JNeurosci
THE JOURNAL OF NEUROSCIENCE

*Research Articles: Behavioral/Cognitive*

# The representation of observed actions at the subordinate, basic and superordinate level

https://doi.org/10.1523/JNEUROSCI.0700-22.2023

**Cite as:** J. Neurosci 2023; 10.1523/JNEUROSCI.0700-22.2023

Received: 8 April 2022
Revised: 8 August 2023
Accepted: 6 September 2023

# The representation of observed actions at the subordinate, basic and superordinate level

Tonghe Zhuang, Zuzanna Kabulska & Angelika Lingnau*

Tonghe Zhuang
tonghe.zhuang@ur.de
ORCID: 0000-0002-7876-4962

Zuzanna Kabulska
zuzanna.kabulska@ur.de
ORCID: 0000-0002-6756-691X

*Corresponding author:
Prof. Dr. Angelika Lingnau
angelika.lingnau@ur.de
ORCID: 0000-0001-8620-3009

address:
University of Regensburg
Faculty of Human Sciences, Institute of Psychology, Chair of Cognitive Neuroscience
Universitätsstrasse 31
93053 Regensburg, Germany

**Abstract**

Actions can be planned and recognized at different hierarchical levels, ranging from very specific (e.g., to swim breaststroke) to very broad (e.g., locomotion). Understanding the corresponding neural representation is an important prerequisite to reveal how our brain flexibly assigns meaning to the world around us. To address this question, we conducted an event-related fMRI study in male and female human participants in which we examined distinct representations of observed actions at the subordinate, basic and superordinate level. Utilizing multiple regression representational similarity analysis in predefined regions of interest, we found that the three different taxonomic levels were best captured by patterns of activations in bilateral LOTC, showing the highest similarity with the basic level model. A whole-brain multiple regression RSA revealed that information unique to the basic level was captured by patterns of activation in dorsal and ventral portions of the LOTC and in parietal regions. By contrast, the unique information for the subordinate level was limited to bilateral occipitotemporal cortex, while no single cluster was obtained that captured unique information for the superordinate level. The behaviorally established action space was best captured by patterns of activation in the LOTC and superior parietal cortex, and the corresponding neural patterns of activation showed the highest similarity with patterns of activation corresponding to the basic level model. Together, our results suggest that occipitotemporal cortex shows a preference for the basic level model, with flexible access across the subordinate and the basic level.

**Keywords:** action categorization; action observation; action recognition.

**Significance statement**

The human brain captures information at varying levels of abstraction. It is debated which brain regions host representations across different hierarchical levels, with some studies emphasizing parietal and premotor regions, while other studies highlight the role of the lateral occipitotemporal cortex. To shed light on this debate, here we examined the representation of observed actions at the three taxonomic levels suggested by Rosch et al. (1976). Our results highlight the role of the LOTC, which hosts a shared representation across the subordinate and the basic level, with the highest similarity with the basic level model. These results shed new light on the hierarchical organization of observed actions and provide insights into the neural basis underlying the basic level advantage.

## Introduction

Depending on the circumstances, different aspects of an action become relevant. As an example, we might be interested in the type of punch when watching a boxing match, while we might be more concerned with the broader distinction between attacking and greeting when approaching a stranger at night. How the brain adapts its representational states to achieve this flexibility is a key question in Cognitive Neuroscience.

The hierarchical organization of objects has been studied for decades (e.g. Gauthier et al., 1997; Mack et al., 2008; Carlson et al., 2013; Iordan et al., 2015). Rosch et al., (1976) argued that objects can be organized into the superordinate (e.g., furniture), basic (e.g., chair) and subordinate level (e.g., kitchen chair), depending on the degree of abstraction, and that the basic level plays a central role in categorization, e.g. in terms of the number and types of features used to describe an object, and in terms of the speed of processing (see also Grill-Spector and Kanwisher, 2005; Mack et al., 2008; Macé et al., 2009). Moreover, different taxonomic levels of objects have been shown to be dissociated at the neural level (Kriegeskorte et al., 2008; Iordan et al., 2015; Dehaqani et al., 2016), and it has been proposed that the ventral temporal cortex (VTC) has flexible access to these different levels (Grill-Spector and Weiner, 2014).

Likewise, the planning and control of actions is assumed to be organized hierarchically (Gallivan et al., 2013; Kadmon Harpaz et al., 2014; Krasovsky et al., 2014; Ariani et al., 2015; Gallivan and Culham, 2015; Turella et al., 2020). Similar hierarchies have been proposed to underlie the organization of observed actions. Several authors distinguished between the *How*, *What* and *Why* level (e.g. Vallacher and Wegner, 1985; Wegner and Vallacher, 1986; Spunt et al., 2016). Hamilton and Grafton (2006, 2008) distinguished between the *goal* level (corresponding to the purpose/ outcome of an action), the *muscle level* and the *kinematic level*, while Wurm and Lingnau (2015) distinguished between different levels of abstraction (e.g. opening versus closing a bottle).

It is assumed that areas involved in action recognition should show invariance to the way the actions are performed (e.g. Wurm and Lingnau, 2015; Hamilton and Grafton, 2006, 2008; Oosterhof et al., 2010, 2012). Several studies have highlighted the role of parietal and premotor regions for action representations at the *Goal* level that generalize across the *muscle or kinematic* level (Hamilton and Grafton, 2006, 2008; Majdandić et al., 2009; see also Lanzilotto et al., 2020; Aflalo et al., 2020). Wurm and Lingnau (2015) revealed representations of observed actions at a concrete level (specific for the object and kinematics) in the LOTC, IPL and ventral premotor cortex (PMv), whereas representations at an abstract level (generalizing across object and kinematics) were restricted to the IPL and LOTC (see also Wurm et al., 2016). In sum, previous studies successfully distinguished between observed actions at varying hierarchical levels, with some studies highlighting the role of parietal and premotor regions, whereas other studies emphasize the role of the LOTC. However, to the best of our knowledge, no previous neuroimaging study directly compared the three taxonomic levels proposed by Rosch et al. (1976). The current study aims to fill this gap.

Zhuang and Lingnau (2021) examined the characteristics of observed actions at the three taxonomic levels. Actions at the three levels differed with respect to the number and type of features participants used to describe them, and in their ratings of abstraction. Moreover, participants verified the action category faster at the basic and subordinate level in comparison to the superordinate level. Together, these results suggest that the basic level holds the maximized information, consistent with the basic level advantage reported for objects (Rosch et al., 1976). Given these behavioral results, here we aimed to determine which brain regions (a) represent observed actions at the three taxonomic levels, and (b) which brain regions host a joint representation across these levels.

## Methods

### Overall rationale and hypotheses

To reveal which brain areas represent actions at the three taxonomic levels, we separated twelve daily actions into three action categories at the superordinate level (see also Zhuang and Lingnau, 2021). Each superordinate action category consisted of two types of actions at the basic level, and each basic level action encompassed two actions at the subordinate level (Figure 1). To verify this hierarchy, we used a multi-arrangement experiment (Kriegeskorte and Mur, 2012) combined with inverse multidimensional scaling (MDS) and hierarchical cluster analysis. Next, to determine which brain areas represent observed actions at the three different hierarchical levels, we conducted an fMRI experiment and carried out ROI-based and whole-brain searchlight-based representational similarity analysis (RSA; Kriegeskorte et al., 2008). Specifically, we examined the representation of observed actions at the subordinate, basic and superordinate level, and the representation of the behavioral similarity structure resulting from the multi-arrangement experiment.

We expected that the subordinate level model is represented by patterns of activations in early visual areas, the LOTC and possibly the IPL and the PMv (see also Wurm and Lingnau, 2015). The basic level model was expected to be represented in the LOTC and the IPL, but not in the PMv, whereas the superordinate level was expected to be represented in anterior portions of the LOTC (Wurm and Lingnau, 2015). The behavioral model was expected to be captured by neural patterns of activation in the LOTC and possibly the IPL (Tucciarelli et al., 2019; Tarhan et al., 2021).

