Received: 30 August 2023 | Revised: 21 December 2023 | Accepted: 12 January 2024

DOI: 10.1002/hbm.26605

RESEARCH ARTICLE

WILEY

Overlapping representations of observed actions and action-related features

Zuzanna Kabulska | Tonghe Zhuang | Angelika Lingnau

Faculty of Human Sciences, Institute of Psychology, Chair of Cognitive Neuroscience, University of Regensburg, Regensburg, Germany

Correspondence
Angelika Lingnau, Faculty of Human Sciences, Institute of Psychology, Chair of Cognitive Neuroscience, University of Regensburg, Regensburg, Germany.
Email: angelika.lingnau@ur.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Numbers: LI 2840/1-1, LI 2840/2-1

Abstract

The lateral occipitotemporal cortex (LOTC) has been shown to capture the representational structure of a smaller range of actions. In the current study, we carried out an fMRI experiment in which we presented human participants with images depicting 100 different actions and used representational similarity analysis (RSA) to determine which brain regions capture the semantic action space established using judgments of action similarity. Moreover, to determine the contribution of a wide range of action-related features to the neural representation of the semantic action space we constructed an action feature model on the basis of ratings of 44 different features. We found that the semantic action space model and the action feature model are best captured by overlapping activation patterns in bilateral LOTC and ventral occipitotemporal cortex (VOTC). An RSA on eight dimensions resulting from principal component analysis carried out on the action feature model revealed partly overlapping representations within bilateral LOTC, VOTC, and the parietal lobe. Our results suggest spatially overlapping representations of the semantic action space of a wide range of actions and the corresponding action-related features. Together, our results add to our understanding of the kind of representations along the LOTC that support action understanding.

KEYWORDS
LOTC, multi-dimensional action space, representational space

1 | INTRODUCTION

We are constantly surrounded by various types of actions and can recognize them without effort. However, understanding them is a complex task, relying on multiple sources of information. One of the key challenges is unraveling the mental representations of actions and the degree to which these explain behavior. A growing number of recent studies suggest that actions can be depicted as data points in a multidimensional action space (e.g., Dima et al., 2022; Kabulska & Lingnau, 2022; Lingnau & Downing, 2023; Thornton & Tamir, 2022; Tucciarelli et al., 2019; Watson & Buxbaum, 2014), in line with corresponding ideas in the object perception literature (Beymer & Poggio, 1997; Edelman, 1998; Kriegeskorte et al., 2008). Understanding the dimensions underlying this action space and the corresponding neural implementation thus is key to understanding the human ability to perceive and recognize actions.

The dimensions spanning the space of actions have been investigated by several behavioral studies. For instance, in the realm of tool usage, Watson and Buxbaum (2014) demonstrated that tools can be sorted into distinct groups based on two dimensions: one associated

with the hand configuration and the other with the magnitude of the arm movement. Tucciarelli et al. (2019) showed that daily-life actions can be mapped onto dimensions reflecting the type of change induced by the action, and the type of need to be fulfilled by the actions (ranging from basic, physiological needs to higher social needs). Furthermore, social importance has emerged as a prominent factor in various other studies, either as the main factor in judgment of action similarity (Dima et al., 2022) or as one of the factors, together with semantic dimensions (e.g., food, work, home life) and visual information (scene setting; Dima et al., 2023). A recent study of Vinton et al. (2023) suggested that actions might be projected onto four dimensions: two related to facial traits and emotions (e.g., friendly—unfriendly) and two others unique to actions (e.g., planned—unplanned). Another important dimension that emerged is the actor's goals (Tarhan et al., 2021). Additionally, using large text data, Thornton and Tamir (2022) revealed six abstract dimensions including *Abstraction*, *Creation*, and *Spiritualism*. Finally, Kabulska and Lingnau (2022) highlighted the importance of the valence of an action, that is, the differentiation between pleasant (e.g., sport-related) and unpleasant (e.g., aggressive) actions.

Brain areas that play a role in action recognition should capture the similarity structure between actions, i.e. actions that are judged to be more similar to each other should also be more similar to each other with respect to the corresponding activity patterns (see also Lingnau & Downing, 2023; Tucciarelli et al., 2019; Wurm & Caramazza, 2022). To reveal areas with such properties, several previous studies examined the neural representations of a semantic action space established on the basis of action similarity judgments as well as the underlying action dimensions. As an example, Tucciarelli et al. (2019) demonstrated that a semantic action space model of 27 different actions depicted as static images is captured by patterns of activation in the lateral occipitotemporal cortex (LOTC). Likewise, using videos of 60 different actions, Tarhan et al. (2021) obtained significant correlations between neural activation patterns obtained during the observation of different actions depicted as videos and a semantic action space model along the ventral and dorsal streams, primary and premotor cortex and the medial parietal lobe. Finally, Zhuang et al. (2023) showed that a semantic action space model for 12 different actions (presented as static images), organized into three taxonomic levels, is reflected by patterns of activation in the LOTC and the superior parietal lobule, with the highest similarity for actions at the basic level.

Several studies examined the neural representations of features and dimensions underlying the organization of observed actions. As an example, Tarhan and Konkle (2020b) revealed five large-scale brain networks associated with action processing: one dedicated to social aspects of actions (such as targeting an agent), and four pertained to a "scale of space" (i.e., near space/far space). Tarhan et al. (2021) proposed a hierarchy in processing actions along the posterior-to-anterior lateral surface of the visual cortex, ranging from information about visual aspects of actions, followed by movement-related information and, finally, the goals of actions, in line with the results of a recent EEG study by Dima et al. (2023). Furthermore, superior and inferior portions of the LOTC have been shown to carry information about actions along the dimensions sociality and transitivity, respectively (Wurm et al., 2017). Overall, these findings contribute to our understanding of the neural substrates underlying the representation of visually presented actions in the human brain. However, most previous studies either used a small set of preselected dimensions, or a rather small stimulus set, which might restrict our understanding of action representation in a real-world environment (for an exception, see the study by Thornton & Tamir, 2022, which, however, was based on large-scale text corpora).

In the current study, we aimed to directly compare the neural representation of a semantic action space model established behaviorally based on action similarity judgments with a model that captures features related to these actions (action feature model). Moreover, we aimed to reveal potential dimensions underlying the organization of these features. For that purpose, we carried out an fMRI experiment in which we presented participants with 100 different actions (four exemplars each). We constructed the semantic action space model on the basis of data resulting from a multi-arrangement task (Kriegeskorte & Mur, 2012) on 100 different actions (Kabulska & Lingnau, 2022). To be able to determine to which degree the neural representation of the semantic action space can be accounted for by a range of different action-related features, we constructed an action feature model on the basis of ratings of 44 different features (see Kabulska & Lingnau, 2022, for details). Finally, in order to determine the dimensions underlying the action feature model, we employed principal component analysis (PCA) and investigated the neural representations of the resulting dimensions (see also Tamir et al., 2016; Tamir & Thornton, 2018; Thornton & Tamir, 2022).

2 | MATERIALS AND METHODS

2.1 | Participants

Here, 23 right-handed participants (11 males; mean age, 23; age range 20–34) participated in the study. All participants had normal or corrected-to-normal vision and no history of neurological or psychiatric disease. Data of three participants were not included in the data analysis due to excessive head motion (translation/rotation bigger than 3 mm; two participants and due to stopping the scan after 5 runs; one participant). The experimental protocol was approved by the ethics committee at the University of Regensburg. Written consent was obtained from all participants before the experiment. Participants were rewarded for taking part in the study.

2.2 | Stimuli

Stimuli consisted of 400 colored images of daily actions that portrayed 100 different daily actions in front of a naturalistic background, such as *running*, *biking*, and *eating* (same as in Kabulska & Lingnau, 2022; for a complete list of stimuli, see Table S1; see

Figure 1 for examples), with four different exemplars per action. Stimuli were carefully chosen on the basis of the following criteria: (a) actions were clearly visible, (b) no other distracting actions were depicted in the image, and (c) the action was embedded in a natural background. The stimulus set was collected from www.shutterstock.de. All selected images were in landscape orientation and were cropped to 600 × 400 pixels. The full set of images used in the study is shown in Figure S1.

2.3 | Experimental design and task

We used a rapid event-related design (see Figure 1) adopting the design used by Tucciarelli et al. (2019). There were eight functional runs in total (approximately 9 min each). Each run started and ended with 12 s fixation period. Each functional run consisted of 100 experimental trials, 20 null trials (4 s long each), and 7 catch trials during which the same action (but not the same exemplar) was shown as during the previous trial. The order of experimental trials was randomized, whereas null trials and catch trials were pseudorandomly interspersed between experimental trials, preventing two consecutive null trials and two consecutive catch trials.

