Shared Neural Circuits for Mentalizing about the Self and Others

Michael V. Lombardo[1], Bhismadev Chakrabarti[1,2],
Edward T. Bullmore[1], Sally J. Wheelwright[1], Susan A. Sadek[1],
John Suckling[1], MRC AIMS Consortium*, and Simon Baron-Cohen[1]

Abstract

■ Although many examples exist for shared neural representations of self and other, it is unknown how such shared representations interact with the rest of the brain. Furthermore, do high-level inference-based shared mentalizing representations interact with lower level embodied/simulation-based shared representations? We used functional neuroimaging (fMRI) and a functional connectivity approach to assess these questions during high-level inference-based mentalizing. Shared mentalizing representations in ventromedial prefrontal cortex, posterior cingulate/precuneus, and temporo-parietal junction (TPJ) all exhibited identical functional connectivity patterns during mentalizing of both self and other. Connectivity patterns were distributed across low-level embodied neural systems such as the frontal operculum/ventral premotor cortex, the anterior insula, the primary sensorimotor cortex, and the presupplementary motor area. These results demonstrate that identical neural circuits are implementing processes involved in mentalizing of both self and other and that the nature of such processes may be the integration of low-level embodied processes within higher level inference-based mentalizing. ■

INTRODUCTION

"Know thyself"… And know others? This ancient Greek aphorism lies at the heart of centuries of scholarly inquiry in many fields such as psychology, philosophy, and most recently, social neuroscience (Decety & Grezes, 2006). One theoretical framework providing the basis for such inquiry is the simulationist accounts of social cognition (Goldman, 2006). Simulationist accounts posit that we gain insight about another's mental experience through the use of privileged access to our own mental states, sensations, emotions, and other embodied representations. One of the main premises forming the basis of simulationist accounts is the simple prediction that the brain engages the same neural systems for self and other referential cognitive processes. Findings supporting this premise, however, are divided across two important neural systems. The first system, coding for low-level embodied/simulative representations, exists within frontal operculum/ventral premotor cortex (FO/PMv) (Gazzola & Keysers, 2009; Gazzola, Aziz-Zadeh, & Keysers, 2006; Iacoboni & Dapretto, 2006; Rizzolatti & Craighero, 2004), somatosensory cortices (SI/SII) (Blakemore, Bristow, Bird, Frith, & Ward, 2005; Keysers et al., 2004), anterior insula (AI), and caudal ACC (cACC) extending into the presupplementary motor area (pre-SMA; Critchley, Wiens, Rotshtein, Ohman, & Dolan, 2004; Singer et al., 2004; Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003; Wicker et al., 2003). The second neural system, dealing with more high-level inference-based mentalizing about both self and other, comprises an independent set of neural regions within the medial prefrontal cortex (MPFC), posterior cingulate cortex/precuneus (PCC), and TPJ (Amodio & Frith, 2006; Mitchell, Macrae, & Banaji, 2006; Saxe, Moran, Scholz, & Gabrieli, 2006; Ochsner et al., 2005). Given the importance of both shared neural systems in social cognition, one puzzling question is why are these two neural systems consistently observed independently of one another? And second, if they are not independent, how do these two crucial shared neural systems interact during social cognitive processes such as mentalizing (Keysers & Gazzola, 2007; Uddin, Iacoboni, Lange, & Keenan, 2007)? These central questions pose key challenges for explaining how we navigate and

[1]University of Cambridge, United Kingdom, [2]University of Reading, United Kingdom

interact with the social world around us. Understanding how, if at all, such shared neural systems interact will help clarify the nature of the underlying processes implemented during mentalizing and canalize further theoretical refinement and empirical work on the topic.

One method for gaining insight into these questions is through functional or effective connectivity approaches (Friston et al., 1997). Functional connectivity can provide a more precise test of whether shared neural representations for mentalizing about the self and other are in fact "shared" in how they interact with the rest of the brain. Previous examples in the domain of shared pain and disgust processing suggest that rather than converging across similar neural circuits, these shared representations actually diverge into distinct functionally connected neural circuits (Jabbi, Bastiaansen, & Keysers, 2008; Zaki, Ochsner, Hanelin, Wager, & Mackey, 2007). These examples are prime illustrations of how functional connectivity can provide additional insight into the nature of the processes underlying shared representations. Extending this idea into the domain of higher level inference-based mentalizing, if functional connectivity patterns are similar for mentalizing about both self and other, this would support the idea that shared mentalizing representations implement similar underlying processes for both self and other. However, the alternative hypothesis would predict that the implementation of processes underlying shared mentalizing representations (indexed by functional connectivity) actually diverge into distinct neural circuits for self and other. This alternative would support the claim that the processes underlying self-mentalizing are distinct from those underlying other-mentalizing.

Second, by identifying the neural space through which shared mentalizing processes are implemented, functional connectivity approaches can also constrain ideas about the psychological/cognitive significance of such underlying processes. Historically speaking, scholars in the field have contrasted high-level inference-based mentalizing processes as an opposite of low-level simulation-based processes (Goldman, 2006; Gopnik & Wellman, 1992; Gordon, 1992). However, a middle ground may exist within the idea that some aspects of both high-level and low-level social cognitive processes are grounded within the framework of embodied cognition (Goldman & de Vignemont, 2009; Barsalou, 1999, 2008). Based on theories of embodied cognition (Barsalou, 1999, 2008), it would be predicted that high-level social-cognitive conceptual representations such as those occurring through explicit mentalizing are built upon through their interactions with low-level embodied/simulative representations. A recent model by Keysers and Gazzola (2007) proposes that this may be implemented in the brain through the integration of information from neural systems for low-level embodied/simulation-based processes (e.g., action–perception mirroring, somatosensory or other embodied representations) and high-level inference-based mentalizing. The integration of these two neural systems makes

the prediction that during high-level inference-based mentalizing, neural systems such as MPFC, PCC, and TPJ should be functionally connected to lower level embodied/simulation-based neural systems like FO/PMv, SI/SII, AI, and cACC/pre-SMA. Such a finding would suggest that during high-level inference-based mentalizing, we use our own lower level embodied/simulated shared representations as the building blocks for making inferences about our own mind as well as other's minds.

We used fMRI and a functional connectivity approach to examine these interactions within 33 healthy male participants. Participants were scanned in a 2 × 2 factorially designed fMRI experiment where participants either made a reflective judgment about the self or a familiar non-close other (the British Queen) in a mentalistic way (self-mentalizing [SM], other-mentalizing [OM]) or physical way (self-physical [SP], other-physical [OP]; see Methods).

METHODS

Participants

Thirty-three healthy male participants were included in this study (mean age = 27.97 years, SD = 6.10 years, range = 18–42 years). Informed consent was obtained for all participants in accord with procedures approved by the Suffolk Local Research Ethics Committee. All participants were right-handed native English speakers with normal or corrected-to-normal vision. Participants reported no history of psychiatric or neurological conditions and were not currently taking any medication.