### Stimulus selection and validation

Stimuli consisted of static images of twelve different actions (600 x 480 pixels, 14.36 x 11.07 degree of visual angle; six exemplars each; see Figure 1 for an overview of stimulus exemplars and corresponding action words). The twelve actions were chosen on the basis of a series of rating and behavioral studies (Zhuang and Lingnau, 2021) that we briefly summarize here. First, we selected action verbs corresponding to the basic level from Levin (1993). Using these action verbs, we carried out a semantic similarity rating, followed by hierarchical cluster analysis. Based on the resulting clusters, we selected a subset of basic level actions, excluding actions that might be hard to portray as a picture (e.g. to learn, to memorize). To select labels for the superordinate level, a new set of participants was provided with the basic level labels of actions belonging to a given cluster revealed by the hierarchical cluster analysis. To select actions belonging to the subordinate level, participants were provided with different action verbs corresponding to the basic level and were asked to generate action verbs corresponding to the subordinate level. Next, another group of participants was asked to rate (a) the relationship between actions at the subordinate and the superordinate level (e.g. between 'swim front crawl' and 'locomotion', or between 'swim front crawl' 'and ingestion') and (b) the degree of abstraction and complexity of each action at the subordinate level. Actions were only included in the final set if they were consistently rated to belong to a given superordinate category, and not to other superordinate categories.

We selected the six different exemplars for each of the twelve actions based on the following criteria: young adult agents of both genders, with an equal representation of three males and three females per action. In addition, we selected three distinct orientations for each agent, including two profile views (facing left and right, respectively) and one frontal view. Note that for the action 'doing the dishes', we replaced frontal views by another profile view exemplar due to the lack of suitable images depicting this action in a frontal view.

Since the rating studies were based on written words, we first wanted to verify how human participants categorize these actions when presented as static images. To this aim, we carried out a multi-arrangement experiment (Kriegeskorte et al., 2008) as implemented in the online platform MEADOWS (https://meadows-research.com) in a group of N=18 participants (12 female; mean age: 27 years; range: 22-31 years) that did not take part in the fMRI experiment. Participants were instructed to judge the degree of similarity between the twelve actions depicted in these static

images, to arrange them accordingly (i.e., the more similar in meaning, the closer they should be positioned on the screen), and to press a button when they were satisfied with the arrangement of the stimuli. In the first trial, all stimuli appeared on the screen (stimulus size: 48 x 38 pixel). In all subsequent trials, an adaptive algorithm chose a subset of all stimuli in order to provide the optimal evidence for pairwise dissimilarity estimates (see Kriegeskorte and Mur, 2012, for details). The experiment continued until the adaptive algorithm reached the required evidence level for pairwise dissimilarities. The full stimulus set contained 72 static images (12 actions x 6 exemplars). Each participant was provided with 12 different actions. Action exemplars were counterbalanced across participants.

Results of the multi-arrangement experiment were collapsed across participants and averaged across image exemplars. We visualized the results by creating a 12 x 12 representational dissimilarity matrix (RDM) where each cell contains a value corresponding to the Euclidean distance between two actions (Figure 2A). Figure 2B shows a 2-dimensional (2D) arrangement derived from multidimensional scaling (metric stress), averaged across participants. Inverse MDS revealed three larger clusters corresponding to the three superordinate action categories. Subsequently, to reveal the corresponding hierarchical structure, we carried out average-linkage hierarchical cluster analysis (using the Matlab function *linkage*) on the results obtained from the multi-arrangement experiment. The results are shown in Figure 2C. This analysis confirmed the hierarchical structure of the selected actions, with three action categories corresponding to the superordinate level (locomotion, ingestion and cleaning), six action categories at the basic level (to swim, to ride, to eat, to drink, to clean the body and to do housework), and twelve actions at the subordinate level (to ride a motorbike, to ride a bike, to swim front crawl, to swim backstroke, to drink water, to drink beer, to eat cake, to eat an apple, to clean windows, to do the dishes, to brush teeth, and to clean the face).

<< Figure 1 >>

<< Figure 2 >>

**Participants**

A group of N=23 participants (17 female; mean age: 26 years; range: 21- 39 years) took part in the fMRI experiment. All participants except two authors of the paper (T.Z. and Z.K.) were naive to the purpose of the study. All participants gave written informed consent before joining the experiment and received either monetary compensation or course credits at the University of Regensburg. All participants had normal or corrected-to-normal vision and reported to have no psychiatric or neurological disorders.

**fMRI experimental design and task**

We used a rapid event-related fMRI design (see Figure 3), programmed in ASF (Schwarzbach, 2011), adopting the design used by Tucciarelli et al., (2019). During each trial, participants were provided with a static image of an action and a central fixation cross superimposed on the image (1 s), followed by a blank screen and a central fixation cross (3 s). Participants were instructed to observe the action while keeping their eyes at fixation. During occasional catch trials (11% of all trials), participants were presented with a phrase depicting an action (e.g. 'to swim?', 1 s) followed by 3 s fixation. During catch trials, participants had to perform a category verification task. Specifically, they were instructed to indicate by button press with the index or middle finger of the right hand whether or not the action shown in the previous trial corresponded to the phrase shown during the catch trial. To make sure that participants were not biased towards answering questions at one of the three different levels, phrases presented during catch trials had an equal probability to address the superordinate (e.g. 'locomotion?'), basic (e.g. 'to swim?') or subordinate ('to swim breaststroke?') level. That is, there was the same number of catch trials for each of the three

taxonomic levels (four catch trials per taxonomic level in each run). Additionally, to improve design efficiency, we included null events (22.2% of all trials) that consisted of 4 s fixation and were presented pseudo-randomly (no consecutive null events and catch trials).

<< Figure 3 >>

Participants performed six runs, each consisting of 72 experimental trials (66.7%), 12 catch trials (11.1%) and 24 null events (22.2%). Additionally, each run included a 10 s fixation period at the beginning and the end. Each run lasted 7.5 minutes. Halfway throughout the experiment, an anatomical scan was performed with a duration of approximately 5 minutes. The whole experiment lasted approximately 50 minutes. To ensure that participants fully understood and followed the instructions, participants performed a short practice run (consisting of 12 trials) prior to entering the scanner.

**Data acquisition**

The experiment was conducted in the MRI laboratory at the University of Regensburg. Data were collected using a 3T full-body Siemens-Prisma scanner with a 64-channel head coil. A T2*-weighted gradient multiband (MB) echo-planar imaging (EPI) sequence was used for acquiring functional images with 64 slices per volume, using the following parameters: repetition time (TR): 2s, Echo Time (TE): 30 ms, flip angle: 75°, excitation pulse duration = 9 ms; echo spacing = 0.58 ms; bandwidth = 2368 Hz/ pixel; Field of view (FoV): 192*192 $mm^2$, partial Fourier = 6/8; voxel resolution: 2.5 $mm^3$, MB-acceleration: 4. Each functional run consisted of 226 volumes and lasted 7 min and 32 s. Between the third and fourth EPI sequence, we acquired a 5 min T1-weighted Magnetization Prepared Rapid Gradient Echo (MPRAGE) structural sequence (TR: 1910 ms, TE: 3.67 ms, FOV: 256*256 mm, voxel size: 1 $mm^3$, flip angle: 9°).

**fMRI data preprocessing**

We used the FMRIB Software Library (FSL 6.0 https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/) to preprocess the data. The first four volumes were deleted from each functional run to ensure to have reached steady-state magnetization. Functional images were slice time corrected, highpass filtered (with a cut-off of 100 s), corrected for head-motion with 7 degrees of freedom (DOF) and the middle volume as reference, and then co-registered to the individual T1 anatomical image. For univariate analysis, functional data were smoothed with a 5-mm full width half maximum (FWHM) kernel. For multivariate analysis, we used unsmoothed data. Data were aligned into Montreal Neurological Institute (MNI) space.

