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

Elahe' Yargholi[1], Gholam-Ali Hossein-Zadeh[1,2], Maryam Vaziri-Pashkam[3]

[1] School of Cognitive Sciences, Institute for Research in Fundamental Sciences, Tehran, Iran

[2] School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran

[3] National Institute of Mental Health (NIMH), Laboratory of Brain and Cognition, MD, USA

Abstract

Humans can recognize others' actions in the social environment. This recognition ability is tolerant to drastic variations in the visual input caused by the movements of people in the environment. What neural underpinnings support this position-tolerant action recognition? In the present study, we aimed to identify regions in the brain that contain position-tolerant representations of actions and explore the representational content of these regions. We recorded fMRI data from twenty-two subjects while they observed video clips of ten different human actions in Point Light Display format. Each stimulus was presented in either the upper or the lower visual fields. We used multivoxel pattern analysis and a searchlight technique to identify brain regions that contain position-tolerant action representation. In a generalization test, linear support vector machine classifiers were trained with fMRI patterns in response to stimuli presented in one position and tested with stimuli presented in another position. Results showed above-chance classification in the left and right lateral occipitotemporal cortex, right inferior intraparietal sulcus, and right superior intraparietal sulcus. To investigate the representational content of these regions, we constructed models based on the objective measures of movements and human subjective judgments about actions. We then evaluated the brain similarity matrix from the cross-position classification analysis based on these models. Results showed cross-position classifications in the lateral occipito-temoiral ROIs was more strongly related to the subjective judgements while those in the dorsal parietal ROIs were more strongly related to the objective movements. An ROI representational similarity analysis futher revealed separation of dorsal and lateral regions. These results suggest the existence of two networks that contain abstract representations of human actions with distinct representational content.

Keywords: 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 view-point, 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: 1) contain information about various human actions, and 2) contain action representations that are tolerant to the variability in the visual input. Here, using functional MRI and multi-voxel pattern analysis, 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 that contain position, size, and view-point 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; Xu and Vaziri-Pashkam, 2020). 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 occipitotemporal and parietal cortices that respond to (Casper et al., 2010; Kalenine 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; Lignau and Downing, 2015; Ferri et al., 2015; Hafri et al., 2017; Wurm 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 view-point in both monkeys (Oram and Perrett, 1996; Vangeneugden et al., 2014; Caggiano et al., 2011; 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., 2016). Results of these human studies provide evidence for view-point tolerance in the lateral occipito-temporal (Grossman et al., 2010; Tucciarelli et al., 2015) and parietal cortices (Ogawa and Inui, 2011; Oosterhof et al., 2012). In contrast to view-point 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 multi-voxel pattern analysis (MVPA, Haxby, 2012) that directly look at the representational content of a region. Roth and Zohary (2015) employed MVPA (correlation and cross-decoding) 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 if 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 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 classifier to search for position tolerant representations of actions in the human brain. Finding these regions, we will then investigate their representational content using a representational similarity analysis (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 occipito-temporal 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 & Poggio, 2003; Peelen et al., 2006; Grossman et al., 2010; Michels et al., 2005; Downing et al., 2006; Vangeneugden et al., 2016 refs from Vangeneugden; Jastorff and Orban, 2009). 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 occipital cortex, Wurm 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 occipito-temporal cortex is correlated with the behavioral similarity judgments.

Investigations to the representational content of parietal action selective regions are fewer in number and their results 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 actions are represented in an effector independent manner in the inferior parietal lobe (Jastorff et al., 2010) as well as superior parietal lobe and interaparietal sulcus (Vingehoets 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 five large-scale networks exist in the human brain that 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; el-sourani et al., 2017; Wurm and Schubotz, 2017; Wurm et al., 2017a; Leshinskaya et al., 2018).

Here, employing fMRI multivoxel pattern analysis, we decoded the position invariant representation of human actions in point-light display (PLD) format. We used controlled stimuli in PLD format to restrict the visual information to the bodily movements and identify regions supporting abstract 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, to devise objective measures for constructing 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

25 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. All subjects were healthy and 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 Point Light Display (PLD) format: Crawl, Cycle, Jumping Jack (henceforth referred to as Jump), Peddle, Play tennis, Salute, Spade, Stir, Walk, and Wave (Figure 1). These actions were selected from a larger set human action videos in PLD format provided by Vanrie and Verfaillie (2004).
We used a block design paradigm. Each run included twenty 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 8 seconds. There was an 8-second blank period at the beginning, end and between stimulus blocks of each run. Each run lasted 328 seconds. Each subject completed one session of 10 runs.

In each stimulus block, there were four repetitions of the same human action video, each repetition lasted 1.5 seconds (frame rate 30) followed by a 0.5 second blank period. In a

randomly chosen repetition in each block, the size of the dots became 50% larger for one second. The point-lights, subtending ~0.25 degrees of visual angle, were in white color against a black background. The center of the stimuli was presented ~4 degrees 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 degrees in width and between 2.38 and 7.7 degrees in height.

The fixation point was a red circle of size ~0.25 degrees of visual angle. 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 multi-dimensional scaling (IMDS) method proposed by (Kriegeskorte & Mur, 2012). At the beginning of the experiment, one snapshot from each of the 10 action videos was presented along with a number 1-10 in two vertical columns at the left border of the screen. On a separate screen, they were provided with numbered action videos and they could watch them as many times as they needed. Participants 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 final Euclidean distances on the circle for each participant.

MRI methods

MRI data were collected at two centers: 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 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 mid-experiment 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 was 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 repetition time (TR) of 2 s, echo time (TE) 26, 90° flip angle, 30 transverse slices, and a voxel size of 3 × 3 × 4 mm^3. A high-resolution T1-weighted structural scan was also acquired from each participant using an MPRAGE pulse sequence (TR = 1800 ms, TE = 3.44 ms, inversion time = 1100 ms, 7° flip angle, 176 sagittal slices, and 1 × 1 × 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 (more than 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 in each subject's hemisphere: V1, V2, V3, hV4, VO1, VO2, MST, hMT, LO2, LO1, V3a, V3b, IPS0, IPS1, IPS2, IPS3, IPS4, IPS5, SPL1 (Figure 3).

