However, since agreeableness has been associated also with brain regions outside the mentalizing network (e.g., Cai et al., 2020; Liu et al., 2019; Riccelli et al., 2017), we opted for a whole-brain parcellation scheme in our analyses, since it offers a nice middle ground between the restrictive ROI-based approach and the heavily corrected and computationally expensive searchlight approach. As a first step, all beta maps were divided into 200 non-overlapping regions using a whole-brain parcellation derived from the meta-analytic functional coactivation of the Neurosynth database (https://identifiers.org/neurovault.image:39711). Next, representational dissimilarities were computed for each participant by correlating social and random response patterns from each parcel and subtracting it from 1 (i.e., higher correlation is reflected as lower dissimilarity). The participants' dissimilarities within each parcel were then correlated with their agreeableness scores using Spearman's rank-order correlations. The statistical significance (p-value) of the resulting correlations was obtained by computing the same correlations after permuting 30,000 times the agreeableness scores and calculating the proportion of permuted correlations that exceeded the ones yielded from non-permuted data. The same representational dissimilarity and behavioral score correlation was repeated for the remaining four trait dimensions. All resulting permuted p-values were corrected for multiple comparisons by dividing the alpha (0.05) by the number of behavioral measures tested (five traits and one accuracy score) and the number of parcels (200).

2.7.4. Univariate analysis

Alternatively, we wanted to assess whether the hypothesized neural dissimilarity associated with agreeableness could be explained by univariate activation differences between social and non-social content. Therefore we calculated the difference between the average levels of activity (averaged betas across all voxels within each parcel) for each condition (social – random contrast), and correlated it with the agreeableness score. All statistical procedures and subsequent multiple comparison corrections were same as in the RSA.

2.7.5. Meta-analytic decoding

The correlational maps from the RSA and the *Z* maps from the GLM analysis were correlated with the meta-analytic activation maps of the

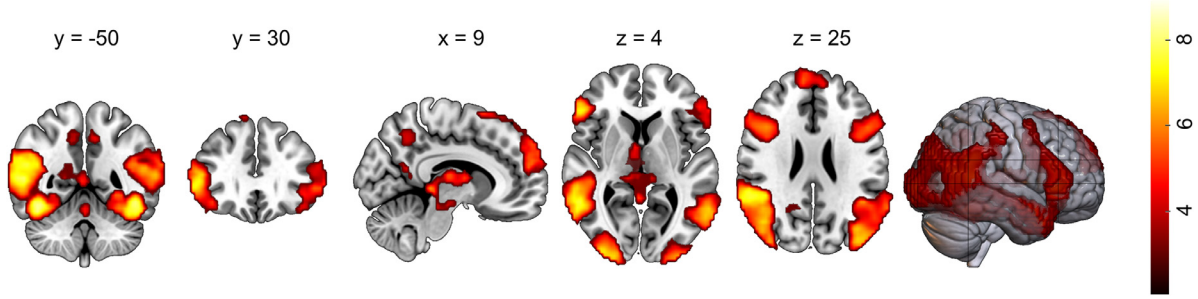

Fig. 1. Social-random contrast activations. Brain activations from the whole brain voxelwise social – random content contrast. The statistical map was assessed with a cluster-based threshold of $Z > 3.1$, corrected at $p = 0.05$ (family-wise error correction) and projected onto a Montreal Neurological Institute (MNI) template. The color bar indicates Z values. The values above the slices indicate the coordinates in MNI space. Right is right. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

terms in the Neurosynth database in order to determine what cognitive terms are most commonly associated with these brain regions in the literature. Since the Neurosynth decoder compares whole-brain images, both comparisons were performed on unthresholded data that contain additional information in the form of continuous values at all voxels (Gorgolewski et al., 2015). For the RSA maps, we retained the top 10 terms with the highest positive correlations referring to cognitive functions (i.e., anatomical terms were excluded), and afterwards we compared their correlations with the correlations observed for the GLM maps, in order to explore what cognitive terms associated with the RSA maps have similar associations with the social-random contrast.

3. Results

3.1. Behavioral results

The accuracy differed significantly between the two types of video content (social accuracy mean: 97.6%, SD: 15.5; random accuracy mean: 88.4%, SD: 32; Wilcoxon signed rank test $V = 353.5$, $p = .0006$, $r = 0.5$), but there were no significant correlations with trait measures (all p's > 0.25; see Supplementary Table 1 for all results). For the RTs, there was no main effect of video type (social RTs mean: 839 ms, SD: 374; random RTs mean: 837 ms, SD: 477; $F(1, 42) = 3.64$, $p = .06$, $\eta_p^2 = 0.08$) and no main effects of or interactions with trait measures (all p's > 0.09; see Supplementary Table 1 for all results).

3.2. Neuroimaging results

3.2.1. Whole-brain group results

The social-random content contrast at the group level showed widespread bilateral activations in regions typically associated with social content processing, comprising the fusiform gyrus, inferior and middle temporal regions, inferior frontal gyrus and superior medial frontal cortex (Fig. 1; Table 1). The results from the random-social contrast are included in the Supplementary material (Supplementary figure 1).