**Region of interest (ROI) definition**

ROI definition was carried out using a combination of functional data and anatomical masks obtained from the Harvard-Oxford Cortical structural and the Jülich Histological atlas (see Table 1). We focused on key areas of the action observation network described in previous studies (e.g., Grill-Spector and Kanwisher, 2005; Hamilton and Grafton, 2008; Grafton and Hamilton, 2009; Kilner, 2011; Binkofski and Buxbaum, 2013; Hétu et al., 2013; Hoffman et al., 2015; Wurm et al., 2017b). Specifically, we selected the bilateral LOTC, IPL, SPL, dorsal premotor cortex (dPM), and IFG. In addition, to be able to compare results with an area concerned with low-level visual analysis, we identified bilateral V1.

ROI definition consisted of several steps. First, we computed the random-effects (RFX) general linear model (GLM) contrast 'all actions versus baseline' with spatially smoothed data (5mm FWHM). The baseline consisted of all events not explicitly modelled in the GLM. The RFX GLM included seventy-two regressors (12 actions x 6 exemplars), one regressor for the catch trials, and six regressors for head motion. The statistical map resulting from the RFX GLM contrast was corrected for multiple

comparisons using Threshold-Free Cluster Enhancement (TFCE, Smith and Nichols, 2009, $p<0.05$, $|z|>1.96$, two tailed, 5000 permutations) as implemented in CoSMoMVPA (Oosterhof et al., 2016; http://cosmomvpa.org/index.html). Second, we defined anatomical masks from the Harvard-Oxford Cortical structural and the Jülich Histological atlas (threshold: 20%; see Table 1 for details). The mask for the LOTC was created by merging the anatomical masks for the inferior lateral occipitotemporal cortex (LOC) and the occipito-temporal cortex. Third, within each resulting mask, we extracted the peak coordinate resulting from the RFX GLM contrast 'all actions vs baseline'. ROIs were defined as spheres (radius: 10 mm) centered around these peaks.

<< Table 1 >>

**Representational similarity analysis**

We carried out RSA using the CoSMoMVPA Toolbox (Oosterhof et al., 2016) and custom written Matlab functions (available at https://osf.io/b6ea4/). We used the following procedure both for the ROI-based and the whole-brain searchlight approach unless otherwise noted.

**ROI-based RSA (partial correlations)**

To examine the representation of observed actions in predefined ROIs, we used RSA with partial correlations (Kriegeskorte et al., 2008). First, we created a model RDM for each of the three taxonomic levels (Figure 1, bottom panel), corresponding to the hierarchical structure shown in Figure 1 (top panel; see also Zhuang and Lingnau, 2021). Since we used 6 exemplars for each of the 12 actions, each model RDM consisted of a 72 x 72 matrix, where each cell in the matrix corresponds to the dissimilarity between a pair of actions. The subordinate level model consists of twelve clusters along the diagonal, with each cluster comprising six exemplars of the same type of action. The basic level model consists of six clusters along the diagonal, while the superordinate model consists of three clusters along the diagonal.

Second, to account for differences between the different action categories in terms of low-level visual properties, we constructed a control model using the 1$^{st}$ layer of a Deep Neural Network (DNN) that has been trained to classify image classes in the ImageNet dataset (ResNet50, He et al., 2015). We chose this layer because the first layer of a deep neural network is generally assumed to learn to detect edges, colours, texture orientations and other simple shapes in input images (Zeiler and Fergus, 2014; Kriegeskorte, 2015; Mahendran and Vedaldi, 2015). For each of the 72 images used in the current experiment, we determined the activations of each unit in the first convolutional layer of the ResNet50 and converted these values into activation vectors, separately for each image. Next, we constructed the 72 x 72 low-level visual control model by computing the dissimilarity (1-correlation) between these activation vectors for each pair of action images (for similar approaches, see Kriegeskorte et al. 2008, 2012). Additionally, to account for features related to the scene or background in which the action took place, we created a 72 x72 scene control model. We used the same procedure as the one described for the construction of the low-level visual control model. The only difference was that we obtained activations in response to each image from the second-to-last layer of the ResNet50 trained on a large image dataset to distinguish between scene categories (Zhou et al., 2017, 2018).

Third, we constructed neural RDMs within each ROI following previous studies (Bonner and Epstein, 2018; Tucciarelli et al., 2019). To do so, separately for each run and each of the 72 action images (12 actions x 6 exemplars), we extracted the beta estimates for each voxel in a given ROI and converted these beta estimates into t-values, resulting in a vector of t-values for each action image, with the length corresponding to the number of voxels. Next, we averaged the t-values across runs and normalized the t-values by subtracting the mean t-values of each voxel from the t-values of each action image (Diedrichsen and Kriegeskorte, 2017), separately for each participant. Finally, for each

of the 72 x 72 pairwise comparisons of action images, we computed the squared Euclidean distance between the corresponding vectors of t-values, resulting in a 72 x 72 neural RDM.

Finally, to determine which of the ROIs captured the similarity between actions at the three taxonomic levels, we computed partial correlations between neural RDMs and each of the three model RDMs, regressing out the low-level visual control model and the scene control model.

**Searchlight-based RSA (multiple regression)**

To determine whether additional areas not captured by the ROI analysis contained information regarding observed actions at the three taxonomic levels, we carried out two different multiple-regression representational similarity analyses (RSAs). In both types of searchlight-based RSAs, we used a spherical neighborhood of 100 voxels which were nearest to each center voxel. As in the ROI analysis, t values were averaged across runs and normalized by subtracting the mean t-value of each voxel across conditions.

To examine risks of multicollinearity, we determined the Variance Inflation Factor (VIF) for the three taxonomic models depicted in Figure 1 (bottom panel), the low-level visual control model and the scene control model. We obtained small to moderate VIFs (subordinate model: 1.90; basic model: 2.55; superordinate model: 1.78; low-level visual control model: 1.31, scene control model: 1.77), suggesting a low risk of multicollinearity (Mason et al., 2003).

In the first multiple regression RSA, computed separately for each taxonomic level, we included the model for a single taxonomic level (e.g. the subordinate level), the low-level visual control model and the scene control model. This allowed us to obtain beta weights for each taxonomic level while regressing out the low-level visual control model and the scene control model.

In the second multiple regression RSA, we aimed to determine the unique contribution of each taxonomic level. To this aim, we set up a multiple regression that included all three taxonomic level models, the low-level visual control model and the scene control model. This way, we were able to obtain the beta weights for each of the three taxonomic levels while regressing out the contribution of the other two taxonomic levels, the low-level visual control model and the scene control model.

Finally, in order to examine the relationship between neural RDMs and the behavioral dissimilarity matrix obtained from the multi-arrangement task (see Figure 2B), we carried out another searchlight-based multiple regression RSA in which we included (a) the behavioral model, the low-level visual control model and the scene control model, and (b) the behavioral model, each of the taxonomic models separately, the low-level visual control model and the scene control model. VIFs were also computed and the results were small to moderate, justifying the use of multiple regression RSA. In particular, the VIFs for the low-level visual control model, the scene control model and the behavioral model were small (low-level visual control model, scene control model and behavioral action space model: VIF=1.31, 1.85, 1.53, respectively). Additionally, the VIFs for regression models including the behavioral action space model, each of the taxonomic level models, the low visual control model and the scene control model were small to moderate (behavioral action space model, subordinate level model, low-level visual control model and scene control model: VIF= 3.84, 3.18, 1.33, 1.88, respectively; behavioral action space model, basic level model, low-level visual control model and scene control model: VIF=6.10, 5.34, 1.32, 1.88, respectively; behavioral action space model, superordinate level model, low-level visual control model and scene control model: VIF=2.63, 2.24, 1.31, 1.87, respectively).