Multivariate pattern analysis (MVPA)

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

***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 cross-validation procedure, samples (t-statistics of vertices) were partitioned to train and test sets. A linear Support Vector Machine (SVM) was employed to perform pairwise 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 all within position pairwise classification accuracies. 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 pairs of actions presented in one position and testing its accuracy in classifying the same pair of actions in the other position. 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, corrected for false discovery rate (Benjamini and Hochberg, 1995) at level q1. The second-level p-value for each vertex was then determined as

$$Pr\ [Binomial(n, R * q1/(n * m)) \geq C]$$

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).

The searchlight group-level statistical analysis results in clusters with significant cross-positions classification accuracy. Contiguous clusters that passed the threshold were selected as ROIs in which decoding of actions could be generalized across positions and used in the next analyses to explore the 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). To do this, 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 with the addition of feature selection. Different ROIs don't 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 didn't qualitatively differ without feature selection.

Exploring the representational content of the ROI

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 divideded the point lights into five groups: trunk (5 dots), left hand (2 dots), right hand (2 dots), left leg (2 dots), and right leg (2 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 five-element vector. Then, the Euclidean distance between these five-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 five-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-pattern 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

(dis)similarity. 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 Wang's ROIs using a similar procedure as that used for the ROIs obtained from 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 corresponding ROI and these distances were used to construct an ROI dissimilarity matrix. The ROI dissimilarity matrix was first computed for individual subjects and then averaged across subjects to acquire group level ROI dissimilarity matrix (Figure 8).

Split-half reliability of ROI dissimilarity matrix was also evaluated; we randomly divided subjects into two equal groups and correlated corresponding group-level ROI dissimilarity matrices as a measure of reliability. This measure was calculated for 10000 random split-half divisions and averaged to produce the 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 2 split-half groups and calculating the reliability for 10000 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 & Xu, 2018). The squared correlation (r^2) between two-dimensional distances in the MDS plots and the original distances in the multi-dimensional 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 occipito-temporal regions and the other passing through the positions of parietal regions. The variance explained by the fitted lines was also measured; two-dimensional 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 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 investigate the representational content of these regions using fMRI multi-voxel pattern analysis. A stimulus set of ten different human action videos in PLD format (Figure 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 support vector machine classifier, within- and cross-position decoding was then applied following a searchlight approach across the whole brain.

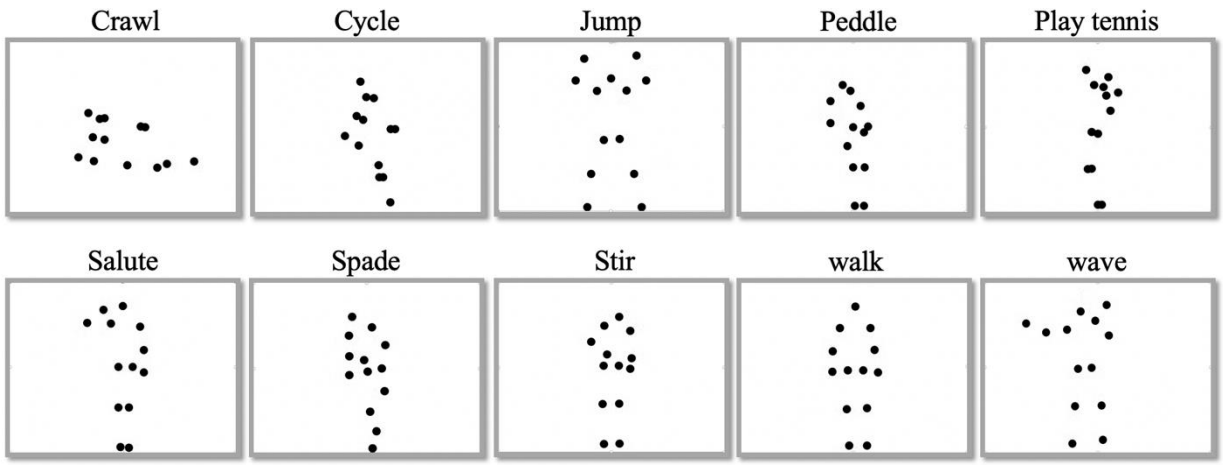

Figure 1. Representative frames of the 10 human action videos in PLD format used in the experiment.

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 for stimuli presented at the upper visual field and tested with fMRI responses for 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 2 A and B depict the result of the group analysis (q1 = 0.05 and q2 = 0.01) for within- and cross-position classification accuracy, respectively. Comparison of these maps reveals that only 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 occipito-temporal cortex, right inferior intraparietal sulcus, and right superior intraparietal sulcus (Figure 2 B).

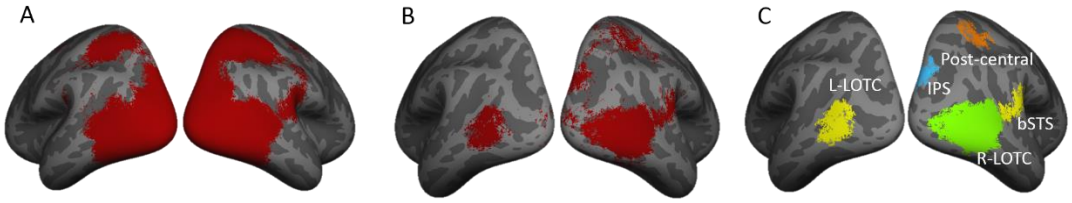

Figure 2. Results of searchlight classification accuracy ***A,*** Significance maps of group analysis for searchlight within-position classification accuracy (thresholded at q1 = 0.05 and q2 = 0.01) ***B***, Significance maps of group analysis for searchlight cross-position classification accuracy (group thresholded at 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).

