Behavioral/Systems/Cognitive

Dissociating Modality-Specific and Supramodal Neural Systems for Action Understanding

Robert P. Spunt and Matthew D. Lieberman

Department of Psychology, University of California, Los Angeles, Los Angeles, California 90095-1563

The neural basis of action understanding in humans remains disputed, with some research implicating the putative mirror neuron system (MNS) and some a mentalizing system (MZS) for inferring mental states. The basis for this dispute may be that action understanding is a heterogeneous construct: actions can be understood from sensory information about body movements or from language about action, and with the goal of understanding the implementation ("how") or motive ("why") of an action. Although extant research implicates the MNS in understanding implementation and the MZS in understanding motive, it remains unknown to what extent these systems subserve modality-specific or supramodal functions in action understanding. While undergoing fMRI, 21 volunteers considered the implementation ("How is she doing it?") and motive ("Why is she doing it?") for actions presented in video or text. Bilateral parietal and right frontal areas of the MNS showed a modality-specific association with perceiving actions in videos, while left-hemisphere MNS showed a supramodal association with understanding implementation. Largely left-hemisphere MZS showed a supramodal association with understanding motive; however, connectivity among the MZS and MNS during the inference of motive was modality specific, being significantly stronger when motive was understood from actions in videos compared to text. These results support a tripartite model of MNS and MZS contributions to action understanding, where distinct areas of the MNS contribute to action perception ("perceiving what") and the representation of action implementation ("knowing how"), while the MZS supports an abstract, modality-independent representation of the mental states that explain action performance ("knowing why").

Introduction

The neural basis of action understanding in humans remains disputed (Gallese et al., 2011). The debate has centered on clarifying the role of the mirror neuron system (MNS), which in humans refers to brain regions that activate during both the observation and execution of actions (Van Overwalle and Baetens, 2009; Rizzolatti and Sinigaglia, 2010; Keysers et al., 2011). Although the human MNS is reliably active during action observation, several studies have shown that it is not sensitive to the demand to explain observed actions; rather, a separate brain system known as the theory-of-mind or mentalizing system (MZS) appears to be critical (Grèzes et al., 2004; Brass et al., 2007; de Lange et al., 2008; Spunt et al., 2011).

Importantly, there still remains little consensus in the literature on what is involved in the act of understanding an action (Gallese et al., 2011; Kilner, 2011; Uithol et al., 2011). This may be because action understanding is a heterogeneous psychological construct that encompasses not one but many ecological contexts in which humans consider the actions of other humans (Vallacher and Wegner, 1987; Kilner, 2011; Uithol et al., 2011). One ecological variable is the observer's comprehension goal. In some cases, action understanding entails understanding how an action is, was, or will be implemented, for instance when attempting to imitate the action of another. In other cases, actions are represented with the goal of understanding the actor's motive—that is, why the action is being performed. A second important ecological variable is the modality through which the action becomes an object of cognition. Specifically, actions are abstract conceptual objects easily understood through language or by simply watching others act. Given these distinct types of inputs, the modality through which an action is apprehended is likely a critical variable in determining the neural systems that support a given instance of action understanding.

These ecological variables are orthogonal: actions, regardless of the modality through which they are apprehended, can be conceived with the goal to understand how or why. Hence, the two factors can be factorially combined to produce a matrix of four contexts of action understanding (see Fig. 1A). Research groups investigating the neural bases of action understanding have typically employed paradigms that capture only one or two of these contexts (Van Overwalle and Baetens, 2009; Zaki and Ochsner, 2009). During action understanding, several studies have manipulated the observer's goal (Decety et al., 1997; de Lange et al., 2008; Spunt et al., 2010, 2011), and one study has manipulated across the sensory and linguistic modalities; however, this study was not designed to directly contrast the modal-

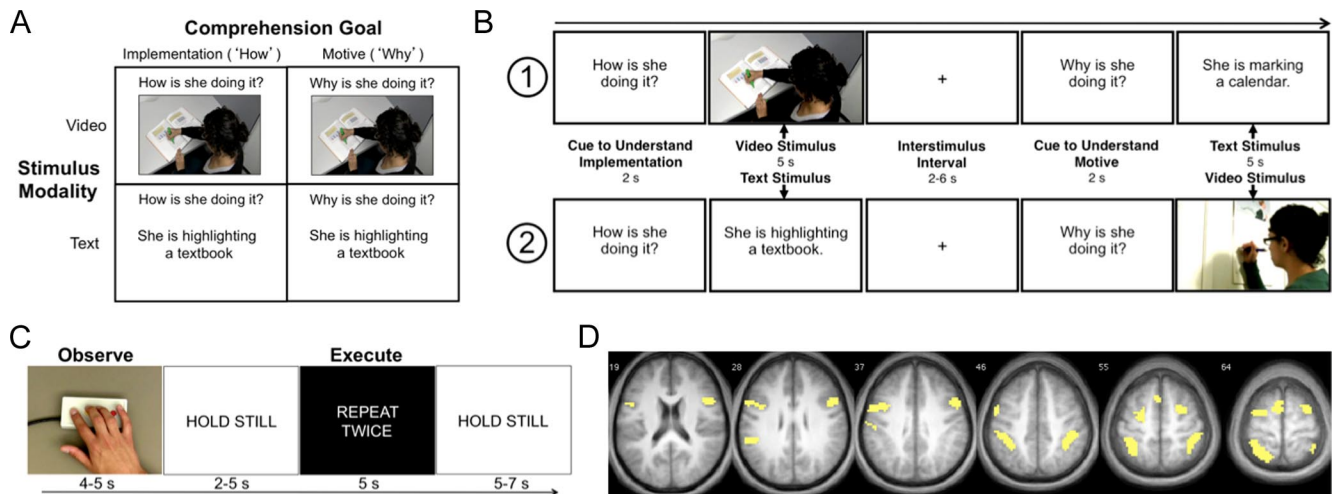**Figure 1.** ***A***, Four contexts in which humans consider the actions of others, and the design of the action understanding task. ***B***, Schematic of the action understanding task. Each row features two trials from one of two versions of the task to which participants were randomly assigned. ***C***, Schematic of the mirror neuron system localizer task. ***D***, Statistical parametric maps of the frontoparietal regions emerging when testing the conjunction null of Observe > Fixation and Execute > Fixation (corrected for multiple comparisons using a voxel-level familywise error rate of 0.05 combined with a voxel extent of 20).

ities (Aziz-Zadeh et al., 2006). To date, no study has simultaneously manipulated goal and modality while attempting to hold action content constant across conditions. Therefore, we designed a novel paradigm for investigating action understanding that faithfully reproduces these four contexts as they might occur in daily life. Within this paradigm, we aimed to determine the neural systems independently sensitive to stimulus modality (sensory vs linguistic) and comprehension goal (how vs why).

Materials and Methods

Participants
Twenty-one right-handed participants (12 females, mean age = 21.71, age range = 19–32) were recruited from the University of California, Los Angeles (UCLA) subject pool and provided written informed consent according to the procedures of the UCLA Institutional Review Board. All participants were native English speakers and were not taking psychoactive medications at the time of the study.

