

# Mentalizing under Uncertainty: Dissociated Neural Responses to Ambiguous and Unambiguous Mental State Inferences

Adrianna C. Jenkins and Jason P. Mitchell

Department of Psychology, Harvard University, Cambridge, MA  
02138, USA

**The ability to read the minds of others (i.e., to mentalize) requires that perceivers understand a wide range of different kinds of mental states, including not only others' beliefs and knowledge but also their feelings, desires, and preferences. Moreover, although such inferences may occasionally rely on observable features of a situation, perceivers more typically mentalize under conditions of "uncertainty," in which they must generate plausible hypotheses about a target's mental state from ambiguous or otherwise underspecified information. Here, we use functional neuroimaging to dissociate the neural bases of these 2 distinct social-cognitive challenges: 1) mentalizing about different types of mental states (beliefs vs. preferences) and 2) mentalizing under conditions of varying ambiguity. Although these 2 aspects of mentalizing have typically been confounded in earlier research, we observed a double dissociation between the brain regions sensitive to type of mental state and ambiguity. Whereas ventral and dorsal aspects of medial prefrontal cortex responded more during ambiguous than unambiguous inferences regardless of the type of mental state, the right temporoparietal junction was sensitive to the distinction between beliefs and preferences irrespective of certainty. These results underscore the emerging consensus that, rather than comprising a single mental operation, social cognition makes flexible use of different processes as a function of the particular demands of the social context.**

**Keywords:** medial prefrontal cortex, mentalizing, neuroimaging, social cognition, theory of mind

## Introduction

Unlike encounters with falling tree branches, stalled cars, or other inanimate objects, an understanding of other people requires the tacit recognition that their behavior is influenced by the contents of their minds (Dennett 1987). However, the ability to infer the nature of those contents—that is, to mentalize—poses a series of nontrivial challenges to human cognition. Perceivers only rarely receive explicit reports about another person's thoughts, feelings, or desires and must instead interpret ambiguous hints about the hidden inner workings of other minds: for example, attempting to uncover the possible significance of an eyebrow raise, sidelong glance, vocal inflection, or sudden departure. Each of these bits of information, in turn, may be clues to a wide range of possible kinds of mental states, such as what a person is thinking (i.e., beliefs), feeling (emotions), desiring (wants and preferences), or intending (goals). Finally, having generated a provisional model of another person's mind, perceivers must also calculate how the contents of that mind are likely to influence the person's behavior.

Given the complexity and diversity of the inferences we make about others, humans likely developed a suite of cognitive processes that, together, allow us to traffic so readily in the mental worlds of other people. Consistent with this possibility that social cognition comprises several distinct processes that meet different computational demands, researchers have identified a set of several brain regions that respond consistently when considering the minds of others: dorsal and ventral aspects of the medial prefrontal cortex (MPFC), the temporoparietal junction (TPJ), medial parietal cortex, and the superior temporal sulcus (Fletcher et al. 1995; Goel et al. 1995; Gallagher et al. 2000, 2002; Mitchell et al. 2002; Saxe and Kanwisher 2003; Van Overwalle 2009). Having identified this constellation of regions involved in human social abilities, researchers have now begun to isolate specific mental processes subserved by each, with the aim of decomposing social cognition into its constituent parts.

Importantly, the main challenge in this enterprise has been delineating the dimensions along which social cognition might be expected to divide. One natural starting place has been the observation that perceivers must infer a variety of different types of mental states, such as beliefs, feelings, and intentions, and indeed, researchers have recently suggested that different brain regions may subserve mentalizing about these different kinds of mental content. For example, a right-lateralized region of TPJ has been implicated specifically in representing others' beliefs (Saxe and Kanwisher 2003; Saxe and Powell 2006), and MPFC has emerged consistently from tasks involving inferences about affective states or preferences (Mitchell, Banaji, Macrae 2005a; Hynes et al. 2006; Mitchell et al. 2006; Vollm et al. 2006; Shamay-Tsoory and Aharon-Peretz 2007). Taken together, these observations have led some commentators to conclude that activation in TPJ and MPFC may be modulated specifically by differences among particular types of mental content to be inferred (e.g., Van Overwalle 2009).

However, in addition to inferring different types of mental states, humans must also mentalize under varying degrees of certainty. In some situations, an inference about the state of another person's mind is all but dictated by given information. For example, when Sarah puts her cookie in the office refrigerator and returns to retrieve it 5 min later, we are fairly confident that she "believes" her cookie is in the refrigerator. Similarly, if Sarah always chooses oatmeal cookies from her many dessert options, we can be fairly confident that Sarah "likes" or "prefers" oatmeal cookies. In these cases, perceivers' inferences can be formulated using a simple set of rules operating over explicit, observable information about a target. To infer where Sarah thinks her cookie will be 5 min after she stashes it in the refrigerator, perceivers may simply apply the rule that people generally can recall easily what

they did 5 min ago. Likewise, perceivers may conclude that Sarah has a particular fondness for oatmeal cookies by applying the rule that if someone freely and consistently chooses an object (e.g., oatmeal cookies) over comparable alternatives, then that person likely prefers that object (Kelley 1972). In both cases, readily observable information can feed into some basic social-cognitive rules to produce fairly unambiguous inferences about another person's mental states.