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 methods) to extract ROIs with above chance cross-decoding accuracy. Five clusters (Figure 2 C) 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.

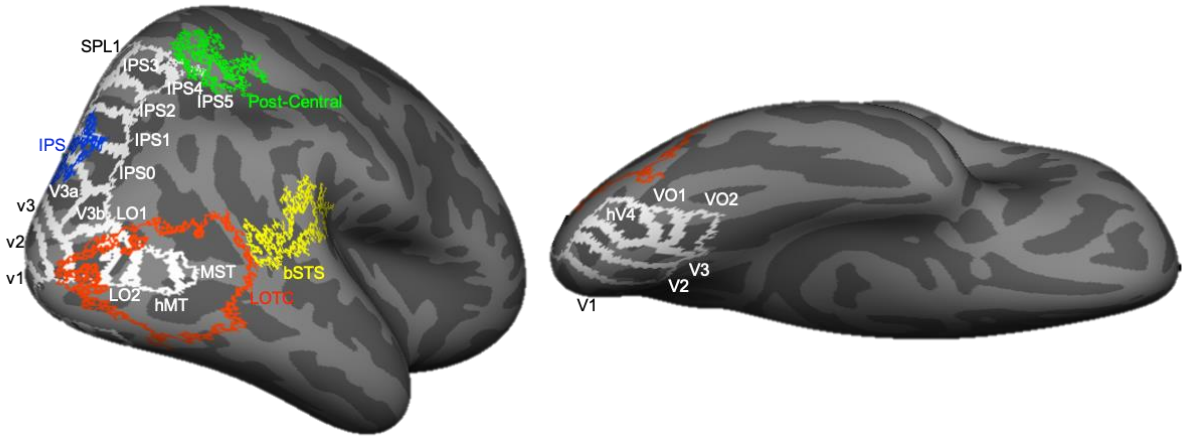

Figure 3. Inflated brain surface from a representative participant showing the ROIs examined in lateral and ventral views with white outlines for Wang's ROIs and colorful outlines for functional ROIs obtained from searchlight analysis.

A 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 4 illustrates the results of within- and cross-position decoding accuracy for each of the ROIs. All ROIs had a significant decoding accuracy for cross- and within-position (group analysis with ($q1 = 0.01$, $q2 = 0.01$). In the left and right lateral-occipito-temporal and parietal ROIs, cross-position decoding accuracy was significantly lower than within-position decoding accuracy (paired t-test, $p < 0.0075$, FDR corrected $q = 0.01$), while this difference was not significant in bSTS and Post-Central regions (paired t-test, $p > 0.13$, FDR corrected $q = 0.01$). These results suggest a reduction in position sensitivity from posterior to anterior regions in both occipito-temporal 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 figure 5. In line with the ROI analysis, this map demonstrates an increase in position tolerance from posterior to anterior regions in both occipito-temporal and parietal cortices.

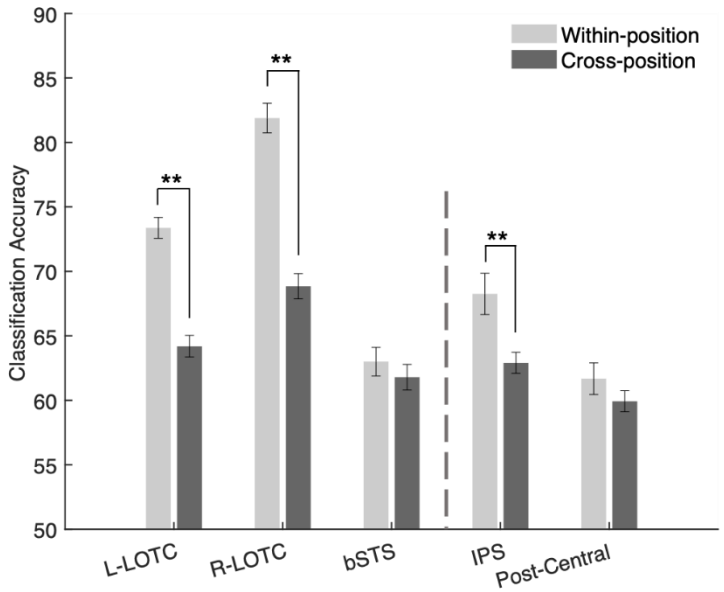

Figure 4. Within and cross-position decoding accuracy for the ROIs obtained from searchlight analysis. Error bars indicate standard errors of the means (** paired t-test, $p < 0.01$, FDR corrected). The vertical dashed line separates occipito-temporal ROIs from parietal ROIs.

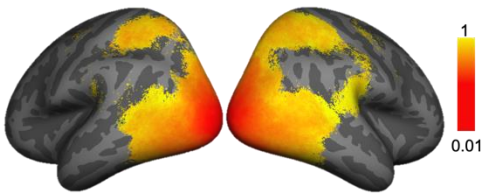

Figure 5. Maps of cross-position divided by within-position classification accuracy obtained from searchlight analysis in vertices with significant within-position classification accuracy.

Revealing the Representational content of the ROIs

To uncover the principles that govern the organization of action representations in each ROI, we used a representational similarity analysis (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 methods) and used correlation analysis to investigate the similarity of action representations between models and ROIs according to their cross-position pairwise decoding accuracies. The reliability of the behavioral dissimilarity matrices was high (Cronbach Alpha = 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-value = 0.2562).

Figure 6 shows the Kendall correlations between the models and cross-position pairwise decoding accuracies across participants for body-part movement and behavioral rating models.

