The Link between Social Cognition and Self-referential Thought in the Medial Prefrontal Cortex

Jason P. Mitchell[1,2], Mahzarin R. Banaji[1], and C. Neil Macrae[2]

Abstract

■ The medial prefrontal cortex (mPFC) has been implicated in seemingly disparate cognitive functions, such as understanding the minds of other people and processing information about the self. This functional overlap would be expected if humans use their own experiences to infer the mental states of others, a basic postulate of simulation theory. Neural activity was measured while participants attended to either the mental or physical aspects of a series of other people. To permit a test of simulation theory's prediction that inferences based on self-reflection should only be made for similar others, targets were subsequently rated for their degree of similarity to self. Parametric analyses revealed a region of the ventral mPFC—previously implicated in self-referencing tasks—in which activity correlated with perceived self/other similarity, but only for mentalizing trials. These results suggest that self-reflection may be used to infer the mental states of others when they are sufficiently similar to self. ■

INTRODUCTION

Recent neuroimaging and neuropsychological research has explored the functional neuroanatomy of social cognition—for instance, by examining brain regions that subserve an understanding of the psychological properties of other people, such as their beliefs, feelings, and personalities. This research has identified the medial prefrontal cortex (mPFC) as a region that supports fundamental aspects of social–cognitive functioning across a wide array of tasks (Blakemore, Winston, & Frith, 2004; Gallagher & Frith, 2003; Frith & Frith, 2001; Adolphs, 1999, 2001), such as judging whether a historical figure would know how to use various objects (Goel, Grafman, Sadato, & Hallett, 1995), making inferences about the mental states of characters in stories or cartoons (Gregory et al., 2002; Gallagher, Happé, et al., 2000; Stone, Baron-Cohen, & Knight, 1998; Fletcher et al., 1995), playing interactive games that require second-guessing an opponent (Gallagher, Jack, Roepstorff, & Frith, 2002; McCabe, Houser, Ryan, Smith, & Trouard, 2001), judging the characteristics of others (Mason, Banfield, & Macrae, 2004; Mitchell, Heatherton, & Macrae, 2002), and encoding information about another's personality (Mitchell, Macrae, & Banaji, 2004).

Despite these repeated observations that the mPFC contributes critically to core aspects of social cognition, surprisingly little is known about the precise functional role that this region plays in the human capacity to understand the minds of others. In part, characterization of mPFC functioning has been complicated by the observation that, together with its role in inferring the mental states of others, this region has also been associated with tasks that require people to reflect on, or introspect about, their own inner mental states. For example, activity in the ventral mPFC has been observed during tasks in which participants report on their own personalities or preferences (Schmitz, Kawahara-Baccus, & Johnson, 2004; Johnson et al., 2002; Kelley et al., 2002; Zysset, Huber, Ferstl, & von Cramon, 2002), adopt a first-person perspective (Vogeley et al., 2004), or reflect on their current affective state (Gusnard, Akbudak, Shulman, & Raichle, 2001), as well as in the memory advantage that emerges when items are encoded in a self-relevant manner (Macrae, Moran, Heatherton, Banfield, & Kelley, 2004).

At first glance, the observation that the mPFC subserves seemingly discrete cognitive functions of mental state attribution and self-reflection poses something of a conundrum. In actual fact, however, such an overlap may be expected if perceivers use their own experience to predict or understand the mental states of others (Gallagher & Frith, 2003; Frith & Frith, 2001). Indeed, some influential theoretical accounts of social–cognitive functioning have suggested just such a possibility. Broadly known as ''simulation'' theory, these accounts posit that one powerful strategy for inferring the mental states of other people is to imagine one's own thoughts,

[1]Harvard University, [2]Dartmouth College

feelings, or behaviors in a similar situation (Adolphs, 2002; Meltzoff & Brooks, 2001; Nickerson, 1999; Gallese & Goldman, 1998; Davies & Stone, 1995a, 1995b; Gordon, 1992; Heal, 1986). Proponents of simulation theory point out that, although never enjoying direct access to the internal workings of another person's mind, one does have continuous, first-hand experience of a highly similar system—namely, one's own mind. Accordingly, one may use self-reflection as a tool to understand or predict the mental states of others, at least under certain conditions. Empirical support for the notion that people frequently use their own experience as a basis for inferring the minds of others (either consciously or unconsciously) comes from demonstrations that perceivers routinely overestimate what others know based on what they themselves know (Fussell & Krauss, 1992; Griffin & Ross, 1991) and see their own beliefs and opinions as representative of other people in general, the so-called false consensus effect (Nickerson, 1999; Ross, Greene, & House, 1977).

Could such simulation accounts help explain the overlap between social–cognitive and self-referential processing in the mPFC? In the current investigation, we tested two empirical predictions derived from the hypothesis that regions of the mPFC contribute to understanding the mental states of other people through the implementation of self-reflective processing. First, although perceivers may introspect about their own mental states when trying to figure out what another person is thinking or feeling, such a strategy would not be useful when making judgments about other, nonmentalistic aspects of other people, such as judgments about their appearance or physical location. That is, regions of the mPFC should be selectively engaged during social–cognitive tasks that prompt attempts to apprehend the mental states of another person, but not during equally challenging tasks that do not require understanding another's mind (Gallagher, Jack, et al., 2002; McCabe et al., 2001).

