Articles in PresS. J Neurophysiol (June 10, 2010). doi:10.1152/jn.00326.2010

Running head: Crossmodal action representations

Surface-based information mapping reveals crossmodal vision-action representations in human parietal and occipitotemporal cortex

Nikolaas N. Oosterhof[1]
Alison J. Wiggett[1]
Jörn Diedrichsen[1,2]
Steven P. Tipper[1]
Paul E. Downing[1]

[1]School of Psychology, Bangor University, United Kingdom
[2]Institute of Cognitive Neuroscience, University College London, UK

Address correspondence to: Nikolaas N. Oosterhof, School of Psychology, Bangor University, Adeilad Brigantia, Penrallt Road, Bangor LL57 2AS, United Kingdom.
Electronic mail: n.oosterhof@bangor.ac.uk

Keywords: mirror neurons, action representation, functional magnetic resonance imaging, multi voxel pattern analysis, information mapping, surface mapping

REVISED MANUSCRIPT

Abstract

Many lines of evidence point to a tight linkage between the perceptual and motoric representations of actions. Numerous demonstrations show how the visual perception of an action engages compatible activity in the observer's motor system. This is seen for both intransitive actions (e.g. in the case of unconscious postural imitation) and for transitive actions (e.g. grasping an object). While the discovery of "mirror neurons" in macaques has inspired explanations of these processes in human action behaviours, the evidence for areas in the human brain that similarly form a crossmodal visual/motor representation of actions remains incomplete. To address this, in the present study, participants performed and observed hand actions while being scanned with fMRI. We took a data-driven approach by applying whole-brain information mapping using a multi-voxel pattern analysis (MVPA) classifier, performed on reconstructed representations of the cortical surface. The aim was to identify regions in which local voxel-wise patterns of activity can distinguish among different actions, across the visual and motor domains. Experiment 1 tested intransitive, meaningless hand movements, while Experiment 2 tested object-directed actions (all right-handed). Our analyses of both experiments revealed crossmodal action regions in the lateral occipitotemporal cortex (bilaterally) and in the left postcentral gyrus/anterior parietal cortex. Furthermore, in Experiment 2 we identified a gradient of bias in the patterns of information in the left hemisphere postcentral / parietal region. The postcentral gyrus carried more information about the effectors used to carry out the action (fingers vs whole hand), while anterior parietal regions carried more information about the goal of the action (lift vs punch). Taken together, these results provide evidence for common neural coding in these areas of the visual and motor aspects of actions, and demonstrate further how MVPA can contribute to our understanding of the nature of distributed neural representations.

Introduction

There is increasing evidence for a direct link between perception and action: perceiving another person's action activates the same representations as does the actual performance of the action. Such common codes between perceiving and producing actions enable humans to embody the behavior of others and to infer the internal states driving it (e.g., Barsalou et al. 2003). That is, by creating common representations between ourselves and another person, we have a deeper understanding of their current states, and are better able to predict their future behaviour, facilitating complex social interactions. However, the basis of the brain's crucial ability to relate one's own actions to those of others remains poorly understood.

One possible contributing neural mechanism is found in macaque single-cell studies of so-called "mirror neurons" (di Pellegrino et al. 1992), which have inspired many theories of the neural basis of a range of human social processes such as theory of mind, language, imitation, and empathy (Agnew et al. 2007; Corballis 2009; Rizzolatti and Fabbri-Destro 2008). Surprisingly, given the extent of such theorizing, the evidence for a human "mirror system" – that is, for brain areas in which the visual and motor aspects of actions are represented in a common code -- is weak (Dinstein et al. 2008b).

Numerous functional neuroimaging studies have identified brain regions that are active during both the observation and the execution of actions (e.g. Etzel et al. 2008; Iacoboni et al. 1999). While these studies show spatial overlap of frontal and parietal activations elicited by action observation and execution, they do not demonstrate representational overlap between visual and motor action representations. That is, spatially overlapping activations could reflect different neural populations in the same broad brain regions (Gazzola and Keysers 2009; Morrison and Downing 2007; Peelen and Downing 2007b). Spatial overlap of activations per se cannot establish whether the patterns of neural response are similar for a given action (whether it is seen or performed) but different for different actions, an essential property of the "mirror system" hypothesis.

Several recent studies have addressed this problem with fMRI-adaptation designs (Grill-Spector and Malach, 2001). Dinstein et al. (2007) used this approach to identify areas (such as the anterior intraparietal sulcus; aIPS) in which the BOLD response was reduced when the same action was either seen or executed twice in a

row. However, none of the areas tested showed adaptation from perception to performance of an action, or vice versa. Two subsequent studies revealed adaptation from performance to observation (Chong et al. 2008), or vice versa (Lingnau et al. 2009), but neither showed bidirectional adaptation across the visual and motor modalities. Most recently, Kilner et al. (2009), using a task that involved goal-directed manual actions, showed adaptation effects bi-directionally in the inferior frontal gyrus (superior parietal cortex was not measured).

Other recent studies have applied multi-voxel pattern analyses (MVPA; Haynes and Rees 2006; Norman et al. 2006) of fMRI data to approach this problem. For example, Dinstein et al. (2008a) found that patterns of activity in aIPS could discriminate, within-modality, among three actions in either visual or motor modalities. However, patterns of activity elicited by viewing actions could not discriminate among performed actions (nor vice versa).

To summarize, neuroimaging studies to date using univariate methods do not provide clear evidence for a brain area (or areas) in which a common neural code represents actions across the visual and motor domains. Likewise, studies using adaptation or MVPA methods also have produced limited and conflicting evidence.

In the present study, in order to identify brain areas in which local patterns of brain activity could discriminate among these actions both within and across modalities, we used MVPA. Unlike the previous MVPA studies reviewed above, each participant's data were analyzed with a whole-cerebrum information mapping ("searchlight") approach (Kriegeskorte et al. 2006). Furthermore, in contrast to the volume-based approach used by most MVPA "searchlight" studies to date, we used surface-based reconstructions of the cortex. This approach improves both the classification accuracy and spatial specificity of the resulting information maps (Oosterhof et al., in press). In this way, we were able to map brain areas that carry crossmodal action representations, without restricting our analysis to pre-defined regions of interest, and in a way that respects cortical anatomy.

Participants were scanned with fMRI while performing and viewing different hand actions. In the first experiment, these were intransitive movements of the hand. Participants viewed a short movie of one of three actions, and then repeatedly either viewed or performed (with their own unseen hand) that action over the length of a block. The aim of this first experiment was to use a simple stimulus set in order to test our methods and to identify candidate visual/motor action representations. This was

followed by a second experiment, in which participants performed or viewed one of four manual actions directed at an object. In this event-related experiment, the actions defined a factorial design, in which either a lift or a punch goal was executed with either the whole hand or with the thumb and index finger. We adopted this design with two aims in mind: to encourage activity in the mirror system by testing actions with object-directed goals (Rizzolatti and Sinigaglia, 2010); and to identify regions in which the local pattern of activity more strongly represents action goals or action effectors.

Experiment 1

Methods

Subjects. Six right-handed, healthy adult volunteers (mean age 29; range = 24-35; 1 female, 5 male) 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 approved by the School of Psychology at Bangor University. Participation was compensated at £30.

Insert Figure 1 about here

Design and procedure. Participants watched short movies (1.5 seconds, 60 frames/second) of simple hand actions, and also performed these actions in the scanner. Supplementary Figure 1 shows the three actions used (labelled A, B, and C). The data were collected across two sessions per subject. There were seven conditions in the main experiment: do-A, do-B, do-C, see-A, see-B, see-C, and null (fixation) trials. Each trial (Figure 1) started with a 500 msec blank screen followed by a 500 msec black rectangle, signifying the beginning of a new trial. For the null trials, a black screen was presented for 24s. For the do and see trials, one of the three actions (A, B, or C) was shown once, followed by an instruction on the screen ("see" or "do") for 2s. After an interval (3.5s), the movie was either repeated eight times ("see" condition), or the participant performed the action eight times ("do" condition). To match the "see" and "do" conditions temporally, a pulsating fixation dot was presented in the middle of the screen during the "do" trials. This fixation dot was presented from 8 until 24 seconds after trial onset and repeatedly changed size with a phase of 2 seconds (large for 1.5 s, followed by small for 0.5 s). Participants were instructed to execute the hand movements in time with the dot. Participants were not able to see their own hand movements while in the scanner.

Each participant was scanned during two sessions, with 8 functional runs per session. Within each of the two sessions, participants were scanned on two sets of four runs, each one preceded by an anatomical scan. Each run started and ended with a 16 s fixation period. The first trial in each run was a repeat of the last trial in the previous run (in runs 1 and 5, it was a repeat of the last trial of runs 4 and 8, respectively) and was not of interest (i.e. regressed out in the analysis; see below). There were 14, 13, 13, and 13 remaining trials of interest (49 in total) for runs 1-4

(respectively), and similarly for runs 5-8. For each set of four runs, the seven conditions were assigned randomly with the constraints that (1) each of the seven trial conditions was preceded by each of the seven trial conditions exactly once, and (2) each condition was present in each of the four runs at least once. Participants completed 16 runs with (in total) 2x2x7x6=168 "do" and "see" trials of interest, that is 28 trials for each action with each task.

To ensure that the actions were executed correctly, participants completed a practice run of the experiment before going in the scanner. They were specifically instructed not to move during "see" and null trials, and to move only their hand and arm during "do" trials. They were told during training to use the viewed actions as a model and to match these as closely as possible during their own performance. Furthermore, we used an MR-compatible video camera (MRC Systems, Heidelberg, Germany) to record participants' hands throughout the scanning session to verify that the actions were carried out correctly and that no movements were executed in the "see" condition and null trials, or during the first 8s of a trial.

Data acquisition. The data were acquired using a 3T Philips MRI scanner with a SENSE phased-array head coil. For functional imaging, a single shot echo planar imaging sequence was used (T2*-weighted, gradient echo sequence; TR=2000 ms, TE=35 ms; flip angle 90°) to achieve near-whole cerebrum coverage. The scanning parameters were as follows: repetition time 2000 ms; 30 off-axial slices; slice pixel dimensions 2 x 2 mm^2; slice thickness 3 mm, no slice gap, FOV 224 x 224 mm^2, matrix 112 x 112, phase encoding direction A-P, SENSE factor = 2. For participants with large brains, where the entire cerebrum could not be covered, we gave priority to covering the superior cortex (including the entire primary motor and somatosensory areas and parietal cortex) at the expense of the inferior cortex (mainly temporal pole). The frontal lobes were covered in all participants. Seven dummy volumes were acquired before each functional run to reduce possible effects of T1 saturation. Parameters for T1-weighted anatomical scans were: 288 x 232 matrix; 1 mm^3 isotropic voxels; TR=8.4 ms, TE= 3.8 ms; flip angle = 8°.

