The Journal of Neuroscience

http://jneurosci.msubmit.net

JN-RM-0188-15R2

Decoding actions at different levels of abstraction

Moritz Wurm, University of Trento
Angelika Lingnau, University of Trento

Commercial Interest: No

Decoding actions at different levels of abstraction

Abbreviated title: Decoding actions concepts

Moritz. F. Wurm[1] & Angelika Lingnau[1,2]

1: Center for Mind/ Brain Sciences, University of Trento, Via delle Regole 101, 38100 Mattarello (TN), Italy
2: Department of Cognitive Sciences, University of Trento, Corso Bettini, 31, 38068 Rovereto (TN), Italy

Corresponding author:
Moritz F. Wurm
Center for Mind/Brain Sciences
University of Trento
Via delle Regole, 101
38100, Mattarello (TN) - ITALY
Tel: ++39 0461 28 2783
Email: moritz.wurm@unitn.it

Number of pages: 30

Number of figures: 4
Number of tables: 3

Number of words:
Abstract: 250
Introduction: 529
Discussion: 1542

Conflict of interest: none

Acknowledgements: We would like to thank Gilles Vannuscorps, Liuba Papeo, Seth Levine, and two anonymous reviewers for helpful comments on the manuscript. This research was supported by the Provincia Autonoma di Trento and the Fondazione Cassa di Risparmio di Trento e Rovereto.

ABSTRACT

Brain regions that mediate action understanding must contain representations that are action-specific and at the same time tolerate a wide range of perceptual variance. Whereas progress has been made in understanding such generalization mechanisms in the object domain, the neural mechanisms to conceptualize actions remain unknown. In particular, there is ongoing dissent between motor-centric and cognitive accounts whether premotor cortex or brain regions in closer relation to perceptual systems, i.e., lateral occipitotemporal cortex, contain neural populations with such mapping properties. To date, it is unclear to which degree action-specific representations in these brain regions generalize from concrete action instantiations to abstract action concepts. However, such information would be crucial to differentiate between motor and cognitive theories. Using ROI-based and searchlight-based fMRI multivoxel pattern decoding, we sought for brain regions in human cortex that manage the balancing act between specificity and generality. We investigated a *concrete level* that distinguishes actions based on perceptual features (e.g., opening vs. closing a specific bottle), an *intermediate level* that generalizes across movement kinematics and specific objects involved in the action (e.g., opening different bottles with cork or screw cap), and an *abstract level* that additionally generalizes across object category (e.g., opening bottles or boxes). We demonstrate that inferior parietal and occipitotemporal cortex code actions at abstract levels whereas premotor cortex codes actions at the concrete level only. Hence, occipitotemporal, but not premotor, regions fulfill the necessary criteria for action understanding. This result is compatible with cognitive theories but strongly undermines motor theories of action understanding.

INTRODUCTION

Neural populations capable of mediating action understanding need to be action-specific and at the same time generalize across perceptual features from concrete observed actions to more abstract levels (Figure 1A). Whereas similar conceptualization problems have been intensively studied in the object recognition domain (e.g., Konen and Kastner, 2008; Mur et al., 2010; Fairhall and Caramazza, 2013; Anzellotti et al., 2014; Cichy et al., 2014), the neural basis of action abstraction remains unexplored.

There is an ongoing debate about the cortical substrates containing neural populations that manage the balancing act between action specificity and feature generality. Mirror neurons in ventral premotor cortex (PMv) were proposed to show such properties and therefore have been suggested to represent the central computational units of action understanding (Rizzolatti and Craighero, 2004; Rizzolatti et al., 2014). However, criticism has been raised that mirror neurons might not show the degree of generality as originally claimed (Kilner, 2011; Cook and Bird, 2013). Alternatively, analogous to conceptualization in the object domain, more posterior regions in closer proximity to the visual system, e.g., lateral occipitotemporal cortex (LOTC), might generalize from perceptually variable instances of actions to abstract action concepts (Oosterhof et al., 2013; Watson et al., 2013). However, the degree of generality of action representations has not yet been established in any of these regions.

Here, we used cross-conditional multivoxel pattern (MVP) analysis of fMRI data to identify action representations at three levels of representation: a *concrete*, an *intermediate*, and an *abstract* level. Participants watched videos of eight actions (open and close two different exemplars of bottles and boxes, each requiring different kinematics) and responded to occasionally occurring catch trials. We decoded *concrete* actions by training a classifier with trials that display the opening or closing of a particular bottle and testing it on different trials from the same conditions, i.e., within the same object exemplars and kinematics (Figure 1B, upper row). To decode actions at an *intermediate* level, we trained the classifier with trials that display the opening or closing of a particular bottle and tested it with trials that display the opening or closing of a different bottle, i.e., across object exemplars and kinematics (Figure 1B, middle row). To decode actions at an *abstract* level, we trained the classifier with trials that display the opening or closing of a bottle and tested it with trials that display the opening or closing of a box, i.e., across object category and kinematics (Figure 1B, lower row).