**Conjunction analysis**

To reveal which brain regions host converging representations across the different taxonomic levels, we conducted a conjunction analysis (Nichols et al., 2005). To this aim, we determined the minimum t-value for each voxel for the whole-brain searchlight maps corresponding to the unique representation of the different taxonomic levels (corresponding to the results shown in Figure 6B).

**Statistics**

For the ROI-based partial correlation analysis, we first used one-sample t-tests to compare the Fisher-transformed partial correlation values against zero. Second, to determine whether the similarity between neural and model RDMs depicted in Figure 4 was modulated by the taxonomic level and ROI, we carried out a two-way repeated measures ANOVA with the factors *taxonomic level* (superordinate, basic, subordinate) and *ROI* (see Table 1), followed by pairwise comparisons. Corrections for multiple comparisons for all factorial combinations (taxonomic level x ROI) was carried out using the false discovery rate (FDR; Benjamini and Hochberg, 1995). Since the assumption of sphericity was violated for the 2-way ANOVA, we report Greenhouse-Geisser-corrected p values (indicated as $p_{GG}$).

For the searchlight-based multiple regression RSA, the beta maps determined separately for each participant were entered into a group statistic by means of a one-sample t-test (one-tailed) against zero. The resulting t maps were corrected for multiple comparisons using threshold free cluster enhancement (TFCE) as implemented in CoSMoMVPA (Oosterhof et al., 2016; number of permutations=5000, corrected $p<0.05$, $z>1.65$, one tailed). Thresholded statistical maps were visualized using BrainNet (https://www.nitrc.org/projects/bnv/).

**Results**

**Behavioral results**

Participants reached a high accuracy (91.1%, standard deviation: 6%) in the category verification task, indicating that they paid attention to the actions depicted in the images.

**ROI-based results**

Figure 4 shows the results from ROI-based RSA for the three taxonomic levels, regressing out low-level visual properties by means of a control model obtained from the first convolutional layer of a Deep Neural Network (ResNet50_conv1; see section ROI-based RSA for details) and the scene control model (based on the last second layer of ResNet50 pretrained with places; see section ROI-based RSA for details). We obtained significant partial correlations between neural RDMs and all three taxonomic level models in most of the examined ROIs, with the exception of the right IFG, where no significant partial correlation was observed between neural RDMs and the subordinate level. Similarities (partial correlation values) between neural RDMs and model RDMs differed between the three taxonomic levels in bilateral LOTC and SPL. In bilateral LOTC, we obtained the highest partial correlations for actions at the basic level in comparison to the other two taxonomic levels. In bilateral SPL, we observed higher partial correlations between the neural RDM and the basic level model in comparison to the superordinate level model.

These observations are supported by the corresponding statistics. In particular, a two-way (taxonomic level x ROI) repeated measures ANOVA revealed a significant main effect of taxonomic level ($F_{(2,44)} = 8.88$, $p_{GG} = .002$, partial $\eta^2 = .29$) and ROI ($F_{(11, 242)} = 32.93$, $p_{GG} < .001$, partial $\eta^2 = .60$), and a significant interaction between taxonomic level and ROI ($F_{(22, 484)} = 3.46$, $p_{GG} < .001$, partial $\eta^2 = .14$). Pairwise comparisons within each ROI (using FDR to correct for multiple comparisons) revealed that neural RDMs in the bilateral LOTC showed the highest similarity with the basic level model in comparison to the other two models (left: basic vs subordinate level, $t_{(1,22)} = 6.11$, $q_{FDR} < .001$; basic vs superordinate level, $t_{(1,22)} = 3.57$, $q_{FDR} = .003$; right: basic vs subordinate level, $t_{(1,22)} =$

5.04, $q_{FDR} < .001$; basic vs superordinate level, $t_{(1,22)} = 4.04$, $q_{FDR} < .001$). Additionally, neural RDMs in bilateral SPL showed a higher partial correlation with the basic level model in comparison to the superordinate level (left: basic vs superordinate level, $t_{(1,22)} = 3.53$, $q_{FDR} = .006$; right: basic vs superordinate level, $t_{(1,22)} = 3.86$, $q_{FDR} = .003$).

<< Figure 4 >>

To illustrate the results obtained in bilateral LOTC in comparison to bilateral early visual cortex (V1), Figure 5 visualizes neural RDMs (left panel) and the corresponding 2D-arrangements obtained from multidimensional scaling analysis (right panel) in bilateral LOTC (top row) and bilateral V1 (bottom row). Using the same approach described for the searchlight-based RSA, we created a 72 X 72 neural RDM in LOTC and V1, and then collapsed the neural RDM across hemispheres and participants. Next, to account for low-level visual features and scene-related properties, we extracted values from the lower triangular part of the neural RDM and regressed out the corresponding lower triangular part of the low-level visual control model and the scene control model. Finally, the resulting residuals were rescaled to values from 0 to 100 (Nili et al., 2014). In the 2D visualization of the MDS (Figure 5B, D), the three superordinate action categories are highlighted in different colors (blue, red, and green), the basic action categories within a superordinate category are shown by open and filled symbols, and different symbols (circles, squares) indicate the different actions at the subordinate level within each basic level category. As can be seen from Figure 5, the neural RDM in bilateral LOTC (Figure 5A) and the corresponding 2-D arrangement resulting from multidimensional scaling (Figure 5B) reveals a broad distinction into actions at the basic level, with open symbols mostly on the right side and filled symbols mostly on the left side, and an additional broad clustering into the three different superordinate categories highlighted in red, green and blue (Figure 5B). By contrast, this pattern is less obvious in early visual cortex (Figure 5C, D), in line with the results of the ROI-based analysis shown in Figure 4.

<< Figure 5 >>

**Whole-brain searchlight results**

Figure 6A reveals areas that capture the similarity between actions at the three different taxonomic levels when low-level visual features and scene-related properties are regressed out via a low-level visual control model and a scene control model (see Section *ROI-based RSA* for details). This analysis reveals a wide set of regions for the subordinate and basic level, with peaks in the right fusiform cortex (Table 2). By contrast, the representation of observed actions at the superordinate level is restricted to a more circumscribed region in occipitotemporal cortex extending into the anterior IPL, with a peak in right occipital pole (Table 2).

Figure 6B shows which areas capture the similarity structure of observed actions that is unique to each of the taxonomic levels (accounting for the other two taxonomic models as well as for the low-level visual features and the scene control model). This analysis reveals that the dissimilarity structure for actions that is unique to the subordinate level is restricted to bilateral occipito-temporal cortex, with a peak in right fusiform cortex, while information that is unique to the basic level is associated with bilateral occipito-temporal cortex, IPL and SPL, with a peak in the right occipital fusiform cortex (see Table 2). By contrast, we obtained no region that represented information that is unique to the superordinate level. We will return to this observation in the discussion.

<< Figure 6 >>

<< Table 2 >>

To identify regions in which actions are jointly represented at the subordinate and basic level, we computed a conjunction analysis on the basis of the statistical maps corresponding to the unique representation of the subordinate and basic level shown in Figure 6B (see Methods section for details). The results are shown in Figure 7. The convergence of observed actions at the subordinate and basic level was located in bilateral OTC.

<< Figure 7 >>

**Relationship between neural RDMs and behavioral dissimilarity structure**

Whereas the previous analyses focused on the representation of the three taxonomic levels, we next aimed to determine which brain areas capture the dissimilarity structure between observed actions resulting from the multi-arrangement task, and to which degree these representations are accounted for by the model RDMs corresponding to the three taxonomic levels. To address these questions, we carried out two additional multiple regression searchlight analyses.

