Viewpoint (In)dependence of Action Representations: An MVPA Study

Nikolaas N. Oosterhof, Steven P. Tipper, and Paul E. Downing

Abstract

■ The discovery of mirror neurons—neurons that code specific actions both when executed and observed—in area F5 of the macaque provides a potential neural mechanism underlying action understanding. To date, neuroimaging evidence for similar coding of specific actions across the visual and motor modalities in human ventral premotor cortex (PMv)—the putative homologue of macaque F5—is limited to the case of actions observed from a first-person perspective. However, it is the third-person perspective that figures centrally in our understanding of the actions and intentions of others. To address this gap in the literature, we scanned participants with fMRI while they viewed two actions from either a first- or third-person perspective during some trials and executed the same actions during other trials. Using multivoxel pattern analysis, we found action-specific cross-modal visual–motor representations in PMv for the first-person but not for the third-person perspective. Additional analyses showed no evidence for spatial or attentional differences across the two perspective conditions. In contrast, more posterior areas in the parietal and occipitotemporal cortex did show cross-modal coding regardless of perspective. These findings point to a stronger role for these latter regions, relative to PMv, in supporting the understanding of others' actions with reference to one's own actions. ■

INTRODUCTION

The human is a social species for which the ability to observe and understand the actions and intentions of others is crucial in everyday interactions. A possible neural mechanism underlying this ability comes from the discovery of "mirror neurons" in the macaque frontal area F5 and parietal areas PF/PFG (Gallese & Goldman, 1998; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; but see Hickok, 2009; Turella, Pierno, Tubaldi, & Castiello, 2009). Mirror neurons show an increase in firing rate both when executing a specific goal-directed action and when observing the same action executed by the experimenter. It has been suggested that similar neurons in humans could explain various complex social human phenomena such as action understanding, imitation, theory of mind, language acquisition, and empathy (e.g., Rizzolatti & Craighero, 2004) and to account for social deficits in autism (Iacoboni & Dapretto, 2006).

Various studies using fMRI have localized areas in human ventral premotor (PMv) and anterior parietal cortex (PCa) that show increased activity both when actions are observed and executed, and these have been suggested to reflect the human homologue of the mirror system (e.g., Gazzola & Keysers, 2008). However, the majority of these studies did not discriminate between action-specific activations and therefore do not rule out general non-action-specific effects of increased activity (Dinstein, Thomas, Behrmann, & Heeger, 2008).

To test more directly for evidence of a human mirror system, recent fMRI studies have used either repetition suppression (RS) or multivoxel pattern analysis (MVPA) to test for the presence of neural populations in PMv and PCa that represent specific actions. When considering only studies that included both observed and executed actions, nearly all failed to find strong evidence for cross-modal coding (i.e., similar coding across the visual and motor domains) in PMv. Furthermore, importantly, the only study (Kilner, Neal, Weiskopf, Friston, & Frith, 2009) that reported full cross-modal action coding in PMv employed only the first-person perspective—where actions were seen as if performed by the participant. This does not test a key proposed feature of the "mirror" system: the ability to generalize to actions observed from third-person perspectives. Indeed, when considering the literature more widely, a significant number of studies only tested actions observed from the first-person perspective (Table 1).

To address this gap in the human neuroimaging literature, we systematically compared cross-modal action-specific representations across first- and third-person perspectives. We focus in particular on PMv as this region has been the center of much empirical and theoretical investigations of the "mirror system" (e.g., Ferrari, Rozzi, & Fogassi, 2005; Iacoboni et al., 2005; Johnson-Frey

Bangor University

Table 1. Overview of fMRI Studies Investigating Action-specific Representations

	Analysis	View	Object	Dyn	vPM			aIPS/IPL		
					Do	See	X	Do	See	X
Majdandžić, Bekkering, & Van Schie, 2009	RS	1st	One	Yes	?	+	?	?	−	?
Ortigue, Thompson, Parasuraman, & Grafton, 2009	RS	3rd	Two[a]	Yes	?	+	?	?	+	?
Kroliczak, Mcadam, Quinlan, & Culham, 2008	RS	1st	Eight	Yes	+		?	+		?
Ramsey & Hamilton, 2010	RS	3rd	Two[a]	Yes	?	+	?	?	+	?
Hamilton & Grafton, 2006	RS	1st	Two[a]	Yes	?	−	?	?	+	?
Hamilton & Grafton, 2008	RS	3rd	One[b]	Yes	?	−[b]	?	?	−[c]	?
Dinstein et al., 2007	RS	1st	No	Yes	+	+	−	+	+	−
Chong, Cunnington, Williams, Kanwisher, & Mattingley, 2008	RS	3rd	No	Yes	?	?	−	?	?	±[d]
Lingnau et al., 2009	RS	1st	No	Yes	−	−	−	−	+	±[d]
Kilner et al., 2009	**RS**	**1st**	**Two[a]**	**Yes**	**?**	**?**	**+**	**?**	**?**	**?**
Dinstein, Gardner, Jazayeri, & Heeger, 2008	MVPA	1st	No	Yes	+	−	−	+	+	−
Ogawa & Inui, 2011	MVPA	1st & 3rd[e]	Two	No	?	+	?	?	−	?
Oosterhof et al., 2010	MVPA	3rd	One	Yes	+	+	−	+	+	+

Studies that investigated action-specific representations of comparable actions using either RS or MPVA are included. Studies that considered activation differences due to manipulated visual or motor familiarity are excluded as they might be confounded by differences in attention. Cross-modal coding in PMv (column with data in **bold font**) was observed in one study (row with data in **bold font**).

1st/3rd = first/third perspective for observed actions; PMv = left PMv; + = significant effect; − = no significant effect; ? = not tested or reported; Do/See/X = action specificity for executed/observed/cross-modal representations.

[a]Two objects were used, each corresponding to a single action.

[b]Changeable object (box with lid).

[c]RS effects were found for "outcome" (box in open or closed position) but not for the kinematics of actions.

[d]Cross-modal repetition suppression effects in one direction (do-then-see, or see-then-do) but not the other direction.

[e]Results collapsed across perspectives.

et al., 2003; Rizzolatti, Fogassi, & Gallese, 2002). Using MVPA, we found action-specific visual–motor representations in this region for first-person but not for third-person perspectives. Our results show that perspective modulates cross-modal visuomotor representations in PMv and suggest that the neural populations representing actions in this region may be unable to generalize across both perspective and modality. In contrast, more posterior areas in the parietal and occipitotemporal cortex did show cross-modal coding regardless of perspective.

METHODS

Subjects

Twenty-nine right-handed, healthy adult volunteers (mean age = 26 years, range = 19–38 years; 20 women) were recruited from the Bangor University community. All participants had normal or corrected-to-normal vision. Participants satisfied all requirements in volunteer screening and gave informed consent. Procedures were approved by the ethics committee of the School of Psychology at Bangor University. Participation was compensated at £15.

Setup

Participants performed and watched object-directed actions in the scanner (Figure 1). The object was cup-shaped and attached with an elastic string to a table located partially inside the scanner bore, approximately above the navel of the participant (the same object was used in Oosterhof, Wiggett, Diedrichsen, Tipper, & Downing, 2010). To avoid trivial confounds of action observation during action execution trials, either a woolen or cardboard screen was attached to the scanner coil in a vertical position above the participant's neck to ensure that the participant was unable to see the table, the object, or their arms. Instructions were presented visually on a projector screen behind the scanner bore, which could be viewed by participants with a tilted, backward reflecting mirror placed on the scanner coil.

Action Stimuli

Participants performed and watched two object-directed actions termed "lift" and "slap" (Figure 1). Actions were watched from either a first- or third-person perspective (in a between-participants design; see below). The first-person perspective (1PP) videos recorded outside the scanner featured a male model in supine position with the table and object placed similarly to the participants in the scanner, with the camera located just behind and above the model's head. The third-person perspective videos were recorded featuring the same model and with a similar position of the table and object, but with the model standing on the opposite side of the table. For all actions, the model placed his hand in a "resting position" on the table next to the object before and after executing the action. Projected on the horizontal plane, the angles between the model's nose, object, and camera were approximately 0 and 135° for the egocentric and allocentric videos, respectively. For each perspective, four exemplar videos were recorded of each of the two actions in alternating order. This introduced small variations in the different videos, which makes identifying the action based on low-level features of the first frames more difficult because the initial hand position is less or not predictive of the action shown in subsequent frames. Furthermore, using multiple exemplars allowed for comparing the action kinematics between actions and perspectives (see below). The third-person perspective videos were scaled and slightly rotated to match the position of the table and objects in the egocentric videos, and all videos were cropped to 514×378 pixels and surrounded by a black background.

