OXFORD

Two distinct networks containing position-tolerant representations of actions in the human brain

Elahé Yargholi (iD)[1,2,*], Gholam-Ali Hossein-Zadeh (iD)[1,3], Maryam Vaziri-Pashkam (iD)[4]

[1]School of Cognitive Sciences, Institute for Research in Fundamental Sciences, Tehran 1956836484, Iran,
[2]Laboratory of Biological Psychology, Department of Brain and Cognition, Leuven Brain Institute, Katholieke Universiteit Leuven, Leuven 3714, Belgium,
[3]School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran 1439957131, Iran,
[4]Laboratory of Brain and Cognition, National Institute of Mental Health (NIMH), Bethesda, MD 20814, United States

*Corresponding author: School of Cognitive Sciences, IPM, Opposite the ARAJ, Artesh Highway, Aghdassieh, Tehran 1956836484, Iran.
Email: elahe.yargholi@gmail.com

Humans can recognize others' actions in the social environment. This action recognition ability is rarely hindered by the movement of people in the environment. The neural basis of this position tolerance for observed actions is not fully understood. Here, we aimed to identify brain regions capable of generalizing representations of actions across different positions and investigate the representational content of these regions. In a functional magnetic resonance imaging experiment, participants viewed point-light displays of different human actions. Stimuli were presented in either the upper or the lower visual field. Multivariate pattern analysis and a surface-based searchlight approach were employed to identify brain regions that contain position-tolerant action representation: Classifiers were trained with patterns in response to stimuli presented in one position and were tested with stimuli presented in another position. Results showed above-chance classification in the left and right lateral occipitotemporal cortices, right intraparietal sulcus, and right postcentral gyrus. Further analyses exploring the representational content of these regions showed that responses in the lateral occipitotemporal regions were more related to subjective judgments, while those in the parietal regions were more related to objective measures. These results provide evidence for two networks that contain abstract representations of human actions with distinct representational content.

Key words: action observation; position invariance; fMRI; multivariate pattern analysis.

Introduction

Humans can rapidly and accurately recognize other people's actions in the social environment. People's movement in the environment introduces often dramatic changes in the viewpoint, size, and position of their image on the retina. This confounding variability renders the action recognition task computationally challenging. Yet, humans' ability to recognize actions is rarely hindered by this variability. A brain region that subserves human action recognition ability is expected to contain (i) information about various human actions and (ii) action representations that are tolerant to the variability in the visual input. Here, using functional magnetic resonance imaging (fMRI) and multivariate pattern analysis (MVPA), we aim to identify such regions using changes in the position of the actor as the source of variability. After identifying these regions, we aim to explore their representational content and determine to what extent they relate to human subjective judgments about actions or objective measures of body movements.

Previous studies that have explored tolerance of visual representations to variability in the input have mostly focused on static pictures of objects and human bodies. Regions in both occipitotemporal (Konen and Kastner 2008; Cichy et al. 2011, 2013; Anzellotti et al. 2013; Ramírez et al. 2014) and parietal cortices (Konen and Kastner 2008) have been identified, which contain position, size, and viewpoint-tolerant representation of static objects (Konen and Kastner 2008; Mur et al. 2010; Cichy et al. 2011, 2013; Anzellotti et al. 2013; Ramírez et al. 2014; Vaziri-pashkam and Xu 2019; Vaziri-Pashkam et al. 2019). This characteristic has been proposed as a defining feature of regions that contribute to human abstract object knowledge (DiCarlo and Cox 2007). In the domain of observed actions, despite the large body of literature that identifies regions in the occipitotemporal and parietal cortices that respond to (Caspers et al. 2010; Kalénine et al. 2010; Grosbras et al. 2012; Watson et al. 2013; Urgesi et al. 2014) and contain information about (e.g. Wheaton et al. 2004; Jastorff et al. 2010; Abdollahi et al. 2013; Lingnau and Downing 2015; Ferri et al. 2015; Hafri et al. 2017; Wurm, Caramazza, et al. 2017b; Urgen et al. 2019; Tucciarelli et al. 2019; Tarhan and Konkle 2020; Urgen and Orban 2021) observed actions, evidence for the tolerance of these representations to changes in the position of the actors is still scant.

Previous studies have extensively investigated the tolerance of action representations to changes in viewpoint in both monkeys (Oram and Perrett 1996; Caggiano et al. 2012; Vangeneugden et al. 2014; Maeda et al. 2015; Maranesi et al. 2015; Barz et al. 2017; Simone et al. 2017; Livi et al. 2019; Albertini et al. 2020; Lanzilotto et al. 2020) and humans (Grossman et al. 2010; Ogawa and Inui 2011; Oosterhof et al. 2012; Tucciarelli et al. 2015; Isik et al. 2017). Results of these human studies provide evidence for viewpoint tolerance in the lateral occipitotemporal (Grossman et al. 2010; Tucciarelli et al. 2015) and parietal cortices (Ogawa and Inui 2011; Oosterhof et al. 2012). In contrast to viewpoint tolerance, position tolerance has not been extensively explored. Grossman et al. (2010) employed an fMRI adaptation approach to investigate position tolerance in the human brain. They found that a region in human STS shows tolerance to variations in the position of observed actions. They did not systematically investigate the position sensitivity of parietal action-selective regions. The fMRI adaptation technique used in this study may not have sufficient power in identifying the full scope of position tolerance in action-selective regions compared to other techniques, such as MVPA, (Haxby 2012), that directly investigate the representational content of a region. Roth and Zohary (2015) employed MVPA correlation and crossdecoding to study position tolerance during observation of tool-grasping movements in humans. Their stimuli were presented in various locations, and position tolerance was examined for presentations in the left and right visual fields. They discovered that position information is gradually lost, and hand/tool identity information is enhanced along the posterior–anterior axis in the dorsal stream. However, their study was limited to grasping movements. It is not obvious whether their results would generalize to full-body movements.

Other studies that have employed MVPA have often used natural stimuli in which actions happen at different positions, but they have averaged the responses across stimuli instead of systematically investigating the tolerance of representations across changes in position. Therefore, it is hard to speculate about the extent of position tolerance in action-selective regions from these studies (Hafri et al. 2017; Wurm, Caramazza, et al. 2017b; Tarhan and Konkle 2020). As such, a systematic investigation of position tolerance across dorsal and ventral action-selective regions is still missing in the literature. In this study, we will use MVPA and cross-position decoding using a support vector machine (SVM) classifier to search for position-tolerant representations of actions in the human brain. Within these regions, we will then investigate their representational content using a representational similarity analysis (RSA; Kriegeskorte et al. 2008).

Several theories have been proposed to characterize the role of individual action-selective regions in the processing of action stimuli. In the occipitotemporal cortex, a division of labor has been suggested between regions

such as EBA, that process the form of the body, and regions such as STS, that process the movement of the body (Giese and Poggio 2003; Michels et al. 2005; Downing et al. 2006; Peelen et al. 2006; Jastorff and Orban 2009; Grossman et al. 2010; Vangeneugden et al. 2014). These studies do not elaborate on the representational content of these regions. A few studies have taken a step further to establish the principles of coding in action-selective regions. Within the lateral occipitotemporal cortex, Wurm, Caramazza, et al. (2017b) have suggested that the neural representation of hand actions is organized based on the extent of sociality and transitivity of these actions. Recently, in a study with a large set of natural images of actions, Tucciarelli et al. (2019) compared the similarity in neural representation with their semantic similarity obtained from behavioral ratings. They showed that the neural organization of observed actions in the lateral occipitotemporal cortex is correlated with the behavioral similarity judgments.

Investigations on the representational content of parietal action-selective regions are fewer in number, and their results are more subject to debate. Shmuelof and Zohary (2006, 2008) have proposed an effector-dependent representation of hand actions in the anterior intraparietal cortex, while others have argued that actions are represented in an effector-independent manner in the inferior parietal lobe (Jastorff et al. 2010) as well as superior parietal lobe and intraparietal sulcus (IPS) (Vingerhoets et al. 2012). One study (Jastorff et al. 2010) suggested that action representations in the parietal cortex are related to the direction of the action relative to the body.

Looking at the entire visual system, Tarhan and Konkle (2020) used an encoding model to predict responses to natural videos. The feature space of their model captured the body parts involved in an action and the action target. Based on the voxel tunning, they suggested that 5 large-scale networks exist in the human brain which represent actions based on their sociality and the spatial extent of their interaction envelope. All these studies have employed natural videos of actions in which often a whole scene and objects are present. Employing natural stimuli may lead to many confounding factors and difficulties in interpreting the results. Even though some of these studies have taken steps to make sure low-level features are not contributing to their results (Tucciarelli et al. 2019; Tarhan and Konkle 2020), there is a possibility their results may be related to the presence of special types of objects, scene contexts, and semantic relations between them (Kourtzi and Kanwisher 2000; Senior et al. 2000; Johnson-Frey 2004; Buxbaum et al. 2006; Wurm et al. 2012; Schubotz et al. 2014; Wurm and Schubotz 2017; Wurm, Artemenko, et al. 2017a; El-Sourani et al. 2018; Leshinskaya et al. 2018).

