

Common Brain Regions with Distinct Patterns of Neural Responses during Mentalizing about Groups and Individuals

Juan Manuel Contreras, Jessica Schirmer, Mahzarin R. Banaji,
and Jason P. Mitchell

Abstract

■ An individual has a mind; a group does not. Yet humans routinely endow groups with mental states irreducible to any of their members (e.g., “scientists hope to understand every aspect of nature”). But are these mental states categorically similar to those we attribute to individuals? In two fMRI experiments, we tested this question against a set of brain regions that are consistently associated with social cognition—medial pFC, anterior temporal lobe, TPJ, and medial parietal cortex. Participants alternately answered questions about the mental states and physical attributes of individual people and groups. Regions previously associated with mentalizing about individuals were also robustly

responsive to judgments of groups, suggesting that perceivers deploy the same social-cognitive processes when thinking about the mind of an individual and the “mind” of a group. However, multivariate searchlight analysis revealed that several of these regions showed distinct multivoxel patterns of response to groups and individual people, suggesting that perceivers maintain distinct representations of groups and individuals during mental state inferences. These findings suggest that perceivers mentalize about groups in a manner qualitatively similar to mentalizing about individual people, but that the brain nevertheless maintains important distinctions between the representations of such entities. ■

INTRODUCTION

One of the cognitive abilities that best separates humans from other animals is the capacity to form complex representations of the mental states of others. Humans see a person cry and assume he is sad. They see a person extend her hand toward an object and infer she intends to reach for it. Spontaneously and without difficulty, humans adopt a theory of mind or intentional stance toward other people (Dennett, 1987; Premack & Woodruff, 1978).

Surprisingly, humans attribute similar mental states to groups of people (Jones, 2010). In ordinary speech, they make statements like “Christians *believe* Adam and Eve were real people,” “scientists *hope* to understand every aspect of nature,” or “companies *think* solely about increasing profits.” These expressions cannot refer to the mental states of each group member; surely, some Christians do not believe Adam and Eve were real people, no single scientist hopes to grasp all aspects of nature, and no actual company executive spends every waking hour fantasizing about profits. These expressions refer to the beliefs, thoughts, and desires of groups of people, although such groups are not conscious entities that can have this kind of internal mental experience. Nonetheless, humans readily use the same mental state vocabulary to describe the

“minds” of groups as they typically do to describe the mental states of individuals.

Perhaps claims about the mental states of groups are merely examples of figurative language that make it easier to communicate about collections of individuals. That is, it may be a linguistic convenience to say “senior citizens want the president to stay in office” rather than “88% of citizens 65 and older polled in a recent representative survey would vote for the president if he were to run for reelection.” Alternatively, references to the mental states of groups may be legitimate instances of theory of mind. In the same way humans endow objects, nonhuman animals, and fictional entities with complex mental states that they do not possess (Epley, Waytz, & Cacioppo, 2007), humans may impute legitimate mental states to groups. That is, humans may naturally perceive a group mind.

The idea of a group mind coheres with previous research on stereotypes as group mental states that inform mentalizing about individual group members (Ames & Mason, 2012; Ames, 2005). In this research, perceivers use stereotypes about the groups to which targets belong to infer the intentions and preferences of these targets (Bottom & Paese, 1997; Plous, 1993; Sagar & Schofield, 1980), especially when they view targets to be substantially different from them (Ames, Weber, & Zou, 2012; Ames, 2004a, 2004b). Further evidence of a link between mentalizing and stereotyping is provided by a neuroimaging

experiment that suggests that these two operations recruit similar brain regions (Contreras, Banaji, & Mitchell, 2012).

These brain region—medial pFC (MPFC), anterior temporal lobe (ATL), TPJ, and medial parietal cortex—respond robustly across a wide range of situations in which perceivers represent mental states (for reviews, see Mitchell, 2009; Saxe, 2009; Frith & Frith, 2006; Gallagher & Frith, 2003). For example, these brain regions show increased activity when participants read stories about others' beliefs (Saxe & Kanwisher, 2003; Fletcher et al., 1995), view cartoons that imply mental states (Brunet, Sarfati, Hardy-Baylé, & Decety, 2000; Gallagher et al., 2000), see objects move intentionally (Wheatley, Milleville, & Martin, 2007; Castelli, Happe, Frith, & Frith, 2000), think about the thoughts of historical characters (Goel, Grafman, Sadato, & Hallett, 1995), and infer the mental states of competitors in strategy games (Gallagher, Jack, Roepstorff, & Frith, 2002; McCabe, Houser, Ryan, Smith, & Trouard, 2001).

If humans adopt a theory of mind about groups, then these brain regions should show robust neural activity when perceivers consider the mental states of groups. Previous research is suggestive: Viewing photographs and videos of social interactions recruits some of the same brain regions that respond to mental state inferences about single individuals (Centelles, Assaiante, Nazarian, Anton, & Schmitz, 2011; Wagner, Kelley, & Heatherton, 2011; Iacoboni et al., 2004). But the participants in these studies were not asked to infer the mental states of groups, and no comparison was made between inferences about the mental states of groups and individuals. Therefore, to test the hypothesis that humans adopt a theory of mind about groups, we conducted an experiment in which participants underwent fMRI while alternately inferring the mental states of groups and individual group members.

EXPERIMENT 1

Methods

Participants

Participants were 25 college students and community members from Cambridge, MA (9 men, 16 women; age range = 19–27 years, $M = 22.0$ years) who were right-handed, had no history of neurological problems, and provided informed consent in a manner approved by the Committee on the Use of Human Subjects in Research at Harvard University. Three additional participants were excluded: one for excessive head movement (more than 100 instances of at least 1 mm of movement or 1° of rotation from one volume to the next) and two for failing to respond to more than 20% of experimental trials.

Stimuli and Behavioral Procedure

Participants were scanned using fMRI while performing a photograph judgment task in which they viewed

80 photographs that depicted groups of people that ranged in size from 3 to 180 individuals ($M = 12.5$). On each trial, participants were asked to consider either the group as a whole or a single member of the group. During group trials, a blue border appeared around the entire photograph, whereas during member trials, the blue border appeared around the face or body of a single person in the group. Participants viewed 40 of the photographs in the member condition, pseudorandomly selected for every participant from the stimuli set of group photographs.

For each photograph, participants were asked to perform one of two tasks. On mental trials, participants were oriented toward the mental states of the targets with the cue, “Enjoy a long car ride?” On group versions of mental trials, participants judged how much the group would enjoy a long car ride together. For member versions of mental trials, participants judged how much the group member would enjoy a long car ride alone. This question was chosen because it could be answered equally well for both groups and individuals, and because group enjoyment depends on the pleasantness of interactions between group members and is not simply reducible to the enjoyment of group members considered individually.

