Activation of the default network during a theory of mind task predicts individual differences in agreeableness and social cognitive ability

Aisha L. Udochi [1] · Scott D. Blain [1] · Tyler A. Sassenberg [1] · Philip C. Burton [1] · Leroy Medrano [1] · Colin G. DeYoung [1]

Abstract

Social cognitive processes, such as emotion perception and empathy, allow humans to navigate complex social landscapes and are associated with specific neural systems. In particular, *theory of mind* (ToM), which refers to our ability to decipher the mental states of others, is related to the dorsal medial prefrontal cortex and temporoparietal junction, which include portions of the default network. Both social cognition and the default network have been linked to the personality trait Agreeableness. We hypothesized that default network activity during a ToM task would positively predict social cognitive abilities and Agreeableness. In a 3T fMRI scanner, participants ($N = 1050$) completed a ToM task in which they observed triangles displaying random or social (i.e., human-like) movement. Participants also completed self-report measures of Agreeableness and tests of intelligence and social cognitive ability. In each participant, average blood oxygen level dependent responses were calculated for default network regions associated with social cognition, and structural equation modeling was used to test associations of personality and task performance with activation in those brain regions. Default network activation in the dorsal medial subsystem was greater for social versus random animations. Default network activation in response to social animations predicted better performance on social cognition tasks and, to a lesser degree, higher Agreeableness. Neural response to social stimuli in the default network may be associated with effective social processing and could have downstream effects on social interactions. We discuss theoretical and methodological implications of this work for social and personality neuroscience.

Introduction

As highly social animals, humans are tasked daily with navigating complex social interactions. In order to succeed in these interactions, we often rely on social cognitive processes that allow us to understand the targets of social interaction, including the ability to perceive and empathize with others' emotions (Barrett et al., 2011; Singer & Klimecki, 2014). In particular, *theory of mind* (ToM), or *mentalization*, describes

Aisha L. Udochi and Scott D. Blain contributed equally to this work.

✉ Aisha L. Udochi
udoch001@umn.edu

✉ Scott D. Blain
blain058@umn.edu

[1] Department of Psychology, University of Minnesota Twin Cities, Elliott Hall, 75 E River Rd, Minneapolis, MN 55455, United States

peoples' ability to recognize and understand the mental states of other people (Gallagher & Frith, 2003). ToM has been positively correlated with social competence (Bosacki & Wilde Astington, 1999; Liddle & Nettle, 2006) and negatively associated with aggressive tendencies (Meier et al., 2006; Mohr et al., 2007). ToM also is positively correlated with social cooperativeness (Paal & Bereczkei, 2007) and social network size, such that individuals with better ToM abilities reported larger networks of friends compared with individuals with less-developed ToM abilities (Liddle & Nettle, 2006; Stiller & Dunbar, 2007). Collectively, these findings suggest the importance of ToM in facilitating successful social interactions.

Although nearly all people are capable of demonstrating ToM to some degree, research on individual differences in ToM can show how variations in ToM ability influence real-world social outcomes. For example, poor ToM performance has been a predictor in mental health research; people with

autism (Baron-Cohen et al., 1985; Pinkham et al., 2008) and schizophrenia (Harrington et al., 2005; Pedersen et al., 2012; Pinkham et al., 2008)—two disorders linked with social impairments such as the inability to form appropriate relationships with others (Sasamoto et al., 2011)—have been found to have poor ToM ability. Poor ToM ability has also been correlated with a lack of understanding of how one's actions affect other people, and with difficulty in accurately assessing others' intentions (Baker, 2003).

The current literature on individual differences in ToM ability is limited. What does exist is largely siloed into subfields, such as developmental or clinical psychology. A large portion of such work focuses on variations in the development of ToM among children (Bowman et al., 2017; Devine & Hughes, 2013; Wang et al., 2016) using false-belief tasks that do not capture the full complexity of ToM (Altschuler et al., 2018; Apperly, 2012; Tager-Flusberg, 2011). Much of the remaining research focuses on ToM deficits in those with mental illness or developmental disabilities (Baron-Cohen et al., 1985; Dahlgren et al., 2010; Kerr et al., 2003; Pedersen et al., 2012), but such studies often are limited in their sample size and methods of assessing ToM ability. Additionally, a majority of research on ToM uses only single tasks to measure this ability, likely failing to capture the breadth of skills and abilities encompassed by ToM (Altschuler et al., 2018; Apperly, 2012). Considering the scarce research on variation among healthy adults, further research on individual differences in ToM ability could help to elucidate the underlying causes for variation in ToM and how this variation might relate to functional outcomes in daily life, eventually paving the way for more effective identification and intervention for those with deficits in these abilities (regardless of whether they meet criteria for specific psychological or developmental disorders).

Theory of Mind and the Default Network

Investigating the neural substrates of individual differences in ToM is one potentially useful approach to increasing scientific understanding of ToM. ToM has been consistently linked to brain regions such as the dorsal medial prefrontal cortex (dmPFC) and temporoparietal junction (TPJ; Allen et al., 2017; Amodio & Frith, 2006; Carrington & Bailey, 2009; Frith & Frith, 2006; Sabbagh et al., 2004; Saxe & Kanwisher, 2003; Saxe & Powell, 2006; Saxe & Wexler, 2005; Schurz et al., 2014; Spunt & Lieberman, 2012; Vogeley et al., 2001; Young et al., 2010). These regions of the brain are included in what is called the default network. Originally conceptualized as the brain's default mode because it was found to be more active when people were merely at rest in the MRI scanner rather than engaged in tasks involving outwardly-directed attention, the default network is now thought of as a network broadly involved in internal simulation, and tasks have been identified that activate it specifically (Andrews-Hanna et al., 2014; Smith et al., 2009). Functions of the default network appear to include simulating the mental states of others and simulating one's own experience during memory, prospection, or fantasy (Allen et al., 2017; Blain, Grazioplene, et al., 2020a; Mars et al., 2012; Meyer, 2019; Nettle & Liddle, 2008; Schilbach et al., 2008, 2012; Schurz et al., 2014; Seitz et al., 2006; Tamir et al., 2016). It is worth noting that the so-called "social brain" encompasses additional regions beyond the default network that are involved in social cognition and social interaction (Adolphs, 2009; Brothers, 1990; Frith & Frith, 2010). Other structures, such as the anterior cingulate cortex and insula, appear to be more involved in affective empathy and processes that allow us to vicariously experience and detect the personal relevance of others' emotions (Bernhardt & Singer, 2012; Jackson et al., 2006). For the purposes of this study, we focused our investigations on the default network, which has been most widely implicated in research on the neural correlates of various kinds of social cognition, and especially theory of mind (Andrews-Hanna et al., 2014; Buckner et al., 2008; Mars et al., 2012; Meyer, 2019; Schilbach et al., 2008, 2012).

Three subsystems of the default network have been identified: 1) core, 2) dorsal medial, and 3) medial temporal subsystems (Figure 1; Andrews-Hanna et al., 2014). Although integrated into a single larger functional network, these three subsystems show some degree of functional specialization, with the dorsal medial subsystem exemplifying the strongest specific associations with ToM tasks (Allen et al., 2017; Buckner et al., 2008; Spreng & Andrews-Hanna, 2015). Nonetheless, the broader default network appears to be important for social cognition and ToM. For instance, the core subsystem of the default network, although primarily active when dealing with personally relevant information and self-related processes, includes several brain regions that have been more specifically implicated in social cognitive processing, including the posterior cingulate cortex (PCC), anterior medial prefrontal cortex (amPFC), and the angular gyrus (Andrews-Hanna et al., 2014; Hyatt et al., 2015; Yeo et al., 2011). Perhaps least related to ToM and social cognitive processing, the default network's medial temporal subsystem is typically associated with autobiographical thoughts and memories, though it also has shown links to the overall default network functions of mental simulation and imagination (Spreng & Andrews-Hanna, 2015). As mentioned, the dorsal medial subsystem is most closely associated with social cognition and contains some of the regions most studied in research on the social brain (i.e., the dmPFC and TPJ), but it is worth mentioning that this subsystem also appears to play an important role in language comprehension generally (Spreng & Andrews-Hanna, 2015). In light of existing research on the role of the default network and its subsystems, we can expect that the default network broadly, but the dorsal medial subsystem in particular, will be active during social processing tasks and that individual differences in the function of these networks might underpin individual differences in social cognitive abilities.

DN-A Core

DN-B Dorsal Medial

DN-C Medial Temporal
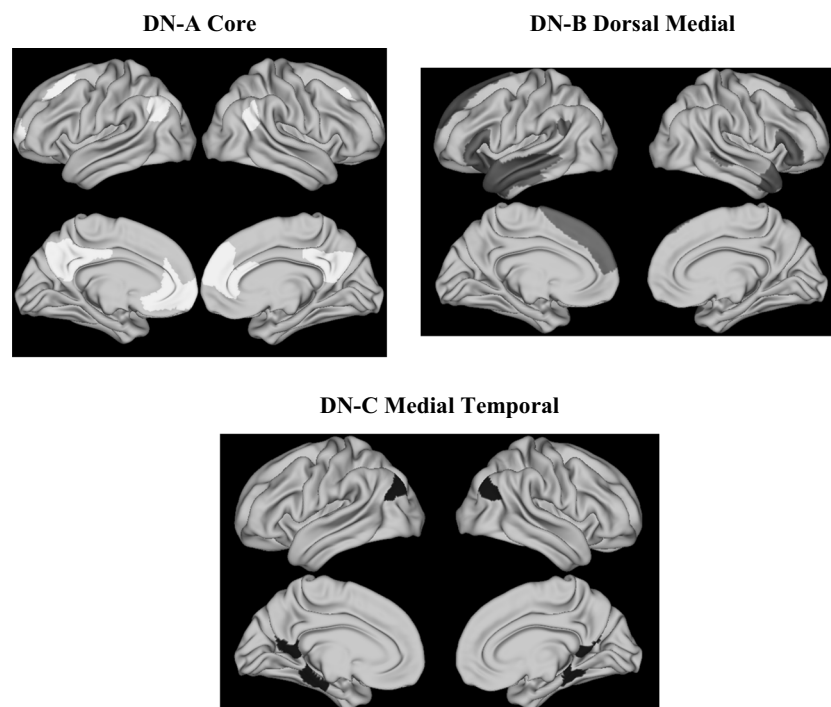
Fig. 1. Default Network Subsystems in the Human Connectome Project. *Note.* DN = Default Network. The above figure displays the three default network subsystems identified by Yeo et al. (2011).

Theory of Mind and Agreeableness

When studying individual differences, it is useful to make connections with the broad personality models that attempt to identify the major domains of psychological variation. The most thoroughly validated and widely used model of personality is the Five Factor Model or Big Five, which describes the major dimensions of covariation among human personality traits (John et al., 2008). One of the Big Five, Agreeableness, which describes traits related to altruism and cooperation, has been associated with variations in ToM ability (Allen et al., 2017; Nettle & Liddle, 2008) and thus provides a particularly useful context for understanding individual differences in social cognitive abilities and associated neural networks. People with high levels of Agreeableness tend to be described as compassionate, polite, and kind, whereas those with low levels of Agreeableness are described as disagreeable, antagonistic, and callous (Allen et al., 2017; Krueger et al., 2012; Laursen et al., 2002). Agreeableness has been shown to correlate positively with many of the same beneficial social outcomes as ToM ability (Allen et al., 2017; Ozer & Benet-Martinez, 2006), indicating the importance of further exploring the relation between ToM and Agreeableness.

Because Agreeableness is one of the traits most strongly related to individual differences in interpersonal behavior (DeYoung et al., 2013; Graziano & Eisenberg, 1997; Koole et al., 2001), better elucidation of Agreeableness and its associated cognitive mechanisms (including social cognitive processes, such as ToM) could allow us to better predict and understand variation in interpersonal behavior and

relationship functioning. This research also has the potential to contribute to theoretical models of personality. Until recently, much more emphasis has been placed on the characterization, rather than explanation, of variation in personality, and this is particularly true for Agreeableness (Nettle & Liddle, 2008). Examining ToM and social cognitive ability as one potential correlate of variation in Agreeableness, and examining the relation of both constructs to underlying variation in the default network, would contribute to neurocognitive accounts attempting to explain individual differences in Agreeableness and associated interpersonal outcomes (Allen & DeYoung, 2017; DeYoung & Weisberg, 2018; Xiao et al., 2019).