Here, employing fMRI MVPA, we searched for position invariant representation of human actions presented in point-light display format . We used controlled stimuli in PLD format to restrict the visual information to the bodily movements and to identify regions supporting abstract
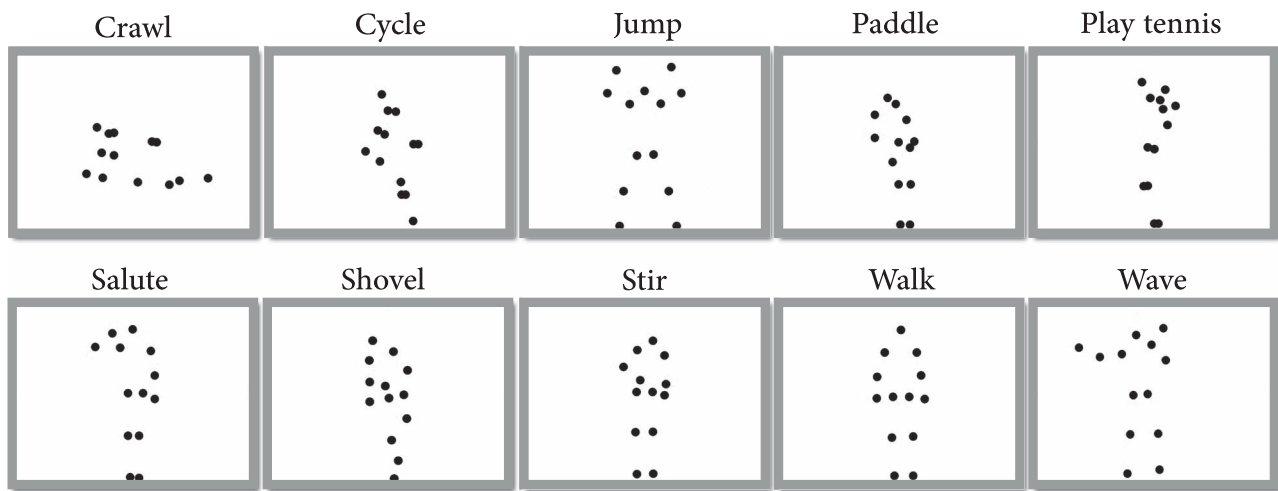
Fig. 1. Representative frames of the 10 human action videos in PLD format used in the experiment.

action representations. Using these controlled action stimuli, we localized regions where action decoding was robust to changes in position. Employing PLD stimuli also allowed us to determine the position and movements of individual limbs and to devise objective measures for estimating similarity between actions. Using these objective measures of similarity as well as subjective measures derived from behavioral experiments, we characterized the representational content of action-selective regions.

Materials and methods
Participants

Twenty-five subjects (14 females, 20–38 years of age) took part in the fMRI experiment and 15 subjects (13 females, 25–38 years of age) took part in the behavioral experiment. Behavioral subjects were independent of fMRI subjects. All subjects were healthy and were right-handed with normal visual acuity. They gave written informed consent and received payment for their participation. The experiments were approved by the Ethics Committee on the use of human subjects at the Iran University of Medical Sciences. Three fMRI subjects were later excluded due to excessive motion (see Data analysis for more details), and the final analyses included the remaining 22 subjects.

fMRI experiment

In this experiment, we used videos of 10 different human actions in PLD format: Crawl, Cycle, Jumping Jack (henceforth referred to as Jump), Paddle, Play tennis, Salute, Shovel, Stir, Walk, and Wave (Fig. 1) (in the original dataset, Paddle and shovel are called Pedal and Spade, respectively). These actions were selected from a larger set of human action videos in PLD format provided by Vanrie and Verfaillie (2004).

We used a block design paradigm. Each run included 20 blocks with two blocks for each human action video. In one of the blocks for each action, the video was presented in the upper visual hemifield, and in the other, it was presented in the lower visual hemifield. The presentation order of the stimuli was counterbalanced across runs. Each block lasted for 8 s. There was an 8-s blank period in the beginning, end, and between stimulus blocks of each run. Each run lasted for 328 s. Each subject completed 1 session of 10 runs.

In each stimulus block, there were 4 repetitions of the same human action video. Each repetition lasted for 1.5 s (frame rate 30), followed by a 0.5-s blank period. The point-lights, subtending ~0.25° of visual angle, were in white color against a black background. The center of the stimuli was presented ~4° of visual angle above/below the fixation point. The size of the area that the dots occupied during movements was between 0.94° and 5.62° in width and between 2.38° and 7.7° in height.

The fixation point was a red circle of size ~0.25° of visual angle. In a randomly chosen repetition in each block, the size of the dots became 50% larger for 1 s. Subjects were instructed to fixate on the fixation point at the center of the screen, watch the videos, detect the change in the size of the point lights, and report it by pressing a response key with their right index finger.

Behavioral experiment

To capture the similarity between our human action stimuli, we conducted a behavioral experiment based on an inverse multidimensional scaling method proposed by (Kriegeskorte and Mur 2012). At the beginning of the experiment, one snapshot from each of the 10 action videos was presented along with a number in the range of 1–10 in two vertical columns at the left border of the screen. On a separate screen, participants were provided with numbered action videos, and they could watch them as many times as they needed. They were asked to rearrange the snapshots on the surface of a gray circle (drag and drop using the mouse) according to the perceived similarity of their corresponding videos. The arrangement was performed based on subjective judgments of

overall similarity, and no other instructions were given regarding the specific aspects they needed to focus on for the arrangement. A behavioral dissimilarity matrix was obtained based on the pairwise Euclidean distances between positions of actions in the final 2D arrangement on the circle for each participant (see Supplementary Fig. S3 for behavioral dissimilarity matrix and MDS).

Magnetic resonance imaging methods

Magnetic resonance imaging (MRI) data were collected at two centers: at the school of Cognitive Sciences, Institute for Research in Fundamental Sciences (Tehran, Iran), using a Siemens 3T Tim Trio MRI scanner and a 32-channel head coil, and at the National Brain Mapping Lab (Tehran, Iran) using a Siemens 3T Prisma MRI scanner and 20-channel or 64-channel head coil. We were forced to switch MRI equipment midexperiment due to scanner/coil malfunction (data of 5 subjects were recorded at the school of Cognitive Sciences, and data of 20 subjects were recorded at National Brain Mapping Lab). We moved scanners to ensure that high-quality data were collected for all subjects. Similar protocols were used across scanners, and the data recorded with different equipment were comparable.

Subjects viewed the visual stimuli through a back-projection screen, and the task was presented using MATLAB and Psychtoolbox-3 (Brainard 1997). Functional images were obtained using a T2*-weighted single-shot gradient-echo EPI sequence with a time repetition (TR) of 2 s, time echo (TE) 26 ms, 90° flip angle, 30 transverse slices, and a voxel size of $3 \times 3 \times 4\,mm^3$. A high-resolution T1-weighted structural scan was also acquired from each participant using an MPRAGE pulse sequence (TR = 1,800 ms, TE = 3.44 ms, inversion time = 1,100 ms, 7° flip angle, 176 sagittal slices, and $1 \times 1 \times 1\,mm^3$ isotropic voxels).

Data analysis

Freesurfer (https://surfer.nmr.mgh.harvard.edu) and FS-FAST (Dale et al. 1999) were employed for data preprocessing and general linear model (GLM) analysis. Preprocessing of fMRI data included motion correction, slice timing correction, linear and quadratic trend removal, and no spatial smoothing. Subjects with excessive head motion (>1.5 mm within a run and 4 mm across runs) were excluded (3 subjects). Functional data were then resampled to the cortical surface of individual subjects.

A run-wise GLM analysis was performed to obtain the beta values and their corresponding *t*-statistic for each human action and each presentation hemifield in each vertex (10 human actions in the upper visual field and 10 human actions in the lower visual field, leading to 20 regressors). Linear and quadratic polynomial nuisance regressors and external regressors from the estimated head movements were also included.

Wang atlas of visual topography (Wang et al. 2014) was used to localize retinotopic areas for each subject: V1, V2,

V3, hV4, VO1, VO2, MST, hMT, LO2, LO1, V3a, V3b, IPS0, IPS1, IPS2, IPS3, IPS4, IPS5, SPL1 (Fig. 5A).

Multivariate pattern analysis

In-house MATLAB code, LibSVM (Chang and Lin 2011), and CoSMoMVPA (Oosterhof et al. 2016) toolboxes were employed for searchlight and Region Of Interest (ROI) analysis.

Searchlight analysis

We performed action classification using a surface-based searchlight procedure to obtain a map of classification accuracy (Oosterhof et al. 2011). In a leave-one-run-out crossvalidation procedure, samples (*t*-statistics of vertices) were partitioned to train and test sets. A multiclass linear SVM was employed to perform brain decoding classification for each presentation position (upper/lower visual hemifield) in individual brains with a searchlight circle of 100 vertices on the surface. The decoding accuracy of a searchlight was calculated as the average of within-position classification accuracies, i.e. decoding was performed for each presentation position (upper/lower visual hemifield) separately, and the average of these two classification accuracy maps was used as the within-position classification accuracy. The resulting maps were then resampled to a common surface for the group-level statistical analysis.