Our design overcomes limitations of recent neuroimaging studies that use object-directed actions to study action representations that generalize across kinematics (Hamilton and Grafton, 2006, 2008), hand posture (Oosterhof et al., 2010), or viewpoint (Oosterhof et al., 2012): By decoding not only across kinematics but also across distinct objects and object categories, action outcomes differ perceptually at intermediate and abstract levels, a condition that is crucial in order to disambiguate whether identified representations are sensitive to the action or to concrete perceptual features of an object's state (e.g., a specific closed box). Importantly, the direct comparison between different levels of abstraction allows more relative estimations of the generalization capacities of action-coding neural populations in different regions that supposedly provide the basis for action understanding.

METHODS

Participants. Twenty-two healthy adults (11 females; mean age, 28 years; age range, 19-41 years) volunteered to participate in the experiment. All participants were right-handed with normal or corrected-to-normal vision and no history of neurological or psychiatric disease. Participants gave written informed consent prior to participation in the study. The experimental procedures were approved by the Ethics Committee for research involving human subjects at the University of Trento, Italy.

Stimuli. The stimulus set consisted of three exemplars of eight actions (24 action videos in total). The actions were opening and closing (two-level factor ACTION) of four different objects (two bottles and two cosmetic boxes; two-level factor OBJECT CATEGORY). One object exemplar of each object category had a screw cap, hence requiring a wrist rotation, the other object exemplar was opened and closed with push and pull kinematics, respectively (two-level factor KINEMATICS). Catch trials consisted of three exemplars of the eight actions that ended with an additional action step (moving, tilting or lifting the object; 24 catch trial videos in total). Action videos were filmed from a 180° third person perspective using a Canon 5D Mark II camera and edited in iMovie (Apple) and Matlab (MathWorks). All 48 videos were identical in terms of action timing, i.e., the videos started with hands on the table moving towards the object, followed by the object manipulation, and ended with hands moving to the same position of the table. Videos were in black and white, had a length of 2 s (30 frames per second), and had a resolution of 400 x 300 pixels.

For intermediate and abstract levels, we aimed targeting neural populations that are capable of differentiating perceptually similar but conceptually dissimilar actions that at the same time generalize across conceptually similar but perceptually dissimilar actions. Hence, “open a water bottle” and “close a water bottle” should be perceptually more similar than “open a water bottle” and “open a wine bottle” (intermediate level) or “open a bottle” and “open a box” (abstract level). To test if our stimuli match these criteria, we estimated the visual similarity between the action videos. To this end, we correlated each video with each other video frame by frame, i.e., we correlated frame 1 of video A with frame 1 of video B, etc. We then averaged the correlation coefficients across frames to obtain a mean correlation matrix of the 24 x 24 action videos (2 actions x 2 kinematics x 2 object categories x 3 action exemplars). In a second averaging step, we computed the means of to-be-classified actions (open vs. close bottle A, open vs. close bottle B, etc.), and of same actions across object exemplar (intermediate; open bottle A vs. open bottle B, close bottle A vs. close bottle B, etc.) and object category (abstract; open bottle A vs. open box B, close bottle A vs. close bottle B, etc.). The results demonstrate that, in line with our criteria, pixelwise similarities of to-be-classified actions were substantially higher (r = 0.54) than similarities of same actions at intermediate (r=0.27) and abstract levels (r = 0.12), suggesting that to-be-classified actions are perceptually more similar (i.e., there is fewer perceptual information that can be exploited by the classifier) than the actions that are generalized at intermediate and abstract levels.

In the scanner, stimuli were back-projected onto a screen (60 Hz frame rate, 1024 x 768 pixels screen resolution) via a liquid crystal projector (OC EMP 7900, Epson Nagano, Japan) and viewed through a mirror mounted on the head coil (video presentation 6.9° x 5.2° visual angle). Stimulus presentation, response collection, and synchronization with the scanner were controlled with ASF (Schwarzbach, 2011) and the Matlab Psychtoolbox-3 for Windows (Brainard, 1997).

Design of the fMRI experiment. Stimuli were presented in an event-related design. In each trial, videos (2 s) were followed by a 1 s fixation period. 18 trials were shown per block. Each of the nine conditions (eight action conditions plus one catch trial) was presented twice per block. Five blocks were presented per run, separated by 12 s fixation periods. Each run started with a 10 s fixation period and ended with a 16 s fixation period. In each run, the order of conditions was first-order counterbalanced (Aguirre, 2007). Each participant was scanned in a single session consisting of 12 functional scans and one anatomical scan. For each of the nine conditions there was a total of 2 (trials per block) x 5 (blocks per run) x 12 (runs per session) = 120 trials per condition.

Task. Participants were instructed to attentively watch the movies. They were asked to press a button with the right index finger on a response button box whenever an action was followed by an additional action step (moving, tilting or lifting the object). Participants could respond either during the movie or during the fixation phase after the movie. To ensure that participants followed the instructions correctly, they completed a practise block outside the scanner.

