Identifying the What, Why, and How of an Observed Action: An fMRI Study of Mentalizing and Mechanizing during Action Observation

Robert P. Spunt[1], Ajay B. Satpute[2], and Matthew D. Lieberman[1]

Abstract

■ Humans commonly understand the unobservable mental states of others by observing their actions. Embodied simulation theories suggest that this ability may be based in areas of the fronto-parietal mirror neuron system, yet neuroimaging studies that explicitly investigate the human ability to draw mental state inferences point to the involvement of a "mentalizing" system consisting of regions that do not overlap with the mirror neuron system. For the present study, we developed a novel action identification paradigm that allowed us to explicitly investigate the neural bases of mentalizing observed actions. Across repeated viewings of a set of ecologically valid video clips of ordinary human actions, we manipulated the extent to which participants identified the unobservable mental states of the actor (*mentalizing*) or the observable mechanics of their behavior (*mechanizing*). Although areas of the mirror neuron system did show an enhanced response during action identification, its activity was not significantly modulated by the extent to which the observers identified mental states. Instead, several regions of the mentalizing system, including dorsal and ventral aspects of medial pFC, posterior cingulate cortex, and temporal poles, were associated with mentalizing actions, whereas a single region in left lateral occipito-temporal cortex was associated with mechanizing actions. These data suggest that embodied simulation is insufficient to account for the sophisticated mentalizing that human beings are capable of while observing another and that a different system along the cortical midline and in anterior temporal cortex is involved in mentalizing an observed action. ■

INTRODUCTION

Bodies are observable; minds are not. Despite this, when people look at their social world, they rarely understand the movements of others as expressions of a body. If we see a person giving a dollar to a homeless person, it is unnatural and incomplete to understand this as "gripping a dollar." Rather, we find higher meaning, we recognize their intention to help, and further, we are quite capable of using this information to infer that they likely *want* to help, *believe* (right or wrong) that money will help, and have a *generous* personality. Although work in the social neurosciences has made considerable progress investigating the neurocognitive mechanisms that enable an observer to recognize *what* others are doing (Rizzolatti & Craighero, 2004; Decety & Grezes, 1999), there is still debate over what mechanisms mediate an observer's higher level understanding of *why* they are doing it (Van Overwalle & Baetens, 2009; Gallese, 2007; Keysers & Gazzola, 2007; Saxe, 2005). That is, once the action is recognized, what enables the attribution of mental states that explain and/or accompany the action? In the present study, we developed a paradigm on the basis of action identification theory (AIT; Vallacher & Wegner, 1987) to investigate the neural bases of this ability to mentalize observed actions.

The most influential theory of the neural bases of action recognition is based on the finding that neurons in the macaque ventral premotor cortex and rostral inferior parietal lobule (rIPL) discharge both when the monkey performs an action and when it observes similar actions performed by others (Rizzolatti & Craighero, 2004). Human homologues to these regions in posterior inferior frontal gyrus (IFG), ventral premotor cortex, and rIPL have also been shown to exhibit this sensorimotor "mirror" property in studies that report regional brain activity common across executing and observing the same actions (e.g., Gazzola & Keysers, 2009). This mirror neuron system (MNS) is believed to implement a mechanism that matches observed actions to one's motor representations of similar actions. Once this match is made, the observer can then understand *what* the other is doing by simulating the actions and outcomes that are likely to follow from the observed motor act (Keysers & Gazzola, 2006).

Such a process of embodied simulation carried out in areas of the MNS may not only enable recognition of *what* others are doing but may also provide observers with an understanding of *why* (Gallese, 2007; Gallese & Goldman, 1998). Evidence for this proposition comes from studies showing that neurons in the macaque brain (Fogassi et al.,

[1]University of California, Los Angeles, [2]Columbia University

© 2010 Massachusetts Institute of Technology

2005) and MNS regions in the human brain (Hamilton & Grafton, 2008; Iacoboni, Molnar-Szakacs, Buccino, Mazziotta, & Rizzolatti, 2005) show sensitivity to features of an observed action's context that cue the actor's intentions. However, although these studies suggest that the MNS participates in mentalizing observed actions, neuroimaging studies that directly investigate mental state inference rarely observe activity in areas of the MNS. Instead, these studies typically observe regions along the cortical midline and in the temporal lobes collectively called the mentalizing system or theory-of-mind network, including areas of the medial pFC, TPJ, temporal poles, posterior cingulate cortex, and posterior STS (Lieberman, 2010; Van Overwalle & Baetens, 2009; Carrington & Bailey, 2009; Amodio & Frith, 2006; Gallagher & Frith, 2004). Why has neuroimaging revealed two systems in the brain for social cognition? One reason may be that the functional properties of these systems have thus far been investigated using dramatically different methods (for a notable exception, see Zaki, Weber, Bolger, & Ochsner, 2009). Researchers studying the MNS have typically used photographs and videos of simple hand–object interactions or whole-body movements performed in contextually impoverished contexts and have not explicitly manipulated and assessed the extent to which participants make mental state inferences during their tasks. In contrast, researchers studying the mentalizing system have typically used verbal or abstract visual stimuli such as cartoons or animations and rarely present participants with stimuli depicting real human behaviors (for recent reviews of these literatures and their methods, see Lieberman, 2010; Carrington & Bailey, 2009; Van Overwalle & Baetens, 2009). Given that the MNS is thought to rely on a mechanism that operates on the perception of embodied actions, it is plausible that the absence of MNS activity reported in extant work on mentalizing is due to the fact that participants in those studies were not presented with observable behaviors. Thus, it is difficult to conclude from existing research which neural system is most critical when participants are explicitly induced to mentalize an observed bodily action.

In the present study, we used a framework from AIT (Vallacher & Wegner, 1987) to control encoding strategies during action observation that varied in the extent to which they induced mentalizing about the observed behavior. AIT is predicated on the insight that the same action (e.g., "riding a bike") can be identified in numerous ways, with higher levels identifying why an action is performed (e.g., "getting exercise") and lower levels identifying how an action is performed (e.g., "gripping handlebars") (Figure 1A). Higher levels of identification refer not to observable motor actions but to unobservable mental states and dispositions that explain those actions and implicate them in a social context (Wegner & Vallacher, 1986), and the level on which another's behavior is identified is strongly associated with the tendency to attribute mentality to them (Kozak, Marsh, & Wegner, 2006). Given this, it is not surprising that researchers investigating action comprehension are becoming increasingly aware of the need for explicit control and/or assessment of the level on which participants encode observed actions (Van Overwalle & Baetens, 2009; Thioux, Gazzola, & Keysers, 2008; Grafton & Hamilton, 2007). Critically, AIT provides a framework within which such control can be achieved by manipulating, for the same action stimulus, the level on which it is identified.