Utility of Latent Variable Modeling

To justify claims regarding the underlying associations among constructs—for example, Agreeableness, social cognition, and default network function—we must first be able to assess each of those constructs individually, in a way that is reliable and valid. Concerns of reliability and validity are especially important when using behavioral tasks, as even tasks that are able to detect robust effects at the group level (e.g., tests of implicit bias or self-regulation) often fail to produce reliable measurement of individual differences (Enkavi et al., 2019; Hedge et al., 2018; Schnabel et al., 2008). Fortunately, questionnaire measures of personality and tests of general or social cognitive ability tend to have better reliabilities than many of the measures commonly used in other areas of psychology

(Hedge et al., 2018; Morrison et al., 2019; Pinkham et al., 2018; Vellante et al., 2013). Nonetheless, we can further increase our ability to reliably measure these constructs and estimate their associations with other variables by using latent variable methods, such as structural equation modeling (SEM), which models the prediction of latent variables by other latent variables.

Latent variables represent the shared variance of multiple measured (or *manifest*) variables (Schumacker & Lomax, 2004). For example, a latent social cognitive ability variable might be modeled as the shared variance of accuracy scores across different social cognitive ability tasks. Assessing variables of interest at the latent level allows for more robust conclusions, as latent variables capture only the shared variance of their indicators, thereby eliminating unsystematic error variance and more accurately capturing variability in the underlying constructs of interest (Keith, 2006). Single-task, performance-based indicators often are limited in their scope and measure constructs narrower than those they purport to represent (Apperly, 2012; Blain, Longenecker, et al., 2020b). Performance on any given task is influenced by a number of task-specific factors, but using multi-indicator designs and latent variable frameworks allows us to move toward measuring constructs more reliably as what is shared across multiple tasks, thereby avoiding underestimation of true effect sizes (Blain, Longenecker, et al., 2020b; Campbell & Fiske, 1959; Eisenberg et al., 2019; Enkavi et al., 2019; Nosek & Smyth, 2007). Modeling social cognitive ability as the shared variance in performance across tasks should give a better representation of true variance in social cognitive ability by factoring out unique task variance (which includes a combination of task-specific variance and error).

SEM also shows promise for analyzing brain function. Such analyses can be facilitated using atlases, such as the cortical parcellation created by Schaefer et al. (2018), where each parcel (a functionally homogeneous region of the cortex) is assigned to one of the large-scale functional networks identified by Yeo et al. (2011). This local-global parcellation scheme provides an ideal opportunity for the implementation of SEM, as the activation of a given neural network can be modeled as the shared variance of activation scores for its constituent parcels. In SEM, parcels with variance more representative of the overall network receive higher weighting in the computation of a latent variable representing overall network activation. These latent variables, representing brain activity in a given network, can then be examined as predictors of various behavioral variables, such as personality or task performance. In the current research, we leverage the advantages of SEM to investigate brain-behavior associations.

Current Study

Research on the default network and Agreeableness provides a promising avenue to improve understanding of individual

differences in ToM ability. The current study used functional magnetic resonance imaging (fMRI) to investigate relations among these constructs. Specifically, we investigated whether individual differences in activity of the default network during a ToM task (Abell et al., 2000; Castelli et al., 2000) was related to individual differences in social cognitive ability and Agreeableness. We hypothesized that neural activity in the default network, and in particular its dorsal medial subsystem, would be greater when participants were engaged in ToM (social) versus nonsocial animations (*Hypothesis 1*). Furthermore, we hypothesized ToM-related activity in the default network would positively predict participants' ToM ability as indicated by accuracy on the triangles task (*Hypothesis 2a*) and by the shared variance of performance on multiple social cognitive tasks (*Hypothesis 2b*). Finally, we expected that the same ToM-related activity in the default network would be positively associated with the personality trait of Agreeableness (*Hypothesis 3*).

Method

Data and materials for the current study are available on the Human Connectome Project's website: https://www.humanconnectome.org/study/hcp-young-adult. Additionally, we have made scripts and model specifications available in an Open Science Framework repository: https://osf.io/tf5sh/?view_only=bbe63663daf6443493ab1b330bfd3f55.

Participants

The current study included 1,050 participants (564 females) from the Human Connectome Project (HCP; Van Essen et al., 2013) young adult sample. Specifically, our subsample was taken from the full 1,206 participant HCP sample and contained all participants with fMRI data for the ToM task. This sample included individuals between the ages of 22 and 37 years ($M = 28.8$, $SD = 3.7$). Exclusion criteria for the HCP included a history of severe psychiatric, neurological, or medical disorders; however, participants were not excluded on the basis of mild psychopathology (i.e., mental illness without active psychosis or mania, medication use, or treatment for a period longer than 1 year). Informed consent was obtained for all participants (consent procedure is further detailed in Van Essen et al., 2013), and all study protocols were approved by the Institutional Review Board of Washington University in St. Louis (IRB # 201204036; "Mapping the Human Connectome: Structure, Function, and Heritability"). Participants completed a large battery of self-report measures and behavioral tasks; however, only measures and tasks relevant to our current research questions (i.e., measures of the Big Five and social cognitive ability) are discussed in this paper.

Measures

NEO Five-Factor Inventory (FFI) The NEO-FFI is a measure of the Big Five personality traits: Conscientiousness, Agreeableness, Neuroticism, Openness to Experience, and Extraversion. It consists of 60 items taken from the longer NEO Personality Inventory, Revised (NEO PI-R; Costa & McCrae, 1992) and uses a five-point Likert scale ranging from 0 ("Strongly Disagree") to 4 ("Strongly Agree"). Examples of Agreeableness items included "I generally try to be thoughtful and considerate," "Most people I know like me," and "If I don't like people, I let them know it (*reversed*)." The other Big Five scales (Conscientiousness, Neuroticism, Openness, and Extraversion) were used for tests of discriminant validity.

Social cognition tasks Based on examinations of the data available in the HCP and comparisons to the existing literature, we originally identified five behavioral tasks as relevant to social cognition: a triangles ToM task, a facial emotion recognition task, a face memory condition from a working memory task, an emotional face matching task, and a moral-of-the-story identification task. After examining accuracy scores from all five of these tasks, we eventually came to focus on the first three tests due to substantial ceiling effects for the face matching and stories tasks, with most participants receiving perfect or nearly perfect scores.

Tricky Triangles Task While in a 3T fMRI scanner, participants were presented with a series of computerized animations of shapes interacting in either a random or social way (Castelli et al., 2000; Wheatley et al., 2007). Originally designed to assess ToM abilities in autism spectrum disorders (Abell et al., 2000), the task required participants to indicate whether each animation was random or social in nature after viewing each 20-second video clip. In the random condition, the shapes did not interact meaningfully with each other but rather moved around purposelessly. In the social condition, the shapes moved in ways that mimicked human behavior, including a variety of interaction types demonstrating particular social sequences such as coaxing, seducing, or mocking. Participants completed a total of 10 task blocks (2 social and 3 random condition video blocks in the first run; 3 social and 2 random condition videos in the second run). Each task block was separated by a 15-second block in which participants observed a fixation cross (with 5 fixation blocks per run). Example stimuli are shown in Figure 2. Participants were asked to identify whether each video clip was random or social in nature and performance was scored for accuracy (i.e., whether participants correctly classified animations as random or social). Scores on the ToM triangles were negatively skewed, and a log transformation (in which scores were reversed before and after transformation to maintain scoring direction) was used to increase their normality.

Penn Emotion Recognition Task (ER40) The emotion recognition task was adapted from Gur et al. (2001). In this task, participants were presented with a series of 40 faces and were asked to identify what emotion each face expressed. Emotion options included "Happy," "Sad," "Angry," "Scared," and "No Feeling." Eight faces were presented for each emotion, half of which were male and the other half female. See Figure 2 for example stimuli and answer choices. Participants' accuracy and reaction times were recorded.

Two-back Task In the two-back task, participants were presented with a series of stimuli from four categories: body parts, faces, places, and tools (Barch et al., 2013). See Figure 2 for example stimuli. In each of the conditions, participants were shown a series of objects and tasked to indicate by pressing a button whenever an object (i.e., face, body part, place, or tool) was presented that had been presented two trials previously. Each block consisted only of one stimulus type. Participants completed a total of 16 blocks (two runs of the two-back for each of the four stimulus types). Each block consisted of 10 trials, lasting 2.5 seconds each. Only the face condition was included in assessing social cognitive ability due to its social relevance compared to the other conditions. Participants were scored for accuracy, separately for each of the object conditions.

Task fMRI Data Acquisition and Processing

Data were obtained that had undergone preprocessing and preliminary analysis by researchers at the HCP. Specifically, we used results of the HCP's level-two, individual-subject, cortical-vertex-based analyses based on fMRI data acquired while participants completed random and social conditions of the tricky triangles task described above (Abell et al., 2000; Castelli et al., 2000). Specifics of the fMRI data acquisition are detailed in previous publications about the HCP (Uğurbil et al., 2013). In summary, whole-brain echo planar imaging acquisitions were acquired with a 32-channel head coil on a modified 3T Siemens Skyra used for all HCP data collection at Washington University in St. Louis (TR = 720 ms, TE = 33.1 ms, flip angle = 52 deg, BW = 2,290 Hz/Px, in-plane FOV = 208 × 180 mm, 72 axial slices, 2.0-mm isotropic voxels, with a multiband acceleration factor of 8). One run of the ToM task used right-to-left phase encoding and the other utilized a left-to-right phase encoding.

Data analysis pipelines for the HCP were primarily built using tools adapted from FreeSurfer and FSL. The first step in processing included application of the HCP "fMRIVolume" pipeline. This process generates "minimally preprocessed" 4D time series data for each run and participant, and the pipeline steps include gradient unwarping, FLIRT-based motion correction, TOPUP-based field map preprocessing using a spin echo field map, brain-boundary-based registration of EPI to

Triangles Task   Penn Emotion Recognition (ER40)   Two-back Conditions
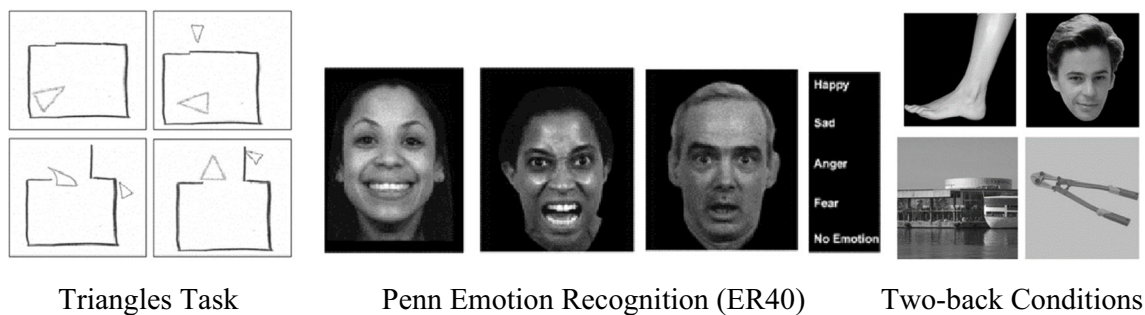
Fig. 2. Social cognition tasks. *Note.* The above figure shows images taken from the three social cognition tasks used in the study.

structural T1-weighted scan, nonlinear (FNIRT) registration into MNI152 space, and grand-mean intensity normalization. The data were then transformed into grayordinate space, which allows for more efficient analysis of brain activation levels for components of the cortical surface. In this process, data from the cortical gray matter ribbon are projected onto the surface and then onto registered surface meshes with a standard number of vertices (in this case, approximately 30,000). Smoothing of the left and right hemisphere time series and autocorrelation estimates (from FILM) were done on the surface using a geodesic Gaussian algorithm.

Activity estimates were computed for the preprocessed functional time series from each run using a general linear model (GLM) implemented in FSL's FILM (FMRIB's Improved Linear Model) with autocorrelation correction. Predictors were convolved with a double gamma "canonical" hemodynamic response function to generate the main model regressors. To facilitate analyses of individual differences in response to given stimuli, GLM predictors were based on the category of each video clip rather than the rating of the individual (i.e., conditions were based on the appropriate response rather than each participants' actual response, and accuracy was not considered in these GLMs). Each predictor covered the duration of a single video clip (20 s) and did not include time during fixation cross-viewing. To compensate for slice-timing differences and variability in the HRF delay across regions, temporal derivative terms derived from each predictor were added to each GLM and were treated as confounds of no interest. Subsequently, both the 4D time series and the GLM design were temporally filtered with a Gaussian-weighted linear high-pass filter with a (soft) cutoff of 200 s.

Fixed-effects analyses were conducted using FEAT to estimate the average effects across runs within-participants. Cross-run statistical comparisons occurred in standard grayordinates space rather than volume space. As in the individual analysis, NIFTI-1 matrices were processed separately for left and right surface and subcortical volume data, and surface outputs were converted to GIFTI at the conclusion of analysis. Participant-level z-statistic maps (computed as z-transformed t-statistics) were combined from left and right hemisphere cortical and subcortical gray matter into the

recently introduced CIFTI data format, with individual z-statistics for each condition output for each cortical vertex.