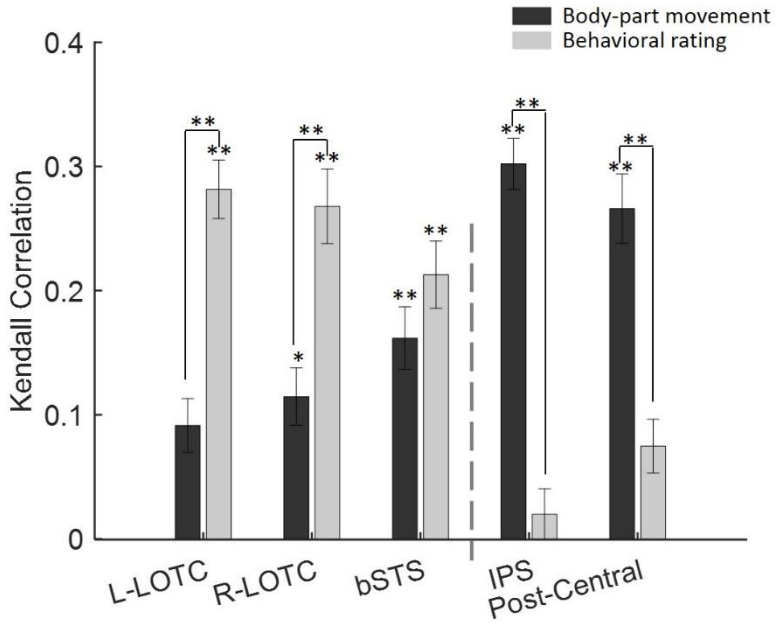

Figure 6. Kendall correlation body-part movement (dark gray) and behavioral rating models (light gray) and cross-position pairwise decoding accuracies for the individual ROIs. Error bars

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 occipito-temporal ROIs from parietal ROIs.

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 $p < 0.01$ for bSTS, IPS, Post-Central) and with the behavioral rating model in all ROIs except for IPS and Post-Central (permutation test, $p < 0.01$).

A two-way repeated-measures ANOVA with model (body-part movement and behavioral rating) and ROI as independent variables and correlation coefficient as the dependent variable was performed. The effect of ROI ($F(4,210) = 0.57$, $p = 0.6847$), and the effect of model ($F(1,210) = 1.05$, $p = 0.3072$) were not significant while the interaction between the two ($F(4,210) = 37.5$, $p = 0$) 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 post-central (all $t(21) > 3.7117$, all $p < 0.0016$, FDR-corrected, 5 comparisons).

We also constructed models based on the the average of movement across all body-parts (body-movement-average) and based on the pattern of movements in body parts (body-movement-pattern) separately (see methods). Figure 7 shows the average of Kendall correlation between the models and cross-position pairwise decoding accuracies across participants for body-movement-pattern and body-movement-average models. As figure 7 shows, all ROIs show 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 Post-Central, FDR corrected, 5 comparisons), but body-movement-average models only shows correlations with the dorsal ROIs (permutation test, $p < 0.01$, FDR corrected, 5 comparisons) and showed no significant correlation with the lateral ROIs (permutation test, $p > 0.05$, FDR corrected, 5 comparisons).

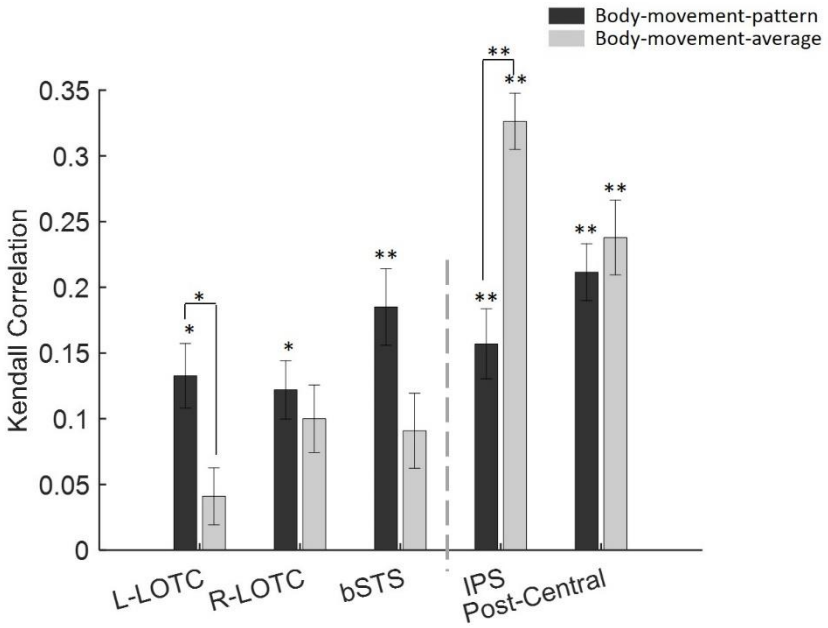

Figure 7. Kendall correlation between body-movement-average (light gray) and body-movement-pattern (dark gray) models and cross-position pairwise decoding accuracies for individual ROIs. Error bars indicate standard errors of the mean (permutation test, *p<0.05, **p<0.01, FDR corrected, 5 comparisons). The vertical dashed line separates occipito-temporal ROIs from parietal ROIs.

We next applied a two-way repeated measure ANOVA with model and ROI as independent variables and correlation coefficient as the dependent variable. Its result showed no significant effect of model ($F(1,210)=0.02, p=0.8778$) while the effect of ROI ($F(4,210)=15.1$, $p=0$), and interaction between the two ($F(4,210)=9.32$, $p = 0$) were significant. The significant effect of ROI points to greater correlation coefficients in dorsal 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, 5 comparisons) and for the average(body-part movement) model in IPS ($t(21) = 4.9394$, $p = 3.4583e-04$, FDR-corrected, 5 comparisons).