Data acquisition. Functional and structural data were collected using a 4 T Bruker MedSpec Biospin MR scanner and an 8-channel birdcage head coil. Functional images were acquired with a T2*-weighted gradient echo-planar imaging (EPI) sequence with fat suppression. Acquisition parameters were a repetition time of 2 s, an echo time of 21 ms, a flip angle of 75°, a field of view of 192 mm, a matrix size of

64 x 64, and voxel resolution of 3 x 3 x 2 mm. We used 43 slices, acquired in ascending interleaved order, with a thickness of 2 mm and 15 % gap (0.3 mm). Slices were tilted to run parallel to the superior temporal sulcus. We thereby covered the full temporal lobe including the poles. In few participants the most superior part of prefrontal and parietal cortex (approx. 1 cm) was not covered. In each functional run, 172 images were acquired. Before each run we performed an additional scan to measure the point-spread function (PSF) of the acquired sequence to correct the distortion expected with high-field imaging (Zaitsev et al., 2004).

Structural T1-weigthed images were acquired with an MPRAGE sequence (176 sagittal slices, TR = 2.7 s, inversion time = 1020 ms, FA = 7°, 256 x 224 mm FOV, 1 x 1 x 1 mm resolution).

Preprocessing. Data were analyzed using BrainVoyager QX 2.4 (BrainInnovation) in combination with the BVQX Toolbox and custom software written in Matlab (MathWorks).

Distortions in geometry and intensity in the echo-planar images were corrected on the basis of the PSF data acquired before each EPI scan (Zeng and Constable, 2002). The first 4 volumes were removed to avoid T1 saturation. The first volume of the first run was aligned to the high-resolution anatomy (6 parameters). Data were 3D motion corrected (trilinear interpolation, with the first volume of the first run of each participant as reference), followed by slice time correction and high-pass filtering (cutoff frequency of 3 cycles per run). Spatial smoothing was applied with a Gaussian kernel of 8 mm FWHM for univariate analysis and 3 mm FWHM for MVPA. Note that smoothing up to 8 mm FWHM can increase the sensitivity in MVP correlation analysis whereas, in MVP decoding, smoothing between 0 and 8 mm showed no substantial increases or decreases in decoding accuracy (Op de Beeck, 2010). A recent study, however, revealed that smoothing between 2 and 3 mm FWHM had best effects on MVP decoding (Gardumi et al., 2014). For group analysis, both anatomical and functional data were transformed into Talairach space using trilinear interpolation.

Cortex-based alignment. For each hemisphere and participant, surface meshes of the border between grey and white matter were segmented and reconstructed. Resulting surfaces were smoothed and inflated. In addition, spherical surface meshes were generated and morphed to a standard spherical surface. On the basis of multiscale surface curvature maps (which reflect the gyral/sulcal folding pattern) with four coarse-to-fine levels of smoothing, the standardized spherical surfaces of all participants were aligned to an average spherical surface using a coarse-to-fine moving target approach (Fischl et al., 1999; Goebel et al., 2006). Transformation matrices of the established correspondence mapping were used to align surface maps entering statistical group analyses. In addition, average folded and inflated group surfaces of both hemispheres were created. Statistical maps were projected onto these group surfaces.

MVP analysis. Multivoxel pattern analysis was carried out using a linear support vector machine (SVM) classifier as implemented by LIBSVM (Chang and Lin, 2011). MVP analysis was carried out both ROI- and searchlight-based. The ROI analysis (see section ROI Analysis) was used to directly investigate the level of abstractness (concrete, intermediate, abstract) represented in core regions involved in action observation, i.e., ventral premotor cortex (PMv), anterior intraparietal sulcus / inferior parietal lobe (IPL), and lateral occipitotemporal cortex (LOTC). Note that we included IPL due to its prominent role in action observation, despite the fact that it is not well suited to differentiate between motor and cognitive theories. One the one hand, IPL is typically counted to the motor system as its homologue in the monkey has been reported to contain mirror neurons (Fogassi et al., 2005) and is suggested to encode motor and visuospatial aspects of actions such as object affordances (Fagg and Arbib, 1998) as well as action outcomes and intentions (Fogassi et al., 2005; Rizzolatti et al., 2014). On the other hand, IPL is considered to belong to a supramodal semantic system (Binder and Desai, 2011) and shows high degrees of abstraction in object recognition and thus classical properties of the ventral “what” stream (Konen and Kastner, 2008; Fairhall and Caramazza, 2013). Motor and cognitive views therefore do not offer opposing predictions regarding generalization capacities in IPL. The searchlight analysis (see section Surface-based Searchlight Analysis) was carried out to identify putative additional regions representing action concepts.

