NeuroImage 310 (2025) 121113

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

# Touching-untouching patterns organize action representation in the inferior parietal cortex

Jennifer Pomp [a,b,*], Moritz F. Wurm [c], Rosari N. Selvan [a,b], Florentin Wörgötter [d], Ricarda I. Schubotz [a,b]

[a] Department of Psychology, University of Münster, Germany
[b] Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Münster, Germany
[c] Center for Mind/Brain Sciences (CIMeC), University of Trento, Rovereto, Italy
[d] Institute for Physics 3 – Biophysics and Bernstein Center for Computational Neuroscience, (BCCN), University of Göttingen, Germany

ARTICLE INFO

*Keywords:*
Action observation
Representational Similarity Analysis
Inverse MultiDimensional Scaling
Semantic Event Chain
aIPL
Object-directed action

ABSTRACT

At an abstract temporospatial level, object-directed actions can be described as sequences of touchings and untouchings of objects, hands, and the ground. These sparse action codes can effectively guide automated systems like robots in recognizing and responding to human actions without the need for object identification. The aim of the current study was to investigate whether the neural processing of actions and their behavioral classification relies on the action categorization derived from the touching-untouching structure. Here we show, using a representational similarity analysis of functional MRI data from two experiments, that action representations in left anterior intraparietal sulcus (aIPS) are particularly associated with this categorization of touching-untouching structures. Within the examined action observation network, only the touching-untouching category model selectively correlated with the representational profile of the left aIPS. The behavioral results showed a significant relation between the touching-untouching structure and the observers' judgments on the similarity of actions with weakly-informative objects. Extending prior research on touchings and untouchings as meaningful anchor points for explicit action segmentation, our findings suggest that touching-untouching sequences serve as an organizing principle in inferior parietal action representation.

## 1. Introduction

Action recognition is crucial in many modern applications like video surveillance, human-computer interaction, web-video search and retrieval, robotics, elderly care, and sports analytics (Herath et al., 2017). Accordingly, automatic action recognition is a popular and promising field of basic and applied research. However, despite having similarities to static image analysis, video data analysis is far more complicated (Jiao et al., 2022). The main challenges of this continuous process arise due to the movement of objects and changes in perspective leading to changing size and appearance, blurriness, and changing light intensities. Moreover, current deep learning networks still depend heavily on extensive pretraining (Han et al., 2021), and object recognition remains particularly difficult due to the complexity and variability within object categories (Liang and Wan, 2020). Therefore, the ultimate goal is to enable machines to learn actions directly from video observations without human intervention. Interestingly, while these tasks pose significant challenges for machines, the human brain excels at action recognition effortlessly, and researchers are actively exploring the underlying mechanisms that enable this ability.

In the attempt to let a robot recognize manipulations performed by a human and let it execute these itself, Aksoy et al. (2011) developed the concept of semantic event chains (SECs). This approach formalizes object-directed actions as sequences of relational changes in the form of touchings (T) and untouchings (U) of objects, hands, and the ground (TUs, hereafter). Thus, SECs or TU sequences encode the contact states between surfaces, without prioritizing the hand over objects or the table surface. Most importantly, robots can recognize and execute manipulations using this SEC-based representation without prior object knowledge (Aksoy et al., 2011). It showed that computer vision using the SEC approach was able to distinguish between 30 different one-handed object manipulations typical of everyday life (Wörgötter et al., 2013). Against this background, the question arose whether these TUs, that are useful for robots to recognize actions, could be also used in

* Correspondence author at University of Münster, Fliednerstrasse 21, Münster 48149, Germany.
*E-mail address:* jennifer.pomp@uni-muenster.de (J. Pomp).

https://doi.org/10.1016/j.neuroimage.2025.121113
Received 20 November 2024; Received in revised form 31 January 2025; Accepted 3 March 2025
Available online 9 March 2025

human action recognition. And indeed, previous research has shown that TUs are meaningful anchor points for individual action segmentation (Pomp et al., 2024, 2021). Thus, TUs have been shown to be objective and meaningful event boundaries, and certain sequences of TUs emerged as particularly relevant for individuals to determine action steps (Pomp et al., 2024, 2021). Moreover, we found that the observation of Ts and the observation of Us were each associated with different brain activation patterns, emphasizing their importance in the analysis of observed actions.

Interestingly, taking the entire TU sequence of an action into account, also action categories arise from the SEC formalism. For example, turning, pulling, and poking an object all share the same TU sequence and belong to the action category termed "rearrange". Their TU sequence differs from, for example, breaking, ripping-off, and uncovering by picking and placing, that also share a similar TU sequence and build another category termed "break" (Wörgötter et al., 2013). In the current study, we built on these findings and investigated whether the action categories derived from the TU structure of an action might be informative for the brain, and whether they are also reflected in categorization behavior. Specifically, the question was whether there are brain regions that reflect action categories as predicted by their full TU sequence and whether behaviorally determined categories resemble them. To address this question, we employed fMRI-based representational similarity analyses as well as inverse multidimensional scaling (MDS).

Based on the SEC-based ontology of manipulation actions, video stimuli were recorded that belonged to six separate action categories. In two separate experiments, two non-overlapping groups of participants passively watched these videos during the MRI scan. Importantly, the two experiments were similar except for the manipulated objects in the videos. In the first experiment, manipulation actions were performed with daily life objects (e.g., a calculator, a cup, or a piggy bank) and in