In contrast, many inferences about human minds take place under conditions of far greater ambiguity. When Steve arrives home and hears voices inside his apartment, will he believe that he is being robbed, that he accidentally left the TV on, that his parents have made a surprise visit, or something else? Similarly, if Steve always arrives late to lecture when the only available seats are in the back of the room, we cannot be particularly confident that he really does prefer to sit far away from the professor. Because the information in such situations is insufficient to constrain one's inferences fully, perceivers must make do with provisional hypotheses about a target's mental states, which remain ambiguous until further clues about their contents are discerned. Although perceivers do make assumptions about other minds even under conditions of relative ambiguity (Gilbert 1998), it is unlikely that they do so using the kind of rule-based processes that can be brought to bear more fruitfully for inferences of greater certainty. Rather, given a scarcity of suitably definitive inputs to our social-cognitive rules, mentalizing under uncertainty likely relies on an alternative, more flexible, and internally generated system for making sense of other minds.

In attempting to identify the dimensions along which social cognition dissociates, most extant research has confounded differences in mentalizing about varying types of internal states with differences in mentalizing under varying degrees of certainty. For example, although the TPJ has been specifically linked to a particular type of mental state—beliefs—the information provided in typical belief mentalizing tasks essentially dictates the mental state of the protagonist, making perceivers' inferences unambiguous. In the bulk of experiments identifying the TPJ with beliefs (Saxe and Kanwisher 2003; Samson et al. 2004; Saxe and Wexler 2005; Saxe and Powell 2006), perceivers read stories based on the classic "Sally-Anne" problem developed for use in children: perceivers watch Sally place her ball in a basket and then, while Sally is away and unaware, they watch Anne surreptitiously move the ball to a second location, at which point they are asked where Sally will look for her ball when she returns. This kind of situation contains all the information needed for an unambiguous, rule-based inference about what Sally believes or thinks (i.e., that the ball is still safely hidden in its original location). In contrast, the information provided in typical preference or affective mentalizing tasks leaves inferences much more open-ended (Hynes et al. 2006; Mitchell et al. 2006; Vollm et al. 2006; Shamay-Tsoory and Aharon-Peretz 2007). For example, participants might be told that Sarah is politically liberal and subsequently be asked whether she would prefer to go hiking or go to the beach (Mitchell et al. 2006). The frequent conflation of these 2 dimensions raises the possibility that findings previously attributed to differences in type of mental state, such as the preferential engagement of MPFC during inferences about others' preferences, may in fact be better attributed to differences in the certainty with which such mental state inferences can be made.

Indeed, a substantial amount of other research supports the possibility that MPFC may subserve mentalizing under uncertainty rather than inferences about particular types of mental states per se. Recently, this region has been implicated in processes supporting the ability to draw on elements of relevant past experiences in order to formulate novel predictions (Addis et al. 2007; Buckner and Carroll 2007), as well as in the use of one's own experience to mentalize about others (Mitchell et al. 2006; Jenkins et al. 2008). When inferences are relatively underspecified by situational constraints, perceivers may find it especially useful to mentalize on the basis of such simulated, internally generated information, whether that information arises from their own firsthand experience or from having observed similar circumstances in the past. That is, perceivers may find it particularly useful to rely on associations formed through past experiences as they generate predictions about what another person may be thinking or feeling in ambiguous or uncertain situations (Mitchell forthcoming). In contrast, such a process may be less useful under circumstances in which another person's mental state could be inferred simply by applying general "rules" about human minds.

In the current experiment, we investigated the extent to which regions associated with mentalizing would be modulated independently by type of mental state (beliefs vs. preferences) and the uncertainty surrounding one's inference about it. Participants were scanned using functional magnetic resonance imaging (fMRI) as they read short vignettes that supported either an unambiguous or ambiguous inference about a person's beliefs or preferences. Unambiguous versions of each vignette were written such that the information in the scenario would strongly suggest the mental state of the protagonist, whereas ambiguous vignettes implied that the protagonist's mental state could be any one of multiple possibilities. For each vignette, participants were obliged to consider either the protagonist's beliefs or his or her preferences, thus allowing us to dissociate brain regions that were sensitive to differences in mental state type (belief and preference) from those sensitive to differences in mentalizing certainty (ambiguous and unambiguous). Although interested in the potential effects of these dimensions across the brain, we were particularly interested in examining the extent to which MPFC contributions to social cognition are better characterized as subserving inferences about affectively laden mental states (such as preferences) or as more generally subserving ambiguous inferences under uncertainty. Moreover, this design also allowed us to test earlier claims that the TPJ specifically subserves inferences about a particular type of mental state (i.e., beliefs).

## Materials and Methods

### *Participants*

Fifteen right-handed college undergraduates (9 females, age range 18–22 years, mean age 19.8 years) with no history of neurological problems participated in exchange for pay or course credit. Participants provided informed consent in accordance with the guidelines maintained by Massachusetts General Hospital.