Each trial consisted of an action image (1 s) with a superimposed central fixation cross, displayed on a uniform gray background, followed by a fixation cross (3 s). Each action was presented once in a run in a random order. Throughout the scanning session, each exemplar was shown twice (each in a separate run). Throughout the experiment, participants performed a one-back task. Prior to entering the scanner, they received written instructions, asking them to attentively watch the images while keeping their eyes at fixation and to press a button with the right index finger whenever there was a consecutive repetition of the same action. Responses during these catch trials were used offline to calculate response time and accuracy (see Results: Behavioral data analysis). To ensure that participants understood the task, they completed a practice run before entering the scanner.

Inside the scanner, stimuli were back-projected onto a screen (resolution 1024 × 768 at 60 Hz; viewing distance 106 cm, 12.98 × 8.53 degree of visual angle) and viewed via a mirror mounted on the radiofrequency (RF) coil. Stimulus presentation and response collection were controlled with *A Simple Framework* (Schwarzbach, 2011), a toolbox based on the MATLAB Psychtoolbox-3 for Windows (Brainard, 1997).

2.4 | Post-session questionnaire

At the end of the experiment, participants filled out a questionnaire in which they were asked to judge on a 6-point Likert scale how (1) comfortable and (2) tired they felt inside the scanner, (3) to which degree they internally verbalized the actions presented in the pictures, and (4) to which degree they concentrated exclusively on the repetition of the actions.

2.5 | Data acquisition

Functional and structural data were collected using a 3 T Siemens Prisma MRI scanner and a 64-channel RF head coil at the University of Regensburg. Functional images were acquired with a T2*-weighted gradient echoplanar imaging sequence (voxel resolution: 2.5 × 2.5 × 2.5 mm; 60 axial slices that cover the whole brain; repetition time [TR]: 2 s, echo time [TE]: 30s, flip angle [FA]: 75°, field of view: 192 mm, matrix size: 96 × 96, 265 volumes per run). Structural T1-weighted images were acquired halfway through the scanning session (i.e., after the fourth functional run) using an MPRAGE sequence (voxel resolution: 1 × 1 × 1 mm, 160 axial slices, TR: 1910 ms, TE: 3.67 s, FA: 9°, matrix size: 256 × 256).

2.6 | Data analysis

Data preprocessing and univariate analyses were performed using FEAT (FMRI Expert Analysis Tool; Woolrich et al., 2004; Woolrich et al., 2001) which is a part of FSL (FMRIB's Software Library, Jenkinson et al., 2012). FSL was also used for the extraction of information about the clusters of the statistical maps (command: *cluster*), creating ROIs, smoothing the maps and performing high-pass filtering

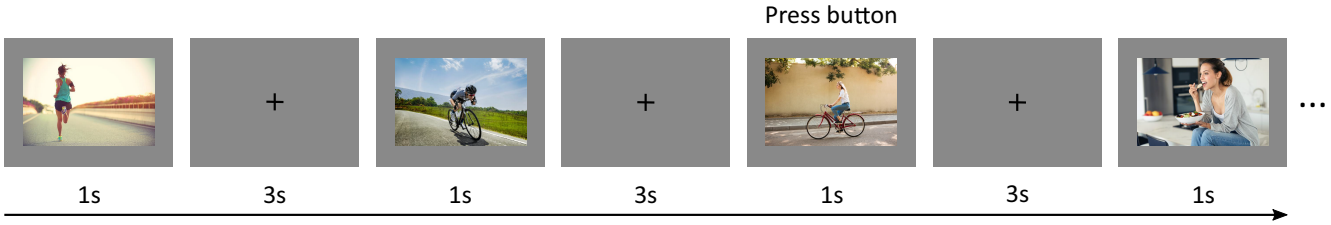

FIGURE 1 Example trial sequence and experimental design. We conducted an fMRI experiment using a rapid event-related design. Each trial consisted of the presentation of an image depicting an action (e.g., running, biking, eating; 1 s) followed by a gray screen (3 s). Throughout the experiment, a central fixation cross was presented on the screen. Participants were instructed to attentively observe the actions while keeping their eyes at fixation and to press a button with their right index finger whenever they saw a repetition of the same action in two subsequent trials (here: biking). Each functional run lasted approximately 9 min and included 100 experimental trials, 7 catch trials and 20 null trials (see Methods for details). The whole fMRI session consisted of eight functional runs.

(command: *fslmaths*). All further analyses were conducted in MATLAB (The MathWorks Inc.) using specific toolboxes mentioned below and custom written scripts (available on https://osf.io/efn3w/?view_only=c2b87331de8b45aab23bf182b9921a57).

2.7 | Preprocessing

The preprocessing of functional data included: (1) removal of the first four volumes, (2) slice time correction, (3) head motion correction (trilinear interpolation) with respect to the first volume of the first run for each participant (using MCFLIRT), (4) BET brain extraction, (5) spatial smoothing with a Gaussian kernel of 5 mm FWHM, and (6) high-pass filtering (cutoff frequency of 100 MHz). Note that step (5) was carried out for reliability-based voxel selection (following Magri et al., 2021; Park et al., 2022; Thornton & Tamir, 2024), whereas this step was omitted for representational similarity analysis (RSA).

Data were linearly registered using FMRIB's Linear Image Registration Tool (Jenkinson et al., 2002; Jenkinson & Smith, 2001), first to each participant's 3D T1-weighted image (7 degrees of freedom) and then to the MNI152 standard brain (12 degrees of freedom).

2.8 | First-level univariate fMRI analysis

We performed the first-level univariate analysis for the reliability-based voxel selection on spatially smoothed data (see previous section), whereas we used unsmoothed data for the RSA. For both types of analysis, a general linear model was used to model the obtained data series. We included 100 regressors of interests (one for each action), with each trial modeled as an epoch lasting from the onset to the offset of the image (1 s). In addition, we included one regressor for the catch trials, and six regressors resulting from 3D motion correction (x, y, z translation and rotation). Each regressor of interest was convolved with a standard dual gamma hemodynamic response function (Friston et al., 1998).

2.9 | Reliability-based voxel selection

To ensure that analyses are performed within a set of voxels that systematically respond during the processing of observed actions, we selected voxels based on their reliability following Tarhan and Konkle (2020a) using their code available on OSF (https://osf.io/m9ykh/). With this approach, the voxels are considered as reliable when they (a) show systematic differences in activation across the different experimental conditions (in our case, actions), and that (b) show similar activation levels across conditions in different sets (i.e., different exemplars) of the stimuli. To select reliable voxels, we first computed *voxel reliability values* for each voxel and each condition by correlating the corresponding vectors of β weights in response to each condition between odd and even runs. Next, for a range of reliability thresholds (from $r = 0$ to $r = .95$), we computed the *condition multi-voxel reliability*, a measure of the stability of the pattern of responses corresponding to a single condition. Condition multi-voxel reliability was computed for each condition separately, by correlating β weights between even and odd runs for all the voxels exceeding a given voxel-reliability threshold.

To implement this method, we performed the second- and group-level univariate analysis on spatially smoothed data split into odd and even runs (averaged across runs within each split). Next, we plotted the group-level condition multi-voxel reliabilities for a range of different voxel-reliability thresholds (Figure S2a) and selected the beginning point of the plateau as the reliability cutoff (Figure S2b). We also provide results of condition multi-voxel reliabilities obtained for each single subject, averaged across conditions for better readability (Figure S2c). Based on the group pattern reliability plot (Figure S2b), we decided on a voxel-reliability threshold equal to 0.25 (which is comparable to the threshold used by Tarhan & Konkle, 2020a). All subsequent analyses (both at the level of single subjects and at the group level) were performed within voxels exceeding this threshold (Figure S2d).

2.10 | Representational similarity analysis

To identify brain areas that represent (a) the semantic action space model and (b) the action feature model, we performed searchlight-based RSA (Kriegeskorte et al., 2006; Kriegeskorte et al., 2008) using the CoSMoMVPA Toolbox (Oosterhof et al., 2016). As input, we used (unsmoothed) t-maps (1 for each of the 100 actions) calculated from β estimates obtained from first-level univariate analysis. RSA was performed using a searchlight sphere (radius: 10 mm) within voxels exceeding the voxel-reliability threshold (see previous paragraph). For each searchlight sphere, a neural representational dissimilarity matrix (RDM) was created by computing pairwise distances (*squared Euclidean* distance) between t-scores of each pair of actions. The resulting neural RDM was correlated (Pearson correlation) with a selected target RDM (see *RSA: Model RDMs* for details) and the correlation value was assigned to the center voxel of each sphere, resulting in a correlation map.