We scanned participants' brains using fMRI while they passively observed fifteen 5-sec video clips of a male actor performing ordinary actions in natural settings; following this, participants watched each of the clips three more times at a different level of identification (LI) each time. During different blocks, participants were instructed to covertly identify *what* the target is doing (intermediate level), *why* he is doing it (high level), or *how* he is doing it (low level) (Figure 1B). Following the scan, participants were cued with still frames from each clip and asked to write down the descriptions they had generated while being scanned (Figure 1C). These descriptions were coded for LI by two individuals trained in the principles of AIT, and these codes were used to compute a post hoc LI parameter, which indexed the average LI in participants' descriptions for each block of the task. Given that this parameter reflects the degree of mental state content in participants' descriptions, it enables a determination of regions selectively sensitive to mental state inference during action observation.

METHODS

Participants

Eighteen right-handed, native English-speaking participants (9 women, mean age = 19.44 years, SD = 1.76 years) were recruited from the University of California, Los Angeles, subject pool and received financial compensation for participating. Data from three male participants were excluded before statistical analysis (n = 2, no response on all trials; n = 1, excessive head motion), leaving 15 participants (9 women, mean age = 19.47 years, SD = 1.88 years) in the statistical analysis.

Action Stimuli

Stimuli were selected from 35 video clips of a male actor (pictured in Figure 1) performing ordinary actions in natural scenes. The actor was instructed to perform each action in the manner he normally would while maintaining a neutral facial expression. After filming, each action clip was edited so that it was 5 sec long, silent, and included an object-directed hand action. To ensure that the actions depicted in the clips were familiar and easy to identify, we conducted a pilot study (n = 27; 21 women, mean age = 20.00 years, SD = 1.64 years) requiring participants to identify each of the 35 clips on the three levels

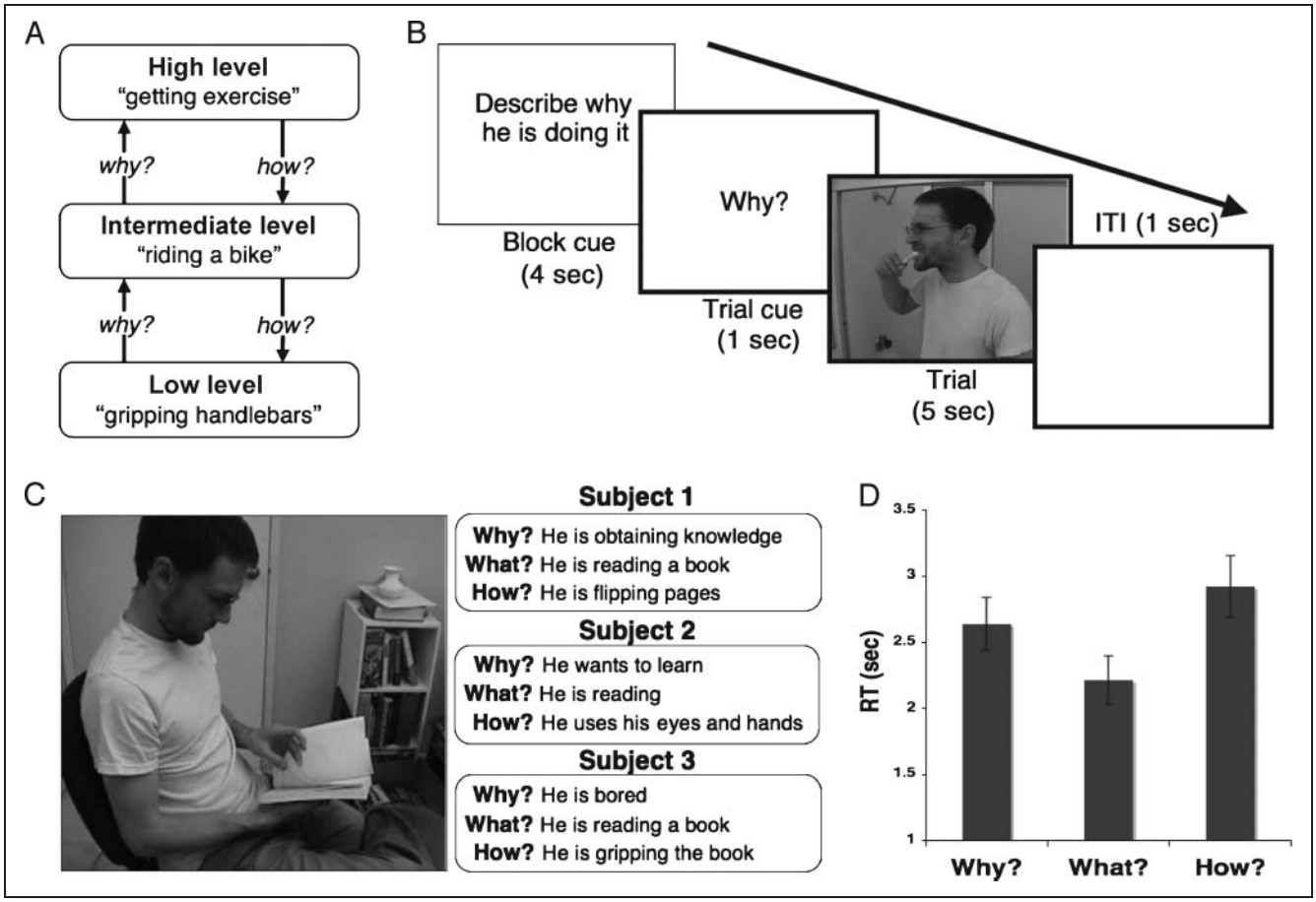

Figure 1. (A) An illustrative act identity structure. (B) The structure of a sample block (windows in bold represent the structure of one trial; each block contained five trials). (C) Depicts a frame from one stimulus and actual responses for three subjects at the three levels of identification. (D) Mean response time (RT) per trial as a function of level of identification. Error bars represent standard errors.

of the action hierarchy. For each clip, RTs were recorded, and participants rated how difficult it was for them to produce each identity. A set of 15 clips that featured the lowest ratings of difficulty and the fastest RTs at the three levels of identification was selected for inclusion in the neuroimaging study.