Experimental stimuli
Action understanding task. The complete set of stimuli used in the action understanding task consisted of 48 video–text pairs. These stimulus pairs were produced using the following procedure. First, we filmed a female actor perform 78 common object-directed actions in natural scenes. During filming, the actor was instructed to maintain neutral affect, and each clip was framed to make salient at least one object-directed hand action. After filming, all clips were edited to be silent and 5 s in duration. Our next step was to produce empirically matched text descriptions of the actions in each clip. To do so, we had 26 UCLA undergraduates view the 78 edited clips while seated at a computer station. For each clip, participants verbally identified the action in each clip by typing a response to the question "What is she doing?" The only constraint put on responses was that they begin with the string "She is." We then selected the 48 clips that produced the highest interobserver agreement in these verbal identifications, and the modal response for each clip was then used to generate the matched text stimulus. For the final set of 48 clips, each of the paired text stimuli was provided by at least 65% of the pilot participants, and the average percentage of observers who displayed agreement was 85%.

MNS localizer task. In addition to the primary action understanding task, all participants performed an additional task that allowed us to independently define the MNS in our sample. The stimulus set consisted of 12 videos of a male actor performing button press sequences on a four-button box. Each clip was filmed from a first-person perspective

and narrowly framed on the hand action. After filming, each clip was edited to be silent and 4–5 s in duration.

Experimental design
Action understanding task. Participants were subjected to a 2 (Stimulus Modality: Video vs Text) × 2 (Comprehension Goal: Implementation vs Motive) within-subjects factorial design, resulting in four experimental conditions (Fig. 1*A*). To manipulate stimulus modality, the 48 video–text pairs described above were evenly divided into two sets. Each participant was randomly assigned to receive one set in video format and the other in text format. For example, approximately half of the participants received the action "She is brushing her teeth" as a video and "She is playing a guitar" as a text stimulus, while the other half received "She is brushing her teeth" as a text and "She is playing a guitar" as a video stimulus (Fig. 1*B*). This was done to counterbalance the action identities across the two stimulus modalities, so that across the group, the video and text conditions featured the same actions.

To manipulate comprehension goal, participants were cued before each trial to answer one of two questions. When the goal was to understand implementation, participants were presented with the question "How is she doing it?" and were asked to silently think of one important part of performing the action. When the goal was to understanding the actor's motive, participants were presented with the question "Why is she doing it?" and were asked to silently think of one plausible motive the actor has for performing the action.

MNS localizer task. Participants underwent two conditions (Fig. 1*C*). In the Observation condition, participants were instructed to passively observe a video clip of a hand performing a sequence of button presses on the same MR-compatible four-button box that they were holding in their right hand. In the Execution condition, participants were cued to repeat the sequence with their right hand.

Experimental procedure
Before scanning, participants were introduced to and trained in both experimental tasks. For the action understanding task, this included performing three trials from each condition (using stimuli not featured in the scanner task) while the experimenter monitored performance. During debriefing, all participants reported complete comprehension of the experimental tasks.

The structure of the action understanding task is depicted in Figure 1*B*. Before each trial, participants were cued to either understand implementation or motive. Once the stimulus appeared, participants were instructed to silently think of their response as quickly as possible and to make a right index finger button press once they had their response in

Table 1. Five most frequent verbs used in response to the task as a function of observer comprehension goal, along with examples of use

Comprehension goal			
Understand motive ('Why')		Understand implementation ('How')	
Verb	Example of use	Verb	Example of use
To want	Wants to be fit	To use	Using her hand to brush
To need	Needs to rest	To hold	Holding the knife
To go	Going to a party	To put	Putting into the mailbox
To get	Getting ready for prom	To move	Moving iron back and forth
To have	Has a job interview	To grip/grasp	By gripping the controller

Responses were collected post-scan.

Table 2. All regions observed in the group conjunction of Observe > Fixation and Execute > Fixation in the mirror neuron system localizer task

Anatomical region	k	L/R	MNI x	y	z	t	p_{FWE}
Whole-brain analysis							
Posterior inferior frontal gyrus	166	R	54	11	31	5.91	<0.001
Rostral inferior parietal lobule	240	R	45	− 34	40	5.90	<0.001
Anterior intraparietal sulcus	−	R	39	− 52	55	5.45	0.001
Posterior inferior frontal gyrus	174	L	− 54	2	37	5.80	<0.001
Posterior inferior frontal gyrus	−	L	− 51	8	13	5.26	0.003
Superior parietal lobule	367	L	− 24	− 61	64	5.78	<0.001
Anterior intraparietal sulcus	−	L	− 42	− 37	46	5.73	<0.001
Rostral inferior parietal lobule	−	L	− 54	− 22	34	4.83	0.017
Anterior insula/frontal operculum	35	R	36	23	7	5.73	<0.001
Rostral inferior parietal lobule	56	L	− 57	− 40	22	5.69	<0.001
Supplementary motor area	264	L	− 9	2	67	5.63	<0.001
Dorsal premotor cortex	−	L	− 36	− 1	58	5.51	0.001
Dorsal premotor cortex	−	L	− 21	− 13	52	5.38	0.002
Supplementary motor area	−	L	− 6	14	46	4.90	0.012
Dorsal premotor cortex	123	R	18	5	67	5.30	0.002
Cerebellum	29	R	36	− 67	− 23	5.07	0.006

All results are cluster-level corrected using a familywise error (FWE) rate of 0.05. Coordinates are all local maxima observed that were separated by at least 20 mm. L/R, left and right hemispheres; x, y, and z, Montreal Neurological Institute (MNI) coordinates in the left–right, anterior–posterior, and inferior–superior dimensions, respectively; t, t statistic value at those coordinates; k, cluster voxel extent; p_{FWE}, voxel-level FWE for the local peak.

mind. Response time (RT) was recorded at this button press. Each stimulus remained onscreen for a maximum duration of 5 s; if the participant responded before 5 s, the stimulus was replaced with a fixation cross for the remainder of the trial. The order of trials was optimized for both estimation and contrast efficiency using a genetic algorithm (Wager and Nichols, 2003). Trials were separated by a fixation screen of variable duration (sampled from an exponential distribution; range = 2000–6000 ms, mean = 3000 ms). Following the scan, participants performed the task a second time and typed their answers to the Why and How questions on a keyboard.

The structure of the MNS localizer task is depicted in Figure 1C. Each trial always began with the participant passively observing a button press sequence. Participants were then presented with a screen displaying the words "HOLD STILL." Next, participants were given 5 s to accurately repeat the sequence observed in the clip either once or twice. To facilitate task engagement, feedback was provided (with the words "Correct" or "Incorrect" centered onscreen) following each execution period. Each trial ended with a screen displaying the words "HOLD STILL."

The MATLAB (The MathWorks) Psychophysics Toolbox (Brainard, 1997) version 7.4 was used to present the stimuli to participants and record their responses. Participants viewed the task through MR-compatible LCD goggles.

Image acquisition
Imaging data were acquired using a Siemens Trio 3.0 tesla MRI scanner at the UCLA Ahmanson-Lovelace Brain Mapping Center (Los Angeles, CA). For each participant we acquired 1216 functional T2*-weighted echo planar image volumes (EPIs; slice thickness = 3 mm, gap = 1 mm, 36 slices, TR = 2000 ms, TE = 25 ms, flip angle = 90°, matrix = 64 × 64, FOV = 200 mm). The action understanding task was performed in two runs (each collecting 290 volumes). The MNS localizer task was performed in a single functional run (152 measurements). The final 484 volumes were collected for the purposes of another investigation. We also acquired a T2-weighted, matched-bandwidth anatomical scan (same parameters as EPIs, except: TR = 5000 ms, TE = 34 ms, flip angle = 90°, matrix = 128 × 128) and a T1-weighted, magnetization-prepared, rapid-acquisition, gradient echo anatomical scan (slice thickness = 1 mm, 176 slices, TR = 2530 ms, TE = 3.31 ms, flip angle = 7°, matrix = 256 × 256, FOV = 256 mm).