On physical trials, participants were oriented toward physical properties of the targets with the cue, “Stay afloat?” On group versions of physical trials, participants judged how well the group would stay afloat in a life raft. For member versions of physical trials, participants judged how well the group member around would stay afloat in a pair of arm flotation devices. Before the start of the experiment, participants saw photographs of these two flotation devices and read descriptions about them: A Boeing life raft 6 ft in diameter and capable of holding up to 500 pounds and inflatable arm bands that are recommended for children but are capable of holding up to 120 pounds. This question was chosen because it could be answered equally well for both groups and individuals, and because it requires participants to evaluate targets without considering their mental states.

Each trial began with one of the two cue phrases. After 500 msec, the photograph appeared under the cue, along with a 4-point Likert scale (1 = *least*, 4 = *most*). The cue, photograph, and scale remained onscreen for an additional 3250 msec, during which participants indicated their response using a button box in their left hand. For the last 250 msec of a trial, a white fixation cross appeared in the middle of the screen. To optimize estimation of the event-related fMRI response, trials were intermixed in a pseudorandom order and separated by a variable stimulus interval (0–14 sec) during which participants passively viewed a fixation crosshair (Dale, 1999). Trials were segregated into four functional runs, each of which consisted of 60 trials (20 trials in each group condition and 10 trials in each member condition).

Functional Imaging Procedure

Imaging data were acquired on a 3.0-T Siemens Tim Trio scanner (Siemens, Erlangen, Germany) with a standard head coil in the Center for Brain Science at Harvard University. Functional runs used a gradient-echo, EPI pulse sequence (repetition time = 2000 msec, echo time = 30 msec, flip angle = 85°, field of view = 216 × 216 mm, matrix = 72 × 72, in-plane resolution = 3 × 3 mm, slice thickness = 4 mm). Thirty-one interleaved axial slices parallel to the AC–PC line were obtained to cover the whole cerebrum. The photograph judgment task consisted of four runs of 160 volume acquisitions each. Each of the functional runs was preceded by 8 sec of gradient and radio-frequency pulses that allowed the scanner to reach steady-state magnetization. After the functional runs in each experiment, a high-resolution T1-weighted structural scan (multi-echo MPRAGE) was conducted.

Functional Imaging Data Analysis

fMRI data were preprocessed and analyzed using SPM8 (Wellcome Department of Cognitive Neurology, London, United Kingdom) and in-house MATLAB code (MathWorks, Natick, MA) written by Dylan Wagner (Dartmouth College, Hanover, NH). To correct for head movement, a rigid body transformation realigned images within each run and across all runs using the first functional image as a reference. Realigned images were unwarped to reduce any additional distortions caused by head movement. Unwarped data were normalized into a stereotaxic space (2-mm isotropic voxels) based on the SPM8 EPI template that conforms to the ICBM 152 brain template space and approximates the Talairach and Tournoux atlas space. Normalized images were spatially smoothed using a Gaussian kernel (8 mm FWHM) to maximize signal-to-noise ratio and reduce the impact of individual differences in functional neuroanatomy. Finally, individual runs were analyzed on a participant-by-participant basis to find outlier volumes with Artifact Detection Toolbox (ART; McGovern Institute for Brain Research, Cambridge, MA). Outliers were defined as volumes in which participant head movement exceeded 0.5 mm or 1° and volumes in which overall signal were more than three standard deviations outside the mean global signal for the entire run.

For each participant, a general linear model (GLM) was constructed to include task effects and nuisance regressors (run mean, linear trend to account for signal drift over time, six movement parameters computed during realignment, and, if any, outlier scans identified by ART and trials in which participants did not provide a response). To compute unweighted (β) and weighted (t) parameter estimates for each condition at each voxel, the GLM was convolved with a canonical hemodynamic response function as well as its temporal and spatial derivatives. These derivatives explain a significant portion of

BOLD variability above and beyond the canonical hemodynamic response function (Henson, Rugg, & Friston, 2001). Trials were modeled as events of durations equal to their respective RTs to account for differences in RTs across conditions (Grinband, Wager, Lindquist, Ferrera, & Hirsch, 2008).

Comparisons of interest were implemented as linear contrasts. Given the large sample size, significant voxels were identified using a voxel-wise statistical criterion of $p < 10^{-5}$. ROIs, defined using MarsBar (Centre IRMF, Marseille, France) and in-house MATLAB code, were required to exceed 32 voxels in extent, establishing an experiment-wide statistical threshold of $p < .05$, corrected for multiple comparisons, on the basis of Monte Carlo simulations (Slotnick, Moo, Segal, & Hart, 2003). Voxels at the intersection of ROIs from different contrasts were identified using xjView (Stanford University, Palo Alto, CA). Contrast maps were overlaid on the same anatomical template without recalculation of the statistical thresholds used to generate each contrast map. Therefore, conjunctions use conservative statistical thresholds that can increase our confidence in them. Additional statistical comparisons were conducted in MATLAB using paired-sample t tests on the parameter estimates associated with each trial type.

Results

Behavioral Data

Means and standard deviations of responses and RTs are displayed in Table 1. On mental trials, participants judged that groups ($M = 3.11$) would enjoy a long car ride together less than single members ($M = 3.39$), $t(24) = 2.84$, $p < .01$, Cohen's $d = 0.58$. On physical trials, participants did not judge groups ($M = 2.79$) differently from single members ($M = 2.84$), $t(24) = 0.29$, $p = .78$, $d = 0.06$. Response latencies were closely matched for mental judgments in the group ($M = 1.87$) and member ($M = 1.89$) conditions, $t(24) = 0.87$, $p = .39$, $d = 0.18$, but participants responded more quickly to physical judgments about groups ($M = 1.77$) than single members ($M = 1.92$), $t(24) = 3.62$, $p < .01$, $d = 0.74$.