Action Identification Paradigm

The neuroimaging paradigm presented participants with each of the 15 clips four times for a total of 60 trials. Figure 1B depicts the block and trial structure. Trials were arranged in 12 blocks of five clips, and action orientation (observation vs. identification) as well as LI (high vs. intermediate vs. low) was manipulated across these blocks. During the first presentation of the clips (Blocks 1–3), we induced passive action observation by instructing participants to *watch what he is doing*. For the remaining nine blocks, we induced three levels of action identification by instructing participants to *describe what he is doing* (intermediate level), to *describe why he is doing it* (high level), or to *describe how he is doing it* (low level). For each of these trials, participants were instructed to covertly describe the clip on the level defined by the block, begin each description with the word "He," and make a right index finger button press once they completed their description. RT to the onset of the clip was recorded at button press. The order of identification levels was counterbalanced both within and across participants to control for order effects, and the use of the same stimuli in all four conditions controlled for stimulus effects. Blocks were separated by 15 sec of rest during which participants were instructed to attend to a fixation cross centered on screen.

Before scanning, participants performed a demonstration version of the task (featuring five video clips not included in the scanner paradigm) to ensure that they understood the task. Immediately following the scanning session, participants were given a booklet containing still frames from each stimulus and were asked to write down the descriptions they generated for each stimulus at each LI while in the scanner. Participants were instructed to skip responses they could not remember.

Image Acquisition

Images were acquired on a Siemens Allegra 3.0-T MRI scanner at the Ahmanson-Lovelace Brain Mapping Center,

University of California, Los Angeles. Stimulus presentation was implemented using MacStim (WhiteAnt Publishing, Melbourne, Australia), and stimuli were presented via magnet-compatible video goggles. Two functional scans lasting 304 and 318 sec were acquired (echo-planar T2-weighted gradient-echo, repetition time = 2000 msec, echo time = 25 msec, flip angle = 90°, matrix size = 64 × 64, 36 interleaved axial slices, field of view = 200 mm, 3 mm thick, voxel size = 3.1 × 3.1 × 3.0 mm). A set of high-resolution structural T2-weighted echo-planar images were acquired coplanar with the functional scans (spin-echo; repetition time = 5000 msec, echo time = 34 msec, matrix size = 128 × 128, 36 sagittal slices, field of view = 20 cm, 3 mm thick, voxel size = 1.6 × 1.6 × 3.0 mm).

Behavior Analysis

Participants' postscan descriptions were coded for LI by two independent raters. Raters were introduced to the concept of an action hierarchy and were then asked to code their perceived LI of all responses. The 54 responses to each stimulus (18 participants × 3 levels of identification) were presented in a random order along with a frame from the corresponding clip. Ratings were made on a 5-point scale (1 = *low*, 3 = *intermediate*, 5 = *high*). Interrater reliability was high ($r = .96$), and both raters provided ratings for all descriptions. We then used the average of the two coders' ratings for each response to compute an LI parameter for each participant, with values representing the mean LI for responses in each block. We also computed an RT parameter for each participant, with values representing the mean response time to trials in each block.

Image Analysis

The imaging data were analyzed using Statistical Parametric Mapping (SPM5, Wellcome Department of Cognitive Neurology, London). Images from each participant were realigned to correct for motion, normalized into the Montreal Neurological Institute space, and smoothed with an 8-mm Gaussian kernel, FWHM.

For each participant, two first-level models were specified with blocks modeled as boxcars convolved with a canonical hemodynamic response function. In the first model, observation and identification blocks were modeled separately, and we modulated the height of the hemodynamic response function for each identification block as a function of the average of the five LI and RT values per block. Appropriate linear contrasts were applied to the design to enable determination of regions selectively associated with the LI parameter controlling for values on the RT parameter. In the second model, each condition (observe, high level, intermediate level, and low level) was modeled separately, and appropriate linear contrasts were applied to the design to enable determination of regions active in the conjunction of all three levels of identification compared with passive observation. All first level contrast images were subjected to a random effects analysis to investigate effects at the group level. Unless otherwise reported, all results were interrogated with an uncorrected p value of .001 combined with a cluster size threshold of 30 voxels (Forman et al., 1995). Our use of cluster size thresholding combined with an uncorrected p value was based on the fact that our whole-brain searches were for regions identified a priori on the basis of past research investigating action understanding and mental state inference. All coordinates are reported in Montreal Neurological Institute space. For the purposes of visual presentation, the results in Figure 2 are overlaid on a canonical brain template, whereas the results in Figure 3 were surfaced rendered using the SPM toolbox SurfRend (http://spmsurfrend.sourceforge.net/).

ROI analyses were conducted in SPM5 in conjunction with the toolbox Marsbar (http://marsbar.sourceforge.net). For the analyses depicted in Figure 2C, single estimates for each ROI were produced by averaging the parameter estimates of all voxels included in the image. The ROIs depicted in Figure 2C were defined functionally on the basis of the analyses depicted in Figure 2A–B. Given that clusters in bilateral anterior temporal cortex did not exclusively occupy the temporal poles, we used the Wake Forest University Pickatlas anatomical toolbox (http://fmri.wfubmc.edu/cms/software#PickAtlas; Maldjian, Laurienti, Burdette, & Kraft, 2003) to create a bilateral temporal pole mask, which was used to isolate left and right temporal pole ROIs. The lateral parietal ROIs graphed in Figure 3B were extracted by masking the conjunction analysis depicted in Figure 3A with an ROI mask produced by growing a 10-mm sphere around parietal coordinates reported by Hamilton and Grafton (2008), whereas the right IFG ROI was created by masking the analysis in Figure 3A with a bilateral IFG mask taken from Iacoboni et al. (2005). The graph in Figure 2C displays parameter estimates for the comparison of each experimentally defined LI (i.e., as defined by the instruction given to participants at the beginning of each block) to fixation baseline. The graph in Figure 2B displays for the comparison of each experimentally defined LI to passive observation.