### *Stimuli and Behavioral Procedure*

#### *Mentalizing Task*

During scanning, participants read short vignettes relating the events of an everyday scenario. Vignettes conveyed information about either

a protagonist's beliefs or preferences (see Table 1 for examples and Supplementary Material for full stimulus set). Unambiguous versions of each scenario were written such that the information in the scenario would strongly suggest, but not state explicitly, the belief or preference of the protagonist. Such scenarios relied heavily on perceptual truisms about human beings (e.g., that they generally perceive objects in the environment and generally remember what they have recently seen). Ambiguous versions of each scenario were written such that the protagonist's belief or preference could plausibly be any one of multiple possibilities under the circumstances provided, that is, the information given did not dictate a correct response but rather left the inference more open-ended. A slight change in what would otherwise be an unambiguous scenario might render deterministic rules about the human mind inapplicable and the scenario therefore ambiguous: for example, if Sarah, on her way out the door having just put her cookie in the refrigerator, hears Tom tell her he's moving her cookie but he does not say where (belief) or if Sarah always eats an oatmeal cookie after dinner but there are never any other options because Tom always buys dessert (preference), we can be less certain about Sarah's mental states given the information provided. Unambiguous belief vignettes were created in both "true belief" and "false belief" versions; however, no differences were observed between true and false beliefs, and analyses were therefore collapsed across this dimension.

Stimuli in all 4 groups (ambiguous preference, ambiguous belief, unambiguous preference, and unambiguous belief) were matched for length (mean number of characters = 213.5). Matched control stories in which participants inferred the content of physical representations (such as those in photographs or on maps) were used for comparison (Zaitchik 1990). For example, participants might read about a tree house that was photographed before being painted blue and be asked to identify the color in which it would have appeared in the photo (Table 1).

Following each mentalizing scenario, participants answered a single multiple-choice question about the protagonist's belief or preference; following each nonsocial scenario, participants answered a question about a physical representation (such as a map or photograph). For all scenarios, the story and question remained onscreen together for a total of 10 s, at which point the story disappeared and 4 response choices were presented for 4 s. In all conditions, participants were asked to formulate an answer to every question before any response choices appeared. Accordingly, to allow for the possibility that participants

generated ideas other than those represented by our answer choices, the fourth response option was always a none-of-the-above possibility (e.g., "Somewhere else"). Each trial was followed by 12 s of fixation. Each participant completed a total of 60 mentalizing scenarios and 12 nonsocial scenarios across 4 functional runs, with presentation randomized across participants such that no participant ever encountered both an ambiguous and an unambiguous version of the same story.

### Imaging Procedure

fMRI data were collected using a 3-T Siemens Trio scanner across 4 functional runs of 234 volume acquisitions (26 axial slices, 5 mm thick, 1 mm skip). Functional imaging used a gradient-echo echo planar pulse sequence (time repetition = 2 s, time echo = 35 ms,  $3.75 \times 3.75$  in-plane resolution). Prior to the functional scans, we collected a high-resolution  $T_1$ -weighted structural scan (magnetization-prepared rapid gradient echo). PsyScope software for Mac OS X (L. Bonatti, International School of Advanced Studies, Trieste, Italy) was used to project stimuli onto a screen at the end of the magnet bore, which participants viewed via a mirror mounted on the head coil. A pillow and foam cushions were placed inside the coil to minimize head movement.

fMRI data were preprocessed and analyzed using SPM2 (Wellcome Department of Cognitive Neurology, London, UK). First, functional data were time corrected for differences in acquisition time between slices for each whole-brain volume and realigned to correct for head movement. Functional data were then transformed into a standard anatomical space (3-mm isotropic voxels) based on the ICBM 152 brain template (Montreal Neurological Institute). Normalized data were then spatially smoothed (8 mm full width at half maximum) using a Gaussian kernel.

Statistical analyses were performed using the general linear model in which the blocked design was modeled using a boxcar function and additional covariates of no interest (a session mean and a linear trend). This analysis was performed individually for each participant, and contrast images for each participant were subsequently entered into a second-level analysis treating participants as random effect. Peak coordinates were identified using a statistical criterion of 25 or more contiguous voxels at a voxelwise threshold of  $P < 0.0001$ . Monte Carlo simulations (S. Slotnick, Boston College) of our brain volume confirmed that these criteria provided a brainwise alpha level of  $P < 0.05$ , corrected for multiple comparisons.