Volume preprocessing. Using AFNI (Cox 1996), for each participant and each functional run separately, data was despiked (using AFNI's 3dDespike with default settings), time-slice corrected, and motion corrected (relative to the "reference volume": the first volume of the first functional run) with trilinear interpolation. The percent signal change was computed by dividing each voxel's time-course signal by

200 the mean signal over the run and multiplying the result by a hundred. The four
201 anatomical volumes were aligned with 3dAllineate, averaged, and aligned to the
202 reference volume (Saad et al. 2009).
203 Although we took measures to limit motion-related artifacts including data
204 "spikes" (e.g. by using short-trajectory hand movements, as far from the head as
205 possible) it is very likely that there were more movement artifacts in the "do" than
206 "see" trials. However, one benefit of the crossmodal analyses on which we focus our
207 attention is that such incidental uncontrolled differences between "see" and "do" trials
208 can only work against our hypothesis. That is, they will tend to reduce the similarity
209 between activity patterns elicited in the "see" and "do" conditions, and hence make it
210 more difficult for a classifier to discriminate among actions crossmodally.
211 Univariate volume analyses. A General Linear Model analysis was performed
212 using the AFNI 3dDeconvolve program in order to estimate the BOLD responses for
213 each do and see action trial (16 s each). Beta coefficients were estimated separately
214 for each of the do and see action trials by convolving a boxcar function (16 s on,
215 starting 8 s after trial onset) with the canonical hemodynamic response function
216 (HRF). The beta coefficients from the first trial in each run were not of interest (see
217 above), while beta coefficients from the other trials were used in the multi-voxel
218 pattern analysis (MVPA; see below). For each run, predictors of no interest were
219 included to regress out potential effects from the instruction part from each trial, also
220 by convolving a boxcar function (3.5 s on, starting 1.0 s after trial onset) with the
221 canonical HRF. To remove low frequency trends, predictors of no interest for
222 constant, linear, quadratic, and cubic trends were included in the model as well.
223 Surface preprocessing. For each participant and hemisphere, anatomical
224 surface meshes of the pial-grey matter ("pial") and smoothed grey matter-white
225 matter ("white") boundaries were reconstructed using Freesurfer (Fischl et al. 2001),
226 and these were used to generate an inflated and a spherical surface. Based on surface
227 curvature, the spherical surfaces of all participants were aligned to a standard
228 spherical surface (Fischl et al. 1999). Using AFNI's MapIcosehedron, these spherical
229 surfaces were resampled to a standardized topology (an icosehedron in which each of
230 the twenty triangles is subdivided into 10,000 triangles), and the pial, white, and
231 inflated surfaces were then converted to the same topology. This ensured that each
232 node on the standardized surfaces represented a corresponding surface location across
233 participants; therefore, group analyses could be conducted using a node-by-node

analysis. The affine transformation from Freesurfer's anatomical volume to the aligned anatomical volume was estimated (using AFNI's 3dAllineate) and applied to the coordinates of the standardized pial and white surfaces to align them with the reference volume.

For each participant, we also estimated the required affine transformation to bring the anatomical volume into Talairach space (Talairach and Tournoux 1988), and applied this transformation to the surfaces. The pial and white surfaces in Talairach space were averaged to construct an intermediate surface, that was used to measure distances (described below) and surface areas in a manner that was unbiased to the participants brain size. To limit our analysis to the cortex, and to improve statistical power when correcting for multiple comparisons, an exclusion mask covering the subcortical medial structures was drawn on the group map. This mask was subsequently used in the searchlight analyses.

Intra-participant surface-based "searchlight" multi-voxel pattern analyses. To investigate which regions represent information about which of the three actions (A, B, and C) was perceived or performed, we combined a searchlight (Kriegeskorte et al. 2006) with multi-voxel pattern analysis (MVPA; Haynes and Rees 2006; Norman et al. 2006) implemented in Matlab® (the Mathworks Ltd., Cambridge, UK) using a geodesic distance metric on the surface meshes (see Figure 2). For each participant and hemisphere, in the intermediate surface a "center node" was chosen and all nodes within a 12 mm radius circle on the surface (using a geodesic distance metric; Kimmel and Sethian 1998) were selected using the Fast Marching Toolbox (Peyre 2008). For each selected node on the intermediate surface, a line was constructed that connected the corresponding nodes on the standardized pial and white surfaces, and on each line, ten equidistant points were constructed. The searchlight contained all voxels that intersected at least one point from at least one line.

Insert Figure 2 about here

Each selected voxel in the searchlight was associated with 168 beta estimates, one from the final 16 s of each "do" or "see" trial of interest. These beta estimates were partitioned into 56 chunks (2 modalities x 28 occurrences of each action), so that each chunk contained three beta estimates of action A, B and C in that modality. To account for possible main effect differences between modalities or specific trials, for

each voxel and chunk separately, the three beta esimates were centered by subtracting the mean of three beta estimates.

Based on these centered responses, a multiclass Linear Discriminant Analysis classifier was used to classify trials using 28-fold cross validation. Because typically the number of voxels in selected regions was larger than the number of beta-estimates from the GLM, the estimate of the covariance matrix is rank deficient. We therefore regularized the matrix by adding the identity matrix scaled by one percent of the mean of the diagonal elements. For each of the two modalities, the classifier was trained on the beta estimates from 27 chunks in that modality, and tested on the remaining chunk in the same modality (unimodal classification), and also tested on the corresponding chunk in the other modality (crossmodal classification). This procedure was repeated for all 28 chunks.

For each of the four combinations of train and test modality (train ("do","see") x test ("do","see")), raw accuracies were computed by dividing the number of correctly classified trials by the total number of trials. For statistical inference in the group analysis (see below), raw accuracies were converted to z-scores based on their binomial distribution under the null-hypothesis of chance accuracy (1/3). For the crossmodal classification, accuracies from the two cross-modal classifications (train on "see", test on "do"; and vice versa) were combined before computing the z-score. This procedure was repeated for all of the 100,002 nodes in the intermediate surface. That is, each node was taken as the center of a circle and classification accuracy was computed using the surrounding nodes within the selection radius.

Surface-based group analysis. A random effects analysis was used to find regions where classification accuracy was above chance, by applying (for each node) a t-test against the null hypothesis of zero mean of the accuracy z-score (i.e. classification accuracy at chance level) and applying a node-wise threshold of p=0.05 (two-tailed).To find clusters that were significant while correcting for multiple comparisons, we employed a bootstrap procedure (Nichols and Hayasaka 2003). For a single bootstrap sample, we took six individual participant maps randomly (sampled with replacement). For each of the six maps, the sign of the z-score was negated randomly with probability of 50%, which is allowed under the null hypothesis of chance accuracy (z-score of 0). We note that the data in the bootstrap sample is unbiased with respect to the spatial autocorrelation structure in the original group map. A t-test was conducted on the resulting six maps and the resulting map was

clustered with the same threshold as the original data. This procedure was repeated a hundred times (i.e. we took a hundred bootstrap samples), and for each bootstrap sample the maximum cluster extent (in mm^2) across the surface was computed, yielding a distribution of maximum cluster extent values under the null hypothesis of chance accuracy. For each cluster in the original group results map, the α-level (significance) was set at the number of times that the maximum cluster extent value across bootstrap samples was larger than the observed cluster extent, divided by the number of bootstrap samples (100). Clusters are only reported for which $\alpha \leq 0.05$. For each cluster, its center-of-mass coordinates were computed by taking the average coordinates of its nodes, relatively weighted by each node's area.

Results

The crossmodal information map revealed significant clusters of crossmodal information about intransitive actions in and around the junction of the left intraparietal and postcentral sulci, and also in the lateral occipitotemporal cortex bilaterally (Figure 3a; Table 1). Two smaller below-chance clusters were also found, possibly due to the small number of subjects tested. For reference, in Figure 3b we show unthresholded t-maps, and in Figures 3c and 3d we present the data in terms of mean raw accuracy (chance = 33.3%). We found approximately equivalent crossmodal information when classifiers trained with "see" data and tested with "do" data, and *vice versa*, were tested separately (Supplementary Figure 2).

Insert Figure 3 about here

Insert Table 1 about here

For the unimodal information maps, we found that both for observing and for performing actions, large areas in the brain contained distributed above-chance information about which action was seen or performed. The highest classification accuracies were found in the expected visual and motor regions for "see" and "do" trials respectively (Supplementary Figure 3).

Discussion

Patterns of BOLD activity in the left anterior parietal cortex, and in lateral occipitotemporal cortex bilaterally, carry information that can discriminate among meaningless intransitive actions across the visual and motor domains. These findings suggest that in these areas the distinguishing properties of actions are represented in a distributed neural code, and that at least some aspects of this code are crossmodal. That is, some features of the patterns that code the actions must be common across the visual and motor modalities. Because the actions were meaningless and intransitive, it is unlikely that these codes reflect action semantics, and the results of Experiment 1 could not have been driven by the features of a target object (cf. Lingnau et al. 2009).

The property of representing intransitive actions in a common vision/action code may be functional in its own right, e.g. to support the learning of movements by observation alone. Aside from explicit, intentional learning, there are several demonstrations of what might be called social "contamination" effects – e.g. situations in which an observer spontaneously adopts the postures or movements of another individual. These automatic mirroring responses appear to facilitate social interactions and social bonding (Chartrand and Bargh 1999; Van Baaren et al. 2003) and may mediate interactive or collaborative actions. Additionally, crossmodal intransitive representations may contribute to the understanding of object-directed actions, for which the underlying movements may themselves be key elements.

Our analyses of unimodal information identified widespread areas that carried weak but significantly above-chance information about either which action was viewed or was performed. Importantly, in contrast to the critical crossmodal test, in the unimodal analyses the stimulus (or motor act) was essentially identical across training and test data sets. In such situations, MVPA can be a highly sensitive method, potentially making use of many sources of congruency between the neural events elicited by repeated instances of a given stimulus (and not necessarily the sources of interest to the investigators) such as commonalities in motion (Kamitani and Tong 2006; Serences and Boynton 2007), thoughts (Stokes et al. 2009), intentions (Haynes et al. 2007) or stimulus orientation (Kamitani and Tong 2005). This means that in general, proper interpretation of an informative brain region requires control conditions that test to what extent representations generalize. In the present study, this is much less a concern in the crossmodal conditions, given the great differences at the

sensory / motor level between seeing an action and performing that action out of view.

Because of the novelty of our methods and of some of the findings (e.g. crossmodal action information in lateral occipitotemporal cortex) we set out to replicate and extend the results of Experiment 1 before attempting to interpret them. First, in order to extend our findings to goal-directed behaviours, in Experiment 2 we tested transitive actions. It has been proposed that the "mirror" system is more effectively engaged by object-directed actions (e.g., Rizzolatti et al. 1996a) and we speculated that testing such actions could increase the recruitment of ventral premotor cortex. Second, we adopted an event-related design. Although such a design carries the risk of reducing statistical power, we reasoned that it would greatly increase participants' engagement in the task (compared to Experiment 1) by requiring more frequent attention to task cues and more frequent switching between conditions. Third, we tested more participants, which increases statistical power in the random effects and bootstrap analyses. Finally, we introduced a monitoring task in the "see" conditions, which required participants to attend actively to the viewed hand movements, as compared to passive viewing, as in Experiment 1.