First, we ran a multiple-regression RSA between neural RDMs and the behavioral dissimilarity structure shown in Figure 2A, regressing out the low-level visual control model and the scene control model. As shown in Figure 8A, this analysis revealed large clusters in bilateral occipitotemporal cortex, IPL, and SPL that were associated with the behavioral action space model.

Second, separately for each of the taxonomic levels, we carried out another multiple-regression RSA in which we included the behavioral action space model, the low-level visual control model, the scene control model and the respective taxonomic level (e.g., the subordinate level). A map of the beta weights corresponding to the behavioral action space model, after regressing out the low-level visual control model, the scene control model and each of the taxonomic level models are shown in Figure 8B-D. As can be seen, after regressing out the subordinate level, the low-level visual control model and scene control model, the behavioral action space model is associated with neural patterns in bilateral occipitotemporal cortex (Figure 8B). After regressing out the superordinate model (Figure 8D), the low-level visual control model and the scene control model, large regions of bilateral occipitotemporal cortex, IPL and SPL remained significantly associated with the behavioral action space model. By contrast, after regressing out the basic level model, the low-level visual control model and the scene control model, a small cluster in the temporal occipital fusiform cortex (bilaterally) and left superior LOC remained that captured the behavioral action space model (Figure 8C). Whereas these latter results need to be interpreted with caution, they are in line with the view that the cortical representations of the behavioral action space model best reflects information at the basic level (see also Figure 4 and 5). We will return to this observation in the discussion.

<< Figure 8 >>

**Control analysis: Whole-brain searchlight analysis for the low-level visual control model**

To examine which brain regions show a significant correlation between neural activation patterns and the control model capturing low-level visual features, we carried out an additional whole-brain searchlight RSA with the low-level visual control model (*ResNet50_conv1*). The results are shown in Figure 8-1. As expected, this analysis revealed clusters in early visual cortex (V1 and V2) bilaterally and additionally in a small portion of the superior LOC, temporal occipital fusiform cortex and inferior LOC (right hemisphere). In other words, this analysis suggests that whereas some of the regions capturing low-level visual features overlap with the regions capturing the three taxonomic levels and the behavioral action space model, the majority of voxels revealed by the taxonomic level models and the behavioral action space model do not overlap with the regions capturing low-level visual features.

## Discussion

In the current study we aimed to examine the neural representation of observed actions at different taxonomic levels. Utilizing a hierarchical stimulus set and representational similarity analysis, we observed the highest similarity between neural patterns and the basic level model in bilateral LOTC. A searchlight RSA revealed that the similarity between observed actions at the subordinate and the basic level was captured in a widespread set of occipitotemporal, parietal and frontal areas, whereas neural patterns corresponding to the superordinate level model were obtained in a more circumscribed region in bilateral occipitotemporal cortex. Unique information corresponding to the basic level was captured by patterns of activation in lateral and ventral occipitotemporal cortex and bilateral SPL, while unique information corresponding to the subordinate level was restricted to bilateral occipitotemporal cortex. For the superordinate model we did not obtain any cluster that captured unique information. Additionally, we found that bilateral occipitotemporal cortex jointly hosted representations that are unique to the subordinate and the basic level. Finally, the behavioral action space model was captured by patterns of activation in occipitotemporal and SPL, and these neural patterns showed high similarity with the basic level model. Together, our results are in line with the view that lateral and ventral portions of the occipitotemporal cortex have flexible access to the representational space of observed actions at the subordinate and basic level, with a special role for the basic level (see also Rosch et al., 1976; Zhuang and Lingnau, 2021). In the following, we will discuss these points as well as limitations and directions for future studies in more detail.

### The representational space of observed actions in the LOTC

The LOTC has been shown to distinguish between different observed actions at the basic level on the basis of patterns of activation, with a generalization across the subordinate level (Wurm et al., 2016; Hafri et al., 2017; Wurm and Lingnau, 2015). We previously reasoned that if the LOTC is involved in processing observed actions at a conceptual-semantic level, it should not only distinguish between two actions A and B, but also capture the similarity structure of a wider range of actions. Several previous studies have demonstrated that this is indeed the case (Tucciarelli et al., 2019; Tarhan et al., 2021), and Wurm and Caramazza (2019) furthermore showed that representations in the LOTC generalize across visual stimuli and verbal descriptions of the same actions. In line with this view, several authors proposed a posterior-to-anterior concrete-to-abstract gradient in the LOTC (Lingnau and Downing, 2015; Wurm et al., 2017b; Papeo et al., 2019; Wurm and Caramazza, 2021).

The results of the current study extend these findings, demonstrating that the LOTC captures (a) the behavioral similarity structure of a different set of actions and (b) the unique representation at the subordinate and basic level, with a preference for the basic level (see also Iordan et al., 2015). Our results demonstrate that the LOTC plays a crucial role in representing actions at multiple levels in a situation in which participants are not biased towards processing one of these levels. An important next step for future studies will be to examine the impact of the observer's goal, emphasizing one level over one other. For instance, our goals (such as acquiring a new skill versus predicting the intention of another agent) may determine our focus on either concrete representations at the subordinate level or more abstract representations at the basic or superordinate level. Moreover, these results are consistent with the view that the lateral and ventral portion of the OTC hosts and integrates different action components (such as visual motion, body part, manipulation of tools, e.g. Lingnau and Downing, 2015; Tucciarelli et al., 2019) at varying levels of abstraction (see also Wurm and Lingnau, 2015; Wurm and Caramazza, 2021).

### The importance of the basic level for the representation of observed actions

Previous behavioral studies suggested that the basic level holds the maximized information for objects (Rosch et al., 1976; Murphy, 2004) and actions (Zhuang and Lingnau, 2021). The results of the current study provide insights into the neural basis underlying these behavioral findings. First, we found that patterns of activation in bilateral LOTC and SPL showed a higher similarity with the

basic level model in comparison to the subordinate and the superordinate level (Figure 4). Second, information that is unique to the basic level was captured in more widespread regions, including OTC, IPL and SPL (Figure 6B), in comparison to information that is unique to the subordinate level (Figure 6A), while we did not obtain any area that captured information that is unique to the superordinate level. Third, the behavioral action space model showed the highest similarity with neural patterns capturing the basic level model (see Figure 8C). Together, the results of the current study on observed actions are consistent with results by Iordan et al. (2015) demonstrating that objects at the basic level show the strongest similarity with patterns of activation in the lateral occipital cortex (LOC).

**The representation of actions at the superordinate level**

In contrast to the subordinate and basic level, we obtained no region that represented information that is unique to the superordinate level. Given that this result is based on the absence of evidence, this finding needs to be interpreted with care. There are several reasons that might account for this observation: (1) lack of power; (2) there is no region that host the unique information at the superordinate level; (3) information at the superordinate level is based on the combination of information at the subordinate and basic level; (4) the representation of actions at the superordinate level is more distributed than the representation of actions at the subordinate and basic level, making it unlikely to be revealed by methods that rely on local patterns of activation (searchlight, ROI analysis). In line with the latter interpretation, Abdollahi, Jastorff, and Orban (2013) found that different classes of observed actions (e.g., manipulation, locomotion and climbing, corresponding to the superordinate level) recruit different parts of parietal cortex. This observation might explain why we failed to obtain evidence for a unique – local - representation capturing the similarity between the different actions at the superordinate level. Further studies will be required to distinguish between these alternatives more systematically.