**Table 1**  
Stimulus examples

	Unambiguous	Ambiguous
Belief	Pam is an avid gardener. The weather was so warm today that all the tulips in Pam's backyard suddenly bloomed. The tulips next to Pam's office still have not yet flowered, though. Pam has been at work all day.  What does Pam think? 1. Her tulips have bloomed 2. Her tulips have not bloomed yet 3. Her tulips have died 4. Something else	Pam is an avid gardener and is particularly fond of her tulips. It's early spring, and a few of her flowers have begun to bloom. When Pam got home from work today, her neighbor told her she might want to take a look at her tulip beds.  What does Pam think? 1. Her tulips have bloomed 2. Her tulips have not bloomed yet 3. Her tulips have died 4. Something else
Preference	Erin has 2 classes on Tuesdays. Today was the last day of Tuesday classes. In both of her classes, Erin is usually one of the first people there, and she always sits in the back.  Where does Erin like to sit in class? 1. In the front 2. In the back 3. In the middle 4. Somewhere else	Erin has 2 classes on Tuesdays. Today was the first day of Tuesday classes for the semester. In both of her classes, the room was quite full when Erin arrived, and she sat in the back.  Where does Erin like to sit in class? 1. In the front 2. In the back 3. In the middle 4. Somewhere else
Nonsocial	The color printer cartridge just ran out of blue ink, but it kept printing anyway. It printed a picture of a healthy grass lawn from a computer screen.  In the printed picture, what color is the grass? 1. Yellow 2. Green 3. Blue 4. Something else	

Note. Mentalizing scenarios support inferences that differ the type of mental state to be inferred (belief vs. preference) and the certainty with which the inference can be made (unambiguous vs. ambiguous). Nonsocial scenarios support inferences without mental content.

We first identified regions of interest from the comparison of mentalizing > nonsocial (i.e., all unambiguous and ambiguous belief and preference stories vs. nonsocial control stories). These regions were then interrogated for differences among the mentalizing scenarios by comparing the parameter estimates associated with the 4 mentalizing trial types: unambiguous belief, unambiguous preference, ambiguous belief, and ambiguous preference. To confirm the results of the region-of-interest analysis, we also conducted whole-brain, random-effects analyses of unambiguous versus ambiguous scenarios (collapsing across content type) and belief versus preference scenarios (collapsing across ambiguity).

## Results

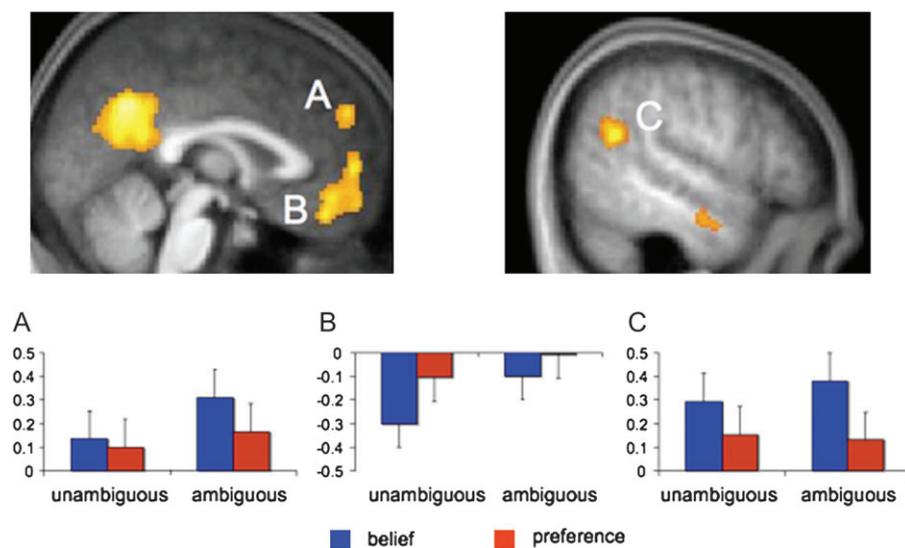
### Behavioral Results

To confirm our conditionalization of scenarios as unambiguous and ambiguous, we first examined the distribution of participants' responses to each question as a function of ambiguity. Specifically, for each question, we calculated the proportion of participants who chose the most commonly selected of the 3 possible contentful answers, excluding "none of the above" responses. Participants overwhelmingly converged on a single answer to each unambiguous scenario, choosing the modal response 87% of the time. In contrast, responses were more variable for ambiguous scenarios, with participants agreeing on a single answer only 51% of the time,  $\chi^2(1, n = 15) = 8.14, P < 0.005$ . Moreover, for unambiguous questions, participants chose "none of the above" less than 1% of the time, whereas for ambiguous questions, participants made use of this option 30% of the time,  $\chi^2(1, n = 15) = 27.13, P < 0.0001$ . This pattern of responding was observed for both beliefs (88% agreement for unambiguous belief vs. 53% agreement for ambiguous belief inferences) and preferences (84% agreement for unambiguous preference vs. 51% agreement for ambiguous preference inferences).