Beyond these largely methodological improvements, we introduced new variables to the design of Experiment 2. We orthogonally varied two aspects of the actions that were viewed and performed by participants. One factor concerned the *effectors* used to make contact with the object during action execution. Half of the actions involved the tips of the thumb and index finger, while the other half involved the whole hand. Orthogonally, we manipulated action *goals*. Half of the actions involved grasping and lifting an object on to a platform in front of the participant. The other half of the actions required the participant to "punch" the side of the object so that it leaned away from the participant before returning to the upright position. By virtue of this factorial manipulation, we were able to test not only for brain regions in which patterns carried crossmodal visuo-motor action representations, but also to further test the nature of these representations (cf. similar efforts in extrastiate cortex, e.g. Aguirre 2007; Haushofer et al. 2008; Op de Beeck et al. 2008). Specifically, we tested whether a given area carries relatively more (crossmodal) information about the effector used to manipulate the object or about the goal of actions on the object.

In Figure 4, we illustrate a simple scheme for thinking about how patterns of cortical activity relate to different types of informational content in a given region.

The scheme centres on assessing the similarity of patterns elicited by particular combinations of seen and performed actions in Experiment 2. (We note that the matrices in Figures 4a-c are congruent with how accuracies were computed in Experiment 1, but with three actions instead of four). Each row and each column (for training set and test set, respectively) represents one of the eight conditions in the experiment, formed by the combination of modality (see, do) x effector (finger, hand) x goal (lift, punch). Where fMRI activity patterns are predicted to be similar (across training and test set, for a given brain region and a given participant), a cell matrix is marked with a pink square. Conversely, trials that were used in the cross-validation scheme but where no similarity between patterns is predicted, are indicated with a grey square. Different matrix arrangements illustrate predicted similarity patterns for within-modality representations (Figure 4a,b), for a visual / motor crossmodal representation (Figure 4c), and for representations biased in favour of either action effectors or goals (Figures 4d,e,f).

Insert Figure 4 about here

Experiment 2

Methods

Subjects. 11 right-handed, healthy adult volunteers 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 approved by the School of Psychology at Bangor University. Participation was compensated at £20.

Design and Procedure. Participants either performed or watched object-directed actions in the scanner (Figure 5). 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 (Figure 5a,b). Earphones delivered auditory instructions to the participants, in the form of words spoken by Apple Mac OS X 10.5 text-to-speech utility "say" using the voice of "Alex". Participants could see the table and the object through a forward-looking mirror mounted on the scanner coil. An experimenter of the same gender as the participant (AJW or NNO) was present in the scanner room to perform real-time actions on the object, which were then observed by the participant through the mirror. Visual instructions for the experimenter were projected on a wall in the scanner room, invisible to the participant.

Insert Figure 5 about here

The action instructions varied orthogonally on the effector used ("finger" for thumb and index finger, or "hand" for the whole hand) and on the goal of the action ("lift" to raise the object, or "punch" to push the object on its side). Thus, the experimental design was 2 (modality: "do" vs. "see") x 2 (effector: "finger" vs. "hand" x 2 (goal: "lift" vs. "punch"). Figure 5c shows the four actions, from the approximate perspective of the participant while executed by the experimenter.

Insert Figure 6 about here

There were nine conditions in the main experiment: eight for which an action was seen or performed, and one null (no action) condition. Each trial (Figure 6)

started with an auditory instruction “close” (for “do” and null trials), or “open” (for “see” trials). Participants were instructed to open or close their eyes according to the instruction, and compliance was monitored using a scanner-compatible eye tracking system. Simultaneously, a visual instruction was given to the experimenter to indicate whether or not (s)he should perform an action. Two seconds after trial onset, for “do” trials, another auditory instruction was given to the participant to indicate the specific action to be executed, in the order goal-effector (e.g. “lift finger”, “punch hand”). For “see” trials, no auditory instruction was given to the participant, but they had to monitor the action executed by the experimenter. To ensure the attention of the participant during these trials, occasionally (twice per run, on average) the experimenter repeated the action twice in rapid succession (“catch trial”), and participants were instructed to knock on the table to indicate that they had observed such a repeat. For both “do” and “see” trials, the names of the action goal and effector were presented visually to the experimenter: for “do” trials, so that (s)he could verify that the participant executed the correct action, and for “see” trials, so that (s)he knew which action to execute. Each trial lasted for seven seconds.

Each participant was scanned during a single session with eight functional (F) runs and three anatomical (A) scans, in the order AFFFAFFFAFF. For two participants, only six functional runs could be acquired due to participant discomfort and technical difficulties with the table-object attachment, respectively. First order counterbalancing was achieved by partitioning the functional runs in (three or four) sets of two runs each. For each set of two runs, the order of the conditions was randomly assigned with the constraints that (1) each of the nine conditions was preceded by each of the nine conditions exactly once, and (2) each condition was present in each of the two runs four or five times. To reinstate potential carryover effects from one trial the next at run boundaries, the first four and last four trials in a run were a repeat of the last four and first four trials, respectively, of the other run in the same set. The first two and last two trials in each run, trials during which participants executed the wrong action, and catch trials were all marked as trials of no interest and modelled separately in the General Linear Model (see below). The first trial started two seconds after the beginning of the run.

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 leave enough space in between their hand and the object so

that the experimenter could execute the actions on the object without touching their hand; to keep their left hand and arm under the table, out of view; and after a "close" instruction, to keep their eyes closed until they were instructed to open them again. To ensure that participants followed the instructions correctly, they completed two practice runs of the experiment: the first before going in the scanner, the second in the scanner during the first anatomical scan. Participants were told during training to use the viewed actions as a model and to match these as closely as possible during their own performance.

Data acquisition. The data were acquired as in Experiment 1, with a variation in some of the scanning parameters for functional imaging: repetition time 2500 ms; 40 off-axial slices; 2.5 mm^3 isotropic voxels, no slice gap, FOV 240 x 240 mm^2, matrix 96 x 96.

Univariate volume analyses. Volume preprocessing was identical to Experiment 1, and univariate analyses very similar to Experiment 1 except for the following. For each run separately, eight beta coefficients of interest (corresponding to the four "do" and four "see" action conditions) were estimated with a General Linear Model by convolving a boxcar function (3 s on, starting 2 s after trial onset) with the canonical hemodynamic response function (HRF). Each trial of no interest (see above) was regressed out with a separate regressor of the same shape. To remove low frequency trends, predictors of no interest for constant, linear, quadratic, and cubic trends were included in the model as well.

Intra-participant surface-based "searchlight" multi-voxel pattern analyses. Before MVPA, surfaces were preprocessed as in Experiment 1. Surface-based MVPA was also performed similarly to Experiment 1, with the only difference that the beta estimates were partitioned in two chunks per run corresponding to the two modalities ("do" and "see"), so that cross-validation was 8-fold for both unimodal and crossmodal classification. In other words, data from one run was used to test the classifier, while data from the other runs was used to train it. Based on the matrices in Figure 4, accuracies were computed as follows. Trials for which the combination of corresponding (training and test) condition in the matrix was coloured red were considered as correctly classified; those for which this combination was marked (red or grey) were counted to yield the total number of trials. Raw accuracy and accuracy z-scores were computed as Experiment 1, while taking into account the chance level (¼ or ½, depending on the contrast: the number of red squares divided by the number

of marked squares in each column). Accuracy z-scores for the "effector" vs. "goal" contrast (Figure 4f) were the node-wise difference of accuracy z-scores for "effector" and "goal" (Figures 4d and 4e). Surface-based group-level analyses were carried out as in Experiment 1.

Results

In Experiment 2, we identified significant clusters of crossmodal action information in the left hemisphere, in and around the anterior parietal cortex including the postcentral gyrus. We also observed clusters bilaterally in the lateral occipitotemporal cortex (Figure 7; Table 2). This result was similar when the two train-test directions (train with "see" data, test with "do" data, and vice versa) were examined separately (Supplementary Figure 4). Unlike Experiment 1, however, the unimodal "do" but not the "see" analysis revealed areas carrying within-modality information about the actions (Supplementary Figures 5 and 6).

Insert Figure 7 about here

Insert Table 2 about here

In order to identify regions in which the crossmodal information content was biased either for action goals or for effectors, we first applied a mask to include only locations for which crossmodal information, averaged across both train-test directions, was significant (as in Figure 7). Each remaining vertex was coloured (Figure 8) according to whether it showed stronger discrimination of: effectors (blue, cyan); goals (red, yellow); or no bias (green). In the left hemisphere parietal and postcentral gyrus regions, this map revealed a gradient of biases in crossmodal action information. Specifically, posteriorly, similarity patterns favoured the distinction between action *goals* over *effectors*. That is, the patterns for lift and punch goals were less similar to each other relative to the patterns for finger and whole hand actions. In contrast, moving anteriorly towards the precentral gyrus, activation patterns favoured the representation of effectors. Finally, in the lateral occipitotemporal clusters, the representations appeared to show no strong bias. Supplementary Figure 7 provides maps showing separately areas that are biased for either the representation of goals or of effectors.

Insert Figure 8 about here

Discussion

The main results of Experiment 2 were highly similar to those of Experiment 1, in spite of several changes to the experimental task, design, and stimuli. (Note however that these differences preclude direct statistical comparisons of the two experiments). We were able to achieve these results with MVPA in spite of the reduced statistical power provided by an event-related design (which may nonetheless have improved the psychological validity of the task). Our principal finding was that patterns of activity across the dorsal and anterior parietal cortex, postcentral gyrus, and lateral occipitotemporal cortex carry significant crossmodal information about transitive actions. The lateral occipitotemporal regions were significant in both hemispheres in both studies, suggesting a crossmodal action representation that is perhaps not tied to the laterality of the specific limb used to perform the task. In contrast, the parietal/postcentral clusters were largely confined to the left hemisphere. It may be that the action representations identified here are specific to the hand that was used to perform the actions, rather than being abstracted across the midline. However, previous reports have identified left-lateralized activity in response to the planning and execution of goal-directed actions performed by either the left or right hand (e.g. Johnson-Frey et al., 2005). Further tests comparing left- and right-handed actions will be needed to resolve questions about the laterality of the regions identified here.

The other significant finding of Experiment 2 is that we were able to identify a gradient of information content extending across the anterior parietal cortex and the postcentral gyrus. This was achieved by using a factorial design that independently varied the effector and the goal of the actions that were performed and observed. At the posterior edge of this gradient, patterns of fMRI activity showed more information about the goals of the action (lift vs punch), while towards the anterior edge, into the postcentral gyrus, the bias shifted to favour the effector used to execute this action (finger vs hand). Note that this pattern was observed for crossmodal analyses testing the similarity of patterns across vision and action. Generally this bias is consistent with previous conceptions of the postcentral gyrus as consisting of somatosensory representations (closely tied to the body surface), while anterior parietal areas represent actions in terms of more abstract hand-object interactions such as different

forms of grasp to achieve specific goals. More specifically the aIPS region in particular has been implicated in object-directed grasp as opposed to reach (e.g. Culham et al., 2003, Frey et al., 2005), comparable to the "lift" vs "punch" distinction tested here. This finding shows that the techniques devised here have the potential to reveal not only regions in which actions are coded similarly across the visual and motor domains, but also to reveal more detailed information about these representations.

General Discussion

The present results succeed in the aim to use fMRI to identify human brain regions that construct, at the population level, representations of actions that cross the visual and motor modalities. Specifically, we show that the distributed neural activity in the regions identified here encodes both seen and performed actions in a way that is at least partially unique for different actions. Thus these broad codes share an essential property of macaque mirror cells – although given the grossly different measures employed, any comparison between the present findings and mirror neurons can only be at an abstract level.