**The contribution of high-level visual features and semantic knowledge to the categorization of actions**

Since we did not compare the processing of visually presented actions with the processing of the corresponding action verbs or phrases, we cannot clearly separate between the contribution of higher-level visual features (or perceptual action precursors, see Wurm and Caramazza, 2021) and semantics. As pointed out above, a study that focused on this direct comparison revealed shared representations between patterns of activations for different actions depicted as videos and as written sentences found this to be the case exclusively in the lateral posterior temporal cortex (Wurm and Caramazza, 2019). That said, to be able to tolerate varying degrees of variability, we assume that the categorization of visually presented actions requires access to semantic knowledge for all three taxonomic levels, while the amount of variability is likely to differ between the levels. Specifically, members of the same subordinate level are assumed to have a higher number of shared features that can be exploited to categorize actions at the subordinate level (e.g. the body posture to distinguish between front crawl and backstroke; see also Zhuang and Lingnau, 2021). At the basic level, which is considered to be the most informative level (see e.g. Rosch et al., 1976), categorization has been proposed to rely on more common features that are shared across members of a category (e.g. applying some liquid to the body with hands or a tool to categorize an image to belong to the basic level category ‘cleaning the body’). Finally, to distinguish actions at the superordinate level, an even higher degree of generalization across high-level visual features is required (e.g. bringing an object to the mouth for the category ingestion).

**Comparison of object and action representations**

There exists a long tradition both in human (see e.g. Bach et al., 2014, Wurm et al., 2017b; Livi et al., 2019; Wurm and Caramazza, 2021) and monkey studies (e.g. Bonini et al., 2014) demonstrating that objects provide strong clues regarding the actions they afford, and that objects play an important

role during action observation. Consequently, the representation of objects and actions have been proposed to follow similar, though not identical, principles of organization (see e.g. Pillon and d'Honincthun, 2011; Wurm et al., 2017b). As an example, Wurm, Caramazza, and Lingnau (2017) demonstrated that features corresponding to actions directed towards persons and objects are preferentially represented in dorsal and ventral portions of the LOTC, parallel to the organization of information related to persons versus inanimate objects, respectively. Given these assumed similar principles of organizations, it is notoriously difficult to dissociate the representation of actions and objects. That said, Tucciarelli et al. (2019) showed that patterns of activation in the LOTC capture the behavioral similarity structure of actions, over and above variability due to action components such as objects involved in the action. Likewise, several previous studies demonstrated representations of observed actions in the LOTC that generalized across the object (Wurm and Lingnau, 2015; Wurm et al., 2016). More specifically, Wurm et al. (2016) demonstrated that it is possible to distinguish between observed opening and closing actions, irrespective of the object, in the LOTC, IPL and PMv, whereas decoding between objects across actions was restricted to clusters in the ventral stream (Wurm et al., 2016, Figures 3, 4 and Supplementary Figure 5). While we cannot clearly dissociate between the representation of action functions and object states/ object affordances in the current study, these previous studies indicate that action representations in the LOTC can be dissociated from representations of objects that are being manipulated.

**The impact of scene-related information**

Like objects, due to frequent co-occurrences, scenes can provide important cues regarding the kind of actions that are likely to be experienced (e.g. food-related actions in a kitchen scene and sport-related actions in a gym; see also (Wurm and Schubotz, 2012, 2017; Wurm et al., 2017a). To be able to account for the potential contribution of scene-related information in the current study, we used the activations of the second-to-last layer of a convolutional neural network (ResNet50) trained to distinguish between scene categories. A promising future step will be to systematically investigate the impact of the relationship between scene and action information on the representational space of actions.

**Limitations and future directions**

Previous studies revealed that posterior portions of the LOTC are recruited both by static images (Hafri et al., 2017; Tucciarelli et al., 2019) and by dynamic videos depicting actions (Hafri et al., 2017; Wurm et al., 2017b; Wurm and Lingnau, 2015). In addition, Hafri et al. (2017) successfully decoded observed actions across static and dynamic input in a number of regions, including the LOTC. That said, we cannot rule out that additional aspects relevant to the processing of the depicted actions, in particular, the corresponding kinematics, are not well captured in the current study.

Moreover, while we accounted for the impact of low-level visual and scene-related features, another important limitation of the current study is related to the fact that with the current stimulus material we cannot discount the impact of variance shared with object-related features involved in the actions. Moreover, it will be important to demonstrate to which degree the behavioral action space examined in the current study generalizes to a wider range of actions. Further studies are required to address these points.

**Conclusion**

Our results offer a neural perspective on categorical distinctions of observed actions across taxonomic levels, yielding insights into the mechanisms underlying behavioral flexibility aligned with the observer's goals. Lateral and ventral portions of the LOTC appear to capture the unique similarity of observed actions at the subordinate and basic level, with a preference for the basic level. Together, our results offer new perspectives on the hierarchical organization of observed actions and the neural basis of the basic level advantage.

## Figure captions

**Figure 1. Top panel**: Stimulus set and corresponding hierarchical structure (based on Zhuang and Lingnau, 2021). Left column: example stimuli (one out of six exemplars per subordinate action). 2nd-4th column: English labels of actions at the subordinate, basic and superordinate level. **Bottom panel**: model RDMs for observed actions at the subordinate (left), basic (middle), and superordinate (right) level. Each model RDM consists of a 72 x 72 matrix (12 actions, with 6 exemplars per action), where each cell in the matrix corresponds to the dissimilarity between a pair of actions. Yellow: high dissimilarity, blue: low dissimilarity.

**Figure 2.** Stimulus selection and validation. **A.** Behavioral action space model, averaged across N=18 participants and across six exemplars per action. **B.** 2-dimensional representation of the results shown in panel A, resulting from multidimensional scaling analysis (Borg and Groenen, 2005). For color code see legend in panel C. The numbers refer to the twelve actions at the subordinate level (1: 'to ride a motorbike', 2: 'to ride a bike', 3: 'to swim front crawl', 4: 'to swim backstroke', 5: 'to drink beer', 6: 'to drink water', 7: 'to eat an apple', 8: 'to eat cake', 9: 'to clean windows', 10: 'to brush teeth', 11: 'to do the dishes', and 12: 'to clean the face'. **C.** Dendrogram resulting from hierarchical cluster analysis, confirming the clusters corresponding to the three superordinate categories (locomotion, ingestion, cleaning) and the six basic level categories (see also Figure 1).

**Figure 3.** Experimental procedure used in the fMRI experiment. We used a rapid event-related design, where each trial consisted of a static image (1 s) depicting one of the twelve actions (see Table 1, 2nd column), followed by a fixation cross (3 s). During occasional catch trials (11 % of all trials), participants had to perform a category verification task, targeting the action presented in the previous trial. Questions to be answered during catch trials targeted the superordinate (locomotion?), basic (to swim?) or subordinate (to swim breaststroke?) level with an equal probability. Moreover, null events (22.2% of all trials) were included to enhance design efficiency (see text for details).

**Figure 4.** Partial correlations between neural RDMs and the models corresponding to the three taxonomic levels (see also Figure 1), regressing out a low-level visual control model, based on the first convolutional layer of a deep neural network (ResNet50, He et al., 2015), trained on a large scale image data set (ImageNet dataset) to distinguish image classes and a scene control model, based on the second-to last layer of ResNet50, trained to distinguish between different scene categories; Zhou et al., 2017 and 2018; see Section ROI-based RSA for details). Error bars show the standard error of the mean. Asterisks illustrate statistical significance (*: $q_{FDR} < 0.05$, **: $q_{FDR} < 0.01$, ***: $q_{FDR} < 0.005$) of one-sample t-tests against zero with FDR correction for all possible combinations of taxonomic level and ROI. Since a 2-factorial (level x ROI) repeated measures ANOVA revealed a significant interaction between taxonomic level and ROI, partial correlation values between the three taxonomic levels within each ROI were compared using pairwise comparisons (corrected for multiple comparisons using FDR).

**Figure 5.** Neural RDM (left panel) and the corresponding 2D visualization of the MDS results (right panel) obtained in bilateral LOTC (A and B) and bilateral V1 (C and D) after regressing out the low-level visual control model (based on the first convolutional layer of ResNet50, trained to distinguish image classes) and the scene control model (based on the second-to last layer of ResNet50, trained to distinguish between scene categories; Zhou et al., 2017 and 2018; see Section ROI-based RSA for details), using the same ROIs as those shown in Figure 4. Colors to indicate actions belonging to the three superordinate level categories (red-locomotion, blue-ingestion, and green-cleaning) are the same as in Figure 2. Opened versus filled symbols indicate different actions at the basic level, whereas circles versus squares are used to distinguish actions at the subordinate level.