Second, simulation accounts of mental state attribution suggest that perceivers only use self-reflection as a strategy to predict the mental states of others when these individuals are in some way similar to self. In one of the first theoretical formulations of the simulation hypothesis, Heal (1986) pointed out that, in order to justify simulating the minds of other people, one must make, ''one simple assumption,'' namely, ''that they are like me in being thinkers, that they possess the same fundamental cognitive capacities and propensities that I do'' (p. 137). That is, one can successfully use self-reflection to provide insight into the internal states of another person only to the extent that one's own beliefs, feelings, and behaviors are deemed to be applicable or relevant to the individual in question. When this condition is not satisfied (i.e., self and other are dissimilar), one is less likely to rely on self-reflection to understand the other person. Of course, if such a strategy is in operation, part of the neural system that supports social cognition must be sensitive to the perceived similarity between self and other in order to modulate self-reflective processing. Against the backdrop of earlier neuroimaging research on the neural basis of self-reflection (Macrae et al., 2004; Schmitz et al., 2004; Vogeley et al., 2004; Johnson et al., 2002; Kelley et al., 2002; Zysset et al., 2002; Gusnard et al., 2001), we expect that the ventral mPFC may serve precisely this function during tasks that require understanding the mental states of other people.

In the current experiment, participants underwent fMRI scanning while judging either the mental or physical aspects of a series of faces. Specifically, participants either judged how pleased the target person was to have his or her photograph taken (mentalizing task) or how symmetrical the face appeared (nonmentalizing task). After scanning, participants viewed each photograph again and indicated the degree to which they perceived the other person to be similar to themselves. These ratings enabled trials to be retroactively conditionalized on the basis of perceived similarity, thereby allowing the identification of brain regions in which the hemodynamic response correlated with self/other similarity as a function of the processing goals of the initial orienting task (i.e., mentalizing or nonmentalizing). Overall, we expected to observe greater mPFC engagement during mentalizing trials compared with nonmentalizing trials (Gallagher & Frith, 2003; Frith & Frith, 2001). In addition, to the extent that self-reflection guides the understanding of others, we expected greater ventral mPFC engagement for targets that were identified as similar compared to dissimilar to self, but only on the mentalizing trials.

RESULTS

Behavioral Data

Mean similarity ratings were comparable for targets that were initially encountered during mentalizing ($M = 1.98$) and nonmentalizing ($M = 1.99$) trials, $t(17) < 1$, *ns*. Table 1 displays the distribution of similarity ratings across the two orienting tasks. Participants were more likely to rate a target as dissimilar than similar, as evidenced by a significant decrease in the proportion of items across increasing levels of similarity [$F(3,51) = 9.64$, $p < .0001$]. Moreover, this trend was comparable across both mentalizing and nonmentalizing tasks [Orienting task × Similarity interaction: $F(3,51) = 1.17$, *ns*]. In addition, across both orienting tasks, participants did not consider same-sex targets to be more similar to self than other-sex targets: no main effect of target sex (same sex as participant vs. other sex as participant) was observed and target sex did not interact with orienting task (both Fs < 1.90, *ns*). In other words, mean similarity ratings did not differ significantly across

Table 1. Mean Proportion (and Standard Deviation) of Mentalizing and Nonmentalizing Trials as a Function of Subsequent Similarity Rating

	Similarity Rating			
Task	*1*	*2*	*3*	*4*
Mentalizing	0.35 (0.27)	0.37 (0.18)	0.21 (0.11)	0.06 (0.12)
Nonmentalizing	0.34 (0.25)	0.40 (0.18)	0.21 (0.11)	0.06 (0.09)

Subjects judged the similarity of targets on a 4-point scale anchored by 1 = very dissimilar from me; 2 = somewhat dissimilar from me; 3 = somewhat similar to me; 4 = very similar to me.

mentalizing same-sex (M = 2.50), mentalizing other-sex (M = 2.44), nonmentalizing same-sex (M = 2.44), or nonmentalizing other-sex trials (M = 2.39). Moreover, none of the fMRI results were qualified by the match between sex of participant and sex of target.

fMRI data

We adopted several complementary analytic strategies to examine differences in neural activation across the orienting tasks. First, we directly contrasted mentalizing and nonmentalizing trials, regardless of target similarity. Consistent with earlier research, the contrast of *mentalizing* > *nonmentalizing* yielded a distributed set of brain regions that have been associated with mental state attribution, including the dorsal aspect of the mPFC (Figure 1A), lateral parietal cortex regions in both hemispheres that included the temporo-parietal junction, the right superior temporal sulcus, and the left amygdala extending into the anterior hippocampus (see Table 2). In contrast, *nonmentalizing* > *mentalizing* yielded loci of activation that were restricted to posterior brain regions, including the bilateral inferior temporal cortex, the inferior parietal gyrus, and the occipital cortex (see Table 3).

Second, to examine how neural activity varied as a function of perceived self/other similarity, we examined brain regions in which the BOLD signal correlated with subsequent similarity ratings; that is, regions in which greater neural activity was observed during the processing of targets that a participant rated as similar to self (see Methods section for details). The relation between BOLD signal and similarity ratings was stronger for mentalizing than for nonmentalizing trials at a single locus, located in the ventral mPFC directly anterior to the genu of the corpus callosum (x = 9, y = 57, z = 3). As displayed in the rightmost column of Figure 1B,