Although the nature of the MVPA technique prevents pinpointing the anatomical source of the crossmodal information with great precision, previous findings shed some light on the neural representations that are likely to underlie the crossmodal clusters identified here. The left lateral occipitotemporal region has long been implicated in the understanding of action (Martin et al. 1996). Also, the clusters identified here fall close to a number of functionally-defined brain regions that are found bilaterally, including: the dorsal/posterior focus of the lateral occipital complex (LO; Grill-Spector et al. 1999) which is involved in visual object perception; the body-selective extrastriate body area (EBA; Downing et al., 2001; Peelen and Downing 2007a); and motion-selective areas including proposed human homologues of MT (Tootell et al. 1995) and MST (Huk et al. 2002). Accordingly it is difficult to assess which of these neural populations, if any, may contribute to the crossmodal information identified here. For example, area MST, which responds to both visual motion and tactile stimulation (Beauchamp et al. 2007), may carry neural responses that are crossmodally informative about actions. Further, EBA has been proposed to have a role in the guidance of unseen motor behaviour and even to play a part in the human mirror "network", and hence might play a crossmodal role in action representation (Astafiev et al. 2004; Jackson et al. 2006; but see Candidi et al., 2008; Kontaris et al. 2009; Peelen and Downing 2005; Urgesi et al., 2007).

Many findings converge on the idea that the parietal cortex generally codes aspects both of the position of the body and its movements, and of visual information, particularly regarding stimuli that are the targets of action. In human neuroimaging studies, activations in the general region of aIPS have frequently been identified in tasks involving either executing or observing human actions, typically those that are object-directed (Tunik et al. 2007; Van Overwalle and Baetens 2009). Evidence of

this kind has led some researchers to the conclusion that this region is part of a human mirror system, although recent work with adaptation and MVPA methods has not supported this hypothesis (Dinstein et al. 2007, 2008a). The present results provide positive evidence for anterior parietal cortex carrying a genuinely crossmodal action code.

The left parietal crossmodal clusters extend substantially into the postcentral gyrus, implicating a role for somatosensory representations in the visual/motor representation of actions. This pattern was especially apparent in Experiment 2, which (unlike Experiment 1) required finely controlled actions as the hand interacted with the object in different ways. Previous work has shown somatosensory activation by seeing others reach for and manipulate objects (e.g. Avikainen et al. 2002; Cunnington et al. 2006), as well as during passive touch (e.g., Keysers et al. 2004). The role of somatosensation in representing sensory aspects during haptic object exploration (e.g., Miquee et al. 2008) suggests its role in action simulation during observation is based on the sensory-tactile aspects of skin-object interactions (e.g., Gazzola and Keysers 2009; see also Keysers et al., 2010).

In Experiment 2 we tested the hypothesis that meaningful, object-directed actions would be more effective than intransitive actions in engaging the ventral premotor cortex (PMv), as found in previous single-unit studies of the macaque and in univariate fMRI studies of the human (e.g., Rizzolatti et al. 1996a, b, 2001). This hypothesis was not confirmed, and indeed in neither experiment did we find significant crossmodal information in PMv. Previous evidence for common coding of vision and action in human PMv was based on overlapping activations in univariate analyses, and as noted above, this could be due to separate but overlapping neural codes for visual and motor action properties in the same brain region.

On its face, however, that argument is not consistent with the findings of Kilner et al. (2009), who found adaptation in PMv from vision to action and vice versa. Note, however, that the visual stimuli in Kilner et al (2009) were depicted from an egocentric view that matched the participant's own viewpoint, rather than the typical view seen of another person's actions. In contrast, in our study the visual stimuli were clearly views of another person's actions. Further studies should test whether MVPA approaches detect crossmodal action information in PMv when the visually-presented actions are seen egocentrically (and also whether adaptation effects are found when actions are presented allocentrically). If MVPA and adaptation effects

664 in PMv are found only for egocentric views, this would limit the proposed homology
665 between BOLD activity in this region in humans and single-cell findings in the
666 macaque.
667 Setting aside the above considerations, it could of course be the case that
668 crossmodal visual/motor action properties are represented jointly in human PMv from
669 any viewing perspective, but on a spatial scale that is not well matched by the
670 combination of imaging resolution and MVPA methods adopted here (cf. Swisher et
671 al., 2010). It is difficult to draw conclusions from a null effect, and we do not take the
672 absence of significant clusters in PMv (and other) regions in the present study as
673 strong evidence against the presence of crossmodal visual-motor representations in
674 those regions.
675 As reviewed in the Introduction, recent evidence on visual/motor action
676 representations from repetition-suppression methods is mixed. One possible
677 hypothesis is that the relevant neural populations may not adapt in the same way as do
678 neurons in other regions such as visual cortex. Previous single-cell studies in
679 macaques support this proposal. For example, Leinonen et al. (1979) measured neural
680 activity in aIPS, noting that "Cells that responded to palpation or joint movement
681 showed no marked habituation on repetitive stimulation". Similarly, Gallese et al.
682 (1996) mentioned that for mirror neurons in frontal area F5, "[t]he visual stimuli most
683 effective in triggering mirror neurons were actions in which the experimenter's hand
684 or mouth interacted with objects. The responses evoked by these stimuli were highly
685 consistent and did not habituate".
686 However, several imaging adaptation studies have shown within-modality
687 adaptation effects, and/or unidirectional cross-modal adaptation (Chong et al. 2008;
688 Hamilton and Grafton 2006). In some cases (e.g. Chong et al. 2008), this could reflect
689 adaptation of semantic representations instead of (or in addition to) visuo-motor
690 representations, although in other paradigms this possibility can be ruled out (Lingnau
691 et al. 2009). Most recently, as noted above, Kilner et al. (2009) have shown fully
692 crossmodal adaptation effects.
693 A potentially important consideration is that the repetition suppression studies
694 to date have focused on short-term repetition, which relates in uncertain and
695 potentially complex ways to single-unit spiking activity (Sawamura et al. 2006) and to
696 long-term priming (Epstein et al. 2008). This emphasis on the short-term changes in
697 activity resulting from repetition stands in contrast to the present approach of

identifying those aspects of activation patterns that remain constant over relatively long time scales on the order of tens of minutes. Clearly, further studies will need to directly compare MVPA and adaptation measures (both short-term and long-term) of crossmodal action representations.

As noted above, there have been previous attempts to identify crossmodal visual/motor action representations with MVPA, most notably by Dinstein et al. (2008a). That study used an event-related fMRI paradigm and a "rock-paper-scissors" task, in which participants freely chose to perform one of three actions on each trial in a simulated competition against a computer opponent. MVPA revealed that activity in left and right aIPS could discriminate, within-modality, among both perceived and performed actions, but in contrast to the present findings this did not extend to the crossmodal case. While there are some similarities between Dinstein et al (2008a) and the present study that can be excluded as causing the divergent results (e.g. both used similar linear discriminant analysis classifiers; both tested hand movements), there are several differences between the approaches used. For one, Dinstein et al (2008a) used functionally-defined regions of interest and so may have missed areas that do not necessarily exhibit strong responses in the univariate sense (see below). Alternatively, task characteristics may be important. The "rock-paper-scissors" task has the advantage over other paradigms that participants freely choose their own actions to perform. However, in it, actions are also performed in a competitive context, which may alter or inhibit representations of the opponent's actions.

Our findings underscore the benefits of whole-brain analyses for MVPA. The use of standardized coordinates does not take into account inter-subject variability in the anatomical structure of the brain, while the using functional localisers to identify *a priori* regions of interest relies on the assumption that higher gross activation levels (e.g., for doing and seeing actions) in a region are a necessary condition for identifying representations of individual actions in that region. Our novel combination of surface reconstruction and information mapping (Oosterhof et al., in press) provides a data-driven map for the whole brain, featuring voxel selection and inter-participant alignment that respect cortical anatomy (Fischl et al. 1999). In this way we have identified areas of potential interest – specifically the lateral occipitotemporal cortex – that were not examined by previous studies of crossmodal visual/motor action representation.

Finally, one general issue that must be confronted is that of mental imagery. It is possible in principle that areas that appear to carry crossmodal vision/action information are actually unimodal, with the additional assumption that one type of task (e.g. performing actions) elicits imagery in another modality (e.g. visual imagery for actions) that is highly similar to a real-world percept (e.g. seeing actions performed). Indeed, studies that explicitly compare actual performance and imagined performance of actions do find overlapping areas of brain activity (e.g. in parietal cortex; Filimon et al., 2007; Lui et al., 2008). This issue is not only relevant to the present work but also to a wide range of previous studies on action perception / performance. Indeed it could apply still more generally across other studies of multimodal cognition: for example, brain areas active for reading words, or for hearing meaningful sounds, or for tactile perception of textures could all in principle reflect visual imagery for their referent objects. The present study does not resolve this question. One avenue for future research would be to adapt the methods used here to test for crossmodal action representations when visual action depictions are presented under conditions of divided attention (which would presumably make imagery more difficult) or even under subliminal conditions (which would make it impossible).

Conclusion

The present results open the way for future studies using MVPA to explore the neural "space" of action representation. Furthermore, the approach developed here could be adopted to test the boundaries of cross-modal action matching. For example, the preceding discussion raised a question about the extent to which the neural activity patterns elicited by observing actions is modulated by variations in viewpoint (cf. Vogt et al. 2003). Additionally, we can ask what role attention and task set play in the construction of crossmodal action representations (cf. Reddy et al., 2009; Esterman et al., 2009). Finally, combining transcranial magnetic stimulation with fMRI would open the possibility of disrupting information-bearing areas, such as those identified here, in order to assess the consequent effects on behaviour and on remote, interconnected brain regions.

Acknowledgments

This research was supported by the ESRC (grant to SPT and PED), the Wellcome Trust (grant to SPT), and the Wales Institute of Cognitive Neuroscience. NNO is supported by a fellowship awarded by the Boehringer Ingelheim Fonds. Thanks to Tobias Wiestler for help with MVPA, to Giuseppe di Pellegrino, Marius Peelen, Tobias Wiestler, and India Morrison for helpful discussions, to Marius Peelen for helpful comments on an earlier draft of this manuscript, and to Steve Johnston and Paul Mullins for technical support.

References

Agnew ZK, Bhakoo KK, Puri BK. The human mirror system: a motor resonance theory of mind-reading. *Brain Res Rev* 54: 286-293, 2007.

Aguirre GK. Continuous carry-over designs for fMRI. *Neuroimage* 35: 1480-1494, 2007.

Astafiev SV, Stanley CM, Shulman GL, Corbetta M. Extrastriate body area in human occipital cortex responds to the performance of motor actions. *Nat Neurosci* 7: 542-548, 2004.

Avikainen S, Forss N, Hari R. Modulated activation of the human SI and SII cortices during observation of hand actions. *Neuroimage* 15: 640-646, 2002.

Barsalou LW, Niedenthal PM, Barbey AK, Ruppert JA. Social embodiment. *Psychology of Learning and Motivation* 43: 43-92, 2003.

Beauchamp MS, Yasar NE, Kishan N, Ro T. Human MST but not MT responds to tactile stimulation. *J Neurosci* 27: 8261-8267, 2007.

Candidi M, Urgesi C, Ionta S, Aglioti SM. Virtual lesion of ventral premotor cortex impairs visual perception of biomechanically possible but not impossible actions. *Soc Neurosci* 3: 388-400, 2008.