Task Design

The study design was a 2 × 2 within-subjects factorial block design where participants were asked to make either reflective "mentalizing" or "physical" judgments about two target individuals: the "self" or a familiar non-close "other" (the British Queen). For SM blocks, participants judged on a scale from 1 (*not at all likely*) to 4 (*very likely*) how likely they themselves would personally agree with opinion questions that focused on mental characteristics (e.g., "How likely are *You* to think that keeping a diary is important"). On OM blocks, the same mentalizing judgments were made, except this time it was in reference to how likely the British Queen would agree with the opinion questions (e.g., "How likely is the *Queen* to think that keeping a diary is important"). During SP blocks, participants judged how likely they would personally agree to questions about their own physical characteristics (e.g., "How likely are *You* to sneeze when a cat is nearby"). Conversely, the same physical judgments were made during OP blocks, except that participants rated these questions with the Queen as the target person (e.g., "How likely is the *Queen* to sneeze when a cat is nearby").

All participants completed one scanning session with one functional imaging run. Within this run, there were 20 trials within each condition and five blocks per condition. Each trial type was presented in blocks of four trials, and the duration of each trial was 4 sec (16 sec per block). After each block, there was a rest period of 16 sec where participants fixated on a cross in the middle of the screen. All trials within blocks and all blocks throughout the functional run were presented in pseudorandom order. All opinion questions were acquired from Jason Mitchell's laboratory and have been used in previous studies on reflective judgments about the self and others (Jenkins, Macrae, & Mitchell, 2008; Mitchell et al., 2006). Stimulus presentation was implemented with the DMDX software, and the stimulus presentation computer was synchronized with the onset of the functional run to ensure accuracy of event timing.

fMRI Acquisition

Imaging was performed using a 3-T GE Signa Scanner (General Electric Medical Systems, Milwaukee, WI) at the Cambridge Magnetic Resonance Imaging and Spectroscopy Unit (MRIS Unit). Our functional imaging run consisted of 325 whole-brain functional T2*-weighted EPIs (slice thickness = 3 mm; 0.8 mm skip; 33 axial slices; repetition time = 2000 msec; echo time = 30 msec; flip angle = 90°; matrix = 64 × 64; field of view = 240 mm; sequential slice acquisition). The first five time points of the run were discarded to allow for T2 stabilization effects. In addition, a high-resolution three-dimensional spoiled gradient (SPGR) anatomical image was acquired for each participant for registration purposes.

Data Analysis

Behavioral RT data were analyzed with a repeated measures ANOVA in SPSS 16 (http://www.spss.com). fMRI data preprocessing and statistics were implemented using SPM5 (Wellcome Trust Centre for Neuroimaging, http://www.fil.ion.ucl.ac.uk/spm). The preprocessing steps were conducted in the following manner: Functional data were slice timing corrected and realigned to the mean functional image. Next, the realigned and slice-timing-corrected functional data were coregistered to the high-resolution SPGR. The high-resolution SPGR was then segmented into cerebrospinal fluid, gray, and white matter. The normalization transformation matrix from the segmentation step was then applied to the functional and structural images, thus transforming it into standard anatomical space based on the ICBM 152 brain template (Montreal Neurological Institute [MNI]) at a resolution of 2 mm of isotropic voxels. Normalized functional data were then smoothed with a FWHM of 8 mm.

Whole-brain statistical analysis was performed using the general linear model in SPM5. Each trial was convolved with the canonical hemodynamic response function. High-pass temporal filtering with a cutoff of 128 sec was applied to remove low-frequency drift in the time series, and global changes were removed by proportional linear scaling. Serial autocorrelations were estimated with a restricted maximum likelihood algorithm with an autoregressive model of order 1. Factorial contrast images were outputted automatically in the first-level single subject analysis.

At the group level of analysis, three conjunction analyses were performed. All conjunction analyses use a logical "AND" masking procedure whereby a map of false discovery rate (FDR)-corrected suprathreshold voxels in contrast A is overlaid on a map of FDR-corrected suprathreshold voxels in orthogonal contrast B and only the voxels that overlap across contrast A "AND" B are extracted. The first conjunction analysis identified overlap among mentalizing representations for self and other (SM > SP "AND" OM > OP). The second and the third conjunction analyses identified general self-biases (SM > OM "AND" SP > OP) or other biases (OM > SM "AND" OP > SP). To test for whole-brain interaction effects among the four conditions, we conducted a paired samples t test comparing the contrasts of SM > SP to OM > OP (i.e., [SM > SP] > [OM > OP]). This analysis was thresholded at $p < .05$, FDR corrected.

Functional connectivity analyses were implemented with psychophysiological interaction (PPI) analyses within SPM5 (Friston et al., 1997). Three seed regions were defined for the PPI analyses. These seeds were the ventromedial prefrontal cortex (vMPFC), the PCC, and the right TPJ (RTPJ) and were each functionally defined as the entire cluster of suprathreshold voxels in each respective region identified from the group-level conjunction analysis as shared mentalizing representations under both SM > SP "AND" OM > OP contrasts (see the first three clusters in Table 1). Time courses from each seed region were extracted and multiplied by a condition vector of 0, 1, or −1, where mentalizing trials (e.g., in SM > SP, this is the SM condition) were coded as 1, physical trials were coded as −1 (e.g., in SM > SP, this is the SP condition), and all other events were coded as 0. This product vector of [time courses × condition vector] was our PPI vector. The time course, the condition vector, and the PPI vector were entered as regressors into single subject analyses, and contrast maps were computed for the PPI regressor. Single subject PPI contrast maps were then entered into a second-level group analysis.

At the group level, one-sample t tests were computed on the PPI contrast images and thresholded at $p < .05$, FDR corrected. These thresholded maps were then used in conjunction analyses (using the aforementioned logical "AND" masking procedure) to identify areas of common functional connectivity to both SM > SP "AND" OM > OP. These conjunction analyses were run on each seed region independently. The resulting conjunction maps for each seed region were then used to compute a final conjunction map that isolated shared functional connectivity for SM and OM that was common to every seed region.

Table 1. Shared Neural Mentalizing Representations for Self AND Other

Region	Hemisphere	SM > SP			Cluster Size	OM > OP		
		BA	MNI (x,y,z)	t		BA	MNI (x,y,z)	t
vMPFC	L	10, 11	−6, 54, −8	4.39	56	10, 11	−4, 50, −8	4.73
PCC	L	30	−4, −56, 12	10.65	2285	19	−16, −52, −2	8.25
	B	30	6, −54, 10	7.64		30	−6, −58, 16	7.33
	R	19	16, −50, −2	5.95		19	16, −52, −4	6.67
TPJ	R	37	54, −58, 14	6.29	333	37	52, −60, 14	7.27
TPJ	L	39	−44, −66, 18	4.8	247	39	−46, −64, 18	5.23
ATL	L	20	−56, −10, −18	3.37	46	21	−56, −4, −26	4.68
	L	21	−54, 4, −28	3.08		21	−54, 4, −28	3.94
	L	21	−54, −6, −26	2.91		20	−58, −12, −18	3.91
SI/MI	L	4	−34, −34, 54	3.76	88	3	−32, −32, 58	3.96
	L	4	−32, −22, 58	3.54				
Cerebellum	R		26, −38, −24	4.93	93		28, −40, −20	4.56

Shared neural representations for self and other during mentalizing judgments. All coordinates are reported from the conjunction analysis of overlap between the SM > SP contrast and the OM > OP contrast (p < .05, FDR-corrected threshold for each contrast).