To determine regions showing position tolerance, we performed cross-position decoding by training a classifier to discriminate actions presented in one position and testing its accuracy in classifying the same actions in the other position. To match the analysis procedure of within- and cross-position accuracies, we also employed a leave-one-run-out crossvalidation procedure for this analysis, training on 9 runs and testing on a left-out run. We performed the analysis in both directions (training on upper visual field and testing on lower visual field and vice versa) and averaged the results across the two directions. A cross-position decoding accuracy greater than the chance level would indicate position tolerance.

To obtain group-level statistics, *P*-values of classification accuracies were computed using a binomial test, which was corrected for false discovery rate (Benjamini and Hochberg 1995) at level $q1$. The second-level *P*-value for each vertex was then determined as

$$\Pr\left[\mathrm{Binomial}\left(n, R * q1/\left(n * m\right)\right) \geq C\right],$$

with C denoting the number of participants for which that vertex was significant, R denoting the sum of the counts of C across all vertices, n denoting the number of participants, and m denoting the number of vertices. Under the null hypothesis, C has a binomial distribution with size n and a probability that is approximately bounded by $R * q1/(n * m)$. Finally, the derived second-level *P*-values were thresholded at the FDR level $q2$ to acquire the significant vertices at the group level (McMahon et al. 2019). This approach is more appropriate

than the *t*-test for comparing classification accuracies against chance level (see Allefeld et al. 2016 for a discussion of why the *t*-test is not suitable in such cases).

From the searchlight group-level statistical analysis, we obtained clusters with significant cross-positions classification accuracy. Thereafter, significant clusters with >100 vertices were hand-selected based on anatomical landmarks to define the five ROIs in which decoding of actions could be generalized across positions and could be used in the next analyses to explore their representational structure. To ascertain that the ROI selection and further analyses are statistically independent and prevent double-dipping, we used a leave-one-subject-out approach (Etzel et al. 2013). The group-level statistical analysis was applied to the searchlight results of all subjects except one, and the resulting ROI was projected on the left-out subject to select the ROI in that subject. The procedure was repeated for each subject. The same clusters were found consistently across all iterations.

Comparing within- and cross-position classification accuracies

To directly compare within- and cross-position classification accuracies, we performed SVM classifications within each ROI obtained from the leave-one-subject-out analysis described above. The procedure for obtaining within- and cross-position classifications was the same as the searchlight analysis except that we employed two-class instead of multi-class linear SVM and calculated the average of all pair-wise classification accuracies. Different ROIs do not include the same number of vertices. This variation in the number of vertices across ROIs could influence classification accuracies. To avoid this potential confounding factor, the pairwise classifications were performed using the 100 most informative vertices in each ROI. To select the most informative vertices, a *t*-test was applied on the training set, and 100 vertices with the lowest *P*-values for discriminating between the conditions of interest in the training set were chosen (Mitchell et al. 2004). It is noteworthy that the classification results did not qualitatively differ without feature selection.

Exploring the representational content of the ROIs

To explore the representational content of the ROIs, we constructed models based on the body-part movements and behavioral ratings and examined the correlation between these models and the cross-position decoding accuracies for each ROI.

To obtain the body-part movement model, we divided the point lights into five groups: trunk (five dots), left hand (two dots), right hand (two dots), left leg (two dots), and right leg (two dots). We obtained the sum of the displacements of point lights in each group for each action video. From the total displacements in each group, we obtained a 5-element vector. Then, the Euclidean distance between these 5-element vectors was used to obtain a dissimilarity matrix based on the body-part movement. This body-part movement model includes

information on both the pattern of body-part movement and the average of their movement for each action. Hence, for further investigation, we examined two other models: body-movement-pattern and body-movement-average. To obtain the body-movement-pattern model for each action, we demeaned the corresponding 5-element vector of the body-part movement model and obtained the Euclidean distance between these new demeaned vectors to build a dissimilarity matrix. The difference between the average values of body-part movement vectors was also used to obtain the body-movement-average model.

To obtain the behavioral dissimilarity model, the results from the behavioral experiment were used. In the behavioral experiment, the distance between stimuli indicates their dissimilarity. The obtained dissimilarity matrix for each subject was normalized by dividing each value in the matrix by the maximum value. The pooled behavioral dissimilarity matrix was then computed as an average of individual normalized behavioral dissimilarity matrices.

To obtain the correlation between the models and the action representations in each ROI, the off-diagonal of the matrix obtained from the pairwise cross-position decoding accuracies was vectorized. The off-diagonal of the dissimilarity matrices from the models were also vectorized. The Kendall rank correlation between dissimilarity vectors and decoding accuracy vectors was then calculated, and the correlation values were compared to characterize the representational content of each ROI.

ROI similarity analysis

We also performed the cross-position decoding accuracies for ROIs obtained from the Wang atlas using a similar procedure as that used for the ROIs obtained from the searchlight analysis. Following a leave-one-run-out approach, a classifier was trained to decode pair of actions presented in one position and tested with the same pair of actions in another position. We applied a *t*-test on the training set to choose 100 vertices with the lowest *P*-values for discriminating between the actions of interest in each ROI. Then, to investigate the representational similarity between ROIs (Wang's ROIs and ROIs obtained from searchlight analysis), correlations were computed between the vectorized matrices of pairwise cross-position decoding accuracy. One minus these correlation values were used to obtain the distance between the ROI pairs, and these distances were used to construct an ROI dissimilarity matrix. The ROI dissimilarity matrix was first computed for individual subjects and was then averaged across subjects to acquire a group-level ROI dissimilarity matrix (Fig. 5B).

Split-half reliability of ROI dissimilarity matrix was also evaluated; we randomly divided subjects into two equal groups and correlated the corresponding group-level ROI dissimilarity matrices as a measure of reliability. This measure was calculated for 10,000 random split-half divisions and was averaged to produce the
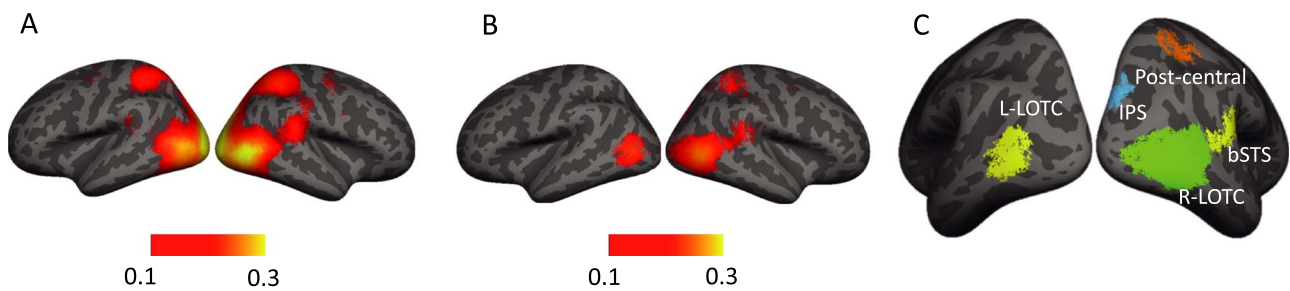
Fig. 2. Results of searchlight classification accuracy (A) average within-position classification accuracy across subjects, significant at the group level ($q1 = 0.05$ and $q2 = 0.01$); (B) average cross-position classification accuracy across subjects, significant at the group level ($q1 = 0.05$ and $q2 = 0.01$); (C) five clusters with significant cross-position classification accuracy used as ROIs (note that the exact ROIs differ slightly across iterations of the leave-one-subject-out procedure).

final reliability measure. To examine the significance of this reliability measure, we obtained the bootstrapped null distribution of reliability by random shuffling of the labels in the correlation matrix separately for the two split-half groups and by calculating the reliability for 10,000 random samples.

An MDS analysis was then performed on this ROI-dissimilarity matrix, and the first two dimensions that captured most of the variance were used to produce an MDS diagram in which the distance between each pair of ROIs represents the similarity between them (Vaziri-Pashkam and Xu 2019). The squared correlation (r^2) between 2D distances in the MDS plots and the original distances in the multidimensional space was used to quantify the variance explained by the first two dimensions. We then fit two regression lines employing the total least square method, with one line passing through the positions of the occipitotemporal regions and the other passing through the positions of parietal regions. The variance explained by the fitted lines was also measured; 2D distances between regions based on the predicted positions on the lines were obtained, and the r^2 between these distances and the original distances based on the input matrix to the MDS analysis was calculated. Similarly, we computed the r^2 between distances based on the MDS plots and those based on the line predictions to determine the variance explained by the two lines on the MDS plot.

Results

In the present study, we aimed to identify regions containing position-tolerant representations of actions and to investigate the representational content of these regions using fMRI MVPA. A stimulus set of 10 different human action videos in PLD format (Fig. 1) was presented at two different positions in the visual display (upper and lower visual hemifields), and *t*-values were extracted in each voxel of the brain for each of the actions and each position. Employing a multiclass SVM classifier, within- and cross-position decoding was then applied following a searchlight approach across the whole brain.