Chartrand TL, Bargh JA. The chameleon effect: the perception-behavior link and social interaction. *J Pers Soc Psychol* 76: 893-910, 1999.

Chong TT, Cunnington R, Williams MA, Kanwisher N, Mattingley JB. FMRI adaptation reveals mirror neurons in human inferior parietal cortex. *Curr Biol* 18: 1576-1580, 2008.

Corballis MC. Mirror neurons and the evolution of language. *Brain Lang* 112:25-35, 2010.

Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29: 162-173, 1996.

Culham JC, Danckert SL, DeSouza JF, Gati JS, Menon RS, Goodale MA. Visually guided grasping produces fMRI activation in dorsal but not ventral stream brain areas. *Exp Brain Res* 153: 180-189, 2003.

Cunnington R, Windischberger C, Robinson S, Moser E. The selection of intended actions and the observation of others' actions: a time-resolved fMRI study. *Neuroimage* 29: 1294-1302, 2006.

di Pellegrino G, Fadiga L, Fogassi L, Gallese V, Rizzolatti G. Understanding motor events: a neurophysiological study. *Experimental brain research Experimentelle Hirnforschung Expérimentation cérébrale* 91: 176-180, 1992.

Dinstein I, Gardner JL, Jazayeri M, Heeger DJ. Executed and observed movements have different distributed representations in human aIPS. *J Neurosci* 28: 11231-11239, 2008a.

Dinstein I, Thomas C, Behrmann M, Heeger DJ. A mirror up to nature. *Curr Biol* 18: R13-8, 2008b.

Dinstein I, Hasson U, Rubin N, Heeger DJ. Brain areas selective for both observed and executed movements. *J Neurophysiol* 98: 1415-1427, 2007.

Downing PE, Jiang Y, Shuman M, Kanwisher NG. A cortical area selective for visual processing of the human body. *Science* 293: 2470-2473, 2001.

Epstein RA, Parker WE, Feiler AM. Two kinds of FMRI repetition suppression? Evidence for dissociable neural mechanisms. *J Neurophysiol* 99: 2877-2886, 2008.

Esterman M, Chiu YC, Tamber-Rosenau BJ, Yantis S. Decoding cognitive control in human parietal cortex. *Proc Natl Acad Sci U S A* 106: 17974-17979, 2009.

Etzel JA, Gazzola V, Keysers C. Testing simulation theory with cross-modal multivariate classification of fMRI data. *PLoS ONE* 3: e3690, 2008.

Filimon F, Nelson JD, Hagler DJ, Sereno MI. Human cortical representations for reaching: mirror neurons for execution, observation, and imagery. *Neuroimage* 37: 1315-1328, 2007.

Fischl B, Liu A, Dale AM. Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Trans Med Imaging* 20: 70-80, 2001.

Fischl B, Sereno MI, Tootell RB, Dale AM. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp* 8: 272-284, 1999.

Frey SH, Vinton D, Norlund R, Grafton ST. Cortical topography of human anterior intraparietal cortex active during visually guided grasping. *Brain Res Cogn Brain Res* 23: 397-405, 2005.

Gallese V, Fadiga L, Fogassi L, Rizzolatti G. Action recognition in the premotor cortex. *Brain* 119: 593-609, 1996.

Gazzola V, Keysers C. The observation and execution of actions share motor and somatosensory voxels in all tested subjects: single-subject analyses of unsmoothed fMRI data. *Cereb Cortex* 19: 1239-1255, 2009.

Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzchak Y, Malach R. Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24: 187-203, 1999.

Grill-Spector K, Malach R. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol (Amst)* 107: 293-321, 2001.

Hamilton AF, Grafton ST. Goal representation in human anterior intraparietal sulcus. *J Neurosci* 26: 1133-1137, 2006.

Haushofer J, Livingstone MS, Kanwisher NG. Multivariate patterns in object-selective cortex dissociate perceptual and physical shape similarity. *PLoS Biol* 6: e187, 2008.

Haynes JD, Rees G. Decoding mental states from brain activity in humans. *Nat Rev Neurosci* 7: 523-534, 2006.

Haynes JD, Sakai K, Rees G, Gilbert S, Frith C, Passingham RE. Reading hidden intentions in the human brain. *Curr Biol* 17: 323-328, 2007.

Huk AC, Dougherty RF, Heeger DJ. Retinotopy and functional subdivision of human areas MT and MST. *J Neurosci* 22: 7195-7205, 2002.

Iacoboni M, Woods RP, Brass M, Bekkering H, Mazziotta JC, Rizzolatti G. Cortical mechanisms of human imitation. *Science* 286: 2526-2528, 1999.

Jackson PL, Meltzoff AN, Decety J. Neural circuits involved in imitation and perspective-taking. *Neuroimage* 31: 429-439, 2006.

Johnson-Frey SH, Newman-Norlund R, Grafton ST. A distributed left hemisphere network active during planning of everyday tool use skills. *Cereb Cortex* 15: 681-695, 2005.

Kamitani Y, Tong F. Decoding the visual and subjective contents of the human brain. *Nat Neurosci* 8: 679-685, 2005.

Kamitani Y, Tong F. Decoding seen and attended motion directions from activity in the human visual cortex. *Curr Biol* 16: 1096-1102, 2006.

Keysers C, Kaas JH, Gazzola V. Somatosensation in social perception. *Nat Rev Neurosci* 11: 417-428, 2010.

Keysers C, Wicker B, Gazzola V, Anton JL, Fogassi L, Gallese V. A touching sight: SII/PV activation during the observation and experience of touch. *Neuron* 42: 335-346, 2004.

Kilner JM, Neal A, Weiskopf N, Friston KJ, Frith CD. Evidence of mirror neurons in human inferior frontal gyrus. *J Neurosci* 29: 10153-10159, 2009.

Kimmel R, Sethian JA. Computing geodesic paths on manifolds. *Proc Natl Acad Sci U S A* 95: 8431-8435, 1998.

Kontaris I, Wiggett AJ, Downing PE. Dissociation of extrastriate body and biological-motion selective areas by manipulation of visual-motor congruency. *Neuropsychologia* 47: 3118-3124, 2009.

Kriegeskorte N, Goebel R, Bandettini P. Information-based functional brain mapping. *Proc Natl Acad Sci U S A* 103: 3863-3868, 2006.

Leinonen L, Hyvarinen J, Nyman G, Linnankoski I. I. Functional properties of neurons in lateral part of associative area 7 in awake monkeys. *Exp Brain Res* 34: 299-320, 1979.

Lingnau A, Gesierich B, Caramazza A. Asymmetric fMRI adaptation reveals no evidence for mirror neurons in humans. *Proc Natl Acad Sci U S A* 106: 9925-9930, 2009.

Lui F, Buccino G, Duzzi D, Benuzzi F, Crisi G, Baraldi P, Nichelli P, Porro CA, Rizzolatti G. Neural substrates for observing and imagining non-object-directed actions. *Soc Neurosci* 3: 261-275, 2008.

Martin A, Wiggs CL, Ungerleider LG, Haxby JV. Neural correlates of category-specific knowledge. *Nature* 379: 649-652, 1996.

Miquee A, Xerri C, Rainville C, Anton JL, Nazarian B, Roth M, Zennou-Azogui Y. Neuronal substrates of haptic shape encoding and matching: a functional magnetic resonance imaging study. *Neuroscience* 152: 29-39, 2008.

Morrison I, Downing PE. Organization of felt and seen pain responses in anterior cingulate cortex. *Neuroimage* 37: 642-651, 2007.

Nichols T, Hayasaka S. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Stat Methods Med Res* 12: 419-446, 2003.

Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci* 10: 424-430, 2006.

Oosterhof N, Wiestler T, Downing P, Diedrichsen J. A comparison of volume-based and surface-based multi-voxel pattern analysis. *Neuroimage*, in press.

Op de Beeck HP, Torfs K, Wagemans J. Perceived shape similarity among unfamiliar objects and the organization of the human object vision pathway. *J Neurosci* 28: 10111-10123, 2008.

Peelen MV, Downing PE. Is the extrastriate body area involved in motor actions? *Nat Neurosci* 8: 125; author reply 125-125; author reply 126, 2005.

Peelen MV, Downing PE. The neural basis of visual body perception. *Nat Rev Neurosci* 8: 636-648, 2007a.

Peelen MV, Downing PE. Using multi-voxel pattern analysis of fMRI data to interpret overlapping functional activations. *Trends Cogn Sci* 11: 4-5, 2007b.

Peyre G. Toolbox fast marching - a toolbox for fast marching and level sets computations. Retrieved on 16 May 2009 from http://www.ceremade.dauphine.fr/ peyre/matlab/fast-marching/content.html.

Reddy L, Kanwisher NG, VanRullen R. Attention and biased competition in multi-voxel object representations. *Proc Natl Acad Sci U S A* 106: 21447-21452, 2009.

Rizzolatti G, Fabbri-Destro M. The mirror system and its role in social cognition. *Curr Opin Neurobiol* 18: 179-184, 2008.

Rizzolatti G, Fadiga L, Gallese V, Fogassi L. Premotor cortex and the recognition of motor actions. *Brain Res Cogn Brain Res* 3: 131-141, 1996a.

Rizzolatti G, Fadiga L, Matelli M, Bettinardi V, Paulesu E, Perani D, Fazio F. Localization of grasp representations in humans by PET: 1. Observation versus execution. *Exp Brain Res* 111: 246-252, 1996b.

Rizzolatti G, Fogassi L, Gallese V. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat Rev Neurosci* 2: 661-670, 2001.

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

Saad ZS, Glen DR, Chen G, Beauchamp MS, Desai R, Cox RW. A new method for improving functional-to-structural MRI alignment using local Pearson correlation. *Neuroimage* 44: 839-848, 2009.

Sawamura H, Orban GA, Vogels R. Selectivity of neuronal adaptation does not match response selectivity: a single-cell study of the FMRI adaptation paradigm. *Neuron* 49: 307-318, 2006.

Serences JT, Boynton GM. The representation of behavioral choice for motion in human visual cortex. *J Neurosci* 27: 12893-12899, 2007.

Stokes M, Thompson R, Cusack R, Duncan J. Top-down activation of shape-specific population codes in visual cortex during mental imagery. *J Neurosci* 29: 1565-1572, 2009.

Swisher JD, Gatenby JC, Gore JC, Wolfe BA, Moon CH, Kim SG, Tong F. Multiscale pattern analysis of orientation-selective activity in the primary visual cortex. *J Neurosci* 30: 325-330, 2010.

Talairach J, Tournoux P. *Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System - an Approach to Cerebral Imaging*. New York: Thieme Medical Publishers, 1988.

Tootell RB, Reppas JB, Kwong KK, Malach R, Born RT, Brady TJ, Rosen BR, Belliveau JW. Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J Neurosci* 15: 3215-3230, 1995.

Tunik E, Rice NJ, Hamilton A, Grafton ST. Beyond grasping: representation of action in human anterior intraparietal sulcus. *Neuroimage* 36 Suppl 2: T77-86, 2007.

Urgesi C, Candidi M, Ionta S, Aglioti SM. Representation of body identity and body actions in extrastriate body area and ventral premotor cortex. *Nat Neurosci* 10: 30-31, 2007.