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

Its result for the body-part-average and behavioral rating models showed no significant effect of ROI ($F(4,210) = 0.59$, $p = 0.4421$), and model ($F(1,210) = 0.55$, $p = 0.7007$) while the interaction between the two was significant ($F(4,210) = 43.95$, $p = 0$). 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 Post-Central (all $t(21) > 3.5686$, all $p < 0.0018$, FDR-corrected, 5 comparisons).

For the body-part-pattern and behavioral rating models results showed no significant effect of model ($F(1,210) = 0.38$, $p = 0.5163$) while the effect of ROI ($F(4,210) = 8.2$, $p = 0$), and interaction between the two ($F(4,210) = 16.4$, $p = 0$) were 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-movement-pattern model in IPS and Post-Central (all $t(21) > 4.8254$, all $p < 1.1327e-04$, FDR-corrected, 5 comparisons).

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

ROI similarity analysis

In addition to the 5 clusters obtained from the searchlight analysis, we obtained the cross-position decoding accuracy of retinotopic regions of interest across occipito-temporal and parietal cortex (see 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 1 minus the correlation between their cross-position decoding accuracy (Figure 8). The resulting ROI dissimilarity matrix had a high split-half reliability of 0.8346, significantly higher than chance (bootstrapped null distribution 95% CI [−0.0011,0.0012]). To visualize region-wise distances we then applied an MDS analysis on this matrix. Figure 9 depicts the first two dimensions that captured most of the variance (r^2 = 0.5522). These results are consistent with the representational similarity analysis 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 towards the parietal cortex with the bSTS region landing right between the two pathways (the average classification accuracies of Wang's ROIs are provided in the supplemental figure S2). This observation was quantified through fitting two total least square lines to the points on the MDS plot: one line was passed through occipito-temporal regions (V1, V2, V3, hV4, VO1, VO2, MST, hMT, LO2, LO1, V3b, V3a, L-LOTC, R-LOTC, bSTS) and the other line was passed through parietal regions (IPS0, IPS1, IPS2, IPS3, IPS4, IPS5, SPL1, Parietal, Post-Central). These lines could explain 0.7950 percent of the variance of the positions of the regions on the MDS plot and 0.4798 percent of the variance of the full multi-dimensional space.

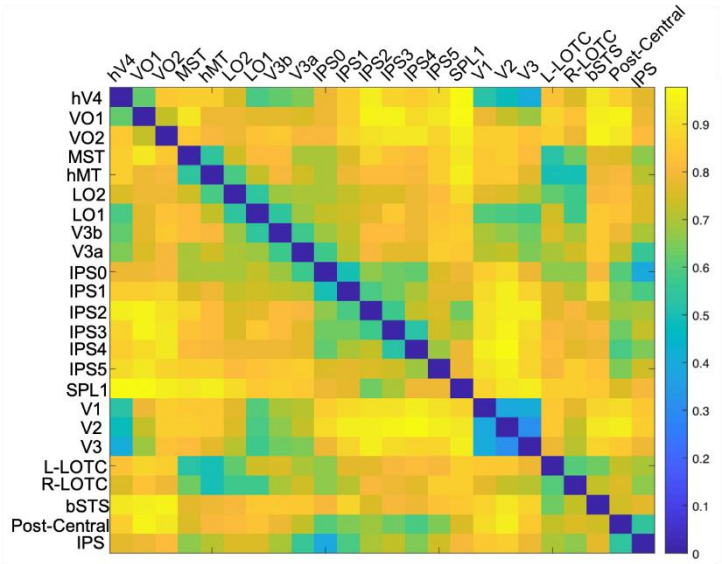

Figure 8. ROI dissimilarity matrix based on the correlation distance between their cross-position decoding accuracy for Wang's ROI and ROIs obtained from searchlight analysis. Lighter colors show higher correlations.

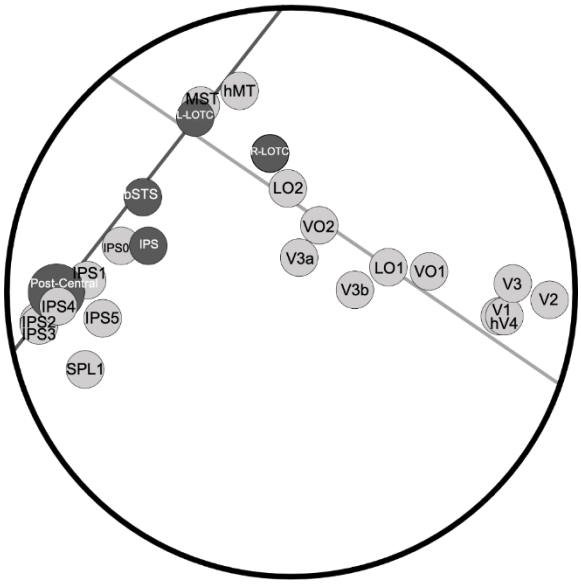

Figure 9. MDS plot of Wang's ROIs (light gray) and ROIs obtained from searchlight analysis (dark gray) and a total least square regression line through the occipito-temporal regions (the light gray line) and another through the parietal regions (the dark gray line). The MDS plot and the lines could capture 0.5522 and 0.4798 percent of the variance of the ROI dissimilarity matrix.

Discussion

In this article, we used point-light displays 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 Left and right lateral occipitotemporal cortex, right inferior intraparietal sulcus, and right superior intraparietal sulcus contain position tolerant representations of action stimuli. Additionally, we investigated the representational content of these regions and found that representations in parietal regions were related to the movements

of the body-parts while those in occipitotemporal regions were more related to human subjective judgments about actions.

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) that 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 cross-decoding 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 didn't observe such a gradient in the lateral stream. This may be due to the significant activation of parietal regions during observation of grasping actions (Culham et al., 2003).