**Figure 6.** Results of the searchlight-based multiple regression RSA for the subordinate (left panel), basic (middle panel) and superordinate (right panel) model. **A**: Multiple regression RSA, regressing out the low-level visual control model and the scene control model (see section Searchlight-based RSA for details). **B**: Multiple regression RSA, regressing out the low-level visual control model, the scene control model and the remaining two taxonomic models. For group statistics (N=23), beta weights resulting from the multiple regression RSA were entered into one-sample t-tests against zero. The statistical t-value maps were corrected for multiple comparisons using TFCE (z>1.65, TFCE-corrected p <.05, one-tailed) as implemented in CoSMoMVPA (Oosterhof et al., 2016).

**Figure 7.** Conjunction map for the unique representation of observed actions at the subordinate and basic level (see Figure 6B). See text for details.

**Figure 8.** Brain areas capturing the representational space of actions obtained from behavioral ratings (see Figure 2A). **A**: Brain areas capturing the behavioral dissimilarity structure after regressing out the low-level visual control model and the scene control model. **B-D**: Brain areas capturing the behavioral dissimilarity structure after regressing out the low-level visual control model, the scene control model and models for the (B) subordinate, (C) basic and (D) superordinate level. For the results of a control analysis examining the areas revealed by a whole-brain searchlight RSA for the low-level visual control model, see Figure 8-1.

**Figure 8-1.** Results of the whole-brain searchlight RSA for the low-level visual control model. See methods section for details.

## Tables

**Table 1.** Overview Regions of Interest. 1st column: abbreviations used in the following figures (LH: left hemisphere, RH: right hemisphere). 2nd column: peak MNI coordinates (identified based on the RFX GLM contrast 'all actions vs. baseline'; see text for details). 3rd and 4th column: label and anatomical atlas used for ROI selection.

| ROIs | MNI 152 | | | Label | Atlas |
|---|---|---|---|---|---|
| | X | Y | Z | | |
| V1, LH | -14 | -92 | -4 | visual cortex (V1 BA 17L) | Jülich Histological atlas |
| V1, RH | 12 | -90 | 0 | visual cortex (V1 BA 17R) | Jülich Histological atlas |
| LOTC, LH | -42 | -66 | 0 | combination of inferior lateral occipital cortex and temporooccipital part | Harvard-Oxford Cortical structural |
| LOTC, RH | 46 | -74 | 0 | combination of inferior lateral occipital cortex and temporooccipital part | Harvard-Oxford Cortical structural |
| SPL, LH | -30 | -52 | 52 | superior parietal lobule | Harvard-Oxford Cortical structural |
| SPL, RH | 20 | -56 | 52 | superior parietal lobule | Harvard-Oxford Cortical structural |
| IPL, LH | -58 | 10 | 24 | inferior parietal lobule (PFL) | Jülich Histological atlas |
| IPL, RH | 62 | -38 | 12 | inferior parietal lobule (PFR) | Jülich Histological atlas |
| dPM, LH | -6 | 10 | 44 | premotor cortex (BA6) | Jülich Histological atlas |
| dPM, RH | 2 | 10 | 46 | premotor cortex (BA6) | Jülich Histological atlas |
| IFG, LH | -40 | 10 | 24 | inferior frontal gyrus pars opercularis | Harvard-Oxford Cortical structural |
| IFG, RH | 44 | 6 | 24 | inferior frontal gyrus pars opercularis | Harvard-Oxford Cortical structural |

**Table 2.** Peak locations for the three taxonomic levels in the whole-brain searchlight

| | | MNI 152 | | | Labels of brain regions |
|---|---|---|---|---|---|
| | | x | y | z | |
| Regressing out low-level visual control model and scene control model only | Subordinate level | 31 | -48 | -9 | R Temporal occipital fusiform cortex |
| | Basic level | 36 | -62 | -6 | R occipital fusiform cortex |
| | Superordinate level | 30 | -94 | -4 | R occipital pole |
| Regressing out low-level visual control model, scene control model and the other two models | Subordinate level | 24 | -58 | -12 | R Temporal occipital fusiform cortex |
| | Basic level | 40 | -62 | -10 | R occipital fusiform cortex |
| | Superordinate level | - | - | - | - |

## References


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

Aflalo T, Zhang CY, Rosario ER, Pouratian N, Orban GA, Andersen RA (2020) A shared neural substrate for action verbs and observed actions in human posterior parietal cortex. Sci Adv 6.

Ariani G, Wurm MF, Lingnau A (2015) Decoding internally and externally driven movement plans. J Neurosci 35:14160–14171.

Bach P, Nicholson T, Hudsons M (2014) The affordance-matching hypothesis: How objects guide action understanding and prediction. Front Hum Neurosci 8:1–13.

Benjamini Y, Hochberg Y (2001) The control of the false discovery rate in multiple testing under dependency. Ann Stat:1165–1188.

Binkofski F, Buxbaum LJ (2013) Two action systems in the human brain. Brain Lang 127:222–229.

Bonini L, Maranesi M, Livi A, Fogassi L, Rizzolatti G (2014) Space-dependent representation of objects and other's action in monkey ventral premotor grasping neurons. J Neurosci 34:4108–4119.

Bonner MF, Epstein RA (2018) Computational mechanisms underlying cortical responses to the affordance properties of visual scenes. PLoS Comput Biol 14:1–31.

Borg I, Groenen PJF (2005) Modern Multidimensional Scaling -- Theory and Applications. New York: Springer.

Carlson T, Tovar DA, Alink A, Kriegeskorte N (2013) Representational dynamics of object vision: The first 1000 ms. J Vis 13:1–19.

Dehaqani MRA, Vahabie AH, Kiani R, Ahmadabadi MN, Araabi BN, Esteky H (2016) Temporal dynamics of visual category representation in the macaque inferior temporal cortex. J Neurophysiol 116:587–601.

Diedrichsen J, Kriegeskorte N (2017) Representational models : A common framework for understanding encoding ,. PLoS Comput Biol:1–33.

Gallivan JP, Culham JC (2015) Neural coding within human brain areas involved in actions. Curr Opin Neurobiol 33:141–149.

Gallivan JP, McLean DA, Flanagan JR, Culham JC (2013) Where one hand meets the other: Limb-specific and action-dependent movement plans decoded from preparatory signals in single human frontoparietal brain areas. J Neurosci 33:1991–2008.

Gauthier I, Anderson AW, Tarr MJ, Skudlarski P, Gore JC (1997) Levels of categorization in visual recognition studied using functional magnetic resonance imaging. Curr Biol 7:645–651.

Grafton ST, Hamilton AF d. C (2007) Evidence for a distributed hierarchy of action representation in the brain. Hum Mov Sci 26:590–616.

Grill-Spector K, Kanwisher N (2005) Visual RecognitionVisual recognition: as soon as you know it is there, you know what it is. Psychol Sci 16:152–160.

Grill-Spector K, Weiner KS (2014) The functional architecture of the ventral temporal cortex and its role in categorization. Nat Rev Neurosci 15:536–548.

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

Hamilton AF d. C, Grafton ST (2006) Goal representation in human anterior intraparietal sulcus. J Neurosci 26:1133–1137.

Hamilton AF d. C, Grafton ST (2008) Action outcomes are represented in human inferior frontoparietal cortex. Cereb Cortex 18:1160–1168.

He K, Zhang X, Ren S, Sun J (2015) Deep residual learning for image recognition. 2016 IEEE Conf Comput Vis Pattern Recognit:770–778.