Behavior analysis
MATLAB was used to analyze all behavioral data. For each participant, we computed the mean RT for each of the four conditions and used a repeated measures ANOVA to test main effects of modality and goal and their interaction on RT. In addition, we used custom MATLAB software to examine the verbs used in the four conditions. First, we combined all participants' post-scan responses and computed the frequency of each verb (summing across different forms of each verb). For each goal, we selected the five most frequently used verbs (Table 1). Next, we used paired samples t tests to test for frequency of use differences of each verb across the modalities.

Image analysis
Functional data were analyzed using Statistical Parametric Mapping (SPM8, Wellcome Department of Cognitive Neurology, London, UK) operating in MATLAB. Within each functional run, image volumes were realigned to correct for head motion, normalized into Montreal Neurological Institute space (resampled at 3 × 3 × 3 mm) using the SPM segmentation routine, and smoothed with an 8 mm Gaussian kernel, full width at half-maximum.

For both tasks, we defined a general linear model for each participant separately. For the MNS localizer, the model included two regressors of interest: Observe and Execute. Trials were modeled as a variable epoch (Grinband et al., 2008) spanning stimulus onset to offset (Observe) or final button press (Execute) and convolved with the canonical (double-gamma) hemodynamic response function (HRF). We included the six motion parameters as covariates of no interest. For the action understanding task, the model included four regressors of interest: WhyVideo (WV), WhyText (WT), HowVideo (HV), and HowText (HT). Trials were modeled as a variable epoch spanning stimulus onset to button press and convolved with the canonical HRF. Additional covariates of no interest included regressors modeling skipped trials (defined as the absence of a button press) and the six motion parameters. For both tasks, the time series was high-pass filtered using a cutoff period of 128 s, and serial autocorrelations were modeled as an AR(1) process.

Following estimation, we first sought to define an MNS mask based on the MNS localizer. To investigate group-level effects for the MNS localizer, we entered contrast images of the effects of each regressor of interest for each participant into a random-effects analysis using a flexible factorial repeated-measures ANOVA (within-subject factor: task; blocking factor: subject). Within this model, we tested the conjunction null (Nichols et al., 2005) of Observe and Execute. The resulting SPM was conservatively thresholded using a voxel-level familywise error (FWE) rate of 0.05 combined with a cluster extent of 20 voxels. This revealed bilateral activity in regions associated with the MNS, namely, areas in the premotor cortex including posterior inferior frontal gyrus (pIFG), ventral premotor cortex (vPMC) and dorsal premotor cortex (dPMC), supplementary motor area (SMA), and clusters spanning the rostral inferior parietal lobule (rIPL) and anterior intraparietal sulcus (aIPS) (Fig. 1D and Table 2). This map will henceforth be referred to as the MNS mask.

To investigate group-level effects for the action understanding task, we entered participants' contrast images for the effects of each regressor of interest into a random-effects analysis using a flexible factorial repeated-measures ANOVA (within-subject factors: goal, modality; blocking factor: subject). Within this model, we tested the following four effects against the conjunction null (Nichols et al., 2005): (1) goal-independent effect of video versus text (WV > WT & HV > HT); (2) goal-independent effect of Text versus Video (WT > WV & HT > HV); (3)

modality-independent effect of How versus Why (HV > WV & HT > WT); and (4) modality-independent effect of Why versus How (WV > HV & WT > HT). The resulting SPMs were interrogated using a two-pass procedure. In the first pass, we examined the whole brain using a cluster-level FWE rate of 0.05, with clusters of activation defined using a voxel-level p value of 0.0001 (uncorrected). In the second-pass, we restricted our examination to the MNS mask, using a cluster-level FWE rate of 0.05 with a cluster-defining threshold of $p < 0.001$ (uncorrected).

The conjunction analyses reported above test for regions in which the presence of the effect of one factor (e.g., modality) does not depend on the level of the other factor (e.g., goal). However, they do not test whether the magnitude of the effect of one factor depends on the level of the other factor; hence, it remains possible that regions identified in the conjunction analyses will show a modality-by-goal interaction. Therefore, we interrogated an F contrast coding the modality-by-goal interaction within a mask of all regions that demonstrated conjunction effects by using a cluster-level FWE rate of 0.05 with a cluster-defining threshold of $p < 0.001$ (uncorrected).

We used psychophysiological interactions (PPIs) (Friston et al., 1997) to test the hypothesis that the functional relationship among supramodal areas for understanding motive and modality-specific areas for action perception would depend on stimulus modality. PPI enables determination of brain regions whose activity shows a change in correlation with a seed region (the "physiological" component of the PPI) as a function of a change in the participants' psychological state (the "psychological" component of the PPI). The analysis was performed using the SPM generalized PPI toolbox (http://www.martinos.org/~mclaren/ftp/Utilities_DGM). As seed, we used the area of dorsomedial prefrontal cortex (dmPFC) that showed the strongest supramodal association with understanding motive from actions (see Fig. 3). The dmPFC seeds were defined using a two-step procedure. First, we defined a binary mask of dmPFC (see Fig. 4A) by thresholding the group-level, supramodal effect of understanding motive [WV > HV & WT > HT] at $p < 0.00001$ (a more conservative threshold was used to restrict the mask to the dorsal mPFC; the resulting area spanned 276 voxels). Then, for each participant we used an automated algorithm to define regions with the dmPFC mask of at least 20 voxel extent that demonstrated the supramodal effect of understanding motive at $p < 0.05$ (uncorrected). This yielded 16 participants with valid seed regions.

We set up one PPI model for each participant, which included four PPI regressors, one for the effect of each condition. These regressors were created in the following way: (1) for each participant, we first defined the time series of their seed region as the first eigenvariate (adjusted for effects of interest); (2) the time series was deconvolved to estimate the underlying neural activity using the deconvolution algorithm in SPM8 (Gitelman et al., 2003); (3) the deconvolved time series was multiplied by the predicted time series (pre-convolved) of each condition, resulting in one "neural" PPI for each condition; and (3) each neural PPI was then convolved with the canonical HRF, yielding the four PPI regressors. As covariates of no interest, these models also included the time series of the seed region, the time series of each condition convolved with the canonical HRF, and the six motion parameters.

To investigate group-level PPI effects, we entered contrast images of the PPI effects for each participant into a random-effects analysis using a flexible factorial repeated-measures ANOVA (within-subject factors: goal, modality; blocking factor: subject). In interrogating this model, we focused on comparisons involving the two conditions demanding the inference of motive: WV > Fixation Baseline, WT > Fixation Baseline, and WV > WT. We restricted our examination to a mask consisting of the three MNS regions observed to be associated with action perception (see Fig. 2) by using a cluster-level FWE rate of 0.05 with a cluster-defining threshold of $p < 0.001$.

For all analyses, regions of activation were labeled based on a combination of visual comparison to functional regions identified in existing meta-analyses (Van Overwalle and Baetens, 2009; Caspers et al., 2010) and by reference to probabilistic cytoarchitechtonic maps of the human brain using the SPM anatomy toolbox (Eickhoff et al., 2005). For visual presentation, thresholded t statistic maps were either: (1) surface rendered using the Surfrend toolbox version 1.0.2

(http://spmsurfrend.sourceforge.net); or (2) overlaid on the average of the participants' T1-weighted anatomical images. Percent signal change for regions of interest (ROIs) was calculated using the Mars-Bar software (http://marsbar.sourceforge.net). ROIs from clusters that included multiple subregions were defined by growing 4 mm spheres around local peaks.