To be able to account for the variability explained by additional models capturing low-level visual features, mid-level scene-related features (GIST), or action features (action feature model; see section *RSA: Model RDMs* for details), we performed a multiple regression RSA (see, e.g., Proklova et al., 2016; Tucciarelli et al., 2019, for similar approaches). To test the suitability of this approach, we determined the degree of multicollinearity between the variables using the variance inflation factor (VIF). The VIFs were small, both when including three models (semantic action space model: 1.01, low-level visual control model: 1.01, GIST model: 1.00; action feature model: 1.01, low-level visual control model: 1.01, GIST model: 1.06) and when including four models (semantic action space model: 1.09, action feature model: 1.15, low-level visual control model: 1.02, GIST model: 1.07), indicating a low risk of multicollinearity between the different

models (see also Figure S3 for Pearson's correlation coefficients between the models).

The obtained β maps were subsequently spatially smoothed with a 5 mm FWHM kernel and entered into a one-sample *t* test. Statistical significance for the group-level analyses was determined by correcting the β maps for multiple comparisons using threshold-free cluster enhancement (TFCE, Smith & Nichols, 2009) in combination with cluster level correction ($p = .05$, one-tailed, $z = 1.65$, 5000 iterations).

To reveal areas that capture the semantic similarity between actions, we carried out two multiple regression RSAs with the semantic action space model. In the first multiple regression RSA, to account for low-level visual features and mid-level scene-related features, we regressed out the low-level visual control model and the GIST model. In the second multiple regression RSA, to account for low-level visual features, mid-level scene-related features and information corresponding to a range of different action-related features, we regressed out the low-level visual control model, the GIST model and the action feature model.

To be able to compare the topography of the areas capturing the semantic action space model and the action feature model, we computed another multiple regression RSA for the action feature model, regressing out the low-level visual control model and the GIST model.

For visualization purposes, we displayed the resulting thresholded t-maps onto an inflated standard surface map provided by BrainNet Viewer (Xia et al., 2013).

2.10.1 | RSA: Model RDMs

The semantic action space model and the action feature model were derived on the basis of a number of behavioral experiments (Kabulska & Lingnau, 2022), whereas the low-level visual control model and the GIST model were established on the basis of image properties. The procedures are briefly summarized below.

Semantic action space model

This model was used to determine which brain areas capture the semantic similarity space of actions resulting from behavioral judgments of action similarity. Following previous studies (Dima et al., 2022; Tucciarelli et al., 2019), we derived this model from a multi-arrangement paradigm (Kriegeskorte & Mur, 2012). In short, 20 participants were asked to arrange 100 images of daily actions (same set of actions as used in the current study) on an arena, where between-action distances reflected action similarity (for details, see Kabulska & Lingnau, 2022). The model was created based on the resulting pairwise distances between the actions, averaged across participants.

Action feature model

We established this model in order to examine to which degree the semantic action space can be accounted for on the basis of the similarity of a wide range of features. First, using a free feature-listing experiment, we asked $N = 40$ participants to list at least 5 features per action which resulted in approximately 6000 collected responses describing a set of 100 daily actions presented as verbs (same actions as used in the current study). Second, we reduced that list of features to 44 key action features (e.g., *Upper/Lower limbs*, *Targeting a person/tool*, *Pace*, *Duration*, *Valence*; see Table S2 for a list of all 44 key action features) and, from another set of $N = 273$ participants, obtained feature-based ratings for the same set of 100 actions. The averaged and rescaled ratings were subsequently used to create a feature model by computing pairwise distances between actions (Euclidean distance).

Note that the action feature model was built using ratings obtained for actions depicted as action verbs (rather than static pictures). This was done to avoid that ratings were driven by particular exemplars depicting an action. A consequence of the use of verbal material as prompts for action feature ratings might be that we underestimated the degree to which the action feature model can account for the semantic action space model.

Low-level visual control model

We constructed this model to be able to account for low-level visual features. Since representations of objects in early layers of artificial neural networks have been shown to resemble neural activity within early visual cortex (Güçlü & van Gerven, 2015; Lindsay, 2021) we decided to use the first convolutional layer from ResNet50, a deep convolutional network with 50 layers, pretrained on object categories (He et al., 2016) and fine-tuned on 339 action categories from the Moments in Time dataset (Monfort et al., 2020). We fed the ResNet50 model with the 400 action images (100 actions with 4 exemplars each) which we used in the fMRI experiment. Next, we (1) determined the activations within the first convolutional layer and stored them as vectors and (2) averaged the resulting vectors across action exemplars, resulting in 100 activation vectors (one vector per action). (3) Next, we computed 1-Pearson's R correlation for each pairwise combination of vectors resulting in a 100×100 RDM. We also created an RDM based on the first layer of AlexNet (Krizhevsky et al., 2017), pretrained on the ImageNet dataset (Russakovsky et al., 2015) and conducted an RSA using the corresponding RDM as a low-level visual control model. Since AlexNet is another frequently used convolutional neural network (e.g., Kietzmann et al., 2019; Lee Masson & Isik, 2021), we wished to determine whether we obtain similar results when low level visual features are determined on the basis of AlexNet instead of ResNet50. The corresponding results are shown in Figures S4 and S5.

GIST model

To account for the similarity between observed actions that is due to the similarity of the scenes in which these actions are performed, we employed the GIST model (Torralba & Oliva, 2001). This computational model extracts information about scenes based on several dimensions that have been shown to be related to specific scene categories, such as naturalness, openness, and roughness. We generated GIST descriptors for all 400 action images (100 actions with 4 exemplars each) using the default parameters for the number of

orientations at which the Gabor filters are applied, and the filter used to reduce illumination effects of input images. Subsequently, we averaged the descriptors across action exemplars, resulting in a set of 100 GIST descriptors, one for each action. To construct the GIST RDM, we computed pairwise distances between the actions using the Euclidean distance metric.

2.11 | Principal component analysis

The action feature model contained information about all 44 action features reported by Kabulska and Lingnau (2022) (see *section RSA: Model RDMs* for details). To be able to reduce this large number of features to a smaller set of dimensions, we conducted a PCA on the 44 feature-based ratings of 100 actions (same ratings as used to create the feature RDM, see *RSA: Behavioral RDMs*). The components were derived using varimax rotation which maintains orthogonality between them. We identified 11 components with eigenvalues greater than one (Table S3). Based on the screen plot combined with the "elbow method" (Figure S6), we chose eight dimensions, accounting for approximately 64.1% in total of the variability in the feature ratings.

2.11.1 | RSA with principal components

Subsequently, we wanted to determine which brain regions best capture these dimensions. To address this question, we performed a regression-based RSA, separately for each of the eight dimensions while regressing out the low-level visual control model and the GIST model (see *section RSA: Model RDMs* for details). In order to construct the dimension-based RDMs, the first step involved multiplying the action feature ratings (obtained in Kabulska & Lingnau, 2022) by loadings on a given dimension. Subsequently, we took the resulting 100 vectors (one per action) of weighted feature ratings and computed pairwise distances (*Euclidean* distance) between them. Prior to conducting multiple regression RSA, we computed the VIF. The VIFs were below 4 for all the models indicating low multicollinearity between them (PC1: 2.34; PC2: 2.91; PC3: 3.16, PC4: 3.46, PC5: 3.16, PC6: 2,40, PC7: 3.33, PC8: 2.30, low-level visual model: 1.04, GIST model: 1.15). Pearson's correlation coefficients between the models are shown in Figure S3.

2.12 | Winner-takes-all with principal components

To visualize the most dominant dimension for each voxel, we calculated a winner-takes-all map following Tarhan et al. (2021) within the voxels exceeding the reliability-based voxel threshold. We only included the six (out of eight) principal components (PCs) for which the multiple regression RSA revealed significant clusters of voxels that survived correction for multiple comparisons. We assigned a unique color to each voxel to the dimension with the highest correlation.

3 | RESULTS

3.1 | Behavioral results

We performed an fMRI experiment with 100 daily actions (four exemplars per action; see Figure S1 for a complete overview of all stimuli) using a rapid event-related design (see *Methods*, section *Experimental design and task* for details). Mean reaction time for correct responses was 959.84 ms (±43.60 ms SEM). Participants identified catch trials with a mean error rate of 24.73% (±2.74% SEM), corresponding to approximately 14 out of 56 catch trials per participant.

The post-session questionnaire revealed that on average the participants were reasonably concentrated on the task of identifying catch trials (mean = 4.09; std = 1.12; 1 = *not concentrated at all*, 6 = *concentrated exclusively on the task*), and that they felt reasonably comfortable inside the scanner (mean = 4.3; std = 0.88; 1 = *very uncomfortable*, 6 = *very comfortable*). The questionnaire also revealed that participants felt neither completely rested nor very tired throughout the experiment (mean = 3.43; std = 1.04; 1 = *not tired at all*, 6 = *very tired*), and that they verbalized the stimuli to some degree to perform the task (mean = 4.7; std = 1.11; 1 = *not naming at all*; 6 = *quietly naming*).