Van Baaren RB, Holland RW, Steenaert B, van Knippenberg A. Mimicry for money: Behavioral consequences of imitation. *Journal of Experimental Social Psychology* 39: 393-398, 2003.

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

Vogt S, Taylor P, Hopkins B. Visuomotor priming by pictures of hand postures: perspective matters. *Neuropsychologia* 41: 941-951, 2003.

Figure Captions

1. Schematic illustration of the trial structure in Experiment 1. Each block began with a warning signal, followed by a 1.5 s movie showing one of three simple, intransitive manual actions. A task cue ("see" or "do") and a blank interval then followed. On "see" trials, the same movie was then presented eight times in succession, with a 0.5 sec blank interval between each movie presentation. On "do" trials, a central fixation dot grew larger for 1.5 sec and then shrank again for 0.5 sec, in a cycle that repeated eight times and that was matched to the cycle of movie presentations in the "see" condition. In the "do" condition, participants were required to perform the action that had appeared at the start of the block, in synchrony with the expansion of the fixation point.

2. Comparison of voxel selection methods in information mapping. (a) Schematic representation of a brain slice, with white matter, grey matter, and matter outside the brain indicated. The curved lines represent the white matter/grey matter boundary, the grey matter/pial surface boundary, and the skull. With the traditional volume-based voxel selection method for multivoxel pattern analysis, a voxel (blue) is taken as the center of a sphere (red; represented by a circle), and all voxels within the sphere are selected for further pattern analysis. (b) An improvement over (a), in that only grey matter voxels are selected. The grey matter can either be defined using a probability map, or using cortical surface reconstruction. A limitation however is that voxels close in Euclidian distance but far in geodesic distance (i.e. measured along the cortical surface) are included in the selection, as illustrated by the three voxels on the left. (c) Using surface reconstruction, the white matter-grey matter and grey matter-pial surfaces are averaged, resulting in an intermediate surface that is used to measure geodesic distances. A node on the intermediate surface (blue) is taken as the center of a circle (red; represented by a solid line), the corresponding circles on the white-grey matter and grey matter-pial surfaces are constructed (red dashed lines), and only voxels in between these two circles are selected.

3. Group crossmodal surface information map for Experiment 1, generated using multivoxel pattern analysis with an LDA classifier with training and test data from different ("see" vs. "do") modalities. **(a)** The coloured brain clusters (see Table 1)

indicate vertices where grey matter voxels within the surrounding circle on the cortical surface show above-chance crossmodal information (random effects analysis, thresholded for cluster size). Crossmodal visuo-motor information about intransitive manual actions is found in the left hemisphere at the junction of the intraparietal and postcentral sulci, and bilaterally in lateral occipitotemporal cortex. For each node this is based on two classifications, in which either the data from the "see" condition was used to train the classifier and the data from the "do" condition was used as test data, or vice versa. Insets: detailed view of the significant clusters. **(b)** The same map as (a), but without cluster thresholding. The color map legend (bottom left) shows the t-value of the group analysis against chance accuracy for panels (a) and (b). **(c)** As (a), except that mean classification accuracy values (chance = 33.3%) are depicted. **(d)** As (c), without cluster thresholding. The color map legend (bottom right) shows the accuracy scale for panels (c) and (d). Abbreviations: CS, central sulcus; PoCS, post-central sulcus; IPS, intraparietal sulcus; STS, superior temporal sulcus.

4. Similarity matrices for evaluation of Experiment 2 cross-validation classification results. Each row and each column (for training set and test set, respectively) represents one of the eight conditions in the experiment, formed by the combination of modality (see, do) x effector (finger, hand) x goal (lift, punch). Where fMRI activity patterns are predicted to be similar (across training and test set, for a given brain region and a given participant), a cell matrix is marked with a pink square. Conversely, trials that were used in the cross-validation scheme but where no similarity between patterns is predicted, are indicated with a grey square. **(a)** This example represents predicted similarity for within-modality "do" action represention. The fMRI activity patterns elicited by performing a given action are predicted to be similar across multiple executions of that action, compared to a different action. **(b,c)** Similarity matrices for within-modality "see" and cross-modal action representation. In the crossmodal case (c), the prediction is that the fMRI activity pattern elicited by performing a given action will be similar to that elicited by seeing that action (relative to other actions), and *vice versa*. **(d,e)**. Similarity matrices for representation of goal irrespective of effector, and vice versa. Note that both cases reflect information carried across modalities. **(f)** Similarity matrix for the contrast of goal vs. effector, where blue squares indicate similarity of patterns, but with a negative weight. Note that this matrix is the difference between the matrices in (d) and (e). Also note that the

matrices in (a-c) are equally applicable to Experiment 1, but with three actions in each modality instead of four.

5. Experimental stimuli from Experiment 2. (**a**) Frame capture from video recording during Experiment 2, showing the position of the participant's hand, experimenter's hand, and the target object during a null (no action) trial. **(b)** Similar to (a), but the experimenter performs a "punch hand" action that is observed by the participant. (**c**) Frames illustrating each of the four actions used in the experiment, formed by crossing effector (finger, hand) with goal (lift, punch).

6. Schematic of the trial structure for Experiment 2. The top row shows the series of events in "see" trials, and the bottom row events in "do" trials.

7. Group crossmodal surface information map for Experiment 2. **(a)** Cluster-thresholded map (conventions as in Figure 3) of crossmodal visuo-motor information about transitive manual actions is found in the left hemisphere, around the junction of the intraparietal and postcentral sulci, and in lateral occipitotemporal cortex bilaterally (see Table 2). **(b)** The same map as (a), without cluster thresholding. The color map legend (bottom left) shows the t-value of the group analysis against chance accuracy for panels (a) and (b). **(c)** As (a), except that mean classification accuracy values (chance = 25%) are depicted. **(d)** As (c), without cluster thresholding. The color map legend (bottom right) shows the accuracy scale for panels (c) and (d). Abbreviations: CS, central sulcus; PoCS, post-central sulcus; IPS, intraparietal sulcus; STS, superior temporal sulcus.

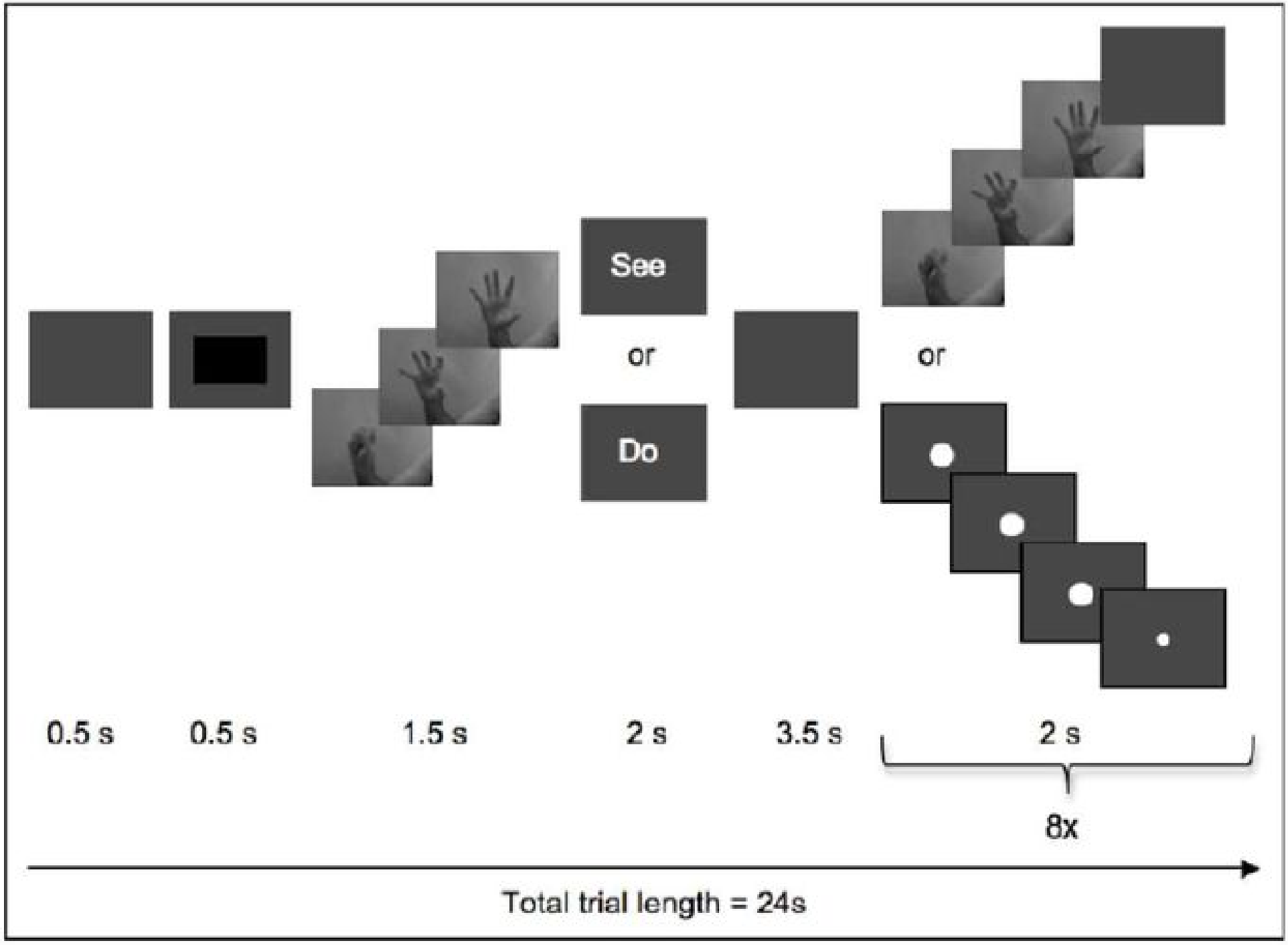
8. Regions in which representations are biased for effector or goal, Experiment 2. These data were first masked to select regions for which accuracy in the overall crossmodal analysis (Figure 7) was above chance. Vertices are coloured to indicate a bias in favour of either discrimination of the action effector (blue / cyan) or discrimination of the action goal (red / yellow). Areas with no bias are shown in green. Note a gradient in the bias from effector (postcentral gyrus) to action (superior parietal cortex).