Group Prior Individualized Parcellation

Network activation was identified using group prior individualized parcellation (GPIP), an approach that begins with a standard atlas of parcels for all participants but adjusts the boundaries of each parcel for each individual to correspond to their unique cortical organization. This is an effective solution to the problem that cortical functional organization is not identically related to anatomical landmarks in each person. For each participant, fMRI BOLD time-series acquired during the tricky triangles task in subject native space were resampled to the fsaverage5 cortical surface mesh (Dale et al., 1999) and normalized at each vertex. The resulting subject surface data were mapped onto a pre-defined group atlas with 400 functionally distinct regions (Schaefer et al., 2018) that align well with the 17-network atlas defined by Yeo et al. (2011). An iterative algorithm utilizing two Bayesian priors was applied to model connectivity between parcels and adjust parcel boundaries (Chong et al., 2017). Through this process, parcel boundaries were modified to reflect each participant's unique patterns of functional connectivity. This method permits a more accurate approximation of individuals' unique functional topography during the social cognition task while maintaining correspondence of parcels across all participants and with the atlas. Previous research evaluating GPIP has demonstrated that individualized parcels exhibit greater network coherence and better segregation of task activation compared to the parcel locations from the initial group atlas (Chong et al., 2017), and a growing body of research has reported robust associations of the parameters of individualized parcels with a variety of measures of individual differences (Anderson et al., 2021; Kong et al., 2019; Mwilambwe-Tshilobo et al., 2019).

For each participant, individualized parcels were resampled to grayordinate space to permit comparisons between parcel assignment and task activation values for each vertex. Following output of z-statistics for each cortical vertex for the social and random conditions for the ToM task (the processed data obtained from HCP's database) and generation of

individualized parcellation mappings using GPIP, we mapped the vertex-wise individual participant data onto each participants' individualized parcels. We then computed parcel activation variables for each condition, for each cortical parcel associated with the default network, by averaging the z-statistics for vertices associated with each cortical parcel. Our parcellation activation variables were sorted by default network subsystem and included 34 parcels associated with the core subsystem, 32 parcels associated with the dorsal medial subsystem, and 13 parcels associated with the medial temporal subsystem. We then reduced the number of variables for each of these subsystems, creating composite activation variables for cortical parcels that were anatomically adjacent. This left us with a total of 9 parcels (per condition) for the core subsystem, 9 for the dorsal medial subsystem, and 6 for the medial temporal subsystem. These parcels were used for subsequent analyses.

Statistical Analysis

Structural equation models (SEMs) were used to examine whether variation in brain activation during the social vs. random condition, for each subnetwork of the default network, predicted social cognitive ability and Agreeableness. Separate social- and random-activation latent variables were derived for each of the three subsystems, using all the corresponding cortical parcels for each subsystem as indicators. The latent variables produced by this procedure represent the shared variance among their indicators and thus can be interpreted as reflecting variation in the tendency toward activation, in a given condition, for each subnetwork as a whole. The core subsystem latent variables had a total of nine indicators: right temporal, right IPL, right PCC/Precuneus, right dPFC, right mPFC, left IPL, left PCC/Precuneus, left dPFC, and left mPFC. The dorsal medial subsystem latent variables had a total of nine indicators: right temporal, right anterior-temporal, right dPFC, right vPFC, left temporal, left IPL, left dPFC, left vPFC, and left lPFC. The medial temporal subsystem latent variables had a total of six indicators: right IPL, right parahippocampal cortex, right retrosplenial cortex, left IPL, left parahippocampal cortex, and left retrosplenial cortex.

Using the same approach to parcellation, we also created latent variables representing activation in the frontoparietal control network (FPCN), as identified by Yeo et al. (2011), to test for discriminant validity. The FPCN makes a good contrast to the default network because it also is involved in complex cognitive processes, such as working memory and intelligence (Santarnecchi et al., 2017), but has not been strongly linked to social cognition. These frontoparietal variables were indicated by parcels located in the right PCC, right PFC, right temporal, right parietal, left PCC, left PFC, left orbitofrontal cortex, left temporal, and left parietal activation. (In Schaefer et al.'s parcellation scheme, each parcel is

assigned to only one network of Yeo et al., so there was no overlap between indicators for FPCN and default network.)

In all of our models, residuals from anatomically identical manifest variables were allowed to correlate (e.g., the random and social manifest variables for right PFC activation in the dorsal medial models). Maximum likelihood estimation was used and common fit indices were computed, including the chi square, Tucker Lewis index, and root mean squared error of approximation (RMSEA). The Latent Variable Analysis (LAVAAN) package for R was used for estimating all models (Rosseel, 2012).

Based on these neural activation measurement models, we first tested for latent mean differences in activation for each of the latent variables representing the default network's three primary subsystems during the social vs. random animation conditions. Subsequently, additional SEMs were used to assess the relations among latent variables representing Agreeableness, social cognitive ability, and neural activation in the three default network subsystems and the FPCN, during the social and random animation conditions. More specifically, SEMs were conducted to test the effects of default network activation during the social animation condition on 1) accuracy on the triangles task, 2) latent social cognitive ability, representing accuracy on a variety of social cognitive tasks, and 3) Agreeableness. Separate models were computed for the core, dorsal medial, and medial temporal subsystems (and for subnetworks of the FPCN) because of a high degree of multicollinearity among the neural variables, leading to failures of model convergence if all networks were included at once. Post-hoc analyses were conducted separately for temporal and prefrontal components of the default network's dorsal medial subnetwork, given the particular relevance of these regions to social cognition and related personality traits and the fact that this network was shown to be significantly more active during social vs. random animations. Intelligence, sex (coded as 0 = female, 1 = male), age, and neural activation during the random condition (in the appropriate subnetwork) were included as covariates in all models.

In our analyses, including activation in the random condition as a covariate replaces using a contrast of the two conditions. Despite the ubiquity of contrast scores (differences in activation between two conditions) as variables of interest in neuroimaging research, this approach suffers from many of the problems that have been noted regarding the use of difference scores instead of including both conditions of interest in analyses (Allison, 1990; Edwards, 1994; Edwards, 1996; Wittenborn, 1951). When using a difference score, as is the case in a traditional fMRI contrast, variation in the effect of interest can either be due to the control condition (e.g., random animations) or the condition of interest (e.g., social animations). Difference scores do not capture any information about the association between scores on the two conditions of interest, instead imposing a linear restriction on their slopes when

predicting outcome variables of interest (Allison, 1990; Edwards, 1994; Edwards, 1996; Wittenborn, 1951). Thus, if we were to use difference scores, we would not be able to identify the specific influence of social activation (vs. random activation) on our behavioral variables of interest. Including activation for both conditions as predictors allows us to partial out variance in activation that is shared between the conditions, such that we see the effect of our condition of interest after controlling for the baseline provided by the random condition. The effect of interest therefore indicates how much each subject's activation deviates in the condition of interest from the activation that would be expected based on the control condition. Thus, we can accurately estimate the unique associations of activation during the social condition (as well as activation during the random condition) with our behavioral variables of interest.

Latent social cognitive ability was modeled using accuracy variables for the triangles task (correct vs. incorrect responses for the random and social animations), ER40, and the face memory condition from the two-back task. In addition to this latent social cognitive ability variable, we also conducted a test using accuracy on only the ToM triangles task as manifest criterion variable. This manifest outcome variable test was included because our neural variables measured activation specifically during this task. For all relevant models, intelligence was modeled using tests of Picture Vocabulary, Matrix Reasoning, and English Reading, as well as a hierarchical working memory factor using the four two-back task conditions, onto which the face memory variable was allowed to cross load.

Results

Descriptive statistics for self-report measures and task performance are presented in Table 1. Pearson correlations among all variables are presented in Tables S1, S2, and S3. Weak positive zero-order correlations were observed between social cognition accuracy measures and NEO Agreeableness. Social cognition measures were also positively correlated with intelligence measures, with stronger magnitudes.

Hypothesis 1

SEM was used to test for mean differences in default network activation for the social vs. random conditions of the triangles task. Fit statistics for all SEMs are presented in Table 2. Results are visualized in Figure 3. For these models, our dependent variables are calculated as weighted averages of activation across the regions of a given subnetwork and can be interpreted as reflecting activation (for each given condition) in that subnetwork as a whole. Activation of the default network's dorsal medial subsystem was significantly greater

during the social condition vs. the random condition (Figure 3; $z = 5.70$, $p < 0.001$); this pattern of activation held both for regions centered on the temporoparietal junction and temporal pole (Figure 3; $z = 6.26$, $p < 0.001$) and for regions in the prefrontal cortex (Figure 3; $z = 14.37$, $p < 0.001$). Activation was significantly less for the social condition vs. the random condition in the medial temporal subsystem (Figure 3; $z = -5.68$, $p < 0.001$) and core subsystem (Figure 3; $z = -9.42$, $p = 0.089$).

Hypotheses 2a, 2b, and 3

SEM also was used to test for associations of default network activation during the social condition of the ToM task with social cognitive ability and Agreeableness (Table 3). Activation of the dorsal medial, core, and medial temporal subsystems of the default network were positively associated with accuracy on the triangles task, controlling for age, sex, intelligence, and neural activation in the random condition (Table 3). Similarly, activation of all three default network subsystems was positively associated with shared variance in accuracy across various tests of social cognitive ability.[1] Across models, intelligence was a significant positive predictor of social cognitive ability. Furthermore, sex was a significant predictor of social cognitive ability, with females performing better on social cognition tasks on average. A full measurement and structural model is presented for the prefrontal component of the dorsal medial subsystem in Figure 4.

Activation during the social condition significantly predicted Agreeableness for the medial temporal subsystem and for the prefrontal cortex component of the dorsal medial subsystem (Table 3). As with our social cognitive ability variable, sex was a significant predictor of Agreeableness; females had higher levels of Agreeableness. A full measurement and structural model is presented for the prefrontal cortex component of the dorsal medial subsystem in Figure 5.

A final set of analyses was conducted to examine discriminant validity. We found that activation in FPCN subnetworks during the social animations condition also significantly predicted accuracy on the triangles task and latent social cognitive ability, but not Agreeableness. Effects were generally, but not always, weaker compared to effects observed in the default network models (Table 4). Activation in the default network during the social condition blocks did not predict personality traits other than Agreeableness (i.e., Conscientiousness, Neuroticism, Extraversion, or Openness). The full results of these analyses are displayed in Table 4.

---

[1] At the request of a reviewer, we reran models without including the face condition of the two-back task, to ensure that including this variable and its cross-loading from the social cognition latent variable was not unduly influencing results. Findings were substantively equivalent across models when this variable was excluded.

Table 1. Descriptive statistics for self-report and task measures

	Mean (SD)	Skew	[Minimum, Maximum]
NEO Agreeableness	33.3 (5.8)	-0.3	[10, 48]
Triangles task accuracy	80.7 (12.1)	-1.8	[5, 100]
Emotion recognition accuracy	35.5 (2.6)	-1.0	[24, 40]
Two-back overall accuracy	89.3 (10.6)	-1.3	[25, 100]

Discussion

The current study used a large fMRI sample, multiple behavioral tasks, and SEM to investigate how ToM-related activity in the default network and its subsystems was related to social cognitive ability and Agreeableness. Findings largely confirmed our three main hypotheses. Neural activity in the dorsal medial subsystem of the default network was significantly greater during the viewing of social animations compared to random animations (*Hypothesis 1*). Activity in the dorsal medial subsystem—while participants viewed the social animations—positively predicted performance on the triangles ToM task (*Hypothesis 2a*). This was true with or without controlling for covariates, such as intelligence, suggesting that the association is robust. This positive association also was found for the default network's core and medial temporal subsystems, as well as components of the frontoparietal control network. Likewise, neural activity in these regions, during the social condition of the task, positively predicted social cognitive ability, modeled using a latent variable indicated by accuracy scores on three different social cognition tasks (*Hypothesis 2b*). Finally, neural activity during the social animations positively predicted individual differences in the personality trait Agreeableness, for prefrontal regions of the dorsal medial subsystem and for the medial temporal subsystem (*Hypothesis 3*). Associations with Agreeableness were not seen for another neural network involved in complex cognition (the FPCN), nor were other personality traits associated with default network activation, suggesting specificity for the associations among default network activation and Agreeableness.