Action decoding across the whole brain

To identify regions that showed selectivity for our action stimuli within each presentation position (upper/lower visual fields), classifiers were trained and tested with fMRI responses for stimuli presented within the same position. To identify regions showing position tolerance, we performed a cross-position classification. Classifiers were trained with fMRI responses to stimuli presented at the upper visual field and were tested with fMRI responses to stimuli presented at the lower visual field and vice versa. Both within- and cross-position classification were applied using a surface-based searchlight method.

Figure 2A and B depict the average within- and cross-position classification accuracy across subjects, significant at the group level ($q1 = 0.05$ and $q2 = 0.01$), respectively. Comparison of these maps reveals that a subset of the regions that show above chance within-position classification also demonstrate generalization across positions. The group analysis showed significantly above-chance cross-position classification accuracy in the left and right lateral occipitotemporal cortices (LOTC), right IPS, and right postcentral gyrus (Fig. 2B).

Action decoding in ROIs

To investigate the extent to which changes in position affect the representations in individual regions, we used a leave-one-subject-out procedure (see Materials and methods) to extract ROIs with above-chance crossdecoding accuracy. Five clusters (Fig. 2C) were selected based on a searchlight analysis on all but one subject, and the selected clusters were used as ROIs for the left-out subject: LOTC, bank of right superior temporal sulcus (bSTS), right IPS, and right postcentral gyrus. A small number of voxels passed the threshold in other regions, including regions in the right premotor (28 ± 14 vertices) and the left parietal cortex (3 ± 2 vertices) as well, but since the number of vertices were too low and did not pass the 100 vertex threshold, these regions were not selected.

A two-class linear SVM was employed for within- and cross-position classification within each ROI using the 100 most informative vertices in each ROI (Mitchell et al. 2004). Figure 3A illustrates the results of within- and
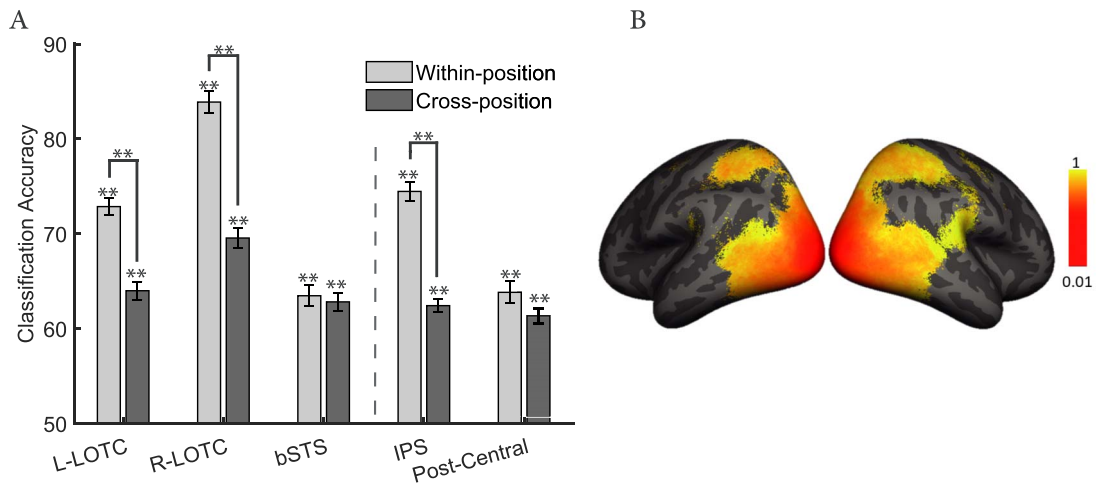
Fig. 3. (A) Within and cross-position decoding accuracy for the ROIs obtained from searchlight analysis. Error bars indicate standard errors of the means (**group analysis at $q1 = 0.01$, $q2 = 0.01$ or paired t-test, $P < 0.01$, FDR-corrected). The vertical dashed line separates occipitotemporal ROIs from parietal ROIs. (B) Maps of cross-position divided by within-position classification accuracy obtained from searchlight analysis in vertices with significant within-position classification accuracy.

cross-position decoding accuracy for each of the ROIs. All ROIs had a significant decoding accuracy for cross- and within-position (group analysis at $q1 = 0.01$, $q2 = 0.01$). In the left and right lateral occipitotemporal and parietal ROIs, cross-position decoding accuracy was significantly lower than within-position decoding accuracy (paired t-test, $Ps < 0.0075$, FDR-corrected $q = 0.01$), while this difference was not significant in bSTS and postcentral regions (paired t-test, $Ps > 0.13$, FDR-corrected $q = 0.01$). These results suggest a reduction in position sensitivity from posterior to anterior regions in both occipitotemporal and parietal cortices. To visualize this finding on the whole brain, we divided the cross-position by within-position classification accuracy in all vertices with above-chance within-position accuracy. The resulting map averaged across subjects is depicted in Fig. 3B. In line with the ROI analysis, this map demonstrates an increase in position tolerance from posterior to anterior regions in both occipitotemporal and parietal cortices. We have also compared the differences in within-position classification accuracy between upper and lower visual field stimuli for these ROIs. Classification accuracies were higher for the lower visual field stimuli, but we did not find any particular differences between parietal and occipitotemporal regions in this regard (see Supplementary Fig. S1).

No eye tracker was used inside the scanner, so we do not have a direct record of eye movements during the scan. However, participants practiced the task outside the scanner under supervision and we made sure they are able to perform the task fixating on the central fixation point. In addition, there is good evidence in the results to demonstrate that subjects were following the fixation instructions: (i) V1 did not show above-chance accuracies in the cross-position classification analysis. If the participants were looking at the stimuli directly, we would have observed high cross-position decoding in V1. (ii) There are significant differences between the

accuracies for upper and lower visual field stimuli. These differences would not be observed if the participants were not fixating. (iii) We performed a contrast analysis comparing upper and lower visual field stimuli. The result maps in early visual cortex indicate strong activation in the dorsal and ventral early visual cortex for the lower and upper visual field stimuli, respectively (see Supplementary Fig. S2 for the maps of contrast analysis comparing upper and lower visual field stimuli).

Revealing the representational content of the ROIs

To uncover the principles that govern the organization of action representations in each ROI, we used an RSA (Kriegeskorte et al. 2008). We constructed models based on the body-part movements and behavioral similarity ratings performed on the videos by human observers (see Materials and methods). We then used correlation analysis to investigate the similarity of action representations between models and ROIs according to their cross-position pairwise decoding accuracies (see Supplementary Figs. S3–S5 for dissimilarity matrices of models and decoding accuracies). The reliability of the behavioral dissimilarity matrices was high (0.7654). Models of stimuli (distance matrices obtained from models) were also compared with each other; the Kendall correlation was calculated, and its significance was examined applying permutation tests. The behavioral and body-part models were not significantly correlated (Kendall correlation = 0.0394, $P = 0.2562$).

Figure 4A shows the Kendall correlations between the models and cross-position pairwise decoding accuracies across participants for body-part movement and behavioral rating models (performing this analysis using regression did not qualitatively change the results).

Significant correlations were found with the body-part model in all ROIs except for L-LOTC (permutation test, $P < 0.05$ for R-LOTC and $Ps < 0.01$ for bSTS, IPS, postcentral) and with the behavioral rating model in all
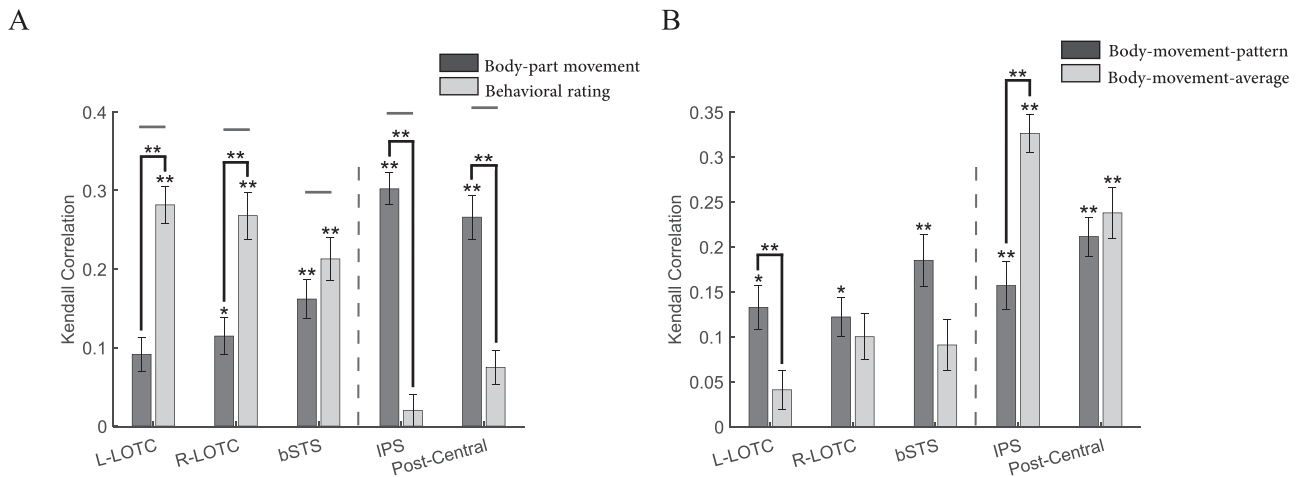
A

B

Fig. 4. Brain-Model correlations: (A) Kendall correlation body-part movement (dark gray) and behavioral rating models (light gray) and cross-position pairwise decoding accuracies for the individual ROIs. Error bars indicate standard errors of the mean (permutation test or paired *t*-test, *$P < 0.05$, **$P < 0.01$, FDR-corrected). The vertical dashed line separates occipitotemporal ROIs from parietal ROIs. The horizontal lines indicate the lower limit of the noise ceiling. (B) Kendall correlation between body-movement-average (light gray) and body-movement-pattern (dark gray) models and cross-position pairwise decoding accuracies for individual ROIs. Error bars indicate standard errors of the mean (permutation test or paired *t*-test, *$P < 0.05$, **$P < 0.01$, FDR-corrected). The vertical dashed line separates occipitotemporal ROIs from parietal ROIs.