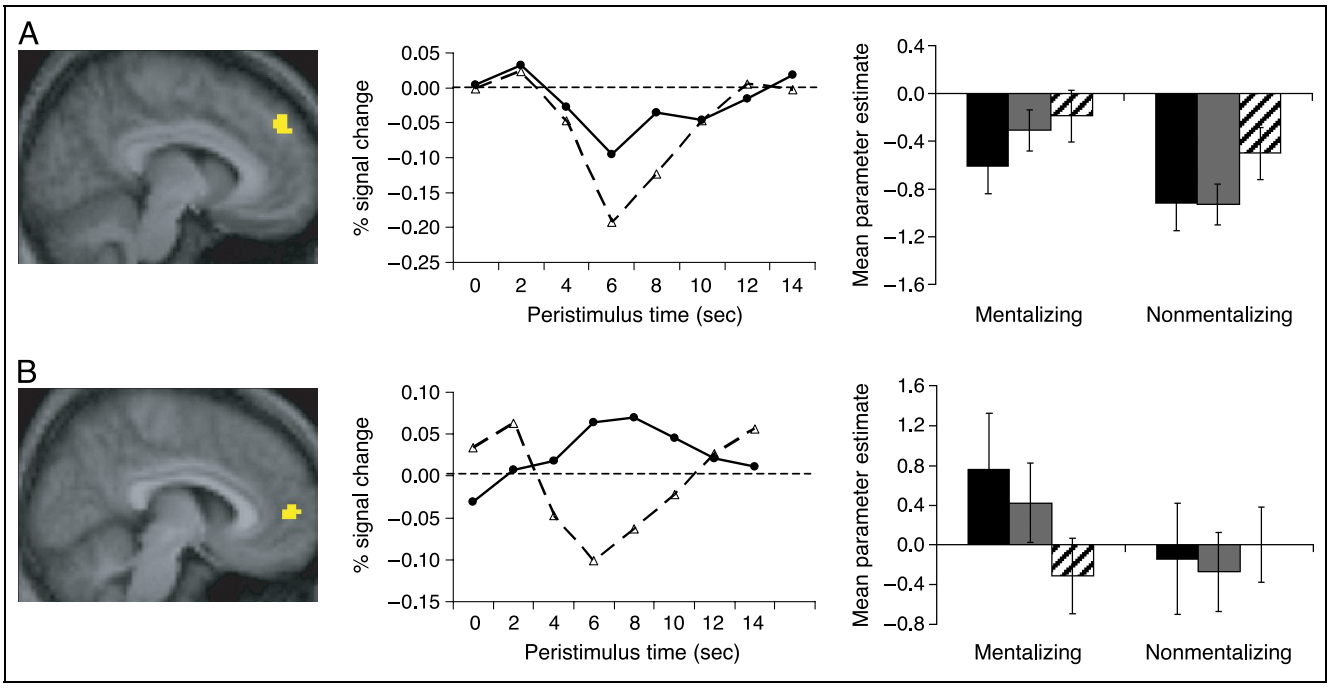

Figure 1. Two mPFC regions were identified in the current study. First, the contrast of *mentalizing* > *nonmentalizing* yielded a region of the dorsal mPFC. (A) displays this region on a sagittal (x = −9) slice of participants' mean normalized brain. The middle section of the panel displays hemodynamic timecourses extracted from this dorsal mPFC region, representing the BOLD response associated with mentalizing (solid circles) and nonmentalizing (dashed open triangles) trials. Second, parametric analyses of fMRI data identified a region of the ventral mPFC (x = 9, y = 57, z = 3) in which BOLD signal correlated with participants' subsequent similarity ratings for targets in the mentalizing task, but not in the nonmentalizing task. (B) displays this region on a sagittal (x = 6) slice of participants' mean normalized brain. For both mPFC regions, the rightmost graphs display parameter estimates obtained for mentalizing and nonmentalizing trials at three levels of similarity: high (rating of 3 or 4; leftmost solid black bars), moderate (rating of 2; middle solid gray bars), and low (rating of 1; rightmost striped bars). Error bars represent the standard error of the mean.

Table 2. Peak Voxel and Number of Voxels for Regions of Interest Obtained from the Contrast of *Mentalizing* > *Nonmentalizing* ($p < .05$, corrected)

Region	*Hemisphere*	*x*	*y*	*z*	*Max t*	*Voxels*
Dorsal mPFC	L	−9	51	36	5.09	30
Inf. frontal gyrus	L	−45	12	18	6.04	151
Mid. frontal gyrus	L	−36	6	51	5.72	26
Amygdala	L	−24	−3	−27	5.60	41
Sup. temporal sulcus	R	45	−6	−15	6.59	110
Insula	L	−30	−18	6	5.85	60
Cingulate sulcus	–	0	−21	48	5.54	147
Postcentral gyrus	L	−24	−42	69	5.13	63
	R	24	−57	69	6.54	107
Parietal cortex	L	−51	−48	3	8.34	783
	R	57	−51	9	10.62	849
Occipital cortex	L	−9	−102	12	9.75	477
	R	9	−96	15	8.50	588

t tests reflect the statistical difference between the two conditions, as computed by SPM99. Coordinates refer to the Montreal Neurological Institute stereotaxic space. Inf. = inferior; Mid. = middle; Sup. = superior.

whereas activity in this region was linearly modulated by target similarity for mentalizing trials (highest for the most similar targets, lowest for the most dissimilar targets), no effect of similarity was observed on nonmentalizing trials. Repeated-measures analysis of variance demonstrated a significant 2 (orienting task) × 3 (level of similarity) interaction in the ventral mPFC [$F(2,34) = 4.37$, $p < .03$]. Confirming that the effect of target similarity was restricted to the mentalizing task, analysis of the simple main effects demonstrated a significant difference across level of similarity for mentalizing trials [$F(2,34) = 5.08$, $p < .02$], but not for nonmentalizing trials [$F(2,34) = 0.21$, *ns*]. In addition, activity in this region was marginally greater for mentalizing than for nonmentalizing trials ($p < .07$, one-tailed). In no brain region was the relation between BOLD signal and target similarity stronger for nonmentalizing than for mentalizing trials.