Functional imaging data. Trials from the photo judgment task were conditionalized on the basis of question (mental, physical) and target (group, member), resulting in four conditions of interest. A whole-brain contrast identified voxels in which BOLD activity was greater during trials in which participants answered questions about the mental states of groups than their physical appearance (*group mental > group physical*). This contrast identified ventral and dorsal aspects of MPFC, bilateral ATL, bilateral TPJ, and medial parietal cortex (Table 2). Consistent with earlier research, a similar whole-brain contrast of single member trials (*member mental > member physical*) identified the same brain

Table 1. Participants' Responses and Response Latencies for Tasks in Experiments 1 and 2

Target	Responses		Response Latencies	
	Mental	Physical	Mental	Physical
<i>Experiment 1: Photograph Judgment Task</i>				
Group	3.11 ^a (0.48)	2.79 ^b (0.61)	1.87 ^a (0.35)	1.77 ^b (0.30)
Member	3.39 ^c (0.54)	2.84 ^{ab} (0.78)	1.89 ^a (0.37)	1.92 ^a (0.35)
<i>Experiment 2: Photograph Judgment Task</i>				
Group	2.97 ^{ab} (0.38)	2.88 ^{ab} (0.51)	1.88 ^{ab} (0.33)	1.81 ^{abc} (0.32)
Member	3.03 ^{ab} (0.31)	2.94 ^a (0.55)	1.88 ^a (0.33)	1.87 ^a (0.21)
Individual	3.50 ^c (0.38)	3.05 ^b (0.59)	1.79 ^{bc} (0.31)	1.73 ^c (0.37)
	Responses		Response Latencies	
	Belief	Photo	Belief	Photo
<i>Experiment 2: False Belief Localizer</i>				
	.93 ^a (.10)	.87 ^b (.13)	3.06 ^a (1.05)	3.29 ^b (0.85)

Means and standard deviations (in parentheses). In the photograph judgment task, participants' responses refer to their judgments on a 4-point Likert scale (1 = *least*, 4 = *most*). In the false belief localizer, participants' responses refer to their proportion of correct responses to the questions about the stories. RTs are displayed in seconds. For each dependent variable, means sharing a superscript (^a, ^b, ^c) do not differ significantly at $p < .05$, as computed in paired-samples t tests.

regions (Table 2). For a formal test that these two contrasts identified overlapping brain regions, we identified voxels in the intersection of the ROIs defined by each contrast. This analysis identified voxels in dorsal and ventral MPFC, bilateral ATL, bilateral TPJ, and medial parietal cortex (Figure 1).

As an additional test of the hypothesis that brain regions recruited by mental state inferences about individuals are also recruited by mental state inferences about groups, the parameter estimates of the group contrast (*group mental* > *group physical*) were extracted for the brain regions defined by the member contrast (*member mental* > *member physical*), averaging over all voxels in each ROI (Table 3). To test that inferences about the mental states of groups also recruit these brain regions, we examined whether BOLD activity in these parameter estimates were significantly higher than zero. This was the case for every ROI, all $t_s(24) > 5.00$, all $p_s < 10^{-4}$, suggesting that perceivers also recruit these regions when they represent the mental states of groups.

Moreover, to test whether inferences about the mental states of groups recruit these brain regions as robustly as

inferences about the mental states of single members, we examined whether differences in BOLD activity between the parameter estimates of the group and member condition (*member mental* > *member physical*) were statistically equivalent (Table 3). Although ventral MPFC showed a marginally larger response to groups than single members, $t(24) = 1.84$, $p = .08$, every other ROI showed equivalent responses to mental state inferences about groups and single members, all $t_s(24) < 1.60$, all $p_s > .12$, suggesting that perceivers recruit these brain regions as strongly when they represent the mental states of groups as when they perform similar inferences about individuals.

Discussion

We hypothesized that, if humans adopt a theory of mind about groups, then brain regions engaged by theory of mind about individuals—MPFC, bilateral ATL, bilateral TPJ, and medial parietal cortex—should show robust neural activity when perceivers consider the mental states of groups. In support of this hypothesis, these brain regions were more responsive to inferences about mental states of both groups and individuals than to inferences about the physical appearance of both groups and individuals.

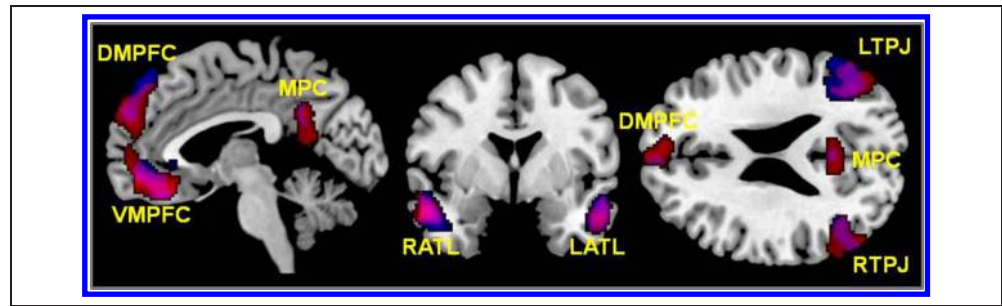
As such, the results of the present experiment suggest that common brain regions respond during inferences about mental states of groups and individuals. But this

Table 2. Brain Regions Identified in Whole-brain, Random-effects Contrasts (*Group Mental* > *Group Physical*, *Member Mental* > *Member Physical*) in the Photograph Judgment Task of Experiment 1, $p < .05$, Corrected for Multiple Comparisons

Region	<i>Group Mental</i> > <i>Group Physical</i>				<i>Member Mental</i> > <i>Member Physical</i>			
	<i>x</i>	<i>y</i>	<i>z</i>	<i>k</i>	<i>x</i>	<i>y</i>	<i>z</i>	<i>k</i>
TPJ								
Right	54	-51	16	1218	56	-57	14	891
Left	-46	-77	30	811	-42	-63	28	1471
MPFC								
Ventral	-4	53	-10	1075	-4	47	-8	579
Dorsal	-8	61	36	749	-6	45	50	490
Anterior temporal cortex								
Left	-56	-9	-22	798	-54	-11	-16	1133
Right	56	-1	-22	769	42	19	-32	897
Medial parietal cortex	-8	-55	16	524	0	-49	28	95

From left to right, columns list the names of regions obtained from whole-brain, random-effects contrasts, the stereotaxic Montreal Neurological Institute coordinates of their peak voxels, and their size in number of voxels (k).

Figure 1. *t* maps computed by whole-brain, random-effects contrasts *group mental* > *group physical* (blue) and *member mental* > *member physical* (red) in the photo judgment task of Experiment 1. Voxels at the intersection of the two *t* maps are displayed in purple. *t* Maps are displayed in sagittal ($x = -4$), coronal ($y = 0$), and axial ($z = 24$) slices of a high-resolution anatomical template. These contrasts identified medial pFC (VMPFC), dorsomedial pFC (DMPFC), bilateral ATL (RATL and LATL), bilateral TPJ (RTPJ and LTPJ), and medial parietal cortex (MPC).



result is puzzling given that people do not consider the minds of single individuals and the “minds” of groups to be equivalent. For instance, people are more likely to attribute intentions than emotions to groups, but they do not show such a bias in imputing mental states to individuals (Knobe & Prinz, 2008). Additionally, people attribute a mind to a group only when they cannot attribute differentiable minds to the individuals that comprise the group (Waytz & Young, 2012). Therefore, although common brain regions may respond preferentially to theory of mind about individuals and groups, they may have distinct representations of mental state inferences about groups and mental state inferences about individuals.