### fMRI Results

The primary question of interest was the extent to which brain regions involved in mentalizing would be sensitive to differences

in content type and in ambiguity. To identify brain regions involved in mentalizing, we first conducted a whole-brain, random-effects analysis of all "mentalizing > nonsocial" scenarios. This contrast revealed a set of regions commonly associated with social cognition, including both dorsal and ventral MPFC (vMPFC), right and left TPJ, the superior temporal sulcus, the temporal poles, and medial parietal cortex (Fig. 1). We then interrogated these regions of interest for their sensitivity to content type, ambiguity, and the interaction between these 2 factors. The response of 3 regions was modulated by ambiguity and/or content (see Table 2). First, dorsal MPFC (dMPFC) demonstrated greater response during ambiguous than unambiguous inferences,  $F_{1,14} = 7.44, P < 0.02, d = 0.73$ , but did not differentiate between preferences and beliefs,  $F_{1,14} = 1.72, P > 0.21, d = 0.35$ . In contrast, right TPJ was characterized by greater activation during belief than preference scenarios,  $F_{1,14} = 9.74, P < 0.01, d = 0.83$ , but did not differentiate between scenarios as a function of ambiguity,  $F_{1,14} = 0.79, P > 0.38, d = 0.24$ , consistent with suggestions that right TPJ contributes specifically to mentalizing about beliefs (Saxe and Kanwisher 2003; Saxe and Powell 2006). Finally, activation in vMPFC was characterized by main effects of both ambiguity,  $F_{1,14} = 7.10, P < 0.02, d = 0.71$ , and content,  $F_{1,14} = 12.26, P < 0.0005, d = 0.94$ , such that the region responded more during ambiguous than unambiguous inferences and also responded more during inferences about preferences than during inferences about beliefs. All 3 regions showed no evidence of an ambiguity  $\times$  content interaction, all  $F$  values  $< 1.65, P$  values  $> 0.22$ . In contrast, a marginally significant interaction between type and ambiguity was observed in medial parietal cortex,  $F_{1,14} = 3.03, P < 0.10, d = 0.47$ , such that the region responded more during unambiguous than ambiguous inferences about beliefs but more during ambiguous than unambiguous inferences about preferences; however, neither the main effect of ambiguity ( $P > 0.45$ ) nor the main effect of content ( $P > 0.76$ ) approached significance in this region. Moreover, the presence of a significant 2-way interaction of region  $\times$  content,  $F_{2,42} = 9.10, P < 0.001$ , confirmed that the pattern of response to beliefs and preferences differed across



**Figure 1.** Average parameter estimates in dMPFC (A), vMPFC (B), and right TPJ (C) as a function of type of mental state (belief vs. preference) and mentalizing certainty (unambiguous vs. ambiguous). Activation in both dorsal and vMPFC was characterized by a main effect of certainty, such that both regions responded more during ambiguous than unambiguous inferences, regardless of content. In contrast, activation in right TPJ was characterized only by a main effect of content type, such that it responded more during inferences about beliefs than during inferences about preferences. Error bars represent confidence interval for within-subject designs (Loftus and Masson 1994).

these 3 regions; however, the 2-way interaction of region  $\times$  ambiguity did not reach significance,  $F_{2,42} = 1.79$ ,  $P > 0.17$ .

To confirm these findings, we also conducted a whole-brain, random-effects contrast of “ambiguous > unambiguous” scenarios. Consistent with the region-of-interest analysis, the sole region to emerge from this contrast was dMPFC. Additionally, whole-brain, random-effects contrasts of belief versus preference scenarios underscored the differential engagement of right TPJ and vMPFC as a function of content. Whereas right TPJ emerged from the contrast of “belief > preference,” vMPFC emerged from the contrast of “preference > belief” (Table 3).

## Discussion

The human ability to apprehend the mental states of others requires solutions to a host of cognitive challenges. The current findings add to the emerging empirical consensus that these challenges are met by an equally varied set of distinct cognitive processes rather than a single, monolithic “theory-of-mind” module. Replicating earlier research (Saxe and Kanwisher 2003; Saxe and Powell 2006), mentalizing about others’ beliefs was associated with greater activity in right TPJ compared with

mentalizing about others’ preferences or to nonsocial processing. That understanding that others’ beliefs would rely on such specialized processing has been anticipated by a number of commentators, who have pointed out that such inferences place unique demands on cognition, including a requirement to understand representational aspects of others’ minds and to suspend attention to one’s own knowledge in favor of understanding the unique knowledge possessed by another person (Apperly et al. 2005; Saxe 2006; Mitchell 2009).

In contrast, regardless of the type of mental state under consideration, both dorsal and ventral aspects of MPFC responded more during ambiguous, underspecified inferences than during unambiguous, well-constrained inferences. Comparisons across past studies have observed greater MPFC activation during relatively ambiguous inferences about preferences than during relatively unambiguous inference about beliefs, concluding that the relevant difference was in the type of mental state being considered (Van Overwalle 2009). However, the current results suggest a different conclusion. Here, dMPFC did not distinguish between inferences about beliefs and preferences when such inferences were matched for ambiguity, suggesting that what primarily drives the engagement of this region is not the type of mental state being inferred but rather the computational demands associated with constructing novel predictions from minimal information (Johnson-Laird 1994, 2001; Mitchell forthcoming).