Finally, the self/other similarity effects observed in the ventral mPFC were not mirrored in the dorsal mPFC. Instead, the pattern of BOLD response in the dorsal mPFC for mentalizing trials demonstrated a trend toward an inverse correlation with similarity ratings: mentalizing about dissimilar others was associated with the highest response in this region, whereas mentalizing about similar others was associated with the lowest response in this region. Although this inverse correlation only reached marginal significance in the dorsal mPFC [$F(2,34) = 1.77$, $p < .10$, one-tailed], this pattern did differ significantly from that observed in the ventral mPFC for mentalizing trials [Similarity × Region interaction: $F(2,34) = 7.87$, $p < .002$].

DISCUSSION

Judgments about the mental state of another person were associated with a distributed set of brain regions that overlap considerably with those observed in earlier studies that have examined the functional neuroanatomy of social cognition. As reviewed above, the dorsal mPFC has been implicated in a wide range of tasks that require understanding the mental states of others (Blakemore et al., 2004; Mason et al., 2004; Mitchell, Macrae, et al., 2004; Gallagher & Frith, 2003; Gallagher, Jack, et al., 2002; Mitchell, Heatherton, et al., 2002; Frith & Frith, 2001; McCabe et al., 2001; Adolphs, 1999, 2001; Gallagher, Happé, et al., 2000; Fletcher et al., 1995; Goel et al., 1995). Likewise, a number of other regions observed in the current study are thought to contribute importantly to various aspects of social cognition, including the superior temporal sulcus (biological motion, gaze detection) (Allison, Puce, & McCarthy, 2000; Puce, Allison, Bentin, Gore, & McCarthy, 1998), the amygdala (emotional processing) (Adolphs, 2002; Morris et al., 1996), and the temporo-parietal junction (which has been linked to the representation of others' beliefs) (Samson, Apperly, Chiavarino, & Humphreys, 2004; Saxe & Kanwisher, 2003). Activation of these regions was modulated by the relative mentalizing demands of the orienting tasks across a common set of stimulus faces. Specifically, activity indicative of social–cognitive processing was only observed when participants were induced to consider the targets as mental agents. When the initial orienting task instead encouraged participants to view the targets in a nonmentalistic manner, areas of the brain associated with object-based processing were differentially activated (e.g., inferotemporal cortex). These results suggest that the set of brain regions

Table 3. Peak Voxel and Number of Voxels for Regions of Interest Obtained from the Contrast of *Nonmentalizing* > *Mentalizing* ($p < .05$, corrected)

Region	*Hemisphere*	*x*	*y*	*z*	*Max t*	*Voxels*
Inf. temporal gyrus	L	−39	−63	−3	5.70	74
	R	51	−57	−12	6.48	128
Inf. parietal gyrus	L	−24	−69	39	5.76	37
Occipital cortex	L	−30	−87	6	6.64	97
	R	36	−87	0	8.53	507

Inf. = inferior.

associated with the current mentalizing task (including the dorsal mPFC) may specifically implement perceivers' attempt to understand the behavior, attributes, and proclivities of others in an agentic manner, but that these regions do not respond globally to all tasks that entail person processing (i.e., when one's task does not require mental state attribution).

When fMRI analyses were further conditionalized on the basis of postscanning ratings of self/other similarity, a correlation between activity in the ventral mPFC and ratings of similarity was observed. Critically, however, this correlation only emerged for mentalizing trials. Of theoretical importance, the peak of this ventral mPFC activation was remarkably similar to the coordinates reported by a number of earlier studies that have examined the neural basis of self-referential processing (Macrae et al., 2004; Schmitz et al., 2004; Vogeley et al., 2004; Johnson et al., 2002; Kelley et al., 2002; Zysset et al., 2002; Gusnard et al., 2001). Table 4 lists the ventral mPFC coordinates reported in a number of such experiments in which subjects have been asked to engage in self-referencing tasks that require reporting their personality characteristics (e.g., judging how well they are described by the word ''curious'') (Macrae et al., 2004; Schmitz et al., 2004; Johnson et al., 2002; Kelley et al., 2002; Zysset et al., 2002); tasks that contrast first- and third-person perspectives (Vogeley et al., 2004); or tasks that require reporting on their current emotions (Gusnard et al., 2001). On average, these peak coordinates were approximately 4 voxels from the peak ventral mPFC activation observed in the current study (mean Euclidean distance = 4.1 voxels, range = 1.1–6.9); by comparison, the distance between the activation peaks of the ventral and dorsal mPFC regions we observed was greater than 12 voxels. Thus, despite considerable differences between the explicit self-referencing paradigms that have been used in earlier research and the current mentalizing task (e.g., no mention of similarity was made in the current experiment until after scanning), both sets of studies observed activity in highly overlapping regions of the ventral mPFC.

Table 4. Ventral mPFC Regions Observed in Earlier Studies of Self-referencing

Study	*Comparison*	*x*	*y*	*z*
Gusnard et al. (2001)	analysis of scenes: emotional > perceptual	−3	42	11*†
Johnson et al. (2002)	self-referencing > semantic analysis	0	54	8
Kelley et al. (2002)	self-referencing > judgments of others	10	52	2
Macrae et al. (2004)	self-referencing: memory hits > misses	0	50	8
Schmitz et al. (2004)	self-referencing > semantic analysis	6	56	4
Vogeley et al. (2004)	first-person > third-person perspective	−2	59	10†
Zysett et al. (2002)	self-referencing > semantic analysis	−6	56	17†

*Ventral-most extent of a number of mPFC loci obtained from the comparison of judging whether a photograph made one feel pleasant or unpleasant versus indicating whether a photograph depicted an indoor or outdoor scene.