RESULTS

Behavioral Performance

The coding of postscan descriptions confirmed that participants were capable of discriminating among high ($M = 4.89$, $SD = .19$), intermediate ($M = 3.00$, $SD = .06$), and low ($M = 1.21$, $SD = .22$) levels of identification, $F(2, 280) = 1627.34$, $p < .001$, $\eta_p^2 = .59$, all post hoc ts(14) > 33.00, ps < .001. Average RT also differed as a function of LI, $F(2, 28) = 20.38$, $p < .001$, $\eta_p^2 = .59$. As depicted in Figure 1D, participants were faster to identify at intermediate ($M = 2.21$ sec, $SD = .71$ sec) than at both high

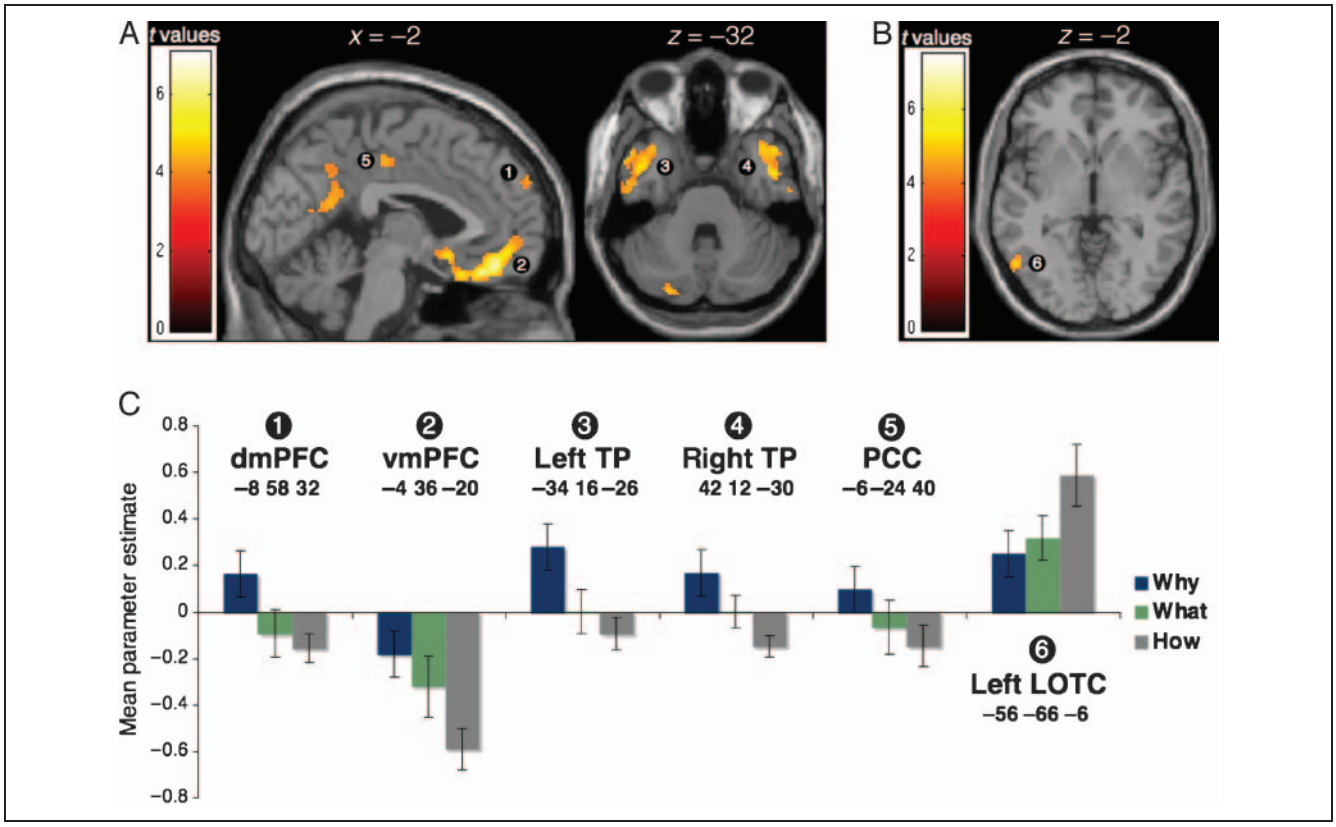

Figure 2. (A) Regions of the mentalizing system associated with increasing level of identification. (B) The region in left LOTC whose activity was associated with decreasing level of identification (numbers correspond to the ROI analysis below). (C) Mean parameter estimates in ROIs for each experimentally defined level of identification compared with fixation baseline. Error bars represent standard errors. dm = dorsomedial; vm = ventromedial; TP = temporal pole; PCC = posterior cingulate cortex; LOTC = lateral occipito-temporal cortex.

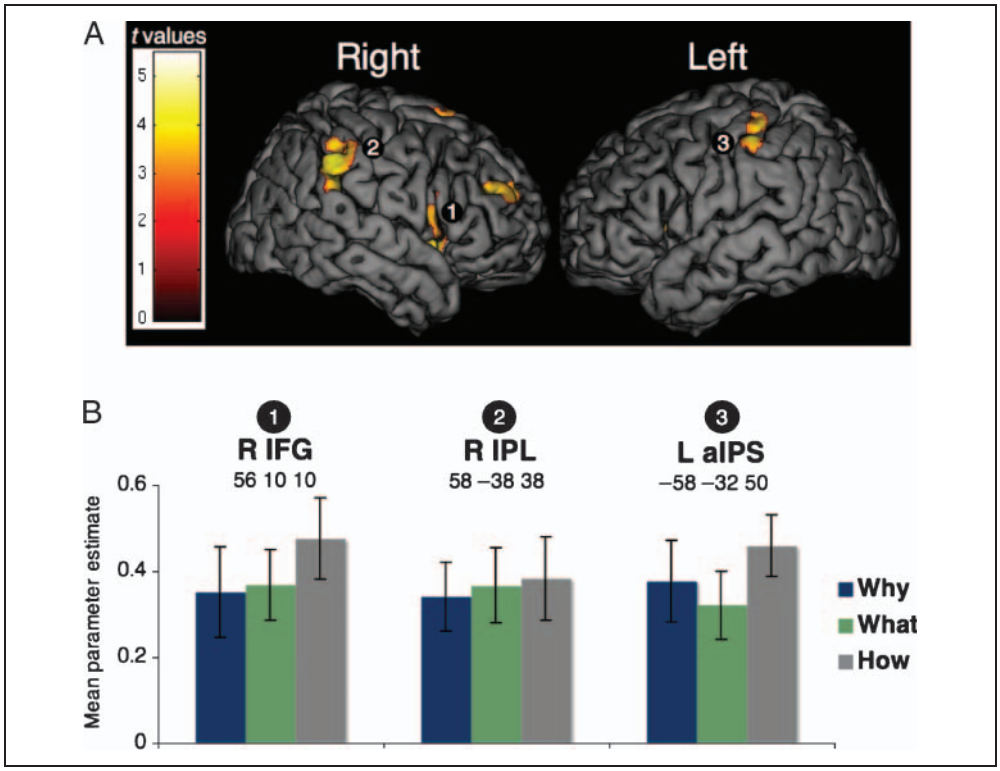

Figure 3. (A) Right IFG, IPL, and left aIPS engagement during all three levels of action identification compared with passive observation (numbers correspond to the ROI analyses presented below). (B) Mean parameter estimates in putative MNS ROIs for each level of identification compared with passive observation. Error bars represent standard errors. IFG = inferior frontal gyrus; IPL = inferior parietal lobule; aIPS = anterior intraparietal sulcus; L = left; R = right.