Collectively, our results suggest that individual differences in both Agreeableness and ToM are related to variation in the same underlying neural network. Our findings reinforce previous research tying the default network—and more specifically its dorsal medial subsystem—to ToM and the ability to understand the mental states and emotions of others (Allen et al., 2017; Amodio & Frith, 2006; Carrington & Bailey, 2009; Frith & Frith, 2006; Sabbagh et al., 2004; Saxe & Kanwisher, 2003; Saxe & Powell, 2006; Saxe & Wexler, 2005; Schurz et al., 2014; Spreng & Andrews-Hanna, 2015; Spunt & Lieberman, 2012; Vogeley et al., 2001; Young et al., 2010). The present study extends previous findings by demonstrating that activity in the dorsal medial subsystem not only predicted performance on a single task, but also performance on a variety of social cognitive tasks modeled as a latent variable, thereby providing evidence for a positive association between individual differences in broad social cognitive ability and default network function. This relationship could provide insight into how both cognitive and neural variation contribute to individual differences in social functioning.

Our study demonstrated that ToM-related brain activity in prefrontal regions of the dorsal medial subsystem positively predicted individual differences in Agreeableness, a personality trait linked to social cognition and especially relevant for understanding social interactions. Previous research evaluating the association between personality and social functioning has linked both Extraversion and Agreeableness with interpersonal tendencies and trait affiliation (Côté & Moskowitz, 1998; DeYoung et al., 2013; DeYoung & Weisberg, 2018). Each of the Big Five traits can be thought of as relating to particular motivational, cognitive, and affective mechanisms (DeYoung, 2015; DeYoung & Blain, 2020). For example, pattern detection and curiosity for Openness-Intellect(Bainbridge et al., 2019; Blain, Longenecker, et al., 2020b; DeYoung et al., 2012; Silvia & Christensen, 2020) and reward sensitivity for Extraversion (Blain, Sassenberg, et al., 2020c; Lucas et al., 2000; Smillie et al., 2012). Agreeableness appears to reflect tendencies related to navigating social norms and coordinating with the needs of others (DeYoung, 2015; DeYoung & Weisberg, 2018; Koole et al., 2001). Agreeableness in particular has been associated with prosociality (Caprara et al., 2010; Habashi et al., 2016), higher levels of satisfaction in relationships (Malouff et al., 2010; Weidmann et al., 2017), and less prejudicial behavior towards others (Sibley & Duckitt, 2008). Although Agreeableness has been positively associated with many desirable social outcomes, its underlying mechanisms remain understudied among the Big Five personality traits, with few studies having investigated its neurocognitive correlates (DeYoung & Blain, 2020). Social cognitive ability and default network function, however, appear to be promising candidates for understanding the substrates of individual differences in Agreeableness (Allen et al., 2017; Arbula et al., 2021).

**Table 2.** Fit statistics for all structural equation models

Models	df	χ^2	p	TLI	RMSEA	95% CI
Primary Models						
DN Overall — Social vs. Random	6	405.8	<0.001	0.911	0.252	[0.231, 0.273]
DN Core — Social vs. Random	126	1397.8	<0.001	0.913	0.098	[0.093, 0.103]
DN Dorsal Medial — Social vs. Random	96	1665.5	<0.001	0.876	0.125	[0.120, 0.130]
DN Dorsal Medial (Temporal)— Social vs. Random	6	23.5	0.001	0.990	0.053	[0.031, 0.076]
DN Dorsal Medial (Prefrontal)— Social vs. Random	30	818.8	<0.001	0.854	0.158	[0.149, 0.168]
DN Medial Temporal — Social vs. Random	48	704.1	<0.001	0.901	0.114	[0.107, 0.122]
DN Overall — Triangles Accuracy	88	363.1	<0.001	0.973	0.055	[0.049, 0.061]
DN Core — Triangles Accuracy	328	1899.5	<0.001	0.911	0.068	[0.065, 0.071]
DN Dorsal Medial — Triangles Accuracy	278	2087.0	<0.001	0.885	0.079	[0.076, 0.083]
DN Dorsal Medial (Temporal)— Triangles Accuracy	88	347.1	<0.001	0.948	0.053	[0.048, 0.059]
DN Dorsal Medial (Prefrontal)— Triangles Accuracy	152	1071.2	<0.001	0.893	0.077	[0.072, 0.081]
DN Medial Temporal — Triangles Accuracy	190	1117.8	<0.001	0.904	0.069	[0.065, 0.073]
DN Overall — Social Cognition Accuracy	101	361.6	<0.001	0.974	0.050	[0.044, 0.056]
DN Core — Social Cognition Accuracy	353	1913.1	<0.001	0.912	0.065	[0.063, 0.068]
DN Dorsal Medial — Social Cognition Accuracy	301	2098.8	<0.001	0.886	0.076	[0.073, 0.079]
DN Dorsal Medial (Temporal)— Social Cognition Accuracy	101	343.1	<0.001	0.953	0.048	[0.043, 0.054]
DN Dorsal Medial (Prefrontal)— Social Cognition Accuracy	169	1075.9	<0.001	0.896	0.072	[0.068, 0.076]
DN Medial Temporal — Social Cognition Accuracy	209	1131.3	<0.001	0.906	0.065	[0.062, 0.069]
DN Overall — Agreeableness	159	597.4	<0.001	0.961	0.051	[0.047, 0.056]
DN Core — Agreeableness	447	2161.9	<0.001	0.905	0.061	[0.058, 0.063]
DN Dorsal Medial — Agreeableness	389	2344.9	<0.001	0.879	0.069	[0.067, 0.072]
DN Dorsal Medial (Temporal)— Agreeableness	159	565.3	<0.001	0.924	0.049	[0.045, 0.054]
DN Dorsal Medial (Prefrontal)— Agreeableness	239	1292.5	<0.001	0.882	0.065	[0.061, 0.068]
DN Medial Temporal — Agreeableness	285	1271.4	<0.001	0.901	0.057	[0.054, 0.061]
Discriminant Validity Models						
FPCN — Triangles Accuracy	88	336.7	<0.001	0.973	0.052	[0.046, 0.058]
FPCN — Social Cognition Accuracy	101	334.6	<0.001	0.975	0.047	[0.042, 0.053]
FPCN — Agreeableness	159	568.8	<0.001	0.960	0.050	[0.045, 0.054]
FPCN A — Triangles Accuracy	382	2835.2	<0.001	0.863	0.079	[0.076, 0.082]
FPCN A — Social Cognition Accuracy	409	2860.3	<0.001	0.864	0.076	[0.074, 0.079]
FPCN A — Agreeableness	509	3065.2	<0.001	0.859	0.069	[0.067, 0.072]
FPCN B — Triangles Accuracy	382	2116.5	<0.001	0.911	0.066	[0.064, 0.069]
FPCN B — Social Cognition Accuracy	409	2133.3	<0.001	0.912	0.064	[0.061, 0.067]
FPCN B — Agreeableness	509	2383.9	<0.001	0.905	0.059	[0.057, 0.062]
FPCN C — Triangles Accuracy	118	419.5	<0.001	0.961	0.050	[0.045, 0.055]
FPCN C — Social Cognition Accuracy	133	420.1	<0.001	0.964	0.046	[0.041, 0.051]
FPCN C — Agreeableness	197	651.4	<0.001	0.945	0.047	[0.043, 0.051]
DN Dorsal Medial Prefrontal — Conscientiousness	239	1619.8	<0.001	0.861	0.074	[0.071, 0.078]
DN Dorsal Medial Prefrontal — Neuroticism	239	1344.8	<0.001	0.891	0.066	[0.063, 0.070]
DN Dorsal Medial Prefrontal — Openness	239	1745.8	<0.001	0.840	0.078	[0.074, 0.081]
DN Dorsal Medial Prefrontal — Extraversion	239	1628.1	<0.001	0.852	0.075	[0.071, 0.078]

Note. DN = default network, *FPCN* = frontoparietal control network

Synthesizing Current Findings and Previous Work

Our findings were consistent with previous work, demonstrating positive associations of default network function with social cognitive ability and Agreeableness. For instance, previous studies with high statistical power have demonstrated positive associations of resting state functional connectivity within the default network and ToM ability and questionnaire
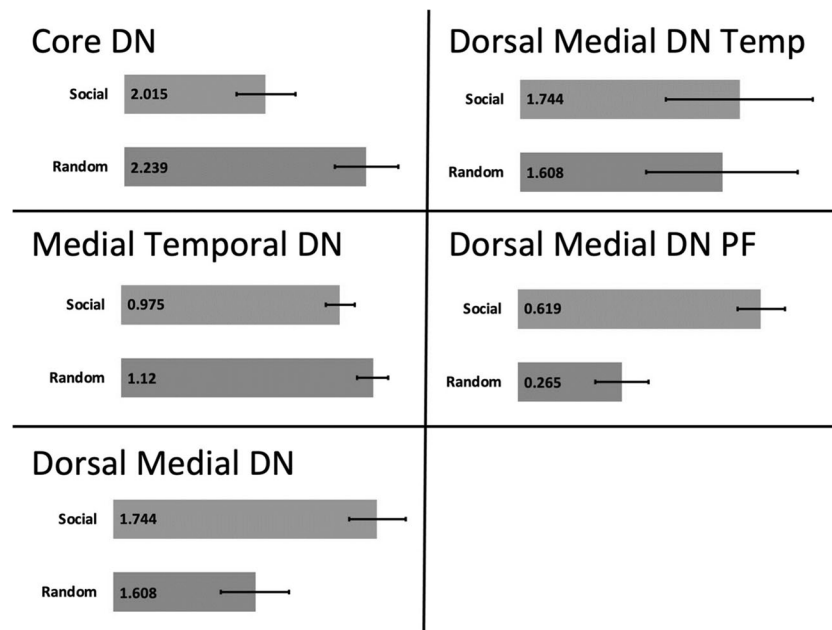
Fig. 3. Social vs. random brain activation during the triangles task. *Note.* DN = default network. Figure shows standardized latent means for activation in DN subsystems and the FPCN for the social vs. random conditions of the triangles task. Our dependent variables were calculated (via structural equation modeling) as weighted averages of activation across the regions of a given subnetwork and can be interpreted as reflecting activation in that subnetwork as a whole (during each condition).

measures of trait empathy and compassion (Allen et al., 2017; Takeuchi et al., 2014). The current study extends this work to look at brain activity during a ToM task rather than just during rest, and suggests that ToM-related brain activity in the dorsal medial subsystem of the default network may be associated with both Agreeableness and ToM abilities.

Together with the previous work, our findings suggest a possible explanation for why people high in Agreeableness tend to demonstrate better interpersonal outcomes than less agreeable people; highly agreeable people may have better social abilities because of differences in the function of specific brain networks, including the default network and particularly its dorsal medial subsystem. Our study also accounts for possible alternative explanations by including covariates, such as intelligence, sex, and age. Although some covariates also were related to our variables of interest, nonetheless controlling for them did not eliminate the hypothesized effects.

In light of the current results, which suggest significant associations of social cognitive ability and Agreeableness with default network activation, it is worth mentioning contrasting findings from a recent study utilizing the same HCP dataset (Weiss et al., 2021). Weiss et al. found no meaningful relations between personality variables and neural activity during the same triangles task data we analyzed in the current study. The authors attribute the lack of significant associations to methodological issues such as the questionable validity of the social cognition task and test–retest reliability of functional biomarkers. Our utilization of better methods, such as the individualized parcellation approach of GPIP and latent variable modeling for behavioral and neural variables, can increase reliability and thus the ability to detect true associations among variables (Blain, Longenecker, et al., 2020b; Campbell & Fiske, 1959; Chong et al., 2017; Eisenberg et al., 2019; Enkavi et al., 2019; Keith, 2006; Kong et al., 2021; Nosek & Smyth, 2007). This increased reliability of our variables of interest is a likely explanation for why we were able to detect significant associations among social cognition, personality, and default network activation in the current work, in contrast to the null effects observed by Weiss et al. (2021).

In the current study, although all default network subsystems showed robust relations to individual differences in social cognitive ability, only the dorsal medial subsystem was significantly more active during the social condition of the ToM task compared with the random condition, and only activity in the medial temporal subsystem and prefrontal regions of the dorsal medial subsystem significantly predicted individual differences in Agreeableness. Though we should avoid overinterpreting this specificity of the dorsal medial subsystem and its prefrontal regions, as effect sizes were fairly similar in magnitude across the subsystems, the current pattern of results is in line with research suggesting the dorsal medial subsystem may be more strongly linked to social cognition than the other two default network subsystems. The dorsal medial subsystem also appears to have broader functions in language processing, which can be argued to be inherently social (Spreng & Andrews-Hanna, 2015).