†Coordinates transformed from the atlas space of Talairach and Tournoux (1988) to MNI space.

Why then does the task of thinking about others modulate cortical areas more commonly associated with self-reflection? Simulation theory may provide some preliminary answers to this question. Although formulations of the simulation approach have generally advanced few concrete empirical predictions, one important corollary of this viewpoint is that simulation is only appropriate—and may therefore only be attempted—when one believes oneself to be an appropriate model from which to understand another's mind (i.e., when another person is believed to be sufficiently similar to oneself). The close overlap between the ventral region of the mPFC observed in the current investigation and those previously reported during self-referential processing is therefore consistent with two predictions made by simulation theory: (i) people sometimes use self-knowledge to infer the mental states of others and (ii) the extent of this simulation is dependent on the degree to which self and other are perceived to be similar. Importantly, because activity in the ventral mPFC did not correlate with similarity ratings during a nonmentalizing task (i.e., symmetry judgments), these findings demonstrate that the ventral mPFC is not ubiquitously sensitive to self/other similarity. Instead, similarity appears to be especially relevant to tasks that require perceivers to understand the mind of another person.

In the current study, two distinct regions of the mPFC were associated with mentalizing. A dorsal region, similar to the mPFC loci reported in several earlier investigations of social cognition (Mitchell, Macrae, et al., 2004; Gallagher, Jack, et al., 2002; Mitchell, Heatherton, et al., 2002; Goel et al., 1995), was differentially engaged during mentalizing trials compared with nonmentalizing trials. In addition, a ventral aspect of the mPFC, overlapping with loci reported in earlier studies of self-referential processing (Macrae et al., 2004; Schmitz et al., 2004; Vogeley et al., 2004; Johnson et al., 2002; Kelley et al., 2002; Zysset et al., 2002; Gusnard et al., 2001), distinguished between mentalizing and nonmentalizing trials as a function of target similarity, but was only marginally more engaged by mentalizing than nonmentalizing trials. Of course, because overall task comparisons included all trials regardless of perceived self/other similarity (and most targets were judged by participants to be relatively dissimilar), it is unsurprising that this ventral region of the mPFC was not obtained

from the direct contrast of *mentalizing* > *nonmentalizing* (at our a priori statistical threshold). However, when trials were segregated on the basis of perceived self/other similarity, activity in the ventral mPFC was indeed greater for mentalizing than for nonmentalizing trials (Figure 1B, middle column), confirming its role in social–cognitive processing.

Although the current results suggest that ventral aspects of the mPFC contribute to the simulation of other minds via self-reflection, little is known about the role that such dorsal aspects of the mPFC play in social–cognitive processing. The self/other similarity effects observed in the ventral mPFC were not mirrored in dorsal regions of the mPFC; indeed, the dorsal mPFC demonstrated a trend towards an inverse correlation with similarity ratings, such that mentalizing about dissimilar others was associated with the highest response in this region, whereas mentalizing about similar others was associated with the lowest response in this region. Although this inverse correlation only reached marginal significance in the dorsal mPFC and should therefore be interpreted with caution, we note that the dorsal mPFC response during mentalizing trials as a function of similarity differed significantly from that in the ventral PFC (as evidenced by the significant Region × Similarity interaction, detailed above), suggesting a possible dissociation between the social–cognitive contributions of these two subregions of the mPFC. We speculate that, whereas the ventral mPFC may guide the understanding of others' mental states through contemplation of one's own, the dorsal mPFC may instead instantiate more universally applicable social–cognitive processes that can aid mentalizing even when simulation is inappropriate (e.g., for dissimilar others). Indeed, the fact that earlier studies have almost exclusively asked perceivers to mentalize about highly dissimilar others (e.g., Christopher Columbus, cartoon figures) or about an unspecified person (e.g., an unseen opponent) may be one reason that the dorsal mPFC has been associated with social–cognitive tasks more frequently than ventral subregions of the mPFC. One goal for future research should be to further delineate the distinct cognitive operations implemented by various subregions of the mPFC during attempts to infer the mental states of other people.

One intriguing, but poorly understood, feature of activity in both the dorsal and ventral mPFC consists of the ''direction'' of change typically observed in these regions. As in the current study, modulations in the mPFC frequently occur as negative deflections in activity, or ''deactivations'' from the resting baseline state (Gusnard & Raichle, 2001; Shulman et al., 1997). Although relatively little is understood about their functional significance, deactivations from baseline typically occur in those cortical regions with the highest resting metabolic rates (Raichle et al., 2001), prompting some observers to suggest that such negative deflections represent suspension from a default state of cognitive processing that includes self-reflection (Gusnard & Raichle, 2001). Interestingly, although activations above resting baseline have occasionally been reported in dorsal aspects of the mPFC (Mitchell, Macrae, et al., 2004; Gusnard et al., 2001; Gusnard & Raichle, 2001), the current study provides among the first observations of true activations in the ventral mPFC (when mentalizing about highly similar others), suggesting that activity in this region may indeed increase over baseline when social–cognitive processing is applied to people believed to be similar to oneself.