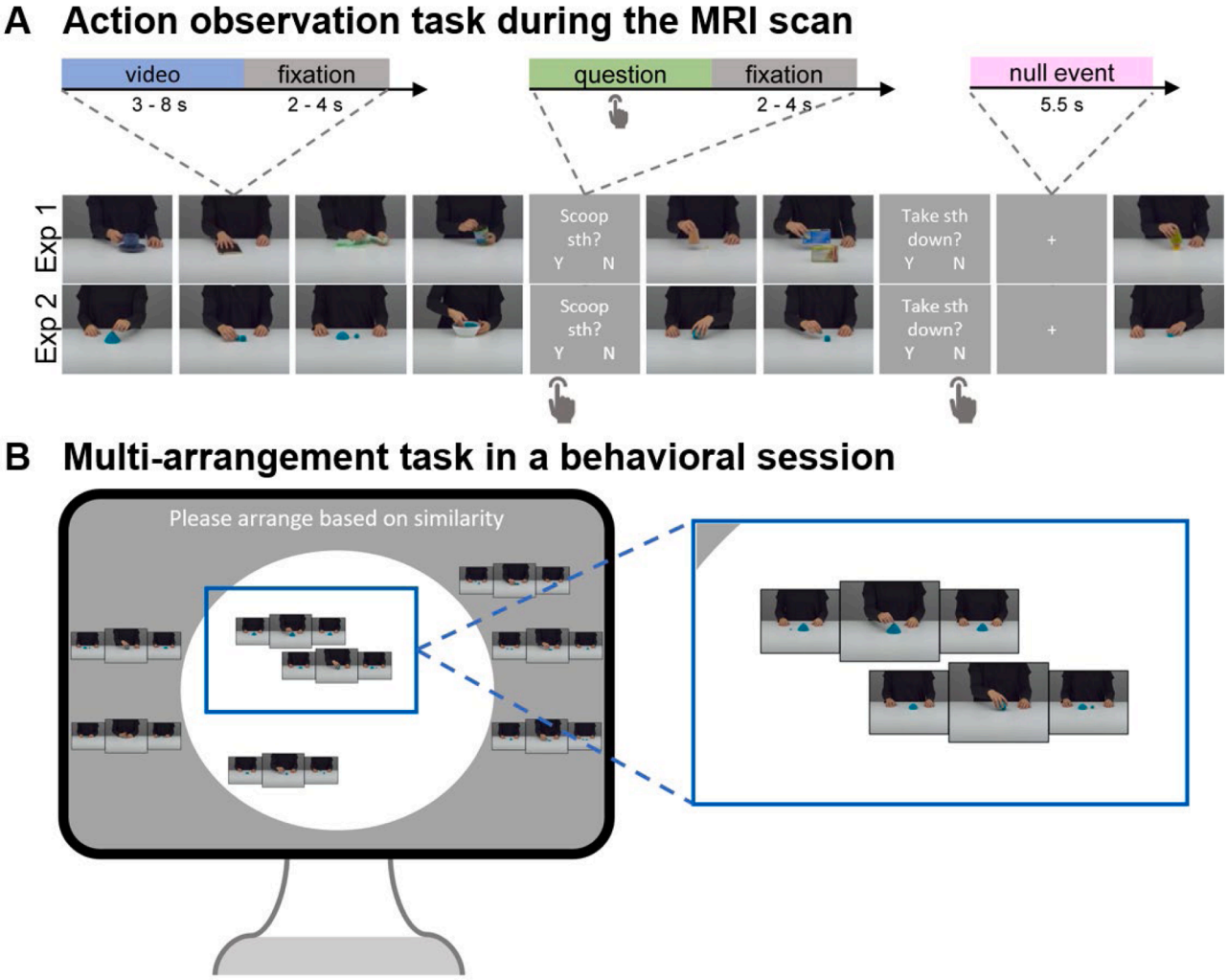


**Fig. 1.** Experimental task designs. (A) During the fMRI scan, video trials (action videos followed by a jittered inter-stimulus-interval that showed a white fixation cross) and null event trials (showing a white fixation cross) were passively attended to. Question trials (question followed by a jittered inter-stimulus-interval that showed a white fixation cross) required participants to confirm or reject by button press an action description with regard to the preceding action video. The question disappeared only after button press and followed 14% of the action videos. For the video trials, here, each single frame image represents a full action video plus inter-stimulus-interval as indicated by the dashed lines. In sum, 308 videos, 42 questions, and 49 null events were presented to each participant, split into seven blocks with short breaks in between. The task design was equivalent in experiment 1 (Exp 1) and experiment 2 (Exp 2), but the action videos differed, as shown in the lower part of (A). Example videos are provided via the Action Video Corpus Münster (AVICOM, https://www.uni-muenster.de/IVV5PSY/AvicomSrv/), where the entire video stimulus material is available upon request. (B) In a subsequent behavioral session, each participant did a multi-arrangement task with the time course of the action videos being represented by a sequence of three single frame images (see blue box zooming in the items). The participants spatially arranged subsets of these items within a 2D white circular arena on a 27″ computer screen using the computer mouse. (Dis-)similarity between items was expressed through the relative spatial position of the items. The task was implemented in the Meadows web-based platform for psychophysical experiments (http://meadows-research.com). The figure shows an exemplary screen for the evaluation of similarity relationships in the stimulus material of experiment 2.

the second experiment, these objects were replaced by formed play dough pieces that did not resemble meaningful objects (see Fig. 1). This enabled us to focus on the actions' common TU sequences and to control for potential effects associated to object identity. For both experiments, we created a similarity model that captured the action categories as defined by the TU structure given by the SEC framework (referred to as "TU model", hereafter). Furthermore, to investigate to which extent the TU structure of actions influenced how participants subjectively categorize actions, we created a rating-based behavioral similarity model and compared it to the TU model. This model was created using the similarity ratings of a behavioral post-MRI multi-arrangement task that employed the inverse MDS approach by Kriegeskorte and Mur (2012), and we refer to it as "MDS model" hereafter.

With regard to brain activation, we reasoned that if two actions are defined as similar in TU structure, they should evoke similar activation patterns in brain areas coding actions in terms of SEC-like information. Accordingly, we performed a specific form of multivoxel pattern analysis known as representational similarity analysis (RSA) to investigate which brain regions reflect the action categories as predicted by their TU structure. We used a searchlight RSA for a brain-wide analysis and a region of interest (ROI) RSA focused on the action observation network, respectively. Here, we specifically focused on the left anterior inferior parietal lobule (aIPL), as this region has been consistently shown to be sensitive to (observed) hand-object interactions (Murata et al., 2016; Vingerhoets, 2014), observation of touch (Chan and Baker, 2015), physical scene understanding (e.g., how objects rest on each other and how colliding objects behave; Fischer et al., 2016), and reasoning on physical object properties (Reynaud et al., 2016). Moreover, lesion studies have shown that the left parietal lobe plays a major role in apraxia, which comes with severe impairments to associate objects to appropriate actions (Buxbaum and Randerath, 2018). While the aIPL apparently enables a range of closely related functions rather than a single homogeneous one, this profile seemed best suited to represent the TU structure of actions.

## 2. Methods

For the current investigation, we used fMRI data that was analyzed and published in two previous works (Pomp et al., 2024, 2021). These original datasets included also unpublished data of a post-fMRI multi--arrangement task, which were the focus of the current work. Furthermore, in contrast to the preceding analyses, fMRI-based RSA was used here to identify neural representations of action categories. In several passages in the methods section, we refer to the preceding publications for detailed descriptions. Yet, we repeat details if necessary for immediate understanding of the current study.

### 2.1. Participants

#### 2.1.1. Experiment 1

As reported in Pomp et al. (2021), 31 participants ($M_{age}$ = 23.84 years, SD = 3.01, age range = 18 - 31 years, 25 women, 6 men) participated in experiment 1. One additional data set was excluded from the analyses as the participant misunderstood the instructions. The participants were all right-handed as determined by the Edinburgh Handedness Inventory (Oldfield, 1971), had intact color perception, normal or corrected-to-normal vision, reported no history of neurological or psychiatric diseases, and self-reportedly met the criteria for MRI scanning. The experiment was conducted according to the Declaration of Helsinki and approved by the local Ethics Committee of the Faculty of Psychology (University of Münster, Germany). The participants provided informed consent and either received course credits (29 of the participants were students of the University of Münster) or were paid for their participation.

#### 2.1.2. Experiment 2

As reported in Pomp et al. (2024), 33 right-handed participants ($M_{age}$ = 23.03 years, SD = 3.06, age range = 18–29 years, 28 women, 5 men) took part in experiment 2. The participants reported having no history of neurological or psychiatric disorders, intact color perception, and had not taken part in related precursor studies. In the course of the behavioral part of the experiment one participant dropped out; hence, this participant's data set was not included in the behavioral model construction but was included in the fMRI data set. The behavioral analysis comprised the data of 32 participants (27 women, 5 men) aged between 18 and 29 years ($M_{age}$ = 22.88, SD = 3.13). The participants gave written informed consent in voluntarily participating in the experiment and were self-reportedly suitable for MRI scanning. They were either paid for their participation or received course credits. The experiment was conducted according to the Declaration of Helsinki and approved by the local Ethics Committee of the Faculty of Psychology (University of Münster, Germany).

### 2.2. Stimulus material

The object-directed manipulation actions for the video stimuli were chosen according to the SEC framework (Wörgötter et al., 2013). This framework includes transitive actions involving one active hand and one or two objects. Twelve of these actions were selected for the current studies that belonged to six action categories: Rearrange (turn; pull), break (rip off; uncover), destroy (cut; scoop), destruct (take down; take away), construct (put on top; put together), and hide (put over; put into). For experiment 1, each action was recorded using four different objects which resulted in 48 object manipulations (6 action categories x 2 actions x 4 objects). For experiment 2, all 12 actions were performed with formed pieces of blue dough (6 action categories x 2 actions).

Action videos were recorded using an industrial camera and the created video material showed the actress from the front up to the shoulders performing the action on a white table (see Fig. 1). Subsequently, the videos were vertically mirrored so that the actions looked like being performed by the left hand. Each participant saw 50% of the action videos mirrored. For more details on the creation of the video material see Pomp et al. (2021, 2024). In order to control the transition probabilities of the video trials in the experiment, the stimulus sequence was designed as a second-level counterbalanced De Bruijn sequence with seven conditions (6 action categories + null condition) using the De Bruijn cycle generator by Aguirre et al. (2011) and NeuroDebian 8.0.0 (Halchenko and Hanke, 2012). See Pomp et al. (2021, 2024) for details.

For the behavioral multi-arrangement task, each action was represented as a sequence of three video images, i.e., single frame images from the start, middle, and end of the action video (Fig. 1B). Therefore, the sequence of images showed the start state, the manipulation, and the end state of the action whereby the middle one was enlarged and highlighted. The image triplets were supposed to represent the course of action of the respective video, that had repeatedly been seen.

### 2.3. Experimental procedure and tasks

In both experiments, the participants completed three experimental sessions. The first session comprised the MRI session. As described earlier (Pomp et al., 2024, 2021), the participants passively attended to the action videos during the MRI scan. Attention capturing questions followed 14% of the videos asking whether an action description was appropriate for the just seen action video. Participants responded by pressing one of two response keys with their right index and middle finger to reject or affirm the given description. Their response was necessary for the experiment to continue ensuring that participants engaged in attending and recognizing the actions shown in the videos. See Fig. 1A for the experimental trial design. See Pomp et al. (2021, 2024) for details on stimulus presentation in the scanner and the procedure at the scanner.