ROI Definition. ROIs were defined separately for each participant on the basis of univariate statistical maps using a similar approach as described in Oosterhof et al. (Oosterhof et al., 2010). In brief, to constrain peak cluster identification in individual contrast maps and thus to avoid possibly arbitrary selection decisions of the experimenter (Oosterhof et al., 2012), individual ROIs were defined as circles around the peak vertex of individual statistical surface maps that lie within a circle of 12 mm radius centered around the group peak vertex. To this end, we first computed a group random-effects (RFX) general linear model (GLM). Design matrices contained predictors of the 8 (ACTION x OBJECT CATEGORY x KINEMATICS) conditions, catch trials, and of 6 parameters resulting from 3D motion correction (x, y, z translation and rotation). Each predictor was convolved with a dual-gamma hemodynamic impulse response function (Friston et al., 1998). Each trial was modeled as an epoch lasting from video onset to offset (2 s). The resulting reference time courses were used to fit the signal time courses of each voxel. To identify the group peak vertices, we contrasted all eight conditions vs. baseline (where baseline is defined as all time points not modeled in the design matrix). The resulting group contrast was projected on the cortex-based aligned group surface and peak vertices were identified in anatomically defined cortical regions in both hemispheres (ventral precentral gyrus, anterior intraparietal sulcus, posterior middle temporal gyrus). To identify individual peak vertices, we computed single-subject GLM contrasts [all eight conditions vs. baseline] in volume space using design matrices as described above. After projecting the resulting individual maps on the surface, peak vertices were identified within circles of 12 mm radius centered around the group peak vertices. Finally, disc-shaped ROIs (12 mm radius) were defined around the individual peak vertex of each participant.

ROI MVPA. The following steps were done for each participant and ROI separately. Within each individual ROI (230 vertices on average), beta weights were estimated on the basis of 5 trials per condition and run resulting in two beta estimates per condition and run. Design matrices thus contained 16 predictors of action conditions, 2 catch trials predictors, and the 6 predictors of the 3D motion correction parameters. Predictors were orthogonal to each other (highest correlation in any of the runs and any of the participants was $R^2 = 0.12$; overall mean: $R^2 = 0.009$). In total, there were 24 beta estimates per condition for each vertex (i.e., 24 multivariate beta patterns per condition). Classification accuracies were computed using leave-one-out cross validation, i.e., the classifier was trained using the data of 23 patterns and tested on its accuracy at classifying the unseen data from the remaining pattern. This procedure was carried out in 24 iterations, using all possible combinations of training and test patterns. The classification accuracies from the 24 iterations were averaged to give a mean accuracy score per test. To decode actions at the concrete level, the classifier was trained to discriminate between open and close bottle A and tested on open vs. close bottle A (Figure 1B, upper row). The same classification procedure was repeated for the remaining 3 objects and the mean of all 4 tests was computed. To decode actions at the intermediate level, the classifier was trained to discriminate between open and close bottle A and tested on open vs. close bottle B (Figure 1B, middle row). Again, the classification procedure was repeated for bottle B → bottle A, box A → box B, box B → box A, and the mean of the 4 tests were computed. Decoding at the intermediate level therefore targeted action representations that generalize across object exemplars (exemplar A and B) and kinematics (screw and push/pull). To decode actions at the abstract level, the classifier was trained to discriminate between open and close bottle A and tested on open vs. close box B (Figure 1B, lower row). The classification procedure was repeated for bottle B → box A, box A → bottle B, box B → bottle A, and the mean of the 4 tests was computed. Decoding at the abstract level therefore targeted action representations that generalize across object categories (bottles and boxes) and kinematics (screw and push/pull). For the intermediate and abstract levels ("across object" classification) we also used the leave-one-out cross validation procedure to ensure that the results are as comparable as possible to the results of the concrete level ("within object" classification). The mean classification accuracy for each abstraction level, ROI, and participant was entered into a one-tailed one-sample t-test against the classification expected by chance (50%). Statistical results were FDR-corrected for the number of one sample *t* tests, i.e., 6 ROIs x 3 levels = 18 tests (Benjamini and Yekutieli, 2001).

To assess statistical significance of the differences between decoding accuracies of different abstraction levels and regions, a repeated measures ANOVA with ABSTRACTION LEVEL, ROI, and HEMISPHERE and post hoc paired samples *t* tests were used. Within each region, we considered the following three possible scenarios:

(1) *"Concrete only" regions*. A region encodes action information at the concrete level but not at intermediate and/or abstract levels (Figure 1C). In this case, three criteria must be met: (a) Significant decoding at the concrete level, (b) no significant decoding at intermediate and/or abstract levels, and (c) significant differences between concrete and intermediate and/or abstract levels, respectively.

(2) *"All levels" regions*. A region encodes action information at all levels of abstraction (Figure 1D). In this case, two criteria must be met: (a) Significant decoding at concrete, intermediate, and abstract levels, and (b) significant differences between concrete and intermediate and between intermediate and abstract levels, respectively. A stepwise decrease from concrete to abstract is expected because in the concrete decoding, action information from all three levels can be exploited by the classifier whereas for the intermediate decoding only information from the intermediate and abstract level can be exploited and for the abstract decoding only abstract action information can be exploited.