As little is currently known about which particular aspects of similarity determine the extent to which an individual will simulate the experience of another person (Gordon, 1992; Heal, 1986), self/other similarity was indexed in an open-ended manner in the current study (i.e., the dimension along which to make similarity judgments was not specified). Presumably, the extent to which a perceiver uses simulation to infer another person's mental states will depend on how ''like-minded'' the other person is considered to be (e.g., ''does s/he think like me?''). This representation of conceptual similarity may include abstract information about the person, such as knowing that she was raised in the same country or has a similar educational background as oneself. However, in the absence of such semantic knowledge about a person, perceivers may rely on other potentially relevant dimensions—such as physical similarity with the target—as a cue to or proxy for like-mindedness. Although the results of the current study provide evidence that perceived self/other similarity modulates the neural activity associated with mentalizing, additional research is required to clarify exactly which dimensions of similarity are critical to the emergence of this effect.

By providing evidence for one important prediction of simulation theory—that understanding the mental states of similar others can draw on self-reflection—the current results contribute to an emerging literature providing empirical support for simulation accounts of mental state attribution. Most notably, research on so-called mirror neurons has repeatedly demonstrated that perceiving the actions of others shares a neural basis with actually performing those behaviors, both at the level of analysis provided by human neuroimaging studies as well as at the level of single neurons (Rizzolatti & Craighero, 2004). Recently, Wicker et al. (2003) have suggested that the correspondence between perception and experience extends to some forms of affective reactions by demonstrating that identical subregions of the human insula are selectively engaged both by feeling disgust and seeing others experience the same emotion. Some observers have suggested that this coupling of mere perception of others to actual, first-person experience may support a rudimentary system for simulating the minds of other people (Gallese & Goldman, 1998).

Simulation theory has frequently been portrayed as mutually exclusive with ''theory–theory'' accounts of mental state attribution, the notion that an understanding of other people arises from combinatorial processing of basic social rules [either akin to the way natural language emerges from the operation of a grammatical ''rule'' (Stich & Nichols, 1992) or through explicit hypothesis testing and theorizing during development (Gopnik & Meltzoff, 1996; Gopnik & Wellman, 1992; Perner, 1991)]. Our own view is that such either–or theorizing may obscure an important aspect of social thought; namely, that making inferences about the mind of another person is a complex, shifting problem, the solution of which depends on the specific information that is available to perceivers in any given situation. For example, although one can infer transient shifts in another's mental states using a host of visual and auditory cues during one-on-one interactions (e.g., facial expressions, body posture, speech prosody, etc.), one must solve the same problem solely on the basis of auditory input when conversing on the telephone. Quite how one goes about drawing inferences about the mental states of other people will therefore be determined in part by the availability of relevant cues. Indeed, the observation in the current study that mentalizing was associated with multiple brain regions—only one of which was also sensitive to self/other similarity—suggests that inferring the mental states of other people may make use of a variety of different cognitive processes. Although a substantive treatment of this issue is well beyond the scope of the current article, we see no reason that support for simulation accounts rules out the possibility that, in the right circumstance, mental state attribution may also be guided by other types of information processing, including rule-based processing (Heal, 1996; Perner, 1996).

Moreover, we note that the effects of perceived self/other similarity on mentalizing may not be inconsistent with all theory–theory views. Indeed, some formulations of the theory–theory account do not insist that rule-based processing of social information be cognitively impenetrable (as would be the case for rule-based processing of linguistic information in language). As such, these versions of theory–theory (e.g., Botterhill, 1996; Gopnik & Meltzoff, 1996; Stone & Davies, 1996; Gopnik & Wellman, 1992; Perner, 1991) leave open the possibility that knowledge about the self could provide a useful basis for theorizing about another person. For example, such views might hypothesize that the quality of the prediction of another person's behavior will depend on the accuracy of the particular content (relevant beliefs, desires, etc.) that enters the theorizing process, and that this content may be based more strongly upon information about the self when the other person is perceived to be similar, than when the other is perceived to be dissimilar. We acknowledge that the results of the current study cannot arbitrate between simulation accounts of mentalizing and this account of theory–theory.

Finally, physical judgments of faces (''how symmetrical is this face?'') were differentially associated with activation in a set of brain regions that included the bilateral inferotemporal cortex. Similar inferotemporal regions have been implicated in a range of tasks that require semantic knowledge about the attributes of nonsocial stimuli, such as naming pictures of inanimate objects (Chao, Haxby, & Martin, 1999; Martin, Wiggs, Ungerleider, & Haxby, 1996). However, subsequent neuroimaging research has suggested that accessing comparable semantic knowledge about the psychological terms used to describe other people does not engage this region (Mason et al., 2004; Mitchell, Heatherton, et al., 2002). The current results extend this work by suggesting that the inferotemporal cortex may in fact subserve some judgments of other people, at least those that require attention to the physical aspects of a person. In other words, the perception of other people does not invariably engage the brain systems involved in social cognition; rather, such processing can be suspended when one's judgment of another person does not involve mental state attribution.

The ability to explain and predict other people's behavior through mental state attribution lies at the very heart of human social cognition. Extending extant work on this topic and consistent with the predictions of simulation theory, the current investigation suggests that self-reflection may represent one strategy for understanding the minds of others (either consciously or unconsciously), at least when these individuals are perceived to be similar to self. By demonstrating that part of the system for understanding others draws predictably on a system believed to play an important role in understanding oneself, these data also address the seemingly paradoxical observation that subregions of the mPFC subserve both self-reflection and mentalizing about others, suggesting that the ventral mPFC may instantiate the self-reflective processing that guides an agentic interpretation of others.