As second part of the third session, participants did a multi-arrangement task (Fig. 1B). We adapted the multi-arrangement method proposed by Kriegeskorte and Mur (2012), which uses inverse MDS to obtain a distance matrix from multiple spatial arrangements of subsets of items within a 2D space. The participants did this task in a behavioral laboratory using the Meadows web-based platform for psychophysical experiments (http://meadows-research.com). Participants arranged the actions, as represented by a sequence of three single frame images, in a two-dimensional space (on a computer screen of 27″) thereby expressing (dis-)similarity through the relative spatial position of the items. Due to the limited screen size, only a subset of actions was arranged per trial. In experiment 1 the subset included a maximum of 11 actions being simultaneously presented and in experiment 2 a maximum of six actions. These numbers were chosen to give participants enough space to arrange items while having enough items per trial to efficiently gather pairwise similarity ratings in a reasonable total number of trials. In experiment 1, in total the pairwise similarity of 48 actions (12 manipulations x 4 objects) was estimated and in experiment 2, as the object dimension was absent, the pairwise similarity of 12 actions was assessed. Regarding the subset of actions per trial, the concrete items for the second trial's subsets of stimuli (and all subsequent) were determined using an adaptive algorithm that provided the optimal evidence for the pairwise similarity estimates that were inferred from the 2D arrangement of the items on the screen (see Kriegeskorte & Mur, 2012, for details). Therefore, some of the trials included fewer actions than the set maximum, as determined by the algorithm, which allowed participants to refine their judgments within the given arena space. The participants were instructed to drag and drop the stimuli within a circular arena using the computer mouse and place similar actions closer together and dissimilar ones further apart. No explicit instruction was given on which feature to use for similarity. The relative inter-item distances, rather than the absolute screen distances, represented dissimilarities between the items from trial to trial. All items had to be placed in the arena. The task terminated automatically either when a sufficient signal to noise ratio was achieved (the minimum required evidence weight was set to 0.5), or when the maximum session length of 60 minutes was reached. Subsequently, the participants completed a short survey. Please note that the participants knew the action videos very well as they also saw the videos in a behavioral test-retest regime and manually segmented meaningful action steps by button press in the

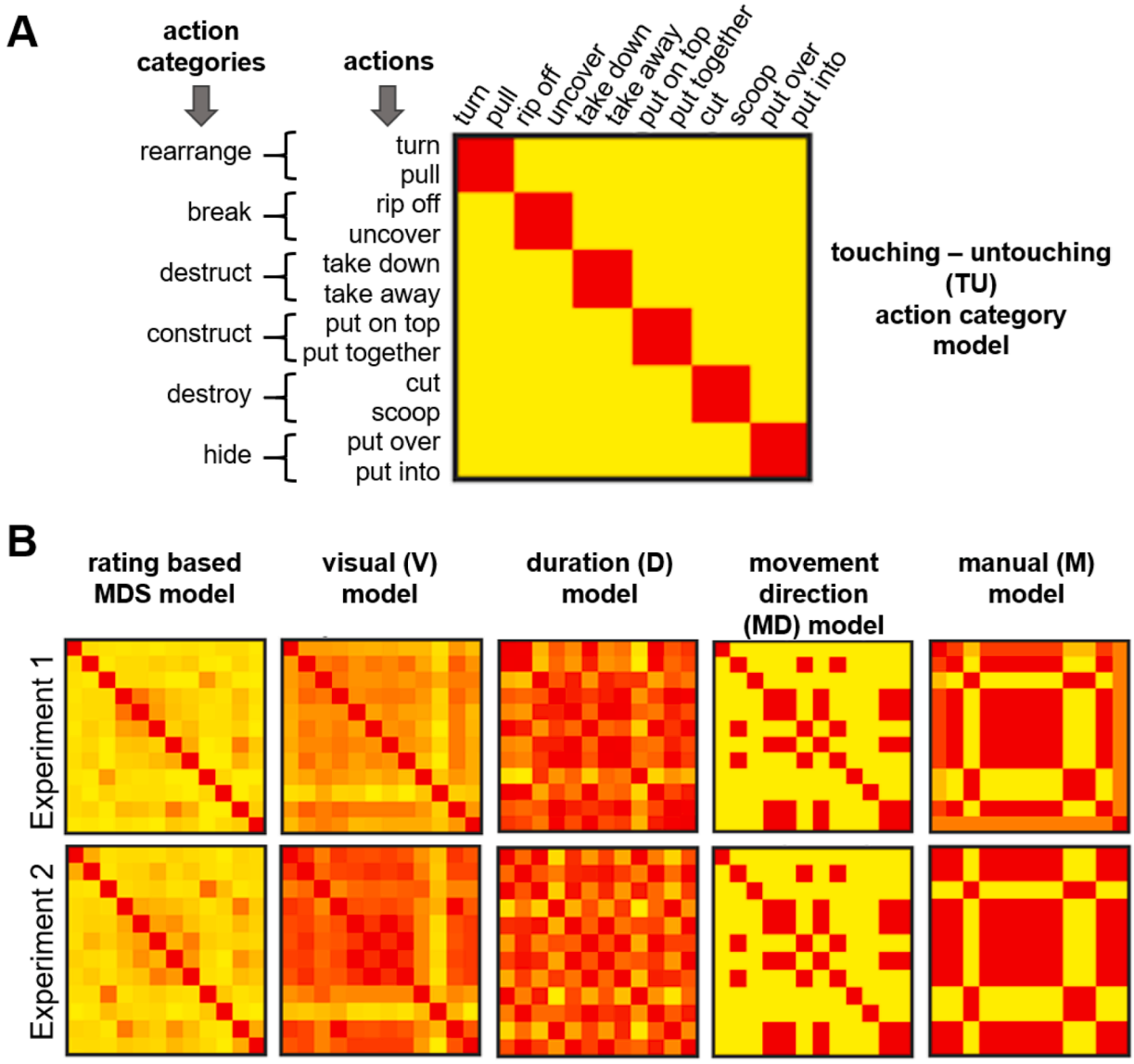


**Fig. 2.** Model Representational Dissimilarity Matrices (RDMs). (A) The touching-untouching (TU) action category model shown as model RDM with row and column labels, which are representative of the other models. Action categories that contain the concrete actions are given on the left side. (B) Visualizations of the second to sixth model RDMs for experiment 1 (top row) and experiment 2 (bottom row). The scale is from dissimilar (yellow) to similar (red).

second and third experimental session (see Pomp et al., 2021, 2024). We therefore assumed that a triplet representation (i.e., three single frame images) of each action video was sufficient to remind the participants of the respective action video and to let them judge the similarity based on the plot of the action (i.e., the interaction with the object).

### 2.4. Representational dissimilarity matrices (RDMs)

We created six model RDMs to use them in the RSA (Fig. 2). The first and second models were of main interest for the current study while the third to sixth models were created to control for perceptual stimulus features that might covary with the first and second one. The second model based on participants' judgments. The other models were objective and were derived directly from the video stimuli either through experimenters' judgments (MD model, M model) or feature extraction (V model, D model).

(1) The first model created was a dissimilarity model that captures the action categories as defined by their TU structure according to the SEC framework. We compared each of the 12 actions with each other in terms of whether they belong to the same SEC category (dissimilarity = 0) or not (dissimilarity = 1). This *TU model* was identical for experiment 1 and 2. All following models were created separately for experiment 1 and 2. Importantly, TU categories did not systematically covary with the position of the object relative to the subject. Similarly, the TU categories did not covary with specific grip types, as the same TU sequence was performed with different objects, each requiring a different type of grip.
(2) To investigate to which extent TUs influenced how participants subjectively categorize actions, we created a second model based on similarity ratings. The inverse MDS results from the post-fMRI multi-arrangement task were used for this *MDS model*. The resulting pairwise dissimilarity estimates, i.e., the Euclidean distance between the two actions of a pair, were averaged across trials, resulting in a 48 × 48 (Exp. 1) or a 12 × 12 (Exp. 2) dissimilarity matrix for each participant. Each dissimilarity matrix was normalized by dividing each value by the maximum value of the matrix. Finally, the dissimilarity matrices were averaged across participants and across the exemplars of each of the 12 actions.
(3) To create a low-level visual similarity or *V model*, we computed the pixelwise similarity of each action video. In a first step, we averaged the frames of each action video. We then vectorized the resulting average frames to obtain a pixelwise vector for each action exemplar. The pixelwise action vectors were correlated with each other, resulting in a 12 × 12 correlation matrix. This matrix was transformed into a dissimilarity matrix by subtracting it from 1 (1 – r). Computing pixelwise similarity is a common practice in RSA to capture general low-level visual representations of stimuli (e.g., Kriegeskorte et al., 2008; Peelen and Caramazza, 2012), including videos (de Vries and Wurm, 2023; Wurm and Lingnau, 2015). To capture more specific levels of low-to-midlevel representations upstream of basic pixelwise similarity, several options exist, including FSIM (others are e.g. Radon, silhouette, DNN layers), and specifically for videos, it is also possible to use models based on optical flow and motion energy, as well as more sophisticated models such as Dense Trajectory models and Space-Time-Interest Points (Urgen et al., 2019). However, such higher-level models were beyond the purpose of this study, since we only wanted to control for basic low-level visual similarity of our stimuli.
(4) Since the videos differed in terms of their duration, we computed a video duration or *D model* by computing the absolute difference between video durations.
(5) We created a movement direction or *MD model* that captured whether an object was moved away from the body toward another object or vice versa, i.e., away from another object toward the body. We compared each of the actions with each other with respect to whether the actions were made in the same direction (0 = similar) or in different directions (1 = dissimilar). This model turned out to be similar for experiment 1 and 2. Please note, this model was independent of left-right movements, as these were balanced by showing the stimuli also vertically mirrored (see *Section 2.2. Stimulus material*).
(6) Finally, some of the actions involved both hands, as the non-dominant hand stabilized the object for manipulation, whereas other actions were unimanual, i.e., the non-dominant hand remained still on the table without touching an object. We computed a manual or *M model* by pairwise comparing each of the actions with each other with respect to whether they both were unimanual or bimanual (0 = similar) or one action was unimanual and the other action was bimanual (1 = dissimilar).