3.2.2. Representational similarity results

The representational similarity analysis testing whether the degree of agreeableness correlates with the dissimilarity in neural representations of social and random content showed a significant positive correlation in the dorsomedial prefrontal cortex (dmPFC; $r = 0.596$, $p < .008$) (Fig. 2). Importantly, no other personality trait correlated either positively or negatively with the social-random dissimilarity measure (Supplementary figure 2). As an additional control analysis, we included age, IQ (assessed with an abbreviated Raven test (Bilker et al., 2012) and gender as covariates in the model predicting dmPFC pattern dissimilarity from agreeableness. Pattern dissimilarity in the dmPFC remained uniquely associated with agreeableness [$F(1, 44) = 21.62$, $p < .0001$, $\eta_p^2 = 0.329$]. To examine whether the representational similarity of social content, and separately of random content, in the dmPFC varied with levels of agreeableness, we computed the average similarity between social videos and between random videos, and correlated those with the agreeableness scores. Although neither was significant (social: $r = -0.039$; random $r = -0.067$), we show the representational similarity matrices (RSM) within and between conditions in the dmPFC with a median split on agreeableness data (HA: high agreeableness, LA: low agreeableness; median = 3.7) for visualization purposes (Fig. 3A), to display the differences in social-random similarities between high and low agreeable individuals. We assessed the within and between conditions similarities between the two groups with a 2-way ANOVA (Fig. 3B), and found a significant difference between the two groups only for the between condition similarity (group × condition: $F(2, 50) = 11.14$, $p < .001$, $\eta_p^2 = 0.31$). As expected, the low agreeableness (LA) group had significantly higher representational similarity between social and non-social content than the high agreeableness group (HA) (Tukey adjusted $p < .0001$). Finally, the dissimilarity correlation with task accuracy showed no significant results (Supplementary figure 2).

3.2.3. Univariate results

The univariate analysis assessing whether the observed neural pattern dissimilarity associated with agreeableness could also be explained by univariate activation differences between social and non-social content showed no significant results (Supplementary figure 3).

3.2.4. Meta-analytic decoding results

The meta-analytic decoding was performed to provide an empirical interpretation of the pattern observed in the RSA unthresholded map. Out of the top 10 terms referring to cognitive constructs, five were related to social cognition, and all five had similarly high correlations for the GLM unthresholded map (Fig. 4), suggesting that the cognitive terms associated with variations in agreeableness are also correlated with activity patterns during social vs. random processing.

4. Discussion

Agreeableness trait is one of the five major dimensions of personality indexing individual variations in empathy, altruism and cooperation. Previous work has linked agreeableness with socio-cognitive abilities and, in particular, with the ability to infer and reason about others' mental states, also known as Theory of Mind (ToM) or mentalization (Allen et al., 2017; Nettle and Liddle, 2008). Recently, a new piece of evidence has strengthened this hypothesis by showing that individual differences in both agreeableness and ToM are related to variation in the same underlying neural network (Udochi et al., 2020). Here we complement and extend these findings by investigating how neural encoding of social information is related to agreeableness, therefore bridging the gap between neural and behavioral features of this trait. A

Table 1
Significant clusters from the whole brain voxelwise social – random content contrast.

Anatomical region	MNI x	y	z	Peak *Z*	Cluster level size	p
R. Inferior Temporal Gyrus	43	−55	−15	9.34	5933	<0.001
R. Fusiform Gyrus	42	−42	−15	8.54		
R. Middle Temporal Gyrus	60	−45	9	8.47		
R. Amygdala	21	−4	−13	8.46		
L. Fusiform Gyrus	−33	−42	−15	8.65	3089	<0.001
L. Inferior Temporal Gyrus	−36	−36	−15	8.58		
L. Inferior Occipital Gyrus	−26	−98	−10	8.26		
L. Hippocampus	−24	−12	−12	7.47	1062	<0.001
L. Amygdala	−27	0	−18	6.5		
L. Superior Temporal Pole	−42	18	−24	6.31		
L. Inferior Frontal Gyrus (operc.)	−48	15	24	5.92		
L. Inferior Frontal Gyrus (triang.)	−51	30	0	5.63		
L. Superior Medial Frontal Gyrus	−6	57	30	6.55	415	<0.001
R. Superior Medial Frontal Gyrus	6	54	30	6.27		
R. Supplementary Motor Area	9	15	66	5.35		
L. Middle Temporal Pole	−54	0	−18	6.57	210	<0.001
L. Precuneus	−9	−51	45	4.7	114	<0.001
R. Precuneus	3	−60	42	4.22		
L. Middle Frontal Gyrus	−39	3	54	4.87	61	0.0013
Vermis	0	−48	−21	4.87	38	0.0176