Individual error rates and answers provided in the post-session questionnaire are provided in Table S4.

3.2 | Reliability map

Following Tarhan and Konkle (2020a), we used a reliability-based voxel selection (see Methods section for details). This analysis revealed voxels with high reliability in occipital brain areas, covering both ventral and dorsal visual streams, and part of the parietal lobe (see Figure S2D), whereas voxel reliability was lower in frontal areas. All subsequent analyses were performed within the reliability map.

3.3 | Searchlight-based RSA

To determine which brain areas reflect the semantic action space model, corresponding to the categorical organization obtained from the multi-arrangement task while accounting for variability due to low- level visual features, mid-level scene-related features, and high-level action features, we performed a multiple regression searchlight-based RSA (see Methods, section *Representational similarity analysis* for details). Moreover, to determine the spatial correspondence between the regions capturing the semantic action space and the regions capturing a feature-based organization, we carried out an additional searchlight-based RSA on the action feature model.

The resulting searchlight maps for the semantic action space model (while regressing out the low-level visual control model and the GIST model) revealed significant correlations between neural patterns of activation and the action space model in bilateral occipitotemporal and fusiform cortex, as well as in small portions of the superior

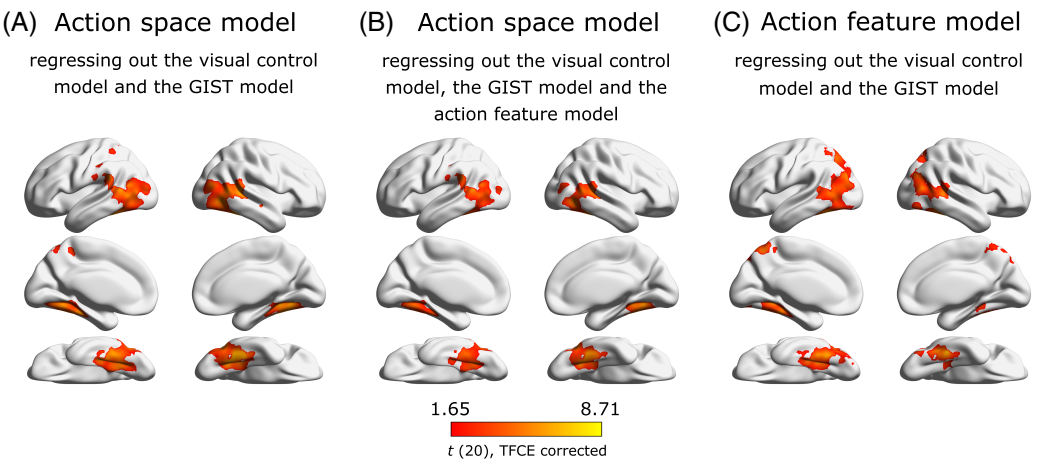

FIGURE 2 Results of the group-level searchlight-based representational similarity analysis (RSA) for: (a) the semantic action space model (regressing out the low-level visual control model and the GIST model); (b) the semantic action space model (regressing out the low-level visual control model, the GIST model and the action feature model); and (c) the action feature model (regressing out the low-level visual control model and the GIST model). Statistical maps show t-values thresholded at a z-score of 1.65, corresponding to $p < .05$ (one-tailed), corrected for multiple comparisons (TFCE, $p < .05$, 5000 Monte Carlo permutations).

parietal lobe (Figure 2a). Additionally regressing out the action feature model resulted in a qualitatively similar, but less widespread map (Figure 2b) that was limited to bilateral occipitotemporal and temporal occipital fusiform cortex.

The action feature model was captured by patterns of activation in a comparable, but slightly more widespread set of regions, including the bilateral occipitotemporal and fusiform cortex, as well as the superior parietal lobe (see Figure 2c). The resulting brain maps show a similar pattern when the low-level visual control model is generated on the basis of the first layer of AlexNet (Figure S4).

3.4 | Principal component analysis

PCA on the 44 feature ratings for the 100 different actions revealed eight components that explained 64.1% of the variance (see *Methods* section for details). We labeled these components on the basis of the features belonging to each component (see Table S3). We labeled the first component that explained most of the variance (21.6%) ***Movements of any type*** due to high positive loadings for features linked to various types of movements, such as *Lower limb movements*, *Change of location*, *Use of force* and negative loadings on features associated with the absence of movement, that is, *No movement* and *Sitting*. The second component was mostly related to different arm movement kinematics (e.g., rotating, sweeping, circular) and was therefore labeled ***Arm movement kinematics***. We used the label ***Object manipulation involving the upper* body** for the third component for several reasons. First, the component consisted of features related to the upper body. Second, we wished to capture the aspect "Object manipulation" in the label since the features *Goal-directedness* and *Targeting a non-manipulable object/a tool* indicated that the actions have a clear end-goal and are aiming at specific objects or tools. The subsequent components were associated with features related to the ***Context*** of the actions (*Indoor*, *Outdoor*, *Season-dependence*), the ***Posture*** of the agents performing the actions, ***Contact with others*** (i.e., whether or not the action involved direct or indirect contact with another person), and **Object-directedness** (i.e., whether or not the action targeted a manipulable object, as well as *Concentration*). The last component referred to the features *Noise*, *Harm*, and *Negative valence* and thus was labeled ***Negative Emotions***.

3.5 | RSA on dimensions resulting from PCA

To determine which brain areas represent the information captured by each of the dimensions resulting from PCA on the feature ratings, we conducted a searchlight-based RSA, separately for each of the eight dimensions, regressing out the low-level visual control model and the GIST model. The results of this analysis are shown in Figure 3. For the dimension *Movements of any type*, that explains the largest amount of variance (21.63%), we identified significant clusters in bilateral temporal occipital fusiform cortices and lateral occipital cortices, extending toward the superior parietal lobules. The dimension *Object manipulation involving the upper body* was captured by clusters in the left inferior temporal gyrus, bilateral temporal fusiform cortices, and lateral occipital cortices. For the dimension *Context*, we obtained clusters in bilateral lateral occipital cortices and the left temporal occipital fusiform cortex. Clusters in bilateral temporal occipital fusiform cortices and superior parietal lobules corresponded to the dimension *Posture*. The dimension *Contact with others* was associated with clusters in the left lateral occipital cortex (superior and inferior division) and a smaller cluster in the right lateral occipital cortex (inferior division), as well as a cluster in the left temporal occipital fusiform cortex. The dimension *Object-directedness* showed a significant correlation with activation patterns within clusters in bilateral temporal occipital fusiform cortices and lateral occipital cortices (superior division). In sum,

this analysis revealed a substantial degree of overlap between the different dimensions along the ventral visual stream and the superior parietal lobe, in particular for the dimensions *Movements of any type*, *Context*, *Posture*, and *Object-directedness*. By contrast, the dimensions *Object manipulation involving the upper body* and *Contact with others* were associated with more circumscribed clusters of voxels. The resulting brain maps show a similar pattern when the low-level visual control model is based on the first layer of AlexNet rather than ResNet50 (see Figure S5). To explore the spatial arrangement of these dimensions, we carried out a Winner-takes-all analysis (see next section).

3.6 | Winner-takes-all map

To explore clusters of voxels with a preference for individual PCs, we calculated a winner-takes-all map (see Methods for details). Note that since we provide no additional statistics for these maps, this analysis merely serves as an additional visualization of the results shown in Figure 3. That said, the winner-takes-all analysis highlights multiple clusters displaying the highest correlation with the dimension *Movements of any type* in a prominent portion of the right LOTC as well as the left dorsal LOTC (Figure 4, blue). The dimension labeled *Object-manipulation involving the upper body* showed the highest correlations