Regarding models three to six, each of the perceptual dissimilarity matrices were averaged across the exemplars of each of the 12 actions where applicable.

### 2.5. Behavioral data analysis

#### 2.5.1. Behavior during the MRI scan

During the MRI scan, attention-maintaining questions irregularly followed the action videos. We evaluated the accuracy and calculated the median reaction time of the correct responses.

#### 2.5.2. Analyses regarding the MDS model

To assess the inter-subject reliability of the MDS models, obtained from the multi-arrangement task, we correlated each participant's pairwise similarities (i.e., the lower triangle of the correlation matrix) with the averaged pairwise similarities of the remaining subjects using leave-one-subject-out cross-validation in MATLAB (https://www.mathworks.com). The resulting correlation values were averaged.

Moreover, to ensure that the MDS models of experiment 1 and 2 reliably captured the behavioral similarity across different types of stimuli (i.e., with and without real objects), we correlated these two models in MATLAB.

Finally, we tested whether the participants' subjective similarity ratings could be explained by the TU model. To this end, we first calculated a multiple regression analysis in JASP (JASP Team, 2024) including the four control models (V, D, MD, and M) as predictor variables, and the subjective similarity ratings as outcome variable. Predictor variables that did not significantly predict the outcome variable were eliminated from the regression equation. Subsequently, the remaining control models were added to the null model to investigate the portion of variance explained by the alternative model which included the TU model as predictor variable.

### 2.6. fMRI data analysis

#### 2.6.1. fMRI data acquisition, preprocessing, and design specification

For both experiments, MRI data were acquired at the Translational Research Imaging Center (TRIC) of the University Hospital Münster using a 3-Tesla Siemens Magnetom Prisma MR tomograph with a 20-channel head coil. First, high-resolution T1-weighted images were obtained by a 3D-multiplanar rapidly acquired gradient-echo (MPRAGE) sequence (scanning parameters: 192 slices, TR = 2130 ms, TE = 2.28 ms, slice thickness = 1 mm, FoV = 256 × 256 mm2, flip angle = 8°). Subsequently, a blood-oxygen-level-dependent (BOLD) contrast was measured by gradient-echo echoplanar imaging (EPI). Seven EPI sequences measured the seven experimental blocks (scanning parameters: 33 slices, TR = 2000 ms, TE = 30 ms, slice thickness = 3 mm, FoV = 192

× 192 mm2, flip angle = 90°). The fMRI data have been used in two previous papers (Pomp et al., 2024, 2021).

The anatomical and functional images were preprocessed using the Statistical Parametric Mapping software (SPM12; The Wellcome Centre for Human Neuroimaging, London, UK) implemented in MATLAB R2019a. The preprocessing included slice time correction to the first slice, realignment to the mean image, co-registration of the individual structural scan to the mean functional image, normalization into the standard anatomical MNI space (Montreal Neurological Institute, Montreal, QC, Canada) on the basis of segmentation parameters, as well as spatial smoothing using an isotropic 3 mm full-width at half maximum (FWHM) Gaussian kernel. A 128 s temporal high-pass filter was applied to the time-series of functional images to remove low-frequency noise.

The functional images were statistically analyzed using SPM12 applying a general linear model (GLM) for serially autocorrelated observations (Friston et al., 1994; Worsley and Friston, 1995) and a convolution with the canonical hemodynamic response function (HRF). As regressors of no interest, the six subject-specific rigid-body transformations obtained from realignment were included. To allow for T1-equilibrium effects, the volumes of the first two video presentations of each EPI were discarded. The constructed GLM included 14 regressors of interest coding for onsets and durations of the 12 action video types, null events, and question trials.

#### 2.6.2. Representational similarity analysis (RSA)

The RSA (Kriegeskorte et al., 2008) was carried out using the CoSMoMVPA toolbox (Oosterhof et al., 2016) and the representational similarity toolbox (Nili et al., 2014). For each participant and run, beta weights of the experimental conditions were estimated using design matrices containing predictors of the 12 action conditions, null trials, question trials, and of 6 parameters resulting from 3D motion (translation and rotation) correction. Each predictor was convolved with a dual-gamma hemodynamic impulse response function (Friston et al., 1998). Each trial was modelled as an epoch lasting from video onset to offset. The resulting reference time courses were used to fit the signal time courses of each voxel. The resulting beta weights were averaged across the seven runs to obtain one beta weight per condition and voxel. The searchlight (Kriegeskorte et al., 2006) and ROI RSA were performed in volume space using spherical ROIs with a radius of 12 mm.

*2.6.2.1. ROI RSA.* For the ROI generalized linear models (GLM) RSA, ROIs were defined in the action observation network based on the peak coordinates of the univariate contrast of all actions vs. null condition (for coordinates see Table 1). While in experiment 2, six bilateral ROIs were defined, experiment 1 did only yield five bilateral and one unilateral ROI (i.e., no univariate peak in right ventral premotor cortex was detected). For each participant, ROI, and condition, we extracted and vectorized the beta values of the ROI to obtain one vector of beta values per action. For each vector, we demeaned the beta values across voxels by subtracting the mean beta value from each individual beta value. Next, we correlated the vectors with each other resulting in a 12 × 12 correlation matrix per ROI and participant. The neural correlation matrices were transformed into a neural RDM (1 – r). The pairwise action comparisons of neural and model RDMs were vectorized, z-scored, and entered as independent and dependent variables, respectively, into a multiple regression RSA. Resulting beta coefficients were entered into a repeated measures analysis of variance (ANOVA) with between-subject factor Experiment(2), and within-subject factors ROI(11), and MODEL(6) to see whether ROIs dissociated before entering the beta coefficients into one-sided signed-rank tests across participants (Nili et al., 2014). Statistical results were corrected for the number of ROIs and tested models (11 ROIs * 6 models = 66 tests) using the false discovery rate (FDR) at q = 0.05 (Benjamini and Yekutieli, 2001).

*2.6.2.2. Searchlight RSA.* The searchlight GLM RSA was performed using identical parameters as reported above. For all searchlight analyses, individual beta coefficient maps were Fisher transformed and entered into one-sample *t*-tests (Oosterhof et al., 2016). Statistical maps were corrected for multiple comparisons using an initial voxelwise threshold of p = .001 and 10,000 Monte Carlo simulations as implemented in the CoSMoMVPA toolbox (Oosterhof et al., 2016). Resulting z maps were used to threshold statistical maps (at p = .05 at the cluster level), which were projected on a cortex-based aligned group surface for visualization.

To test for multicollinearity between the models, we computed condition indices (CI), variance inflation factors (VIF), and variance decomposition proportions (VDP) using the colldiag function for MATLAB. The results of these tests (Exp. 1: CI<4, VIF<2.2, DVP<0.8; Exp. 2: CI<4, VIF<2.6, DVP<1.0) revealed no indications of multicollinearity that could give rise to potential estimation problems (Belsley et al., 1980).

## 3. Results

### 3.1. Behavioral results

#### 3.1.1. Behavior during the MRI scan

To control whether the participants performed the task in the scanner accurately, their performance was analyzed: In experiment 1, participants responded on average in 88.5 % correct (SD = 7.5) with a median reaction time of 1615 ms; and in experiment 2, the mean accuracy was 90.7 % (SD = 6.6) with a median reaction time of 1511 ms.