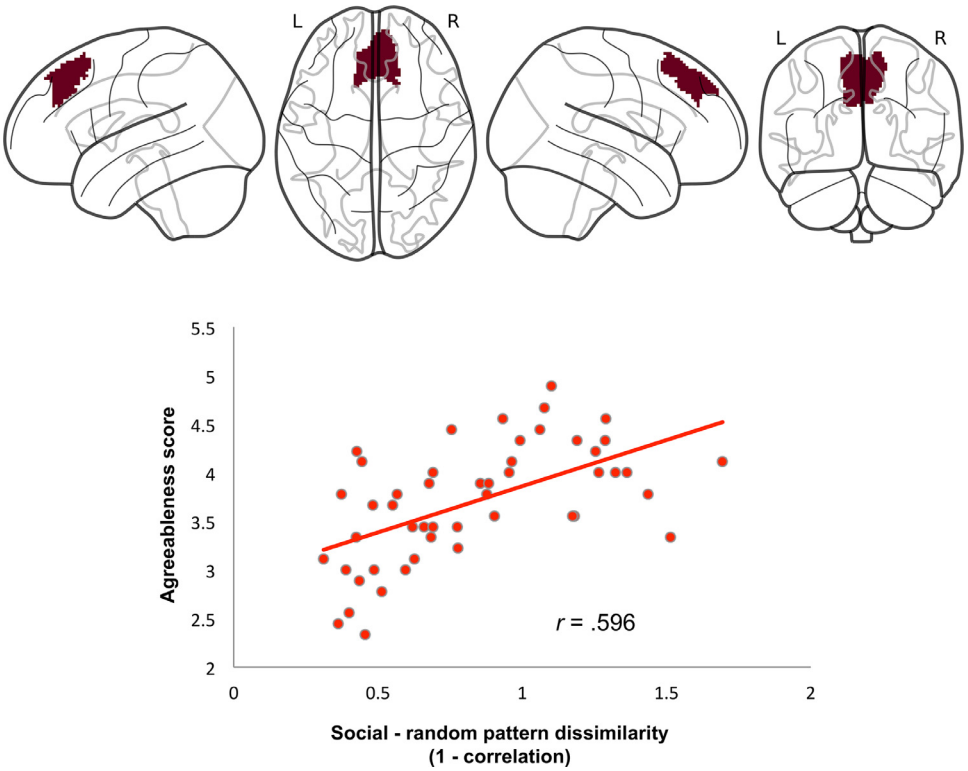

Fig. 2. Agreeableness correlation with social-random pattern dissimilarity. Results from the whole brain representational similarity analysis assessing the correlation between the degree of agreeableness and the dissimilarity in neural representations of social and random content. The upper panel shows the dorsomedial prefrontal cortex (dmPFC), the only region showing a significant positive correlation (permutation based *p*-values were Bonferroni-corrected over all tested regions - 200). The scatterplot in the lower panel shows the correlation between the participants' agreeableness trait scores and their social – random content pattern dissimilarity. The latter was calculated by correlating social and random response patterns and subtracting it from 1 (i.e., higher correlation is reflected as lower dissimilarity).

straightforward prediction is that individuals high in agreeableness encode social and non-social information in a more dissimilar fashion with respect to individuals low in agreeableness, and primarily in brain regions involved during mentalization. To test this prediction we adopted a representational similarity analysis (RSA) approach that derives the degree of similarity between patterns of activation for different stimuli (Kriegeskorte et al., 2008). Since we were interested in assessing whether the degree of similarity between social and non-social content neural representation can predict the degree of agreeableness, we compared on an individual basis patterns of brain activation during a classic ToM animation task in which different shapes could interact in socially meaningful way, or randomly (Castelli et al., 2000; Wheatley et al., 2007). Consistent with our predictions, we found that more agreeable individuals had more distinct encodings of social and random animations, and correspondingly these encodings were more similar in individuals with low agreeableness. Critically, this positive association between agreeableness and representational dissimilarity of social and random content was significant only in the dmPFC, a brain region consistently related to attributions of mental states (Molenberghs et al., 2016; Schurz et al., 2014). This assumption was also confirmed by the meta-analytic decoding of our RSA results, which showed strongest associations with socio-cognitive topics.

The dmPFC has been largely acknowledged as one of the key regions in the mentalizing network, although there is still no consensus regarding its functional role. Indeed, numerous meta-analyses have delineated the dmPFC involvement within different social, affective and

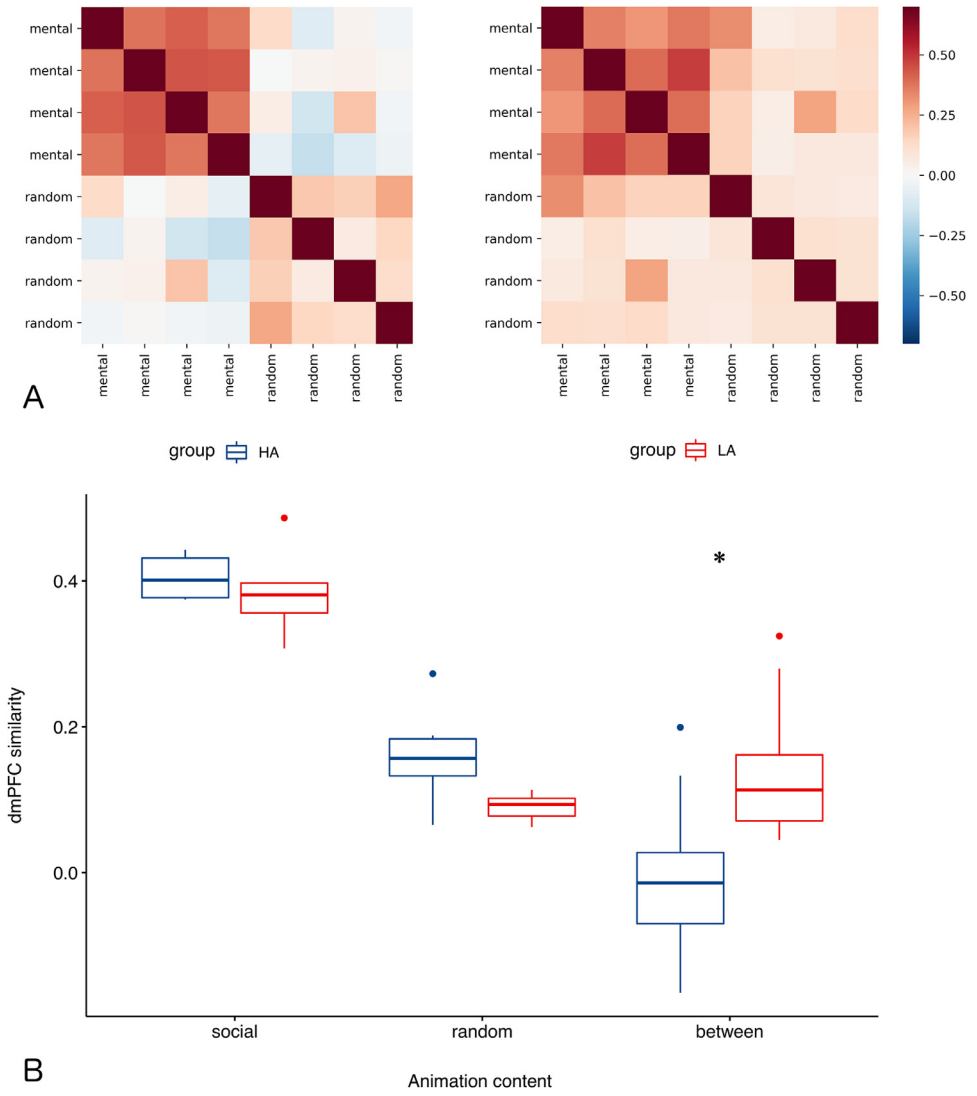

Fig. 3. High agreeableness (HA) and low agreeableness (LA) individuals' representational similarity in the dmPFC. (A) The representational similarity matrices (RSM) within and between conditions in the dmPFC with a median split on agreeableness data (median = 3.7) for visualization purposes. The color bar indicates the Pearson's *r*. (B) Average within and between conditions similarities (Pearson correlation) in the HA and LA groups (* $p < .0001$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