(3) *"Abstract only" regions*. A region encodes action information at the abstract level only (Figure 1E). In this case, two criteria must be met: (a) Significant decoding at concrete, intermediate, and abstract levels, and (b) no significant differences between concrete and intermediate and between intermediate and abstract levels, respectively. No differences between the three levels are expected because in all three levels the same (abstract) information is picked up by the classifier.

Across regions, we further examined double dissociations of abstraction level and region. To do so, we considered the following scenarios: In case region X encodes concrete action information only and region Y encodes abstract action information only (Figure 1C and E, respectively), an interaction of ABSTRACTION LEVEL and ROI is expected. However, in case region X encodes concrete action information only and region Y encodes actions at both concrete and more abstract levels (Figure 1C and D, respectively), *no interaction* of ABSTRACTION LEVEL and ROI is expected. This is because region Y should show higher decoding accuracies for concrete compared to abstract levels. Importantly, the relative differences between concrete and intermediate/abstract levels can be similar in region X and Y, in which case no interaction would be observed. Hence, for the case that region Y, but not region X, encodes actions at intermediate and abstract levels the following criteria must be met: (a) significant decoding accuracies for concrete, intermediate and abstract levels in region Y, (b) a significant main effect of ABSTRACTION LEVEL, (c) a significant main effect of ROI, and (d) significant decoding differences between region X and Y at intermediate and abstract levels.

Surface-based Searchlight MVPA. To identify any additional regions coding actions at different levels of abstraction we carried out a surface-based (Oosterhof et al., 2010) searchlight pattern classification (Kriegeskorte et al., 2006). For each participant and hemisphere, we transformed volume time courses into surface mesh time courses. Volume time courses were sampled along the mesh vertex normal from -1 to 3 mm. GLM computation and MVPA classification was carried out using identical parameters and procedures as for the ROI MVPA. The classification accuracy was assigned to the central vertex. Resulting individual surface accuracy maps were anatomically aligned using the transformation parameters of the cortex-based alignment. Aligned maps were entered into a one-sample t-test to identify vertices where classification was significantly above chance. We reasoned that wherever actions can be decoded at the intermediate level (action classification across object exemplar) actions should also be decodable at the concrete level (action classification within object exemplar). Likewise, wherever actions can be decoded at the abstract level (action classification across object class) actions should also be decodable at both the concrete (action classification within object exemplar) and the intermediate level (action classification across object exemplar). We therefore entered statistical maps for the intermediate and abstract levels into a conjunction analysis: For the intermediate level, a conjunction of the maps for concrete and intermediate level was computed. For the abstract level, a conjunction of the maps for concrete, intermediate, and abstract level was computed. Conjunctions were computed by outputting the minimum t value for each vertex of the input maps (Nichols et al., 2005). Finally, maps were corrected for multiple comparisons at $p = 0.05$ at the cluster level, using a cluster size algorithm (Forman et al., 1995) based on Monte Carlo simulations (1000 iterations) as implemented in BrainVoyager 2.4. An initial voxelwise threshold of $p < 0.005$ and an estimate of the spatial correlation of voxels of the statistical maps were used as input in the simulations.

RESULTS

Behavioral results. All participants identified catch trials with high accuracy. Mean error rates were 5.0 ± 0.8%, (SEM). Reaction times for correct responses (measured with respect to video onset) were 1953 ± 25 ms (SEM).

Univariate fMRI results. To determine ROIs for subsequent MVP analysis, we computed a group contrast of all eight conditions (ACTION x OBJECT CATEGORY x KINEMATICS) vs. baseline (see section ROI Definition in the Methods). This revealed widespread activations within left and right ventral and dorsal premotor cortex, IPS, and occipitotemporal cortex extending dorsally into posterior IPS and ventrally into middle and inferior temporal gyrus. Peak Talairach coordinates identified in the group contrast for the ROI MVPA were: -47/0/27 (left PMv), 53/0/36 (right PMv), -43/-36/39 (left IPL), 35/-35/46 (right IPL), -43/-69/-2 (left LOTC), and 43/-65/1 (right LOTC).

In addition, we computed a univariate contrast “open” vs. “close” (collapsed across object category and kinematics) to test for putative univariate effects. This contrast revealed no significant effects (even after applying very liberal correction thresholds of $p = 0.05$ at the voxel level). The lack of significant differences in the univariate contrast suggests that the activation levels were comparable over the two actions.

ROI MVPA results. In a ROI-based multivoxel pattern analysis, we investigated the degree of generality of action representations (see Methods for details of the procedure) in regions typically associated with action observation, i.e., PMv, IPL, and LOTC (Figure 2).