#### 3.1.2. Results regarding the MDS model

In a first step, we aimed at assessing the reliability of the MDS models obtained from the multi-arrangement task. In both experiments, we observed robust correlations between the ratings of individual subjects: For experiment 1, we found a mean inter-subject correlation of $r(64) = 0.66$, and all but two subjects' data correlated significantly with the group mean (all *ps* < 0.01, *one-sided; for the two remaining subjects:* $p = .09$, and $p = .45$; *one-sided*). For experiment 2, the mean inter-subject correlation was $r(64) = 0.53$, and the data of all the subjects

**Table 1**
ROI coordinates.

| | Experiment 1 | | | | | | Experiment 2 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | L | | | R | | | L | | | R | | |
| ROI | x | y | z | x | y | z | x | y | z | x | y | z |
| aIPS | -51 | -25 | 41 | 45 | -22 | 41 | -48 | -25 | 41 | 42 | -28 | 44 |
| LOTC | -45 | -70 | 8 | 51 | -67 | -1 | -42 | -73 | -4 | 48 | -64 | 2 |
| pIPS | -24 | -76 | 35 | 30 | -67 | 32 | -24 | -73 | 32 | 27 | -73 | 38 |
| SPL | -24 | -55 | 59 | 30 | -49 | 59 | -27 | -55 | 62 | 21 | -64 | 59 |
| PMd | -27 | -10 | 53 | 30 | -7 | 59 | -24 | -7 | 56 | 24 | -4 | 53 |
| PMv | -54 | 5 | 38 | – | – | – | -57 | 8 | 29 | 60 | 8 | 32 |

Note. Spheres of 12 mm. ROI = Region of interest, L = left hemisphere, R = right hemisphere, x, y, z = MNI coordinates, aIPS = anterior intraparietal sulcus, LOTC = lateral occipitotemporal cortex, pIPS = posterior intraparietal sulcus, SPL = superior parietal lobule, PMd = dorsal premotor cortex, PMv = ventral premotor cortex.

correlated significantly with the group mean (all *ps* < 0.01, *one-sided*). Moreover, the MDS models of the two experiments strongly correlated with each other ($r(64) = 0.82$, $p < .001$, *one-sided*). This finding suggests that the MDS models reliably captured the subjective similarity across different types of stimuli, i.e., with and without real objects.

Subsequently, to test whether the participants' subjective similarity ratings were predicted by the TU model independent of the four control models (V, D, MD, and M), we conducted a multiple regression analysis. For experiment 1, the test revealed that the D model as well as the V model were no significant predictors (D: $\beta = 0.131$, $t(65) = 1.377$, $p = .174$; V: $\beta = 0.075$, $t(65) = 0.722$, $p = .473$). Therefore, these two predictor variables were deleted from the equation. Then, the remaining two models (MD and M) were included in the null model of the regression analysis. This test revealed that the null model explained 49.0 % of the variance in the subjective similarity ratings ($R^2 = 0.490$, $F(2,63) = 30.224$, $p < .001$) and the alternative model including the TU model as predictor variable explained additional 2.0 % of the variance ($R^2 = 0.510$, $F(1,62) = 2.563$, $p = .115$). The predictor variable of the TU model was no significant predictor of the participants' subjective similarity ratings, $\beta = 0.144$, $t(65) = 1.601$, $p = .115$. This means that the subjectively judged similarity barely changed for a higher TU similarity in experiment 1.

For experiment 2, we proceeded in the same way and found slightly different results: As in experiment 1, the initial test of the control models revealed that the D model as well as the V model were no significant predictors of the subjective similarity ratings (D: $\beta = 0.165$, $t(65) = 1.943$, $p = .057$; V: $\beta = 0.209$, $t(65) = 1.692$, $p = .096$). Therefore, these two predictor variables were deleted from the equation. Then, the MD model and the M model were included in the null model of the regression analysis. The test indicated that the null model explained 49.8 % of the variance in the subjective similarity ratings ($R^2 = 0.498$, $F(2,63) = 31.288$, $p < .001$) and the alternative model including the TU model as predictor variable explained additional 5.1 % of the variance ($R^2 = 0.549$, $F(1,62) = 6.943$, $p = .011$). Thus, the predictor variable of the TU model was a significant predictor of the participants' subjective similarity ratings in experiment 2, $\beta = 0.229$, $t(65) = 2.635$, $p = .011$. This means that the subjectively judged similarity increased for a higher TU-structure similarity in experiment 2.

In sum, a small portion of the variance in the participants' subjective similarity ratings was significantly explained by the TU model in experiment 2 only. In both experiments, the M model and the MD model were significant predictors of the similarity ratings.

### 3.2. Neuroimaging results

As expected, the searchlight GLM RSA revealed that in both, experiment 1 and experiment 2, the TU model predicted the representational organization of actions in the left anterior intraparietal sulcus (aIPS). The clusters in aIPS found in the two experiments strongly overlapped, peaking in the ventral postcentral gyrus and extending anteriorly into the central sulcus and posteriorly into the supramarginal gyrus (Fig. 3). Experiment 1 revealed an additional cluster for the TU model in the left occipital cortex.

The effect of the TU model in left aIPS indicated that the similarity in terms of TU sequence explained the representational variance in this area over and above the control models (for the searchlight result maps of the other models, see the Supplementary Material).

To test whether the aIPS differs significantly from other regions of the action observation network in this respect, we performed a GLM RSA in ROIs of the action observation network (based on univariate cluster peaks), which allows a more sensitive quantitative comparison of RSA effects in the different ROIs. This was done as it could be that also other regions of the action observation network are sensitive to the TU sequence, but failed to survive the conservative correction for multiple comparisons in the searchlight analysis.

Preparatory for the ROI RSA, to see whether the effects in the ROIs dissociate, a repeated measures ANOVA with between-subject factor Experiment(2), and within-subject factors ROI(11), and Model(6) was carried out. The Greenhouse-Geisser corrected results revealed significant main effects for ROI ($F(10,620) = 7.529$, $p < .0001$, $\eta^2_G = 0.006$) and Model ($F(3.39,210.07) = 19.919$, $p < .0001$, $\eta^2_G = 0.094$) as well as a significant interaction effect between ROI and Model ($F(17.78,1102.61) = 6.7$, $p < .0001$, $\eta^2_G = 0.060$) but no main effect for Experiment ($F(1,62) = 0.363$, $p = .549$, $\eta^2_G = 0.0002$). All remaining interaction effects were significant (all *ps* < 0.001).

The ROI RSA (Fig. 4) revealed that in the left aIPS, not only the TU

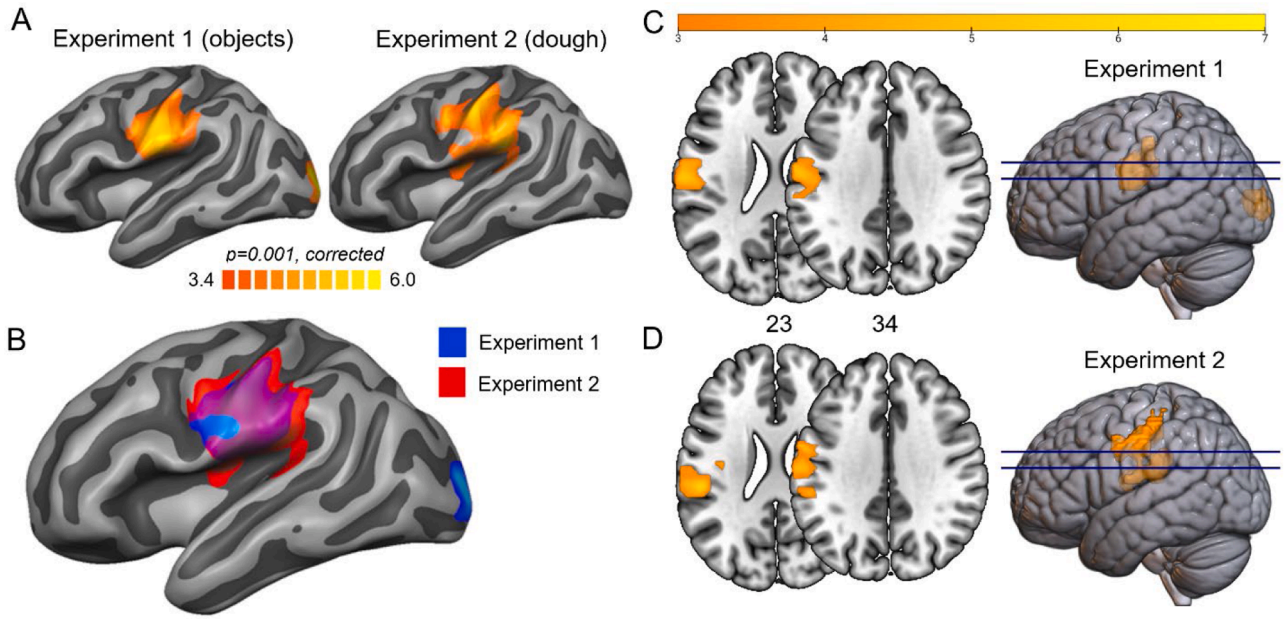


**Fig. 3.** Searchlight GLM RSA result for the TU model separately for experiment 1 and experiment 2 (A), and overlapping (B). Furthermore, (C) and (D) show the results for experiment 1 and 2, respectively, in axial slices (z coordinates given between slices). The peak coordinates are given in the Supplementary Material. Maps were thresholded using Monte Carlo correction for multiple comparisons.