cognitive functions (Binder et al., 2009; Kober et al., 2008; Kogler et al., 2020; Molenberghs et al., 2016; Schurz et al., 2014; Spreng et al., 2009; Van Overwalle, 2009), making it even more difficult to reconcile all the observed data into a single picture. A different approach gathering evidence from causal and reverse inference studies, which are more suited for testing structure-function relationships, showed the dmPFC strongest functional link with social cognition and mental state inference processes (Lieberman et al., 2019). In particular, there have been a growing number of studies adopting multivariate pattern analyses that have characterized and/or decoded the neural representation of social information within the dmPFC (e.g., Corradi-Dell'Acqua et al., 2014; Dungan et al., 2016; Skerry and Saxe, 2015; see Wagner et al., 2018 for a revew). In a more comprehensive study, Tamir et al. (2016) investigated what psychological dimensions that organize the understanding of mental states shape also their neural representations in the brain, and one specific component, "rationality", loading highly on dimensions such as emotion, experience and warmth, predicted reliably the neural patterns in the dmPFC. Although in our study we did not consider different dimensions of social information that would allow us to capture what particular dimension is represented differently in individuals with different levels of agreeableness, we speculate that social videos conveyed emotional information that could be inferred from the interactions between the shapes (e.g., fear, surprise, compassion, happiness, irritation, etc.). Highly agreeable individuals might have spontaneously inferred these emotional states, increasing the pattern dissimilarity in the dmPFC between social and non-social videos, contrarily to individuals low in agreeableness. This speculation would also be in line with the supramodal representations of perceived emotions observed in the dmPFC (Peelen et al., 2010; Skerry and Saxe, 2014). This hypothesis should however be tested with an appropriate task design, in particular by modulating the emotional content across stimuli and assessing its impact on neural representations in individuals with different levels of agreeableness.

Previous research on inter-individual differences in mentalization ability has typically focused on neurological and psychiatric conditions, evidencing its importance in every day functioning (Baron-Cohen, 1995; Kerr et al., 2003; Richell et al., 2003; Snowden et al., 2003; Stuss et al., 2001). Nevertheless, it is well acknowledged that there are substantial mentalization differences also in healthy adults, yet these are difficult to capture behaviorally, mainly due to ceiling effects observed on standard laboratory tasks (Koster-Hale and Saxe, 2013). To overcome this issue, one possible strategy is to assess neural variations to ambiguous stimuli that induce spontaneous mentalization, which we implemented in our study. Few other studies have adopted this approach to investigate inter-subject variability in mentalization (Moessnang et al., 2017; Moriguchi et al., 2006; Udochi et al., 2020; Wagner et al., 2011), and most of them observed modulation of activity in the dmPFC in relation to trait-like scores. Specifically, higher empathy (Wagner et al., 2011)

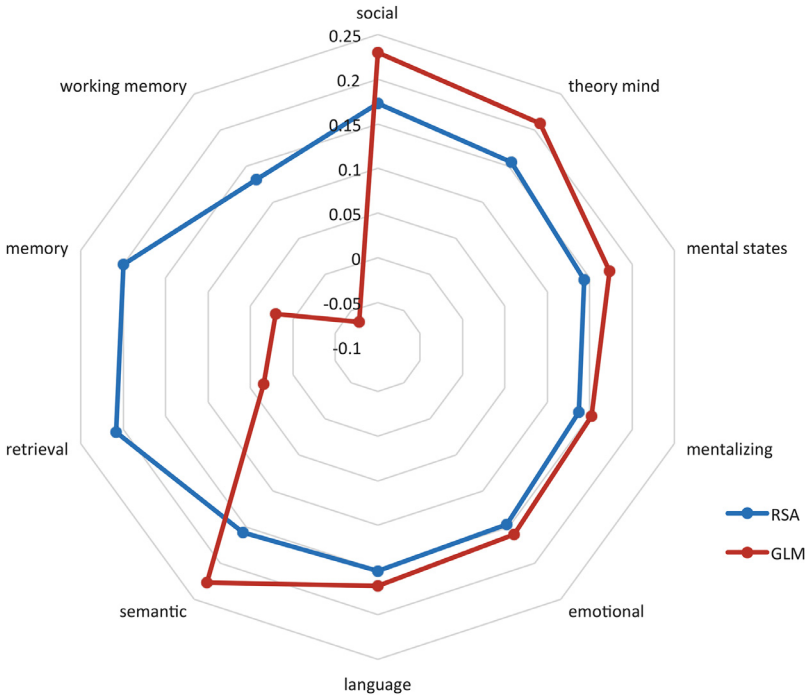

Fig. 4. Meta-analytic decoding. Unthresholded RSA and GLM maps show similar associations with social cognitive terms from the Neurosynth database.