Most regions that contained position tolerant representations of actions in our study were located in the right hemisphere. This laterality toward the right hemisphere is in line with the results of previous studies. Beauchamp et al. (2003) investigated human motion versus tool motion (natural and PLD video clips). They found that category-related responses were lateralized with a significantly greater number of tool-preferring voxels in the left hemisphere and a significantly greater number of human-preferring voxels in the right hemisphere. Grosbras et al. (2012) performed a meta-analyses (activation likelihood estimation with random effect) on three categories of human motion including the movement of the whole body, hands and face. The conjunction of the three meta-analysis maps showed convergence in the right posterior superior temporal sulcus and the bilateral junction between middle temporal and lateral occipital

gyri. Right laterality in parietal regions has also been previously observed for observed limb movements (Grosbras et al., 2012; Pelphrey et al., 2005).

After identifying the position-tolerant regions, we investigated their representational content. In occipito-temporal 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) that 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 just 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 occipito-temporal action representations.

Tarhan and Konkle (2020) used behavioral rating to construct the feature space of their encoding model which included the body parts engaged in each action and the action target. They compared the behavioral model with a low-level gist model in predicting fMRI responses. Their behavioral model outperformed the low-level model in both dorsal and ventral pathways (except the early visual cortex). However, the behavioral rating in Tarhan and Konkle (2020) is substantially different from our behavioral experiment. We asked participants to arrange observed actions according to perceived similarity and didn’t give them any instruction for this arrangement; they were not limited to thinking about specific aspects in the videos. Due to this fundamental difference, result comparison is not feasible.

Investigating the representations of actions in the dorsal stream Buccino et al. (2001) suggested that dorsal 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 representation in dorsal 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 dorsal 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 dorsal regions suggests that dorsal regions contain a more wholistic representation of body parts during action observation.

Eventhough lateral regions show higher correlations with the subjective model than the body part model, their correlation with the body part movement was still above zero. Therefore lateral regions show signatures of both objective and subjective measures. Nevertheless, more detailed analysis of the body part model spearting the patterns of movement from average movements revealed further differences between the dorsal and lateral regions. While the dorsal regions showed sensitivity to both pattern and average of movements, the lateral regions showed only a correlation with a model based on patterns of movement. These findings are preliminary as with the limited number of stimuli in our set we cannot perform a more neuanced analysis of the differences in the sensitivity to various body parts between dorsal and lateral regions. Future experiments with an expanded set of PLD stimuli could illucidate potential differences between the two networks in their body part representations.

In the domain of object representations, recent studies have revealed that both dorsal and ventral stream regions contain object information indepentend of position of objects (Vaziri-Pashkam and Xu, 2019; Almeida et al., 2018), but data-driven analysis of object representation still reveals the separation of these dorsal and ventral regions in their representational (Vaziri-Pashkam and Xu, 2019). Here, focusing on action observation, we observed that action

representations becomes progressively less position-dependent from posterior to anterior along both dorsal and lateral pathways. Moreover, employing either data-driven (looking at ROI similarities) or model-based approaches we observed a clear distinction between representational content in two pathways. Looking more closely at the arrangement of ROIs in the MDS plot, we can see that ROIs along the lateral surface follow a continuous trajectory from early visual areas while dorsal ROIs don't follow a similar arrangement and are located away from both early visual and lateral ROIs. These results provides 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 & 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 & Craighero, 2004). In our results we failed to find position tolerant represenations in prefrontal and premotor regions and the representions in our parietal regions did not show correlation with the model based on subjective judgements on actions. These results suggest that the motor, premotor and parietal regions are unlikely to contribute to human's 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 immitations (Caspers et al., 2010) through analysis of body part movements. Our results suggest distinct roles for dorsal and lateral regions in action observation.

In summary, we found regions capable of representing highly abstract forms of observed human actions, namely the point light display video clips of actions across changes in position. These regions located in lateral occipito-temporal and parietal cortices contained distinct representational content. The lateral regions reflected more strongly the human subjective knowledge about objects while the dorsal 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.

References

Abdollahi, R. O., Jastorff, J., & Orban, G. A. (2013). Common and segregated processing of observed actions in human SPL. *Cerebral cortex*, *23*(11), 2734-2753.

Albertini, D., Gerbella, M., Lanzilotto, M., Livi, A., Maranesi, M., Ferroni, C. G., & Bonini, L. (2020). Connectional gradients underlie functional transitions in monkey pre-supplementary motor area. *Progress in neurobiology*, *184*, 101699.

Allefeld, C., Görgen, K., & Haynes, J. D. (2016). Valid population inference for information-based imaging: From the second-level t-test to prevalence inference. *Neuroimage*, *141*, 378-392.

Almeida, J., Amaral, L., Garcea, F. E., Aguiar de Sousa, D., Xu, S., Mahon, B. Z., & Martins, I. P. (2018). Visual and visuomotor processing of hands and tools as a case study of cross talk between the dorsal and ventral streams. *Cognitive neuropsychology*, *35*(5-6), 288-303.

Anzellotti, S., Fairhall, S. L., & Caramazza, A. (2013). Decoding representations of face identity that are tolerant to rotation. *Cerebral Cortex*, *24*(8), 1988-1995.

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

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), 289-300.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, *10*(4), 433-436.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., ... & Freund, H. J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European journal of neuroscience*, *13*(2), 400-404.

Buxbaum, L. J., Kyle, K. M., Tang, K., & Detre, J. A. (2006). Neural substrates of knowledge of hand postures for object grasping and functional object use: Evidence from fMRI. *Brain Research*, *1117*(1), 175-185.

Caramazza, A., Anzellotti, S., Strnad, L., & Lingnau, A. (2014). Embodied cognition and mirror neurons: a critical assessment. *Annual review of neuroscience*, *37*, 1-15.

Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *Neuroimage, 50(3)*, 1148-1167.

Chang, C. C., & Lin, C. J. (2011). LIBSVM: A library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, *2*(3), 27.

Cichy, R. M., Chen, Y., & Haynes, J. D. (2011). Encoding the identity and location of objects in human LOC. *Neuroimage*, *54*(3), 2297-2307.