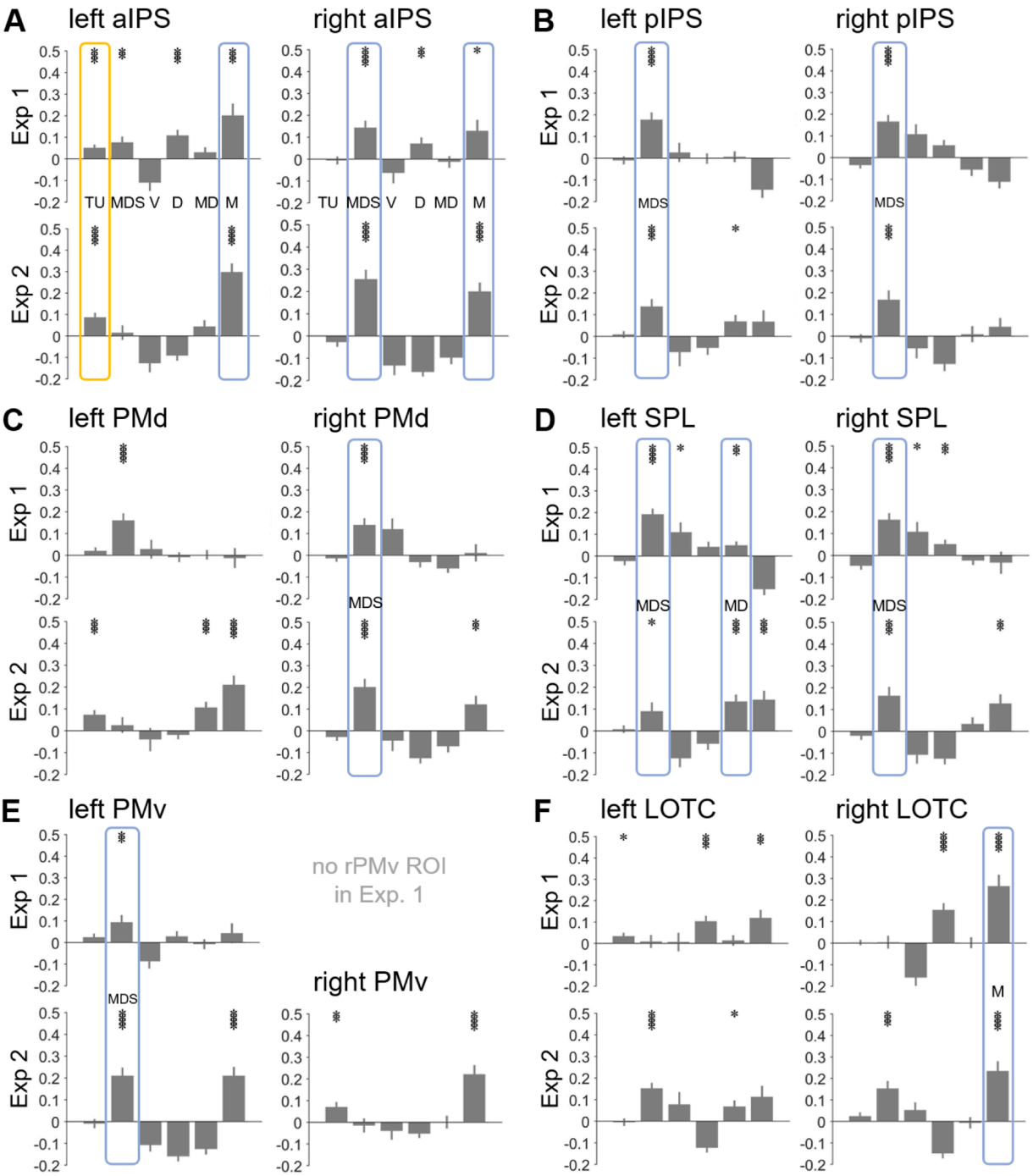


**Fig. 4.** ROI RSA results. Each bar chart shows the correlation of a neural RDM with the six model RDMs. A – F show the results for each bilateral ROI in experiment1 (Exp 1, top row) and experiment 2 (Exp 2, bottom row). For experiment 1, no right PMv ROI was defined as represented by blank space in the figure. Colored frames highlight matching results between experiment 1 and experiment 2. The yellow frame highlights common results regarding the TU model. Blue frames highlight matches for all the other models. Model abbreviations: TU - touching-untouching category model, MDS - inverse multidimensional scaling model, V - low-level visual model, D - duration model, MD - movement direction model, and M - manual model. aIPS = anterior intraparietal sulcus, pIPS = posterior intraparietal sulcus, PMd = dorsal premotor cortex, SPL = superior parietal lobule, PMv = ventral premotor cortex, LOTC = lateral occipitotemporal cortex. * $p <= 0.05$, ** $p <= 0.01$, *** $p <= 0.001$, **** $p <= 0.0001$.

model explained the representational variance. In fact, most variance was explained by the M model. This did not come as a real surprise, since the aIPS is a critical region for object manipulation and tool use, which usually mainly draws on the dominant hand but often also requires the coordinated interplay of the dominant hand with the passive, stabilizing hand.

Furthermore, experiment 1 revealed additional effects in the left aIPS for the MDS model and the D model. Interestingly, the representational profiles in other areas clearly dissociated from that of the left aIPS. Thus, the right aIPS also revealed effects for the M model, but not for the TU model. Instead, we observed a clear effect for the MDS model. Interestingly, common effects were revealed for the MDS model in several ROIs: right aIPS, left PMv, bilateral pIPS, bilateral SPL, and right PMd. Right lateral occipitotemporal cortex (LOTC) revealed an effect for the M model and left SPL for the MD model in both experiments. In contrast, the left LOTC as well as the left PMd revealed no effects that were

consistent for both experiments.

## 4. Discussion

At a fundamental temporospatial level, object-directed actions can be described as touching-untouching (TU) sequences of hands, objects, and the ground. While actions that we observe can be classified in various ways, our study aimed to investigate whether our brain employs this TU structure as a sparse classification code for actions, and whether this code is reflected in explicit action categorization. Findings were consistent across two separate experiments involving either real objects or weakly-informative dough objects: Neuroimaging results revealed that the TU structure of actions is represented in the left anterior intraparietal sulcus, independent of other perceptual and semantic factors that may co-vary with this structure. At the behavioral level, we identified a subtle yet significant correlation between the TU structure and observers' judgments of action similarity with the dough objects. This suggests that the TU structure of actions can predict action categorization, particularly in contexts where object information is limited.

Just as technology is sometimes inspired by nature, the current brain activation study was inspired by robotics. Sequences of TUs are highly informative and useful for robots to recognize and execute object manipulations (Aksoy et al., 2011). Our findings suggest that they provide an organizing principle in human action recognition, too. There could be various reasons for this. For instance, TUs are particularly salient and thus easily recognizable incidents, since touching is always accompanied by deceleration and untouching announces acceleration of our movements. Accordingly, TUs are also particularly informative. At the same time, the recognition of TU sequences does not require the ability to distinguish between hands, objects, and the ground or to identify different objects. It therefore appears to be an ideal starting point for learning categories of action even before critical object expertise is built up (see Hunnius & Bekkering, 2010, for the early development of object knowledge). In this vein, it has been proposed that object-action association may develop earlier than object-word association (Eiteljoerge et al., 2019). Accordingly, preverbal infants might use TU sequences to more efficiently encode and more easily recognize everyday object manipulations they observe.

In our previous studies, we showed that the TUs in an object-manipulation action video are important and reliable anchor points for subjective event boundaries (Pomp et al., 2024, 2021). In addition to this behavioral relevance of TUs, we showed specific neural processing patterns differentiating between touchings and untouchings. Specifically, the difference in brain activity between touchings and untouchings suggested distinct cognitive processing roles: touchings strongly engaged visual regions, likely reflecting bottom-up visual processing, while untouchings recruited broader regions involved in updating expectations, highlighting the brain's response to action transitions and anticipation of what comes next.