SM = self-mentalizing; SP = self-physical; OM = other-mentalizing; OP = other-physical; vMPFC = ventromedial prefrontal cortex; PCC = posterior cingulate cortex/precuneus; ATL = anterior temporal lobe; SI/MI = primary sensorimotor cortex.

To identify areas of functional connectivity specific to SM or OM, interaction effect analyses were conducted on the PPI contrasts using paired samples t tests. These paired samples t test compared the PPI contrasts of SM > SP to OM > OP and were done individually on all seed regions. All paired samples t tests were thresholded at $p < .05$, FDR corrected.

To corroborate the direction of the relationships identified with PPI, we used the following procedure. Briefly, the raw time courses of the seed and target ROIs (target ROIs of FO, PMv, pre-SMA, and SI/MI were defined from the final conjunction map of shared connectivity among all seed regions) were extracted and multiplied by an HRF-convolved task vector of each condition. Correlations were computed separately for each condition among the seed and target ROIs. Correlations coefficients were converted into Fisher's z scores (Steiger, 1980) for the purposes of visual comparisons of correlations between conditions.

RESULTS

Behavioral Data

Participants were significantly faster to make judgments about the self in a mentalizing context (SM, mean = 2413.65, SD = 56.72) compared with all other conditions (OM, mean = 2564.49, SD = 58.87; SP, mean = 2523.14, SD = 52.31; OP, mean = 2565.45, SD = 57.63; $p < .001$).

RTs for all other conditions were statistically equivalent to one another ($p > .05$).

Shared Neural Representations

Using conjunction analyses, we defined shared mentalizing representations as the overlap between the SM > SP contrast and the OM > OP contrast (each thresholded at $p < .05$, FDR corrected). These areas of overlap consisted of the vMPFC (Brodmann's area [BA] 10/11), PCC (BA 30/19), and bilateral TPJ (BA 37/39) as well as the left anterior temporal lobe (ATL) along the middle temporal gyrus (BA 20/21), left primary sensorimotor cortex (SI/MI; BA 3/4), and cerebellum (see Figure 1A and Table 1).

To identify any interaction effects, we examined the paired samples t test comparing the SM > SP contrast images to the OM > OP contrast images ([SM > SP] > [OM > OP]). This analysis revealed no significant results. The absence of any interactions within our factorial design combined with the results of the conjunction analyses signals that mentalizing representations about the self and others are largely recruiting identical neural circuitry.

Self–Other Distinction

Although mentalizing representations shared the same neural space for self and other, three of the shared mentalizing regions (vMPFC, PCC, and RTPJ) were also identified

Figure 1. Conjunction analyses activation results. This figure plots the activation results for the conjunction analysis of (A) SM compared with SP judgments (red voxels) superimposed on the results of the OM compared with OP judgments (blue voxels). The white voxels show the overlap in mentalizing representations for both self and other. (B) Activation results for the self-conjunction analysis of SM judgments compared with OM judgments (red voxels) superimposed on the results of SP judgments compared with OP judgments (blue voxels). Again, white voxels denote the overlap for self-judgments compared with other-judgments. (C) Results of the other-conjunction analysis for OM compared with SM judgments (red voxels) superimposed on the contrast of OP compared with SP judgments (blue voxels). The white voxels denote the overlap for other judgments. Each contrast is thresholded independently at $p < .05$, FDR corrected.
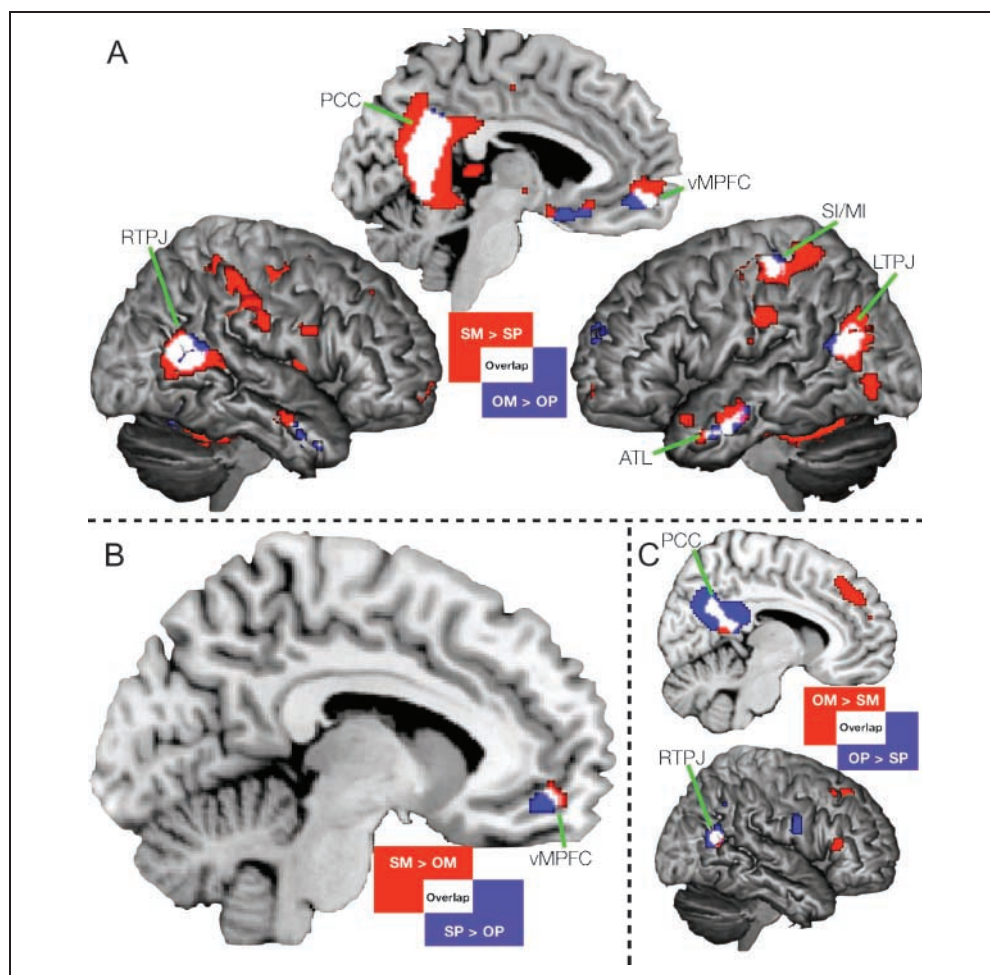
as biased toward target-specific information processing. In another set of whole-brain conjunction analyses of the SM > OM and SP > OP contrasts (thresholded at $p < .05$, FDR corrected), vMPFC (BA 10/11) was the only region identified as having such a self > other bias (see Figure 1B and Table 2A). In contrast, the conjunction of OM > SM and OP > SP contrasts revealed a general other > self bias in both right PCC (BA 23/29/30) and RTPJ (BA 37; see Figure 1C and Table 2B).

Shared Functional Connectivity

Next, we ran PPI analyses to identify areas that are more functionally connected during SM and OM compared with connectivity during physical judgments. Three seed regions were specified in vMPFC, PCC, and RTPJ. These seed regions consisted of all suprathreshold voxels in each respective region that were identified from the SM > SP and OM > OP group-level conjunction analysis. For each seed, two separate PPI analyses were conducted; one identified increased functional connectivity during SM > SP, whereas the second identified increased functional connectivity during OM > OP. Each PPI analysis was thresholded at $p < .05$, FDR corrected. Conjunction

maps of the overlap between the PPI results for SM > SP and OM > OP were examined to identify connectivity common to both SM and OM within each seed region. To test for any distinct connectivity patterns as a function of either SM or OM, paired sample t tests ([SM > SP] > [OM > OP]) for each seed region were computed and thresholded at $p < .05$, FDR corrected.