To assess how well the temporal characteristics of actions across videos were matched, five key frames were identified in all videos: (BH) the last frame before the model's hand started to move from the resting position, (BO) the last frame before the object moved, (XX) the frame in which the object was in the most extreme position (lifted: highest position, slapped, most angled position), (AO) the first frame after the object had moved, and (AH) the first frame the model put his hand back in the resting position. Frames were chosen so that frame (XX) was always presented at 1.4 sec after the start of the video, and the videos were made of equal length (2.8 sec) by adding duplicates of frames (BH) and (AH) of the hand in the resting position before and after these respective frames. To compare the temporal characteristics across videos, relative onset times between frame (XX) and each of the frames (BH), (BO), (AO), and (AH) was computed, yielding four relative onset times for each video exemplar. These onset times were then compared between the different videos (four exemplars per condition) in a 2 (view: first-person perspective, third-person perspective) $\times$ 2 (action: slap, lift) MANOVA with pooled (co)variance estimates. The MANOVA showed a significant difference between the relative onsets of the lift and slap actions (Wilk's $\Lambda = 0.1411$, Roy's $F(4, 9) = 13.69$, $p < .001$) but, more importantly, not between the first-person perspective and third-person perspective views ($\Lambda = 0.5153$, $F(4, 9) = 2.12$, $p = .16$) and no interaction ($\Lambda = 0.6053$, $F(4, 9) = 1.47$, $p = .29$). In other words, there were no apparent differences between the first-person perspective and third-person perspective videos with respect to the temporal characteristics. Subsequent post hoc univariate two-sample t tests comparing the relative onsets of the lift and slap actions (collapsed across the first-person perspective and third-person perspective videos) showed that the former started earlier (onset (BH), $t_{12} = 6.05$, $p < .001$, not corrected for multiple comparisons; onset (BO), $t_{12} = 1.54$, $p = .14$)) and took longer to complete
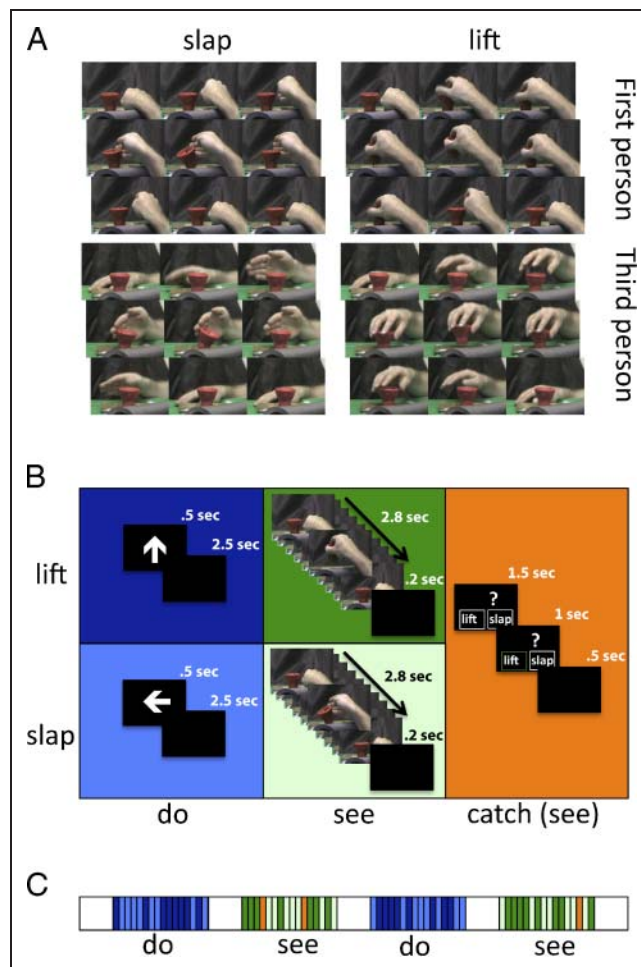
Figure 1. Paradigm and stimuli. (A) Still frames of representative videos used for two observed actions ("slap" and "lift" in the left and right columns, respectively) for first-person (top row) and third-person (bottom row) perspective conditions. For each type of video, four similar exemplar videos were used. (B) Overview of conditions in the paradigm. Participants either executed (left column) or observed (center column) the "slap" and "lift" conditions. Occasionally a "catch" trial (right column) was introduced after an observed action. (C) Overview of a single "run," consisting two "execute" and two "observe" blocks of 48 sec each, separated by null trials of 16 sec each.

(onset (AO), $t_{12} = -4.05$, $p = .0012$; onset (AO), $t_{12} = -6.67$, $p < .001$).

Experimental Design and Task

The main experiment had a 2 (action: lift, slap) × 2 (modality: do, see) within-participants design. The actions were viewed in first-person perspective in Experiment 1 and in third-person perspective in Experiment 2. We used different perspectives for different participants to avoid potential "contamination" effects across perspectives (cf. Poulton, 1982).

The "see" trials (Figure 1) were presented in blocks of 16 trials of 3 sec each. A video of either the lift or slap action (randomly selected from the four exemplars; see above) was shown for 2.8 sec, followed by a 0.2-sec black screen. To ensure attention of the participants, once or twice during a block a "catch" trial was presented instead of an action video, which consisted of a question mark displayed centrally and the words "lift" and "slap" at the left and right bottom of the screen. Catch trials were assigned in each "see" block randomly with the following constraints: (1) the number of catch trials (one or two) was random; (2) catch trials could not be assigned to the first trial of a block; (3) a catch trial could not be followed immediately by another catch trial; and (4) if two catch trials were used in a block, then the two preceding trials showed one lift and one slap action. During a catch trial, participants had to indicate which of the two actions was the last one they observed before the question mark appeared, using a button press with their middle (action word on the left) or index (action word on the right) finger of the left hand. Feedback was given by either a green (correct response) or red (incorrect response) square that surrounded the selected action word from the moment the button was pressed until 2.5 sec after trial onset. If no response was given within 1.5 sec after trial onset, both action words were surrounded by a red square. To prevent potential finger motor planning strategies for catch trials, the position of the action words "lift" and "slap" (left/right or right/left) was chosen randomly. Both catch and noncatch trials lasted 3 sec each.

Similar to the "see" trials, the "do" trials were also presented in blocks of 16 trials of 3 sec each. Action instruction cues consisted a white arrow on a black background for 0.5 sec at trial onset, followed by a 2.5-sec black screen. The arrow pointed either upward or leftward, indicating a lift or slap action, respectively. We note that since the action instruction cues were symbolic, rather than linguistic, the "do" trials did not require verbal strategies by the participants.

Each run consisted two "chunks," where each chunk contained a single "do" block and a single "see" block in random order. Thus, each run contained two "do" (D) and two "see" (S) blocks of 48 sec each, with either order DSDS, DSSD, SDDS, and SDSD equally likely. This allowed us to perform cross validation (see below) with twice the number of chunks than runs. These blocks were preceded and followed by 16-sec baseline blocks during which the projection screen was not illuminated (black screen), resulting in runs of 272 sec in total. In each "do" and "see" block, the order of the two types of action trial (lift and slap) was randomized with the constraint that the order was first-, second-, and third-order counterbalanced (cf. Aguirre, 2007).

Each participant was scanned in a single session, consisting an anatomical scan followed by eight functional scans. Participants were instructed as follows: to rest their right hand on the table, on the right-hand side of the object (from their perspective); to only move their right hand during "do" trials; to ensure the object was not touched except during "do" trials; and to keep their left hand and arm under the table. To ensure that participants followed the instructions correctly, they completed a practice run of the experiment while the anatomical scan was acquired, and compliance to the instructions including proper action execution was monitored using an MRI-compatible camera attached to the scanner bore. Participants were instructed to use the viewed actions as a model and to match these as closely as possible during their own performance.

Data Acquisition

Data were acquired using a 3-T Philips MRI scanner with a SENSE phased-array head coil. For functional imaging, a T2*-weighted single shot gradient EPI sequence was used to achieve partial brain coverage. The scanning parameters were as follows: repetition time = 2000 msec; echo time = 35 msec, flip angle = 90°, 31 off-axial slices, $(2.5 \text{ mm})^3$ isotropic voxels, no slice gap, field of view = 240 × 240 mm^2, matrix = 96 × 96, anterior–posterior phase encoding, SENSE factor = 2. Slices were tilted approximately 15° in the anterior–superior direction from the anterior commissure–posterior commissure axis to achieve coverage of the inferior frontal, inferior parietal, superior temporal, and occipital cortices. Seven dummy scans were acquired before each functional scan to reduce possible effects of T1 saturation. Parameters for whole-brain T1-weighted anatomical scans were as follows: matrix = 256 × 256, 175 coronal slices, 1 mm^3 isotropic voxels, repetition time = 8.4 msec, echo time = 3.8 msec; flip angle = 8°.

Volume Preprocessing

Using AFNI (Cox, 1996), data were preprocessed (despiked with 3dDespike, slice-time corrected with 3dTshift and motion corrected with 3dvolreg) for each participant and each functional scan separately. For motion correction, the functional volumes were aligned to the "reference volume": the first volume in the first functional scan. The anatomical volume was aligned to the reference volume using 3dAllineate (Saad et al., 2009).