METHODS

Participants

Participants were 18 (11 women) right-handed, native English speakers with no history of neurological problems (mean age, 20 years; range, 17–25). Informed consent was obtained in a manner approved by the Committee for the Protection of Human Subjects at Dartmouth College.

Stimuli and Behavioral Procedure

Stimuli consisted of 480 black-and-white photographs of faces collected from a number of publicly available face

databases. All targets were Caucasian adults (260 women, 220 men), photographed displaying a neutral facial expression. Images were resized to a width of 7.6 cm and height of 8.4 cm. During the acquisition of fMRI scans, a randomly selected subset of 240 faces were each presented once for 3 sec. Each face was immediately preceded by a 1-sec presentation of one of two cues (''How pleased?'' or ''How symmetrical?'') that indicated, respectively, whether the mentalizing or nonmentalizing task was to be performed on that trial. For mentalizing trials, participants were instructed to judge how pleased the person in the photograph seemed to be to have his or her photograph taken. For nonmentalizing trials, participants were instructed to judge how symmetrical each face appeared. Participants used a 4-point scale for both types of judgment. For each participant, half the faces were randomly assigned to the mentalizing condition and the remaining half to the nonmentalizing condition. To optimize estimation of the event-related fMRI response, trials were intermixed in a pseudorandom order and separated by a variable interstimulus interval (500–7500 msec) (Dale, 1999), during which participants passively viewed a fixation crosshair.

Approximately 30 min after completing the last functional run, participants performed a similarity-rating task. During the similarity-rating task, each of the faces was presented again, and participants were asked to use a 4-point scale to judge how similar they believed each target was to themselves (1 = very dissimilar to me; 2 = somewhat dissimilar to me; 3 = somewhat similar to me; 4 = very similar to me). Participants were not explicitly instructed about the dimension on which to base their similarity judgments.

Imaging Procedure

Imaging was conducted using a 1.5-T GE Signa scanner. Functional scanning used a gradient-echo, echo-planar pulse sequence (TR, 2 sec; TE, 35 msec; 3.75 × 3.75 in-plane resolution). Participants completed three functional runs of 200 acquisitions (25 axial slices; 4.5 mm thick; 1 mm skip). Stimuli were projected onto a screen at the end of the magnet bore that participants viewed by way of a mirror mounted on the head coil. A pillow and foam cushions were placed inside the head coil to minimize head movements.

SPM99 software (Wellcome Department of Cognitive Neurology, London, UK) was used for slice timing and motion correction, normalization to the MN1305 stereotactic space (interpolating to 3 mm^3 voxels), and spatial smoothing (8-mm Gaussian kernel). Statistical analyses were performed using the general linear model in which the event-related design was modeled using a canonical hemodynamic response function and covariates of no interest (a session mean and a linear trend). In addition, for each trial, the value of the participant's subsequent similarity rating was included as a linear parametric modulator. That is, each trial was coded by both trial type (mentalizing, nonmentalizing) as well as subsequent similarity rating, allowing us to identify regions in which BOLD signal increases were associated with increased target similarity, separately for mentalizing and nonmentalizing trials. Because of the relative infrequency of ''highly similar'' responses, ratings of 3 and 4 were combined into a single level.

Comparisons of interest were implemented using a random-effects model. First, tasks were compared directly, for example, *mentalizing* > *nonmentalizing* ($p < .001$, 25 contiguous voxels, providing an overall alpha level of $p < .05$, corrected). Second, the effect of target similarity was examined parametrically by identifying brain regions in which a significantly stronger correlation was observed between the BOLD signal and similarity ratings for mentalizing than for nonmentalizing trials ($p < .005$, 10 contiguous voxels). Statistical comparisons between conditions were conducted using analysis-of-variance procedures on the parameter estimates associated with each trial type.

Acknowledgments

We thank L. Davachi, W. Kelley, T. Laroche, A. Maril, M. Mason, and A. Schein for advice and assistance.

Reprint requests should be sent to Jason Mitchell, Department of Psychology, Harvard University, William James Hall 1568, 33 Kirkland Street, Cambridge, MA 02138, or via e-mail: jmitchel@wjh.harvard.edu.

The data reported in this experiment have been deposited with The fMRI Data Center archive (www.fmridc.org). The accession number is 2-2004-1189A.

REFERENCES

Adolphs, R. (1999). Social cognition and the human brain. *Trends in Cognitive Sciences, 3,* 469–479.

Adolphs, R. (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology, 11,* 231–239.

Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology, 12,* 169–177.

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences, 7,* 267–278.

Blakemore, S. J., Winston, J., & Frith, U. (2004). Social cognitive neuroscience: Where are we heading? *Trends in Cognitive Sciences, 8,* 216–222.

Botterhill, G. (1996). Folk psychology and theoretical status. In P. Carruthers & P. K. Smitch (Eds.), *Theories of theory of mind.* Cambridge, UK: Cambridge University Press.

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience, 2,* 913–919.

Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping, 8,* 109–114.

Davies, M., & Stone, T. (Eds.). (1995a). *Folk psychology: The theory of mind debate.* Oxford, UK: Blackwell Publishers.