Within each of the three shared mentalizing seed regions (vMPFC, PCC, and RTPJ), robust increased functional connectivity was observed during SM > SP and OM > OP in a widely distributed neural system consisting of the medial pre-SMA extending into SMA and cACC (BA 6/24/32), left anterior temporal pole extending into AI, ventrolateral prefrontal cortex, and lateral OFC (BA 47), left FO and dorsal premotor cortex (PMd) and PMv (BA 6/44/48), left SI/MI (BA 3/4), and adjacent left anterior intraparietal sulcus (BA 19) as well as visual cortex (BA 18) extending into the cerebellum. Hippocampal formation connectivity was also increased during SM > SP, but only for the PCC and RTPJ seeds (see Figure 2A and Table 3A–C). Interestingly, connectivity within AI and hippocampal formation during OM > OP was below the FDR threshold among all three seeds.

Table 2. Biases for Self or Other

(A) Self-bias [SM > OM] and [SP > OP]

			SM > OM				SP > OP	
Region	Hemisphere	BA	MNI (x,y,z)	t	Cluster Size	BA	MNI (x,y,z)	t
vMPFC	L	10, 11	−8, 52, −8	5.9	13	10, 11	−6, 50, −8	5.75

(B) Other-bias [OM > SM] and [OP > SP]

			OM > SM				OP > SP	
Region	Hemisphere	BA	MNI (x,y,z)	t	Cluster Size	BA	MNI (x,y,z)	t
PCC	R	29	8, −44, 14	4.56	196	23	10, −60, 32	6.51
	R	23	8, −58, 28	4.43		30	8, −54, 20	4.86
	R					29	6, −42, 14	4.76
TPJ	R	37	50, −60, 14	4.89	55	37	52, −62, 14	4.35

Target-specific biases in processing for self (A) or other (B). All coordinates are reported from the conjunction analysis of the overlap between the SM > OM contrast and the SP > OP contrast (Table 2A) or the conjunction of the OM > SM contrast and OP > SP contrast (Table 2B). Each contrast was thresholded at $p < .05$, FDR corrected.

SM = self-mentalizing; SP = self-physical; OM = other-mentalizing; OP = other-physical; vMPFC = ventromedial prefrontal cortex; PCC = posterior cingulate cortex/precuneus.

Corroborating the idea that the difference in functional connectivity during SM > SP was statistically indistinguishable across the whole brain to the difference in functional connectivity during OM > OP, a paired sample *t* test of [SM > SP] > [OM > OP] revealed no significant results.

Finally, we computed another conjunction map to identify shared mentalizing connectivity that is common to all three seed regions. This was done by first computing a conjunction map of overlap among all seed regions within the SM > SP PPI analyses. The same procedure was carried out for OM > OP PPI analyses. A final conjunction map extracted the overlap between the SM > SP and the OM > OP PPI conjunction maps. Thus, what this final conjunction map reflects is the shared connectivity for both SM and OM that is common to all three seed regions. This conjunction map highlighted pre-SMA, FO, PMv, SI/MI, anterior intraparietal sulcus, and visual cortex/cerebellum as convergent areas of increased functional connectivity during both SM and OM that are common to all three seed regions (see Figure 2B).

Given that the PPI results are a measure of connectivity change between two experimental conditions, a positive PPI result could have been the result of a positive correlation between the seed and a target region for mentalizing whereas the physical condition exhibits a negative or a nonsignificant correlation. However, a positive PPI result could have also resulted from a strong negative correlation in the physical condition and a nonsignificant or less negative correlation in the mentalizing condition. To disambiguate the PPI results, we followed the PPI analyses by extracting the correlations for the seed and the target ROIs for each condition separately (see Methods). Visual examination of the correlations separately for each condition confirmed that each correlation between seed and target ROIs was always more positive in the mentalizing condition compared with the physical condition (see Figure 3A and B).

DISCUSSION

Previous research has emphasized the crucial importance of shared neural representations in demonstrating how we understand others via ourselves (Gazzola & Keysers, 2009; Amodio & Frith, 2006; Decety & Grezes, 2006; Gazzola et al., 2006; Goldman, 2006; Iacoboni & Dapretto, 2006; Mitchell et al., 2006; Saxe et al., 2006; Blakemore et al., 2005; Ochsner et al., 2005; Keysers et al., 2004; Rizzolatti & Craighero, 2004; Singer et al., 2004; Carr et al., 2003; Wicker et al., 2003). The current study provides independent replication of this and also provides many substantial new insights. First, with our factorial design, we were able to test the interaction effect of whether mentalizing or physical representations recruit distinct regions for the self or other. The absence of such an effect in our large sample suggests that mentalizing representations about the self or others are not recruiting distinct and independent neural systems. Instead, we found robust evidence that mentalizing representations are distributed across similar neural systems with respect to self and other.