Surface Preprocessing

For each participant and hemisphere, anatomical surface meshes of the pial-gray matter ("pial") and smoothed gray matter–white matter ("white") boundaries were reconstructed using Freesurfer's recon-all program (Fischl, Sereno, Tootell, & Dale, 1999), and these were used to generate inflated and spherical surface meshes. On the basis of surface curvature, the spherical surfaces of all participants were aligned to a standard spherical surface (Fischl et al., 1999), which has been shown to provide better intersubject alignment than typical volume-based alignment. Using AFNI's MapIcosahedron (Saad, Reynolds, & Argall, 2004), these spherical surfaces were resampled to a standardized topology (an icosahedron in which each of the 20 triangles is subdivided in 10,000 small triangles), and the pial, white, and inflated surfaces were then converted to the same topology. This ensured that each node on the standardized surfaces represented a corresponding surface location across participants; therefore, group analyses could be conducted using a node-by-node analysis. Using the "surfing" toolbox (Oosterhof, Wiestler, & Diedrichsen, 2010), the affine transformation from Freesurfer's anatomical volume to the aligned anatomical volume was estimated using AFNI's align_epi_anat.py and applied to the coordinates of the surfaces to align them with the reference volume, which ensured alignment between the surfaces and the motion-corrected functional time series data.

The functional time series obtained after motion correction were projected to the surface using 3dVol2Surf as follows. Across the pial and white surfaces, line segments were constructed between the corresponding nodes. On each segment, 10 points were defined with equal distance. The value of the projected time series on the surface for each segment was based on the average value of the time series across the voxels containing the points on the corresponding segment.

The projected time series were then smoothed on the intermediate surface using a heat kernel as implemented in SurfSmooth (Chung, 2004) to obtain a smoothness of 5-mm FWHM, where the original time series was detrended with 12 polynomial basis functions before the initial smoothness was estimated.

For both volume-based and surface-based analyses, time series were converted to percent signal change by dividing the signal for each time point by 1% of the mean signal over the run before they were analyzed with the general linear model.

Univariate Analyses

A general linear model was used to estimate the BOLD response for the different conditions. We used different design matrices for activation and information mapping, but they shared the same regressors of no interest: Legendre polynomials up to the third degree in each run to remove low-frequency trends, motion parameter estimates (three for translation and three for rotation), and predictors for each catch trial separately. We used separate predictors for catch trials because the psychological and neural responses might differ across catch trials. All "do," "see," and catch trials were modeled by a 3-sec boxcar convolved with the AFNI's BLOCK4 canonical hemodynamic response function.

For activation mapping, the design matrix contained two regressors of interest for "do" and "see" trials (irrespective of the type of action; "lift" or "slap"). For information mapping, the design matrix contained two "do" and two "see" regressors of interest—corresponding to the two actions (lift and slap)—for each "chunk" (see above). Thus, 2 (action regressors per modality) × 2 (modalities per chunk) × 2 (chunks per run) × 8 (runs) = 64 action regressors of interest were used for this analysis.

Whole-brain Activation Mapping

To identify areas that responded more during the observation and execution of actions than during baseline blocks, surface-based activation maps were generated based on the spatially smoothed time series for each individual and the design matrices described earlier. These maps were subsequently analyzed at a group level with a standard t test against zero of the beta estimates. To identify areas commonly activated or deactivated across the "do" and "see" conditions, a signed conjunction group map was computed by taking, for each node separately, the minimum of the absolute t value across the "do" and "see" group maps, which was then multiplied by 1 if both "do" and "see" values were positive, by −1 if both values were negative, and by 0 otherwise.

ROI MVPA

On the basis of the univariate activation conjunction group map based on the data from all participants (i.e., collapsed across the two perspectives), we defined the center of PMv at the group level as the node with maximum conjunction value near the ventral precentral gyrus. Because there is potential anatomical and/or functional variability in the location of action representation areas across participants that might prevent such areas to be identified on a group map, voxels for multivariate pattern analyses (MVPA; Haynes & Rees, 2006; Norman, Polyn, Detre, & Haxby, 2006; Haxby et al., 2001; Edelman, Grill-Spector, Kushnir, & Malach, 1998) were selected based on the magnitude of their response in the univariate activation analysis. For each participant, the following steps were taken: first, a circle with 15-mm radius centered at the group peak node was defined on the individual's surface, and the node with the highest "do" and "see" conjunction value in that circle was selected as the individual's peak. We note that the size of the radius (15 mm) is somewhat arbitrary but in the same order of magnitude as the variability in location across studies that identified PMv (see below; Figure 5). Second, in the

volume, the voxel enclosing the individual's peak node was taken as the center of sphere with 10-mm radius. Third, the 100 voxels that showed the highest conjunction value in this sphere were selected for MVPA. Importantly, this algorithmic approach has the advantage that peak node and voxel selection is fully reproducible within and across studies and not biased by possibly arbitrary decisions of the experimenter to choose between multiple peaks.

Using these selected voxels, MVPA was conducted using a standard support vector machine as implemented by LIBSVM (Chang & Lin, 2011). For the unimodal "do" MVPA, action discriminability between lift and slap actions was computed using take-one-chunk-out cross validation, where the support vector machine classifier was tested on the "do" lift and slap t value estimates of one chunk, after it had been trained on all the other chunks. Then, by taking all chunks as a test chunk once, the classifier made an unbiased prediction for each action in each chunk, and classification accuracies (chance level is 50%, distributed binomially under the null hypothesis of no information) were converted to z scores (chance level: $z = 0$). Unimodal "see" MVPA was conducted similarity. For cross-modal MVPA, training and testing was based on activation estimates from different modalities (train on "do" and test on "see," and vice versa) and the accuracies averaged.

To obtain a more reliable and continuous z score that was not based on a single cross-validation procedure, we took a random subspace approach (similar to Kuncheva et al., 2010, with the only difference that we averaged accuracies rather than classification predictions) with 100 iterations, where in each iteration 50% of the highest 100 voxels in the conjunction analysis were selected randomly and used for MVPA; the resulting classification z scores were then averaged over iterations. To exclude the possibility that any effects could be due to main activation differences across modalities, for each voxel and each modality (do and see) separately, the mean activation across chunks was subtracted in each ROI before MVPA. Repeating this procedure with another set of random subsets and computing correlations across participants for both unimodal ("do" and "see") and cross-modal classification results showed that this method was highly reliable (min(r) = .98, max(p) = 10^{-17}).

We stress that even though voxel selection and MVPA was based on the same data, the MVPA results are not affected by circular analysis problems (Kriegeskorte et al., 2009; Vul, Harris, Winkielman, & Pashler, 2009), because the selection voxel criteria were based on univariate analyses where the two actions (lift and slap) were modeled by the same regressor. In other words, the voxel selection procedure used, by construction, no information about which action was performed or observed during each trial.

Whole-brain Surface-based MVPA

To identify other areas than our ROIs that potentially represent the two actions differently, we conducted a whole-brain surface-based "searchlight" analysis (Kriegeskorte, Goebel, & Bandettini, 2006) using the "surfing" toolbox similar to earlier work (Oosterhof, Wiestler, Downing, & Diedrichsen, 2011). Briefly, a searchlight was defined as a circle with variable radius that contained a hundred voxels in the gray matter (i.e., voxels that intersect the pial or white surface and voxels in between). A given node on the surface was taken as the center of a searchlight circle, the corresponding voxels used for MVPA—similar to ROI analyses described above—and classification z scores accuracy assigned to that center node. This procedure was repeated for every node on the surface, yielding a whole-brain information map.

To correct for multiple comparisons without the need of choosing an a priori uncorrected threshold, the resulting information maps were subjected to group analysis as follows: first, the average accuracy across participants for each node was used to compute a threshold-free cluster enhancement (TFCE) group map, based on formula (1) in Smith and Nichols (2009) with the recommended values of $h_0 = 0$, $E = 0.5$, $H = 2$ and dh = 0.1. Second, a null hypothesis TFCE map distribution (corresponding to classification accuracy at chance level, i.e., $z = 0$) was computed using a bootstrap procedure. In this procedure, the classification z scores across participants were sampled randomly with replacement, and their signs were randomly inverted with 50% chance (which is allowed on the null hypothesis of $z = 0$). We note that this approach preserved spatial smoothness in individual participant's information maps. A null hypothesis TFCE group map was computed based on these values, and the maximum TFCE value across the map was taken. This procedure was repeated a thousand times to obtain a null hypothesis distribution of maximum TFCE values. Third, statistical significance of nodes in the original TFCE group map were computed by dividing the number of times the node's value exceeded the TFCE values in the null hypothesis distribution by the number of iterations (a thousand).