ROIs except for IPS and postcentral (permutation test, $Ps < 0.01$).

A two-way repeated-measures ANOVA with model (body-part movement and behavioral rating) and ROI as independent factors and correlation coefficient as the dependent variable was performed. The effect of ROI ($F(4,84) = 0.9303$, $P = 0.4504$) and the effect of model ($F(1,84) = 0.7636$, $P = 0.3921$) were not significant, while the interaction between the two ($F(4,84) = 36.8$, $P < 0.0001$) was significant. Individual comparisons within ROIs showed that correlation coefficients were greater for the behavioral rating model in L-LOTC and R-LOTC and for the body-part movement model in IPS and postcentral (all $t(21) > 3.7117$, all $P < 0.0016$, FDR-corrected). To measure the reliability of the data, we used the lower limit of the noise ceiling, which gives an estimate of the highest correlation we can expect in each region when correlating neural and behavioral dissimilarity. For each region, it was computed as the crossvalidated correlation of each subject's dissimilarity matrix with the mean of the remaining subjects' dissimilarity matrices.

Figure 4B shows the average Kendall correlation between the models and cross-position pairwise decoding accuracies across participants for body-movement-pattern and body-movement-average models. As Fig. 4B shows, all ROIs showed correlations with the body-movement-pattern model (permutation test, $P < 0.05$ for L-LOTC and R-LOTC, $P < 0.01$ for bSTS, IPS and postcentral, FDR-corrected, 5 comparisons), but body-movement-average models only showed correlations with the parietal ROIs (permutation test, $P < 0.01$, FDR-corrected) and showed no significant correlation with the lateral occipitotemporal ROIs (permutation test, $P > 0.05$, FDR-corrected).

A two-way repeated-measure ANOVA with model and ROI as the independent factors and correlation

coefficient as the dependent variable showed no significant effect of model ($F(1,84) = 0.0117$, $P = 0.9148$), while the effect of ROI ($F(4,84) = 22.5280$, $P < 0.0001$) and the interaction between the two ($F(4,84) = 11.5105$, $P < 0.0001$) were significant. The significant effect of ROI points to greater correlation coefficients in parietal ROIs. Individual comparisons within ROIs showed that correlation coefficients were greater for the pattern model in L-LOTC ($t(21) = 2.6553$, $P = 0.0370$, FDR-corrected) and for the average model in IPS ($t(21) = 4.9394$, $P = 3.4583e - 04$, FDR-corrected).

We also compared the behavioral rating model with the average and pattern models by applying two separate two-way repeated-measure ANOVAs with model and ROI as independent variables and correlation coefficient as the dependent variable.

Results for the body-movement-average and behavioral rating models showed no significant effect of ROI ($F(4,84) = 1.1066$, $P = 0.3589$), and model ($F(1,84) = 0.3031$, $P = 0.5878$), while the interaction between the two was significant ($F(4,84) = 36.4416$, $P < 0.0001$). Individual comparisons within ROIs showed that correlation coefficients were greater for the behavioral rating model in L-LOTC and R-LOTC and for the body-movement-average model in IPS and postcentral (all $t(21) > 3.5686$, all $P < 0.0018$, FDR-corrected).

For the body-movement-pattern and behavioral rating models, results showed no significant effect of model ($F(1,84) = 0.4335$, $P = 0.5174$), while the effect of ROI ($F(4,84) = 8.8253$, $P = 0.0001$) and interaction between the two ($F(4,84) = 37.6654$, $P < 0.0001$) were significant. Individual comparisons within ROIs showed that correlation coefficients were greater for the behavioral rating model in L-LOTC and R-LOTC and were greater for the body-movement-pattern model in IPS and postcentral (all $t(21) > 4.8254$, all $P < 0.0001$, FDR-corrected).
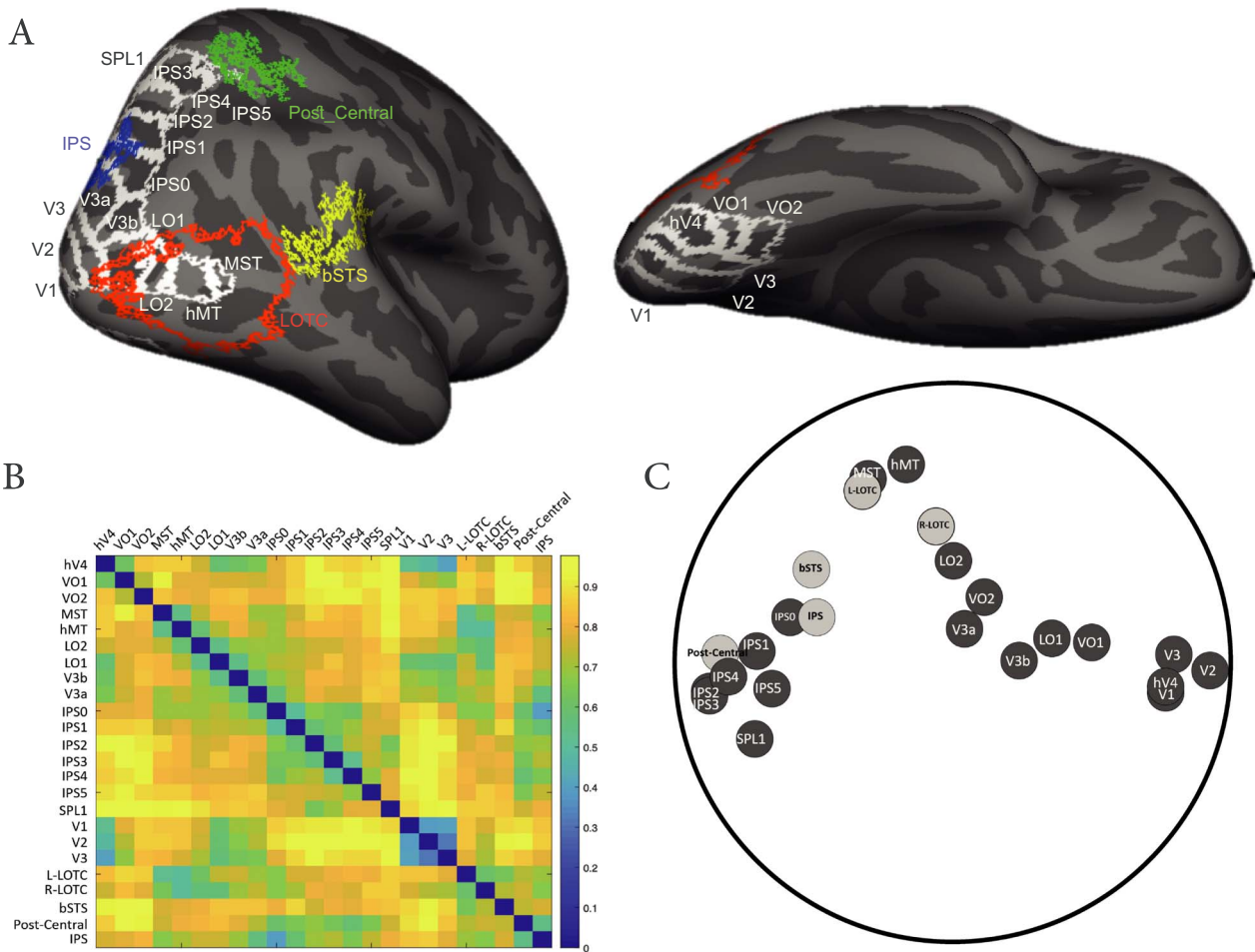
Fig. 5. (A) Inflated brain surface from a representative participant showing the ROIs examined in lateral and ventral views with white outlines for Wang's ROIs and colorful outlines for functional ROIs obtained from searchlight analysis. (B) ROI dissimilarity matrix based on the distance between their cross-position decoding accuracy for Wang's ROI and ROIs obtained from searchlight analysis. Lighter colors show higher distance. (C) MDS plot of Wang's ROIs (light gray) and ROIs obtained from searchlight analysis (dark gray) and a total least square regression line through the occipitotemporal regions (the light gray line) and another through the parietal regions (the dark gray line). The MDS plot and the lines could capture 0.5522% and 0.4798% of the variance of the ROI dissimilarity matrix.