In the domain of high-level inference-based mentalizing, many of the areas identified in the current study have
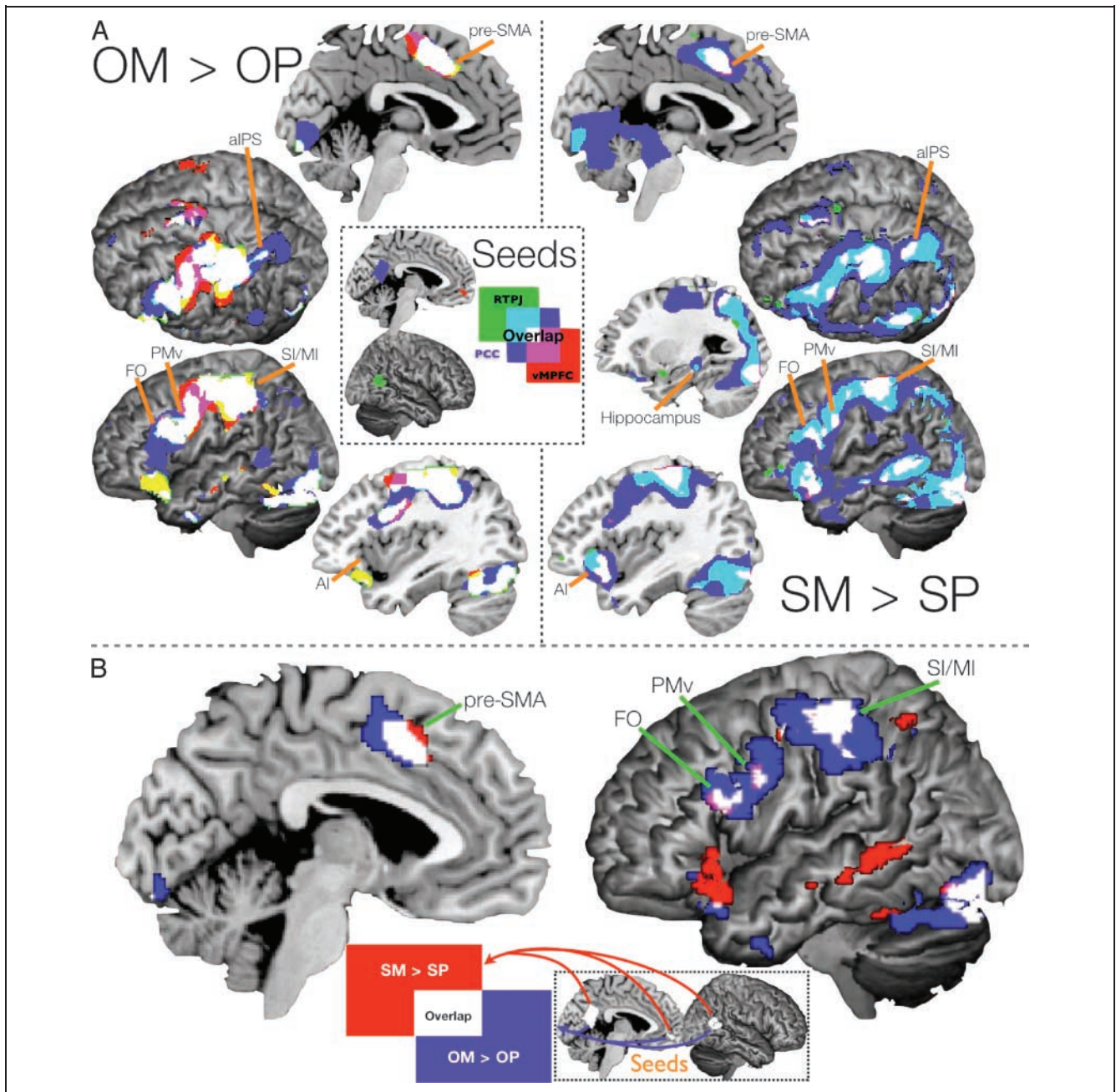
Figure 2. Shared mentalizing connectivity from vMPFC, PCC, and RTPJ. (A, left side) Increases in functional connectivity during OM judgments compared with connectivity during OP judgments (OM > OP) from vMPFC (red voxels), PCC (blue voxels), and RTPJ (green voxels). (A, right side) Same results, but for when connectivity is increased during SM judgments compared with connectivity during SP judgments (SM > SP). Purple and turquoise voxels denote overlapping connectivity among two of three seed regions. White voxels denote connectivity patterns that are common to all three seed regions. (B) Increases in functional connectivity common to all three seed regions for SM compared with connectivity during SP judgments (SM > SP; red voxels) superimposed on top of connectivity increases common to all three seed regions under OM compared with connectivity during OP judgments (OM > OP; blue voxels). White voxels are areas where functional connectivity converges among all three seeds for both SM and OM.

been previously identified as crucial for mentalizing, namely, the vMPFC, PCC, RTPJ, and ATL (Amodio & Frith, 2006; Mitchell et al., 2006; Saxe & Powell, 2006; Saxe et al., 2006; Ochsner et al., 2005; Frith & Frith, 2003; Saxe & Kanwisher, 2003). However, what was particularly interesting and new was the observation that left SI/MI

was also sensitive to mentalizing about both the self and other. The role of somatosensory cortex in low-level shared representations of touch (Blakemore et al., 2005; Keysers et al., 2004), self-experienced pain (Singer et al., 2004), and action–perception mirroring (Gazzola & Keysers, 2009; Gazzola et al., 2006) is well established,

Table 3. Shared Mentalizing Connectivity

Region	Hemisphere	SM > SP			Cluster Size	OM > OP		
		BA	MNI (x,y,z)	t		BA	MNI (x,y,z)	t
(A) vMPFC Connectivity [SM > SP] and [OM > OP]								
Pre-SMA/SMA/cACC	B	32	−8, 14, 42	4.47	431	32	−8, 14, 48	4.9
	B	32	8, 18, 40	3.5				
	B	6	10, 8, 48	3.21				
SI/MI	L	3	−36, −28, 52	3.22	441	3	−46, −22, 52	5.3
	L	3	−48, −24, 54	3.08		4	−38, −22, 60	4.86
	L	3	−40, −30, 60	3.07		3	−38, −34, 50	4.63
FO	L	48	−44, 16, 22	3.02	46	44	−46, 10, 26	4.67
PMv	L	6	−56, 2, 32	2.85	38	4	−56, 0, 30	3.77
	L					6	−48, 0, 32	3.7
PMd	L	6	−32, −6, 46	3.57	91	6	−36, −4, 52	4.43
TP	L	38	−40, 22, −18	3.66	122	38	−42, 26, −18	4.68
	L	38	−50, 26, −14	3.28		38	−52, 28, −12	4.45
Visual cortex/cerebellum	L	18	−28, −84, −14	3.08	383	19	−34, −84, −22	4.02
	L	19	−42, −78, −14	2.71		19	−34, −78, −14	3.93
	L					18	−22, −88, −12	3.81
Visual cortex/cerebellum	R	19	32, −82, −8	3.02	93	18	20, −84, −12	4.39
	R	18	22, −84, −10	2.85		18	26, −86, −18	4.2
	R	19	32, −84, −18	2.57		19	36, −82, −18	3.74
IPS	L	7	−30, −56, 52	3.24	13	7	−32, −54, 48	3.25
(B) PCC Connectivity [SM > SP] and [OM > OP]								
Pre-SMA/SMA/cACC	B	32	−2, 14, 46	5.69	971	32	−4, 8, 48	5.63
	B	32	10, 10, 44	4.04				
	B	6/24	−4, −6, 50	3.92				
SMA	L	6	−4, −8, 66	2.67	18	6	−6, −6, 66	4.79
SI/MI/IPS/PMd/PMv/FO/AI	L	3	−44, −28, 52	7.51	4959	3	−38, −34, 48	6.52
	L	6	−52, 4, 36	6.79		6	−36, −8, 58	6.31
	L	6	−44, −4, 50	5.9		6	−44, 0, 42	6.25
TP	L	38	−44, 18, −18	4.15	42	38	−40, 22, −22	3.58
ATL	L	20	−44, 0, −40	3.71	52	20	−46, 0, −36	4.21
	L	20	−48, 6, −34	3.17				
Visual cortex/cerebellum	B	18	14, −86, −10	7.56	3691	18	−26, −90, −8	6.26
	B	18	−16, −90, 4	6.85		19	−34, −86, −14	6.08
	B	18	−8, −88, −10	6.66		19	−40, −80, 14	6.05
MTG	L	21	−54, −44, 0	6.22	152	21	−58, −52, 4	4.94
IPS	R	19	30, −66, 36	3.99	66	19	34, −66, 36	3.35
dMPFC	L	9	−10, 56, 30	2.84	12	10	−6, 56, 28	3.2

Table 3. (continued)