RESULTS

Behavioral Results

Motion estimates across the functional scans exceeded 4-mm translation or 4° rotation in four participants. One other participant performed at exactly chance level (50% correct) during catch trials. These five participants were excluded, and all subsequent analyses were conducted using the remaining participants ($n_{1st} = 11$, $n_{3rd} = 13$).

One participant's left hand was mispositioned on the button box during the first three runs so that he pressed the wrong button during catch trials but, after realizing this, performed well (92% correct) during the remaining runs. All other participants performed well during catch trials (median(μ) = 88% correct, binominal median

$(p) = 7.4 \times 10^{-6}$, $\min(\mu) = 68\%$ correct, $\min(p) = .047$). No differences were found between participants that viewed actions from first-person ($\mu_{1st} = 88\%$) and third-person ($\mu_{3rd} = 88\%$) perspectives, Wilcoxon rank-sum $p = 1$.

PMv Localization

The conjunction group map of "do versus baseline" and "see versus baseline," collapsed across perspective (Figure 2), allowed us to localize left PMv near the precentral gyrus. Using the peak ROI coordinates on the conjunction group map, we identified a nearby maximum (within a 15-mm circle on the surface) in each participant's individual conjunction map (Table 2). We compared the individual's ROI center coordinates across participants between the two perspectives (Figure 3) using a one-way MANOVA on the x, y, and z coordinates and found no differences between the two groups (Wilk's $\Lambda = .84$, $\chi_3^2 = 3.50$, $p = .32$). Finally, we compared PMv's coordinates (-56, 1, 37) in our study with those reported in other studies. The meta-analysis by Van Overwalle et al. describes 22 univariate studies that report a group-based PMv area active during imitation. Although imitation is not identical to the task requirements of the present study, it is similar in that it requires both the execution and observation of actions. Therefore, an area showing a response to both the observation and execution of actions—as defined by the conjunction analysis to define PMv—is also expected to be active during imitation (although the reverse is not necessarily true). The coordinates in our study did not differ from those reported in a recent meta-analysis (Van Overwalle & Baetens, 2009) of imitated hand or finger

actions (Mahalanobis $D^2 = 4.19$, $p = .24$; see Figure 4). Similarly, Dinstein, Hasson, Rubin, and Heeger (2007) listed studies that report group-averaged coordinates of univariate studies showing a PMv area active during either action execution or observation, and also these coordinates did not differ from the PMv coordinates in the present study (Mahalanobis $D^2 = 4.22$, $p = .24$). To summarize, these analyses suggest that the location of the PMv as defined here is consistent with other studies.

PMv MVPA

Using the individual's PMv ROI centers defined above, for MVPA we selected voxels in their neighborhood that were maximally activated in the conjunction analysis (see Methods for details). Classification accuracies were significantly above chance (Figure 5; Table 2) for the unimodal "do" and "see" analyses with no interaction ($t(22) = 1.379$, $p = .18$) between the unimodal views ("do" and "see") and the perspective (first or third person). In the cross-modal analysis, a difference between the two perspective conditions was observed ($t(22) = 3.45$, $p = .002$; nonparametric rank-sum test $p = .007$): first-person observed actions showed cross-modal information (one-tailed $t(10) = 5.14$, $p = .0002$; nonparametric sign test $p = .001$), whereas the third-person observed actions did not (one-tailed $t(12) = .71$, $p = .25$; nonparametric sign test $p = .27$). We note that, although the nonparametric tests can be less sensitive than parametric tests, they are also more conservative in that they do not require normality assumptions. In other words, the consistency of our results across the parametric and nonparametric tests excludes

Figure 2. ROI voxel selection method. (A) Surface-based conjunction map of observed and executed actions versus baseline from data collapsed across the first- and third-person perspective conditions. The blue cross indicates the peak in PMv. The inset shows a detailed view of the PMv peak and surrounding cortex (CS, central sulcus; pCG, precentral gyrus). (B) Conjunction maps of two individual participants. The group peak is projected on the individual's brain taken as the center of a circle and the individual's peak (denoted by △ and ○, respectively) determined. (C) Within a sphere centered around the individuals' peak, the voxels in the volume that were most active in the conjunction analysis are selected for subsequent MPVA (see Methods for details).
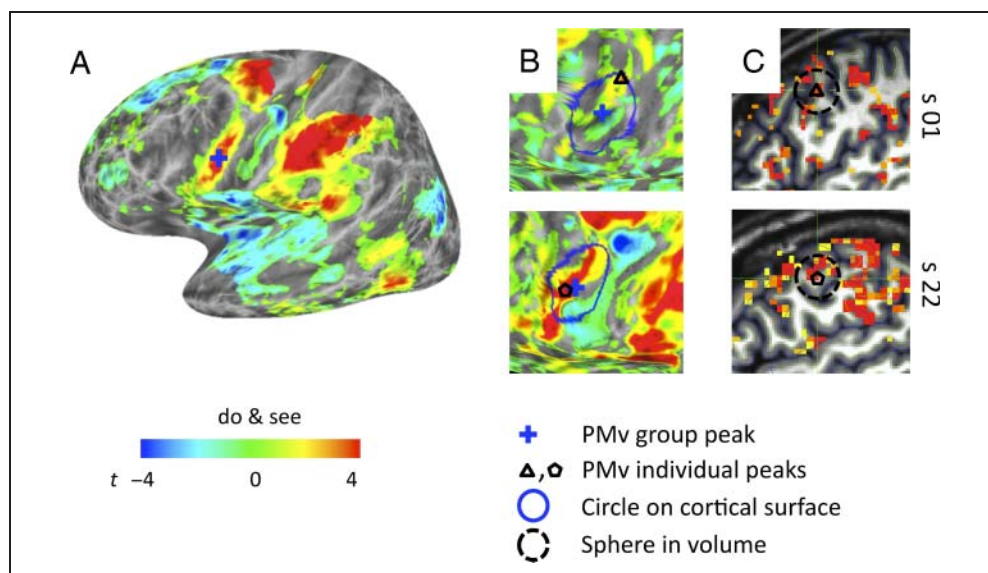
Table 2. PMv Talairach Coordinates and MPVA Classification *z* Scores Shown in Figure 5 for Action Representations in Individual Participants

	Coordinates			MVPA z Score		
View	x	y	z	Do	See	Cross-modal
1	−49	−4	43	0.25	0.66	0.53
1	−54	−4	46	0.72	1.18	0.58
1	−61	−5	36	0.3	0.28	0.06
1	−54	6	39	0.86	−0.22	1.04
1	−60	1	44	1.03	1.33	0.89
1	−58	−1	40	0.71	0.9	0.16
1	−52	2	40	0.35	0.8	0.58
1	−56	−1	45	1.35	0.67	0.88
1	−64	3	31	−0.92	−0.42	0.1
1	−58	−8	38	1.56	−0.56	0.54
1	−57	1	27	0.35	1.36	0.93
3	−56	−5	36	0.9	−0.11	0.32
3	−58	2	35	−0.48	0.88	0.3
3	−51	3	39	0.25	0.77	0.25
3	−54	5	28	0.5	1.09	−0.11
3	−56	0	43	2.88	0.79	−0.24
3	−55	7	35	0.48	−0.34	−0.48
3	−51	11	17	0.91	0.65	0.14
3	−60	11	44	0.04	0.19	−0.32
3	−60	−7	26	−0.03	−0.01	0.14
3	−51	6	51	2.01	−0.65	0.86
3	−62	−1	29	0.81	−0.14	0.1
3	−52	−9	35	1.2	0.09	0.22
3	−62	6	36	2.7	0.35	−0.06

the possibility that our effects are driven by a few "outlier" participants (see also the dots in Figure 5, which represent the results of individual participants).

To assess the robustness of this effect, we varied number of voxels initially selected (50, 100, or 150) and the percentage of the voxels with highest conjunction values used for random subsets (25%, 50%, or 75%). As shown in Figure 6, we found that the effect was reliable across these parameters: All comparisons between first- and third-person perspectives showed that the former showed higher cross-modal classification accuracies, and this effect was significant except for ROIs with a low number of voxels. Finding weaker classification results with smaller sets of voxels is as expected and conforms to previous findings (Oosterhof et al., 2011; Cox & Savoy, 2003).

Additional Analyses

An alternative explanation for cross-modal coding differences between first- and third-person perspectives in PMv is modulation by attention or depth of processing when participants viewed the actions. For example, participants may have attended less to actions presented in the third-person perspective condition than to those in the first-person condition, which might have led to weaker action-specific coding. To rule out such explanations, we conducted several additional analyses.

First, attentional effects have been shown to modulate overall BOLD response (e.g., Kanwisher & Wojciulik, 2000). We considered the voxels used for MVPA and computed their average activity for the executed and observed actions separately. As shown in Figure 7, there were no perspective effects in the peak or median response for executed (peak $t(22) = -1.45$, $p = .16$; median $t(22) = 0.76$, $p = .45$) or observed (peak $t(22) = -0.44$, $p = .65$; median $t(22) = 0.02$, $p = .98$) actions. Also the signed conjunction value did not show a difference (peak $t(22) = -1.01$, $p = .32$; median $t(22) = 0.07$, $p = .95$).