In sum, these results reveal dissociations between the classification accuracies in the lateral occipitotemporal and parietal ROIs. In lateral occipitotemporal ROIs, the cross-position decoding accuracies are more correlated with behavioral ratings than any of the body-part movement models, while those in parietal ROIs are more correlated with the objective movement models.

ROI similarity analysis

In addition to the five clusters obtained from the searchlight analysis, we obtained the cross-position decoding accuracy of retinotopic regions of interest across occipitotemporal and parietal cortex (see Materials and methods). We calculated cross-position classification accuracies for these ROIs following a similar procedure as that used for the ROIs obtained from searchlight analysis. Then, to examine the representational similarity across all ROIs (including retinotopic ROIs and ROIs obtained from searchlight analysis), we computed the one minus the correlation between their cross-position decoding accuracy (Fig. 5B). The

resulting ROI dissimilarity matrix had high split-half reliability of 0.8346, significantly higher than chance (bootstrapped null distribution 95% confidence interval [−0.0011, 0.0012]). To visualize region-wise distances, we then applied an MDS analysis on this matrix. Figure 5C depicts the first two dimensions that captured most of the variance ($r^2 = 0.5522$). These results are consistent with the RSA and further suggest the presence of two separate networks for processing actions. One starting from the early visual cortex going to the lateral surface of the brain, and another splitting from this pathway toward the parietal cortex with the bSTS region landing right between the two pathways. This observation was quantified through fitting two total least square lines to the points on the MDS plot: one line was passed through occipitotemporal regions (V1, V2, V3, hV4, VO1, VO2, MST, hMT, LO2, LO1, V3b, and V3a from the Wang atlas, and L-LOTC, R-LOTC, and bSTS from our ROI analysis) and the other line was passed through parietal regions (IPS0, IPS1, IPS2, IPS3, IPS4, IPS5, and SPL1 from Wang atlas and IPS and postcentral from our ROI analysis). These lines

could explain 0.79% of the variance of the positions of the regions on the MDS plot and 0.48% of the variance of the full multidimensional space. We repeated the same analysis, including only the searchlight identified ROIs, and the result again showed a clear separation of parietal and occipitotemporal regions (Supplementary Fig. S6).

Discussion

In this article, we used PLDs of human actions presented in either the upper or the lower visual fields and examined the extent of position tolerance as well as the representational content of action-selective regions in the human brain. Using cross-position decoding analysis, we found that regions in the LOTC, right IPS, and right postcentral gyrus contain position-tolerant representations of action stimuli. Additionally, we investigated the representational content of these regions and found that representations in parietal regions were more related to the movements of the body parts, while those in occipitotemporal regions were more related to human subjective judgments about actions.

To the best of our knowledge, this is the first multivariate pattern analysis study that has used PLDs to characterize action representations in the human brain. Previous multivariate studies of action representation have used natural videos of actions (Hafri et al. 2017; Wurm, Caramazza, et al. 2017b; Tarhan and Konkle 2020). Employing natural action stimuli may have the problem that it is unclear what drives the responses. The responses may be related to actions in the videos, but other correlated features, such as objects, scenes, number of people in the videos, etc., could contribute to the responses. Using the PLD format allows us to restrict the visual content of action stimuli to the movements of the limbs and removes other confounding factors. Studies that use natural videos often reveal larger swaths of the cortex as action-selective compared to our results (Hafri et al. 2017; Wurm, Caramazza, et al. 2017b; Tarhan and Konkle 2020). These differences could be related to the fact that PLD stimuli are a more specific probe of action selectivity in the cortex. But these studies are also different from ours in that they did not directly test for position tolerance. In fact, our within-position classification does demonstrate action selectivity bilaterally in larger parts of the cortex. But this selectivity is much more restricted and mostly right-lateralized when carefully testing for position tolerance. Future studies that directly compare position-tolerant selectivity to PLD stimuli and natural videos on the same population of subjects could further elucidate the differences between responses to PLDs and videos.

We did not find a strong selectivity in left anterior inferior parietal lobe commonly found in studies of action observation. A careful survey of previous literature (Goldenberg and Spatt 2009; Buxbaum and Kalénine 2010; Grosbras et al. 2012) reveals that this region is mostly activated in the presence of hand actions and

hand/object interactions. In our PLD stimuli, we did not have hands, and the object interactions mostly involved visible arm and torso movements. Future studies with a focus on hand actions may be better suited to test the position tolerance in the left anterior IPL.

We found regions in the left parietal cortex that showed above-chance within-position classification but did not show a strong position tolerance in our cross-position classification analysis. Could this lack of position tolerance observed here be due to lower overall classification accuracies in this region? Note that the bSTS region has even lower within-position classification accuracies, yet it does show significant cross-position classification. Therefore, we are inclined to attribute the chance level cross-position accuracies in the left parietal cortex to the lack of position tolerance in these regions.

Our results, for the first time, provide a comprehensive understanding of position-tolerant action representations in the human visual cortex. Our findings in the lateral occipital cortex are consistent with those of Grossman et al. (2010) who applied fMRI adaptation and showed position-tolerant representation of actions in STS. However, they failed to find other position-invariant regions found in our study, especially the regions in the parietal cortex. This discrepancy may be related to differences in design. Our block design paradigm and the use of SVM classifiers may have provided us with a higher power and sensitivity to detect regions that contain position-tolerant action representation (Coutanche et al. 2016). Roth and Zohary (2015) investigated position invariance during observation of natural tool-grasping clips using MVPA (correlation and crossdecoding classifier). They observed a gradual loss of position information accompanied by enhancement of hand/tool identity along the posterior–anterior axis in the dorsal stream. This result is consistent with our finding for the parietal cortex, but they did not observe such a gradient in the lateral stream. Their focus on grasping actions may have lowered their chance of observing the full extent of position tolerance in the lateral occipitotemporal cortex.

After identifying the position-tolerant regions, we investigated their representational content. In the occipitotemporal cortex, we found that a model based on the behavioral ratings of similarity was the best predictor of the cross-position pattern classification accuracies. These results are consistent with a study by Tucciarelli et al. (2019), which used natural images of different actions and showed that responses in the LOTC were correlated with the behavioral ratings of semantic similarity between actions. Similar to our results, they also showed a lower correlation in LOTC with a model based on body movements. On the other hand, they failed to find a significant correlation with a body-part model in the parietal cortex. This could be because they used human ratings of body-part movements as opposed to the objective measures we used based on actual measures of body position. Humans may fail

to consider the movement of all body parts and only focus on the main effector of action for their ratings. In addition, they also employed an event-related design that has less power than our block design paradigm. Thus, our results extend their findings and further elucidate the differences between parietal and occipitotemporal action representations.

In a study with a large set of human action videos, Tarhan and Konkle (2020) used behavioral rating to construct an encoding model to investigate action representations in the human brain. Their behavioral ratings included questions related to body parts involved in the action as well as the action target. Using clustering analysis, they divided voxels into four networks, suggesting that one is driven by the social aspects of actions and four are driven by the scale of the interaction envelope. Consistent with our study, their clusters spanned both occipitotemporal and parietal regions, with notable differences in the response profile between the two networks. However, it is difficult to be more specific in comparing their results to our findings due to methodological differences. They used natural videos of everyday scenes containing actions as stimuli. Other than human bodies or body parts performing the action, their stimuli also contained objects and other scene elements. The use of natural stimuli makes it difficult to determine if the clusters reflect responses to actions or the scene elements correlated with the actions. Also, since we used a block design paradigm and aimed at determining the extent of position invariance in action-selective regions, we could only collect data from a small set of actions. Future studies with a larger set of controlled PLD stimuli could test the extent to which sociality and interaction envelope determine the responses in position-invariant action-selective regions in the brain.

Investigating the representations of actions in the dorsal stream, Buccino et al. (2001) suggested that parietal regions represent actions based on action effectors. Jastorff et al. (2010), Abdollahi et al. (2013), and Vannuscorps et al. (2018), on the other hand, provided evidence that action representations in parietal regions are organized based on the category of actions independent of the effectors. Here, we show that a model based on body-part movement shows high correlations with the pattern of classification accuracies in the parietal stream regions. Our body-part movement model was constructed based on the movement of all body parts, including the trunk and the limbs. This model included more information than just the action effector. High correlations of this model with the classification accuracies in the parietal regions suggest that parietal regions may contain a more holistic representation of body parts during action observation.

Even though lateral regions show higher correlations with the subjective model than the body-part model, their correlation with the body-part movement was still greater than zero. Therefore, lateral regions show signatures of both objective and subjective measures.

Nevertheless, a more detailed analysis of the body-part model separating the patterns of movement from average movements revealed further differences between the lateral occipitotemporal and parietal regions. While the parietal regions showed sensitivity to both pattern and average of movements, the lateral occipitotemporal regions showed only a correlation with a model based on patterns of movement. With the limited number of stimuli in our set, we cannot perform a more nuanced analysis of the differences in the sensitivity to various body parts between lateral occipitotemporal and parietal regions. Future experiments with an expanded set of PLD stimuli could reveal potential differences between the two networks in their body-part representations.