**Table 3.** Results for default network activation models

Models	z	β	p
Triangles Accuracy			
DN Overall Social Activation	3.11	0.13	0.002
DN Overall Random Activation	-1.59	-0.07	0.113
DN Core Social Activation	2.69	0.11	0.007
DN Core Random Activation	-1.30	-0.05	0.193
DN Dorsal Medial Social Activation	2.50	0.10	0.013
DN Dorsal Medial Random Activation	-1.38	-0.06	0.168
DN Dorsal Medial (Temporal) Social Activation	2.22	0.09	0.026
DN Dorsal Medial (Temporal) Random Activation	-0.32	-0.01	0.751
DN Dorsal Medial (Prefrontal) Social Activation	2.52	0.10	0.012
DN Dorsal Medial (Prefrontal) Random Activation	-1.74	-0.07	0.083
DN Medial Temporal Social Activation	2.88	0.12	0.004
DN Medial Temporal Random Activation	-1.75	-0.07	0.080
Social Cognition Accuracy			
DN Overall Social Activation	2.83	0.24	0.005
DN Overall Random Activation	-0.39	-0.03	0.700
DN Core Social Activation	2.58	0.22	0.010
DN Core Random Activation	-0.26	-0.02	0.792
DN Dorsal Medial Social Activation	2.28	0.19	0.023
DN Dorsal Medial Random Activation	-0.14	-0.01	0.889
DN Dorsal Medial (Temporal) Social Activation	2.39	0.20	0.017
DN Dorsal Medial (Temporal) Random Activation	0.41	0.03	0.679
DN Dorsal Medial (Prefrontal) Social Activation	2.64	0.22	0.008
DN Dorsal Medial (Prefrontal) Random Activation	-0.80	-0.07	0.423
DN Medial Temporal Social Activation	2.62	0.22	0.009
DN Medial Temporal Random Activation	-1.05	-0.08	0.296
Agreeableness			
DN Overall Social Activation	1.38	0.06	0.168
DN Overall Random Activation	0.29	0.01	0.769
DN Core Social Activation	1.42	0.07	0.157
DN Core Random Activation	0.19	0.01	0.849
DN Dorsal Medial Social Activation	1.54	0.07	0.125
DN Dorsal Medial Random Activation	-0.76	0.00	0.939
DN Dorsal Medial (Temporal) Social Activation	1.28	0.06	0.201
DN Dorsal Medial (Temporal) Random Activation	-0.08	0.00	0.935
DN Dorsal Medial (Prefrontal) Social Activation	1.99	0.09	0.046
DN Dorsal Medial (Prefrontal) Random Activation	-0.50	-0.02	0.621
DN Medial Temporal Social Activation	1.98	0.09	0.047
DN Medial Temporal Random Activation	-0.08	0.00	0.940

Note. DN = default network.

The core subsystem includes regions of the brain associated with social cognitive functions (Leech & Sharp, 2014; Spreng et al., 2009; Spreng & Andrews-Hanna, 2015) but also other cognitive functions that are not as specifically social in nature, such as the retrieval of autobiographical memory and personal knowledge (Moran et al., 2013; Spreng et al., 2009;

Spreng & Andrews-Hanna, 2015). Similarly, the medial temporal subsystem is involved in more non-social cognitive functions, again including episodic memory (Buckner et al., 2008; Spreng & Andrews-Hanna, 2015). Future work should more specifically investigate the coordination of these three subsystems, as the joint activation of these subsystems and
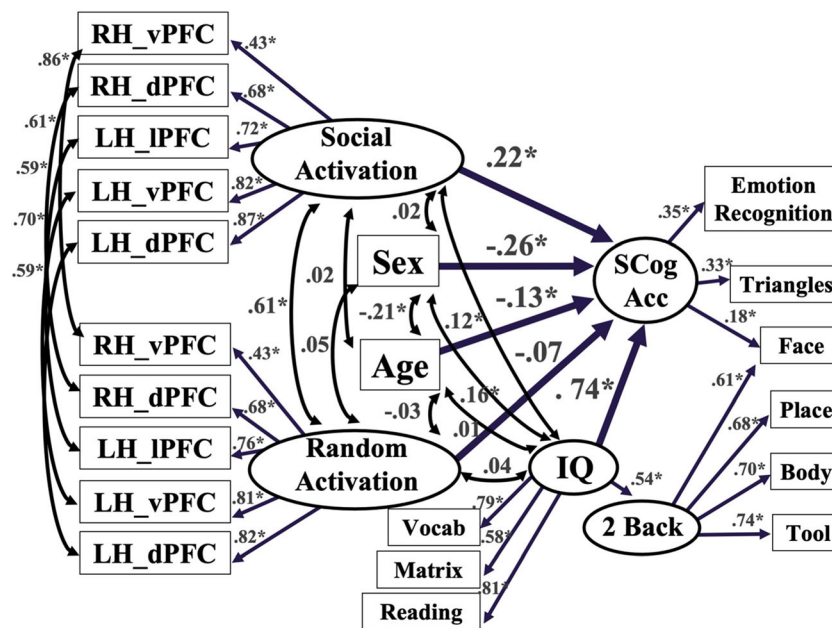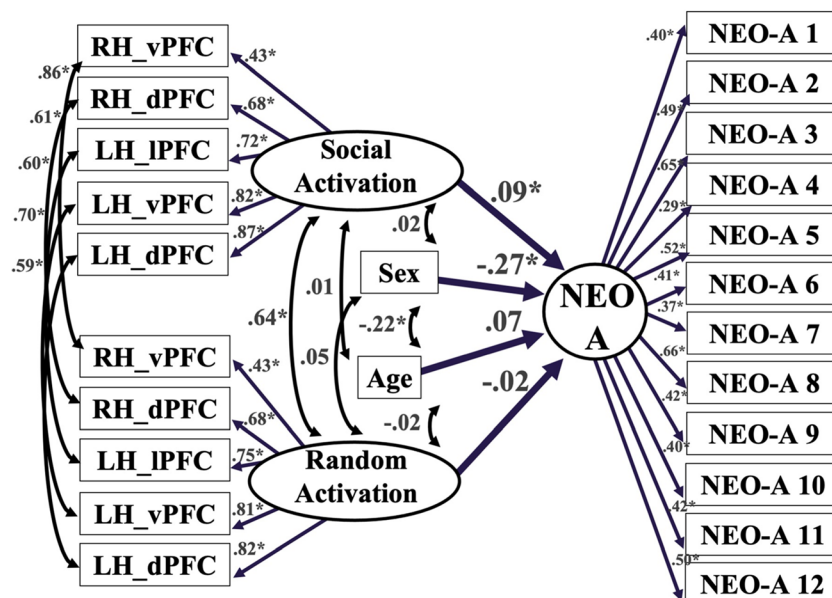
Fig. 4. Relation between dorsal medial prefrontal activation and social cognitive ability. *Note.* RH = right hemisphere, LH = left hemisphere, PFC = prefrontal cortex, v = ventral, d = dorsal, l = lateral, SCog Acc = accuracy on the social cognition tasks. Structural equation modeling was used to test the association of dorsal medial default network activation during the social condition of the theory of mind task with social cognitive ability. Age, sex, intelligence, and neural activation in the random condition were included as covariates. Activation of the default network's dorsal medial subsystem (modeled as the shared variance among activation scores in spatially contiguous regions) was positively associated with shared variance in accuracy across various tests of social cognitive ability.

functional connectivity between the systems appears to be particularly relevant to social cognition and corresponding individual differences (Allen et al., 2017; Spreng & Andrews-Hanna, 2015; Takeuchi et al., 2014).

Role of General Cognitive Ability and Sex Differences

Social cognitive ability was significantly predicted by general cognitive ability and sex, as well as by Agreeableness and

Fig. 5. Relation between dorsal medial prefrontal activation and Agreeableness. *Note.* RH = right hemisphere, LH = left hemisphere, PFC = prefrontal cortex, v = ventral, d = dorsal, l = lateral, NEO A = Agreeableness. Structural equation modeling was used to test the association of dorsal medial default network activation during the social condition of the theory of mind task with Agreeableness. Age, sex, and neural activation in the random condition were included as covariates. Activation of the default network's dorsal medial subsystem (modeled as the shared variance among activation scores in spatially contiguous regions) was positively associated with shared variance in Agreeableness items from the NEO Five Factor Inventory.

Table 4. Results for discriminant validity models

Models	Z	β	P
Triangles Accuracy			
FPCN Social Activation	2.63	0.11	0.009
FPCN Random Activation	-1.48	-0.06	0.140
Social Cognition Accuracy			
FPCN Social Activation	2.38	0.20	0.017
FPCN Random Activation	-0.29	-0.02	0.770
Agreeableness			
FPCN Social Activation	1.38	0.07	0.168
FPCN Random Activation	0.22	0.01	0.827
Triangles Accuracy			
FPCN A Social Activation	2.74	0.12	0.006
FPCN A Random Activation	-1.69	-0.07	0.090
Social Cognition Accuracy			
FPCN A Social Activation	2.31	0.20	0.021
FPCN A Random Activation	-0.60	-0.05	0.550
Agreeableness			
FPCN A Social Activation	1.29	0.06	0.198
FPCN A Random Activation	0.49	0.02	0.623
Triangles Accuracy			
FPCN B Social Activation	2.78	0.12	0.005
FPCN B Random Activation	-1.52	-0.06	0.129
Social Cognition Accuracy			
FPCN B Social Activation	2.35	0.20	0.019
FPCN B Random Activation	-0.23	-0.02	0.821
Agreeableness			
FPCN B Social Activation	0.99	0.05	0.323
FPCN B Random Activation	0.38	0.02	0.705
Triangles Accuracy			
FPCN C Social Activation	2.33	0.11	0.020
FPCN C Random Activation	-1.40	-0.06	0.163
Social Cognition Accuracy			
FPCN C Social Activation	1.90	0.17	0.057
FPCN C Random Activation	-0.08	-0.01	0.939
Agreeableness			
FPCN C Social Activation	1.11	0.06	0.265
FPCN C Random Activation	0.22	0.01	0.828
Conscientiousness			
DN Dorsal Medial - Prefrontal Social Activation	0.07	0.00	0.942
DN Dorsal Medial - Prefrontal Random Activation	-0.13	-0.01	0.897
Neuroticism			
DN Dorsal Medial - Prefrontal Social Activation	-0.90	-0.04	0.367
DN Dorsal Medial - Prefrontal Random Activation	-0.47	-0.02	0.639
Openness			
DN Dorsal Medial - Prefrontal Social Activation	1.12	0.05	0.264
DN Dorsal Medial - Prefrontal Random Activation	-0.89	-0.04	0.374
Extraversion			
DN Dorsal Medial - Prefrontal Social Activation	0.81	0.04	0.417
DN Dorsal Medial - Prefrontal Random Activation	-0.72	-0.03	0.471

Note. DN = default network, FPCN = frontoparietal control network.

brain activity. The association between social cognitive ability and general cognitive ability is not surprising, because utilizing social cognitive processes likely also engages other general cognitive processes, such as working memory (Phillips

et al., 2008; Spreng, 2013; Thornton & Conway, 2013), attentional processes (Holmes et al., 2003; Leslie et al., 2004; Schultebraucks et al., 2016), and nonverbal communication skills (Morrison et al., 2019). Indeed, previous research suggests strong positive correlations between general and social cognitive abilities (Allen et al., 2017; Landy, 2005; Thorndike & Stein, 1937).

What is perhaps more interesting is the potential role general intelligence might play in the association between sex and social cognitive ability, as well as how our findings might provide some explanation for why previous research has shown mixed results for sex differences in social cognitive ability (Di Tella et al., 2020). Without controlling for age or intelligence, associations between sex and the indicators for our latent social cognitive ability variable suggested little sex difference. Once general intelligence and age were introduced as covariates; however, a significant negative association appeared between sex and our latent social cognitive ability variable, indicating that females displayed higher social cognitive ability than males. Further, zero-order correlations between sex and all of the indicators of our latent IQ variable (i.e., Picture Vocabulary, English Reading, Matrix Reasoning, and all four conditions of the two-back task) show that males significantly outperformed females in general cognitive ability in our sample. Thus it would make sense that females' greater ability in social cognitive tasks specifically might be suppressed when not controlling for general cognitive ability. Considering this possibility, and the fact that few studies looking at sex differences in social cognitive ability have controlled for general intelligence (Navarra-Ventura et al., 2018), our study suggests that mixed results in previous research may stem from confounding sex differences in *general* cognitive ability with those specific to *social* cognitive ability. Our findings are consistent with the wealth of literature suggesting that females empathize with others more (Hoffman, 1977; Mestre et al., 2009) and are more accurate at interpreting the emotional states of others (Montagne et al., 2005; Nettle, 2007; Stiller & Dunbar, 2007; Wingenbach et al., 2018). This is in line with research that indicated that females are higher in Agreeableness (Costa Jr et al., 2001; Weisberg et al., 2011).