Second, using the same procedure as for PMv, we localized several other areas that have been suggested as part of an action representation network and that were significantly activated in our univariate conjunction analysis: bilateral occipital temporal cortex (OT), PCa, and dorsal
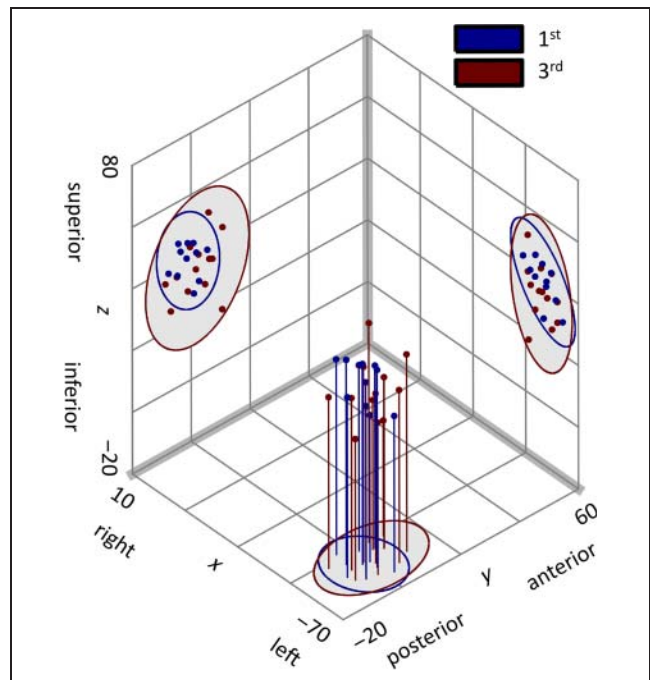
Figure 3. Comparison of PMv Talairach (*x, y, z*) coordinates for participants in the first-person (blue spheres) and third-person (red spheres) perspectives. Spheres connected to lines indicate the location in 3-D space; the lower end of each line is positioned on the "floor" *xy* plane (*z* = −20). Spheres not connected to lines indicate projections of the same coordinates on the *xz* (*y* = 60) and *yz* (*x* = 10) planes. Gray ellipses show 2-D projections of the 95% confidence ellipsoids for the first- and third-person view conditions.

Figure 4. Comparison of PMv coordinates in the present study (large red sphere) with coordinates of other studies (smaller blue spheres) as reported by Van Overwalle and Baetens (2009; left) and Dinstein et al. (2007; right). Conventions as in Figure 3.
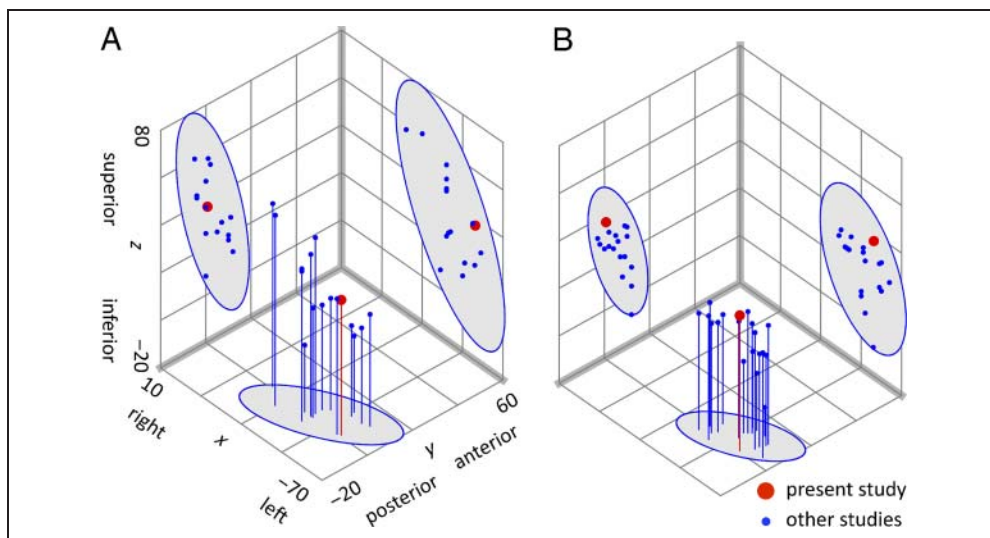
premotor cortex (PMd). All areas showed reliable activation for both observing and executing actions ($ps < .001$ in the conjunction analysis, not corrected for multiple comparisons). If attention modulates cross-modal coding, then one would expect a similar modulation of perspective in these action representation areas. However, we found no such cross-modal differences in these areas (Figure 8; Table 3). It is unlikely that this is due to lack of power, because PCa and OT showed, consistent with earlier work, reliable cross-modal coding in both perspectives. Furthermore, several of these areas showed a (nonsignificant; $min(p) = .22$) stronger cross-modal coding in the third-person perspective than in the first-person perspective, an effect in the opposite direction as would be expected for general weaker encoding in the latter perspective. This lack of differences across conditions in the other areas makes it unlikely that the differences in PMv are due to differences in attention, task strategies, or depth of processing.

Third, we conducted a surface-based searchlight analysis to identify cross-modal areas in the whole brain, unrestricted by a priori assumptions of areas involved in action representation. As shown in the group map correcting for multiple comparisons ($p = .05$), while PMv

did not survive correction for multiple comparisons for the first-person condition, we found a reliable cluster in left PCa for both perspectives (Figure 9; Table 4). Importantly, the location and size of this cluster was consistent across the two groups.

Together with the earlier observation that we found no differences in behavioral performance between the two perspectives, this makes an explanation due to attentional effects unlikely.

DISCUSSION

Using an ROI-based MVPA approach, we showed that perspective modulates cross-modal visuomotor representations in PMv, a key region of the putative human mirror system. Actions perceived from a first-person perspective showed reliable cross-modal coding, but the same actions perceived from a third-person perspective did not.

At a neural level, our results may seem incongruent with macaque studies of mirror neurons. For example, the first study that quantified the properties of neurons in macaque area F5—the putative homologue of human PMv—showed reliable coding in mirror neurons for actions that were executed by an experimenter and observed by