To test this hypothesis, we used multivoxel pattern analysis (MVPA), which takes as its unit of analysis not the responses of individual voxels but the responses of

multiple voxels or multivoxel patterns (for a review, see Weil & Rees, 2010). Differences in multivoxel patterns across stimuli index differences in the underlying representations of these stimuli. For example, face and house percepts elicit distinct multivoxel patterns in ventral temporal cortex, suggesting that the visual system contains distinct representations of faces and houses (Haxby et al., 2001). In the present experiment, we identified brain regions in which multivoxel patterns differentiate mental state inferences about groups from similar inferences about individual group members. In other words, this analysis aimed to identify brain regions that contain distinct representations of mental state inferences about groups and individuals.

We conducted this analysis in an experiment in which participants performed the same photograph judgment

Table 3. Mean Parameter Estimates of the Contrasts *Group Mental* > *Group Physical* and *Member Mental* > *Member Physical*, Extracted from Brain Regions Identified in Whole-brain, Random-effects Contrasts in the Photograph Judgment Task of Experiment 1 (*Member Mental* > *Member Physical*) and Experiment 2 (*Individual Mental* > *Individual Physical*)

Region	<i>Member Mental</i> > <i>Member Physical</i>		<i>Individual Mental</i> > <i>Individual Physical</i>		<i>d</i>
	β_{group}	β_{member}	β_{group}	β_{member}	
TPJ					
Right	0.55 (0.09)	0.39 (0.08)	0.44 (0.09)	0.39 (0.08)	0.13
Left	0.48 (0.09)	0.45 (0.08)	0.40 (0.11)	0.46 (0.12)	0.11
MPFC					
Ventral	0.83 (0.13)	0.50 (0.11)	0.47 (0.08)	0.23 (0.12)	0.47
Dorsal	1.00 (0.19)	0.78 (0.15)	0.35 (0.10)	0.34 (0.08)	0.03
Anterior temporal cortex					
Left	0.37 (0.06)	0.37 (0.05)	0.25 (0.05)	0.23 (0.05)	0.08
Right	0.41 (0.06)	0.35 (0.05)	0.32 (0.07)	0.26 (0.06)	0.19
Medial parietal cortex	0.49 (0.10)	0.30 (0.08)	0.25 (0.08)	0.19 (0.07)	0.15

From left to right, columns list the names of regions obtained from whole-brain, random-effects contrasts and the parameter estimates of the group (*group mental* > *group physical*) and member (*member mental* > *member physical*) contrasts, averaging over all voxels in each region. The rightmost column lists the effect size (Cohen’s *d*) of the difference between parameter estimates of the group and member contrasts in Experiment 2. Numbers in parentheses denote standard errors. Parameter estimates are reliably larger than zero in every region, all *ps* < .01. Group and member mean parameters do not differ reliably in any region, all *ps* > .05.

task from Experiment 1. But whereas participants in Experiment 1 only viewed some of the photographs in their member condition, participants in Experiment 2 viewed all photographs in their group and member condition. The larger number of trials in the member condition as well as the equal number of trials in the group and member conditions allows the use of MVPA. Specifically, this modification ensures that the parameter estimates in the multivoxel patterns of the group and member conditions are equal in statistical power and reliability. To ascertain that we examined the brain regions that have been identified as responding preferentially to mentalizing, this version of the task included an additional set of trials that required participants to make mental state inferences about individuals outside a group context. These trials were used as a functional localizer of brain regions engaged by mentalizing. Additionally, participants completed a false belief task (Saxe & Kanwisher, 2003), which provided an independent functional localizer of brain regions engaged by mentalizing.

EXPERIMENT 2

Method

Participants

Participants were 14 individuals (eight men, six women; age range = 18–23 years, $M = 19.9$ years) sampled from the same population as Experiment 1. One participant withdrew from the experiment after becoming ill during data collection.

Stimuli and Behavioral Procedure

Participants were scanned using fMRI while performing the same photograph judgment task with two modifications. First, whereas participants in Experiment 1 viewed 40 of the photographs in the member condition, participants in Experiment 2 viewed all 80 photographs in their member version. Second, the task included photographs that depicted a single individual (40 photographs). These photographs, drawn from the same database, depicted individual people in a variety of settings. Each photograph was presented twice. In one presentation, participants judged how much the individual would enjoy a long car ride alone in one set of trials. In the other presentation, they judged how well the group member would stay afloat in a pair of arm flotation devices. Trials were segregated into eight functional runs, each of which consisted of 50 trials (10 trials in each group and member condition, 5 trials in each individual condition).

Participants in Experiment 2 also completed two runs of a *false belief* localizer used to identify brain regions that represent mental states (Saxe & Kanwisher, 2003). During this task, participants read 12 vignettes that referred to the false belief of a person (*belief blocks*). For example, one belief vignette read, “Jenny put her choco-

late away in the cupboard. Then she went outside. Alan moved the chocolate from the cupboard into the fridge. Half an hour later, Jenny came back inside.” The corresponding cue read, “Jenny expects to find her chocolate in the:” and the answer choices read “fridge” and “cupboard.” Participants also read 12 vignettes that referred to an outdated physical representation, such as a photograph (photo blocks). These vignettes had the same logical structure as belief stories, but referred to outdated physical, rather than mental, representations (Zaitchik, 1990). For example, one physical vignette read, “A photograph was taken of an apple hanging on a tree branch. The film took half an hour to develop. In the meantime, a strong wind blew the apple to the ground.” The corresponding cue read, “The developed photograph shows the apple on the:” and the corresponding answer choices read “ground” and “branch.” Each vignette was presented for 10 sec, after which it was replaced by the cue sentence. Participants had 6 sec to respond to the cue before the end of the story. Each cue was followed by a fixation crosshair lasting 10 sec. Stories were segregated into two functional runs, each of which consisted of 6 belief and 6 photo stories, intermixed in pseudorandom order.

Functional Imaging Procedure

The functional imaging procedures were identical to those of Experiment 1. However, the photograph judgment task consisted of eight runs of 130 volume acquisitions each. The false belief localizer consisted of two runs of 174 volume acquisitions each. Significant voxels in the photograph judgment task were identified using a voxel-wise statistical criterion of $p < .005$, and ROIs were required to exceed 59 voxels in extent. Given the block design of the false belief localizer, the thresholds in the analysis of this data were $p < 10^{-5}$ and $k = 32$. These thresholds established an experiment-wide statistical threshold of $p < .05$, corrected for multiple comparisons, on the basis of Monte Carlo simulations (Slotnick et al., 2003).