Relevance to Psychopathology

Findings from the current study could potentially be extended in future research to benefit understanding of various forms of psychopathology. As previously mentioned, poor ToM performance has been associated with a variety of psychopathology dimensions and disorders such as schizophrenia (Abram et al., 2016; Pedersen et al., 2012), schizotypy (Blain et al., 2017; Bora, 2020), autism (Baron-Cohen et al., 1985; Brune & Brune-Cohrs, 2006), autistic traits (Best et al., 2008; Blain et al., 2017), and Williams syndrome (Tager-Flusberg & Sullivan, 2000). Likewise, low Agreeableness (i.e.,

Antagonism) has been associated with a host of personality disorders and negative real-life social outcomes (Anderson et al., 2018; Krueger et al., 2012).

Our results might serve to inform research done in these clinical populations and are consistent with recent dimensional and transdiagnostic frameworks for understanding psychopathology, such as the National Institute of Mental Health's Research Domain Criteria (RDoC; Insel et al., 2010) and the Hierarchical Taxonomy of Psychopathology (HiTOP; Kotov et al., 2017). These frameworks seek to understand psychopathology in terms of underlying dimensions rather than diagnostic categories. Agreeableness and ToM, especially when considering dysfunctionally low levels of functioning, are two such promising dimensions that could be useful in clinical research and practice. Future intervention research could explore ToM deficits and low Agreeableness as transdiagnostic targets for intervention. Likewise, as the default network and its dorsal medial subsystem appear to be implicated in ToM and Agreeableness, neurostimulation research could explore whether electrical or magnetic stimulation of brain regions such as the TPJ and dmPFC might lead to changes in social cognitive ability and relevant social outcomes (Johnson et al., 2013).

Methodological Considerations

Compared with much of the previous work done on the topic, the current study uses a large dataset conferring relatively high statistical power. Although a number of existing studies have investigated possible associations between individual differences in default network function and individual differences in social cognitive ability and related traits, the majority of these studies have used sample sizes in the range of 10 to 70 individuals (Hughes et al., 2019; Inagaki & Meyer, 2020; Kaplan & Iacoboni, 2006; Song et al., 2009; Tamir et al., 2016; Wagner et al., 2011; Waytz et al., 2012; Zhang et al., 2019), which are not optimal for detecting reliable estimates of between-subjects effects (Button et al., 2013). Considering this, the design of the present study should yield more reliable findings and contribute to the robustness of the field (while still keeping in mind the limitations of our study detailed below).

The approach employed in this study to measure brain function could also benefit future research on the neurobiology of individual differences. We used a network-based approach by incorporating atlases based on patterns of functional connectivity in large samples (Schaefer et al., 2018; Yeo et al., 2011). Each participants' data were individually mapped onto a 400-parcel atlas that aligned within 17 broader functional networks identified by Yeo et al. (2011), using GPIP to ensure that parcels were adjusted to the optimal location for each participant. Each of the networks described in Yeo et al. (2011) also summarize regions of the brain that tend to be synchronously active in patterns that can be consistently identified across samples. Understanding how specific cortical parcels map onto these broad networks can be used to unify and better understand previous findings for individual regions of interest from the social neuroscience literature (Tompson et al., 2018).

Our approach should be more effective for studying individual differences that will generalize across samples, compared with the typical use of contrasts for identifying brain regions on a voxel-wise or cluster-wise basis. By focusing on well-established, large-scale brain networks, the identified regions of interest represent a broad set of brain structures with *a priori* relevance for a given construct of interest (in this case, social cognition) rather than specific voxels or clusters that might be most strongly associated with that construct only by chance in any given sample (Vul et al., 2009; Yarkoni, 2009). Moreover, approaches that focus on broad networks may be more reflective of how the brain typically functions, relative to a more localized or modular approach. Brain-behavior associations appear to be more extensive than once believed, in contrast to the relatively small clusters or regions of the brain that are often reported in underpowered samples (Yarkoni et al., 2010). A majority of brain regions are involved in multiple psychological processes, and many psychological processes involve multiple different regions of the brain, not just in the case of social neuroscience (Poldrack, 2010; Yeo et al., 2011). A network-based approach allows researchers to capture a wider picture of brain function and its relation to behavioral constructs of interest; this approach, in conjunction with a large sample size, should lend itself to reproducibility and generalizability (Yarkoni, 2009).

Limitations

Despite the multiple strengths of the current study, there are some important limitations. First, although one advantage of the current study was using multiple tasks in defining our social cognition accuracy variable, we still only utilized neuroimaging data from a single ToM task in computing our neural activation variables. Future research could use an SEM approach to model how variance in brain activity during a variety of different ToM tasks completed in the scanner might predict social cognitive ability and personality. Moreover, although the current data provide evidence that individual differences in default network function are associated with social cognitive abilities and related personality traits, the causal direction and dynamics of these associations cannot be established with the current study design. Future work incorporating methods such as neuromodulation, dynamic causal modeling, experience sampling, and long-term longitudinal data collection could help to establish more clearly the causal pathways involved in the neurobiology of personality and individual differences.

Finally, even though we found a significant correlation between ToM-related activation in the dorsal medial subsystem and Agreeableness, this correlation is likely attenuated in its effect size by the personality measure used. Although the NEO-FFI is a reasonably effective brief measure for evaluating the Big Five personality traits, it was not designed to allow for the assessment of personality at lower levels of the hierarchy, including personality aspects and facets. Given possible differential associations of subdimensions within Agreeableness with social cognition and default network function (Allen et al., 2017), a measure that can distinguish between subdimensions of Agreeableness would be optimal. Future research examining the relation between these variables should include personality measures that can assess personality at multiple levels of the trait hierarchy to better discern which specific dimensions contribute to brain-behavior associations.

Conclusions

Our findings in a very large neuroimaging sample confirm and extend the current literature linking ToM, the default network, and Agreeableness. Given that ToM-related activation in prefrontal regions of the dorsal medial subsystem positively predicted both latent levels of Agreeableness and social cognitive ability, it appears that the functions of the default network may help account for the link between Agreeableness and ToM. These findings may inform future research that seeks to understand how normal functioning goes awry in psychopathology involving social deficits, and how individual differences in social cognition and related traits affect real-world relationship success, social network quality, and interpersonal functioning. In sum, the current research furthers work on the neural and personality correlates of individual differences in social cognition while demonstrating effective methods in social cognitive neuroscience research. We recommend that researchers consider using individualized parcellation methods, network-based hypotheses, and latent variable techniques such as SEM, rather than voxel-wise analyses, when designing future studies assessing individual differences in brain data.

References

Abell, F., Happe, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. Cognitive Development, 15(1), 1–16.

Abram, S. V., Wisner, K. M., Fox, J. M., Barch, D. M., Wang, L., Csernansky, J. G., MacDonald III, A. W., & Smith, M. J. (2016). Fronto-temporal connectivity predicts cognitive empathy deficits and experiential negative symptoms in schizophrenia. Human Brain Mapping, 38(3), 1111–1124. https://doi.org/10.1002/hbm.23439.

Adolphs, R. (2009). The social brain: Neural basis of social knowledge. Annual Review of Psychology, 60, 693–716.

Allen, T. A., & DeYoung, C. G. (2017). Personality neuroscience and the Five-Factor Model. In T. A. Widiger (Ed.) Oxford Handbook of the Five Factor Model of Personality (p. 319–349). Oxford University Press.

Allen, T. A., Rueter, A. R., Abram, S. V., Brown, J. S., & DeYoung, C. G. (2017). Personality and neural correlates of mentalizing ability. European Journal of Personality, 31(6), 599–613.

Allison, P. D. (1990). Change scores as dependent variables in regression analysis. Sociological Methodology, 20, 93–114.

Altschuler, M., Sideridis, G., Kala, S., Warshawsky, M., Gilbert, R., Carroll, D., Burger-Caplan, R., & Faja, S. (2018). Measuring individual differences in cognitive, affective, and spontaneous theory of mind among school-aged children with autism spectrum disorder. Journal of Autism and Developmental Disorders, 48, 3945–3957.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. Nature Reviews Neuroscience, 7(4), 268–277. https://doi.org/10.1038/nrn1884

Anderson, J. L., Sellbom, M., & Salekin, R. T. (2018). Utility of the Personality Inventory for DSM-5-Brief Form (PID-5-BF) in the measurement of maladaptive personality and psychopathology. Assessment, 25(5), 596–607.

Anderson, K. M., Ge, T., Kong, R., Patrick, L. M., Spreng, R. N., Sabuncu, M. R., Yeo, B. T. T., & Holmes, A. J. (2021). Heritability of individualized cortical network topography. PNAS, 118(9), e2016271118. https://doi.org/10.1073/pnas.2016271118

Andrews-Hanna, J. R., Smallwood, J., & Spreng, R. N. (2014). The default network and self-generated thought: Component processes, dynamic control, and clinical relevance. Annals of the New York Academy of Sciences, 1316(1), 29–52.

Apperly, I. A. (2012). What is "theory of mind"? Concepts, cognitive processes and individual differences. The Quarterly Journal of Experimental Psychology, 65(5), 825–839.

Arbula, S., Pisanu, E., & Rumiati, R. I. (2021). Representation of social content in dorsomedial prefrontal cortex underlies individual differences in agreeableness trait. NeuroImage, 235, 118049. https://doi.org/10.1016/j.neuroimage.2021.118049

Bainbridge, T. F., Quinlan, J. A., Mar, R. A., Smillie, L. D., & Fajkowska, M. (2019). Openness/intellect and susceptibility to pseudo–profound bullshit: A replication and extension. European Journal of Personality, 33(1), 72–88.

Baker, J. (2003). Social Skills Training: For Children and Adolescents with Asperger Syndrome and Social-Communication Problems. Autism Asperger Publishing Company.

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.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? Cognition, 21, 37–46.

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. Current Directions in Psychological Science, 20(5), 286–290.

Bernhardt, B. C., & Singer, T. (2012). The neural basis of empathy. Annual Review of Neuroscience, 35, 1–23.

Best, C. S., Moffat, V. J., Power, M. J., Owens, D. G. C., & Johnstone, E. C. (2008). The boundaries of the cognitive phenotype of autism: Theory of mind, central coherence and ambiguous figure perception in young people with autistic traits. Journal of Autism and Developmental Disorders, 38(5), 840–847. https://doi.org/10.1007/s10803-007-0451-8

Blain, S. D., Peterman, J. S., & Park, S. (2017). Subtle cues missed: Impaired perception of emotion from gait in relation to schizotypy and autism spectrum traits. Schizophrenia Research, 183, 157–160.

Blain, S. D., Grazioplene, R. G., Ma, Y., & DeYoung, C. G. (2020a). Toward a neural model of the Openness-Psychoticism dimension: Functional connectivity in the default and frontoparietal control networks. Schizophrenia Bulletin, 46(3), 540–551.

Blain, S. D., Longenecker, J. M., Grazioplene, R. G., Klimes-Dougan, B., & DeYoung, C. G. (2020b). Apophenia as the disposition to false positives: A unifying framework for Openness and Psychoticism. Journal of Abnormal Psychology, 129(3), 279–292.

Blain, S. D., Sassenberg, T. A., Xi, M., Zhao, D., & DeYoung, C. G. (2020c). Extraversion but not depression predicts implicit reward sensitivity: Revisiting the measurement of anhedonic phenotypes. Online preprint: https://doi.org/10.31234/osf.io/vzt6d. Accessed 8 June 2020

Bora, E. (2020). Theory of mind and schizotypy: A meta-analysis. Schizophrenia Research, 222, 97–103. https://doi.org/10.1016/j.schres.2020.04.024

Bosacki, S., & Wilde Astington, J. (1999). Theory of mind in preadolescence: Relations between social understanding and social competence. Social Development, 8(2), 237–255.

Bowman, L. C., Thorpe, S. G., Cannon, E. N., & Fox, N. A. (2017). Action mechanisms for social cognition: Behavioral and neural correlates of developing theory of mind. Developmental Science, 20(5), e12447.

Brothers, L. (1990). The social brain: A project for integrating primate behavior and neurophysiology in a new domain. Concepts in Neuroscience, 1, 27–51.

Brune, M., & Brune-Cohrs, U. (2006). Theory of mind—evolution, ontogeny, brain mechanisms and psychopathology. Neuroscience & Biobehavioral Reviews, 30(4), 437–455.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. Annals of the New York Academy of Sciences, 1124, 1–38.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafo, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. Nature Reviews. Neuroscience, 14, 1–14.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56(2), 81–105.