The rfxplot toolbox (Gläscher, 2009) was used to compute peristimulus time histograms (PSTHs) of the event-related response to WV and WT trials in the subsample included in the PPI analysis (see Fig. 4C). This was performed for the single subject ROIs of dmPFC used as seeds, and the group-level area of right pIFG/vPMC that was found to be functionally coupled to dmPFC. The extracted PSTHs spanned the peristimulus period −2 to 12 s, and data were split in 2 s time bins corresponding to the TR. To investigate whether the peak response occurred significantly later in dmPFC during WV compared to WT trials, as well as in dmPFC compared to right pIFG/vPMC during WV, we defined the time of peak in each region on a subject-by-subject basis as the bin containing the maximum value in the peristimulus period 2 to 10 s. Paired sample t tests were then used to compare the identified time to peaks.

Results
Behavioral results
A repeated-measures ANOVA examining the effects of goal and modality on response time revealed a nonsignificant main effect of goal, $F_{(1,20)} = 0.650$, $p = 0.430$, and marginally significant effects of both modality, $F_{(1,20)} = 4.278$, $p = 0.052$, and the interaction of goal and modality, $F_{(1,20)} = 4.329$, $p = 0.051$. Examination of the simple effects revealed that these marginal effects were driven by increased RT to HowText trials (M = 3.57 s, SD = 0.28) compared to the three other trial types (HowVideo: M = 3.34 s, SD = 0.28; WhyText: M = 3.44 s, SD = 0.24; WhyVideo: M = 3.36 s, SD = 0.26). As described above, we accounted for RT differences using a variable-epoch model of the neural response to each trial (Grinband et al., 2008).

Table 1 displays the five most commonly used verbs for each of the two goals along with examples of their use by participants. When comparing frequency of use of each verb across the modalities, no significant differences were observed, all t values <1.41, p values >0.15. These data confirm the face validity of the goal manipulation and suggest that the content of responses across the two modalities likely showed minimal differences.

Modality-specific effects
We first sought to determine modality-specific effects on the BOLD response during action understanding, that is, effects of presentation modality that were independent of the observer's comprehension goal. As displayed in Figure 2 and listed in Table 3, observing videos compared to text descriptions of actions was associated with a widespread network, including occipitotemporal and occipitoparietal areas known to be associated with the perception of human form and motion (Allison et al., 2000; Grèzes and Decety, 2001; Peelen and Downing, 2007); the amygdala bilaterally, which is critically involved in social perception (Adolphs, 2009); and in three areas believed to be part of the human MNS (Keysers et al., 2011; Van Overwalle and Baetens, 2009): bilateral anterior intraparietal sulcus (aIPS) and an area of right frontal cortex spanning the posterior inferior gyrus (pIFG) and ventral premotor cortex (vPMC). When performing the analysis within the MNS mask, we also observed the areas in bilateral aIPS and right pIFG/vPMC. These results demonstrate that bilateral parietal and right frontal MNS areas are associated with action perception regardless of the observer's explicit comprehension goal.
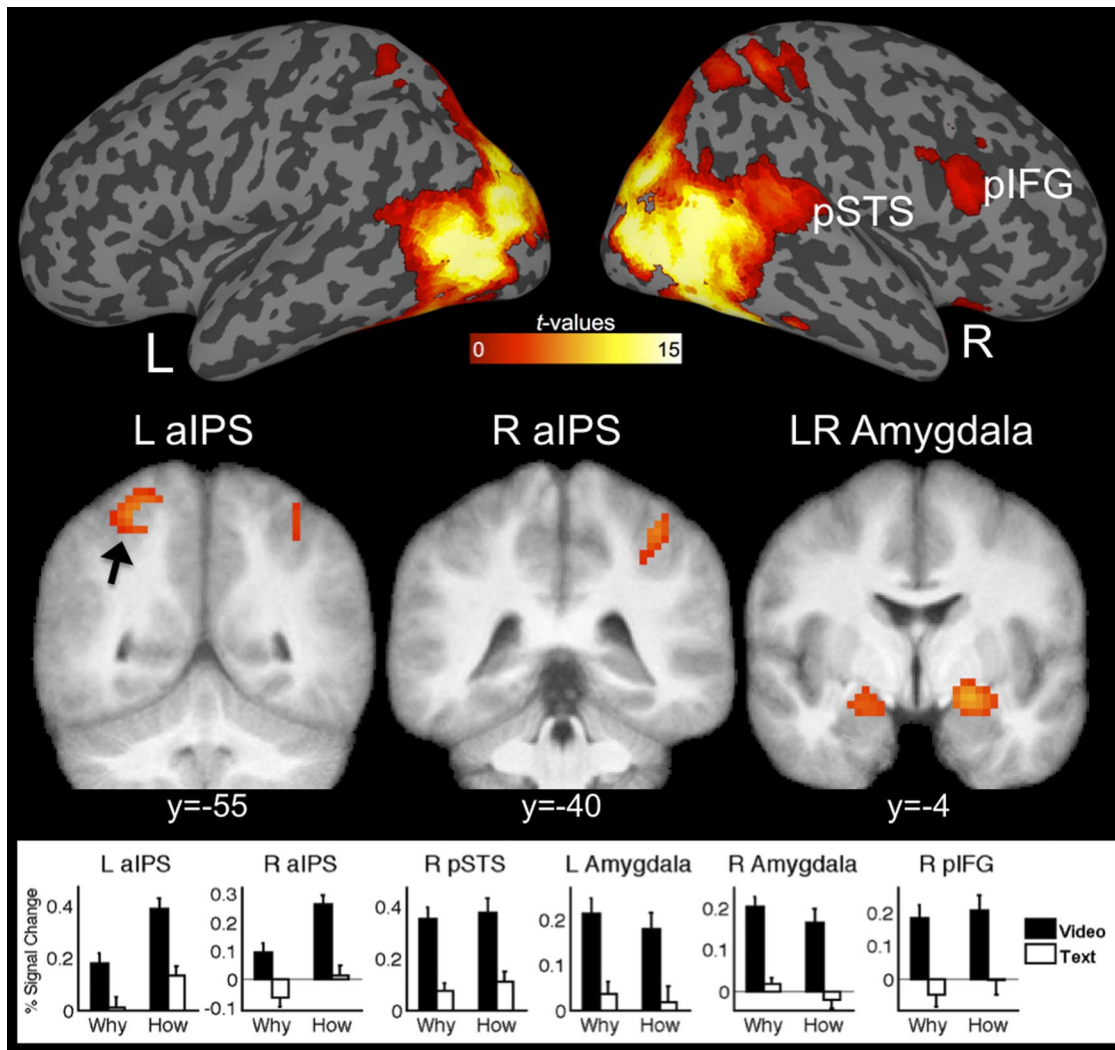
Figure 2. Modality-specific (goal-independent) effects of action perception (cluster-level corrected with a familywise error rate of 0.05). Results are from testing the conjunction null of WhyVideo > WhyText and HowVideo > HowText. The bottom row depicts percent signal change from fixation baseline for regions of interest as a function of modality and goal. L, Left. R, right; pIFG, posterior inferior frontal gyrus; aIPS, anterior intraparietal sulcus; pSTS, posterior superior temporal sulcus.

When examining the modality-specific effect of encoding text descriptions compared to videos of actions, no regions were observed, even at a more liberal voxelwise threshold of $p < 0.001$.