Functional Imaging Data Analysis

The functional imaging data analysis was identical to that of Experiment 1, with one exception. Information-based functional brain mapping with a multivariate spherical searchlight was conducted for each participant (Kriegeskorte, Goebel, & Bandettini, 2006). For this analysis, fMRI data from the photo judgment task were preprocessed as before, but they were spatially smoothed with a smaller Gaussian kernel (4 mm FWHM). GLMs were also constructed as before, but each group and member condition was split into one condition with trials from odd runs and another condition with trials from even runs (e.g., group mental odd, group mental even). Contrast images were created by comparing the parameter estimates of each mental condition to those of its corresponding physical condition in a linear contrast (e.g., *group mental odd > group*

physical odd). Following Misaki, Kim, Bandettini, and Kriegeskorte (2010), the multivariate searchlight analysis used the resulting contrast images (*group mental odd > group physical odd, group mental even > group physical even, member mental odd > member physical odd, member mental even > member physical even*) because they reduce the influence of noisy voxels.

For each voxel, we extracted the parameter estimates of each contrast within a spherical neighborhood (8-mm radius; neighborhood size in resampled voxels, $M = 248$, $SD = 21$) similar in shape to those used by Kriegeskorte and colleagues (2006). As such, each neighborhood was associated with four vectors, one for each contrast. These vectors were correlated in four different ways: two same-target correlations (*group odd with group even, member odd with member even*) and two different-target correlations (*group odd with member even, member odd with group even*). These correlations were Fisher-transformed to z values ($z = \frac{1}{2} * \ln(\frac{1+r}{1-r})$) and then averaged to yield a single same-target correlation and a single different-target correlation. For each neighborhood, we subtracted the average same-target correlation from the average different-target correlation and assigned the difference to the center voxel. This analysis yielded a correlation difference map expressed in z scores for each participant, indexing the degree to which each voxel exists in a neighborhood in which group trials correlate with each other more than with single member trials and, vice versa, single member trials correlate more with each other than group trials. Finally, a univariate, random-effects analysis identified brain regions that showed higher same- than different-target correlations across participants. For each voxel, we performed a right-tailed one-sample t test against zero with the corresponding z values from all participants (*same > different*).

Results

Behavioral Data

Means and standard deviations of responses and RTs are displayed in Table 1. On mental trials, participants indicated that groups ($M = 2.97$) would enjoy a long car ride together as much as single members ($M = 3.03$), $t(13) = 0.90$, $p = .39$, $d = 0.25$. On physical trials, participants did not judge groups ($M = 2.88$) differently from single members ($M = 2.94$), $t(13) = 0.28$, $p = .78$, $d = 0.08$. Response latencies were closely matched for mental judgments in the group ($M = 1.88$) and member ($M = 1.88$) conditions, $t(13) = 0.19$, $p = .85$, $d = 0.05$, as well as for physical judgments about groups ($M = 1.81$) and single members ($M = 1.89$), $t(13) = 1.41$, $p = .18$, $d = 0.39$.

Functional Imaging Data

Univariate analysis. Trials from the photo judgment task were conditionalized on the basis of question (mental,

physical) and target (group, member, individual), resulting in six conditions of interest. A whole-brain contrast identified voxels in which BOLD activity was greater during trials in which participants answered questions about the mental states of individuals than their physical appearance (*individual mental > individual physical*). Consistent with earlier research, this contrast yielded a set of brain regions that included dorsal and ventral MPFC, bilateral ATL, bilateral TPJ, and medial parietal cortex (Table 4).

The parameter estimates of the group and member conditions (*group mental > group physical, member mental > member physical*) were extracted for each of these brain regions, averaging over all voxels in each ROI (Table 3). To test that inferences about the mental states of groups recruit these brain regions, we examined whether BOLD activity in the parameter estimates of the group conditions were significantly higher than zero. This was the case for every ROI, all $ts(13) > 2.98$, all $ps < .01$, suggesting that people also recruit these regions when they represent the mental states of groups. To test that inferences about the mental states of groups recruit these brain regions as robustly as inferences about the mental states of single members, we examined whether differences in BOLD activity between the group and member parameter estimates were statistically unreliable. This was the case for every ROI, all $ts(13) < 1.70$, all $ps > .11$, suggesting that people recruit these brain regions as strongly when they represent the mental states of groups as when they perform similar inferences about individuals.

These results were replicated when ROIs were defined by the false belief localizer. A whole-brain contrast identified voxels in which BOLD activity was greater during blocks in which participants read and answered questions about belief stories than closely-matched stories about physical representations (*belief > photo*). Consistent with earlier research, this contrast yielded a set of brain regions that included dorsal MPFC, right ATL, bilateral TPJ, and medial parietal cortex (Table 4).

The parameter estimates of the group and member conditions (i.e., *group mental > group physical, member mental > member physical*) were extracted for each of these brain regions, averaging over all voxels in each ROI. To test that inferences about the mental states of groups recruit these brain regions, we examined whether BOLD activity in the parameter estimates of the group conditions were significantly higher than zero. This was the case for dorsal MPFC, right ATL, and bilateral TPJ, all $ts(13) > 2.18$, all $ps < .03$, suggesting that people also recruit these regions when they represent the mental states of groups. Medial parietal cortex showed a similar effect, but the difference was statistically reliable at a marginal level, $t(13) = 1.60$, $p = .06$. To test that inferences about the mental states of groups recruit these brain regions as robustly as inferences about the mental states of single members, we examined whether differences in BOLD activity between the group and member parameter estimates were statistically unreliable. This was

Table 4. Brain Regions Identified in Whole-brain, Random-effects Contrasts in the Photograph Judgment Task (*Individual Mental > Individual Physical*) and the False Belief Localizer (*Belief > Photo*) of Experiment 2, $p < .05$, Corrected for Multiple Comparisons

Region	<i>Individual Mental > Individual Physical</i>				<i>Belief > Photo</i>			
	<i>x</i>	<i>y</i>	<i>z</i>	<i>k</i>	<i>x</i>	<i>y</i>	<i>z</i>	<i>k</i>
MPFC								
Dorsal	-8	45	52	1221	8	61	26	699
Ventral	-8	55	2	807	-	-	-	-
Medial parietal cortex	-2	-51	22	860	10	-51	34	2491
Anterior temporal cortex								
Left	-48	-17	-24	813	-	-	-	-
Right	62	-7	-18	123	54	-3	-22	1484
TPJ								
Left	-54	-67	32	166	-60	-59	20	561
Right	66	-49	26	80	58	-53	30	888
Cerebellum								
Right	30	-83	-36	376	-	-	-	-
Left	-26	-79	-36	99	-24	-77	-38	333
STS	44	-45	6	263	-	-	-	-
Parahippocampal gyrus	-34	-31	-18	130	-	-	-	-
Inferior frontal gyrus (orbital)	-48	29	-14	116	-	-	-	-

From left to right, columns list the names of regions obtained from whole-brain, random-effects contrasts, the stereotaxic Montreal Neurological Institute coordinates of their peak voxels, and their size in number of voxels (*k*).

the case for every ROI, all $t_s(13) < 1.30$, all $p_s > .21$, suggesting that people recruit these regions as strongly when they represent the mental states of groups as when they perform similar inferences about individuals.