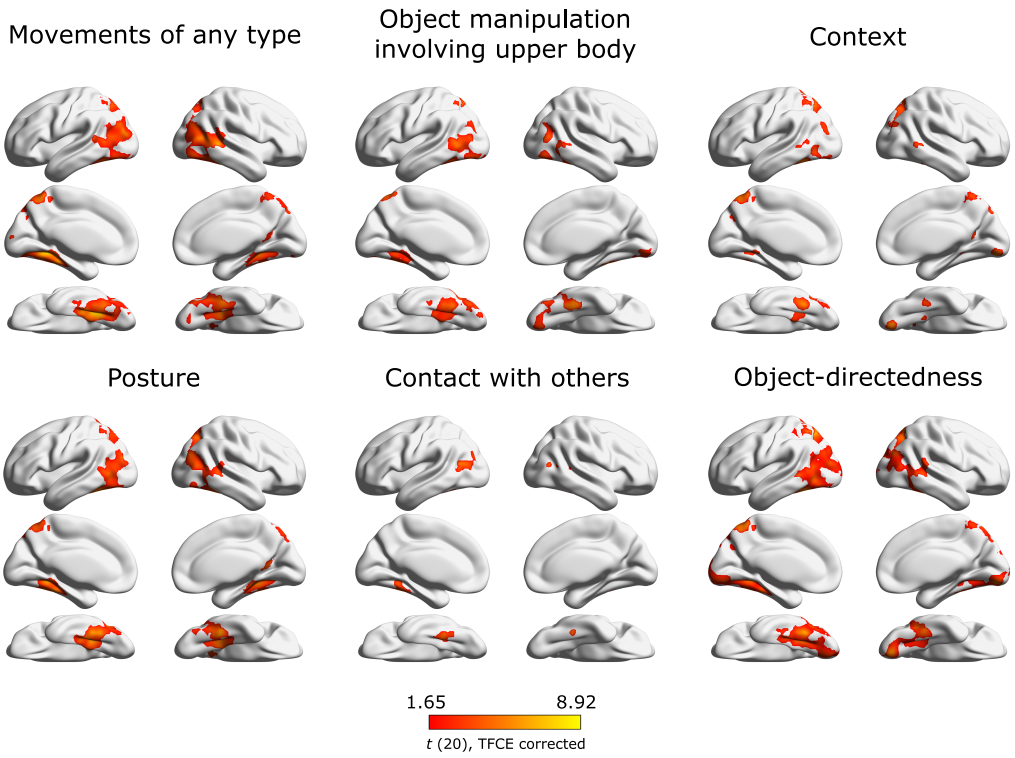

FIGURE 3 Results of the searchlight representational similarity analysis (RSA), carried out separately for each of the eight dimensions (regressing out the low-level visual control model and the GIST model). Six out of eight dimensions showed a significant correlation with neural activation patterns after correction for multiple comparisons (TFCE, $p < .05$, 5000 Monte Carlo permutations). Statistical maps show t-maps thresholded using threshold-free cluster enhancement (TFCE) at a z-score of 1.65. The remaining two dimensions, namely arm movement kinematics and negative emotions, did not survive the correction.

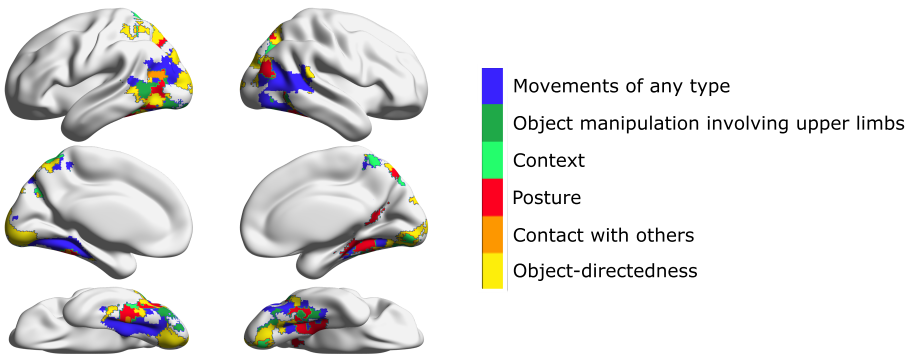

FIGURE 4 Results of the winner-takes-all analysis with maps for six different dimensions obtained from the searchlight-based representational similarity analysis (RSA) (see Figure 2; see also Tarhan et al., 2021). Each voxel was assigned a color corresponding to the dimension that showed the strongest correlation with the corresponding activity patterns (see legend on the right for the assignment of colors to each dimension).

with patterns of activation in a more anterior portion of the left middle LOTC (dark green). The dimension *Context* formed small clusters in bilateral superior parietal lobe (light green). The information related to *Posture* was encoded in the left middle posterior LOTC, right dorsal LOTC and portions of bilateral VOTC (red), whereas the *Contact with others* dimension exhibited a significant correlation with patterns of activation in the left dorsal LOTC (orange). Finally, this analysis highlighted that the *Object-directedness* dimension exhibits the highest correlation with activation patterns in the superior parietal lobe (bilaterally), a small portion in the inferior parietal lobe (bilaterally), as well as portions of visual cortex (bilaterally), ranging from V1 to V4 (yellow).

4 | DISCUSSION

In this study, we investigated the neural architecture underlying the organization of a wide range of observed actions. For that purpose, we conducted an fMRI experiment with static images depicting 100 different human actions. Using multiple regression RSA in which we accounted for variability due to low-level visual features and mid-level scene-related features, we identified shared representations of a semantic action space and a high-level action feature model in lateral and ventral occipitotemporal cortex. Using PCA, we found that these action features can be reduced to eight dimensions, including movements of any type, object-directedness and context that explained 64.8% of the variance of the data. RSA with these dimensions revealed distinct, but partially overlapping clusters for six out of eight dimensions within the LOTC, the VOTC and the superior parietal lobe that were further distinguished using a winner-take-all analysis. In the following we discuss these results in the context of existing studies on this topic and point out future directions.

4.1 | Neural representation of the semantic action space

In the current study, we aimed to determine which brain regions reflect the similarity between observed actions, captured in the semantic action space model. We followed the assumption that activation patterns in areas that store conceptual representations of actions show a significant correlation with the similarity structure captured in the semantic action space model. Our results are in line with the results by Tucciarelli et al. (2019) who reported that a behaviorally determined action space assumed to capture the semantic similarity of a set of 27 actions is reflected by patterns of activation in the LOTC. To account for additional action components that might covary with the semantic action space, Tucciarelli et al. (2019) regressed out nine additional models capturing diverse aspects, including the similarity of objects, body parts and the distance between the observer and the actor. These additional action components partially overlapped with the cluster capturing the semantic action space. The current study advances the findings of Tucciarelli et al. (2019) in two important ways. First, we demonstrated that the results of Tucciarelli et al. (2019) generalize to a significantly wider range of actions (i.e., 100 instead of 27 actions). Second, our whole-brain searchlight RSA revealed the highest similarity between the semantic action space model and patterns of activation in dorsal and ventral portions of the LOTC, even after regressing out (a) a low-level visual control model derived from the first convolutional layer of a neural network (ResNet50), (b) mid-level spatial information of the scenes captured by the GIST model and (3) high-level information related to 44 different action features. In line with this view, the posterior middle temporal gyrus, a subportion of the LOTC, has been shown to be involved in the processing of action semantics (e.g., Kable et al., 2002; Kemmerer et al., 2008; Papeo et al., 2015), and lesions to temporal and parietal regions, but not to frontal regions, have been shown to have an impact on the ability to associate an action with an appropriate semantic label (Kalénine et al., 2010). Together, our results provide an important extension of a growing number of studies suggesting that the LOTC gathers not only perceptual evidence on the basis of action features, but also more conceptual action aspects (Hafri et al., 2017; Oosterhof et al., 2010, 2012; Wurm et al., 2015; Zhuang et al., 2023; for reviews, see Wurm & Caramazza, 2022, Lingnau & Downing, 2015, 2023).

4.2 | Dimensions underlying the organization of action features

The action feature model used in the current study is based on ratings obtained for 44 action features carried out for 100 different actions (Kabulska & Lingnau, 2022). PCA revealed eight dimensions underlying the organization of these features (*Movements of any type*, *Arm movement kinematics*, *Object manipulation involving the upper body*, *Context*, *Posture*, *Contact with others*, *Object-directedness*, and *Negative Emotions*). These dimensions align remarkably well with those previously proposed and examined. Movements and posture are undeniably crucial aspects of actions, as they are sufficient to identify a wide range of actions (Johansson, 1973, see, e.g., Beauchamp et al., 2003; Grossman et al., 2000; Papeo et al., 2017 for studies with point-light displays). Moreover, numerous actions involve the use of tools (e.g., Buxbaum, 2001; Chao & Martin, 2000; Watson & Buxbaum, 2014) or are directed toward specific objects (e.g., Bach et al., 2014; Wurm et al., 2017). Additionally, contact with other people and social actions play a vital role in our daily lives, enabling successful communication with others, while comprehending and interpreting emotions is a crucial part in this process (e.g., Isik et al., 2017; Papeo, 2020; Poyo Solanas et al., 2020). Moreover, in contrast to objects that can be understood in isolation, understanding actions involves information that extends beyond the body itself (e.g., information regarding the scene; see Wurm & Schubotz, 2012; Wurm & Schubotz, 2017). Hence, using a wide range of actions and action features, our study revealed a set of dimensions that have been

proposed in previous studies but, to the best of our knowledge, have not been collectively investigated before. Our approach allowed us to determine the degree to which the neural representation of these action features contributes to the neural representation of the semantic action space model (see section *Neural representation of the semantic action space*), and to examine to which degree the topographies corresponding to the neural representation of the different action dimensions spatially overlap with the neural territory capturing the semantic action space. We discuss these results in more detail in the following paragraphs.