Cichy, R. M., Sterzer, P., Heinzle, J., Elliott, L. T., Ramirez, F., & Haynes, J. D. (2013). Probing principles of large-scale object representation: Category preference and location encoding. *Human Brain Mapping*, *34*(7), 1636-1651.

Coutanche, M. N., Solomon, S. H., & Thompson-Schill, S. L. (2016). A meta-analysis of fMRI decoding: Quantifying influences on human visual population codes. *Neuropsychologia*, *82*, 134-141.

Culham, J. C., Danckert, S. L., De Souza, J. F., Gati, J. S., Menon, R. S., & Goodale, M. A. (2003). Visually guided grasping produces fMRI activation in dorsal but not ventral stream brain areas. *Experimental brain research*, *153*(2), 180-189.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in cognitive sciences*, *11*(8), 333-341.

Downing, P. E., Peelen, M. V., Wiggett, A. J., & Tew, B. D. (2006). The role of the extrastriate body area in action perception. *Social Neuroscience*, *1*(1), 52-62.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage*, *9*(2), 179-194.

El-Sourani, N., Wurm, M. F., Trempler, I., Fink, G. R., & Schubotz, R. I. (2018). Making sense of objects lying around: How contextual objects shape brain activity during action observation. *Neuroimage*, *167*, 429-437.

Etzel, J. A., Zacks, J. M., & Braver, T. S. (2013). Searchlight analysis: promise, pitfalls, and potential. *Neuroimage, 78*, 261-269.

Ferri, S., Rizzolatti, G., & Orban, G. A. (2015). The organization of the posterior parietal cortex devoted to upper limb actions: An fMRI study. *Human brain mapping*, *36*(10), 3845-3866.

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*, *4*(3), 179-192.

Grosbras, M. H., Beaton, S., & Eickhoff, S. B. (2012). Brain regions involved in human movement perception: A quantitative voxel-based meta-analysis. *Human brain mapping*, *33*(2), 431-454.

Grossman, E. D., Jardine, N. L., & Pyles, J. A. (2010). fMR-adaptation reveals invariant coding of biological motion on human STS. *Frontiers in human neuroscience*, *4*, 15.

Hafri, A., Trueswell, J. C., & Epstein, R. A. (2017). Neural representations of observed actions generalize across static and dynamic visual input. *Journal of Neuroscience*, *37*(11), 3056-3071.

Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: the early beginnings. *Neuroimage*, *62*(2), 852-855.

Isik, L., Koldewyn, K., Beeler, D., & Kanwisher, N. (2017). Perceiving social interactions in the posterior superior temporal sulcus. *Proceedings of the National Academy of Sciences*, *114*(43), E9145-E9152.

Jastorff, J., & Orban, G. A. (2009). Human functional magnetic resonance imaging reveals separation and integration of shape and motion cues in biological motion processing. *Journal of Neuroscience*, *29*(22), 7315-7329.

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

Johnson-Frey, S. H. (2004). The neural bases of complex tool use in humans. *Trends in cognitive sciences*, *8*(2), 71-78.

Kalénine, S., Buxbaum, L. J., & Coslett, H. B. (2010). Critical brain regions for action recognition: lesion symptom mapping in left hemisphere stroke. *Brain, 133(11),* 3269-3280.

Konen, C. S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nature neuroscience*, *11*(2), 224.

Kourtzi, Z., & Kanwisher, N. (2000). Activation in human MT/MST by static images with implied motion. *Journal of cognitive neuroscience*, *12*(1), 48-55.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, *2*, 4.

Kriegeskorte, N. & Mur, M. (2012). Inverse MDS: Inferring Dissimilarity Structure from Multiple Item Arrangements. *Front. Psychol. 3*, 1–13.

Lanzilotto, M., Maranesi, M., Livi, A., Ferroni, C. G., Orban, G. A., & Bonini, L. (2020). Stable readout of observed actions from format-dependent activity of monkey's anterior intraparietal neurons. *Proceedings of the National Academy of Sciences*, *117*(28), 16596-16605.

Leshinskaya, A., Wurm, M. F., Caramazza, A., & Leshinskaya, A. (2018). Concepts of Actions and their Objects. M Gazzaniga, GR Mangun, D Poepped. *The Cognitive Neurosciences*, 757-765.

Lingnau, A., & Downing, P. E. (2015). The lateral occipitotemporal cortex in action. *Trends in cognitive sciences, 19(5),* 268-277.

Livi, A., Lanzilotto, M., Maranesi, M., Fogassi, L., Rizzolatti, G., & Bonini, L. (2019). Agent-based representations of objects and actions in the monkey pre-supplementary motor area. *Proceedings of the National Academy of Sciences*, *116*(7), 2691-2700.

Maeda, K., Ishida, H., Nakajima, K., Inase, M., & Murata, A. (2015). Functional properties of parietal hand manipulation–related neurons and mirror neurons responding to vision of own hand action. *Journal of cognitive neuroscience*, *27*(3), 560-572.

Maranesi, M., Livi, A., & Bonini, L. (2015). Processing of own hand visual feedback during object grasping in ventral premotor mirror neurons. *Journal of Neuroscience*, *35*(34), 11824-11829

McMahon, E. G., Zheng, C. Y., Pereira, F., Gonzalez, R., Ungerleider, L. G., & Vaziri-Pashkam, M. (2019). Subtle predictive movements reveal actions regardless of social context. *Journal of vision*, *19*(7), 16-16.

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

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

Mur, M., Ruff, D. A., Bodurka, J., Bandettini, P. A., & Kriegeskorte, N. (2010). Face-identity change activation outside the face system:“release from adaptation” may not always indicate neuronal selectivity. *Cerebral Cortex*, *20*(9), 2027-2042.

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

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. *Frontiers in neuroinformatics*, *10*, 27.