Supramodal effects

Next, we sought to determine brain areas that show a supramodal association with understanding the implementation compared to the motive of actions. As displayed in Figure 3A and listed in Table 4, we observed robust and exclusively left hemispheric activation in the superior parietal lobule, posterior middle temporal gyrus, and multiple areas believed to be part of the human MNS, namely pIFG/vPMC, aIPS, rostral inferior parietal lobule (rIPL), and dorsal premotor cortex (dPMC). The MNS regions were also observed when restricting the analysis to the MNS mask. This demonstrates that left-lateralized areas of the MNS form a supramodal system for explicitly representing the implementation of actions (Frey, 2008).

We then determined brain areas showing a supramodal association with understanding the motive compared to the implementation of actions. As displayed in Figure 3B and listed in Table 4, this revealed robust and primarily left-lateralized activation of core areas of the MZS as indicated by reviews and meta-

analyses (Frith and Frith, 1999; Van Overwalle and Baetens, 2009; Mar, 2011), namely dorsomedial prefrontal cortex (dmPFC), ventromedial prefrontal cortex (vmPFC), posterior cingulate cortex (PCC) extending into precuneus (PC), left temporoparietal junction (TPJ), and bilateral anterior temporal cortex (aTC). This demonstrates that largely left-lateralized areas of the MZS form a supramodal system for understanding motive from actions.

Modality-by-goal interaction effects

Next, we tested the modality-by-goal interaction within the group of regions that demonstrated either modality-specific or supramodal effects. This analysis revealed no significant interaction effects in these regions. Even in light of this null result, it is important to emphasize that the conjunction analyses reported above demonstrate only that the presence—and not the magnitude—of the effect of one factor (e.g., modality) is not dependent on the level of the other factor (e.g., goal). For instance, the effect in right pIFG/vPMC is goal independent only insofar as the presence of a significant effect in the contrast Video > Text did not depend on whether the observer possessed the goal to understand implementation or motive.

Table 3. All regions observed in the group conjunction showing modality-specific brain regions (WhyVideo > Why Text and HowVideo > HowText)

| Anatomical region | k | L/R | MNI | | | t | p_{FWE} |
			x	y	z		
Understanding actions in videos							
Whole-brain analysis							
Middle occipital gyrus	7965	L	−42	−76	4	20.04	<0.001
Lingual gyrus (BA17)		R	6	−82	−2	19.20	<0.001
Middle temporal gyrus		R	51	−67	13	19.01	<0.001
Fusiform gyrus		R	39	−52	−14	18.92	<0.001
Cuneus (BA18)		R	6	−94	16	18.20	<0.001
Inferior temporal gyrus		R	42	−70	−5	18.08	<0.001
Fusiform gyrus		L	−33	−58	−14	16.18	<0.001
Middle occipital gyrus		R	33	−82	19	15.80	<0.001
Thalamus		L	−18	−31	1	11.76	<0.001
Thalamus		R	18	−31	−2	11.71	<0.001
Dorsal precuneus		R	12	−79	46	9.08	<0.001
Superior occipital gyrus		L	−12	−85	43	8.90	<0.001
Superior temporal gyrus		R	63	−46	19	7.75	<0.001
Superior parietal lobule/IPS		R	30	−49	58	7.71	<0.001
Dorsal precuneus		R	6	−58	67	6.38	0.001
Calcarine sulcus/retrosplenial cortex		R	21	−55	16	5.42	0.015
Superior parietal lobule/IPS		L	−27	−55	58	5.00	0.053
Cerebellum	34	L	−12	−76	−41	8.04	<0.001
Amygdala	122	R	21	−4	−14	7.55	<0.001
Temporal pole		R	36	17	−26	5.36	0.017
Posterior inferior frontal gyrus	163	R	45	11	28	5.69	0.006
Amygdala	29	L	−18	−1	−17	5.11	0.039
Mirror neuron system mask							
Posterior inferior frontal gyrus	123	R	48	11	31	6.59	<0.001
Anterior intraparietal sulcus	100	R	36	−46	55	7.09	<0.001
		R	42	−28	40	4.88	0.003
	99	L	−30	−55	58	6.01	<0.001
Understanding actions in text							
No suprathreshold voxels							

All results are cluster-level corrected using a familywise error (FWE) rate of 0.05. Coordinates are all local maxima observed that were separated by at least 20 mm. L/R, Left and right hemispheres; x, y, and z, Montreal Neurological Institute (MNI) coordinates in the left–right, anterior–posterior, and inferior–superior dimensions, respectively; t, t statistic value at those coordinates; k, cluster voxel extent; p_{FWE}, voxel-level FWE for the local peak.

Dissociating left and right pIFG/vPMC
The results thus far suggest a clear dissociation of the function of similar areas of the pIFG/vPMC in the left and right hemispheres. As the bar graph in Figure 2 demonstrates, right pIFG/vPMC was selectively associated with the perception of actions in videos, regardless of the observer's explicit comprehension goal. Conversely, the bar graphs in Figure 3 demonstrate that left pIFG/vPMC was selectively associated with the explicit representation of action implementation, regardless of the presentation modality. To garner further evidence for this dissociation, we directly contrasted WhyVideo and HowText trials within the bilateral frontal areas of the MNS mask. Consistent with the proposed dissociation, right pIFG/vPMC showed a robust response in the contrast WhyVideo > HowText (peak: $x = 45$, $y = 17$, $z = 22$; voxel extent $= 40$; $t = 4.69$ cluster-level $p_{FWE} = 0.005$), while left pIFG/vPMC showed a robust response in the opposite contrast (peak: $x = −48$, $y = 5$, $z = 25$; voxel extent $= 20$; $t = 3.78$ cluster-level $p_{FWE} = 0.013$).

Modality-specific connectivity among the MNS and MZS during the inference of motive
Finally, we investigated the proposition that, in some contexts but not others, the MNS and MZS may collaborate to enable action understanding. Based on attribution theories from social psychology (Gilbert, 1998), we have elsewhere proposed an Identification–Attribution (I–A) model of MNS and MZS contributions to causal attribution during social perception wherein the

MNS contributes to the (perceptual) identification of motor behaviors that are subsequently attributed to social causes, such as motives and beliefs, in the MZS (Spunt and Lieberman, 2012). This predicts that the MNS and MZS collaborate when individuals infer motive only when actions must be decoded from sensory information about body motion. Thus, we conducted a PPI analysis (Friston et al., 1997) to test the hypothesis that functional connectivity among the MZS and MNS would be greater when motive is inferred from videos compared to text descriptions of actions. As a seed, we used the region observed to be most reliably associated with understanding motive, the dmPFC. We tested the hypothesis that this area would be functionally coupled with the bilateral parietal and right frontal MNS areas found to be selective for action perception (Fig. 2). As displayed in Figure 4A, dmPFC demonstrated significantly increased functional coupling with a cluster in right pIFG/vPMC (peak: $x = 48$, $y = 14$, $z = 31$; voxel extent $= 12$; $t = 3.77$ cluster-level $p_{FWE} = 0.016$) when motive was inferred from videos compared to text descriptions of actions. This demonstrates that the context of action understanding, in addition to modulating activation of the MZS and MNS, also modulates connectivity between the two systems.