In IPL and LOTC, we could decode actions at all levels of abstraction, while in PMv, we could only decode actions at the concrete level (FDR corrected alpha = 0.034). A three-way repeated measures ANOVA with the factors Abstraction Level x ROI x Hemisphere revealed main effects of Abstraction Level ($F(2,378) = 10.23$, $p < 0.001$) and ROI ($F(2,378) = 18.93$, $p < 0.001$). No effects of Hemisphere and no interactions were observed (all $p > 0.1$). Post-hoc paired samples *t* tests revealed that decoding accuracies for intermediate and abstract levels were significantly higher in IPL compared to PMv and in LOTC compared to PMv (Tables 1). In addition, accuracies in PMv differed significantly between concrete and abstract levels (Table 2). Together, these results demonstrate that IPL and LOTC, but not PMv, encode actions at abstract levels of representation.

A second observation is that LOTC showed significantly stronger decoding for the concrete compared to intermediate and abstract levels, whereas IPL showed relatively similar decoding accuracies across all levels (Table 2). This indicates that LOTC contains both concrete and more abstract representations whereas IPL contains abstract action representations only suggesting that generalization from perceptual to conceptual action representations takes place in LOTC (see section ROI MVPA in the Methods for a detailed description of expected patterns of results).

Finally, in all regions, decoding accuracies for intermediate and abstract levels were at similar levels and did not show significant differences (Table 2) suggesting that generalization from concrete (object-specific) to abstract (object-category-independent) action representations does not require an additional, intermediate (object-independent but object-category-specific) abstraction step (see section ROI MVPA in the Methods for a detailed description of expected patterns of results).

Searchlight MVPA results. A searchlight analysis corroborated the findings of the ROI MVPA (Figure 3 and 4, Table 3): At the concrete level, we decoded actions in both hemispheres throughout the occipitotemporal cortex, postcentral sulcus (poCS), intraparietal sulcus (IPS), and ventral as well as dorsal premotor cortex. At intermediate and abstract levels, we decoded actions in bilateral posterior middle temporal gyrus (pMTG) / inferior temporal sulcus (pITS) and poCS (at the junction to anterior IPS), but not in areas anterior to the postcentral sulcus.

DISCUSSION

Our results demonstrate that LOTC, but not PMv, encode the actions “open” and “close” at abstract levels of representation, i.e., independently of the concrete objects and object categories involved in the actions and the kinematics required to manipulate these objects. This finding provides evidence that LOTC and IPL contain neural populations that are action-specific and at the same time generalize across perceptually different instantiations of an action and thus fulfill the necessary criteria for action understanding. On the contrary, PMv codes actions at a concrete level only. We found no regions anterior to postcentral gyrus that contain action representations that generalize across involved object exemplars or categories. The presence of abstract action representations in LOTC and the lack of such representations in premotor cortex seriously questions the motor-centric view that premotor and/or inferior prefrontal cortex provides the basis of action understanding (Rizzolatti et al., 2014). Instead, our results provide clear support for cognitive accounts that suggest action understanding to be associated with perceptual functions, similar to object recognition.

Our searchlight analysis at the intermediate and abstract level revealed a cluster in left LOTC that closely overlapped with the region identified in a meta analysis on conceptual action processing using picture compared to verbal stimuli (Watson et al., 2013). This finding raises the question to which degree information decoded at the intermediate and abstract levels can be regarded perceptual vs. conceptual. Based on our design, we can narrow down a few alternatives: (1) Could decoding be driven by low-level perceptual differences between open and close across decoding levels? Our study was designed to target neural populations that are sensitive to the difference between perceptually similar but conceptually dissimilar actions and at the same time generalize across perceptually dissimilar but conceptually similar actions. Using perceptual similarity analysis (see methods), we ensured that perceptual differences between to-be-decoded actions (e.g. open vs. close water bottle) are smaller than perceptual differences between same actions across decoding levels (e.g. open water bottle vs. open wine bottle; open bottle vs. open box). This makes it unlikely that decoding at intermediate and abstract levels was driven by low-level perceptual similarity. In line with this view, only at the concrete level, where low-level visual features were likely to contribute to the decoding between open and close, we found above chance decoding throughout visual cortex, including early visual areas. (2) Could decoding be driven by similarities of action-specific motion patterns for open and close across decoding levels? Because different kinematics were required for open and close at the intermediate and abstract level (screw vs. push/pull), we can rule out that decoding at these two levels relied on fine-grained motion patterns of hands and fingers. In addition, movements for open and close were actually mirror-like: Not only is open the exact reverse of close, and vice versa, but also in the initial and end phases of each action hand and arm movements are highly similar (hands towards vs. away from object). Therefore, decoding is unlikely to be based on the coarse-grained movement trajectories of arms and hands. So what is the systematic difference between open and close across object exemplars and categories that could be picked up by our classifier? We consider it likely that decoding at the intermediate and abstract level relied on neural populations that are sensitive to the specific change of an object's state (e.g. in case of closing: from open to closed, but not vice versa) independent of the concrete means of the manipulation. However, we do not know if the generalization capacities of these neuronal populations are limited to (a) manual actions (or comprise also the opening of a trashbin with the foot), (b) containers (or comprise also the opening of a door), (c) transitive actions (or comprise also the opening of the mouth or the eye), or (d) physical actions (or comprise also figurative use of action concepts, e.g. opening a business). Finally, (e) we do not know whether the change of the object's state has to be intentionally induced by an actor or whether the same neural populations would also respond to a door that is opened by the wind. These considerations are certainly very exiting and exemplifies our limited knowledge about the architecture underlying action representations. Notably, however, they are of little relevance for the goal of our study, i.e., a comparison of the relative abstraction capacities of regions involved in action observation.