4.3 | Neural representation of action dimensions

The searchlight RSA on the obtained PCs revealed overlapping clusters of voxels along ventral and dorsal portions of the LOTC and the superior parietal cortex for the dimensions *Movements of any type*, *Context*, *Posture*, and *Object-directedness*, and more circumscribed clusters in the LOTC and the fusiform cortex for the dimensions *Object manipulation involving the upper body* and *Contact with others*. Thus, in line with the results reported by Tucciarelli et al. (2019), the LOTC carries information about each of the investigated action dimensions, which we will discuss in more detail in the following sections.

We found that activation patterns in the dorsal LOTC showed the highest similarity with the dimension labeled *Contact with others*, while activation patterns in the ventral LOTC showed the highest similarity with the dimension labeled *Object-directedness*. These results are in line with several recent studies indicating an animate–inanimate organization of dorsal and ventral portions of the LOTC (e.g., Lingnau & Downing, 2015; Wurm et al., 2017; Wurm & Caramazza, 2022). More precisely, it has been shown that dorsal portions of the LOTC have a preference for animate things (Chao et al., 1999; He et al., 2020), body parts (e.g., Downing et al., 2001), movements (Beauchamp et al., 2003), and person-directed actions (Wurm & Caramazza, 2019a, 2019b), while ventral portions have a preference for inanimate things (Chao et al., 1999; He et al., 2020), action-specific tool motion (Beauchamp et al., 2002; Beauchamp et al., 2003), and actions involving objects (e.g., Wurm & Caramazza, 2019a; see Wurm & Caramazza, 2022 for a recent review on the animate–inanimate organization). Note that the clusters representing the dimensions *Contact with others* and *Object-directedness* obtained in the current study (see Figures 2 and 3) are well aligned with the clusters showing a high similarity with the *Sociality and the Transitivity model* reported by Wurm et al. (2017).

It is worth noting that the studies that formed the basis of the idea of the animate–inanimate dimension as one of the organizing principles of the LOTC used material that was quite different from the material used in the current study. Specifically, Martin and Weisberg (2003) used moving geometric shapes, whereas Wurm et al. (2017), Wurm and Caramazza (2019a), and Wurm and Caramazza (2019b) used well-controlled videos of a small set of actions performed by an actor sitting at a table with the upper arms directed at an object or a person. In the current study, we demonstrate that the distinction between person-directed and object-directed actions generalizes across a wide range of actions from a diverse set of categories, involving different body parts and objects depicted in naturalistic scenes as static images.

In contrast to most of the other dimensions which showed significant correlations with the activity patterns within several brain regions, the dimension *Contact with others* was mainly located in the posterior superior temporal sulcus (pSTS). This result is well aligned with a growing number of studies demonstrating that the pSTS carries information about communicative actions (Isik et al., 2017; Pitcher & Ungerleider, 2021; Walbrin et al., 2018). Moreover, the cluster for *Contact with others* was more widespread in the left compared to the right hemisphere, indicating a lateralization in encoding social aspects of actions.

4.4 | High-dimensional spaces in the LOTC

The regions capturing the higher-level action feature model and the underlying dimensions strongly overlapped with those capturing the semantic action space model. The overlap encompassed the LOTC, indicating the pivotal role of this region in representing diverse information about actions (see also Lingnau & Downing, 2015; Wurm & Caramazza, 2022). This raises the question according to which principles this diverse information is represented along the LOTC. Our data are compatible with the proposal put forward by Lingnau and Downing (2015) that diffuse patterns of activation across the LOTC integrate information from more focal, but strongly overlapping selective regions, and that the distribution of these activity patterns might define multiple representational spaces. The organization of the LOTC along multiple dimensions is in line with the idea put forward by the work by Graziano and Aflalo (2007), indicating that the motor cortex is organized along multiple dimensions—such as somatotopic information and information about different types of limb movements. As suggested by Graziano and Aflalo (2007), this structure is not limited to the motor cortex and may extend to any region that processes multidimensional and complex knowledge. Whereas it seems likely that the importance of specific dimensions differs between the dorsal and the ventral stream, the general principle proposed by Graziano and Aflalo (2007) might apply to the LOTC as well. In line with this view, several studies demonstrated that planned actions can be decoded not only from premotor and parietal regions, but also from the LOTC (e.g., Ariani et al., 2015; Gallivan et al., 2013; Gallivan & Culham, 2015; Turella et al., 2020; see also Gallivan, 2014). Moreover, information is assumed to be exchanged between the dorsal and the ventral stream, for example, via connectivity between the posterior parietal cortex and the extrastriate body area (e.g., Hutchison et al., 2014; Zimmermann et al., 2016, 2018). The organization of action-relevant information along high-dimensional spaces might facilitate such an exchange. The current study revealed a number of these dimensions and thus may serve as a starting point for further studies.

4.5 | The contribution of conceptual knowledge to the organization of observed actions

Conceptual knowledge can be divided into function (i.e., "what for") and manipulation (i.e., "how to") knowledge (e.g., Buxbaum et al., 2000; Lesourd et al., 2021; Mahon & Caramazza, 2009), with a specific role of temporal and parietal regions, respectively. Moreover, object concepts have been proposed to be organized taxonomically (e.g., in terms of feature-based similarities), thematically (e.g., in terms of co-occurrences in particular contexts) or both (Kalénine & Buxbaum, 2016). One thus may wonder about the link between the organizing principles for observed actions obtained in the current study and those that are assumed to contribute to the organization of conceptual knowledge.

For the generation of our semantic action space model, we did not explicitly instruct participants to focus on any of these aspects. Moreover, function and manipulation knowledge, and taxonomic and thematic relations are often correlated (as an example, actions belonging to the category "food-related actions" are typically thematically related to the context "kitchen"). Since the current study did not focus on a distinction between function and manipulation knowledge, or between taxonomic and thematic relations, we did not select our stimuli in a way that would allow us to disentangle the contribution of these different types of knowledge. Consequently, it is not straightforward to quantify the contribution of function or manipulation knowledge, or the role of taxonomic versus thematic relations during the generation of the semantic action space model. In fact, we consider it likely that participants based their similarity judgments of observed actions on a combination of several dimensions (see also previous paragraph), including the use of function and manipulation knowledge and taxonomic and thematic relations. Moreover, we assume that the importance of different dimensions during the judgment of action similarities varies as a function of the task (see also Lingnau & Downing, 2023).

Regarding the action feature model, we asked participants to rate the importance of a range of different features, related to sensory, functional, motor, and manipulation knowledge, but also of higher-level, rather abstract features such as "valence" or "harm." The components revealed by PCA applied to the action feature model range from manipulation knowledge (components "movements of any type," and "posture") to thematic relations (e.g., components "context," "object directedness," and "object manipulation involving upper body") to more abstract organizing principles (component "contact with others").

5 | CONCLUSION

Our results provide an important extension of previous studies, suggesting that the LOTC hosts conceptual representations for a wide range of observed actions from several different action categories. Moreover, our results suggest that the areas capturing the semantic action space overlap with areas capturing action components at varying hierarchical levels, in line with the idea that the LOTC, like other areas of the brain such as the motor cortex and parietal cortex, is organized along multiple dimensions (see also Graziano & Aflalo, 2007; Lingnau & Downing, 2015; Wurm & Caramazza, 2022).

ACKNOWLEDGMENTS

The authors are thankful to Leyla Tarhan for providing MATLAB code for computing the reliability maps and for conducting the winner-takes-all analysis. The authors also thank Lucca Scheuermeyer for help with MRI data collection, as well as Marius Zimmermann, Oleg Vrabie, Federica Danaj, Max Reger, Andre Bockes, and Robert Bosek for helpful discussions and comments on the manuscript, and the two anonymous reviewers for their constructive feedback. Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

This research was funded by a Research Grant from the German Research Foundation (Li 2840/1-1). Angelika Lingnau was funded by a Heisenberg Professorship (German Research Foundation, Li 2840/2-1).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest or competing interests.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in OSF at https://osf.io/efn3w/?view_only=c2b87331de8b45aab23bf182b9921a57.