Looking at decoding for different supraordinate action categories would also be interesting. However, since the experiment was not designed for this purpose, there is an imbalance in the number of actions per category and the results of this analysis may be unreliable (see Supplementary Fig. S7 for decoding social actions, object-directed actions, and whole-body-motion actions).

In the domain of object representations, recent studies have revealed that both dorsal and ventral stream regions contain object information independent of the position of objects (Almeida et al. 2018; Vaziri-Pashkam and Xu 2019), but the data-driven analysis of object representation still reveals the separation of these dorsal and ventral regions in their representational content (Vaziri-Pashkam and Xu 2019). Here, focusing on action observation, we showed that action representations become progressively less position-dependent from posterior to anterior along both lateral occipitotemporal and parietal pathways. Moreover, employing either data-driven- (looking at ROI similarities) or model-based approaches, we observed a clear distinction between representational content in the two pathways. Looking more closely at the arrangement of ROIs in the MDS plot, we can see that ROIs along the lateral occipitotemporal surface follow a continuous trajectory from early visual areas, while parietal ROIs are separated away from both early visual and lateral ROIs. These results provide evidence for distinct representational content along the two pathways.

Embodied cognition theories argue that recognition of observed actions relies on sensory-motor simulation and taps into motor representation necessary for executing those actions (Rizzolatti and Craighero 2004; Pulvermüller 2013). Motor, premotor, and parietal regions that contain mirror neurons in macaque monkeys have been proposed as potential regions contributing to action recognition through motor simulation (Rizzolatti and Craighero 2004). In our study, although we did find action selectivity in the premotor cortex in the within-position classification analysis, these representations were not strongly position-tolerant. The representations in our parietal regions did not show correlation with the model based on subjective judgments on actions. These results suggest that the motor, premotor, and parietal regions are

unlikely to contribute to the subjective understanding of actions (Caramazza et al. 2014). Nevertheless, it is still possible that parietal regions would contribute to action observation during coupled motor interactions as well as motor imitations (Caspers et al. 2010) through the analysis of body-part movements. Our results suggest distinct roles for lateral occipitotemporal and parietal regions in action observation.

In summary, we found regions capable of representing highly abstract forms of observed human actions, namely the PLD video clips of actions across changes in position. These regions located in lateral occipitotemporal and parietal cortices contained distinct representational content. The lateral occipitotemporal regions reflected more strongly the human subjective knowledge about actions, while the parietal regions reflected only the objective bodily movements. These results suggest the existence of two distinct networks that contain abstract representations of human actions likely serving different purposes in the visual processing of actions.

Conclusions

Humans can recognize actions rapidly and accurately. To obtain this ability, the human brain should contain abstract representations of actions that are robust to changes in the position of the actor. Here, we measured fMRI responses to video clips of actions presented in different positions. We found multiple brain regions that contained position-tolerant action representations. These regions had distinct representational contents. The responses in occipito-temproal regions were more related to subjective knowledge about actions, and those in the parietal regions were more related to objective movements. These results suggest that the two networks play distinct roles in human action understanding.

Supplementary material

Supplementary material is available at *Cerebral Cortex* online.

References

Abdollahi RO, Jastorff J, Orban GA. Common and segregated processing of observed actions in human SPL. *Cereb Cortex*. 2013:23(11): 2734–2753.

Albertini D, Gerbella M, Lanzilotto M, Livi A, Maranesi M, Ferroni CG, Bonini L. Connectional gradients underlie functional transitions in monkey pre-supplementary motor area. *Prog Neurobiol*. 2020:184:101699.

Allefeld C, Görgen K, Haynes JD. Valid population inference for information-based imaging: from the second-level *t*-test to prevalence inference. *NeuroImage*. 2016:141:378–392.

Almeida J, Amaral L, Garcea FE, Aguiar de Sousa D, Xu S, Mahon BZ, Martins IP. Visual and visuomotor processing of hands and tools as a case study of cross talk between the dorsal and ventral streams. *Cogn Neuropsychol*. 2018:35(5–6):288–303.

Anzellotti S, Fairhall SL, Caramazza A. Decoding representations of face identity that are tolerant to rotation. *Cereb Cortex*. 2013:24(8): 1988–1995.

Barz F, Livi A, Lanzilotto M, Maranesi M, Bonini L, Paul O, Ruther P. Versatile, modular 3D microelectrode arrays for neuronal ensemble recordings: from design to fabrication, assembly, and functional validation in non-human primates. *J Neural Eng*. 2017:14(3):036010.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Methodol*. 1995:57(1):289–300.

Brainard DH. The psychophysics toolbox. *Spat Vis*. 1997:10(4): 433–436.

Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L, Gallese V, Seitz RJ, Zilles K, Rizzolatti G, Freund HJ. Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur J Neurosci*. 2001:13(2):400–404.

Buxbaum LJ, Kalénine S. Action knowledge, visuomotor activation, and embodiment in the two action systems. *Ann N Y Acad Sci*. 2010:1191:201.

Buxbaum LJ, Kyle KM, Tang K, Detre JA. Neural substrates of knowledge of hand postures for object grasping and functional object use: evidence from fMRI. *Brain Res*. 2006:1117(1):175–185.

Caggiano V, Fogassi L, Rizzolatti G, Casile A, Giese MA, Thier P. Mirror neurons encode the subjective value of an observed action. *Proceedings of the National Academy of Sciences*. 2012:109(29):11848–53.

Caramazza A, Anzellotti S, Strnad L, Lingnau A. Embodied cognition and mirror neurons: a critical assessment. *Annu Rev Neurosci*. 2014:37:1–15.

Caspers S, Zilles K, Laird AR, Eickhoff SB. ALE meta-analysis of action observation and imitation in the human brain. *NeuroImage*. 2010:50(3):1148–1167.

Chang CC, Lin CJ. LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol (TIST)*. 2011:2(3):27.

Cichy RM, Chen Y, Haynes JD. Encoding the identity and location of objects in human LOC. *NeuroImage*. 2011:54(3):2297–2307.

Cichy RM, Sterzer P, Heinzle J, Elliott LT, Ramirez F, Haynes JD. Probing principles of large-scale object representation: category preference and location encoding. *Hum Brain Mapp*. 2013:34(7): 1636–1651.

Coutanche MN, Solomon SH, Thompson-Schill SL. A meta-analysis of fMRI decoding: quantifying influences on human visual population codes. *Neuropsychologia*. 2016:82:134–141.

Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage*. 1999:9(2): 179–194.

DiCarlo JJ, Cox DD. Untangling invariant object recognition. *Trends Cogn Sci.* 2007:11(8):333–341.

Downing PE, Peelen MV, Wiggett AJ, Tew BD. The role of the extrastriate body area in action perception. *Soc Neurosci.* 2006:1(1):52–62.

El-Sourani N, Wurm MF, Trempler I, Fink GR, Schubotz RI. Making sense of objects lying around: how contextual objects shape brain activity during action observation. *NeuroImage.* 2018:167: 429–437.

Etzel JA, Zacks JM, Braver TS. Searchlight analysis: promise, pitfalls, and potential. *NeuroImage.* 2013:78:261–269.

Ferri S, Rizzolatti G, Orban GA. The organization of the posterior parietal cortex devoted to upper limb actions: an fMRI study. *Hum Brain Mapp.* 2015:36(10):3845–3866.

Giese MA, Poggio T. Neural mechanisms for the recognition of biological movements. *Nat Rev Neurosci.* 2003:4(3):179–192.

Goldenberg G, Spatt J. The neural basis of tool use. *Brain.* 2009:132(6): 1645–1655.

Grosbras MH, Beaton S, Eickhoff SB. Brain regions involved in human movement perception: a quantitative voxel-based meta-analysis. *Hum Brain Mapp.* 2012:3(2):431–454.

Grossman ED, Jardine NL, Pyles JA. fMR-adaptation reveals invariant coding of biological motion on human STS. *Front Hum Neurosci.* 2010:4:15.

Hafri A, Trueswell JC, Epstein RA. Neural representations of observed actions generalize across static and dynamic visual input. *J Neurosci.* 2017:37(11):3056–3071.

Haxby JV. Multivariate pattern analysis of fMRI: the early beginnings. *NeuroImage.* 2012:62(2):852–855.

Isik L, Koldewyn K, Beeler D, Kanwisher N. Perceiving social interactions in the posterior superior temporal sulcus. *Proc Natl Acad Sci.* 2017:114(43):E9145–E9152.

Jastorff J, Orban GA. Human functional magnetic resonance imaging reveals separation and integration of shape and motion cues in biological motion processing. *J Neurosci.* 2009:29(22): 7315–7329.

Jastorff J, Begliomini C, Fabbri-Destro M, Rizzolatti G, Orban GA. Coding observed motor acts: different organizational principles in the parietal and premotor cortex of humans. *J Neurophysiol.* 2010:104(1):128–140.

Johnson-Frey SH. The neural bases of complex tool use in humans. *Trends Cogn Sci.* 2004:8(2):71–78.

Kalénine S, Buxbaum LJ, Coslett HB. Critical brain regions for action recognition: lesion symptom mapping in left hemisphere stroke. *Brain.* 2010:133(11):3269–80.