Table 1. Results of Whole-brain Analysis of Regions Associated with Increasing and Decreasing Values on the LI Parameter as well as the Simple Contrasts among Experimentally Defined Levels of Identification

Anatomical Region	*BA*		*x*	*y*	*z*	*t*	*k*
Increasing LI							
Cortical midline							
Dorsomedial pFC	9	L	−8	58	32	4.74	102
	6/8	L	−12	34	58	4.54	51
Ventromedial prefrontal cortex	11/32	L	−4	36	−20	7.06	672
Posterior cingulate cortex	31	L	−6	−24	40	4.71	62
Retrosplenial cortex	23/30	R	12	−44	20	5.24	305
Temporal cortex							
Lateral temporal cortex	20/21	L	−62	−4	−24	6.58	$1{,}088_a$
	20/21	R	46	0	−34	5.74	416_b
	21	R	54	−12	−20	5.53	194
Temporal pole	38	L	−34	16	−26	5.33	$1{,}088_a$
	38	L	−34	12	−46	4.56	41
	38	R	42	12	−30	6.16	416_b
Parahippocampal gyrus	30/27	R	14	−44	0	4.57	42
	35	R	24	−24	−16	4.54	38
Other regions							
Angular gyrus	39/19	L	−38	−82	34	4.98	90
	39/19	R	44	−78	38	5.18	95
Dorsolateral pFC	8	L	−28	24	46	4.75	39
	8	L	−26	40	44	4.56	41
Cerebellum	–	L	−24	−84	−32	5.12	42
Midbrain	–	L	−12	−16	−14	4.91	33
Decreasing LI							
Lateral occipito-temporal cortex	37/19	L	−56	−66	−6	7.51	89
Why Minus How							
Cortical midline							
Dorsomedial pFC	9	–	0	62	34	4.83	90
Ventromedial prefrontal cortex	11/32	L	−6	38	−20	10.16	660_a
Medial pFC	10	L	−6	52	−4	5.97	660_a
Posterior cingulate cortex	23	L	−2	−54	22	4.81	191
Temporal cortex							
Temporal pole	38	L	−38	7	−23	6.02	542
	38	R	36	14	−19	5.64	186
Lateral temporal cortex	21/22	L	−56	−38	0	4.89	53
	21/22	R	50	−12	18	5.59	119

Table 1. (*continued*)

Anatomical Region	*BA*		*x*	*y*	*z*	*t*	*k*
Other regions							
Ventral striatum	–	L	−6	14	−7	5.47	660_a
Angular gyrus	39/19	L	−40	−78	34	4.77	89
How Minus Why							
Lateral occipito-temporal cortex	37/19	L	−56	−68	−6	7.24	173
Why Minus What							
Temporal cortex							
Temporal pole	38	L	−44	18	−38	6.70	81
	38	L	−46	18	−16	6.50	337
	38	L	−54	6	−28	5.55	40
	38	R	30	6	−48	4.72	43
Inferior temporal cortex	20	L	−46	−4	−38	6.06	138
Posterior superior temporal sulcus	22/39	L	−50	−50	10	5.47	219
Frontal cortex							
Supplementary motor area	6	L	−10	16	64	6.66	359
Premotor cortex	6/8	L	−34	10	50	6.23	175
Inferior frontal gyrus (triangularis)	45/44	L	−46	18	16	5.47	142
Dorsolateral pFC	10/9	L	−36	54	26	4.79	40
Other regions							
Cerebellum	–	L	−18	−52	−18	5.60	68
Putamen	–	L	−20	8	6	4.92	39
Cuneus	17/19	L	−16	−82	8	4.87	53
What Minus Why							
No suprathreshold clusters							
What Minus How							
Precuneus/posterior cingulate	7/31	R	2	−52	28	5.41	495
Angular gyrus	39/19	R	37	−70	32	4.76	109
How Minus What							
Lateral occipito-temporal cortex	37/19	L	−52	−72	6	6.08	54
Inferior temporal cortex	20	L	−46	−18	−34	5.43	35
Rostral inferior parietal lobule	40	L	−54	−38	28	5.38	74
Supplementary motor area	6	L	−12	16	70	4.64	31
Premotor cortex	6/8	L	−52	10	44	4.40	31

BA = putative Brodmann's area; L and R = left and right hemispheres; *x*, *y*, and *z* = Montreal Neurological Institute coordinates in the left–right, anterior–posterior, and interior–superior dimensions, respectively; *t* = *t* score at those coordinates (local maxima); *k* = cluster size (in voxels); LI = level of identification. Regions with *k*s that share a subscript originate from the same cluster.

(M = 2.64 sec, SD = .77 sec) and low levels (M = 2.92 sec, SD = .90 sec), both $ts(14) > 4.51$, $ps < .001$. This finding is consistent with the AIT proposition that identities produced at this level are more prepotent than those produced at higher or lower levels (Wegner & Vallacher, 1986). Participants were also faster identifying at high than at low levels, $t(14) = 2.28$, $p = .039$. Finally, the number of missed trials did not differ as a function of LI, $F(2, 28) = 1.53$, $p = ns$.

Neural Regions Associated with Level of Action Identification

If areas of the MNS participate in mentalizing observed actions, one would expect a larger MNS response as task demands increase the need for mentalizing. To test this hypothesis, we used the LI parameter as a parametric modulator of block-related activity during action identification. Given that this parameter was based on the content of participants' actual responses to the stimuli, this provides a more sensitive test of regions associated with LI than contrasting activity on the basis of block instructions alone. A whole-brain analysis of regions associated with increasing or decreasing values on the LI parameter produced no activity in the MNS. Instead, several regions of the mentalizing system were associated with increasing values on the LI parameter (Figure 2A; for all regions observed in the whole-brain analysis, see Table 1). These include the dorsomedial pFC, the temporal poles bilaterally, and the posterior cingulate cortex, all of which have been shown to reliably respond to a wide range of tasks that require explicit mental state inferences (Lieberman, 2010). We also observed a large cluster in ventromedial pFC, a finding consistent with work suggesting that this region is important for at least some forms of mentalizing, for example, understanding the affective states of others (Shamay-Tsoory, Tibi-Elhanany, & Aharon-Peretz, 2006; Vollm et al., 2006). Each of the regions depicted in Figure 2A was also observed in the simple contrast of experimentally defined high- and low-level identification (Table 1). Finally, at a more liberal statistical threshold ($p < .005$, voxel extent = 30), we observed subpeaks in the right TPJ in both the parametric analysis ($x = 46, y = -64, z = 30, t = 4.19$) and the simple contrast of high- and low-level identification ($x = 46, y = -62, z = 30, t = 4.05$).