Oosterhof, N. N., Tipper, S. P., & Downing, P. E. (2012). Viewpoint (in) dependence of action representations: an MVPA study. *Journal of Cognitive Neuroscience*, *24*(4), 975-989.

Oosterhof, N. N., Wiestler, T., Downing, P. E., & Diedrichsen, J. (2011). A comparison of volume-based and surface-based multi-voxel pattern analysis. *Neuroimage*, *56*(2), 593-600.

Oram, M. W., & Perrett, D. I. (1996). Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *Journal of neurophysiology*, *76*(1), 109-129.

Peelen, M. V., Wiggett, A. J., & Downing, P. E. (2006). Patterns of fMRI activity dissociate overlapping functional brain areas that respond to biological motion. *Neuron*, *49*(6), 815-822.

Pelphrey, K. A., Morris, J. P., Michelich, C. R., Allison, T., & McCarthy, G. (2005). Functional anatomy of biological motion perception in posterior temporal cortex: an fMRI study of eye, mouth and hand movements. Cerebral cortex, 15(12), 1866-1876.

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

Ramírez, F. M., Cichy, R. M., Allefeld, C., & Haynes, J. D. (2014). The neural code for face orientation in the human fusiform face area. *Journal of Neuroscience*, *34*(36), 12155-12167.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. Annu. Rev. Neurosci., 27, 169-192.

Roth, Z. N., & Zohary, E. (2015). Position and identity information available in fMRI patterns of activity in human visual cortex. *Journal of Neuroscience*, *35*(33), 11559-11571.

Sartori, L., Begliomini, C., Castiello, U. (2013). Motor resonance in left- and right-handers: evidence for effector-independent motor representations. *Front. Hum. Neurosci. 7, 33*.

Schubotz, R. I., Wurm, M. F., Wittmann, M. K., & von Cramon, D. Y. (2014). Objects tell us what action we can expect: dissociating brain areas for retrieval and exploitation of action knowledge during action observation in fMRI. *Frontiers in psychology*, *5*, 636.

Senior, C., Barnes, J., Giampietroc, V., Simmons, A., Bullmore, E. T., Brammer, M., & David, A. S. (2000). The functional neuroanatomy of implicit-motion perception or 'representational momentum'. *Current Biology*, *10*(1), 16-22.

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

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

Simone, L., Bimbi, M., Rodà, F., Fogassi, L., & Rozzi, S. (2017). Action observation activates neurons of the monkey ventrolateral prefrontal cortex. *Scientific reports*, *7*(1), 1-13.

Tarhan, L., & Konkle, T. (2020). Sociality and interaction envelope organize visual action. representations. *Nature Communications*, *11*(1), 1-11.

Tucciarelli, R., Turella, L., Oosterhof, N. N., Weisz, N., & Lingnau, A. (2015). MEG multivariate analysis reveals early abstract action representations in the lateral occipitotemporal cortex. *Journal of Neuroscience*, *35*(49), 16034-16045.

Tucciarelli, R., Wurm, M., Baccolo, E., & Lingnau, A. (2019). The representational space of observed actions. *Elife, 8*, e47686.

Urgen, B. A., Pehlivan, S., & Saygin, A. P. (2019). Distinct representations in occipito-temporal, parietal, and premotor cortex during action perception revealed by fMRI and computational modeling. *Neuropsychologia, 127*, 35-47.

Urgen, B. A., & Orban, G. A. (2021). The unique role of parietal cortex in action observation: Functional organization for communicative and manipulative actions. *bioRxiv*.

Urgesi, C., Candidi, M., & Avenanti, A. (2014). Neuroanatomical substrates of action perception and understanding: an anatomic likelihood estimation meta-analysis of lesion-symptom mapping studies in brain injured patients. *Frontiers in human neuroscience, 8*, 344.

Vangeneugden, J., Peelen, M. V., Tadin, D., & Battelli, L. (2014). Distinct neural mechanisms for body form and body motion discriminations. *Journal of Neuroscience*, *34*(2), 574-585.

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

Vanrie, J., & Verfaillie, K. (2004). Perception of biological motion: A stimulus set of human point-light actions. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 625-629.

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

Wang, L., Mruczek, R. E., Arcaro, M. J., & Kastner, S. (2014). Probabilistic maps of visual topography in human cortex. *Cerebral cortex*, *25*(10), 3911-3931.

Watson, C. E., Cardillo, E. R., Ianni, G. R., & Chatterjee, A. (2013). Action concepts in the brain: an activation likelihood estimation meta-analysis. *Journal of cognitive neuroscience*, *25*(8), 1191-1205.

Wheaton, K. J., Thompson, J. C., Syngeniotis, A., Abbott, D. F., & Puce, A. (2004). Viewing the motion of human body parts activates different regions of premotor, temporal, and parietal cortex. *Neuroimage*, *22*(1), 277-288.

Wurm, M. F., Artemenko, C., Giuliani, D., & Schubotz, R. I. (2017a). Action at its place: Contextual settings enhance action recognition in 4-to 8-year-old children. *Developmental psychology*, *53*(4), 662.

Wurm, M. F., Caramazza, A., & Lingnau, A. (2017b). Action categories in lateral occipitotemporal cortex are organized along sociality and transitivity. *Journal of Neuroscience*, *37*(3), 562-575.

Wurm, M. F., Cramon, D. Y., & Schubotz, R. I. (2012). The context–object–manipulation triad: Cross talk during action perception revealed by fMRI. *Journal of Cognitive Neuroscience*, *24*(7), 1548-1559.

Wurm, M. F., & Schubotz, R. I. (2017). What's she doing in the kitchen? Context helps when actions are hard to recognize. *Psychonomic bulletin & review*, *24*(2), 503-509.

Xu, Y., & Vaziri-Pashkam, M. (2020). The development of transformation tolerant visual representations differs between the human brain and convolutional neural networks. *bioRxiv*.