and perspective-taking scores (Moriguchi et al., 2006), and lower autism trait (Moessnang et al., 2017) were associated with a greater recruitment of dmPFC, suggesting a more enhanced processing of social information. In line with this evidence, our finding provides additional understanding regarding the distinctiveness of social information in the dmPFC and how it underlies individual differences. In particular, by comparing patterns of activity during social and non-social content processing, we observed that the degree to which the two information are distinguished in the dmPFC predicts individual variations in agreeableness, a trait related to mentalizing ability (Allen and DeYoung, 2016). It is important to point out that this result can reflect a different encoding of social and/or non-social information in individuals with different levels of agreeableness, or alternatively it can reflect differences in processes related to the task at hand (e.g., emotion recognition but also attentional processes, cognitive control, etc.). However, given the greater neural similarity for social videos observed in the dmPFC (Fig. 3), and the frequent involvement of this region in processing social interactions (e.g., Wagner et al., 2016), we are confident that differences in social process are a more plausible interpretation. Echoing previous correlational fMRI studies, we addressed a potential alternative account according to which the agreeableness trait could be predicted by average activation differences between social and non-social content. Results from this analysis showed no significant trait prediction, suggesting that individual variations in agreeableness are better reflected within representational differences of social and non-social information than the overall difference in activity they evoke. Still, we do not know if these differences in social content representation are functionally relevant, since we did not observe similar relations with task performance, probably due to an insensitive behavioral measure we adopted. However, it would be plausible to expect more dissimilar encoding of social vs. non-social information in relation to better understanding of the events depicted in the animations. Future studies should implement more detailed task performance measures to investigate the functional importance of the underlying neural representations. One potential limitation of this study, which probably reduced the power of our main brain-behavior correlation analyses, is the moderate sample size. Although we have adopted a sampling strategy (i.e., we selected our participants from a larger sample based on their personality scores), which was found to increase the power to detect moderate effect sizes in small samples without inflating the false positive rate (de Haas, 2018), future studies should try to replicate this finding in larger samples.

Lastly, but equally important, the observed task-related representational difference underlying agreeableness argues for a different, behavior-based approach in studying the cognitive and neural mechanisms behind personality. In personality neuroscience, agreeableness has been one of the least studied traits (Allen and DeYoung, 2016) and until recently its neurobiological substrates have been explored only among stable brain features like structural or resting-state measures (e.g., Cai et al., 2020; Riccelli et al., 2017). Although these studies were rarely underpowered in terms of sample size, they reported very divergent findings, or even null-results (e.g., Avinun et al., 2020; Dubois et al., 2018). One possible explanation for these inconsistent findings may involve the complexity of behaviors associated with agreeableness and its link with various processes such as empathy, self-regulation and motivation (Graziano and Tobin, 2016). This could result in multiple brain systems mediating variability in the agreeableness trait, which would therefore be hardly identifiable with structural and resting-state brain measures. This explanation is in line with a recent functional connectivity study (Liu et al., 2019) in which the agreeableness trait was found to be correlated to a widely distributed connectivity pattern comprising the largest number of connections, almost four times that of other traits. Therefore our finding linking agreeableness to mentalization probably captures only one aspect of agreeableness and future studies should implement more itemized personality measures to identify other cognitive and neural determinants of agreeableness.

Conclusions

Despite its limitations, the present study represents a step forward in characterizing the neural determinants of agreeableness, a personality trait highly related to socio-cognitive abilities. Our results suggest that neural representations of social content in the dmPFC can vary among individuals with different levels of agreeableness in a functionally relevant way: more similar representations of social and non-social content predict lower agreeableness scores, revealing the link between neural and behavioral mechanisms underlying agreeableness. Additionally,

these sorts of connections between personality traits and specific cognitive abilities provide new opportunities for the development of more objective personality measures.

Data statement

Data/code availability statement: Raw MRI data, behavioral data and MRI quality metrics are available on OpenNeuro (https://openneuro.org/datasets/ds003436).

Declarations of Competing Interest

None.

Credit authorship contribution statement

Sandra Arbula: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Elisabetta Pisanu:** Validation, Investigation, Data curation, Writing – review & editing. **Raffaella I. Rumiati:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition.

Acknowledgments

We thank Dr. Tania Cerni and Dr. Luca Piretti for their discussion during the early phases of the project.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.118049.

References

Allen, T.A., DeYoung, C.G., 2016. Personality Neuroscience and the Five Factor Model, Vol. 1 doi:10.1093/oxfordhb/9780199352487.013.26.

Allen, T.A., Rueter, A.R., Abram, S.V., Brown, J.S., DeYoung, C.G., 2017. Personality and neural correlates of mentalizing ability. Eur. J. Pers. 31 (6), 599–613. doi:10.1002/per.2133.

Andersson, J.L.R., Skare, S., Ashburner, J, 2003. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. Neuroimage 20 (2), 870–888. doi:10.1016/S1053-8119(03)00336-7.

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. Med. Image Anal. 12 (1), 26–41. doi:10.1016/J.MEDIA.2007.06.004.

Avinun, R., Israel, S., Knodt, A.R., Hariri, A.R., 2020. Little evidence for associations between the big five personality traits and variability in brain gray or white matter. Neuroimage 220, 117092. doi:10.1016/J.NEUROIMAGE.2020.117092.

Barch, D.M., Burgess, G.C., Harms, M.P., Petersen, S.E., Schlaggar, B.L., Corbetta, M., Glasser, M.F., Curtiss, S., Dixit, S., Feldt, C., Nolan, D., Bryant, E., Hartley, T., Footer, O., Bjork, J.M., Poldrack, R., Smith, S., Johansen-Berg, H., Snyder, A.Z., Van Essen, D.C., 2013. Function in the human connectome: task-fMRI and individual differences in behavior. Neuroimage 80, 169–189. doi:10.1016/J.NEUROIMAGE.2013.05.033.

Baron-Cohen, S., 1995. Mindblindness : an Essay on Autism and Theory of Mind. MIT Press.

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. doi:10.1016/J.NEUROIMAGE.2007.04.042.

Bilker, W.B., Hansen, J.A., Brensinger, C.M., Richard, J., Gur, R.E., Gur, R.C., 2012. Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. Assessment 19 (3), 354–369. doi:10.1177/1073191112446655.

Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., 2009. Where is the semantic system? a critical review and meta-analysis of 120 functional neuroimaging studies. Cereb. Cortex 19 (12), 2767–2796. doi:10.1093/cercor/bhp055.

Cai, H., Zhu, J., Yu, Y., 2020. Robust prediction of individual personality from brain functional connectome. Soc. Cogn. Affect. Neurosci. 15 (3), 359–369. doi:10.1093/scan/nsaa044.

Carlson, T., Goddard, E., Kaplan, D.M., Klein, C., Ritchie, J.B., 2018. Ghosts in machine learning for cognitive neuroscience: moving from data to theory. Neuroimage 180, 88–100. doi:10.1016/J.NEUROIMAGE.2017.08.019.

Castelli, F., Happé, F., Frith, U., Frith, C., 2000. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. Neuroimage 12 (3), 314–325. doi:10.1006/nimg.2000.0612.

Corr, P.J., DeYoung, C.G., McNaughton, N., 2013. Motivation and personality: a neuropsychological perspective. Soc. Personal Psychol. Compass 7 (3), 158–175. doi:10.1111/spc3.12016.

Corradi-Dell'Acqua, C., Hofstetter, C., Vuilleumier, P., 2014. Cognitive and affective theory of mind share the same local patterns of activity in posterior temporal but not medial prefrontal cortex. Soc Cogn Affect Neurosci 9 (8), 1175–1184. doi:10.1093/scan/nst097.

Costa, P.T., McCrae, R.R., 1992. The five-factor model of personality and its relevance to personality disorders. J. Pers. Disord. 6 (4), 343–359. doi:10.1521/pedi.1992.6.4.343.

Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29 (3), 162–173. doi:10.1006/cbmr.1996.0014.

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis: I. Segmentation and surface reconstruction. Neuroimage 9 (2), 179–194. doi:10.1006/NIMG.1998.0395.

de Haas, B., 2018. How to enhance the power to detect brain–behavior correlations with limited resources. Front. Hum. Neurosci. 12, 421. doi:10.3389/fnhum.2018.00421.

DeYoung, C.G., 2010. Personality neuroscience and the biology of traits. Soc. Personal Psychol. Compass 4 (12), 1165–1180. doi:10.1111/j.1751-9004.2010.00327.x.

Dubois, J., Galdi, P., Han, Y., Paul, L.K., Adolphs, R., 2018. Resting-state functional brain connectivity best predicts the personality dimension of openness to experience. Personal. Neurosci. 1, e6. doi:10.1017/pen.2018.8.

Dungan, J.A., Stepanovic, M., Young, L., 2016. Theory of mind for processing unexpected events across contexts. Soc. Cogn. Affect. Neurosci. 11 (8), 1183–1192. doi:10.1093/scan/nsw032.

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. doi:10.1371/journal.pone.0184661.

Esteban, O., Blair, R.W., Nielson, D.M., Varada, J.C., Marrett, S., Thomas, A.G., Poldrack, R.A., Gorgolewski, K.J., 2019a. Crowdsourced MRI quality metrics and expert quality annotations for training of humans and machines. Sci. Data 6 (1), 30. doi:10.1038/s41597-019-0035-4.

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., 2019b. fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat. Methods 16 (1), 111–116. doi:10.1038/s41592-018-0235-4.

Etzel, J.A., Courtney, Y., Carey, C.E., Gehred, M.Z., Agrawal, A., Braver, T.S., 2020. Pattern similarity analyses of frontoparietal task coding: individual variation and genetic influences. Cereb. Cortex 30 (5), 3167–3183. doi:10.1093/cercor/bhz301.

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. doi:10.1016/S1053-8119(09)70884-5.

Gorgolewski, K.J., 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. Front. Neuroinform. 5, 13. doi:10.3389/fninf.2011.00013.

Gorgolewski, K.J., Esteban, O., Ellis, D.G., Notter, M.P., Ziegler, E., Johnson, H., Hamalainen, C., Yvernault, B., Burns, C., Manhães-Savio, A., Jarecka, D., Markiewicz, C.J., Salo, T., Clark, D., Waskom, M., Wong, J., Modat, M., Dewey, B.E., Clark, M.G., & Ghosh, S. (2017). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. 0.13.1. https://doi.org/10.5281/ZENODO.581704

Gorgolewski, K.J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S.S., Maumet, C., Sochat, V.V., Nichols, T.E., Poldrack, R.A., Poline, J.-.B., Yarkoni, T., Margulies, D.S., 2015. NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. Front. Neuroinform. 9, 8. doi:10.3389/fninf.2015.00008.

Gray, J.A., 1982. The Neuropsychology of Anxiety: An Enquiry into the Functions of the Septo-Hippocampal System. Clarendon Press/Oxford University Press.

Graziano, W.G., Habashi, M.M., Sheese, B.E., Tobin, R.M., 2007. Agreeableness, empathy, and helping: a person x situation perspective. J. Pers. Soc. Psychol. 93 (4), 583–599. doi:10.1037/0022-3514.93.4.583.

Graziano, W.G., Tobin, R.M., 2016. Agreeableness and the five factor model. *The Oxford handbook of the Five Factor Model* (Issue April 2018) doi:10.1093/oxfordhb/9780199352487.013.17.

Greve, D.N., Fischl, B., 2009. Accurate and robust brain image alignment using boundary-based registration. Neuroimage 48 (1), 63–72. doi:10.1016/J.NEUROIMAGE.2009.06.060.

Habashi, M.M., Graziano, W.G., Hoover, A.E., 2016. Searching for the prosocial personality: a Big Five approach to linking personality and prosocial behavior. Pers. Soc. Psychol. Bull. 42 (9), 1177–1192. doi:10.1177/0146167216652859.

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. doi:10.1006/NIMG.2002.1132.

John, O.P., Naumann, L.P., Soto, C.J., 2008. Paradigm shift to the integrative Big Five trait taxonomy: history, measurement, and conceptual issues. In: Handbook of personality: Theory and Research. The Guilford Press, pp. 114–158.