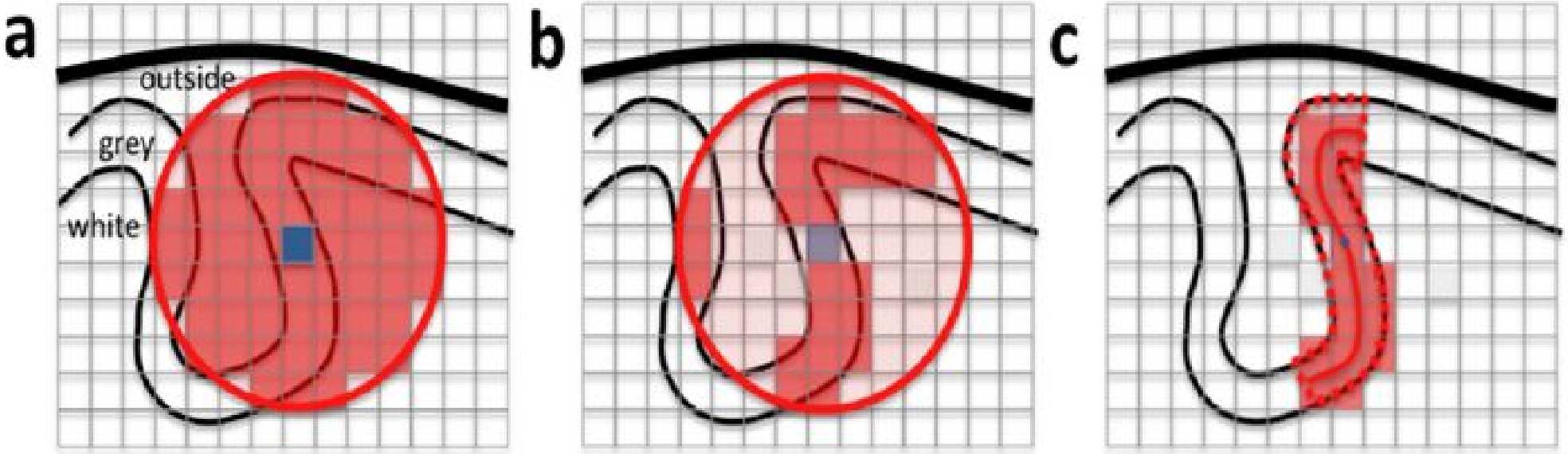
Table 1. Significant clusters in Experiment 1 that carry cross-modal information (see Figure 3). Center of mass is shown in Talairach coordinates. Mean and maximum classification t-values within each cluster are shown. Clusters are thresholded based on a bootstrap-approach (see Methods). Approximate anatomical locations are provided. Abbreviations: IPS, intraparietal sulcus; OT, occipitotemporal cortex; EVC, early visual cortex; MTG, middle temporal gyrus

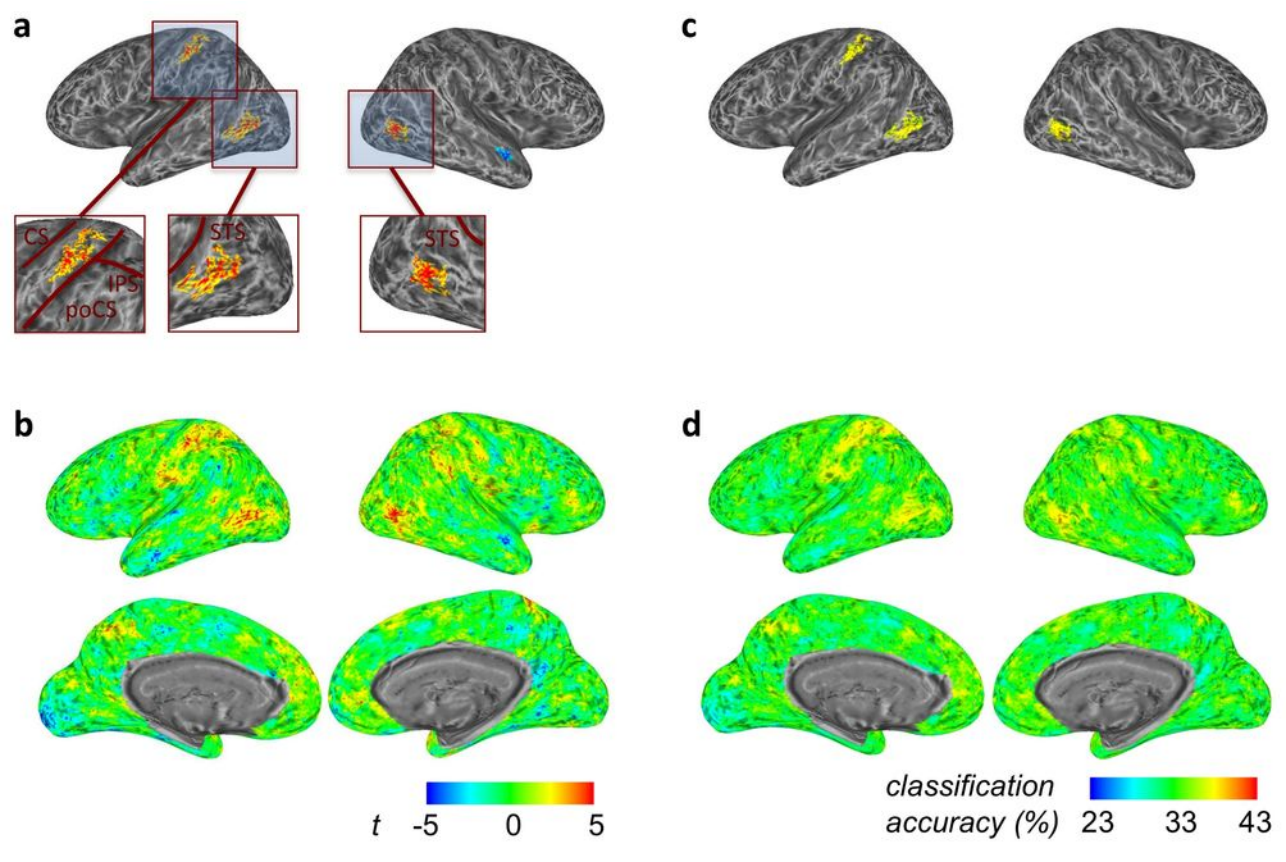
Name	Area (mm^2)	CoM L-R	CoM P-A	CoM I-S	Mean	Max
Left hemisphere						
aIPS	493	-44	-32	47	3.52	10.02
OT	445	-53	-56	3	3.72	9.71
EVC	329	-9	-87	-1	-3.79	-2.45
Right hemisphere						
aIPS	303	45	-62	3	3.89	11.11
MTG	87	48	-7	-10	-3.92	-2.46

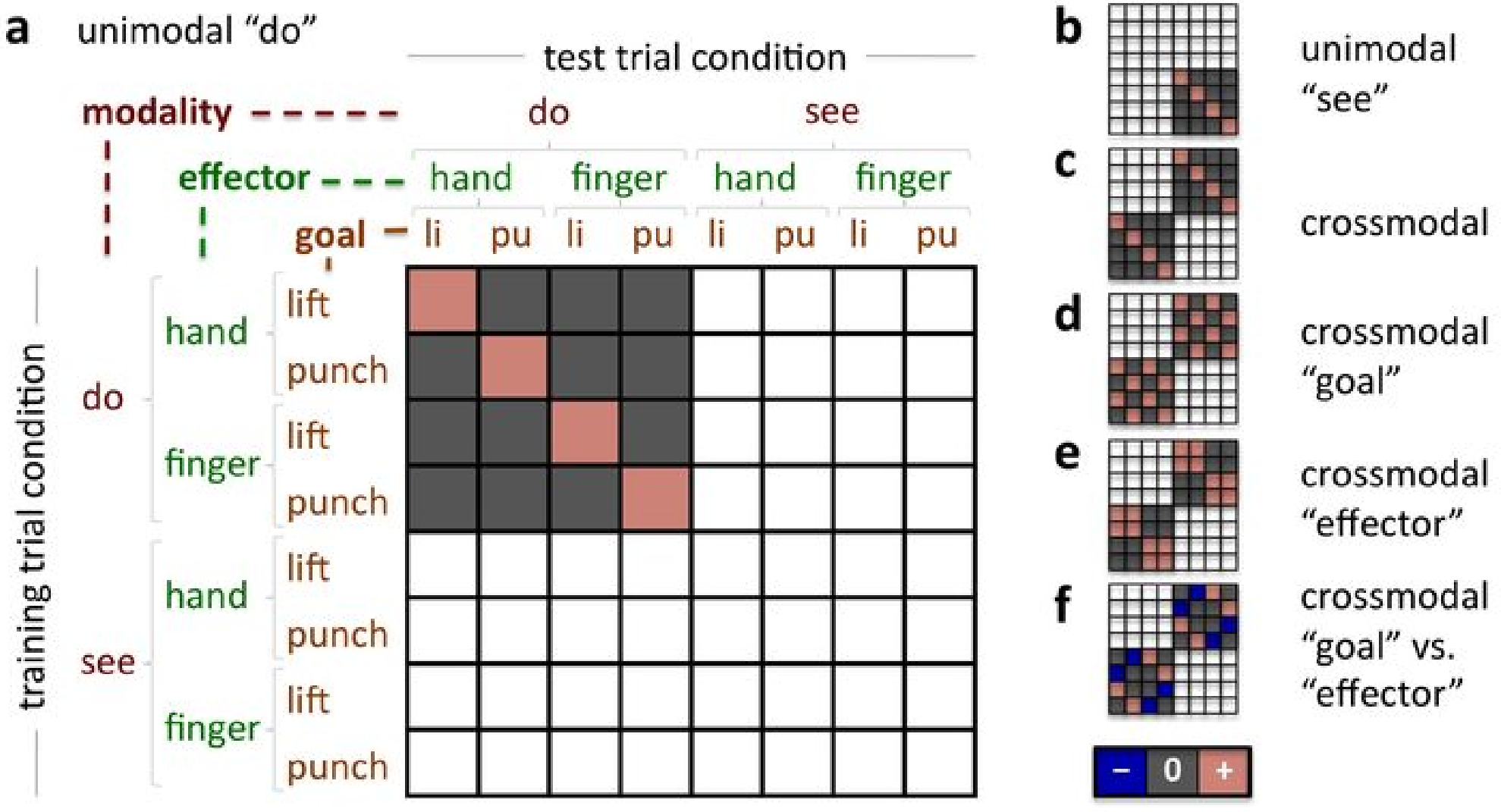
Table 2. Significant clusters in Experiment 2 that carry cross-modal information (see Figure 7). Conventions as in Table 1. Abbreviations: IPS, intraparietal sulcus; OT, occipitotemporal cortex; poCG, postcentral gyrus; SFG, superior frontal gyrus; PCC, posterior cingulate cortex.