Time course analyses
The I–A model described above not only predicts modality-specific connectivity among the MNS and MZS, but also predicts an ordered sequence of mental operations, with MZS-mediated attribution occurring only after motor behaviors are encoded in the MNS. In the present study, this generates two predictions. First, the time to peak response in the MZS should occur later when motive is inferred from videos compared to text descriptions of actions. In addition, when motive is inferred from videos, the time to peak response in the MZS should be later than the same response in the MNS. We indeed observed that in the dmPFC, time to peak was significantly later during WhyVideo (M = 7.63 s, SD = 1.82) than during WhyText (M = 6.00 s, SD = 1.79), $t_{(15)} = 2.93$, $p = .010$ and compared to the time to peak in right pIFG/vPMC during WhyVideo (M = 4.38 s, SD = 1.50), $t_{(15)} = 6.34$, $p < 0.001$ (Fig. 4C). These time course analyses are consistent with the ordered sequence of operations predicted by the I–A model.

Discussion
The present results help resolve the ongoing debate over the neural bases of action understanding. When varying task demands are systematically considered, as in the present study, the roles the MNS and MZS have in action understanding emerges with more clarity. We observed that bilateral parietal and right frontal MNS showed a modality-specific association with perceiving actions ("what is being done"), left hemisphere MNS showed a supramodal association with understanding action implementation ("how it is being done"), and the MZS showed a supramodal association with understanding motive from actions ("why it is being done"). These results support the tripartite model proposed by Thioux et al. (2008) wherein the MNS supports understanding actions at low (how) and intermediate (what) levels of abstraction, whereas the MZS supports understanding actions at high (why) levels of abstraction (Vallacher and Wegner, 1987). Their claims were based on studies that manipulated either the observer's goal or the content of a perceived action (Grèzes et al., 2004; Brass et al., 2007; de Lange et al., 2008). However, these studies could only provide indirect support, as none independently manipulated the level of understanding (goal) and the demand for
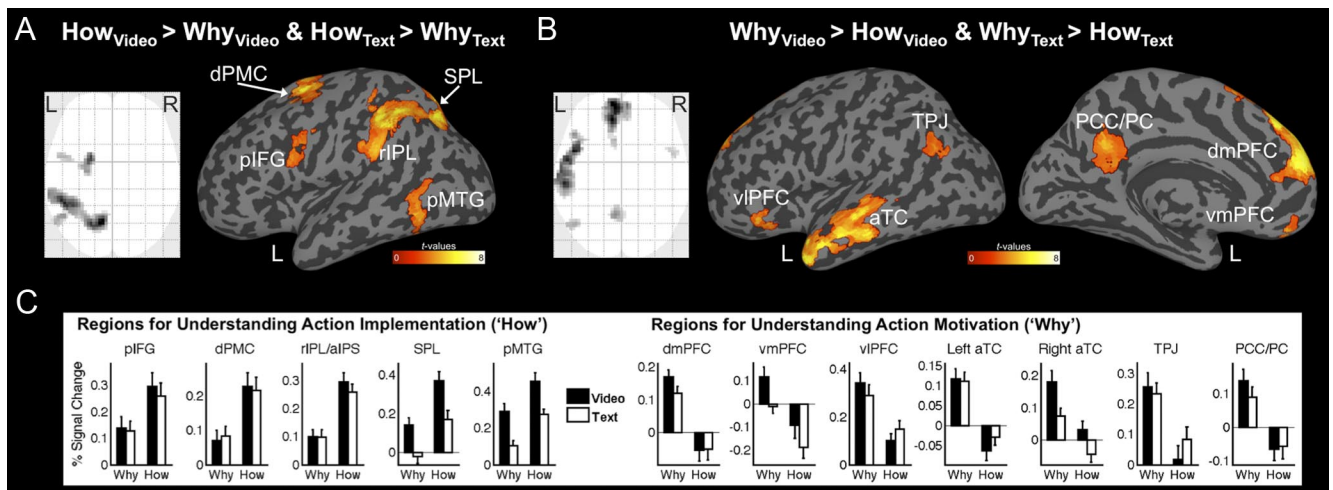
Figure 3. Supramodal effects of goal (cluster-level corrected with a familywise error rate of 0.05). *A*, Glass brain and surface rendering from testing the conjunction null for HowVideo > WhyVideo and HowText > WhyText. *B*, Glass brain and surface rendering testing the conjunction null for WhyVideo > HowVideo and WhyText > HowText. *C*, Percent signal change for regions of interest from *A* and *B*. L, Left; R, right; dPMC, dorsal premotor cortex; pIFG, posterior inferior frontal gyrus; rIPL, rostral inferior parietal lobule; aIPS, anterior intraparietal sulcus; SPL, superior parietal lobule; pMTG, posterior middle temporal gyrus; vlPFC, ventrolateral prefrontal cortex; TPJ, temporoparietal junction; aTC, anterior temporal cortex; PCC, posterior cingulate cortex; PC, precuneus; dmPFC, dorsomedial prefrontal cortex; vmPFC, ventromedial prefrontal cortex.

Table 4. All regions observed in the group conjunction analyses determining supramodal brain regions for understanding implementation (HowVideo > WhyVideo and HowText > WhyText) and motive (WhyVideo > HowVideo and WhyText > HowText)

Anatomical Region	k	L/R	MNI x	y	z	t	p_{FWE}
Understanding Implementation ("How")							
Whole-brain analysis							
Superior parietal lobule/dorsal precuneus	725	L	−12	−64	58	7.30	<0.001
Rostral inferior parietal lobule/aIPS		L	−48	−43	49	6.91	0.001
Supramarginal gyrus		L	−63	−34	34	5.54	0.022
Dorsal premotor cortex	148	L	−21	5	58	6.52	0.001
Posterior middle/inferior temporal gyrus	62	L	−51	−55	−5	5.28	0.046
Posterior inferior frontal gyrus	49	L	−51	5	28	4.74	0.192
Mirror neuron system mask							
Dorsal premotor cortex	93	L	−21	2	58	5.93	<0.001
Anterior intraparietal sulcus	163	L	−39	−43	40	5.89	<0.001
Posterior inferior frontal gyrus	41	L	−48	5	25	4.23	0.060
Understanding motive ("Why")							
Whole-brain analysis							
Dorsomedial prefrontal cortex	476	L	−6	59	22	6.92	<0.001
		L	−9	44	46	6.86	<0.001
		L	−6	23	64	4.82	0.091
Anterior temporal cortex	416	L	−48	11	−35	6.76	<0.001
		L	−57	−22	−8	6.52	<0.001
Ventrolateral prefrontal cortex		L	−48	29	−11	5.19	0.030
Precuneus/posterior cingulate cortex	118	L	−6	−55	34	5.27	0.023
Ventromedial prefrontal cortex	51	L	−6	50	−17	5.11	0.038
Temporal parietal junction/pSTS	38	L	−45	−64	34	4.86	0.082
Anterior temporal cortex	21	R	51	8	−29	4.45	0.249
Mirror neuron system mask							
No suprathreshold voxels							

All results are cluster-level corrected using a familywise error (FWE) rate of 0.05. Coordinates are all local maxima observed that were separated by at least 20 mm. L/R, Left and right hemispheres; x, y, and z, Montreal Neurological Institute (MNI) coordinates in the left–right, anterior–posterior, and inferior–superior dimensions, respectively; t, t statistic value at those coordinates; k, cluster voxel extent; p_{FWE}, voxel-level FWE for the local peak.

action perception (modality) while holding action content constant across conditions. By doing so in the present study, we have provided unambiguous support for a tripartite model of the brain systems supporting action understanding.