The current results expand on these findings and show that TUs are not only meaningful for action segmentation but also serve as meaningful information in neural action category representation. Specifically, the left aIPS represents actions with similar TU structure similarly, which was shown at the whole-brain level and was unique in the examined action observation network. Generally, the aIPS is involved in creating an action plan for reach and grasp actions (see Turella & Lingnau, 2014, for a review), it codes hand and tool actions (Cabrera-Álvarez and Clayton, 2020), represents skills and conceptual action knowledge (Johnson-Frey, 2004), stores abstract representations of specific object-directed actions (Chen et al., 2018), and is involved in physical scene understanding (Fischer et al., 2016). The left aIPS has been described as one of the brain regions that are generally capable of discriminating actions of distinct categories and specifically of object manipulations (Wurm et al., 2017). Recent work by Wurm and Erigüç (2024) found the aIPL to encode abstract representations of cause-effect structures that capture the effect that is induced by an effector-target interaction in both observed human actions and abstract animations. Thus, our current finding that the TU structure of actions is represented in aIPS fits well the current state of research. Given the numerous similar characterizations of aIPS in terms of its role in hand-object interaction, TU sequences might even be a simple explanatory abstraction of this functionality.

Importantly, the present findings do not suggest that the left aIPS is specifically involved in the representation of touchings and untouchings per se. Rather, this area reflects the sequential pattern of touchings and untouchings that characterize different action categories. This is also supported by previous studies in which the contrast between touchings and non-TUs showed no engagement of the aIPS but significant activation in regions associated with visual processing, specifically the cuneus and lingual gyrus (Pomp et al., 2024, 2021). Engagement of these visual areas in response to touchings highlights their role in providing perceptual anchors for action segmentation in visual contexts. In accordance with this, the conjunction of untouchings contrasted to non-TUs in real-object actions and dough-object actions neither showed aIPS activation (Pomp et al., 2024).

In the same vein, our findings do not reflect differences in hand configuration. Previous studies showed the importance of grip similarity for ratings of manipulable objects (Hussain et al., 2024) as well as the influence of similarity in magnitude of arm movement and the hand configuration during use (Watson and Buxbaum, 2014). At first glance, one might assume that TU categories correspond to different grip types; however, this potential confound was ruled out, as TU categories in the present work did not systematically covary with specific hand postures.

Furthermore, to integrate our results into existing research, it is crucial to inspect the relationship between TU sequences and other formal descriptions of actions. A sequence of TUs describes an action on a very basic level. It captures the temporal dynamics of object manipulation—specifically, when and where contact happens or is lost. TU sequences might be a basic, foundational code that can be used to represent and recognize actions from early on and throughout life. Though TU sequences do not differentiate between agents or objects and can code contact states between item surfaces in object interactions without any agent involved (e.g., a branch that breaks off the tree in a storm and separates two apples lying on the ground is coded as destroy). Therefore, these sequences do not directly represent an abstract functional goal itself, but they form the underlying frame-work through which abstract functional goals like "rearranging" or "destroying" are realized and can correspond to them.

At the behavioral level, observers' similarity judgments of actions on minimally informative dough objects significantly related to the TU structure. While real objects might prompt action classification based on categorical associations, the reduced object information in the dough condition highlighted the TU structure, making it a key, informative basis for similarity judgments. This finding aligns with our previous study (Pomp et al., 2024), which showed that when object information is weak, subjective event boundaries fall systematically closer to touchings, emphasizing their perceptual relevance. Interestingly, comparing planning actions with unfamiliar in contrast to familiar objects, Van Elk et al. (2012) suggested a stronger goal-representation for familiar objects and a stronger motor imagery for unfamiliar objects, which points in the same direction as the current results. Also, violation paradigms where the object did not match the grip or goal of an action revealed independent neural temporal dynamics of the integration of motor acts and goal-related information during action observation (Decroix et al., 2020). In a similar vein, Bach et al. (2009) describe at least two partially distinct subprocesses involved in deriving both the motor act and the function of objects, which integrate during recognition into a unified action representation. To judge the similarity of the dough-object actions in the current study, participants may have preferably used the motor act information which is closer to TU structures than functional goals implied by familiar objects.

Remarkably, the MDS models derived from the two behavioral

similarity ratings accounted for a significant proportion of the variance in neural representational profiles across several regions in the action observation network of both experiments: bilateral posterior IPS, bilateral SPL, right aIPS, right PMd, and left PMv. This alignment suggests that behavioral similarity judgments resonate with the neural encoding of action structure, even in regions not directly tied to touching-untouching representation.

### 4.1. Limitations

In the multi-arrangement task, participants arranged triplets of single frame images that were supposed to remind them of the well-known action videos. While it is not particularly uncommon to use static images in action observation paradigms (cf. Caspers et al., 2010), future research might profit from presenting the entire action video in multi-arrangement tasks. Especially when participants are not as familiar with the actions as in the present work.

Furthermore, we applied a range of perceptual and semantic control models to confirm the specific, independent representation of TU structure. Future research could examine whether the TU model primarily reflects TU-based SEC categorization or if it aligns with additional perceptual and semantic principles beyond those assessed in this study. It could also profit from using more sophisticated visual control models than pixelwise similarity. Furthermore, it remains an open question whether the TU structure of an action plays a role in the recognition process or merely emerges as a byproduct of action categorization. This leads to the possibility that observers recognize actions by category, and because these categories covary with TU structure, we observe corresponding effects in both brain and behavior. Ultimately, these findings establish TU structure as a key element in how actions are perceived and recognized, opening new avenues for exploring the cognitive and neural bases of action segmentation.

### 4.2. Conclusion

Describing actions as an abstract sequence of relational changes in the form of touchings and untouchings of objects, hands, and the ground can effectively guide automated systems like robots in recognizing and responding to human actions without relying on object identification. The current study examined whether the neural processing of actions and their behavioral classification relies on the action categorization derived from their TU sequence. Using fMRI-based multivoxel pattern analysis, we identified neural representations of actions in the anterior intraparietal sulcus to be particularly associated to this TU structure. While models of one- or two-handedness and behavioral similarity ratings also explained representational activity in the action observation network, only the TU category model selectively correlated with the representational profile of the aIPS. These findings suggest that sequences of touchings and untouchings serve as a key organizing principle in inferior parietal action representation, extending prior research on meaningful anchor points for subjective event boundaries.

## CRediT authorship contribution statement

**Jennifer Pomp:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Moritz F. Wurm:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Methodology, Formal analysis, Conceptualization. **Rosari N. Selvan:** Writing – review & editing, Validation, Formal analysis. **Florentin Wörgötter:** Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization. **Ricarda I. Schubotz:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

*Acknowledgements*

The authors would especially like to thank Jasper J. F. van den Bosch for his great help in implementing the multi-arrangement task on the meadows research platform. Furthermore, we would like to thank Minija Tamosiunaite and Tomas Kulvicius for their help with the methods setup. Finally, we thank Monika Mertens, Theresa Eckes, Katharina Thiel, Alina Eisele, Mina-Lilly Shibata, Simon Reich, Yuyi Xu, and Annika Garlichs for their assistance during stimulus material creation or data collection.

*Funding*

This work was supported by the German Research Foundation (DFG) [grant numbers SCHU 1439/8–1, WO 388/13–1]. The DFG had no involvement in study design, data collection, analysis and interpretation of the data, writing of the report, or decision to submit for publication.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2025.121113.

## Data availability

The multi-arrangement task stimuli, iMDS data as well as the ROI data have been deposited in an OSF repository (DOI 10.17605/OSF.IO/9V3CQ). The video stimulus material is available via the Action Video Corpus Münster (AVICOM, https://www.uni-muenster.de/IVV5PSY/AvicomSrv/). The raw data of the behavioral multi-arrangement task and the fMRI analyses is available upon reasonable request.

## References

Aguirre, G.K., Mattar, M.G., Magis-Weinberg, L., 2011. De Bruijn cycles for neural decoding. Neuroimage 56, 1293–1300. https://doi.org/10.1016/j.neuroimage.2011.02.005.

Aksoy, E.E., Abramov, A., Dörr, J., Ning, K., Dellen, B., Wörgötter, F., 2011. Learning the semantics of object-action relations by observation. Int. J. Rob. Res. 30, 1229–1249. https://doi.org/10.1177/0278364911410459.

Bach, P., Gunter, T.C., Knoblich, G., Prinz, W., Friederici, A.D., 2009. N400-like negativities in action perception reflect the activation of two components of an action representation. Soc. Neurosci. 4, 212–232. https://doi.org/10.1080/17470910802362546.

Belsley, D.A., Kuh, E., Welsch, R.E., 1980. Regression diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley & Sons, New York.

Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. Ann. Stat. 29, 1165–1188. https://www.jstor.org/stable/2674075.

Buxbaum, L.J., Randerath, J., 2018. Limb apraxia and the left parietal lobe. Handbook of Clinical Neurology, 1st ed. Elsevier B.V. https://doi.org/10.1016/B978-0-444-63622-5.00017-6

Cabrera-Álvarez, M.J., Clayton, N.S., 2020. Neural processes underlying tool use in humans, macaques, and corvids. Front. Psychol. 11, 1–11. https://doi.org/10.3389/fpsyg.2020.560669.