Parametrically decreasing activity (i.e., activity associated with increasingly lower levels of identification) was observed in a single cluster in left lateral occipito-temporal cortex (Figure 2B) in an area believed to be involved in the perception of motor actions, body parts, and tools (Noppeney, 2008; Peelen & Downing, 2007). This region was also the only one observed in the simple contrast of experimentally defined low- and high-level identification (Table 1). The pattern of activity displayed in Figure 2C suggests that activity in this region is subject to both bottom–up and top–down influences: Although it shows an enhanced response to the action stimuli regardless of the level on which they are identified, the magnitude of this response progressively increases as observers turn their attention toward the observable mechanics of the behavior.

To further investigate the possibility that subregions of the MNS were associated with LI, we conducted an ROI analysis of human MNS regions that previous studies suggest encode action goals and outcomes. We restricted the search space to these ROIs, using a small volume corrected p value of .05 (voxel extent = 5 voxels) and observed no evidence of modulation in these regions as a function of increasing or decreasing LI. Even with an uncorrected p value of .05 (voxel extent = 5 voxels), we observed no clusters associated with increasing LI but did observe clusters in bilateral posterior IFG (left: $x = -52$, $y = 12$, $z = 30$, $t = 2.13$, cluster size = 14 voxels; right: $x = 54$, $y = 14$, $z = 18$, $t = 2.27$, cluster size = 14 voxels), left inferior parietal lobule ($x = -52$, $y = -40$, $z = 30$, $t = 2.46$, cluster size = 32 voxels), and left anterior intraparietal sulcus ($x = -60$, $y = -32$, $z = 48$, $t = 2.16$, cluster size = 7 voxels) that were associated with decreasing LI. Although this result is too weak to be considered conclusive, it is consistent with findings showing that MNS activity is enhanced when an observer's attention is directed toward the *means*, compared with the *end*, of object-directed hand actions (Hesse, Sparing, & Fink, 2008).

Neural Regions Associated with Action Identification

Although regions of the MNS were not modulated by LI, a conjunction analysis of all three identification levels relative to passive observation did produce regions of the MNS. Figure 3A displays lateral fronto-parietal regions that were more active when identifying actions at all three levels compared with passively observing the same actions (for all regions observed in the whole-brain analysis, see Table 2). We observed activity in right posterior IFG, right inferior parietal lobule, and left anterior intraparietal sulcus that correspond to clusters reported in previous neuroimaging work on humans associating these regions with goal and outcome understanding during action observation (Hamilton & Grafton, 2008; Iacoboni et al., 2005). Figure 3B displays parameter estimates of activity at all three levels of identification compared with passive observation in the putative MNS regions that emerged in the conjunction analysis. In contrast to the clear modulation of the mentalizing system and lateral occipito-temporal cortex in response to change in LI (Figure 2C), these regions show no evidence of modulation by LI.

DISCUSSION

Taken together, these results suggest that mental state inference depends primarily on the mentalizing system and not areas of the MNS, even when these inferences are drawn from the observation of ecologically valid

Table 2. All Regions Observed in the Whole-brain Conjunction Analysis of All Three Experimentally Defined Levels of Action Identification Compared with Passive Observation of the Same Action Stimuli

Anatomical Region	*BA*		*x*	*y*	*z*	*t*	*k*
Frontal Cortex							
Inferior frontal gyrus (pars opercularis)	44	R	60	14	20	3.82	290_a
Supplementary motor area	6	R	8	10	64	4.81	164
Dorsolateral pFC	9	R	42	52	26	4.53	305
Parietal Cortex							
Inferior parietal lobule	40	R	60	−38	46	5.11	423
Anterior intraparietal sulcus	40	L	−58	−32	50	4.86	141
Other Regions							
Anterior insula	13	L	−36	12	10	5.48	131
	13	R	34	11	9	4.87	290

BA = putative Brodmann's Area; L and R = left and right hemispheres; *x*, *y*, and *z* = Montreal Neurological Institute coordinates in the left–right, anterior–posterior, and interior–superior dimensions, respectively; *t* = *t* score at those coordinates (local maxima); *k* = cluster size (in voxels). Regions with *k*s that share a subscript originate from the same cluster.

video clips of familiar human actions. As LI increased, thereby increasing mental state inference demands, so did activity in the mentalizing system. In contrast, there was no significant increase in MNS activity with increasing LI. Recent studies have produced results consistent with these, showing the involvement of the mentalizing system, but not the MNS, during the observation of moving objects in contexts that either did or did not encourage mental state attribution (Wheatley, Milleville, & Martin, 2007); the observation of actions performed in contexts that were parametrically varied to provide a plausible explanation for why the action was being performed (Brass, Schmitt, Spengler, & Gergely, 2007); judgments of whether actions depicted in photographs were guided by an unusual intention (de Lange, Spronk, Willems, Toni, & Bekkering, 2008); the instruction to attend to the end rather the means of object-directed hand actions (Hesse et al., 2008); and in both recognizing communicative intentions and generating communicative actions (Noordzij et al., 2009). However, this is the first study to simultaneously manipulate and measure LI during the observation of ordinary human actions performed in natural scenes, showing clearly that the mentalizing system, but not the MNS, is associated with mentalizing observed actions.

The regions associated with increasing LI show substantial overlap with a default mode network of the brain that exhibits high metabolic activity during periods of rest but which deactivates during stimulus-dependent, goal-directed tasks (Gusnard & Raichle, 2001). In the present study, low-level identification elicited the longest RTs, suggesting that the observed pattern of results might be explained by changes in task engagement rather than by changes in mental state inferential processing. This explanation is insufficient for two reasons. First, the parametric analysis presented in Figure 1 statistically controlled for variance in RT across blocks, effectively ruling out time on task as an explanatory variable. Second, although low-level identification featured the longest RTs, a default mode explanation would necessitate that high-level identification, which featured the highest amount of activity in default mode areas (Figure 2C), would feature the shortest RTs. Instead, intermediate-level identification featured the shortest RTs. As a result, we conclude that the observed pattern of data is best explained by variance in mental state inferential processing.