Multivariate searchlight analysis. A whole-brain contrast identified voxels in neighborhoods in which multivoxel correlations were higher for inferences about the mental states of the same class of targets (e.g., *group odd with group even*) than different classes of targets (e.g., *group odd with member even*). This *same > different* contrast yielded a set of brain regions that included dorsal MPFC, medial parietal cortex (precuneus and posterior cingulate cortex), left ATL, and left TPJ (Table 5, Figure 2). In these brain regions, the patterns of BOLD activity associated with inferences about the mental states of groups was more similar to the pattern for other groups than it was for individual members; likewise, the pattern of BOLD activity associated with inferences about the mental states of individuals was more similar to other individuals than to groups. The z values that correspond to the differences between same- and different-target correlations were reliably larger than zero, all $t_s(13) > 4.46$, all $p_s < 10^{-4}$. Moreover, the magnitude of these differences did not vary across ROIs, all $t_s < 1.65$, all $p_s > .12$, suggesting that these brain regions were equally sen-

sitive in their discrimination of groups and single members during inferences about mental states.

To ensure that these results were not caused by the group or the member condition exclusively, two additional multivariate searchlight analyses were conducted. A group analysis examined multivoxel correlations in the group condition (*group odd with group even*), whereas a member analysis examined multivoxel correlations in the member condition (*member odd with member even*). By analyzing the two same-target correlations separately, we can examine a correlation difference that detects the presence of a group representation and a correlation that detects the presence of a member representation. Each of these same-target correlations was independently contrasted against the average different-target correlation (*group odd with member even*, *member odd with group even*).

The correlation differences of the group analysis and the member analysis were extracted separately from the ROIs defined by the original multivariate searchlight analysis that averaged these two correlation differences. If these brain regions contain distinct group and member representations, then we should expect each of these two correlation differences to be reliably larger than zero in each ROI. This was the case in most ROIs (Figure 2). The correlation differences from the group analysis were

Table 5. Brain Regions Identified in Whole-brain, Random-effects Contrast *Same > Different* from the Multivariate Searchlight Analysis in Experiment 2, $p < .05$, Corrected for Multiple Comparisons

Region	x	y	z	k	t
Dorsomedial pFC	-6	41	32	133	3.78
Medial parietal cortex					
Precuneus	-18	-69	54	228	3.78
Posterior cingulate cortex	-8	-49	16	282	3.68
Middle frontal gyrus	-30	33	26	342	3.73
Left anterior temporal cortex	-46	9	-34	282	3.58
Postcentral gyrus	40	-27	54	142	3.58
Lateral superior frontal gyrus	22	5	56	775	3.56
	-18	1	66	165	3.53
Supramarginal gyrus	56	-31	34	166	3.56
Inferior frontal gyrus (orbital)	-34	27	-14	250	3.53
Left TPJ	-52	-39	24	133	3.44

From left to right, columns list the names of regions obtained from whole-brain, random-effects contrast, the stereotaxic Montreal Neurological Institute coordinates of their peak voxels, their size in number of voxels (k), and their mean weighted parameter estimate (t).

reliably larger than zero in all ROIs, all $ts(13) > 3.09$, all $ps < 10^{-3}$, suggesting that these brain regions contain distinct group representations. In the member analysis, the correlation differences were reliably larger than zero in medial parietal cortex (precuneus and posterior cingulate cortex), left ATL, and left TPJ, all $ts(13) > 2.84$, all $ps < 10^{-3}$, but not in dorsal MPFC, $t(13) > 1.23$, $p = .12$; however, the correlation difference in dorsal MPFC had the predicted direction. These results suggest that these ROIs contain distinct member representations. Together, the group analysis and the member analysis suggest that these brain regions, with the exception of dorsal MPFC, have distinct representations of groups and individuals.

The multivoxel patterns of these brain regions discriminate inferences about the mental states of groups from similar inferences about single members. To determine whether these brain regions discriminated between groups and single members with their univariate responses, parameter estimates from the contrasts (*group mental odd > group physical odd*, *group mental even > group physical even*, *member mental odd > member physical odd*, *member mental even > member physical even*) were extracted for each of these brain regions, averaging over all voxels in each ROI. Parameter estimates from odd and even runs of the group contrasts were averaged as were the parameter estimates from odd and even runs of the member contrasts. The differences between these two averages were not statistically reliable in any of the ROIs, all $ts(13) < 1.53$, all $ps > .15$, suggesting that the univariate,

unlike the multivariate, responses of these brain regions do not carry information about the target of the mental state inference.

Finally, an additional multivariate searchlight analysis was conducted to ensure that the results of the multivariate searchlight analysis were specific to the group-member distinction. Photographs were pseudorandomly assigned to two sets (X and Y) such that group and member versions of each photograph were always in the same set. Pseudorandom assignments varied across participants. Experimental trials were reassigned from old group member conditions to new X-Y conditions (e.g., if photograph 3 is in set X, then its group and member mental trials in the first run are in condition *X mental odd*). Critically, these new conditions contain an equal number of group and member trials. They were used to build new contrasts (*X mental odd > X physical odd*, *X mental even > X physical even*, *Y mental odd > Y physical odd*, *Y mental even > Y physical even*), which were submitted to a whole-brain multivariate searchlight analysis. At each neighborhood, the average of two different-set correlations (*X odd with Y even*, *X even with Y odd*) was subtracted from the average of two same-set correlations (*X odd with X even*, *Y odd with Y even*) and assigned to the center voxel. A whole-brain random-effects analysis to identify brain regions that showed greater same- than different-set correlations failed to dorsal MPFC, medial parietal cortex (posterior cingulate cortex or precuneus), left ATL, or left TPJ. Instead, it yielded a brain region in left posterior superior temporal gyrus ($[x\ y\ z] = -50, -47, 8$).