ORCID

Zuzanna Kabulska https://orcid.org/0000-0002-6756-691X
Tonghe Zhuang https://orcid.org/0000-0002-7876-4962
Angelika Lingnau https://orcid.org/0000-0001-8620-3009

REFERENCES

Ariani, G., Wurm, M. F., & Lingnau, A. (2015). Decoding internally and externally driven movement plans. *Journal of Neuroscience*, *35*(42), 14160–14171. https://doi.org/10.1523/JNEUROSCI.0596-15.2015

Bach, P., Nicholson, T., & Hudsons, M. (2014). The affordance-matching hypothesis: How objects guide action understanding and prediction. *Frontiers in Human Neuroscience*, *8*, 254. https://doi.org/10.3389/fnhum.2014.00254

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron*, *34*(2), 149–159. https://doi.org/10.1016/S0896-6273(02)00642-6

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2003). fMRI responses to video and point-light displays of moving humans and manipulable objects. *Journal of Cognitive Neuroscience*, *15*(7), 991–1001. https://doi.org/10.1162/089892903770007380

Beymer, D., & Poggio, T. (1997). Image representations for visual learning. *Lecture Notes in Computer Science*, *1206*(37), 143. https://doi.org/10.1007/bfb0015989

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Buxbaum, L. J. (2001). Ideomotor apraxia: A call to action. *Neurocase*, *7*(6), 445–458. https://doi.org/10.1093/neucas/7.6.445

Buxbaum, L. J., Veramontil, T., & Schwartz, M. F. (2000). Function and manipulation tool knowledge in apraxia: Knowing 'what for' but not 'how'. *Neurocase*, *6*(2), 83–97. https://doi.org/10.1093/neucas/6.2.97

Chao, L. L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *NeuroImage*, *12*(4), 478–484. https://doi.org/10.1006/nimg.2000.0635

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, *2*(10), 913–919. https://doi.org/10.1038/13217

Dima, D. C., Hebart, M. N., & Isik, L. (2023). A data-driven investigation of human action representations. *Scientific Reports*, *13*(1), 5171. https://doi.org/10.1038/s41598-023-32192-5

Dima, D. C., Tomita, T. M., Honey, C. J., & Isik, L. (2022). Social-affective features drive human representations of observed actions. *eLife*, *11*, 1–22. https://doi.org/10.7554/eLife.75027

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*(5539), 2470–2473. https://doi.org/10.1126/science.1063414

Edelman, S. (1998). Representation is representation of similarities. *Behavioral and Brain Sciences*, *21*(4), 449–498. https://doi.org/10.1017/S0140525X98001253

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., & Turner, R. (1998). Event-related fMRI: Characterizing differential responses. *NeuroImage*, *7*, 30–40. https://doi.org/10.1006/nimg.1997.0306

Gallivan, J. P. (2014). A motor-oriented organization of human ventral visual cortex? *Journal of Neuroscience*, *34*(9), 3119–3121. https://doi.org/10.1523/JNEUROSCI.0060-14.2014

Gallivan, J. P., & Culham, J. C. (2015). Neural coding within human brain areas involved in actions. *Current Opinion in Neurobiology*, *33*, 141–149. https://doi.org/10.1016/j.conb.2015.03.012

Gallivan, J. P., Chapman, C. S., Mclean, D. A., Flanagan, J. R., & Culham, J. C. (2013). Activity patterns in the category-selective occipitotemporal cortex predict upcoming motor actions. *European Journal of Neuroscience*, *38*(3), 2408–2424. https://doi.org/10.1111/ejn.12215

Graziano, M. S. A., & Aflalo, T. N. (2007). Rethinking cortical organization: Moving away from discrete areas arranged in hierarchies. *The Neuroscientist*, *13*(2), 138–147. https://doi.org/10.1177/1073858406295918

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., & Blake, R. (2000). Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience*, *12*(5), 711–720. https://doi.org/10.1162/089892900562417

Güçlü, U., & van Gerven, M. A. J. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience*, *35*(27), 10005–10014. https://doi.org/10.1523/JNEUROSCI.5023-14.2015

Hafri, A., Trueswell, J. C., & Epstein, R. A. (2017). Neural representations of observed actions generalize across static and dynamic visual input. *The Journal of Neuroscience*, *37*(11), 3056–3071. https://doi.org/10.1523/JNEUROSCI.2496-16.2017

He, C., Hung, S. C., & Cheung, O. S. (2020). Roles of category, shape, and spatial frequency in shaping animal and tool selectivity in the occipitotemporal cortex. *Journal of Neuroscience*, *40*(29), 5644–5657. https://doi.org/10.1523/JNEUROSCI.3064-19.2020

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 770–778. https://doi.org/10.1109/CVPR.2016.90

Hutchison, R. M., Culham, J. C., Everling, S., Flanagan, J. R., & Gallivan, J. P. (2014). Distinct and distributed functional connectivity patterns across cortex reflect the domain-specific constraints of object, face, scene, body, and tool category-selective modules in the ventral visual pathway. *NeuroImage*, *96*, 216–236. https://doi.org/10.1016/j.neuroimage.2014.03.068

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 of the United States of America*, *114*(43), 9145–9152. https://doi.org/10.1073/pnas.1714471114

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*(2), 143–156. https://doi.org/10.1016/S1361-8415(01)00036-6

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1016/S1053-8119(02)91132-8

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). FSL 1. *NeuroImage*, *62*, 782–790.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*, 201–211. https://doi.org/10.3758/BF03212378

Kable, A. J. W., Lease-spellmeyer, J., & Chatterjee, A. (2002). Neural substrates of action event knowledge. *Journal of Cognitive Neuroscience*, *14*(5), 795–805. https://doi.org/10.1162/08989290260138681

Kabulska, Z., & Lingnau, A. (2022). The cognitive structure underlying the organization of observed actions. *Behavior Research Methods*, *55*(4), 1890–1906. https://doi.org/10.3758/s13428-022-01894-5

Kalénine, S., & Buxbaum, L. J. (2016). Thematic knowledge, artifact concepts, and the left posterior temporal lobe: Where action and object semantics converge. *Cortex*, *82*, 164–178. https://doi.org/10.1016/j.cortex.2016.06.008

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. https://doi.org/10.1093/brain/awq210

Kemmerer, D., Castillo, J. G., Talavage, T., Patterson, S., & Wiley, C. (2008). Neuroanatomical distribution of five semantic components of verbs: Evidence from fMRI. *Brain and Language*, *107*, 16–43. https://doi.org/10.1016/j.bandl.2007.09.003

Kietzmann, T. C., Spoerer, C. J., Sörensen, L. K. A., Cichy, R. M., Hauk, O., & Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 21854–21863. https://doi.org/10.1073/pnas.1905544116

Kriegeskorte, N., & Mur, M. (2012). Inverse MDS: Inferring dissimilarity structure from multiple item arrangements. *Frontiers in Psychology*, *3*, 245. https://doi.org/10.3389/fpsyg.2012.00245

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 3863–3868. https://doi.org/10.1073/pnas.0600244103

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4. https://doi.org/10.3389/neuro.06.004.2008

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Lee Masson, H., & Isik, L. (2021). Functional selectivity for social interaction perception in the human superior temporal sulcus during natural viewing. *NeuroImage*, *245*(July), 118741. https://doi.org/10.1016/j.neuroimage.2021.118741

Lesourd, M., Servant, M., Baumard, J., Reynaud, E., Ecochard, C., Medjaoui, F. T., Bartolo, A., & Osiurak, F. (2021). Semantic and action tool knowledge in the brain: Identifying common and distinct networks. *Neuropsychologia*, *159*, 107918. https://doi.org/10.1016/j.neuropsychologia.2021.107918

Lindsay, G. W. (2021). Convolutional neural networks as a model of the visual system: Past, present, and future. *Journal of Cognitive Neuroscience*, *33*(10), 2017–2031. https://doi.org/10.1162/jocn_a_01544

Lingnau, A., & Downing, P. (2023). *Action understanding*. Cambridge Elements.