We observed no evidence for an association of areas of the MNS with mentalizing observed actions. However, this null result might be explained by the proposition that the MNS operates automatically during action observation (Van Overwalle & Baetens, 2009; Gallese, 2007; Keysers & Gazzola, 2006; Iacoboni et al., 2005). By this argument, the MNS always encodes the intention of an observed action, and thus the explicit inducement to draw mental state inferences would not produce any additional MNS activity. However, existing studies suggest the claim that the MNS operates automatically is unwarranted (Engel, Burke, Fiehler, Bien, & Rösley, 2008; Hesse et al., 2008; Jonas et al., 2007; Lee, Josephs, Dolan, & Critchley, 2006). In the present study, we included blocks involving the passive observation of a set of action stimuli as well as blocks involving the active identification of those same stimuli. If the operation of the MNS were automatic, its activity should show no difference across these two kinds of blocks. However, in Figure 3, we show that putative areas of the MNS show a differential response, suggesting non-automatic operation of at least some areas of the MNS.

Whether the operation of the MNS is automatic or not, several studies do suggest that the MNS does play a role in using information about an action's context to infer why the action is being performed (Hamilton & Grafton, 2008; Fogassi et al., 2005; Iacoboni et al., 2005). How are these prior studies to be reconciled with the present study, which suggests that identifying why an observed action is being performed is primarily associated with the mentalizing system? We suggest that this apparent inconsistency may result because these studies investigated action understanding at a relatively low level of abstraction, where the output of the process is not an unobservable mental state of the actor but rather an anticipated change in the state of the observable world, for instance, a physical consequence of the actor's movement (for a related discussion, see Hesse et al., 2008). For example, Hamilton and Grafton (2008) found that the response of putative human MNS to an observed motor action (e.g.,

a hand pushing a lid of a box) was conditioned by the as-yet-unobserved outcome of that action (e.g., a closed box). Here, what is understood is the concrete physical outcome of an observed movement, and as Grafton and Hamilton (2007) specify themselves, the highest level of understanding obtainable within the visuomotor system may be "the physical consequences of an action, for example altering the position or configuration of objects in the world" (p. 605). In a similarly concrete definition of high-level action understanding, Gallese (2007, p. 602) asserts that "…determining *why* a given act (e.g., grasping a cup) was executed [is] equivalent to detecting the goal of the still not executed and impending subsequent act (e.g., bringing the cup to the mouth)" (Gallese, 2007, p. 602). We suggest that it is not incorrect to characterize anticipating consequences such as "closing a box" or "bringing the cup to the mouth" as a primitive form of intention understanding that may be attributable to activity within the MNS, but it is one that we propose sits a relatively low level of abstraction, referring to a specific (observable) physical consequences of a specific (observable) object-directed hand action. On this point, we emphasize the validity of work by Hamilton and Grafton (2008; for a review, see Grafton & Hamilton, 2007), which indicates that the visuomotor system is organized hierarchically, with primarily visual areas in occipito-temporal cortex encoding low-level, kinematic properties of motor actions and areas in putative MNS, including posterior IFG and rIPL, encoding relatively higher level information about the objects of actions and their likely physical consequences. That these regions encode relatively high-level properties of motor actions is not tantamount to asserting that they underlie mental state inference, which requires the use of mental state concepts that by definition cannot be trusted to refer to perceptual and/or motor events (Leslie, 1987). As illustrated in Figure 1C, the high-level representations generated by participants in the present study contained mental state verbs like *try* and *want* and nouns like *knowledge* and *boredom*. As numerous theorists have noted, an embodied simulation mechanism is likely insufficient to represent such high-level concepts, which are defined by being difficult to refer to the physical world of motor events and perceptual objects (Mahon & Caramazza, 2008; Toni, de Lange, Noordzij, & Hagoort, 2008; Keysers & Gazzola, 2006).

In summary, we suggest that our results can be reconciled with these previous studies by the proposition that whereas areas of the MNS are capable of representing motor actions in terms of their goal objects and physical outcomes, the mentalizing system is necessary when an action cannot be understood on the basis of perceptual information alone (Brass et al., 2007), when there is no perceptual information presented regarding the action (as in the majority of previous work directly investigating mental state inference; Lieberman, 2010; Van Overwalle & Baetens, 2009), when the observer's attention is explicitly directed to the end state of an observed action (de Lange et al., 2008; Hesse et al., 2008), or when the observer attempts to identify action-related mental states that cannot be represented simply as changes in the state of the observable world (as when identifying actions on high levels, as in the present study).

Although extant neuroimaging work, including the present study, suggests that the MNS is most critically involved in low-level identification of actions, this does not preclude an MNS involvement in high-level identification of actions. As depicted in Figure 3, we observed engagement of areas of the MNS in response to all three levels of identification compared with passive observation of the same action stimuli, suggesting MNS involvement at all levels of the identification task. This pattern is consistent with recent propositions that the mirror neuron and the mentalizing systems may play complementary roles in understanding the actions of others, with the MNS encoding the observable, perceptual-motor properties of others and the mentalizing system interpreting those properties in terms of unobservable mental states and traits (de Lange et al., 2008; Thioux et al., 2008; Keysers & Gazzola, 2007; Lieberman, 2007; Olsson & Ochsner, 2007; see also Zaki et al., 2009). This model fits with classic theories of social cognition that treat behavior identification as the first step in the inferential process of understanding the mental states and dispositions of others (Heider, 1958; for a review, see Gilbert, 1998). Despite the appeal of such an integrative model, it is important to emphasize that MNS involvement in social cognition is likely limited to situations where perceptual-motor information about others must first be encoded before mental state inferential processing can begin, for instance when attempting to attribute a mental state to an observed behavior, as in the present study. Absent such information, we speculate that the MNS does not participate in social cognition, or at least its role is dramatically minimized. Of course, it should also be emphasized that our observed pattern does not conclusively indicate that the mentalizing system depends on the MNS for its inputs during high-level action identification. Instead, the two systems might be functionally independent. Further research is needed to directly test for functional interdependence of these systems during high-level social cognition.

In conclusion, we report evidence that mentalizing an observed action is carried out by areas of the mentalizing system and suggest that areas of the MNS may be necessary, but not sufficient, when the object of mentalizing is an embodied behavior. Future research should be directed at more clearly delineating the roles of the MNS and mentalizing systems in social cognition. Such research will profit from the insight that the actions of others are complex stimuli that can be represented on multiple levels of an action hierarchy (Van Overwalle & Baetens, 2009; Thioux et al., 2008; Grafton & Hamilton, 2007; Kozak et al., 2006; Vallacher & Wegner, 1987). Actors are not just moving bodies; they are also moving minds. Although the MNS may be sufficient to understand actions as expressions of a body, additional recruitment of the mentalizing system

rather than the MNS appears to be necessary for the sophisticated human ability to understand actions as expressions of an unobservable mind.