What kinds of computational demands might these be? Recently, a number of commentators have suggested that MPFC contributes to a network of regions that subserves the construction of simulated scenarios. For example, in addition to its ubiquitous role in mentalizing, MPFC is consistently engaged by attempts to prospectively imagine the future and to retrospectively remember the past (Addis et al. 2007; Buckner and Carroll 2007; Schacter et al. 2007; Spreng et al. 2009), both of which require perceivers to use internally generated simulations of a situation that is divorced from the current context. Likewise, mentalizing under uncertainty may require perceivers to engage in similar processes of simulation, for example, by imagining their own response to an analogous situation or by drawing on aspects of comparable events from their own life. That is, when ambiguity about another person’s mental states is high, our inferences about other minds may be guided by the contents of our own internal mental experience, mediated by MPFC (Mitchell, Banaji, Macrae 2005b; Mitchell et al. 2006; Jenkins et al. 2008).

Intriguingly, this observation suggests that one reason that MPFC has been so consistently associated with social cognition may be that inferences about the minds of other people are necessarily less constrained than inferences about the physical world. Because the mind of another person is inherently mutable and impossible to perceive directly, inferences about human minds may be fundamentally more ambiguous than inferences about our inanimate, physical surroundings (Mitchell forthcoming). To the extent that MPFC contributes to simulating plausible outcomes for indistinct and shifting phenomena, this region should be expected to participate frequently in understanding the minds of others.

However, such MPFC-mediated processes might also be engaged during nonsocial inferences that likewise require the consideration of multiple, “fuzzy” alternatives based on internally generated simulations. Humans must often make

**Table 2**

Peak voxel and number of voxels for brain regions obtained from the random-effects contrast of all mentalizing scenarios > nonsocial scenarios,  $P < 0.05$ , corrected

	x	y	z	Voxels
dMPFC	0	54	32	186
vMPFC	-8	50	-2	940
R superior temporal sulcus	56	-10	-20	133
L superior temporal sulcus	-68	-34	-4	215
R TPJ	54	-56	22	305
L TPJ	-48	-62	36	272
Medial parietal cortex	-8	-60	18	2477
R occipital cortex	12	-102	8	171
L occipital cortex	-24	-100	6	269

Note. Coordinates refer to the Montreal Neurological Institute stereotaxic space. R, right; L, left.

**Table 3**

Peak voxel and number of voxels for brain regions obtained from random-effects contrasts of certainty and type of mental state  $P < 0.05$ , corrected

	x	y	z	Voxels
Ambiguous > unambiguous				
MPFC	-4	36	40	241
Unambiguous > ambiguous				
No regions observed at $P < 0.05$ , corrected				
Belief > preference				
R TPJ	50	-52	20	25
L TPJ	-50	-52	22	101
Preference > belief				
vMPFC	6	56	0	280
L orbitofrontal cortex	-22	36	-6	30
R insula	50	12	-8	162
	46	-10	-4	35
L inferior frontal gyrus	-46	2	14	30
Midcingulate cortex	-6	-8	30	32
Posterior cingulate cortex	0	-34	30	495
R intraparietal sulcus	30	-42	40	1739
L intraparietal sulcus	-30	-56	54	280
	-30	-36	40	82
	-46	-38	52	48
R middle temporal gyrus	58	-50	-10	111
Cerebellum	18	-64	-46	41
R lateral occipitotemporal sulcus	46	-60	-8	149
Superior parietal gyrus	-8	-78	40	52
Occipital cortex	4	-90	20	46

complex, underdetermined inferences outside the social domain, such as when deciding what kind of weather to expect during an upcoming trip or how the stock market will be affected by lower interest rates. In and of themselves, the current results cannot adjudicate whether MPFC contributions to uncertain inference making are limited to social situations (i.e., mentalizing) or may extend to relatively less social contexts. Indeed, recent findings demonstrate that regions of the right TPJ previously thought to be selective for social cognition also contribute to decidedly nonsocial tasks (Mitchell 2008; Scholz et al. 2009), raising the possibility that MPFC will also prove to participate across both social and nonsocial situations. The possibility that this region subserves processing of ambiguous information across multiple domains awaits future empirical test.

### ***The Flexible Nature of Social Cognition***

The current results also have implications for a longstanding debate over the question of how one person goes about “reading the mind” of another. Psychologists and philosophers have together posited 2 main accounts of the processes by which human beings understand other minds: broadly, those that are “simulationist” (Heal 1986; Gordon 1992) and those that are more “rule based” (also known as “theory” theories; Gopnik and Wellman 1994). Specifically, simulationist theories take as their starting point the observation that, although perceivers can never access the mind of another person directly, they do have constant and direct access to the conscious experience of one mind—their own—which they may be able to use as a model in which to understand the mental experience of another. Such theories suggest that, consciously or unconsciously, perceivers appeal to aspects of their own experience in order to generate insights into other minds. In contrast, rule-based theories emphasize the accumulation over one’s lifetime of probabilistic laws about how human minds work (e.g., “people generally remember what they did 5 min ago”; “when people choose an object freely and consistently, they generally like that object”), which can be applied as relevant situations arise. Although simulationist and rule-based theories of social cognition have often been portrayed as mutually exclusive possibilities for how humans understand the minds of others, the current study suggests a more hybrid view. On one hand, rule-based mentalizing may be a useful strategy when perceivers reason about unambiguous mental states in ways that are strongly guided by explicit contextual information. However, more self-based simulationist processes may be needed to infer mental states under conditions of greater uncertainty, that is, when contextual cues less firmly constrain the possible goings on of another person’s mind.