Figure 5. PMv MVPA results. Classification accuracies for unimodal do, unimodal see, and cross-modal action-specific representations in left PMv are shown for first-person (blue) and third-person (red) observed actions. Accuracies are denoted by z scores, where $z = 0$ denotes chance (no action-specific representations). Colored dots indicate individual participants. $*p < .05$; $**p < .01$; $***p < .001$.
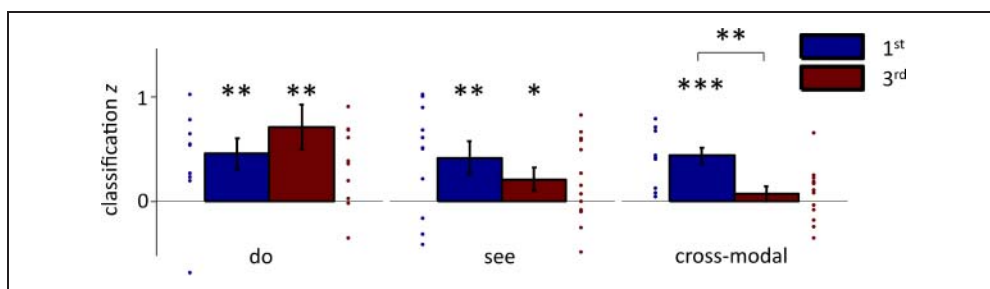
Figure 6. PMv voxel
selection robustness analysis.
Classification accuracies for
cross-modal action-specific
representations in the left
PMv using different voxel
selection parameters for
MVPA. Accuracies are shown
for the highest 50, 100, and
150 (top, middle, and bottom
rows, respectively) with random
subsets of 25%, 50%, and 75%
(left, middle, and right columns,
respectively) of the maximally
active voxels in the univariate
conjunction analysis across
observed and executed actions
(see Methods for details).
Conventions as in Figure 5.
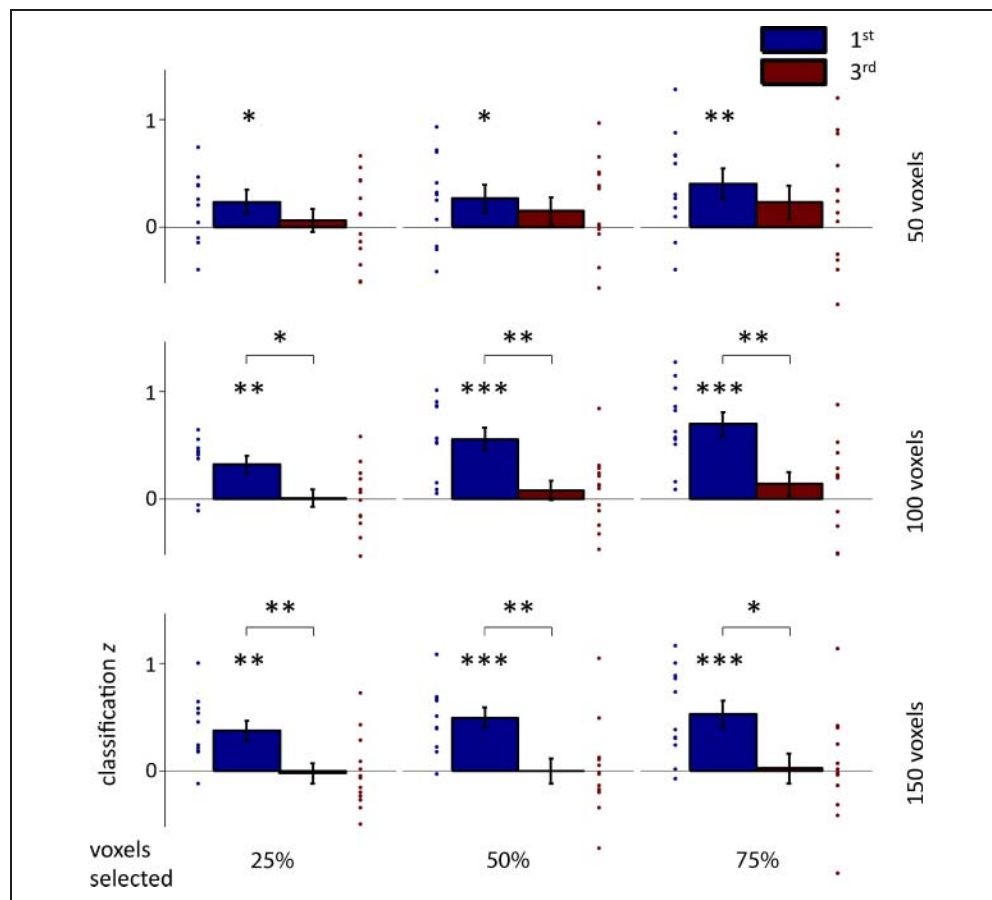
the monkey in a third-person view (Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). Of the 532 neurons reported in that study, 29 (5%) showed "strictly congruent" properties, that is, increased firing for the observation and execution of a specific action with a specific goal. Another 25 (5%) showed "mirror-like" properties, that is, increased firing for the observation of actions but without motor properties. A more recent study (Caggiano et al., 2011) investigated the role of perspective on firing rates of motor neurons in F5, where macaques observed actions from 0,

90°, or 180° perspectives. After having established that action videos did elicit changes in firing rates similar to live actions, the researchers found that different neurons showed different tuning profiles across the different perspectives, with slightly more neurons responding specifically to actions observed from a first-person (0) perspective ($n = 27$) than a third-person (90° and 180°) perspective ($n = 15$ and $n = 18$, respectively).

Taken together, these results suggest that (1) mirror neurons are a minority among other types of neurons in

Figure 7. Univariate PMv
analysis. Univariate maximum
(top) and median (bottom)
univariate conjunction t scores
across observed and executed
actions in left PMv. Conventions
as in Figure 5.
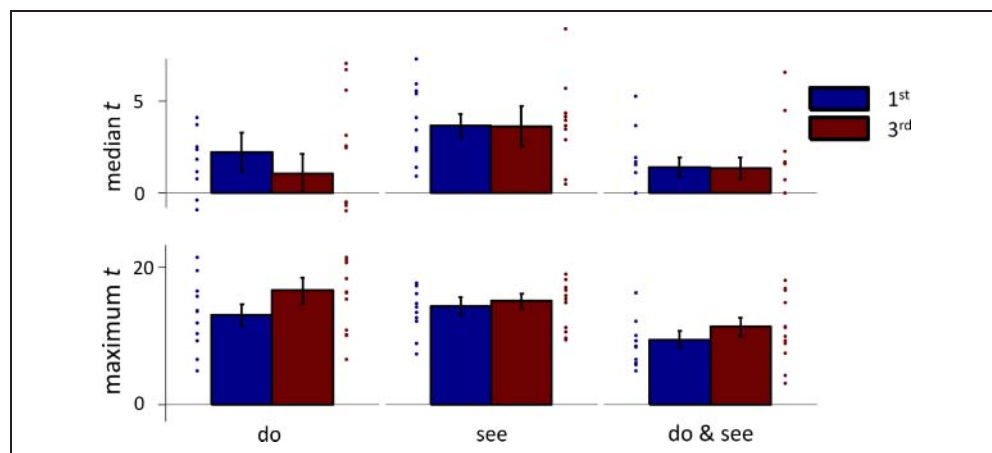
Figure 8. Classification accuracies ($z = 0$ is chance level) for cross-modal action-specific representations in several brain areas suggested as representing actions and showing activity for both observed and executed actions (see Methods). The top and bottom row show accuracies for the left and right hemispheres. Conventions as in Figure 5.
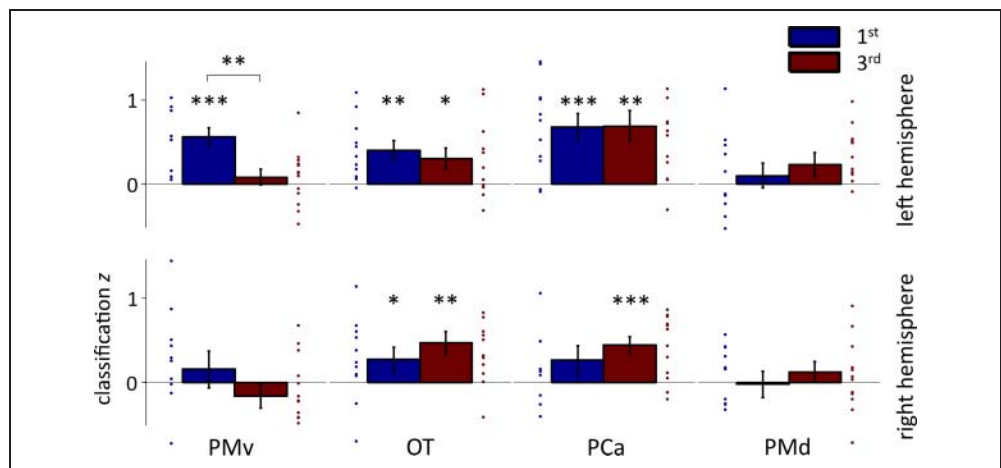
F5, (2) there are dissociable neural populations showing unimodal coding for observed actions (mirror-like neurons) and for cross-modal coding (strictly congruent mirror neurons), and (3) there are dissociable neural populations for actions observed from first- and third-person perspectives, with a potential bias for stronger coding of actions from a first-person perspective. Although extrapolating from these results to human fMRI is not straightforward and must remain speculative, these properties are consistent with the results of others and our own, namely (1) evidence for cross-modal coding in PMv is weak, because of the small proportion of neurons involved; (2) unimodal coding for vision and motor actions is possible without the necessary presence of cross-modal coding in the same area; and (3) observed actions may be coded differently in PMv for first- versus third-person perspective, and the former may be coded more strongly.

Considering studies in humans, the only study reporting cross-modal action coding in PMv so far (Kilner et al., 2009) used a repetition–suppression paradigm. In this approach, an action was either observed (in a first-person view) or executed, followed immediately by either the execution or observation of either the same action or another action. Reduced activity for repeated actions compared with non-repeated actions was found in PMv. Insofar as that study showed cross-modal coding when actions were observed from a first-person perspective, it is consistent with our findings. However, Kilner and colleagues used two different objects as the targets of the two actions, which means that action-specific and object-specific coding could not be dissociated. In contrast, the current study shows that cross-modal coding in PMv can be attributed to specific coding of different actions, not of specific objects, and, more importantly, that this coding is modulated by the perspective of observed actions and does not appear to generalize to the third-person view.

Several other studies have used either RS or MVPA approaches to study specific coding of actions—and failed to find cross-modal visuomotor coding in PMv. Apart from variations in methodology employed in those studies (in-

cluding variations in perspective), the interpretation of such results has varied considerably as well. For example, two studies (Lingnau, Gesierich, & Caramazza, 2009; Chong, Cunnington, Williams, & Kanwisher, 2008) both used RS and found partial evidence for cross-modal coding, that is, RS effects from one modality (do or see) to the other (see or do, respectively), but not vice versa. Chong and colleagues interpreted this result as evidence for human mirror neurons, whereas Lingnau and colleagues made the opposite interpretation, namely that finding only unidirectional RS argued against the existence of a mirror-like mechanism.