Caprara, G. V., Alessandri, G., Giunta, L. D., Panerai, L., & Eisenberg, N. (2010). The contribution of agreeableness and self-efficacy beliefs to prosociality. European Journal of Personality, 24(1), 36–55. https://doi.org/10.1002/per.739

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.

Castelli, F., Happe, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. NeuroImage, 12, 314–325.

Chong, M., Bhushan, C., Joshi, A. A., Choi, S., Haldar, J. P., Shattuck, D. W., Spreng, R. N., & Leahy, R. M. (2017). Individual parcellation of resting fMRI with a group functional connectivity prior.

NeuroImage, 156, 87–100. https://doi.org/10.1016/j.neuroimage.2017.04.054

Costa Jr, P. T., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: robust and surprising findings. Journal of Personality and Social Psychology, 81(2), 322.

Costa, P. T., & McCrae, R. R. (1992). Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) Professional Manual. Psychological Assessment Resources.

Côté, S., & Moskowitz, D. S. (1998). On the dynamic covariation between interpersonal behavior and affect: Prediction from Neuroticism, Extraversion, and Agreeableness. Journal of Personality and Social Psychology, 75(4), 1032–1046.

Dahlgren, S., Dahlgren Sandberg, A., & Hjelmquist, E. (2010). The non-specificity of theory of mind deficits: Evidence from children with communicative disabilities. European Journal of Cognitive Psychology, 15(1), 129–155.

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

Devine, R. T., & Hughes, C. (2013). Silent films and strange stories: Theory of mind, gender, and social experiences in middle childhood. Child Development, 84(3), 989–1003.

DeYoung, C. G. (2015). Cybernetic big five theory. Journal of Research in Personality, 56, 33–58.

DeYoung, C. G. & Blain, S. D. (2020). Personality neuroscience: Explaining individual differences in motivation, emotion, cognition, and behavior. In Corr, P. J., & Matthews, G. (Eds.) The Cambridge Handbook of Personality Psychology (2nd ed.). Cambridge University Press.

DeYoung, C. G., & Weisberg, Y. J. (2018). Cybernetic approaches to personality and social behavior. In M. Snyder & K. Deaux (Eds). Oxford Handbook of Personality and Social Psychology, Second edition (p. 387–414). : Oxford University Press.

DeYoung, C. G., Grazioplene, R. G., & Peterson, J. B. (2012). From madness to genius: The Openness/Intellect trait domain as a paradoxical simplex. Journal of Research in Personality, 46, 63–78.

DeYoung, C. G., Weisberg, Y. J., Quilty, L. C., & Peterson, J. B. (2013). Unifying the aspects of the Big Five, the interpersonal circumplex, and trait affiliation. Journal of Personality, 81(5), 465–475.

Di Tella, M., Miti, F., Ardito, R. B., & Adenzato, M. (2020). Social cognition and sex: Are men and women really different? Personality and Individual Differences, 162, 110045. https://doi.org/10.1016/j.paid.2020.110045

Edwards, J. R. (1994). Regression analysis as an alternative to difference scores. Journal of Management, 20(3), 683–689.

Edwards, J. R. (1996). An examination of competing versions of the person-environment fit approach to stress. Academy of Management Journal, 39(2), 292–339.

Eisenberg, I.W., Bissett, P. G., Enkavi, A. Z., Li, J., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Uncovering the structure of self-regulation through data-driven ontology discovery. Nature Communications, 10(1), 1–13.

Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Large-scale analysis of test–retest reliabilities of self-regulation measures. Proceedings of the National Academy of Sciences, 116(12), 5472–5477.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. Neuron, 50(4), 531–534.

Frith, U., & Frith, C. D. (2010). The social brain: Allowing humans to boldly go where no other species has been. Philosophical Transactions of the Royal Society B, 365, 165–175.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. Trends in Cognitive Sciences, 7(2), 77–83.

Graziano, W. G., & Eisenberg, N. (1997). Agreeableness: A dimension of personality. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.),

Handbook of Personality Psychology (pp. 795–824). Academic Press.

Gur, R. C., Ragland, J. D., Moberg, P. J., Turner, T. H., Bilker, W. B., Kohler, C., Siegel, S. J., & Gur, R. E (2001). Computerized neurocognitive scanning: I. Methodology and validation in healthy people. Neuropsychopharmacology, 25(5), 766–776.

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. Personality and Social Psychology Bulletin, 42(9), 1177–1192.

Harrington, L., Siegert, R., & Mcclure, J. (2005). Theory of mind in schizophrenia: A critical review. Cognitive Neuropsychiatry, 10(4), 249–286. https://doi.org/10.1080/13546800444000056

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. Behavior Research Methods, 50(3), 1166–1186.

Hoffman, M. L. (1977). Sex differences in empathy and related behaviors. Psychological Bulletin, 84(4), 712–722. https://doi.org/10.1037/0033-2909.84.4.712

Holmes, A., Vuilleumier, P., & Eimer, M. (2003). The processing of emotional facial expression is gated by spatial attention: Evidence from event-related brain potentials. Cognitive Brain Research, 16(2), 174–184. https://doi.org/10.1016/S0926-6410(02)00268-9

Hughes, C., Cassidy, B. S., Faskowitz, J., Avena-Koenigsberger, A., Sporns, O., & Krendl, A. C. (2019). Age differences in specific neural connections within the Default Mode Network underlie theory of mind. NeuroImage, 191, 269–277. https://doi.org/10.1016/j.neuroimage.2019.02.024

Hyatt, C. J., Calhoun, V. D., Pearlson, G. D., & Assaf, M. (2015). Specific default mode subnetworks support mentalizing as revealed through opposing network recruitment by social and semantic fMRI tasks. Human Brain Mapping, 36, 3047–3063.

Inagaki, T. K., & Meyer, M. L. (2020). Individual differences in resting-state connectivity and giving social support: implications for health. Social Cognitive and Affective Neuroscience, 15(10), 1076–1085.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. American Journal of Psychiatry, 167(7), 748–751.

Jackson, P. L., Rainville, P., & Decety, J. (2006). To what extent do we share the pain of others? Insight from the neural bases of pain empathy. Pain, 125(1), 5–9.

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 O. P. John, R. W. Robins, & L. A. Pervin (Eds.). Handbook of Personality: Theory and Research (p. 114–158). Guilford.

Johnson, M. D., Lim, H. H., Netoff, T. I., Connolly, A. T., Johnson, N., Roy, A., Holt, A., Lim, K. O., Carey, J. R., Vitek, J. L., & He, B. (2013). Neuromodulation for brain disorders: challenges and opportunities. IEEE Transactions on Biomedical Engineering, 60(3), 610–624.

Kaplan, J. T., & Iacoboni, M. (2006). Getting a grip on other minds: Mirror neurons, intention understanding, and cognitive empathy. Social Neuroscience, 1(3–4), 175–183.

Keith, T. Z. (2006). Multiple Regression and Beyond. Allyn & Bacon.

Kerr, N., Dunbar, R. I. M., & Bentall, R. P. (2003). Theory of mind deficits in bipolar affective disorder. Journal of Affective Disorders, 73(3), 253–259. https://doi.org/10.1016/S0165-0327(02)00008-3

Kong, R., Li, J., Orban, C., Sabuncu, M. R., Liu, H., Schaefer, A., Sun, N., Zuo, X., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2019). Spatial topography of individual-specific cortical networks predicts human cognition, personality, and emotion. Cerebral Cortex, 29, 2533–2551. https://doi.org/10.1093/cercor/bhy123

Kong, R., Yang, Q., Gordon, E., Xue, A., Yan, X., Orban, C., ... & Yeo, B. T. (2021). Individual-specific areal-level parcellations improve functional connectivity prediction of behavior. bioRxiv. https://doi.org/10.1101/2021.01.16.426943

Koole, S. L., Jager, W., van den Berg, A. E., Vlek, C. A. J., & Hofstee, W. K. B. (2001). On the social nature of personality: Effects of Extraversion, Agreeableness, and feedback about collective resource use on cooperation in a resource dilemma. Personality and Social Psychology Bulletin, 27(3), 289–301. https://doi.org/10.1177/0146167201273003

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Hynam, D. R., Markon, K., … Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): a dimensional alternative to traditional nosologies. Journal of Abnormal Psychology, 126(4), 454.

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. Psychological Medicine, 42(9), 1879–1890.

Landy, F. J. (2005). Some historical and scientific issues related to research on emotional intelligence. Journal of Organizational Behavior, 26(4), 411–424.

Laursen, B., Pulkkinen, L., & Adams, R. (2002). The antecedents and correlates of agreeableness in adulthood. Developmental Psychology, 38(4), 591–603. https://doi.org/10.1037/0012-1649.38.4.591

Leech, R., & Sharp, D. J. (2014). The role of the posterior cingulate cortex in cognition and disease. Brain, 137(1), 12–32.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind'. Trends in Cognitive Sciences, 8(12), 528–533. https://doi.org/10.1016/j.tics.2004.10.001

Liddle, B., & Nettle, D. (2006). Higher-order theory of mind and social competence in school-age children. Journal of Cultural and Evolutionary Psychology, 4(3–4), 231–246.

Lucas, R. E., Diener, E., Grob, A., Suh, E. M., & Shao, L. (2000). Cross-cultural evidence for the fundamental features of extraversion. Journal of Personality and Social Psychology, 79(3), 452–468. https://doi.org/10.1037/0022-3514.79.3.452

Malouff, J. M., Thorsteinsson, E. B., Schutte, N. S., Bhullar, N., & Rooke, S. E. (2010). The five-factor model of personality and relationship satisfaction of intimate partners: A meta-analysis. Journal of Research in Personality, 44, 124–127. https://doi.org/10.1016/j.jrp.2009.09.004

Mars, R. B., Neubert, F. X., Noonan, M. P., Sallet, J., Toni, I., & Rushworth, M. F. (2012). On the relationship between the "default mode network" and the "social brain". Frontiers in Human Neuroscience, 6, 189.

Meier, B. P., Robinson, M. D., & Wilkowski, B. M. (2006). Turning the other cheek. Agreeableness and the regulation of aggression-related primes. Psychological Science, 17(2), 136–142.

Mestre, M. V., Samper, P., Frías, M. D., & Tur, A. M. (2009). Are Women More Empathetic than Men? A Longitudinal Study in Adolescence. The Spanish Journal of Psychology, 12(1), 76–83. https://doi.org/10.1017/S1138741600001499

Meyer, M. L. (2019). Social by default: characterizing the social functions of the resting brain. Current Directions in Psychological Science, 28(4), 380–386.

Mohr, P., Howells, K., Gerace, A., Day, A., & Wharton, M. (2007). The role of perspective taking in anger arousal. Personality and Individual Differences, 43, 507–517.

Montagne, B., Kessels, R. P. C., Frigerio, E., de Haan, E. H. F., & Perrett, D. I. (2005). Sex differences in the perception of affective facial expressions: Do men really lack emotional sensitivity? Cognitive

Processing, 6(2), 136–141. https://doi.org/10.1007/s10339-005-0050-6

Moran, J. M., Kelley, W. M., & Heatherton, T. F. (2013). What can the organization of the brain's default mode network tell us about self-knowledge? Frontiers in Human Neuroscience, 7, 391.

Morrison, K. E., Pinkham, A. E., Kelsven, S., Ludwig, K., Penn, D. L., & Sasson, N. J. (2019). Psychometric evaluation of social cognitive measures for adults with autism. Autism Research, 12(5), 766–778.

Mwilambwe-Tshilobo, L., Ge, T., Chong, M., Ferguson, M. A., Misic, B., Burrow, A. L., Leahy, R. M., & Spreng, R. N. (2019). Loneliness and meaning in life are reflected in the intrinsic network architecture of the brain. Social Cognitive and Affective Neuroscience, 14(4), 423–433. https://doi.org/10.1093/scan/nsz021

Navarra-Ventura, G., Fernandez-Gonzalo, S., Turon, M., Pousa, E., Palao, D., Cardoner, N., & Jodar, M. (2018). Gender differences in social cognition: A cross-sectional pilot study of recently diagnosed patients with schizophrenia and healthy subjects. Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie, 63(8), 538–546. https://doi.org/10.1177/0706743717746661

Nettle, D. (2007). Empathizing and systemizing: What are they, and what do they contribute to our understanding of psychological sex differences? British Journal of Psychology, 98(2), 237-255.

Nettle, D., & Liddle, B. (2008). Agreeableness is related to social-cognitive, but not social-perceptual, theory of mind. European Journal of Personality, 22, 323–335.