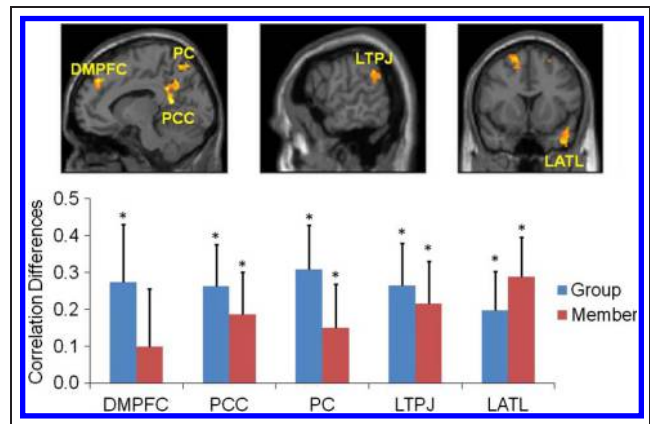


Figure 2. t map computed by the whole-brain, random-effects contrast *same > different* from the multivariate searchlight analysis in Experiment 2 and displayed in sagittal ($x = -10$, $x = -60$) and axial ($y = 13$) slices of a high-resolution anatomical template. The analysis identified dorsomedial pFC (DMPFC), two clusters in medial parietal cortex: precuneus (PC) and posterior cingulate cortex (PCC), left ATL (LATL), and left TPJ (LTPJ). Bar graphs display the mean differences of Fisher-transformed correlation averages for group trials (*group odd with group even > group odd with member even*, *member odd with group even*) and member trials (*member odd with member even > group odd with member even*, *member odd with group even*). Error bars represent 95% confidence intervals in within-subject comparisons (Masson & Loftus, 2003).

GENERAL DISCUSSION

People know that groups do not have minds, but nevertheless speak about groups as if they had mental states like those of individuals. In the present experiments, we observed that brain regions that respond preferentially during inferences about the mental states of individuals also responded robustly during inferences about the “mental states” of groups. Irrespective of how they were identified, dorsal and ventral MPFC, bilateral ATL, bilateral TPJ, and medial parietal cortex were more strongly engaged by inferences about the mental states of groups than by similar inferences about physical aspects of these groups. Moreover, these brain regions responded as robustly to inferences about mental states of groups as they did to similar inferences about individual group members. As a whole, these results suggest that not only do humans adopt a theory of mind about groups, but that it is just as robust as is theory of mind about individuals. As such, the results of the present experiment suggest that common brain regions underlie inferences about mental states of groups and individuals.

However, the fact that these brain regions are involved in representing the mental states of groups and individuals leaves open the question of whether these brain regions represent inferences about the mental states of groups and individuals similarly or differently. To determine whether these brain regions contain such target-specific representations (group vs. individual), we carried out a multivariate searchlight analysis to find brain regions that use multivoxel patterns to discriminate mental states inferences about groups from similar inferences about individuals. This analysis identified dorsal MPFC, medial parietal cortex (posterior cingulate cortex and precuneus), left ATL, and left TPJ as repositories of target identity in mental state inferences. These brain regions differentiated between inferences about the mental states of groups and single members despite the fact that the two conditions were closely matched in behavioral responses, RTs, information on screen, and univariate BOLD activity. Thus, although the same brain regions represent the mental states of groups and individuals, most of them carry information about whether the target of the mental state inference is a group or an individual.

One could argue that mental state inferences about groups and individuals recruit the same brain regions in this study because participants inferred the mental state of a single, representative group member when they were asked to mentalize a group. However, multivoxel patterns in most of these brain regions differentiate group and member trials. If participants were answering questions about an individual member in every trial, then multivoxel patterns would not differentiate between group and member trials. The MVPA suggests that these brain regions have distinct representations of groups and individuals.

In previous research, the brain regions that people recruit in inferences about the mental states of other

humans were also engaged by mentalizing about non-humans, such as objects (Wheatley et al., 2007; Castelli et al., 2000), robots (Krach et al., 2008), or other animals (Mitchell, Banaji, & Macrae, 2005). The present findings extend this body of work by showing that these brain regions can also respond robustly to inferences about the mental states of groups of people. Together, these experiments suggest that theory of mind is a cognitive module that can be flexibly deployed for understanding a diverse array of agents (Epley et al., 2007). Although it is likely that theory of mind evolved to understand the behavior of individual humans (Povinelli & Preuss, 1995), it may have been co-opted by the human mind to increase its understanding of the behavior of nonhuman agents and groups of people.

In summary, the present experiments contribute to our emerging understanding of how the human brain represents information about social groups. Although cognitive neuroscientists have started to document the neural basis of the perception of individuals from different social groups (for reviews, see Kubota, Banaji, & Phelps, 2012; Decety & Cacioppo, 2011; Todorov, Fiske, & Prentice, 2011; Van Bavel & Cunningham, 2011; Cunningham & Van Bavel, 2009; Ito & Bartholow, 2009; Eberhardt, 2005), only a few experiments have examined the functional neuroanatomy of perceiving and thinking about groups qua groups. However, this research gap is growing smaller. For example, recent reports have started to shed light on which brain regions represent stereotypes and other forms of semantic knowledge about social groups (Contreras et al., 2012; Quadflieg et al., 2009). Given the importance of groups to our everyday life as social animals, this burgeoning interest in the neural basis of intergroup cognition promises to become an important program of research in cognitive neuroscience.

Acknowledgments

The authors thank Amrita Dani, Katharine Dobos, Anna Leshinskaya, Joe Moran, and Jennifer Tao for advice and assistance. This work was supported by a grant to J. P. M. and M. R. B. from the National Science Foundation (BCS 0642448). J. M. C. was supported by a Jacob Javits fellowship from the United States Department of Education and a Graduate Research Fellowship from the National Science Foundation.

Reprint requests should be sent to Juan Manuel Contreras, Department of Psychology, Harvard University, Northwest Science Building, 52 Oxford Street, Cambridge, MA 02138, or via e-mail: jmcontr@fas.harvard.edu.