Kerr, N., Dunbar, R.I.M., Bentall, R.P, 2003. Theory of mind deficits in bipolar affective disorder. J. Affect. Disord. 73 (3), 253–259. doi:10.1016/s0165-0327(02)00008-3.

Klein, A., Ghosh, S.S., Bao, F.S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B.,

Reuter, M., Chaibub Neto, E., Keshavan, A., 2017. Mindboggling morphometry of human brains. PLoS Comput. Biol. 13 (2), e1005350. doi:10.1371/journal.pcbi.1005350.

Kober, H., Barrett, L.F., Joseph, J., Bliss-Moreau, E., Lindquist, K., Wager, T.D., 2008. Functional grouping and cortical-subcortical interactions in emotion: a meta-analysis of neuroimaging studies. Neuroimage 42 (2), 998–1031. doi:10.1016/j.neuroimage.2008.03.059.

Kogler, L., Müller, V.I., Werminghausen, E., Eickhoff, S.B., Derntl, B., 2020. Do I feel or do I know? Neuroimaging meta-analyses on the multiple facets of empathy. Cortex 129, 341–355. doi:10.1016/j.cortex.2020.04.031.

Koster-Hale, J., Saxe, R., 2013. Functional neuroimaging of theory of mind. In: Understanding Other Minds. Oxford University Press, pp. 132–163. doi:10.1093/acprof:oso/9780199692972.003.0009.

Kriegeskorte, N., Mur, M., Bandettini, P.A., 2008. Representational similarity analysis – connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2, 4. doi:10.3389/neuro.06.004.2008.

Kuper, N., Käckenmester, W., Wacker, J., 2019. Resting frontal EEG asymmetry and personality traits: a meta-analysis. Eur. J. Pers. 33 (2), 154–175. doi:10.1002/per.2197.

Lewis, G.J., Dickie, D.A., Cox, S.R., Karama, S., Evans, A.C., Starr, J.M., Bastin, M.E., Wardlaw, J.M., Deary, I.J., 2018. Widespread associations between trait conscientiousness and thickness of brain cortical regions. Neuroimage 176, 22–28. doi:10.1016/j.neuroimage.2018.04.033.

Lieberman, M.D., Straccia, M.A., Meyer, M.L., Du, M., Tan, K.M., 2019. Social, self, (situational), and affective processes in medial prefrontal cortex (MPFC): causal, multivariate, and reverse inference evidence. Neurosci. Biobehav. Rev. 99, 311–328. doi:10.1016/j.neubiorev.2018.12.021.

Liu, W., Kohn, N., Fernández, G., 2019. Intersubject similarity of personality is associated with intersubject similarity of brain connectivity patterns. Neuroimage 186, 56–69. doi:10.1016/J.NEUROIMAGE.2018.10.062.

Moessnang, C., Otto, K., Bilek, E., Schäfer, A., Baumeister, S., Hohmann, S., Poustka, L., Brandeis, D., Banaschewski, T., Tost, H., Meyer-Lindenberg, A., 2017. Differential responses of the dorsomedial prefrontal cortex and right posterior superior temporal sulcus to spontaneous mentalizing. Hum. Brain Mapp. 38 (8), 3791–3803. doi:10.1002/hbm.23626.

Molenberghs, P., Johnson, H., Henry, J.D., Mattingley, J.B., 2016. Understanding the minds of others: a neuroimaging meta-analysis. Neurosci. Biobehav. Rev. 65, 276–291. doi:10.1016/j.neubiorev.2016.03.020.

Moriguchi, Y., Ohnishi, T., Lane, R.D., Maeda, M., Mori, T., Nemoto, K., Matsuda, H., Komaki, G., 2006. Impaired self-awareness and theory of mind: an fMRI study of mentalizing in alexithymia. Neuroimage 32 (3), 1472–1482. doi:10.1016/J.NEUROIMAGE.2006.04.186.

Mulders, P., Llera, A., Tendolkar, I., van Eijndhoven, P., Beckmann, C., 2018. Personality profiles are associated with functional brain networks related to cognition and emotion. Sci. Rep. 8 (1), 13874. doi:10.1038/s41598-018-32248-x.

Naselaris, T., Kay, K.N., 2015. Resolving ambiguities of MVPA using explicit models of representation. Trends Cogn. Sci. 19 (10), 551–554. doi:10.1016/J.TICS.2015.07.005.

Nettle, D., Liddle, B., 2008. Agreeableness is related to social-cognitive, but not social-perceptual, theory of mind. Eur. J. Pers. 22 (4), 323–335. doi:10.1002/per.672.

Nostro, A.D., Müller, V.I., Varikuti, D.P., Pläschke, R.N., Hoffstaedter, F., Langner, R., Patil, K.R., Eickhoff, S.B., 2018. Predicting personality from network-based resting-state functional connectivity. Brain Struct. Funct. 223 (6), 2699–2719. doi:10.1007/s00429-018-1651-z.

Omura, K., Todd Constable, R., Canli, T., 2005. Amygdala gray matter concentration is associated with extraversion and neuroticism. Neuroreport 16 (17), 1905–1908. doi:10.1097/01.wnr.0000186596.64458.76.

Owens, M.M., Hyatt, C.S., Gray, J.C., Carter, N.T., MacKillop, J., Miller, J.D., Sweet, L.H., 2019. Cortical morphometry of the five-factor model of personality: findings from the human connectome project full sample. Soc. Cogn. Affect. Neurosci. 14 (4), 381–395. doi:10.1093/scan/nsz017.

Peelen, M.V., Atkinson, A.P., Vuilleumier, P., 2010. Supramodal representations of perceived emotions in the human brain. J. Neurosci. 30 (30), 10127–10134. doi:10.1523/JNEUROSCI.2161-10.2010.