Acknowledgments

The authors thank Raynee Gutting, Leanna Constantini, Adam Zika, and Mariana Preciado for their assistance and Marco Iacoboni and Jonas Kaplan for the frontal MNS ROIs. For generous support, the authors also thank the Brain Mapping Medical Research Organization, the Brain Mapping Support Foundation, the Pierson-Lovelace Foundation, the Ahmanson Foundation, the William M. and Linda R. Dietel Philanthropic Fund at the Northern Piedmont Community Foundation, the Tamkin Foundation, the Jennifer Jones-Simon Foundation, the Capital Group Companies Charitable Foundation, the Robson Family, and the Northstar Fund.

Reprint requests should be sent to Matthew D. Lieberman, Department of Psychology, University of California, Los Angeles, 1285 Franz Hall, Los Angeles, CA 90095-1563, or via e-mail: lieber@ucla.edu.

REFERENCES

Amodio, D. M., & Frith, C. D. (2006). Meeting of the minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7,* 268–278.

Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007). Investigating action understanding: Inferential processes versus action simulation. *Current Biology, 17,* 2117–2121.

Carrington, S. J., & Bailey, A. J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping, 30,* 2313–2335.

de Lange, F. P., Spronk, M., Willems, R. M., Toni, I., & Bekkering, H. (2008). Complementary systems for understanding action intentions. *Current Biology, 18,* 454–457.

Decety, J., & Grezes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends in Cognitive Science, 3,* 172–178.

Engel, A., Burke, M., Fiehler, K., Bien, S., & Rösley, R. (2008). What activates the human mirror neuron system during observation of artificial movements: Bottom–up visual features or top–down intentions? *Neuropsychologia, 46,* 2033–2042.

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersim, F., & Rizzolatti, G. (2005). Parietal lobe: From action organization to intention understanding. *Science, 308,* 662–667.

Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnetic Resonance in Medicine, 33,* 636–647.

Gallagher, H. L., & Frith, C. D. (2004). Functional imaging of "theory of mind." *Trends in Cognitive Sciences, 7,* 77–83.

Gallese, V. (2007). Before and below "theory of mind": Embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 362,* 659–669.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 2,* 493–501.

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

Gilbert, D. T. (1998). Ordinary personology. In D. T. Gilbert, S. T. Fiske, & G. Lindsey (Eds.), *Handbook of social psychology* (4th ed., pp. 89–150). New York: Oxford Univ. Press.

Grafton, S. T., & Hamilton, A. F. C. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Human Movement Science, 26,* 590–616.

Gusnard, D. A., & Raichle, M. E. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nature Reviews Neuroscience, 2,* 685–694.

Hamilton, A. F., & Grafton, S. T. (2008). Action outcomes are represented in human inferior frontoparietal cortex. *Cerebral Cortex, 18,* 1160–1168.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: John Wiley.

Hesse, M. D., Sparing, R., & Fink, G. R. (2008). Ends or means—The "what" and "how" of observed intentional actions. *Journal of Cognitive Neuroscience, 21,* 776–790.

Iacoboni, M., Molnar-Szakacs, I., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror system. *PLoS Biology, 3,* e79.

Jonas, M., Siebner, H. R., Biermanm-Ruben, K., Kessler, K., Baumer, T., Buchel, C., et al. (2007). Do simple intransitive finger movements consistently activate frontoparietal mirror neuron areas in humans? *Neuroimage, 36,* T44–T55.

Keysers, C., & Gazzola, V. (2006). Towards a unifying theory of social cognition. *Progress in Brain Research, 156,* 379–401.

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

Kozak, M. N., Marsh, A. A., & Wegner, D. M. (2006). What do I think you're doing? Action identification and mind attribution. *Journal of Personality and Social Psychology, 90,* 543–555.

Lee, T. W., Josephs, O., Dolan, R. J., & Critchley, H. D. (2006). Imitating expressions: Emotion-specific neural substrates in facial mimicry. *Social Cognitive and Affective Neuroscience, 1,* 122–135.

Leslie, A. M. (1987). Pretense and representation: The origins of "Theory of Mind." *Psychological Review, 94,* 412–426.

Lieberman, M. D. (2007). Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology, 58,* 259–289.

Lieberman, M. D. (2010). Social cognitive neuroscience. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *The handbook of social psychology* (5th ed., pp. 143–193). New York: McGraw-Hill.

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology (Paris), 102,* 59–70.

Maldjian, J. A., Laurienti, P. J., Burdette, J. B., & Kraft, R. A. (2003). An automated method for neuroanatomic and cytoarchitechtonic atlas-based interrogation of fMRI data sets. *Neuroimage, 19,* 1233–1239.

Noordzij, M. L., Newman-Norlund, S. E., de Ruiter, J. P., Hagoort, P., Levinson, S. C., & Toni, I. (2009). Brain mechanisms underlying human communication. *Frontiers in Human Neuroscience, 3,* 1–13.

Noppeney, U. (2008). The neural systems of tool and action semantics: A perspective from functional imaging. *Journal of Physiology (Paris), 102,* 40–49.

Olsson, A., & Ochsner, K. N. (2007). The role of social cognition in emotion. *Trends in Cognitive Sciences, 12,* 65–71.

Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews Neuroscience, 8,* 636–647.

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

Saxe, R. (2005). Against simulation: The argument from error. *Trends in Cognitive Science, 9,* 174–179.
Shamay-Tsoory, Y., Tibi-Elhanany, Y., & Aharon-Peretz, J. (2006). The ventromedial prefrontal cortex is involved in understanding affective but not cognitive theory of mind. *Social Neuroscience, 1,* 149–166.
Thioux, M., Gazzola, V., & Keysers, C. (2008). Action understanding: How, what and why. *Current Biology, 18,* 431–434.
Toni, I., de Lange, F. P., Noordzij, M. L., & Hagoort, P. (2008). Language beyond action. *Journal of Physiology (Paris), 102,* 71–79.
Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review, 94,* 3–15.
Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *Neuroimage, 48,* 564–584.
Vollm, B. A., Taylor, A. N., Richardson, P., Corcoran, R., Stirling, J., McKie, S., et al. (2006). Neuronal correlates of theory of mind and empathy: A functional magnetic resonance imaging study in a nonverbal task. *Neuroimage, 29,* 90–98.
Wegner, D. M., & Vallacher, R. (1986). Action identification. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition* (Vol. 1, pp. 550–582). New York: Guilford.
Wheatley, T., Milleville, S. C., & Martin, A. (2007). Understanding animate agents: Distinct roles for the social network and mirror systems. *Psychological Science, 18,* 469–474.
Zaki, J., Weber, J., Bolger, N., & Ochsner, K. (2009). The neural bases of empathic accuracy. *Proceedings of the National Academy of Sciences, U.S.A., 106,* 11382–11387.