Interestingly, although both vMPFC and dMPFC differentiated between ambiguous and unambiguous inferences, vMPFC also showed greater activation during inferences about preferences than during inferences about beliefs. Analysis of participants’ agreement on a single response for each vignette confirmed that preference scenarios were no more ambiguous than belief scenarios, and no other regions sensitive to ambiguity (e.g., dMPFC) distinguished between preferences and beliefs. As such, this finding replicates earlier studies that demonstrated greater response in vMPFC when mentalizing about others’ affective states than their cognitive states (Hynes et al. 2006; Vollm et al. 2006; Shamay-Tsoory and Aharon-Peretz

2007) but suggests that this region may be sensitive not only to type of mental state being inferred but also the ambiguity of the information on which such an inference can be made (cf., Van Overwalle 2009).

A possible, albeit speculative, explanation for the less selective functional profile observed in vMPFC builds on social psychological research on attribution, which has long distinguished between explanations of behavior that focus on “the person” versus those that focus on “the situation” (Heider 1958). In the current study, unambiguous preference scenarios supported highly certain inferences because of what perceivers knew about their protagonists (i.e., the person), for example, that someone chose a particular item consistently despite having other options. In contrast, unambiguous belief scenarios supported highly certain inferences because they contained strong situational constraints, such that most human beings would be expected to believe the same thing under the same circumstances (Gilbert 1998), for example, that someone put an object in a particular place and returned to retrieve it a few minutes later. One possibility is that activity in vMPFC, which responded more during unambiguous inferences about preferences than during unambiguous inferences about beliefs ( $P < 0.02$ ), could be associated specifically with person-focused attribution (Mitchell et al. 2005). This hypothesis raises the interesting possibility that vMPFC may respond more strongly to stable, idiosyncratic beliefs (such as a person’s belief in ghosts or karma) that provoke high levels of person-based attribution than to transient preferences that depend heavily on the particular context (a person prefers mojitos to mimosas, but not before noon). Because the current study relied primarily on transient beliefs and stable preferences, additional research is needed to determine the specific contributions to mentalizing made by vMPFC, including its involvement in situation- versus person-based attribution.

### **Conclusion**

The current findings continue the ongoing work of cleaving the functional neuroanatomy of social cognition into its constituent parts. Rather than comprising a single, monolithic process for contemplating the minds of others, recent research has increasingly made clear that social cognition decomposes into a number of distinct processes, each contributing some specific function to overall human social competence. Here, we suggest that one fruitful way to divide social cognition follows from the fact that perceivers face a number of uniquely different mentalizing challenges: not only the ability to infer a wide variety of mental states—such as beliefs, knowledge, feelings, and preferences—but also the ability to mentalize under varying degrees of uncertainty and ambiguity. The current results suggest that the human brain appears to respond to such demands by selectively engaging different regions as a function of the particular social-cognitive challenge to be met. Although some regions, such as the right TPJ, appear to contribute to social cognition by subserving inferences about specific types of mental states (i.e., beliefs), other regions, such as the dMPFC, are indifferent to the distinctions between others’ beliefs and preferences. Instead, the MPFC may contribute preferentially to social cognition when making sense of new situations, unfamiliar individuals, or ambiguously motivated behavior—in other words, when mentalizing under uncertainty.

## Funding

National Science Foundation (BCS 0642448); National Center for Research Resources (P41RR14075); Medical Investigation of Neurodevelopmental Disorders Institute.

## Supplementary Material

Supplementary material can be found at <http://www.cercor.oxfordjournals.org/>

## Notes

The authors thank Daniel L. Ames, Daniel Gilbert, Sean Loosli, and Lindsey Powell for advice and assistance. Data were collected at the Athinoula A. Martinos Center for Biomedical Imaging. *Conflict of Interest*: None declared.

Address correspondence to Adrianna C. Jenkins, Department of Psychology, Harvard University, William James Hall, 33 Kirkland Street, Cambridge, MA 02138, USA. Email: [ajenkins@wjh.harvard.edu](mailto:ajenkins@wjh.harvard.edu).