Penner, L.A., Fritzsche, B.A., Craiger, J.P., Freifeld, T.S., 1995. Measuring the prosocial personality. In: Advances in Personality Assessment, Vol. 10. Lawrence Erlbaum Associates, Inc, pp. 147–163.

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. doi:10.1016/J.NEUROIMAGE.2013.08.048.

Riccelli, R., Toschi, N., Nigro, S., Terracciano, A., Passamonti, L., 2017. Surface-based morphometry reveals the neuroanatomical basis of the five-factor model of personality. Soc. Cogn. Affect. Neurosci. 12 (4), 671–684. doi:10.1093/scan/nsw175.

Richell, R.A., Mitchell, D.G.V, Newman, C., Leonard, A., Baron-Cohen, S., Blair, R.J.R, 2003. Theory of mind and psychopathy: can psychopathic individuals read the "language of the eyes"?. Neuropsychologia, Vol. 41 http://docs.autismresearchcentre.com/papers/2003_Richell_etal.pdf.

Ritchie, J.B., Kaplan, D.M., Klein, C., 2019. Decoding the brain: neural representation and the limits of multivariate pattern analysis in cognitive neuroscience. Br. J. Philos. Sci. 70 (2), 581–607. doi:10.1093/bjps/axx023.

Schneider, W., Eschman, A., Zuccolotto, A., 2012. E-Prime 2.0. Psychology Software Tools, Inc.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., Perner, J., 2014. Fractionating theory of mind: a meta-analysis of functional brain imaging studies. Neurosci. Biobehav. Rev. 42, 9–34. doi:10.1016/J.NEUBIOREV.2014.01.009.

Skerry, A.E., Saxe, R., 2014. A common neural code for perceived and inferred emotion. J. Neurosci. 34 (48), 15997–16008. doi:10.1523/JNEUROSCI.1676-14.2014.

Skerry, A.E., Saxe, R., 2015. Neural representations of emotion are organized around abstract event features. Curr. Biol. 25 (15), 1945–1954. doi:10.1016/J.CUB.2015.06.009.

Snowden, J.S., Gibbons, Z.C., Blackshaw, A., Doubleday, E., Thompson, J., Craufurd, D., Foster, J., Happé, F., Neary, D., 2003. Social cognition in frontotemporal dementia and Huntington's disease. Neuropsychologia 41 (6), 688–701. doi:10.1016/s0028-3932(02)00221-x.

Spreng, R.N., Mar, R.A., Kim, A.S.N, 2009. The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. J. Cogn. Neurosci. 21 (3), 489–510. doi:10.1162/jocn.2008.21029.

Stuss, D.T., Gallup, G.G., Alexander, M.P., 2001. The frontal lobes are necessary for "theory of mind.". Brain 124 (2), 279–286. doi:10.1093/brain/124.2.279.

Taki, Y., Thyreau, B., Kinomura, S., Sato, K., Goto, R., Wu, K., Kawashima, R., Fukuda, H., 2013. A longitudinal study of the relationship between personality traits and the annual rate of volume changes in regional gray matter in healthy adults. Hum. Brain Mapp. 34 (12), 3347–3353. doi:10.1002/hbm.22145.

Tamir, D.I., Thornton, M.A., Contreras, J.M., Mitchell, J.P., 2016. Neural evidence that three dimensions organize mental state representation: rationality, social impact, and valence. PNAS 113 (1), 194–199. doi:10.1073/pnas.1511905112.

Tellegen, A., Waller, N.G., 1981. Exploring personality through test construction: development of the multidimensional personality questionnaire. In: The SAGE Handbook of Personality Theory and Assessment: Volume 2 — Personality Measurement and Testing. SAGE Publications Ltd, pp. 261–292. doi:10.4135/9781849200479.n13.

Ubbiali, A., Chiorri, C., Hampton, P., Donati, D., 2013. Italian Big Five inventory. psychometric properties of the italian adaptation of the Big Five inventory (BFI). Appl. Psychol. Bull. 266 (59), 37–48.

Udochi, A.L., Blain, S.D., Burton, P., Medrano, L., & DeYoung, C.G. (2020). Activation of the default network during a theory of mind task predicts individual differences in agreeableness and social cognitive ability. *PsyArXiv*. https://doi.org/10.31234/osf.io/prhau

Van Overwalle, F., 2009. Social cognition and the brain: a meta-analysis. Hum. Brain Mapp. 30 (3), 829–858. doi:10.1002/hbm.20547.

Wagner, D.D., Chavez, R.S., Broom, T.W., 2018. Decoding the neural representation of self and person knowledge with multivariate pattern analysis and data-driven approaches. *Wiley Interdiscipl. Rev.: Cogn. Sci., August* 1–19. doi:10.1002/wcs.1482.

Wagner, D.D., Kelley, W.M., Haxby, J.V., Heatherton, T.F., 2016. The dorsal medial prefrontal cortex responds preferentially to social interactions during natural viewing. J. Neurosci. 36 (26), 6917–6925. doi:10.1523/JNEUROSCI.4220-15.2016.

Wagner, D.D., Kelley, W.M., Heatherton, T.F., 2011. Individual differences in the spontaneous recruitment of brain regions supporting mental state understanding when viewing natural social scenes. Cereb. Cortex 21 (12), 2788–2796. doi:10.1093/cercor/bhr074.

Wheatley, T., Milleville, S.C., Martin, A., 2007. Understanding animate agents: distinct roles for the social network and mirror system. Psychol. Sci. 18 (6), 469–474. doi:10.1111/j.1467-9280.2007.01923.x.

Yarkoni, T., 2015. Neurobiological substrates of personality: a critical overview. Pers. Process. Individ. Diff. 4, 61–83. doi:10.1017/CBO9781107415324.004.

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 Trans. Med. Imaging 20 (1), 45–57. doi:10.1109/42.906424.