Right pIFG/vPMC, in the MNS, was selectively associated with the perception of action regardless of the observer's compre-

hension goal. Additionally, this region was functionally coupled with dmPFC, a core area of the MZS, when participants were prompted to understand the motives driving actions presented in videos but not in text. We recently found that these two areas are functionally coupled when individuals are prompted to infer the cause of observed emotional facial expressions (Spunt and Lieberman, 2012). This functional coupling is consistent with social psychological theorizing stating that before making attributions about the causes of behavior, sensory input about body movements must be translated into attribution-relevant events such as goals, intentions, and emotions (Gilbert, 1998). The right pIFG/vPMC is a strong candidate for being critically involved in this translation function, as previous neuroimaging studies of action perception demonstrate that this region is associated with spontaneous event segmentation (Zacks et al., 2001) and with the encoding of features of the motor context that indicate what the actor is intending to do (Iacoboni et al., 2005; Hamilton and Grafton, 2008). Collectively, this suggests an Identification–Attribution model of MNS and MZS system function during social perception, wherein the MNS translates sensory input about motor behavior into a format that is relevant to attributional processes carried out in the MZS. In line with this ordered sequence of operations, we found that for trials in which participants inferred motive, dmPFC exhibited a significantly delayed response to videos of actions, both compared to its own response to text descriptions of actions and to the response of right pIFG/vPMC to videos of actions. Although the results of these time course analyses are consistent with the Identification–Attribution model, it should be noted that RT to video trials was not longer than RT to text trials, as might be predicted by a two-stage model of inferring motive during action perception. However, the lack of an RT difference does not necessarily rule out this two-stage model. It is possible that the kind of information gathered from the visual analysis of a social scene, including (but not limited to) the decoding of motor intention putatively based in the MNS, facilitates more efficient inferential processing in the MZS. The videos were all filmed in contextually rich natural scenes, and it is plausible that this additional contextual information—not captured by the text descriptions—provides constraints that make the inference of

motive easier. This would explain a lack of difference in total processing speed across the two conditions despite the additional processing stage for videos. Future research is needed to clarify the temporal dynamics of MNS and MZS function during action perception.

In contrast to the right pIFG/vPMC, left pIFG/vPMC was selectively sensitive to the explicit goal to understand the implementation of actions. In fact, we found that left pIFG/vPMC was more active when understanding implementation from text descriptions of actions than when understanding motive from videos of actions. This left-lateralized effect cannot be explained by the fact that this comparison is of verbal to nonverbal stimuli, because as is clear from the bar graphs in Figure 3C, left pIFG/vPMC shows an enhanced response to understanding implementation from videos of actions than to understanding motive from text descriptions of actions. This demonstrates that left pIFG/vPMC is not selective for visual input regarding actions, and that the reliable association of left pIFG/vPMC with action observation in prior work may be due to either spontaneous (e.g., during passive observation) or goal-related (e.g., during active imitation) cognition about action implementation. Indeed, a specific role for this region in the explicit identification of motor events is consistent with neuroimaging studies showing that linguistic material describing actions preferentially activates left pIFG compared to linguistic material not describing actions (Frey, 2008).

The goal to understand motive from action robustly activated the MZS in a modality-independent fashion. Although previous studies have observed an MZS association with explaining actions in terms of mental states (Grèzes et al., 2004; Brass et al., 2007; de Lange et al., 2008; Spunt et al., 2010, 2011), the present study provides the first empirical demonstration that the MZS is a supramodal system for inferring motive from actions, whether seen or read. This demonstration is important given that the majority of studies mapping the MZS have relied on verbal materials (Van Overwalle and Baetens, 2009). With this said, the observation of supramodal function is implied by studies showing that the MZS comes online during mental state inference for a wide variety of different experimental paradigms (Gallagher et al., 2000; Gobbini et al., 2007; Carrington and Bailey, 2009). In addition, our findings add to recent evidence that two areas of the MZS, the mPFC and left pSTS, encode the identity of emotions in a supramodal fashion (Peelen et al., 2010). Hence, the MZS may support an abstract, modality-independent representation of multiple categories of mental state (e.g., emotion vs motive) on the basis of multiple categories of behavior (e.g., emotional expression vs goal-directed action).

Importantly, although we demonstrated that the MZS plays a supramodal role in mental state inference, its functional connectivity with the MNS was modality specific. This suggests that the MNS may indeed play a role in mental state inference, but only on two conditions: (1) the observer possesses the goal to infer mental states on the basis of the target's actions; and (2) the actions must be decoded from sensory information, rather than conveyed verbally. Generally, this underscores the conditional, or context-
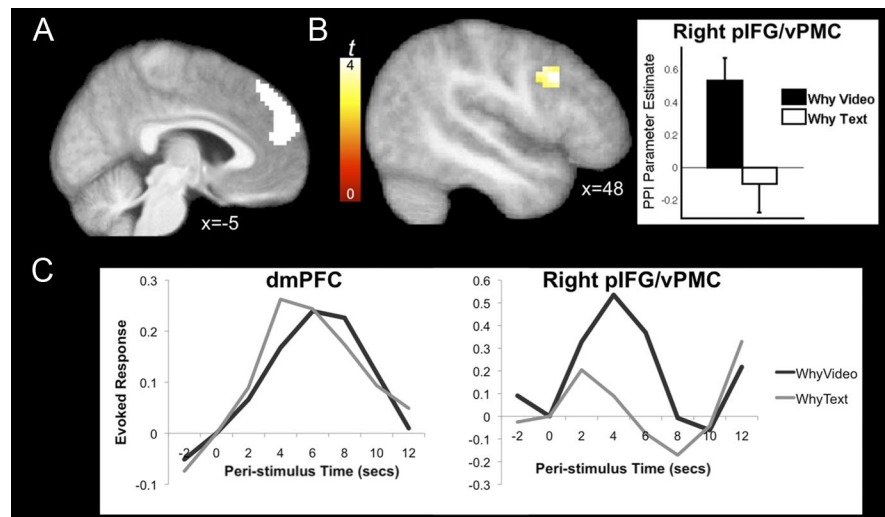

Figure 4. Psychophysiological interaction (PPI) effects. **A**, The group-level dorsomedial prefrontal cortex (dmPFC) seed region used to identify single-subject seeds. **B**, Right posterior inferior frontal gyrus/ventral premotor cortex (pIFG/vPMC) showing increased connectivity with dmPFC during WhyVideo compared to WhyText. The plot depicts the mean parameter estimate of context-dependent connectivity among dmPFC and right pIFG/vPMC. Error bars are SEM. **C**, Peristimulus time course histograms of the response of right pIFG/vPMC and dmPFC to WhyVideo and WhyText trials.

dependent, function of the neural systems that support social cognition, where the MNS, the MZS, and MNS/MZS connectivity may each make independent contributions to the human ability to represent the mental states associated with another's behavior.

We observed an interesting dissociation between posterior and anterior regions of the IFG in their contribution to action representation. Namely, we observed that posterior regions of IFG were associated with action perception (Fig. 2) and understanding action implementation (Fig. 3A), whereas an anterior region of IFG in the ventrolateral PFC was associated with understanding action motive (Fig. 3B). This anterior–posterior distinction in IFG mirrors the rostrocaudal axis believed to exist in the function of IFG (Badre and D'Esposito, 2009; Kilner, 2011), wherein anterior regions encode abstract representations of actions ("why") while more posterior regions encode more concrete representations ("what" and "how").

Finally, these findings converge with two recent studies that used transcranial magnetic stimulation to show that, for both photographs of objects (Pobric et al., 2010) and object-words (Ishibashi et al., 2011), a region of the left IPL participates in the representation of the manipulability of the object—that is, how it is used—while a region of the left aTC contains more abstract, modality-nonspecific information about object function—that is, why it is used (see also Canessa et al., 2008). These regions correspond well with the left IPL and left aTC regions observed to be associated with representing action implementation ("how") and motive ("why"), respectively, in the present study.