Davies, M., & Stone, T. (Eds.). (1995b). *Mental simulation: Evaluations and applications.* Oxford, UK: Blackwell Publishers.
Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., & Frith, C. D. (1995). Other minds in the brain: A functional imaging study of ''theory of mind'' in story comprehension. *Cognition, 57,* 109–128.
Frith, C., & Frith, U. (2001). The biological basis of social interaction. *Current Directions in Psychological Science, 10,* 151–155.
Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology, 62,* 378–391.
Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of ''theory of mind''. *Trends in Cognitive Sciences, 7,* 77–83.
Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: An fMRI study of ''theory of mind'' in verbal and nonverbal tasks. *Neuropsychologia, 38,* 11–21.
Gallagher, H. L., Jack, A. I., Roepstorff, A., & Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *Neuroimage, 16,* 814–821.
Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 2,* 493–501.
Goel, V., Grafman, J., Sadato, N., & Hallett, M. (1995). Modeling other minds. *NeuroReport, 6,* 1741–1746.
Gopnik, A., & Meltzoff, A. (1996). *Words, thoughts, and theories.* Cambridge: MIT Press.
Gopnik, A., & Wellman, H. (1992). Why the child's theory of mind is really a theory. *Mind and Language, 7,* 145–171.
Gordon, R. M. (1992). Folk psychology as simulation. *Mind and Language, 1,* 158–171.
Gregory, C., Lough, S., Stone, V., Erzinclioglu, S., Martin, L., Baron-Cohen, S., Hodges, J. R. (2002). Theory of mind in patients with frontal variant frontotemporal dementia and Alzheimer's disease: Theoretical and practical implications. *Brain, 125,* 752–764.
Griffin, D. W., & Ross, L. (1991). Subjective construal, social inference, and human misunderstanding. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 24, pp. 319–354). New York: Academic Press.
Gusnard, D. A., Akbudak, E., Shulman, G. L., & Raichle, M. E. (2001). Medial prefrontal cortex and self-referential mental activity: Relation to a default mode of brain function. *Proceedings of the National Academy of Sciences, U.S.A., 98,* 4259–4264.
Gusnard, D. A., & Raichle, M. E. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nature Reviews: Neuroscience, 2,* 685–694.
Heal, J. (1986). Replication and functionalism. In J. Butterfield (Ed.), *Language, mind and logic.* Cambridge, UK: Cambridge University Press.
Heal, J. (1996). Simulation, theory, and content. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 75–89). Cambridge, UK: Cambridge University Press.
Johnson, S. C., Baxter, L. C., Wilder, L. S., Pipe, J. G., Heiserman, J. E., & Prigatano, G. P. (2002). Neural correlates of self-reflection. *Brain, 125,* 1808–1814.
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.
Macrae, C. N., Moran, J. M., Heatherton, T. F., Banfield, J. F., & Kelley, W. M. (2004). Medial prefrontal activity predicts memory for self. *Cerebral Cortex, 14,* 647–654.
Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature, 379,* 649–652.
Mason, M. F., Banfield, J. F., & Macrae, C. N. (2004). Thinking about actions: The neural substrates of person knowledge. *Cerebral Cortex, 14,* 209–214.
McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences, U.S.A., 98,* 11832–11835.
Meltzoff, A. N., & Brooks, R. (2001). ''Like me'' as a building block for understanding other minds: Bodily acts, attention, and intention. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition.* Cambridge: MIT Press.
Mitchell, J. P., Heatherton, T. F., & Macrae, C. N. (2002). Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences, U.S.A., 99,* 15238–15243.
Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2004). Encoding specific effects of social cognition on the neural correlates of subsequent memory. *Journal of Neuroscience, 24,* 4912–4917.
Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., & Dolan, R. J. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature, 383,* 812–815.
Nickerson, R. (1999). How we know—and sometimes misjudge—what other know: Imputing one's own knowledge to others. *Psychological Bulletin, 125,* 737–759.
Perner, J. (1991). *Understanding the representational mind.* Cambridge: MIT Press.
Perner, J. (1996). Simulation as explicitation of predication-implicit knowledge about the mind: Arguments for a simulation–theory mix. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 90–104). Cambridge, UK: Cambridge University Press.
Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *Journal of Neuroscience, 18,* 2188–2199.
Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences, U.S.A., 98,* 676–682.
Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27,* 169–192.
Ross, L., Greene, D., & House, P. (1977). The ''false consensus'' effect: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology, 13,* 279–301.
Samson, D., Apperly, I. A., Chiavarino, C., & Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience, 7,* 499–500.
Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: fMRI investigations of theory of mind. *Neuroimage, 19,* 1835–1842.
Schmitz, T. W., Kawahara-Baccus, T. N., & Johnson, S. C. (2004). Metacognitive evaluation, self-relevance, and the right prefrontal cortex. *Neuroimage, 22,* 941–947.
Shulman, G. L., Fiez, J. A., Corbetta, M., Buckner, R. L., Miezen, F. M., Raichle, M. E., & Petersen, S. E. (1997). Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. *Journal of Cognitive Neuroscience, 9,* 648–663.

Stich, S., & Nichols, S. (1992). Folk psychology: Simulation or tacit theory? *Mind and Language, 7,* 35–71.
Stone, T., & Davies, M. (1996). The mental simulation debate: A progress report. In P. Carruthers & P. K. Smitch (Eds.), *Theories of theory of mind.* Cambridge, UK: Cambridge University Press.
Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience, 10,* 640–656.
Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain.* New York: Thieme.
Vogeley, K., May, M., Ritzl, A., Falkai, P., Zilles, K., & Fink, G. R. (2004). Neural correlates of first-person perspective as one constituent of human self-consciousness. *Journal of Cognitive Neuroscience, 16,* 817–827.
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.
Zysset, S., Huber, O., Ferstl, E., & von Cramon, D. Y. (2002). The anterior frontomedian cortex and evaluative judgment: An fMRI study. *Neuroimage, 15,* 983–991.