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, 1148–1167. https://doi.org/10.1016/j.neuroimage.2009.12.112.

Chan, A.W.Y., Baker, C.I., 2015. Seeing is not feeling: posterior parietal but not somatosensory cortex engagement during touch observation. J. Neurosci. 35, 1468–1480. https://doi.org/10.1523/JNEUROSCI.3621-14.2015.

Chen, Q., Garcea, F.E., Jacobs, R.A., Mahon, B.Z., 2018. Abstract representations of object-directed action in the left inferior parietal lobule. Cereb. Cortex 28, 2162–2174. https://doi.org/10.1093/cercor/bhx120.

de Vries, I.E.J., Wurm, M.F., 2023. Predictive neural representations of naturalistic dynamic input. Nat. Commun. 14. https://doi.org/10.1038/s41467-023-39355-y.

Decroix, J., Roger, C., Kalénine, S., 2020. Neural dynamics of grip and goal integration during the processing of others' actions with objects: an ERP study. Sci. Rep. 10, 1–11. https://doi.org/10.1038/s41598-020-61963-7.

Eiteljoerge, S.F.V., Adam, M., Elsner, B., Mani, N., 2019. Word-object and action-object association learning across early development. PLoS One 14, 1–22. https://doi.org/10.1371/journal.pone.0220317.

Fischer, J., Mikhael, J.G., Tenenbaum, J.B., Kanwisher, N., 2016. Functional neuroanatomy of intuitive physical inference. Proc. Natl. Acad. Sci. U. S. A. 113, E5072–E5081. https://doi.org/10.1073/pnas.1610344113.

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.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.-P, Frith, C.D., Frackowiak, R.S.J., 1994. Statistical parametric maps in functional imaging: A general linear approach. Hum. Brain Mapp. 2, 189–210. https://doi.org/10.1002/hbm.460020402.

Halchenko, Y.O., Hanke, M., 2012. Open is not enough. Let's take the next step: an integrated, community-driven computing platform for neuroscience. Front. Neuroinform. 6, 1–4. https://doi.org/10.3389/fninf.2012.00022.

Han, X., Zhang, Z., Ding, N., Gu, Y., Liu, X., Huo, Y., Qiu, J., Yao, Y., Zhang, A., Zhang, L., Han, W., Huang, M., Jin, Q., Lan, Y., Liu, Y., Liu, Z., Lu, Z., Qiu, X., Song, R., Tang, J., Wen, J.R., Yuan, J., Zhao, W.X., Zhu, J., 2021. Pre-trained models: past, present and future. AI Open 2, 225–250. https://doi.org/10.1016/j.aiopen.2021.08.002.

Herath, S., Harandi, M., Porikli, F., 2017. Going deeper into action recognition: A survey. Image Vis. Comput. 60, 4–21. https://doi.org/10.1016/j.imavis.2017.01.010.

Hunnius, S., Bekkering, H., 2010. The early development of object knowledge: a study of infants' Visual anticipations during action observation. Dev. Psychol. 46, 446–454. https://doi.org/10.1037/a0016543.

Hussain, A., Walbrin, J., Tochadse, M., Almeida, J., 2024. Primary manipulation knowledge of objects is associated with the functional coupling of pMTG and aIPS. Neuropsychologia 205, 109034. https://doi.org/10.1016/j.neuropsychologia.2024.109034.

JASP Team, 2024. JASP (Version 0.18.3) [Computer software].

Jiao, L., Zhang, R., Liu, F., Yang, S., Hou, B., Li, L., Tang, X., 2022. New generation deep learning for video object detection: A survey. IEEE Trans. Neural Networks Learn. Syst. 33, 3195–3215. https://doi.org/10.1109/TNNLS.2021.3053249.

Johnson-Frey, S.H., 2004. The neural bases of complex tool use in humans. Trends Cogn. Sci. 8, 71–78. https://doi.org/10.1016/j.tics.2003.12.002.

Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. Proc. Natl. Acad. Sci. U. S. A. 103, 3863–3868. https://doi.org/10.1073/pnas.0600244103.

Kriegeskorte, N., Mur, M., 2012. Inverse MDS: inferring dissimilarity structure from multiple item arrangements. Front. Psychol. 3, 1–13. https://doi.org/10.3389/fpsyg.2012.00245.

Kriegeskorte, N., Mur, M., Bandettini, P., 2008. Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2, 1–28. https://doi.org/10.3389/neuro.06.004.2008.

Liang, Q.A., Wan, T.F., 2020. Difficulty Within Deep Learning Object-Recognition Due to Object Variance. Springer International Publishing, pp. 278–289. https://doi.org/10.1007/978-3-030-63830-6_24.

Murata, A., Wen, W., Asama, H., 2016. The body and objects represented in the ventral stream of the parieto-premotor network. Neurosci. Res. 104, 4–15. https://doi.org/10.1016/j.neures.2015.10.010.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., Kriegeskorte, N., 2014. A toolbox for representational similarity analysis. PLoS Comput. Biol. 10. https://doi.org/10.1371/journal.pcbi.1003553.

Oldfield, R.C., 1971. The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9, 97–113. https://doi.org/10.1016/0028-3932(71)90067-4.

Oosterhof, N.N., Connolly, A.C., Haxby, J.V., 2016. CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in matlab/GNU octave. Front. Neuroinform. 10, 1–27. https://doi.org/10.3389/fninf.2016.00027.

Peelen, M.V., Caramazza, A., 2012. Conceptual object representations in human anterior temporal cortex. J. Neurosci. 32, 15728–15736. https://doi.org/10.1523/JNEUROSCI.1953-12.2012.

Pomp, J., Garlichs, A., Kulvicius, T., Tamosiunaite, M., Wurm, M.F., Zahedi, A., Wörgötter, F., Schubotz, R.I., 2024. Action segmentation in the brain: the role of object–Action associations. J. Cogn. Neurosci. 36, 1784–1806. https://doi.org/10.1162/jocn_a_02210.

Pomp, J., Heins, N., Trempler, I., Kulvicius, T., Tamosiunaite, M., Mecklenbrauck, F., Wurm, M.F., Wörgötter, F., Schubotz, R.I., 2021. Touching events predict human action segmentation in brain and behavior. Neuroimage 243, 118534. https://doi.org/10.1016/j.neuroimage.2021.118534.

Reynaud, E., Lesourd, M., Navarro, J., Osiurak, F., 2016. On the neurocognitive origins of human tool use: A critical review of neuroimaging data. Neurosci. Biobehav. Rev. 64, 421–437. https://doi.org/10.1016/j.neubiorev.2016.03.009.

Turella, L., Lingnau, A., 2014. Neural correlates of grasping. Front. Hum. Neurosci. 8, 1–8. https://doi.org/10.3389/fnhum.2014.00686.

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. https://doi.org/10.1016/j.neuropsychologia.2019.02.006.

Van Elk, M., Viswanathan, S., Van Schie, H.T., Bekkering, H., Grafton, S.T., 2012. Pouring or chilling a bottle of wine: an fMRI study on the prospective planning of object-directed actions. Exp. Brain Res. 218, 189–200. https://doi.org/10.1007/s00221-012-3016-9.

Vingerhoets, G., 2014. Contribution of the posterior parietal cortex in reaching, grasping, and using objects and tools. Front. Psychol. 5, 1–17. https://doi.org/10.3389/fpsyg.2014.00151.

Watson, C.E., Buxbaum, L.J., 2014. Uncovering the architecture of action semantics. J. Exp. Psychol. Hum. Percept. Perform. 40, 1832–1848. https://doi.org/10.1037/a0037449.

Wörgötter, F., Aksoy, E.E., Krüger, N., Piater, J., Ude, A., Tamosiunaite, M., 2013. A simple ontology of manipulation actions based on hand-object relations. IEEE Trans. Auton. Ment. Dev. 5, 117–134. https://doi.org/10.1109/TAMD.2012.2232291.

Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series revisited—Again. Neuroimage 2, 173–181. https://doi.org/10.1006/nimg.1995.1023.

Wurm, M.F., Caramazza, A., Lingnau, A., 2017. Action categories in lateral occipitotemporal cortex are organized along sociality and transitivity. J. Neurosci. 37, 562–575. https://doi.org/10.1523/JNEUROSCI.1717-16.2016.

Wurm, M.F., Erigüç, D.Y., 2024. Decoding the physics of observed actions in the human brain. BioRxiv 2023.10.04.560860. https://doi.org/10.1101/2023.10.04.560860.

Wurm, M.F., Lingnau, A., 2015. Decoding actions at different levels of abstraction. J. Neurosci. 35, 7727–7735. https://doi.org/10.1523/JNEUROSCI.0188-15.2015.