We found that not only LOTC but also IPL encodes action information at abstract levels of representation. Motor and cognitive theories do not offer opposing predictions regarding the generalization capacities in IPL. Our findings in IPL are therefore not suited to differentiate between the two views. However, there seems to be general agreement that IPL is associated with representing action outcomes, either in the sence of proximal physical end states or more distal long-term goals (Hamilton and Grafton, 2007; Oosterhof et al., 2013; Rizzolatti et al., 2014). Recently, anterior IPL has been shown to encode functional knowledge of how to achieve particular outcomes that generalize across motor and sensory information, e.g., decorate room and dress up (Leshinskaya and Caramazza, 2015). This suggests, in line with our findings, that the notion of IPL encoding concrete action outcomes (in the sense of physical end states in the world) is too narrow and needs to be expanded to non-motoric and non-sensory outcomes and purposes.

Notably, the abstraction from concrete actions (open a specific bottle) to intermediate (open bottle) and abstract (open) represents levels of a *conceptual* action hierarchy, which is qualitatively different from hierarchies that describe different levels from muscle activation to movements, goals, and intentions of one and the same *concrete* action (Csibra, 2007; Hamilton and Grafton, 2007; Kilner et al., 2007). Importantly, previous studies that disentangled levels of the latter hierarchy (e.g. the goal vs. the kinematics of an action) were not designed to identify conceptual action representations because the investigated actions always involved the same objects and therefore an action feature that was perceptually constant for the tested representations (for an exception focusing on the performance of tool-related pantomimes, see Chen et al., 2015). These studies cannot disambiguate if an identified representation would be triggered by a concrete action element (e.g. a specific opened box) or by any instantiation of that action independent of the concrete object. Only the latter case fulfills the necessary criteria for action understanding. This ambiguity might explain why some studies found action goal-specific and kinematic-independent representations in premotor cortex (Majdandzic et al., 2009) or inferior frontal gyrus (Hamilton and Grafton, 2008). In the light of this reasoning and our results, it seems likely that representations in these regions code concrete perceptual action features, like the estimated end state of an action, or possibly also even lower level perceptual and motion differences between concrete instantiations of "open" and "close". This interpretation is in line with the observation that mirror neurons in monkey's premotor cortex are not independent of, but in fact modulated by, low-level features of an observed action (Cook and Bird, 2013).

One may argue that, although PMv does not code abstract actions, simulation of the concrete action in PMv is necessary to activate conceptual action information in LOTC and IPL. However, given that premotor cortex receives visual input only indirectly via the dorsal pathway from LOTC and IPL or via the ventral pathway from LOTC and IFG (Kilner, 2011; Nelissen et al., 2011; Turken and Dronkers, 2011) this option seems unparsimonious because it implies that information is first processed in LOTC and IPL, then sent to PMv to enable a motor simulation of the action, and finally sent back to posterior regions where conceptual action information is activated. A more ecological explanation would be that action understanding is a function of LOTC and IPL and action-specific activation of neurons in PMv rather follows or runs in a parallel to action understanding. In line with this view, Papeo et al. (2014) showed that repetitive TMS applied to the posterior middle temporal gyrus (pMTG) abolished the distinction between action and non-action verbs in the precentral gyrus. The hypothesis that action understanding is not causally related to activation of motor circuits in PMv is further corroborated by the observation that congenital absence of motor representations (Vannuscorps et al., 2013) or damage to premotor or motor cortex following stroke (Negri et al., 2007; Kalenine et al., 2010;

but see Pazzaglia et al., 2008) does not necessarily result in deficits in action understanding.

If premotor cortex is not required for action understanding, what role could it play in action observation? One hypothesis is that observed actions activate associated motor responses (Hickok, 2013). Although this is possible, one might argue that similar motor responses should be expected for the intermediate level, i.e., observing the opening of two different bottles should not be associated with two different responses. A different theory suggests that motor circuits are exploited to simulate and anticipate perceptual consequences of observed actions (Csibra, 2007; Kilner, 2011). This view would be in line with the observation of PMv involvement in anticipatory processing of dynamic stimuli in general (Schubotz, 2007; Press and Cook, 2015) and in generating predictions of action outcomes in particular (Jeannerod, 2006; Csibra, 2007). Our finding that PMv codes concrete, but not abstract, action information supports this view.

REFERENCES

Aguirre GK (2007) Continuous carry-over designs for fMRI. Neuroimage 35:1480-1494.