Region	Hemisphere	SM > SP			Cluster Size	OM > OP		
		BA	MNI (x,y,z)	t		BA	MNI (x,y,z)	t
(C) RTPJ Connectivity [SM > SP] and [OM > OP]								
Pre-SMA/SMA/MCC	B	32	−4, 12, 48	5.57	454	6/32	−4, 2, 52	3.68
SI/MI/PMd/PMv/FO	L	3	−40, −30, 58	5.62	1786	2	−46, −34, 50	4.9
	L	2	−48, −32, 50	4.78		48	−48, 16, 28	4.84
	L	6	−54, 6, 34	4.78		40	−38, −34, 46	4.62
TP/AI/VLPFC	L	38	−44, 20, −12	4.25	496	38	−36, 22, −22	4.15
	L	47	−42, 26, −6	3.95		45	−52, 26, 0	3.15
						38	−46, 30, −12	2.79
Visual cortex/cerebellum/IPS	R	18	10, −86, −6	7.36	5002	18	−8, −88, −12	4.93
	L	19	−34, −82, −20	6.18		18	−20, −88, −16	4.78
	L	18	−14, −88, −18	5.85		18	−26, −90, −10	4.76
MTG	L	21	−52, −44, −2	4.84	269	21	−60, −52, 4	2.94

Shared functional connectivity for mentalizing about the self and other for (A) vMPFC, (B) PCC, and (C) RTPJ seed regions. All coordinates are reported from the conjunction analysis of the overlap between the SM > SP PPI contrast and the OM > OP PPI contrast (*p* > .05, FDR-corrected threshold for each contrast).

SM = self-mentalizing; SP = self-physical; OM = other-mentalizing; OP = other-physical; vMPFC = ventromedial prefrontal cortex; PCC = posterior cingulate cortex/precuneus; pre-SMA = presupplementary motor area; cACC = caudal ACC; dMPFC = dorsomedial prefrontal cortex; SI/MI = primary sensorimotor cortex; VLPFC = ventrolateral prefrontal cortex; FO = frontal operculum; AI = anterior insula; PMv = ventral premotor cortex; PMd = dorsal premotor cortex; ATL = anterior temporal lobe; TP = temporal pole; IPS = intraparietal sulcus; MTG = middle temporal gyrus.

and the disruption of somatosensory cortex (via transcranial magnetic stimulation or lesion studies) impairs emotion recognition (Pitcher, Garrido, Walsh, & Duchaine, 2008; Adolphs, Damasio, Tranel, Cooper, & Damasio, 2000) as well as automatic motor- and somatosensory-evoked potentials when viewing actions or others in pain (Avenanti, Bolognini, Maravita, & Aglioti, 2007; Bufalari, Aprile, Avenanti, Di Russo, & Aglioti, 2007; Avenanti, Bueti, Galati, & Aglioti, 2005). Furthermore, in mirror-touch synesthesia, shared representations of touch are heightened in SI/SII (Blakemore et al., 2005), and such individuals are higher on measures of empathy (Banissy & Ward, 2007). Thus, the current observation that SI/MI is also recruited for mentalizing about self and other suggests that such low-level embodied/simulative shared representations computed by SI/MI are also important for the processes underlying more inference-based mentalizing when compared with reflecting on physical characteristics.

Although the neural implementation of both SM and OM was identical, it cannot be said that the brain is agnostic with regard to distinguishing self from other. On the contrary, the brain does make a general self–other distinction and does so within neural circuitry that is also biased toward mentalizing. Replicating past research (Kelley et al., 2002), we found that the vMPFC was biased for self-referential processing in general, whereas the PCC and the RTPJ were biased for other-referential processing (Pfeifer,

Lieberman, & Dapretto, 2007; Saxe et al., 2006; Ruby & Decety, 2001). However, regarding the other > self effect in the PCC, we note that other studies find the opposite effect of self > other in PCC (D'Argembeau et al., 2008; Kelley et al., 2002). One explanation for these opposing observations may be due to the different task demands from studies that find opposing effects. Studies that find an other > self effect (including the current study) probe event-specific judgments about the self and other (Pfeifer et al., 2007; Ruby & Decety, 2001), whereas more generalized trait inferences about self and other are probed in studies that find a self > other effect (D'Argembeau et al., 2008; Kelley et al., 2002). These subtle differences in task demands may alter the underlying memory retrieval processes that are integral to PCC functioning (Cavanna & Trimble, 2006; Maddock, Garrett, & Buonocore, 2001; Henson, Rugg, Shallice, Josephs, & Dolan, 1999). Future work should systematically explore this interaction between the possible deployment of different retrieval processes in PCC during specific versus general information processing about the self and others.

Although the current study isolated several regions coding for shared mentalizing via standard fMRI contrast analyses, one of the main contributions of the current study is in demonstrating that results from standard fMRI contrast analyses are only the beginning of making such an inference. We have demonstrated that functional

Figure 3. Direction of PPI effects. This figure depicts the relationships among the seeds and each target region shown in Figure 2B for the PPI of (A) SM compared with SP judgments (SM > SP) and (B) OM compared with OP judgments (OM > OP). Visual inspection shows that the mentalizing conditions (white bars) are always greater positive correlations compared with the physical conditions (gray bars). Error bars are not plotted here because statistical inference on these comparisons has already being established from the whole-brain PPI analyses.
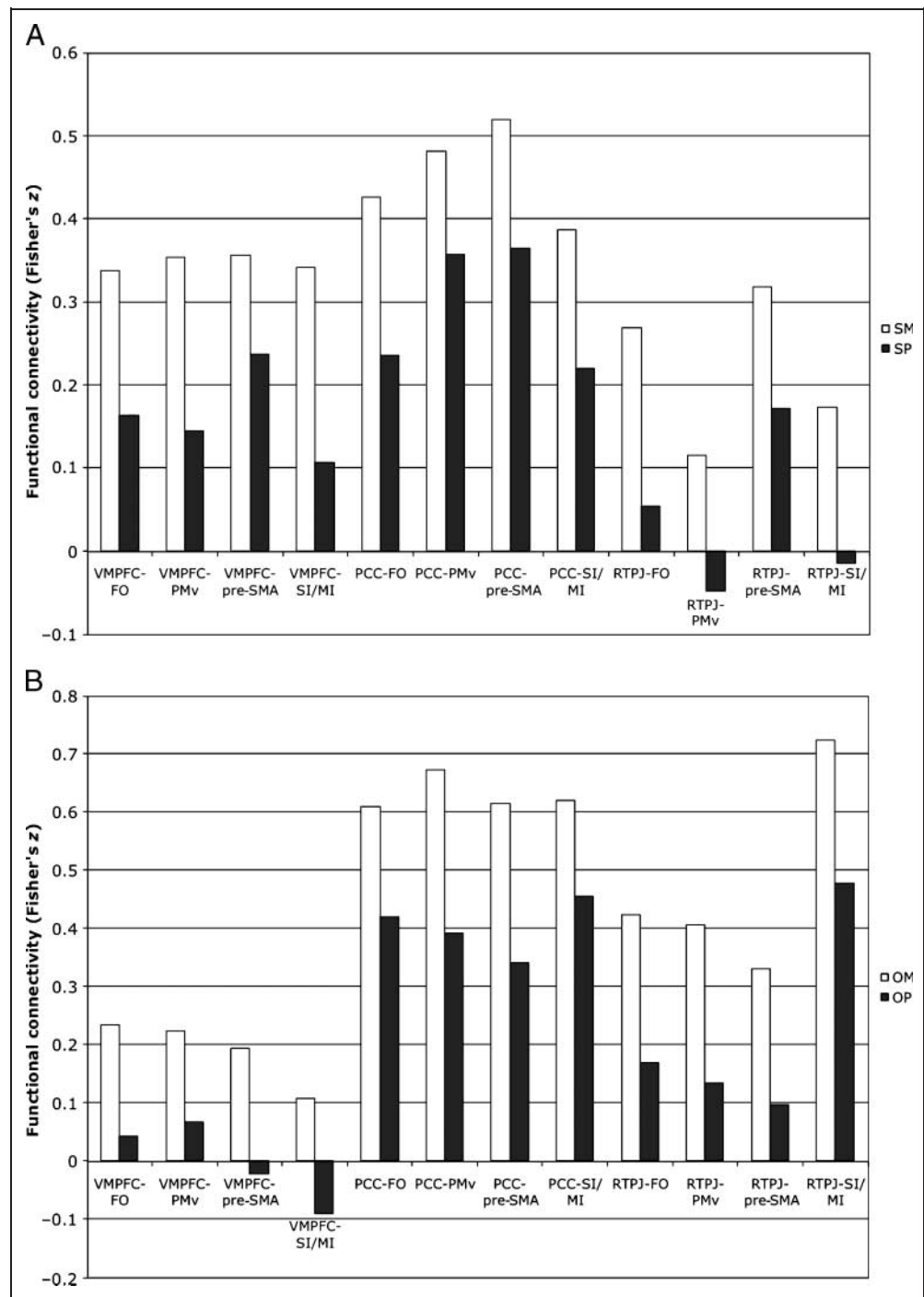
connectivity approaches are necessary for fully qualifying statements that shared neural representations for self and other are in fact participating in the same underlying processes via a similar neural implementation. It is clear from our connectivity analyses that high-level inference-based mentalizing areas such as vMPFC, PCC, and RTPJ are participating in identical neural circuits for both reflecting on one's own mental states and reflecting on another's mental states. Without such insights from connectivity approaches, shared neural representations, while indexing some kind of similarity (e.g., a convergence