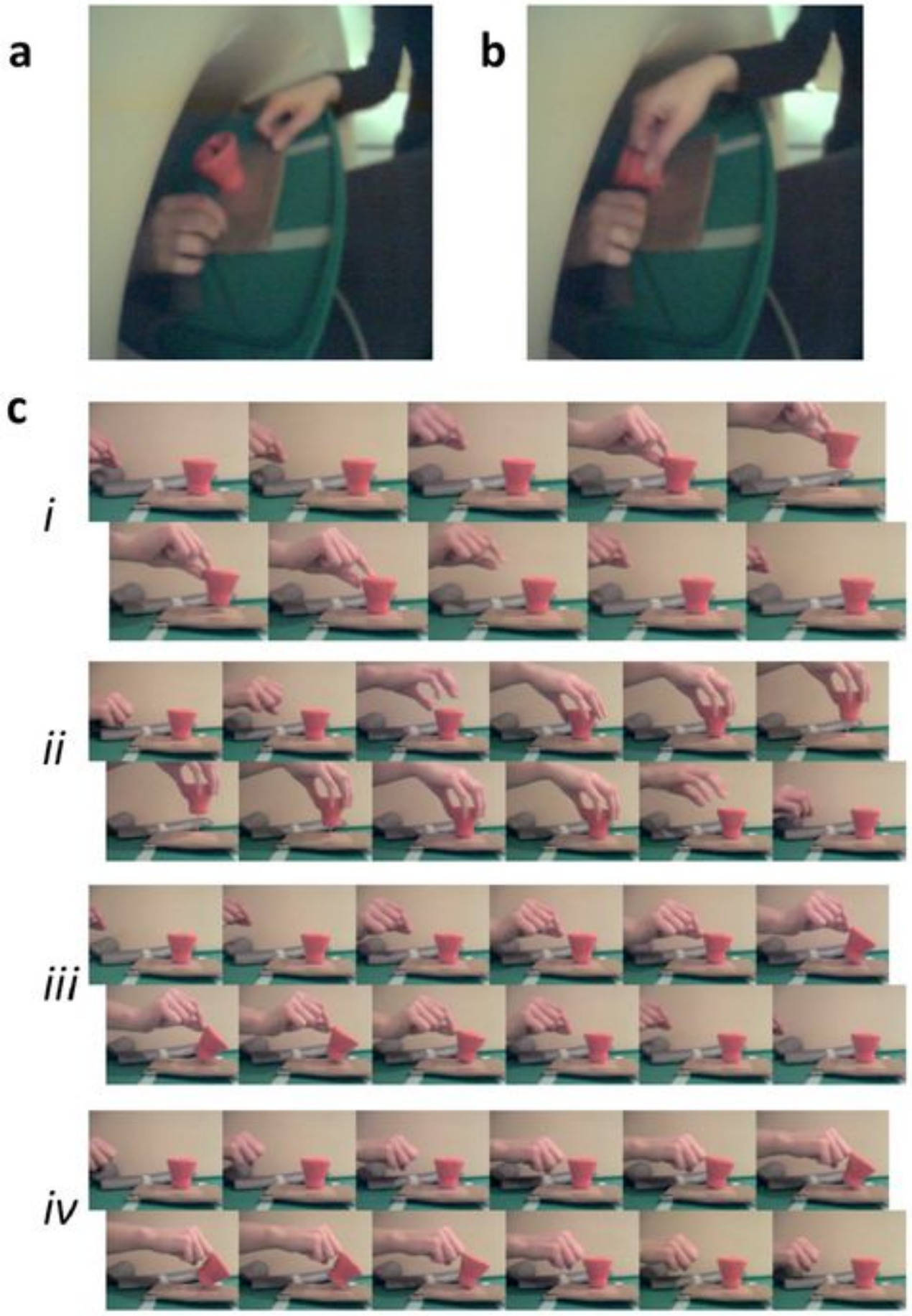
Name	Area (mm^2)	CoM L-R	CoM P-A	CoM I-S	Mean	Max
Left hemisphere						
aIPS	1953	-44	-31	44	3.38	7.61
OT	749	-49	-61	2	3.72	9.60
poCG	532	-52	-19	20	2.80	5.10
SFG	142	-23	54	12	-2.94	-2.20
Right hemisphere						
OT	887	43	-61	-7	3.14	7.59
PCC	217	5	-56	19	-3.19	-2.20


See
or
Do
or
0.5 s
0.5 s
1.5 s
2 s
3.5 s
2 s
8x
Total trial length = 24s

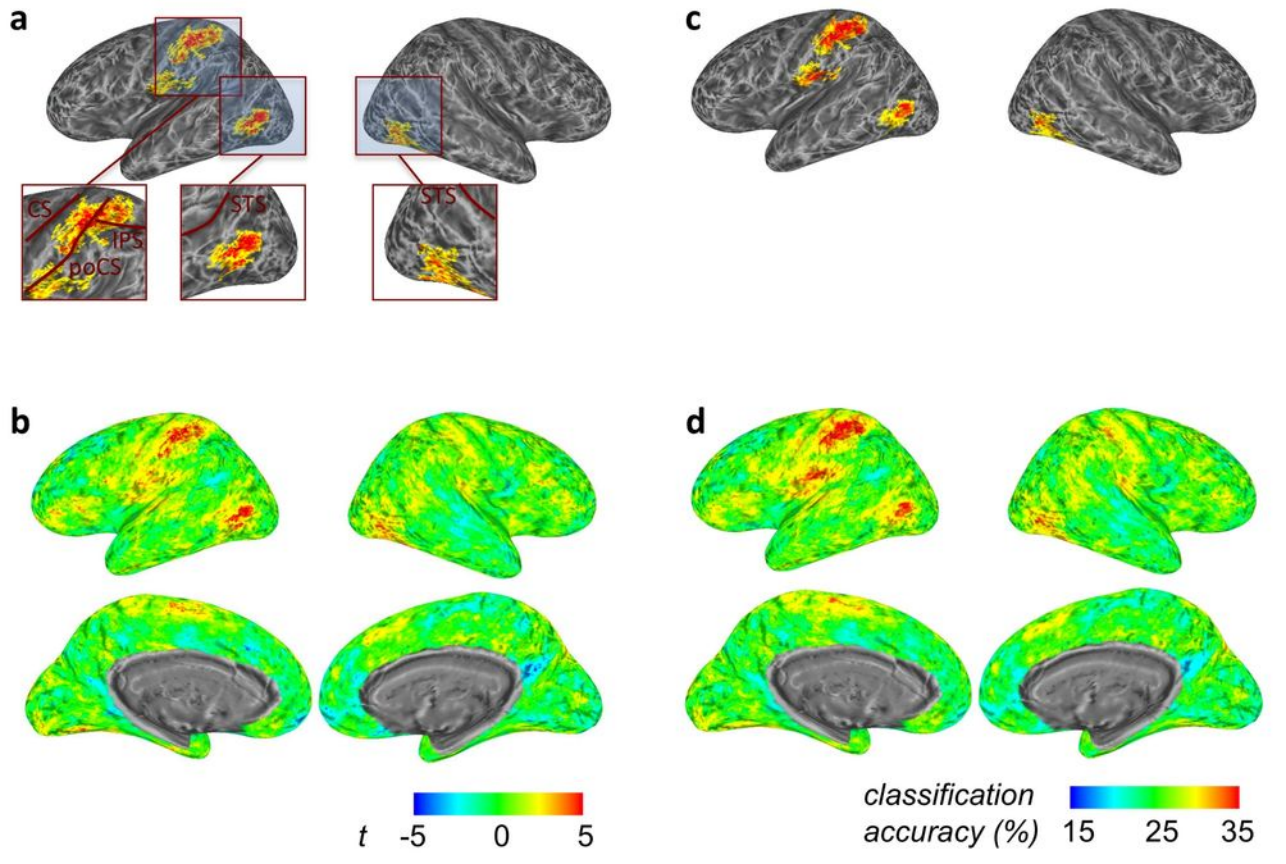

a
outside
grey
white
b
c

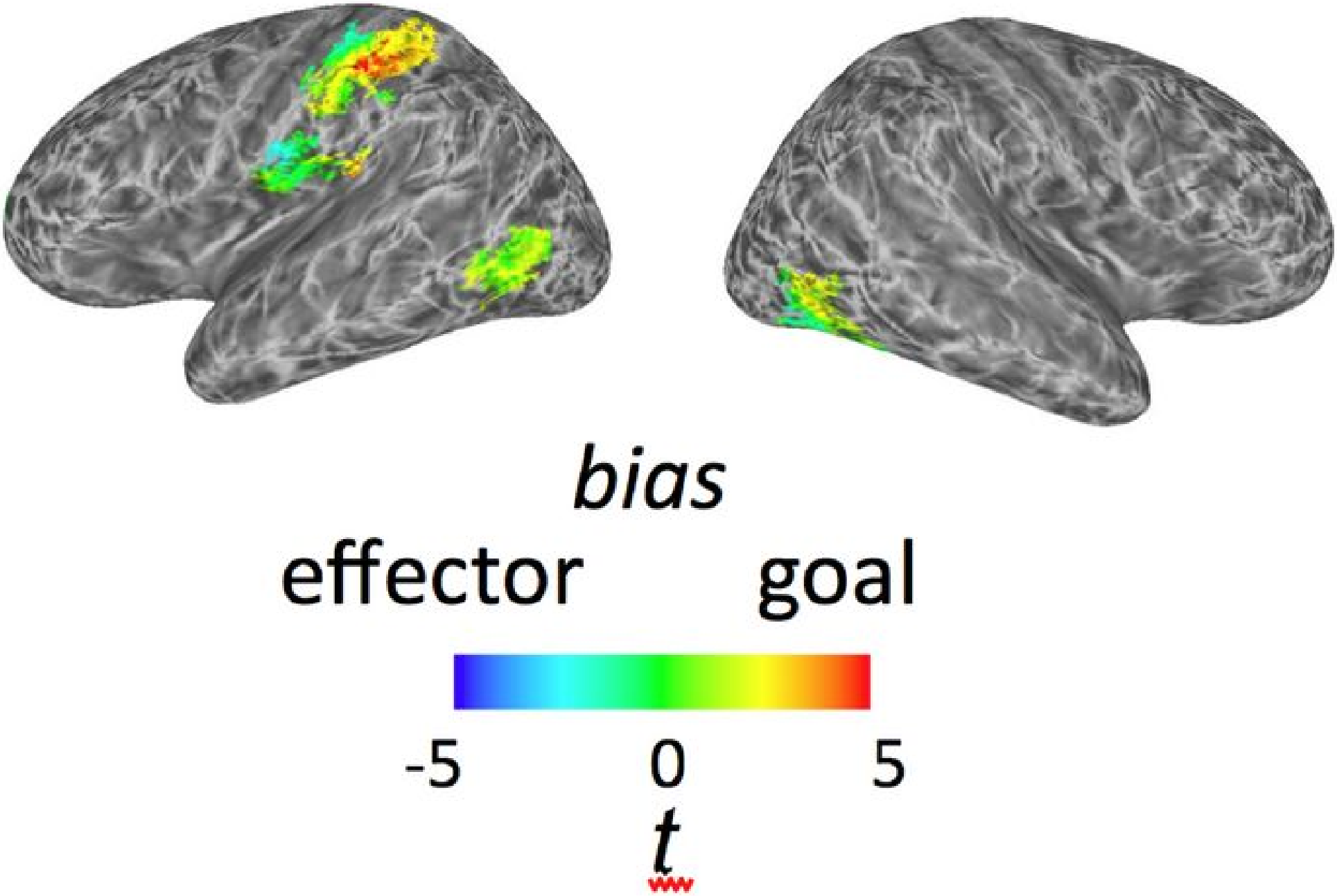

a
CS
IPS
poCS
STS
STS
c
b
d
t -5 0 5
classification
accuracy (%) 23 33 43


a
unimodal "do"
test trial condition
modality
do
see
effector
hand
finger
hand
finger
goal
li
pu
li
pu
li
pu
li
pu
training trial condition
do
see
hand
finger
hand
finger
lift
punch
lift
punch
lift
punch
lift
punch
b
unimodal "see"
c
crossmodal
d
crossmodal "goal"
e
crossmodal "effector"
f
crossmodal "goal" vs. "effector"
−
0
+


a
b
c
i
ii
iii
iv

"see" trials	instruction	"open"	
	action	(opens eyes)	(observes experimenter executing action)
		or	
"do" trials	instruction	"close"	"punch hand"
	action	(closes eyes)	(executes action)


a
CS
IPS
poCS
STS
STS
c
b
d
t -5 0 5
classification
accuracy (%) 15 25 35


bias
effector
goal
-5
0
5
t