Nosek, B. A., & Smyth, F. L. (2007). A multitrait-multimethod validation of the implicit association test. Experimental Psychology, 54(1), 14–29.

Ozer, D. J., & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. Annual Review of Psychology, 57, 401–421.

Paal, T., & Bereczkei, T. (2007). Adult theory of mind, cooperation, Machiavellianism: The effect of mindreading on social relations. Personality and Individual Differences, 43, 541–551.

Pedersen, A., Koelkebeck, K., Brandt, M., Wee, M., Kueppers, K. A., Kugul, H., Kohl, W., Bauer, J., & Ohrmann, P. (2012). Theory of mind in patients with schizophrenia: Is mentalizing delayed? Schizophrenia Research, 137(1–3), 224–229.

Phillips, L. H., Channon, S., Tunstall, M., Hedenstrom, A., & Lyons, K. (2008). The role of working memory in decoding emotions. Emotion, 8(2), 184–191. https://doi.org/10.1037/1528-3542.8.2.184

Pinkham, A. E., Hopfinger, J. B., Pelphrey, K. A., Piven, J., & Penn, D. L. (2008). Neural bases for impaired social cognition in schizophrenia and autism spectrum disorders. Schizophrenia Research, 99(1–3), 164–175.

Pinkham, A. E., Harvey, P. D., & Penn, D. L. (2018). Social cognition psychometric evaluation: Results of the final validation study. Schizophrenia Bulletin, 44(4), 737–748.

Poldrack, R. A. (2010). Mapping mental function to brain structure: How can cognitive neuroimaging succeed? Perspectives on Psychological Science : A Journal of the Association for Psychological Science, 5(5), 753–761. https://doi.org/10.1177/1745691610388777

Rosseel, Y. (2012). LAVAAN: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48, 1–36. URL http://www.jstatsoft.org/v48/i02/

Sabbagh, M. A., Moulson, M. C., & Harkness, K. L. (2004). Neural correlates of mental state decoding in human adults: An event-related potential study. Journal of Cognitive Neuroscience, 16(3), 415–426.

Santarnecchi, E., Emmendorfer, A., & Pascual-Leone, A. (2017). Dissecting the parieto-frontal correlates of fluid intelligence: A comprehensive ALE meta-analysis study. Intelligence, 63, 9–28.

Sasamoto, A., Miyata, J., Hirao, K., Fujiwara, H., Kawada, R., Fujimoto, S., Tanaka, Y., Kubota, M., Sawamoto, N., Fukuyama, H., Takahashi, H., & Murai, T. (2011). Social impairment in

schizophrenia revealed by Autism-Spectrum Quotient correlated with gray matter reduction. Social Neuroscience of Psychiatric Disorders, 6, 548–558.

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.

Saxe, R., & Powell, L. J. (2006). It's the thought that counts: Specific brain regions for one component of theory of mind. Psychological Science, 17, 692–699. https://doi.org/10.1111/j.1467-9280.2006.01768.x.

Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo- parietal junction. Neuropsychologia, 43, 1391–1399. https://doi.org/10.1016/j.neuropsychologia.2005.02.013.

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X. N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. Cerebral Cortex, 28(9), 3095–3114.

Schilbach, L., Eickhoff, S. B., Rotarska-Jagiela, A., Fink, G. R., & Vogeley, K. (2008). Minds at rest? Social cognition as the default mode of cognizing and its putative relationship to the "default system" of the brain. Consciousness and Cognition, 17(2), 457–467.

Schilbach, L., Bzdok, D., Timmermans, B., Fox, P. T., Laird, A. R., Vogeley, K., & Eickhoff, S. B. (2012). Introspective minds: using ALE meta-analyses to study commonalities in the neural correlates of emotional processing, social & unconstrained cognition. PLoS One, 7(2), e30920.

Schnabel, K., Asendorpf, J. B., & Greenwald, A. G. (2008). Assessment of individual differences in implicit cognition: A review of IAT measures. European Journal of Psychological Assessment, 24(4), 210–217.

Schultebraucks, K., Deuter, C. E., Duesenberg, M., Schulze, L., Hellmann-Regen, J., Domke, A., Lockenvitz, L., Kuehl, L. K., Otte, C., & Wingenfeld, K. (2016). Selective attention to emotional cues and emotion recognition in healthy subjects: The role of mineralocorticoid receptor stimulation. Psychopharmacology, 233(18), 3405–3415. https://doi.org/10.1007/s00213-016-4380-0

Schumacker, R. E., & Lomax, R. G. (2004). A Beginner's Guide to Structural Equation Modeling. Psychology Press.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: a meta-analysis of functional brain imaging studies. Neuroscience & Biobehavioral Reviews, 42, 9–34.

Seitz, R. J., Nickel, J., & Azari, N. P. (2006). Functional modularity of the medial prefrontal cortex: involvement in human empathy. Neuropsychology, 20(6), 743.

Sibley, C. G., & Duckitt, J. (2008). Personality and prejudice: A meta-analysis and theoretical review. Personality and Social Psychology Review, 12(3), 248–279. https://doi.org/10.1177/1088868308319226.

Silvia, P. J., & Christensen, A. P. (2020). Looking up at the curious personality: individual differences in curiosity and openness to experience. Current Opinion in Behavioral Sciences, 35, 1–6.

Singer, T., & Klimecki, O. M. (2014). Empathy and compassion. Current Biology, 24(18), R875–878.

Smillie, L. D., Cooper, A. J., Wilt, J., & Revelle, W. (2012). Do extraverts get more bang for the buck? Refining the affective-reactivity hypothesis of Extraversion. Journal of Personality and Social Psychology, 103(2), 306–326.

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., ... & Beckmann, C. F. (2009). Correspondence of the brain's functional architecture during activation and rest. Proceedings of the National Academy of Sciences, 106(31), 13040-13045.

Song, M., Liu, Y., Yuan, Z., Wang, K., Yu, C., & Jiang, T. (2009). Default network and intelligence difference. Autonomous Mental Development, IEEE Transactions On, 2009, 101–109. https://doi.org/10.1109/TAMD.2009.2029312

Spreng, R. N. (2013). Examining the role of memory in social cognition. Frontiers in Psychology, 4, 1–2. https://doi.org/10.3389/fpsyg.2013.00437

Spreng, R. N., & Andrews-Hanna, J. R. (2015). The default network and social cognition. Brain Mapping: An Encyclopedic Reference, 3, 165–169.

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. Journal of Cognitive Neuroscience, 21(3), 489–510.

Spunt, R. P., & Lieberman, M. D. (2012). An integrative model of the neural systems supporting the comprehension of observed emotional behavior. NeuroImage, 59(3), 3050–3059. https://doi.org/10.1016/j.neuroimage.2011.10.005

Stiller, J., & Dunbar, R. I. M. (2007). Perspective-taking and memory capacity predict social network size. Social Networks, 29, 93–104.

Tager-Flusberg, H. (2011). False-belief tasks are distinct from theory of mind. Retrieved March 29th, 2020 from https://www.spectrumnews.org/opinion/viewpoint/false-belief-tasks-are-distinct-from-theory-of-mind/

Tager-Flusberg, H., & Sullivan, K. (2000). A componential view of theory of mind: Evidence from Williams syndrome. Cognition, 76, 59–90.

Takeuchi, H., Taki, Y., Nouchi, R., Sekiguchi, A., Hashizume, H., Sassa, Y., Kotozaki, Y., Miyauchi, C. M., Yokoyama, R., Iizuka, K., Nakagawa, S., Nagase, T., Kunitoki, K., & Kawashima, R. (2014). Association between resting-state functional connectivity and empathizing/systemizing. NeuroImage, 99, 312–322.

Tamir, D. I., Bricker, A. B., Dodell-Feder, D., & Mitchell, J. P. (2016). Reading fiction and reading minds: The role of simulation in the default network. Social Cognitive and Affective Neuroscience, 11(2), 215–224. https://doi.org/10.1093/scan/nsv114

Thorndike, R. L., & Stein, S. (1937). An evaluation of the attempts to measure social intelligence. Psychological Bulletin, 34(5), 275.

Thornton, M. A., & Conway, A. R. A. (2013). Working memory for social information: Chunking or domain-specific buffer? NeuroImage, 70, 233–239. https://doi.org/10.1016/j.neuroimage.2012.12.063

Tompson, S. H., Falk, E. B., Vettel, J. M., & Bassett, D. S. (2018). Network approaches to understand individual differences in brain connectivity: Opportunities for personality neuroscience. Personality Neuroscience, 1(e5), 1–12.

Uğurbil, K., Xu, J., Auerbach, E. J., Moeller, S., Vu, A. T., Duarte-Carvajalino, J. M., Lenglet, C., Wu, X., Schmitter, S., Van de Moortele, P. F., Strupp, J., Sapiro, G., De Martino, F., Wang, D., Harel, N., Garwood, M., Chen, L., Feinberg, D. A., Smith, S. M., … Yacoub, E. (2013). Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. NeuroImage, 80, 80–104.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., Uğurbil, K., & WU-Minn HCP Consortium. (2013), The WU-Minn human connectome project: an overview. NeuroImage, 80, 62–79.

Vellante, M., Baron-Cohen, S., Melis, M., Marrone, M., Petretto, D. R., Masala, C., & Preti, A. (2013). The "Reading the Mind in the Eyes" test: Systematic review of psychometric properties and a validation study in Italy. Cognitive Neuropsychiatry, 18(4), 326–354.

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happé, F., Falkai, P., ... & Zilles, K. (2001). Mind reading: neural mechanisms of theory of mind and self-perspective. Neuroimage, 14(1), 170–181.

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. Perspectives on Psychological Science, 4(3), 274–290.

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. Cerebral Cortex, 21(12), 2788–2796. https://doi.org/10.1093/cercor/bhr074

Wang, Z., Devine, R. T., Wong, K. K., & Hughes, C. (2016). Theory of mind and executive function during middle childhood across cultures. Journal of Experimental Child Psychology, 149, 6–22.

Waytz, A., Zaki, J., & Mitchell, J. P. (2012). Response of dorsomedial prefrontal cortex predicts altruistic behavior. Journal of Neuroscience, 32(22), 7646–7650.

Weidmann, R., Ledermann, T., & Grob, A. (2017). Big Five traits and relationship satisfaction: The mediating role of self-esteem. Journal of Research in Personality, 69, 102–109. https://doi.org/10.1016/j.jrp.2016.06.001

Weisberg, Y. J., DeYoung, C. G., & Hirsh, J. B. (2011). Gender differences in personality across the ten aspects of the Big Five. Frontiers in Psychology, 2, 178.

Weiss, B., Jahn, A., Hyatt, C. S., Owens, M. M., Carter, N. T., Sweet, L. H., ... & Haas, B. W. (2021). Investigating the neural substrates of Antagonistic Externalizing and social-cognitive Theory of Mind: An fMRI examination of functional activity and synchrony. Personality Neuroscience, 4.

Wheatley, T., Milleville, S. C., & Martin, A. (2007). Understanding animate agents: Distinct role for the social network and mirror system. Psychological Science, 18(6), 469–474.

Wingenbach, T. S. H., Ashwin, C., & Brosnan, M. (2018). Sex differences in facial emotion recognition across varying expression intensity levels from videos. PLoS One, 13(1), e0190634. https://doi.org/10.1371/journal.pone.0190634

Wittenborn, J. R. (1951). An evaluation of the use of difference scores in prediction. Journal of Clinical Psychology, 7, 108–111.

Xiao, M., Zhu, W., Wei, J., Lei, X., & Xia, L. (2019). The relationship among resting-state brain activity and connectivity, agreeableness and displaced aggression: Two possible mediation models. Journal of Affective Disorders, 256, 641–649. https://doi.org/10.1016/j.jad.2019.06.046

Yarkoni, T. (2009). Big correlations in little studies: Inflated fMRI correlations reflect low statistical power—Commentary on Vul et al. (2009). Perspectives on Psychological Science, 4(3), 294–298.

Yarkoni, T., Poldrack, R. A., Van Essen, D. C., & Wager, T. D. (2010). Cognitive neuroscience 2.0: Building a cumulative science of human brain function. Trends in Cognitive Sciences, 14(11), 489–496.

Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zollei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. Journal of Neurophysiology, 106, 1125–1165.

Young, L., Dodell-Feder, D., & Saxe, R. (2010). What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. Neuropsychologia, 48, 2658–2664. https://doi.org/10.1016/j.neuropsychologia.2010.05.012

Zhang, M., Savill, N., Margulies, D. S., Smallwood, J., & Jefferies, E. (2019). Distinct individual differences in default mode network connectivity relate to off-task thought and text memory during reading. Scientific Reports, 9(1), 16220. https://doi.org/10.1038/s41598-019-52674-9