zone), might also be participating in functionally distinct neural circuits that may be performing very different types of processes (Jabbi et al., 2008; Zaki et al., 2007).

In addition to demonstrating an identical neural implementation for mentalizing about the self and other, the distribution of where such implementation arose has substantial theoretical implications about the debate between contrasting high-level inference-based models of mentalizing with lower level embodied/simulationist accounts. Historically, scholars have pinned high-level inference-based accounts and low-level embodied/

simulationist accounts as opposites of each other (Gopnik & Wellman, 1992; Gordon, 1992). However, as recent scholars have pointed out (Keysers & Gazzola, 2007; Uddin et al., 2007; Goldman, 2006), perhaps such arguments are missing a third alternative; that is, perhaps higher level inference-based processes are grounded in their interactions with lower level embodied/simulation-based processes (Goldman & de Vignemont, 2009; Barsalou, 1999, 2008). The prediction here would be that high-level inference-based mentalizing systems (e.g., vMPFC, PCC, RTPJ) are integrating their signal with lower level embodied/simulation-based systems (e.g., FO/PMv, IPL, AI, SI/MI, pre-SMA; Keysers & Gazzola, 2007). One piece of evidence supporting this integration hypothesis is based on the earlier observation of the shared activation of SI/MI for SM and OM. In addition, the connectivity analyses went further to directly test the prediction that these two systems were specifically linked during mentalizing more than during physical judgments. The patterns of connectivity strikingly map onto this prediction. Even more striking was that such patterns of connectivity were apparent for both SM and OM. Thus, the patterns of connectivity combined with the identical neural implementation for both SM and OM are the first formal observations of such an integration between two paramount neural systems for social cognition. These results support the idea that during high-level mentalizing, there is a middle ground where both inference-based processes merge with lower level embodied/simulative processes.

Finally, one important point for which the current results have some bearing is on the link between the literature on the default mode network (DMN) of intrinsic functional brain organization (Fox & Raichle, 2007). Although it is tempting to link the activation maps of shared mentalizing results with the DMN, we would suggest two reasons for resisting interpretation of the current findings in relation to the DMN literature. First, as Morcom and Fletcher (2007) point out, the cognitive/psychological nature of what happens in the DMN at rest is unknown, and thus interpretations of well-controlled task-specific differences in signal change (as in the current study) stand relatively independently of what they mean in relation to the unconstrained and unknown cognitive/psychological nature of intrinsic DMN organization.

We also have empirical reasons to resist interpreting the current findings as simply an extension of intrinsic DMN resting state organization. Fox et al. (2005) elegantly showed that the intrinsic connectivity of the DMN is like an internal feedback loop; that is, the time course of signal in MPFC, PCC, and lateral parietal regions is positively correlated with each other. However, Fox et al. extended this by also showing that MPFC, PCC, and lateral parietal regions are negatively correlated with regions similar to those found in our connectivity maps: lateral PMd and PMv, medial SMA/pre-SMA, FO, AI, and intraparietal sulcus. The striking difference here is that whereas the current study's task-specific connectiv-ity analyses observed this same network, we found that connectivity was in the opposite direction (i.e., positive correlations) of what Fox et al. found (i.e., negative correlations). We found that vMPFC, PCC, and RTPJ were positively correlated with a similar network to what Fox et al. found was negatively correlated at rest with MPFC, PCC, and a lateral parietal region near TPJ. If the task-specific shared mentalizing circuits were simply the same network involved in intrinsic resting state organization (with the same kinds of cognitive/psychological interpretations associated with it), we should have observed task-specific increases in mentalizing connectivity from vMPFC, PCC, and RTPJ similar to an internal feedback loop; that is, vMPFC, PCC, and RTPJ should have been positively correlated with each other during SM and OM. Given the observation that this network shows opposing patterns of connectivity between rest and mentalizing, we suggest that the intrinsic mode of functional brain organization is exactly the opposite of what occurs during task-specific shared mentalizing processes for both self and other. Thus, if there are any interpretations to be made about the current findings in relation to the DMN literature, it would be that during mentalizing about the self and other, there is possibly an adaptive reconfiguration of dynamic functional organization from how the brain is naturally functionally organized (Bassett, Meyer-Lindenberg, Achard, Duke, & Bullmore, 2006).

In summary, we used fMRI and a functional connectivity approach to test whether the neural implementation of high-level inference-based mentalizing processes was similar or different with respect to the self and other. Our findings show that such neural implementation is indeed identical across activation contrasts and functional connectivity for SM and OM. Moreover, the tight link between high-level inference-based mentalizing systems and low-level embodied/simulation-based systems suggests that these two neural systems for social cognition are integrated in a task-specific manner for mentalizing about both self and other. These observations provide a first glimpse at how such an integration takes place and provides the groundwork for further theoretical refinement and empirical work as well as translational work into how such systems may be disrupted, as is the case for neurodevelopmental conditions such as autism (Lombardo, Barnes, Wheelwright, & Baron-Cohen, 2007; Dapretto et al., 2006).

Acknowledgments

Competing Interests Statement

E. T. B. is employed half-time by the University of Cambridge and half-time by GlaxoSmithKline plc. None of the other authors have any other biomedical financial interests or potential conflicts of interest. This work was conducted in association with the NIHR-CLAHRC for Cambridgeshire and Peterborough NHS Mental Health Trust.

Reprint requests should be sent to Michael V. Lombardo, Autism Research Centre, Department of Psychiatry, University of Cambridge, Douglas House, 18B Trumpington Road, Cambridge, CB2 8AH, United Kingdom, or via e-mail: ml437@cam.ac.uk.