Table 3. Coordinates of Areas Shown in Figure 7

Area	Perspective	x	y	z
Left PMv	1st	−57	−1	39
	3rd	−56	2	35
Left OT	1st	−49	−73	1
	3rd	−49	−73	2
Left PCa	1st	−39	−37	47
	3rd	−41	−36	48
Left PMd	1st	−35	−8	55
	3rd	−40	−7	55
Right PMv	1st	56	6	37
	3rd	57	1	37
Right OT	1st	52	−63	−1
	3rd	52	−60	2
Right PCa	1st	36	−41	49
	3rd	35	−40	54
Right PMd	1st	44	−4	56
	3rd	46	−2	55

Figure 9. Classification accuracies (50% is chance level) on surface-based information group map for cross-modal action-specific representations in the first-person (top) and third-person (bottom) perspective conditions. Nodes surrounded by a blue line survived TFCE with correction for multiple comparisons at $p = .05$ (see Methods).
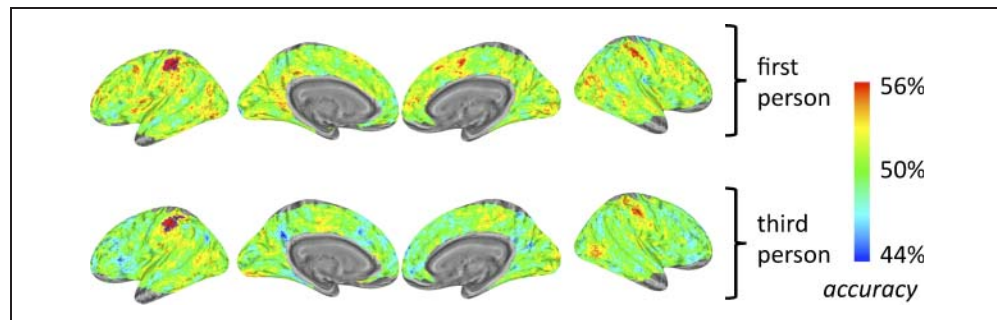
Additionally, a challenge is posed by the differences between the assumptions behind interpretations of RS (Epstein, Parker, & Feiler, 2008; Summerfield, Trittschuh, Monti, Mesulam, & Egner, 2008; Sawamura, Orban, & Vogels, 2006) and MVPA results (Kriegeskorte, Cusack, & Bandettini, 2010; Op de Beeck, 2010). It may be reassuring that a direct comparison of RS and MVPA found that both approaches yielded qualitatively similar results (with MVPA being more sensitive) in visual cortex (Sapountzis, Schluppeck, Bowtell, & Peirce, 2010), but it is unclear whether these results generalize to other brain areas. Taken together, a full understanding of the neural mechanisms underlying RS and MVPA and the implications of results from these paradigms for understanding the human mirror system is unlikely to be resolved with fMRI alone. Future studies may require combination of neurophysiological with fMRI methods.

Apart from methodological considerations in interpreting our findings, few would dispute that first- and third-person perspectives differ in terms of phenomenology, behavior, and neurally. At a phenomenological level, a first-person perspective is important for self-consciousness (Newen, 2003). For example, there is a clear distinction between executing an action yourself and imitating an action performed by someone else. Unlike observing someone else's action, the process of observing one's own action is associated with a planned goal of the actor present before any movement has taken place (Von Hofsten, 2004), requires perceived ownership of the effector used to execute the action (Synofzik, Vosgerau, & Newen, 2008), requires complex coordination of several muscles of the effector that executes the action (Aflalo & Graziano, 2006), and involves visuomotor and proprioceptive neural feedback mechanisms (Balslev, Cole, & Miall, 2007).

Table 4. Brain Areas Showing Cross-modal Action Representations Surviving Multiple-comparison Corrected ($p = .05$, TFCE; see Methods) in a Whole-brain Searchlight as Reported in Figure 9

	x	y	z	Area (mm^2)
1st person PCa	−48	−30	41	768
3rd person PCa	−49	−31	42	819

Various behavioral studies have also shown speed and accuracy advantages for first-person perspectives (e.g., Vogt, Taylor, & Hopkins, 2003; Maeda, Kleiner-Fisman, & Pascual-Leone, 2002). In contrast, the third-person perspective is almost universally experienced when perceiving others' actions. Therefore, perceiving actions from this perspective is more likely to engage mechanisms involved in interpreting others' behavior—such as a "theory of mind" system. Our finding of view invariance in parietal and occipitotemporal regions suggests it may be that these regions (more so than premotor regions) provide input to such mechanisms. Altogether, there are many potential reasons why actions observed from a first-person perspective could, at a neural level, be represented more strongly or distinctly from actions observed from a third-person perspective.

Such an explanation is consistent with hierarchical coding frameworks that describe actions in a motor hierarchy from kinematics to goals and intentions that is implemented in a distributed system of connected areas in the brain (Hamilton & Grafton, 2007; Kilner, Friston, & Frith, 2007; Wolpert, Doya, & Kawato, 2003). In this hierarchy, posterior areas in occipitotemporal cortex encode visual properties of actions, parietal areas code for action goals and intentions, and frontal areas code for precise reach and grasp motor control. Indeed, our results show involvement of all these areas in action-specific coding. First, our results show that OT codes actions cross-modally (irrespective of perspective; Figure 8) replicating earlier findings (Oosterhof et al., 2010), which is consistent with not only a visual role for this general region but also motor coding of actions (Orlov, Makin, & Zohary, 2010; Astafiev, Stanley, Shulman, & Corbetta, 2004), possibly with specific preference for coding manual actions (Bracci, Ietswaart, Peelen, & Cavina-Pratesi, 2010). Second, many lines of evidence point to a role of PCa for integrating perception, action, and cognition (Gottlieb, 2007) and representing goals abstractly—potentially at the top of the action representation hierarchy (Hamilton & Grafton, 2007)—which is consistent with the reliable perspective-independent cross-modal coding found in other studies (see Table 1; Fogassi, Ferrari, Gesierich, Rozzi, Chersi, & Rizzolatti, 2005) and the present results (Figures 8 and 9). Third, PMv has been suggested to be involved in the precise implementation (e.g., type of grasp) of actions (Hamilton

& Grafton, 2007; Kilner et al., 2007) which—as argued above—could explain stronger cross-modal coding for actions observed from a first-person perspective than a third-person perspective.

Although our results do not rule out the existence of mirror neurons in PMv, they emphasize the important modulatory properties of perspective on cross-modal action coding. This is an issue that—despite being central to the proposed function of mirror neurons, namely that they are involved in the coding of the actions of others—has not been investigated systematically in humans so far. Future studies can address which areas code for specific actions in more detail by delineating the representational similarity structure (cf. Kriegeskorte, Mur, & Bandettini, 2008) of different aspects of actions (e.g., effector, trajectory, object, grasp posture, and goal) across different modalities, perspectives, and areas of the brain.

REFERENCES

Aflalo, T. N., & Graziano, M. S. A. (2006). Possible origins of the complex topographic organization of motor cortex: Reduction of a multidimensional space onto a two-dimensional array. *Journal of Neuroscience, 26,* 6288–6297.

Aguirre, G. K. (2007). Continuous carry-over designs for fMRI. *Neuroimage, 35,* 1480–1494.

Astafiev, S. V., Stanley, C. M., Shulman, G. L., & Corbetta, M. (2004). Extrastriate body area in human occipital cortex responds to the performance of motor actions. *Nature Neuroscience, 7,* 542–548.

Balslev, D., Cole, J., & Miall, R. C. (2007). Proprioception contributes to the sense of agency during visual observation of hand movements: Evidence from temporal judgments of action. *Journal of Cognitive Neuroscience, 19,* 1535–1541.

Bracci, S., Ietswaart, M., Peelen, M. V., & Cavina-Pratesi, C. (2010). Dissociable neural responses to hands and non-hand body parts in human left extrastriate visual cortex. *Journal of Neurophysiology, 103,* 3389–3397.

Caggiano, V., Fogassi, L., Rizzolatti, G., Pomper, J. K., Thier, P., Giese, M. A., et al. (2011). View-based encoding of actions in mirror neurons of area F5 in macaque premotor cortex. *Current Biology, 21,* 144–148.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2,* 27.

Chong, T. T. J., Cunnington, R., Williams, M. A., Kanwisher, N., & Mattingley, J. B. (2008). fMRI adaptation reveals mirror neurons in human inferior parietal cortex. *Current Biology, 18,* 1576–1580.

Chong, T. T.-J., Cunnington, R., Williams, M., Kanwisher, N., & Mattingley, J. B. (2008). fMRI adaptation reveals mirror neurons in human inferior parietal cortex—Supplemental material. *Current Biology, 18,* 1576–1580.

Chung, M. (2004). Heat kernel smoothing and its application to cortical manifolds. Technical report 1090. Department of Statistics, University of Wisconsin—Madison. http://www.stat.wise.edu/~mchung/papers/heatkernel_tech.pdf.