## References

- Addis DR, Wong AT, Schacter DL. 2007. Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*. 45:1363-1377.
- Apperly IA, Samson D, Humphreys GW. 2005. Domain-specificity and theory of mind: evaluating neuropsychological evidence. *Trends Cogn Sci*. 9:572-577.
- Buckner RL, Carroll DC. 2007. Self-projection and the brain. *Trends Cogn Sci*. 11:49-57.
- Dennett DC. 1987. *The intentional stance*. Cambridge (MA): MIT Press.
- Fletcher PC, Happe F, Frith U, Baker SC, Dolan RJ, Frackowiak RS, Frith CD. 1995. Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. *Cognition*. 57:109-128.
- Gallagher HL, Happe F, Frunswick N, Fletcher PC, Frith U, Frith CD. 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 HL, Jack AI, Roepstorff A, Frith CD. 2002. Imaging the intentional stance in a competitive game. *Neuroimage*. 16:814-821.
- Gilbert DT. 1998. Ordinary personology. In: Gilbert DT, Fiske ST, Lindzey G, editors. *Handbook of social psychology*. New York: McGraw Hill. p. 89-150.
- Goel V, Grafman J, Sadato N, Hallett M. 1995. Modeling other minds. *Neuroreport*. 6:1741-1746.
- Gopnik A, Wellman HM. 1994. The theory theory. In: Hirschfeld LA, Gelman SA, editors. *Mapping the mind: domain specificity in cognition and culture*. New York: Cambridge University Press. p. 257-293.
- Gordon RM. 1992. Folk psychology as simulation. *Mind Lang*. 1:158-171.
- Heal J. 1986. Replication and functionalism. In: Butterfield J, editor. *Language, mind and logic*. Cambridge (UK): Cambridge University Press. p. 135-150.
- Heider F. 1958. *The Psychology of interpersonal relations*. New York: John Wiley & Sons.
- Hynes CA, Baird AA, Grafton ST. 2006. Differential role of the orbital frontal lobe in emotional versus cognitive perspective-taking. *Neuropsychologia*. 44:374-383.
- Jenkins AC, Macrae CN, Mitchell JP. 2008. Repetition suppression of ventromedial prefrontal activity during judgments of self and others. *Proc Natl Acad Sci USA*. 105:4507-4512.
- Johnson-Laird PN. 1994. Mental models and probabilistic thinking. *Cognition*. 50:189-209.
- Johnson-Laird PN. 2001. Mental models and deduction. *Trends Cogn Sci*. 5:434-442.
- Kelley HH. 1972. Attribution in social interaction. In: Jones EE, Kanouse DE, Kelley HH, Nisbett RE, Valins S, Weiner B, editors. *Attribution: perceiving the cause of behavior*. Hillsdale (NJ): Lawrence Erlbaum and Associates. p. 1-26.
- Loftus GR, Masson MEJ. 1994. Using confidence intervals in within-subject designs. *Psychon Bull Rev*. 1:476-490.
- Mitchell JP. 2008. Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cereb Cortex*. 18:262-271.
- Mitchell JP. 2009. Inferences about other minds. *Philos Trans R Soc Lond B Biol Sci*. 364:1309-1316.
- Mitchell JP. Forthcoming. Social psychology as a natural kind. *Trends Cogn Sci*.
- Mitchell JP, Banaji MR, Macrae CN. 2005a. General and specific contributions of the medial prefrontal cortex to knowledge about mental states. *Neuroimage*. 28:757-762.
- Mitchell JP, Banaji MR, Macrae CN. 2005b. The link between social cognition and self-referential thought in the medial prefrontal cortex. *J Cogn Neurosci*. 17:1306-1315.
- Mitchell JP, Heatherton TF, Macrae CN. 2002. Distinct neural systems subserve person and object knowledge. *Proc Natl Acad Sci USA*. 99:15238-15243.
- Mitchell JP, Macrae CN, Banaji MR. 2005. Forming impressions of people versus inanimate objects: social-cognitive processing in the medial prefrontal cortex. *Neuroimage*. 26:251-257.
- Mitchell JP, Macrae CN, Banaji MR. 2006. Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*. 50:655-663.
- Samson D, Apperly IA, Chiavarino C, Humphreys GW. 2004. Left temporoparietal junction is necessary for representing someone else's belief. *Nat Neurosci*. 7:499-500.
- Saxe R. 2006. Uniquely human social cognition. *Curr Opin Neurobiol*. 16:235-239.
- Saxe R, Kanwisher N. 2003. People thinking about thinking people: fMRI investigations of theory of mind. *Neuroimage*. 19:1835-1842.
- Saxe R, Powell LJ. 2006. It's the thought that counts: specific brain regions for one component of theory of mind. *Psychol Sci*. 17:692-699.
- Saxe R, Wexler A. 2005. Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia*. 43:1391-1399.
- Schacter DL, Addis DR, Buckner RL. 2007. Remembering the past to imagine the future: the prospective brain. *Nat Rev Neurosci*. 8:657-661.
- Scholz J, Triantafyllou C, Whitfield-Gabrieli S, Brown EN, Saxe R. 2009. Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. *PLoS ONE*. 4, doi:10.1371/journal.pone.0004869.
- Shamay-Tsoory SG, Aharon-Peretz J. 2007. Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. *Neuropsychologia*. 45:3054-3067.
- Spreng RN, Mar RA, Kim AS. 2009. The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *J Cogn Neurosci*. 21:489-510.
- Van Overwalle F. 2009. Social cognition and the brain: a meta-analysis. *Hum Brain Mapp*. 30:829-858.
- Vollm BA, Taylor AN, Richardson P, Corcoran R, Stirling J, McKie S, Deakin JF, Elliott R. 2006. Neuronal correlates of theory of mind and empathy: a functional magnetic resonance imaging study in a non-verbal task. *Neuroimage*. 29:90-98.
- Zaitchik D. 1990. When representations conflict with reality: the preschooler's problem with false beliefs and "false photographs". *Cognition*. 35:41-68.