Konen CS, Kastner S. Two hierarchically organized neural systems for object information in human visual cortex. *Nat Neurosci.* 2008:11(2):224.

Kourtzi Z, Kanwisher N. Activation in human MT/MST by static images with implied motion. *J Cogn Neurosci.* 2000:12(1): 48–55.

Kriegeskorte N, Mur M. Inverse MDS: inferring dissimilarity structure from multiple item arrangements. *Front Psychol.* 2012:3:1–13.

Kriegeskorte N, Mur M, Bandettini PA. Representational similarity analysis-connecting the branches of systems neuroscience. *Front Syst Neurosci.* 2008:2:4.

Lanzilotto M, Maranesi M, Livi A, Ferroni CG, Orban GA, Bonini L. Stable readout of observed actions from format-dependent activity of monkey's anterior intraparietal neurons. *Proc Natl Acad Sci.* 2020:117(28):16596–16605.

Leshinskaya A, Wurm MF, Caramazza A, Leshinskaya A. Concepts of actions and their objects. In: Gazzaniga M, Mangun GR, Poepped D, editors. *The cognitive neurosciences.* MIT Press; 2018. pp. 757–765

Lingnau A, Downing PE. The lateral occipitotemporal cortex in action. *Trends in cognitive sciences.* 2015:19(5):268–77.

Livi A, Lanzilotto M, Maranesi M, Fogassi L, Rizzolatti G, Bonini L. Agent-based representations of objects and actions in the monkey pre-supplementary motor area. *Proc Natl Acad Sci.* 2019:116(7): 2691–2700.

Maeda K, Ishida H, Nakajima K, Inase M, Murata A. Functional properties of parietal hand manipulation–related neurons and mirror neurons responding to vision of own hand action. *J Cogn Neurosci.* 2015:27(3):560–572.

Maranesi M, Livi A, Bonini L. Processing of own hand visual feedback during object grasping in ventral premotor mirror neurons. *J Neurosci.* 2015:35(34):11824–11829.

McMahon EG, Zheng CY, Pereira F, Gonzalez R, Ungerleider LG, Vaziri-Pashkam M. Subtle predictive movements reveal actions regardless of social context. *J Vis.* 2019:19(7):16–16.

Michels L, Lappe M, Vaina LM. Visual areas involved in the perception of human movement from dynamic form analysis. *Neuroreport.* 2005:16(10):1037–1041.

Mitchell TM, Hutchinson R, Niculescu RS, Pereira F, Wang XR, Just M, Newman S. Learning to decode cognitive states from brain images. *Mach Learn.* 2004:57:145–175.

Mur M, Ruff DA, Bodurka J, Bandettini PA, Kriegeskorte N. Face-identity change activation outside the face system:"release from adaptation" may not always indicate neuronal selectivity. *Cereb Cortex.* 2010:20(9):2027–2042.

Ogawa K, Inui T. Neural representation of observed actions in the parietal and premotor cortex. *NeuroImage.* 2011:56(2):728–735.

Oosterhof NN, Wiestler T, Downing PE, Diedrichsen J. A comparison of volume-based and surface-based multi-voxel pattern analysis. *NeuroImage.* 2011:56(2):593–600.

Oosterhof NN, Tipper SP, Downing PE. Viewpoint (in) dependence of action representations: an MVPA study. *J Cogn Neurosci.* 2012:24(4):975–989.

Oosterhof NN, Connolly AC, Haxby JV. CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU octave. *Front Neuroinform.* 2016:10:27.

Oram MW, Perrett DI. Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *J Neurophysiol.* 1996:76(1):109–129.

Peelen MV, Wiggett AJ, Downing PE. Patterns of fMRI activity dissociate overlapping functional brain areas that respond to biological motion. *Neuron.* 2006:49(6):815–822.

Pulvermüller F. Semantic embodiment, disembodiment or misembodiment? In search of meaning in modules and neuron circuits. *Brain Lang.* 2013:127(1):86–103.

Ramírez FM, Cichy RM, Allefeld C, Haynes JD. The neural code for face orientation in the human fusiform face area. *J Neurosci.* 2014:34(36):12155–12167.

Rizzolatti G, Craighero L. The mirror-neuron system. *Annu Rev Neurosci.* 2004:27:169–192.

Roth ZN, Zohary E. Position and identity information available in fMRI patterns of activity in human visual cortex. *J Neurosci.* 2015:35(33):11559–11571.

Schubotz RI, Wurm MF, Wittmann MK, von Cramon DY. Objects tell us what action we can expect: dissociating brain areas for retrieval and exploitation of action knowledge during action observation in fMRI. *Front Psychol.* 2014:5:636.

Senior C, Barnes J, Giampietroc V, Simmons A, Bullmore ET, Brammer M, David AS. The functional neuroanatomy of implicit-motion perception or 'representational momentum'. *Curr Biol.* 2000:10(1): 16–22.

Shmuelof L, Zohary E. A mirror representation of others' actions in the human anterior parietal cortex. *J Neurosci*. 2006:26:9736–9742.

Shmuelof L, Zohary E. Mirror-image representation of action in the anterior pari- etal cortex. *Nat Neurosci*. 2008:11:1267–1269.

Simone L, Bimbi M, Rodà F, Fogassi L, Rozzi S. Action observation activates neurons of the monkey ventrolateral prefrontal cortex. *Sci Rep*. 2017:7(1):1–13.

Tarhan L, Konkle T. Sociality and interaction envelope organize visual action representations. *Nat Commun*. 2020:11(1):1–11.

Tucciarelli R, Turella L, Oosterhof NN, Weisz N, Lingnau A. MEG multivariate analysis reveals early abstract action representations in the lateral occipitotemporal cortex. *J Neurosci*. 2015:35(49): 16034–16045.

Tucciarelli R, Wurm M, Baccolo E, Lingnau A. The representational space of observed actions. *elife*. 2019:8:e47686.

Urgen BA, Orban GA. The unique role of parietal cortex in action observation: functional organization for communicative and manipulative actions. *NeuroImage*. 2021:237:118220.

Urgen BA, Pehlivan S, Saygin AP. Distinct representations in occipitotemporal, parietal, and premotor cortex during action perception revealed by fMRI and computational modeling. *Neuropsychologia*. 2019:127:35–47.

Urgesi C, Candidi M, Avenanti A. Neuroanatomical substrates of action perception and understanding: an anatomic likelihood estimation meta-analysis of lesion-symptom mapping studies in brain injured patients. *Front Hum Neurosci*. 2014:8:344.

Vangeneugden J, Peelen MV, Tadin D, Battelli L. Distinct neural mechanisms for body form and body motion discriminations. *J Neurosci*. 2014:34(2):574–585.

Vannuscorps G, Wurm MF, Striem-Amit E, Caramazza A. Large-scale organization of the hand action observation network in individuals born without hands. *Cereb Cortex*. 2018:29(8):3434–3444.

Vanrie J, Verfaillie K. Perception of biological motion: a stimulus set of human point-light actions. *Behav Res Methods Instrum Comput*. 2004:36(4):625–629.

Vaziri-Pashkam M, Taylor J, Xu Y. Spatial frequency tolerant visual object representations in the human ventral and dorsal visual processing pathways. *Journal of cognitive neuroscience*. 2019:31(1): 49–63.

Vaziri-Pashkam M, Xu Y. An information-driven 2-pathway characterization of occipitotemporal and posterior parietal visual object representations. *Cereb Cortex*. 2019:29(5): 2034–2050.

Vingerhoets G, Stevens L, Meesdom M, Honoré P, Vandemaele P, Achten E. Influence of perspective on the neural correlates of motor resonance during natural action observation. *Neuropsychological Rehabilitation*. 2012:22(5):752–67.

Wang L, Mruczek RE, Arcaro MJ, Kastner S. Probabilistic maps of visual topography in human cortex. *Cereb Cortex*. 2014:25(10): 3911–3931.

Watson CE, Cardillo ER, Ianni GR, Chatterjee A. Action concepts in the brain: an activation likelihood estimation meta-analysis. *Journal of cognitive neuroscience*. 2013:25(8): 1191–205.

Wheaton KJ, Thompson JC, Syngeniotis A, Abbott DF, Puce A. Viewing the motion of human body parts activates different regions of premotor, temporal, and parietal cortex. *NeuroImage*. 2004:22(1): 277–288.

Wurm MF, Schubotz RI. What's she doing in the kitchen? Context helps when actions are hard to recognize. *Psychon Bull Rev*. 2017:24(2):503–509.

Wurm MF, Cramon DY, Schubotz RI. The context–object–manipulation triad: cross talk during action perception revealed by fMRI. *J Cogn Neurosci*. 2012:24(7):1548–1559.

Wurm MF, Artemenko C, Giuliani D, Schubotz RI. Action at its place: contextual settings enhance action recognition in 4-to 8-year-old children. *Dev Psychol*. 2017a:53(4):662.

Wurm MF, Caramazza A, Lingnau A. Action categories in lateral occipitotemporal cortex are organized along sociality and transitivity. *J Neurosci*. 2017b:37(3):562–575.