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": Detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage, 19,* 261–270.

Cox, R. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, an International Journal, 29,* 162–173.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events—A neurophysiological study. *Experimental Brain Research, 91,* 176–180.

Dinstein, I., Gardner, J. L., Jazayeri, M., & Heeger, D. J. (2008). Executed and observed movements have different distributed representations in human aIPS. *Journal of Neuroscience, 28,* 11231–11239.

Dinstein, I., Hasson, U., Rubin, N., & Heeger, D. J. (2007). Brain areas selective for both observed and executed movements. *Journal of Neurophysiology, 98,* 1415–1427.

Dinstein, I., Thomas, C., Behrmann, M., & Heeger, D. J. (2008). A mirror up to nature. *Current Biology, 18,* R13–R18.

Edelman, S., Grill-Spector, K., Kushnir, T., & Malach, R. (1998). Toward direct visualization of the internal shape representation space by fMRI. *Psychobiology, 26,* 309–321.

Epstein, R. A., Parker, W. E., & Feiler, A. M. (2008). Two kinds of FMRI repetition suppression? Evidence for dissociable neural mechanisms. *Journal of Neurophysiology, 99,* 2877–2886.

Ferrari, P. F., Rozzi, S., & Fogassi, L. (2005). Mirror neurons responding to observation of actions made with tools in monkey ventral premotor cortex. *Journal of Cognitive Neuroscience, 17,* 212–226.

Fischl, B., Sereno, M. I., Tootell, R. B. H., & Dale, A. M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping, 8,* 272–284.

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

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119,* 593–609.

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. (2008). The observation and execution of actions share motor and somatosensory voxels in all tested subjects: Single-subject analyses of unsmoothed fMRI data. *Cerebral Cortex, 19,* 1239–1255.

Gottlieb, J. (2007). From thought to action: The parietal cortex as a bridge between perception, action, and cognition. *Neuron, 53,* 9–16.

Hamilton, A., & Grafton, S. (2006). Goal representation in human anterior intraparietal sulcus. *Journal of Neuroscience, 26,* 1133–1137.

Hamilton, A., & Grafton, S. (2007). The motor hierarchy: From kinematics to goals and intentions. In P. Haggard, R. Rossetti, & M. Kawato (Eds.), *Attention and performance 22. The sensorimotor foundations of higher cognition* (pp. 590–616). Oxford, UK: Oxford University Press.

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

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping

representations of faces and objects in ventral temporal cortex. *Science, 293,* 2425–2430.

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience, 7,* 523–534.

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience, 21,* 1229–1243.

Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience, 7,* 942–951.

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

Johnson-Frey, S., Maloof, F., Newman-Norlund, R., Farrer, C., Inati, S., & Grafton, S. (2003). Actions or hand-object interactions? Human inferior frontal cortex and action observation. *Neuron, 39,* 1053–1058.

Kanwisher, N., & Wojciulik, E. (2000). Visual attention: Insights from brain imaging. *Nature Reviews Neuroscience, 1,* 91–100.

Kilner, J., Friston, K. J., & Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing, 8,* 159–166.

Kilner, J., Neal, A., Weiskopf, N., Friston, K. J., & Frith, C. D. (2009). Evidence of mirror neurons in human inferior frontal gyrus. *Journal of Neuroscience, 29,* 10153–10159.

Kriegeskorte, N., Cusack, R., & Bandettini, P. (2010). How does an fMRI voxel sample the neuronal activity pattern: Compact-kernel or complex spatiotemporal filter. *Neuroimage, 49,* 1965–1976.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences, U.S.A., 103,* 3863–3868.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience, 2,* 3863–3868.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience, 12,* 535–540.

Kroliczak, G., Mcadam, T. D., Quinlan, D. J., & Culham, J. C. (2008). The human dorsal stream adapts to real actions and 3D shape processing: A functional magnetic resonance imaging study. *Journal of Neurophysiology, 100,* 2627–2639.

Kuncheva, L. I., Rodriguez, J. J., Plumpton, C. O., Linden, D. E. J., & Johnston, S. J. (2010). Random subspace ensembles for fMRI classification. *IEEE Transactions on Medical Imaging, 29,* 531–542.

Lingnau, A., Gesierich, B., & Caramazza, A. (2009). Asymmetric fMRI adaptation reveals no evidence for mirror neurons in humans. *Proceedings of the National Academy of Sciences, U.S.A., 106,* 9925–9930.

Maeda, F., Kleiner-Fisman, G., & Pascual-Leone, A. (2002). Motor facilitation while observing hand actions: Specificity of the effect and role of observer's orientation. *Journal of Neurophysiology, 87,* 1329–1335.

Majdandžić, J., Bekkering, H., & Van Schie, H. (2009). Movement-specific repetition suppression in ventral and dorsal premotor cortex during action observation. *Cerebral Cortex, 19,* 2736–2745.

Newen, A. (2003). Self-representation: Searching for a neural signature of self-consciousness. *Consciousness and Cognition, 12,* 529–543.

Norman, K., Polyn, S., Detre, G., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences, 10,* 424–430.

Ogawa, K., & Inui, T. (2011). Neural representation of observed actions in the parietal and premotor cortex. *Neuroimage, 56,* 728–735.

Oosterhof, N. N., Wiestler, T., & Diedrichsen, J. (2010). *Surfing: A Matlab toolbox for surface-based voxel selection.* Retrieved from surfing.sourceforge.net.

Oosterhof, N. N., Wiestler, T., Downing, P. E., & Diedrichsen, J. (2011). A comparison of volume-based and surface-based multi-voxel pattern analysis. *Neuroimage, 56,* 593–600.

Oosterhof, N. N., Wiggett, A. J., Diedrichsen, J., Tipper, S. P., & Downing, P. E. (2010). Surface-based information mapping reveals crossmodal vision-action representations in human parietal and occipitotemporal cortex. *Journal of Neurophysiology, 104,* 1077–1089.

Op de Beeck, H. P. (2010). Against hyperacuity in brain reading: Spatial smoothing does not hurt multivariate fMRI analyses. *Neuroimage, 49,* 1943–1948.

Orlov, T., Makin, T. R., & Zohary, E. (2010). Topographic representation of the human body in the occipitotemporal cortex. *Neuron, 68,* 586–600.

Ortigue, S., Thompson, J. C., Parasuraman, R., & Grafton, S. T. (2009). Spatio-temporal dynamics of human intention understanding in temporo-parietal cortex: A combined EEG/fMRI repetition suppression paradigm. *PLoS ONE, 4,* e6962. doi:10.1371/journal.pone.0006962

Poulton, E. C. (1982). Influential companions: Effects of one strategy on another in the within-subjects designs of cognitive psychology. *Psychological Bulletin, 91,* 673–690.

Ramsey, R., & Hamilton, A. (2010). Understanding actors and object-goals in the human brain. *Neuroimage, 50,* 1142–1147.

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

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3,* 131–141.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2002). Motor and cognitive functions of the ventral premotor cortex. *Current Opinion in Neurobiology, 12,* 149–154.

Saad, Z., Glen, D. R., Chen, G., Beauchamp, M. S., Desai, R., & Cox, R. (2009). A new method for improving functional-to-structural MRI alignment using local Pearson correlation. *Neuroimage, 44,* 839–848.

Saad, Z., Reynolds, R., & Argall, B. (2004). SUMA: An interface for surface-based intra- and inter-subject analysis with AFNI. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (1510–1513)*.* Arlington VA: IEEE.

Sapountzis, P., Schluppeck, D., Bowtell, R., & Peirce, J. W. (2010). A comparison of fMRI adaptation and multivariate pattern classification analysis in visual cortex. *Neuroimage, 49,* 1632–1640.

Sawamura, H., Orban, G. A., & Vogels, R. (2006). Selectivity of neuronal adaptation does not match response selectivity: A single-cell study of the fMRI adaptation paradigm. *Neuron, 49,* 307–318.

Smith, S. M., & Nichols, T. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage, 44,* 83–98.

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M.-M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience, 11,* 1004–1006.

Synofzik, M., Vosgerau, G., & Newen, A. (2008). I move, therefore I am: A new theoretical framework to investigate

agency and ownership. *Consciousness and Cognition, 17,* 411–424.

Turella, L., Pierno, A. C., Tubaldi, F., & Castiello, U. (2009). Mirror neurons in humans: Consisting or confounding evidence. *Brain and Language, 108,* 10–21.

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *Neuroimage, 48,* 564–584.

Vogt, S., Taylor, P., & Hopkins, B. (2003). Visuomotor priming by pictures of hand postures: Perspective matters. *Neuropsychologia, 41,* 941–951.

Von Hofsten, C. (2004). An action perspective on motor development. *Trends in Cognitive Sciences, 8,* 266–272.

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science, 4,* 274.

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society, Series B, Biological Sciences, 358,* 593–602.