In conclusion, it is possible that, in our evolutionary past, action understanding encompassed a single mental process, perhaps corresponding to the decoding of motor intention from observed behavior (Frith and Frith, 1999). However, the present study demonstrates that attempts to isolate the single brain system that supports action understanding in humans is likely doomed to failure. This is because action understanding can be used to refer to numerous mental processes that are deployed to meet the demands of a complex social world, where social interactions are not always mediated by the perception of moving bodies and where the aims of the observer can and do vary. Mov-

ing forward, research and theory on the neural bases of action understanding will benefit by embracing this variability.

References

Adolphs R (2009) The social brain: neural basis of social knowledge. Annu Rev Psychol 60:693–716.

Allison T, Puce A, McCarthy G (2000) Social perception from visual cues: role of the STS region. Trends Cogn Sci 4:267–278.

Aziz-Zadeh L, Wilson SM, Rizzolatti G, Iacoboni M (2006) Congruent embodied representations for visually presented actions and linguistic phrases describing actions. Curr Biol 16:1818–1823.

Badre D, D'Esposito M (2009) Is the rostro-caudal axis of the frontal lobe hierarchical? Nat Rev Neurosci 10:659–669.

Brainard DH (1997) The psychophysics toolbox. Spat Vis 10:433–436.

Brass M, Schmitt RM, Spengler S, Gergely G (2007) Investigating action understanding: inferential processes versus action simulation. Curr Biol 17:2117–2121.

Canessa N, Borgo F, Cappa SF, Perani D, Falini A, Buccino G, Tettamanti M, Shallice T (2008) The different neural correlates of action and functional knowledge in semantic memory: an FMRI study. Cereb Cortex 18:740–751.

Carrington SJ, Bailey AJ (2009) Are there theory of mind regions in the brain? A review of the neuroimaging literature. Hum Brain Mapp 30:2313–2335.

Caspers S, Zilles K, Laird AR, Eickhoff SB (2010) ALE meta-analysis of action observation and imitation in the human brain. Neuroimage 50:1148–1167.

Decety J, Grèzes J, Costes N, Perani D, Jeannerod M, Procyk E, Grassi F, Fazio F (1997) Brain activity during observation of actions. Influence of action content and subject's strategy. Brain 120:1763–1777.

de Lange FP, Spronk M, Willems RM, Toni I, Bekkering H (2008) Complementary systems for understanding action intentions. Curr Biol 18:454–457.

Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005) A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage 25:1325–1335.

Frey SH (2008) Tool use, communicative gesture and cerebral asymmetries in the modern human brain. Philos Trans R Soc Lond B Biol Sci 363:1951–1957.

Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6:218–229.

Frith CD, Frith U (1999) Interacting minds–a biological basis. Science 286:1692–1695.

Gallagher HL, Happé F, Brunswick N, Fletcher PC, Frith U, Frith CD (2000) Reading the mind in cartoons and stories: an fMRI study of "theory of mind" in verbal and nonverbal tasks. Neuropsychologia 38:11–21.

Gallese V, Gernsbacher MA, Heyes C, Hickok G, Iacoboni M (2011) Mirror neuron forum. Perspect Psychol Sci 6:369–407.

Gilbert DT (1998) Ordinary personology. In: The handbook of social psychology, Vol 2, Ed 4 (Gilbert DT, Fiske T, Lindzey G, eds), pp. 89–150. New York: McGraw Hill.

Gitelman DR, Penny WD, Ashburner J, Friston KJ (2003) Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. Neuroimage 19:200–207.

Gläscher J (2009) Visualization of group inference data in functional neuroimaging. Neuroinformatics 7:73–82.

Gobbini MI, Koralek AC, Bryan RE, Montgomery KJ, Haxby JV (2007) Two takes on the social brain: a comparison of theory of mind tasks. J Cogn Neurosci 19:1803–1814.

Grèzes J, Decety J (2001) Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. Hum Brain Mapp 12:1–19.

Grèzes J, Frith C, Passingham RE (2004) Brain mechanisms for inferring deceit in the actions of others. J Neurosci 24:5500–5505.

Grinband J, Wager TD, Lindquist M, Ferrera VP, Hirsch J (2008) Detection of time-varying signals in event-related fMRI designs. Neuroimage 43:509–520.

Hamilton AF, Grafton ST (2008) Action outcomes are represented in human inferior frontoparietal cortex. Cereb Cortex 18:1160–1168.

Iacoboni M, Molnar-Szakacs I, Gallese V, Buccino G, Mazziotta JC, Rizzolatti G (2005) Grasping the intentions of others with one's own mirror neuron system. PLoS Biol 3:e79.

Ishibashi R, Lambon Ralph MA, Saito S, Pobric G (2011) Different roles for lateral anterior temporal lobe and inferior parietal lobule in coding function and implementation tool knowledge: evidence from an rTMS study. Neuropsychologia 49:1128–1135.

Keysers C, Thioux M, Gazzola V (2011) The mirror neuron system and social cognition. In: The Oxford handbook of social neuroscience, Vol 1, (Decety J, Cacioppo JT, eds), pp 525–550. New York: Oxford UP.

Kilner JM (2011) More than one pathway to action understanding. Trends Cogn Sci 15:352–357.

Mar RA (2011) The neural bases of social cognition and story comprehension. Annu Rev Psychol 62:103–134.

Nichols T, Brett M, Andersson J, Wager T, Poline JB (2005) Valid conjunction inference with the minimum statistic. Neuroimage 25:653–660.

Peelen MV, Downing PE (2007) The neural basis of visual body perception. Nat Rev Neurosci 8:636–648.

Peelen MV, Atkinson AP, Vuilleumier P (2010) Supramodal representations of perceived emotions in the human brain. J Neurosci 30:10127–10134.

Pobric G, Jefferies E, Lambon Ralph MA (2010) Category-specific versus category-general semantic impairment induced by transcranial magnetic stimulation. Curr Biol 20:964–968.

Rizzolatti G, Sinigaglia C (2010) The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. Nat Rev Neurosci 11:264–274.

Spunt RP, Lieberman MD (2012) An integrative model of the neural systems supporting the comprehension of observed emotional behavior. Neuroimage 59:3050–3059.

Spunt RP, Falk EB, Lieberman MD (2010) Dissociable neural systems support retrieval of how and why action knowledge. Psychol Sci 21:1593–1598.

Spunt RP, Satpute AB, Lieberman MD (2011) Identifying the what, why, and how of an observed action: an fMRI study of mentalizing and mechanizing during action observation. J Cogn Neurosci 23:63–74.

Thioux M, Gazzola V, Keysers C (2008) Action understanding: how, what and why. Curr Biol 18:R431–R434.

Uithol S, van Rooij I, Bekkering H, Haselager P (2011) Understanding motor resonance. Soc Neurosci 6:388–397.

Vallacher RR, Wegner DM (1987) What do people think they're doing? Action identification and human behavior. Psychol Rev 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.

Wager TD, Nichols TE (2003) Optimization of experimental design in fMRI: a general framework using a genetic algorithm. Neuroimage 18:293–309.

Zacks JM, Braver TS, Sheridan MA, Donaldson DI, Snyder AZ, Ollinger JM, Buckner RL, Raichle ME (2001) Human brain activity time-locked to perceptual event boundaries. Nat Neurosci 4:651–655.

Zaki J, Ochsner K (2009) The need for a cognitive neuroscience of naturalistic social cognition. Ann NY Acad Sci 1167:16–30.