Hétu S, Grégoire M, Saimpont A, Coll MP, Eugène F, Michon PE, Jackson PL (2013) The neural network of motor imagery: An ALE meta-analysis. Neurosci Biobehav Rev 37:930–949.

Hoffman P, Binney RJ, Lambon Ralph MA (2015) Differing contributions of inferior prefrontal and anterior temporal cortex to concrete and abstract conceptual knowledge. Cortex 63:250–266.

Iordan MC, Greene MR, Beck DM, Li F-F (2015) Basic Level Category Structure Emerges Gradually across Human Ventral Visual Cortex. J Cogn Neurosci 27:1426–1446.

Kadmon Harpaz N, Flash T, Dinstein I (2014) Scale-invariant movement encoding in the human motor system. Neuron 81:452–462.

Kilner JM (2011) More than one pathway to action understanding. Trends Cogn Sci 15:352–357.

Krasovsky A, Gilron R, Yeshurun Y, Mukamel R (2014) Differentiating intended sensory outcome from underlying motor actions in the human brain. J Neurosci 34:15446–15454.

Kriegeskorte N (2015) Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. Annu Rev Vis Sci 1:417–446.

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

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008) Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. Neuron 60:1126–1141.

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

Levin B (1993) English verb classes and alternations: A preliminary investigation. University of Chicago Press.

Lingnau A, Downing PE (2015) The lateral occipitotemporal cortex in action. Trends Cogn Sci 19:268–277.

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

Macé MJM, Joubert OR, Nespoulous JL, Fabre-Thorpe M (2009) The time-course of visual categorizations: You spot the animal faster than the bird. PLoS One 4:e5927.

Mack ML, Gauthier I, Sadr J, Palmeri TJ (2008) Object detection and basic-level categorization: Sometimes you know it is there before you know what it is. Psychon Bull Rev 15:28–35.

Mahendran A, Vedaldi A (2015) Understanding deep image representations by inverting them. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit:5188–5196.

Majdandić J, Bekkering H, Van Schie HT, Toni I (2009) Movement-specific repetition suppression in ventral and dorsal premotor cortex during action observation. Cereb Cortex 19:2736–2745.

Mason RL, Gunst RF, Hess JL (2003) Statistical Design and Analysis of Experiments: With Applications to Engineering and Science.

Murphy G (2004) The big book of concepts. Cambridge,MA: MIT press.

Nichols T, Brett M, Andersson J, Wager T, Poline JB (2005) Valid conjunction inference with the minimum statistic. Neuroimage 25:653–660.

Nili H, Wingfield C, Walther A, Su L, Marslen-Wilson W, Kriegeskorte N (2014) A Toolbox for Representational Similarity Analysis. PLoS Comput Biol 10.

Oosterhof NN, Connolly AC, Haxby J V. (2016) CoSMoMVPA: Multi-modal multivariate pattern analysis of neuroimaging data in matlab/GNU octave. Front Neuroinform 10:1–27.

Oosterhof NN, Tipper SP, Downing PE (2012) Viewpoint ( In ) dependence of Action Representations : An MVPA Study. J Cogn Neurosci:975–989.

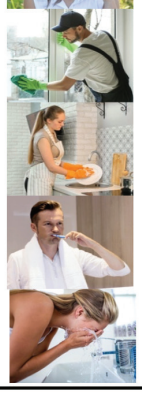
Oosterhof NN, Wiggett AJ, Diedrichsen J, Tipper SP, Downing PE (2010) Surface-Based Information Mapping Reveals Crossmodal Vision – Action Representations in Human Parietal and Occipitotemporal Cortex. J Neurophysiol:1077–1089.

Papeo L, Agostini B, Lingnau A (2019) The large-scale organization of gestures and words in the middle temporal gyrus.

Pillon A, d'Honincthun P (2011) A common processing system for the concepts of artifacts and actions? evidence from a case of a disproportionate conceptual impairment for living things. Cogn Neuropsychol 28:1–43.

Rosch E, Mervis CB, Gray WD, Johnson DM, Boyes-braem P (1976) Basic Objects in Natural Categories. Cogn Psychol 8:382–439.
Schwarzbach J (2011) A simple framework (ASF) for behavioral and neuroimaging experiments based on the psychophysics toolbox for MATLAB. Behav Res Methods 43:1194–1201.
Smith SM, Nichols TE (2009) Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44:83–98.
Spunt RP, Kemmerer D, Adolphs R (2016) The neural basis of conceptualizing the same action at different levels of abstraction. Soc Cogn Affect Neurosci 11:1141–1151.
Tarhan L, De Freitas J, Konkle T (2021) Behavioral and neural representations en route to intuitive action understanding. Neuropsychologia 163:108048.
Tucciarelli R, Wurm MF, Baccolo E, Lingnau A (2019) The representational space of observed actions. Elife 8:e47686.
Turella L, Rumiati R, Lingnau A (2020) Hierarchical Action Encoding Within the Human Brain. Cereb Cortex 30:2924–2938.
Vallacher RR, Wegner DM (1985) A theory of action identification. Hillsdale, NJ: Lawrence Erlbaum Associates.
Wegner DM, Vallacher RR (1986) Action identification. B: Handbook of motivation and cognition: Foundations of social behavior (Sorrentino RM, Higgins ET, ред), pp 550–582. New York: Guilford.
Wurm MF, Ariani G, Greenlee MW, Lingnau A (2016) Decoding Concrete and Abstract Action Representations During Explicit and Implicit Conceptual Processing. Cereb Cortex 26:3390–3401.
Wurm MF, Artemenko C, Giuliani D, Schubotz RI (2017a) Action at its place: Contextual settings enhance action recognition in 4- to 8-year-old children. Dev Psychol 53:662–670.
Wurm MF, Caramazza A (2019) Distinct roles of temporal and frontoparietal cortex in representing actions across vision and language. Nat Commun 10:1–10.
Wurm MF, Caramazza A (2021) Two 'what' pathways for action and object recognition. Trends Cogn Sci:1–14.
Wurm MF, Caramazza A, Lingnau A (2017b) Action Categories in Lateral Occipitotemporal Cortex Are Organized Along Sociality and Transitivity. J Neurosci 37:562–575.
Wurm MF, Lingnau A (2015) Decoding Actions at Different Levels of Abstraction. J Neurosci 35:7727–7735.
Wurm MF, Schubotz RI (2012) Squeezing lemons in the bathroom: Contextual information modulates action recognition. Neuroimage 59:1551–1559.
Wurm MF, Schubotz RI (2017) What's she doing in the kitchen? Context helps when actions are hard to recognize. Psychon Bull Rev 24:503–509.
Yamins DLK, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc Natl Acad Sci U S A 111:8619–8624.
Zeiler MD, Fergus R (2014) Visualizing and understanding convolutional networks. Eur Conf Comput vision Springer:818–833.
Zhou B, Lapedriza A, Khosla A, Oliva A, Torralba A (2018) Places: A 10 Million Image Database for Scene Recognition. IEEE Trans Pattern Anal Mach Intell 40:1452–1464.
Zhou B, Lapedriza A, Torralba A, Oliva A (2017) Places: An Image Database for Deep Scene Understanding. J Vis 17:296.
Zhuang T, Lingnau A (2021) The characterization of actions at the superordinate, basic and subordinate level. Psychol Res.

A

| Stimuli | Subordinate level | Basic level | Superordinate level |
|---|---|---|---|
|  | to ride a motorbike | to ride | locomotion |
| | to ride a bike | | |
| | to swim front crawl | to swim | |
| | to swim backstroke | | |
|  | to drink water | to drink | ingestion |
| | to drink beer | | |
| | to eat cake | to eat | |
| | to eat an apple | | |
|  | to clean windows | to do housework | cleaning |
| | to do the dishes | | |
| | to brush teeth | to clean the body | |
| | to clean the face | | |

B