REFERENCES

Adolphs, R., Damasio, H., Tranel, D., Cooper, G., & Damasio, A. R. (2000). A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *Journal of Neuroscience, 20,* 2683–2690.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7,* 268–277.

Avenanti, A., Bolognini, N., Maravita, A., & Aglioti, S. M. (2007). Somatic and motor components of action simulation. *Current Biology, 17,* 2129–2135.

Avenanti, A., Bueti, D., Galati, G., & Aglioti, S. M. (2005). Transcranial magnetic stimulation highlights the sensorimotor side of empathy for pain. *Nature Neuroscience, 8,* 955–960.

Banissy, M. J., & Ward, J. (2007). Mirror-touch synesthesia is linked with empathy. *Nature Neuroscience, 10,* 815–816.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577–609.

Barsalou, L. W. (2008). Grounded cognition. *Annual Reviews of Psychology, 59,* 617–645.

Bassett, D. S., Meyer-Lindenberg, A., Achard, S., Duke, T., & Bullmore, E. (2006). Adaptive reconfiguration of fractal small-world human brain functional networks. *Proceedings of the National Academy of Sciences, U.S.A., 103,* 19518–19523.

Blakemore, S. J., Bristow, D., Bird, G., Frith, C., & Ward, J. (2005). Somatosensory activations during the observation of touch and a case of vision–touch synaesthesia. *Brain, 128,* 1571–1583.

Bufalari, I., Aprile, T., Avenanti, A., Di Russo, F., & Aglioti, S. M. (2007). Empathy for pain and touch in human somatosensory cortex. *Cerebral Cortex, 17,* 2553–2561.

Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences, U.S.A., 100,* 5497–5502.

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: A review of its functional anatomy and behavioural correlates. *Brain, 129,* 564–583.

Critchley, H. D., Wiens, S., Rotshtein, P., Ohman, A., & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature Neuroscience, 7,* 189–195.

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., et al. (2006). Understanding emotions in others: Mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience, 9,* 28–30.

D'Argembeau, A., Feyers, D., Majerus, S., Collette, F., Van der Linden, M., Maquet, P., et al. (2008). Self-reflection across time: Cortical midline structures differentiate between present and past selves. *Social Cognitive and Affective Neuroscience, 3,* 244–252.

Decety, J., & Grezes, J. (2006). The power of simulation: Imagining one's and other's behavior. *Brain Research, 1079,* 4–14.

Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience, 8,* 700–711.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences, U.S.A., 102,* 9673–9678.

Friston, K. J., Buchel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage, 6,* 218–229.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 358,* 459–473.

Gazzola, V., Aziz-Zadeh, L., & Keysers, C. (2006). Empathy and the somatotopic auditory mirror system in humans. *Current Biology, 16,* 1824–1829.

Gazzola, V., & Keysers, C. (2009). The observation and execution of actions share motor and somatosensory voxels in all tested subjects: Single-subject analyses of unsmoothed fMRI data. *Cerebral Cortex, 19,* 1239–1255.

Goldman, A. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mind reading.* New York: Oxford University Press.

Goldman, A., & de Vignemont, F. (2009). Is social cognition embodied? *Trends in Cognitive Sciences, 13,* 154–159.

Gopnik, A., & Wellman, H. (1992). Why the child's theory of mind really is just a theory. *Mind and Language, 7,* 145–171.

Gordon, R. M. (1992). The simulation theory: Objections and misconceptions. *Mind and Language, 7,* 11–34.

Henson, R. N. A., Rugg, M. D., Shallice, T., Josephs, O., & Dolan, R. J. (1999). Recollection and familiarity in recognition memory: An event-related functional magnetic resonance imaging study. *Journal of Neuroscience, 19,* 3962–3972.

Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience, 7,* 942–951.

Jabbi, M., Bastiaansen, J., & Keysers, C. (2008). A common anterior insula representation of disgust observation, experience, and imagination shows divergent functional connectivity pathways. *PLoS One, 3,* e2939.

Jenkins, A. C., Macrae, C. N., & Mitchell, J. P. (2008). Repetition suppression of ventromedial prefrontal activity during judgments of self and others. *Proceedings of the National Academy of Sciences, U.S.A., 105,* 4507–4512.

Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. *Journal of Cognitive Neuroscience, 14,* 785–794.

Keysers, C., & Gazzola, V. (2007). Integrating simulation and theory of mind: From self to social cognition. *Trends in Cognitive Sciences, 11,* 194–196.

Keysers, C., Wicker, B., Gazzola, V., Anton, J. L., Fogassi, L., & Gallese, V. (2004). A touching sight: SII/PV activation during the observation and experience of touch. *Neuron, 42,* 335–346.

Lombardo, M. V., Barnes, J. L., Wheelwright, S. J., & Baron-Cohen, S. (2007). Self-referential cognition and empathy in autism. *PLoS One, 2,* e883.

Maddock, R. J., Garrett, A. S., & Buonocore, M. H. (2001). Remembering familiar people: The posterior cingulate cortex and autobiographical memory retrieval. *Neuroscience, 104,* 667–676.

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron, 50,* 655–663.

Morcom, A. M., & Fletcher, P. C. (2007). Does the brain have a baseline? Why we should be resisting a rest. *Neuroimage, 37,* 1073–1082.

Ochsner, K. N., Beer, J. S., Robertson, E. R., Cooper, J. C., Gabrieli, J. D., Kihsltrom, J. F., et al. (2005). The neural correlates of direct and reflected self-knowledge. *Neuroimage, 28,* 797–814.

Pfeifer, J. H., Lieberman, M. D., & Dapretto, M. (2007). "I know what you are but what am I?!": Neural bases of self- and social knowledge retrieval in children and adults. *Journal of Cognitive Neuroscience, 19,* 1323–1337.

Pitcher, D., Garrido, L., Walsh, V., & Duchaine, B. C. (2008). Transcranial magnetic stimulation disrupts the perception and embodiment of facial expressions. *Journal of Neuroscience, 28,* 8929–8933.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Reviews of Neuroscience, 27,* 169–192.

Ruby, P., & Decety, J. (2001). Effect of subjective perspective taking during simulation of action: A PET investigation of agency. *Nature Neuroscience, 4,* 546–550.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *Neuroimage, 19,* 1835–1842.

Saxe, R., Moran, J. M., Scholz, J., & Gabrieli, J. (2006). Overlapping and non-overlapping brain regions for theory of mind and self reflection in individual subjects. *Social Cognitive and Affective Neuroscience, 1,* 229–234.

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.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science, 303,* 1157–1162.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87,* 245–251.

Uddin, L. Q., Iacoboni, M., Lange, C., & Keenan, J. P. (2007). The self and social cognition: The role of cortical midline structures and mirror neurons. *Trends in Cognitive Sciences, 11,* 153–157.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in My insula: The common neural basis of seeing and feeling disgust. *Neuron, 40,* 655–664.

Zaki, J., Ochsner, K. N., Hanelin, J., Wager, T. D., & Mackey, S. C. (2007). Different circuits for different pain: Patterns of functional connectivity reveal distinct networks for processing pain in the self and others. *Social Neuroscience, 2,* 276–291.