REFERENCES

- Ames, D. R. (2004a). Inside the mind-reader's tool kit: Projection and stereotyping in mental state inference. *Journal of Personality and Social Psychology*, 87, 340–353.
- Ames, D. R. (2004b). Strategies for social inference: A similarity contingency model of projection and stereotyping in attribute

- prevalence estimates. *Journal of Personality and Social Psychology*, 87, 573–585.
- Ames, D. R. (2005). Everyday solutions to the problem of other minds. In B. F. Malle & S. D. Hodges (Eds.), *Other minds: How human bridge the divide between self and others* (pp. 158–173). New York: Guilford Publications.
- Ames, D. R., & Mason, M. F. (2012). Mind perception. In S. T. Fiske & C. N. Macrae (Eds.), *The SAGE handbook of social cognition* (pp. 115–137). Thousand Oaks, CA: Sage.
- Ames, D. R., Weber, E. U., & Zou, X. (2012). Mind-reading in strategic interaction: The impact of perceived similarity on projection and stereotyping. *Organizational Behavior and Human Decision Processes*, 117, 96–110.
- Bottom, W. P., & Paese, P. W. (1997). False consensus, stereotypic cues, and the perception of integrative potential in negotiation. *Journal of Applied Social Psychology*, 27, 1919–1940.
- Brunet, E., Sarfati, Y., Hardy-Baylé, M. C., & Decety, J. (2000). A PET investigation of the attribution of intentions with a nonverbal task. *Neuroimage*, 11, 157–166.
- Castelli, F., Happe, F., Frith, U., & Frith, C. D. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage*, 12, 314–325.
- Centelles, L., Assaiante, C., Nazarian, B., Anton, J. L., & Schmitz, C. (2011). Recruitment of both the mirror and the mentalizing networks when observing social interactions depicted by point-lights: A neuroimaging study. *PLoS One*, 6, e15749.
- Contreras, J. M., Banaji, M. R., & Mitchell, J. P. (2012). Dissociable neural correlates of stereotypes and other forms of semantic knowledge. *Social Cognitive & Affective Neuroscience*, 7, 764–770.
- Cunningham, W. A., & Van Bavel, J. J. (2009). A neural analysis of intergroup perception and evaluation. In G. G. Berntson & J. T. Cacioppo (Eds.), *Handbook of neuroscience for the behavioral sciences* (pp. 975–984). New York: Wiley.
- Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping*, 8, 109–114.
- Decety, J., & Cacioppo, J. T. (2011). *The Oxford handbook of social neuroscience*. New York: Oxford University Press.
- Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: MIT Press.
- Eberhardt, J. L. (2005). Imaging race. *American Psychologist*, 60, 181–190.
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114, 864–886.
- Fletcher, P., Happe, F., Frith, U., Baker, S., Dolan, R., Frackowiak, R., et al. (1995). Other minds in the brain: A functional imaging study of “theory of mind” in story comprehension. *Cognition*, 57, 109–128.
- Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, 50, 531–534.
- Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of “theory of mind”. *Trends in Cognitive Sciences*, 7, 77–83.
- Gallagher, H. L., Happe, 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.
- Goel, V., Grafman, J., Sadato, N., & Hallett, M. (1995). Modeling other minds. *NeuroReport*, 6, 1741.
- Grinband, J., Wager, T. D., Lindquist, M., Ferrera, V. P., & Hirsch, J. (2008). Detection of time-varying signals in event-related fMRI designs. *Neuroimage*, 43, 509–520.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293, 2425–2430.
- Henson, R., Rugg, M. D., & Friston, K. J. (2001). The choice of basis functions in event-related fMRI. *Neuroimage*, 13, S149.
- Iacoboni, M., Lieberman, M. D., Knowlton, B. J., Molnar-Szakacs, I., Moritz, M., Throop, C. J., et al. (2004). Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. *Neuroimage*, 21, 1167–1173.
- Ito, T. A., & Bartholow, B. D. (2009). The neural correlates of race. *Trends in Cognitive Sciences*, 13, 524–531.
- Jones, T. E. (2010). *What people believe when they say that people believe: Folk sociology and the nature of group intentions*. Lanham, MD: Lexington Books.
- Knobe, J., & Prinz, J. (2008). Intuitions about consciousness: Experimental studies. *Phenomenology and the Cognitive Sciences*, 7, 67–83.
- Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS One*, 3, e2597.
- Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences, U.S.A.*, 103, 3863–3868.
- Kubota, J. T., Banaji, M. R., & Phelps, E. A. (2012). The neuroscience of race. *Nature Reviews Neuroscience*, 15, 940–948.
- Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology*, 57, 203–220.
- 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.
- Misaki, M., Kim, Y., Bandettini, P. A., & Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage*, 53, 103–118.
- Mitchell, J. P. (2009). Inferences about mental states. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, 364, 1309–1316.
- Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). General and specific contributions of the medial prefrontal cortex to knowledge about mental states. *Neuroimage*, 28, 757–762.
- Plous, S. (1993). The nuclear arms race: Prisoner’s dilemma or perceptual dilemma? *Journal of Peace Research*, 30, 163–179.
- Povinelli, D. J., & Preuss, T. M. (1995). Theory of mind: Evolutionary history of a cognitive specialization. *Trends in Neurosciences*, 18, 418–424.
- Premack, D. G., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1, 515–526.
- Quadflieg, S., Turk, D. J., Waite, G. D., Mitchell, J. P., Jenkins, A. C., & Macrae, C. N. (2009). Exploring the neural correlates of social stereotyping. *Journal of Cognitive Neuroscience*, 21, 1560–1570.
- Sagar, H. A., & Schofield, J. W. (1980). Racial and behavioral cues in Black and White children’s perceptions of ambiguously aggressive acts. *Journal of Personality and Social Psychology*, 39, 590–598.
- Saxe, R. (2009). Theory of mind (neural basis). In W. P. Banks (Ed.), *Encyclopedia of consciousness* (Vol. 2, pp. 401–410). Oxford, UK: Elsevier.
- Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in “theory of mind.” *Neuroimage*, 19, 1835–1842.

- Slotnick, S. D., Moo, L. R., Segal, J. B., & Hart, J., Jr. (2003). Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Cognitive Brain Research*, *17*, 75–82.
- Todorov, A. B., Fiske, S. T., & Prentice, D. A. (2011). *Social neuroscience: Toward understanding the underpinnings of the social mind*. New York City: Oxford University Press.
- Van Bavel, J. J., & Cunningham, W. A. (2011). A social neuroscience approach to self and social categorisation: A new look at an old issue. *European Review of Social Psychology*, *21*, 237–284.
- Wagner, D. D., Kelley, W. M., & Heatherton, T. F. (2011). Individual differences in the spontaneous recruitment of brain regions supporting mental state understanding when viewing natural social scenes. *Cerebral Cortex*, *21*, 2788–2796.
- Waytz, A., & Young, L. (2012). The group-member mind trade-off: Attributing mind to groups versus group members. *Psychological Science*, *23*, 77–85.
- Weil, R. S., & Rees, G. (2010). Decoding the neural correlates of consciousness. *Current Opinion in Neurology*, *23*, 649–655.
- Wheatley, T., Milleville, S. C., & Martin, A. (2007). Understanding animate agents: Distinct roles for the social network and mirror system. *Psychological Science*, *18*, 469–474.
- Zaitchik, D. (1990). When representations conflict with reality: The preschooler's problem with false beliefs and "false" photographs. *Cognition*, *35*, 41–68.