Lingnau, A., & Downing, P. E. (2015). The lateral occipitotemporal cortex in action. *Trends in Cognitive Sciences*, *19*(5), 268–277. https://doi.org/10.1016/j.tics.2015.03.006

Magri, C., Konkle, T., & Caramazza, A. (2021). The contribution of object size, manipulability, and stability on neural responses to inanimate objects. *NeuroImage*, *237*(December 2020), 118098. https://doi.org/10.1016/j.neuroimage.2021.118098

Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories: A cognitive neuropsychological perspective. *Annual Review of Psychology*, *60*, 27–51. https://doi.org/10.1146/annurev.psych.60.110707.163532

Martin, A., & Weisberg, J. (2003). Neural foundations for understanding social and mechanical concepts. *Cognitive Neuropsychology*, *20*(3–6), 575–587. https://doi.org/10.1080/02643290342000005

Monfort, M., Andonian, A., Zhou, B., Ramakrishnan, K., Bargal, S. A., Yan, T., Brown, L., Fan, Q., Gutfreund, D., Vondrick, C., & Oliva, A. (2020). Moments in time dataset: One million videos for event understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(2), 502–508. https://doi.org/10.1109/TPAMI.2019.2901464

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. https://doi.org/10.3389/fninf.2016.00027

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. https://doi.org/10.1162/jocn_a_00195

Oosterhof, N. N., Wiggett, A. J., Diedrichsen, J., Tipper, S. P., & Downing, P. E. (2010). Surface-based information mapping reveals crossmodal vision-action representations in human parietal and occipitotemporal cortex. *Journal of Neurophysiology*, *104*(2), 1077–1089. https://doi.org/10.1152/jn.00326.2010

Papeo, L. (2020). Twos in human visual perception. *Cortex*, *132*, 473–478. https://doi.org/10.1016/j.cortex.2020.06.005

Papeo, L., Lingnau, A., Agosta, S., Pascual-Leone, A., Battelli, L., & Caramazza, A. (2015). The origin of word-related motor activity. *Cerebral Cortex*, *25*, 1668–1675. https://doi.org/10.1093/cercor/bht423

Papeo, L., Wurm, M. F., Oosterhof, N. N., & Caramazza, A. (2017). The neural representation of human versus nonhuman bipeds and quadrupeds. *Scientific Reports*, *7*(1), 1–8. https://doi.org/10.1038/s41598-017-14424-7

Park, J., Josephs, E., & Konkle, T. (2022). Ramp-shaped neural tuning supports graded population-level representation of the object-to-scene continuum. *Scientific Reports*, *12*(1), 1–14. https://doi.org/10.1038/s41598-022-21768-2

Pitcher, D., & Ungerleider, L. G. (2021). Evidence for a Third Visual Pathway Specialized for Social Perception. *Trends in Cognitive Sciences*, *25*(2), 100–110. https://doi.org/10.1016/j.tics.2020.11.006.Evidence

Poyo Solanas, M., Vaessen, M. J., & de Gelder, B. (2020). The role of computational and subjective features in emotional body expressions. *Scientific Reports*, *10*, 6202. https://doi.org/10.1038/s41598-020-63125-1

Proklova, D., Kaiser, D., & Peelen, M. V. (2016). Disentangling representations of object shape and object category in human visual cortex: The animate–inanimate distinction. *Journal of cognitive neuroscience*, *28*(5), 680–692.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, *115*(3), 211–252. https://doi.org/10.1007/s11263-015-0816-y

Schwarzbach, J. (2011). A simple framework (ASF) for behavioral and neuroimaging experiments based on the psychophysics toolbox for MATLAB. *Behavior Research Methods*, *43*(4), 1194–1201. https://doi.org/10.3758/s13428-011-0106-8

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*(1), 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061

Tamir, D. I., & Thornton, M. A. (2018). Modeling the predictive social mind. *Trends in Cognitive Sciences*, *22*(3), 201–212. https://doi.org/10.1016/j.tics.2017.12.005

Tamir, D. I., Thornton, M. A., Contreras, J. M., & Mitchell, J. P. (2016). Neural evidence that three dimensions organize mental state representation: Rationality, social impact, and valence. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(1), 194–199. https://doi.org/10.1073/pnas.1511905112

Tarhan, L., & Konkle, T. (2020a). Reliability-based voxel selection. *NeuroImage*, *207*(July), 116350. https://doi.org/10.1016/j.neuroimage.2019.116350

Tarhan, L., & Konkle, T. (2020b). Sociality and interaction envelope organize visual action representations. *Nature Communications*, *11*, 3002. https://doi.org/10.1038/s41467-020-16846-w

Tarhan, L., De Freitas, J., & Konkle, T. (2021). Behavioral and neural representations en route to intuitive action understanding. *Neuropsychologia*, *163*, 108048. https://doi.org/10.1016/j.neuropsychologia.2021.108048

Thornton, M. A., & Tamir, D. I. (2024). Neural representations of situations and mental states are composed of sums of representations of the actions they afford. *Nature Communications*, *15*(1), 620.

Thornton, M. A., & Tamir, D. I. (2022). Six dimensions describe action understanding: The ACT-FASTaxonomy. *Journal of Personality and Social Psychology*, *122*(4), 577–605. https://doi.org/10.1037/pspa0000286

Torralba, A., & Oliva, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*(3), 145–175.

Tucciarelli, R., Wurm, M. F., Baccolo, E., & Lingnau, A. (2019). The representational space of observed actions. *eLife*, *8*, e47686. https://doi.org/10.7554/eLife.47686

Turella, L., Rumiati, R., & Lingnau, A. (2020). Hierarchical action encoding within the human brain. *Cerebral Cortex*, *30*(5), 2924–2938. https://doi.org/10.1093/cercor/bhz284

Vinton, L. C., Preston, C., de la Rosa, S., Mackie, G., Tipper, S. P., & Barraclough, N. E. (2023). Four fundamental dimensions underlie the perception of human actions. *Attention, Perception, & Psychophysics*, *86*, 536–558. https://doi.org/10.3758/s13414-023-02709-1

Walbrin, J., Downing, P., & Koldewyn, K. (2018). Neural responses to visually observed social interactions. *Neuropsychologia*, *112*(February), 31–39. https://doi.org/10.1016/j.neuropsychologia.2018.02.023

Watson, C. E., & Buxbaum, L. J. (2014). Uncovering the architecture of action semantics. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(5), 1832–1848. https://doi.org/10.1037/a0037449

Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. *NeuroImage*, *21*, 1732–1747. https://doi.org/10.1016/j.neuroimage.2003.12.023

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage*, *14*, 1370–1386. https://doi.org/10.1006/nimg.2001.0931

Wurm, M. F., & Caramazza, A. (2019a). Distinct roles of temporal and frontoparietal cortex in representing actions across vision and language. *Nature Communications*, *10*(1), 1–10. https://doi.org/10.1038/s41467-018-08084-y

Wurm, M. F., & Caramazza, A. (2019b). Lateral occipitotemporal cortex encodes perceptual components of social actions rather than abstract

representations of sociality. *NeuroImage*, *202*, 116153. https://doi.org/10.1016/j.neuroimage.2019.116153

Wurm, M. F., & Caramazza, A. (2022). Two 'what' pathways for action and object recognition. *Trends in Cognitive Sciences*, *26*(2), 103–116. https://doi.org/10.1016/j.tics.2021.10.003

Wurm, M. F., & Schubotz, R. I. (2012). Squeezing lemons in the bathroom: Contextual information modulates action recognition. *NeuroImage*, *59*(2), 1551–1559. https://doi.org/10.1016/j.neuroimage.2011.08.038

Wurm, M. F., & Schubotz, R. I. (2017). What's she doing in the kitchen? Context helps when actions are hard to recognize. *Psychonomic Bulletin and Review*, *24*(2), 503–509. https://doi.org/10.3758/s13423-016-1108-4

Wurm, M. F., Ariani, G., Greenlee, M. W., & Lingnau, A. (2015). Decoding concrete and abstract action representations during explicit and implicit conceptual processing. *Cerebral Cortex*, *26*, 3390–3401. https://doi.org/10.1093/cercor/bhv169

Wurm, M. F., Caramazza, A., & Lingnau, A. (2017). Action categories in lateral occipitotemporal cortex are organized along sociality and transitivity. *The Journal of Neuroscience*, *37*(3), 562–575. https://doi.org/10.1523/JNEUROSCI.1717-16.2017

Xia, M., Wang, J., & He, Y. (2013). BrainNet viewer: A network visualization tool for human brain connectomics. *PLoS One*, *8*(7), e68910. https://doi.org/10.1371/journal.pone.0068910

Zhuang, T., Kabulska, Z., & Lingnau, A. (2023). The representation of observed actions at the subordinate, basic and superordinate level. *The Journal of Neuroscience*, *43*, 8219–8230. https://doi.org/10.1523/JNEUROSCI.0700-22.2023

Zimmermann, M., Mars, R. B., de Lange, F. P., Toni, I., & Verhagen, L. (2018). Is the extrastriate body area part of the dorsal visuomotor stream? *Brain Structure and Function*, *223*(1), 31–46. https://doi.org/10.1007/s00429-017-1469-0

Zimmermann, M., Verhagen, L., de Lange, F. P., & Toni, I. (2016). The extrastriate body area computes desired goal states during action planning. *ENeuro*, *3*(2), 1918–1921. https://doi.org/10.1523/ENEURO.0020-16.2016

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Kabulska, Z., Zhuang, T., & Lingnau, A. (2024). Overlapping representations of observed actions and action-related features. *Human Brain Mapping*, *45*(3), e26605. https://doi.org/10.1002/hbm.26605