Anzellotti S, Fairhall SL, Caramazza A (2014) Decoding representations of face identity that are tolerant to rotation. Cereb Cortex 24:1988-1995.

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Ann Stat 29:1165–1188.

Binder JR, Desai RH (2011) The neurobiology of semantic memory. Trends Cogn Sci 15:527-536.

Brainard DH (1997) The Psychophysics Toolbox. Spatial vision 10:433-436.

Chang C-C, Lin C-J (2011) LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2:27.

Chen Q, Garcea FE, Mahon BZ (2015) The Representation of Object-Directed Action and Function Knowledge in the Human Brain. Cereb Cortex.

Cichy RM, Pantazis D, Oliva A (2014) Resolving human object recognition in space and time. Nat Neurosci 17:455-462.

Cook R, Bird G (2013) Do mirror neurons really mirror and do they really code for action goals? Cortex 49:2944-2945.

Csibra G (2007) Action mirroring and action interpretation: An alternative account. In: P. Haggard, Y. Rosetti, & M. Kawato (Eds.), Sensorimotor Foundations of Higher Cognition. Attention and Performance XXII (pp. 435-459). Oxford: Oxford University Press. Oxford: Oxford University Press:435-459.

Fagg AH, Arbib MA (1998) Modeling parietal-premotor interactions in primate control of grasping. Neural Netw 11:1277-1303.

Fairhall SL, Caramazza A (2013) Brain regions that represent amodal conceptual knowledge. J Neurosci 33:10552-10558.

Fischl B, Sereno MI, Tootell RB, Dale AM (1999) High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum Brain Mapp 8:272-284.

Fogassi L, Ferrari PF, Gesierich B, Rozzi S, Chersi F, Rizzolatti G (2005) Parietal lobe: from action organization to intention understanding. Science 308:662-667.

Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 33:636-647.

Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg MD, Turner R (1998) Event-related fMRI: characterizing differential responses. Neuroimage 7:30-40.

Gardumi A, Ivanov D, Hausfeld L, Valente G, Formisano E, Uludağ K (2014) The effect of spatial resolution on decoding accuracy in fMRI multivariate pattern analysis. Poster presented at OHBM 2014, Hamburg, Germany.

Goebel R, Esposito F, Formisano E (2006) Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. Hum Brain Mapp 27:392-401.

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

Hamilton AF, Grafton ST (2007) The motor hierarchy: From kinematics to goals and intentions. In: Haggard P, Rossetti Y, Kawato M, editors Sensorimotor Foundations of Higher Cognition Oxford: Oxford University Press:pp 381–408.

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

Hickok G (2013) Do mirror neurons subserve action understanding? Neuroscience letters 540:56-58.

Jeannerod M (2006) Motor Cognition: What Action Tells the Self. Oxford University Press, New York.

Kalenine S, Buxbaum LJ, Coslett HB (2010) Critical brain regions for action recognition: lesion symptom mapping in left hemisphere stroke. Brain 133:3269-3280.

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

Kilner JM, Friston KJ, Frith CD (2007) Predictive coding: an account of the mirror neuron system. Cogn Process 8:159-166.

Konen CS, Kastner S (2008) Two hierarchically organized neural systems for object information in human visual cortex. Nat Neurosci 11:224-231.

Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. Proc Natl Acad Sci U S A 103:3863-3868.

Leshinskaya A, Caramazza A (2015) Abstract categories of functions in anterior parietal lobe. Neuropsychologia.

Majdandzic 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.

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

Negri GA, Rumiati RI, Zadini A, Ukmar M, Mahon BZ, Caramazza A (2007) What is the role of motor simulation in action and object recognition? Evidence from apraxia. Cognitive neuropsychology 24:795-816.

Nelissen K, Borra E, Gerbella M, Rozzi S, Luppino G, Vanduffel W, Rizzolatti G, Orban GA (2011) Action observation circuits in the macaque monkey cortex. J Neurosci 31:3743-3756.

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

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

Oosterhof NN, Tipper SP, Downing PE (2013) Crossmodal and action-specific: neuroimaging the human mirror neuron system. Trends Cogn Sci 17:311-318.

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 104:1077-1089.

Op de Beeck HP (2010) Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses? Neuroimage 49:1943-1948.

Papeo L, Lingnau A, Agosta S, Pascual-Leone A, Battelli L, Caramazza A (2014) The Origin of Word-related Motor Activity. Cereb Cortex.

Pazzaglia M, Smania N, Corato E, Aglioti SM (2008) Neural underpinnings of gesture discrimination in patients with limb apraxia. J Neurosci 28:3030-3041.

Press C, Cook R (2015) Beyond action-specific simulation: domain-general motor contributions to perception. Trends Cogn Sci.

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

Rizzolatti G, Cattaneo L, Fabbri-Destro M, Rozzi S (2014) Cortical mechanisms underlying the organization of goal-directed actions and mirror neuron-based action understanding. Physiological reviews 94:655-706.

Schubotz RI (2007) Prediction of external events with our motor system: towards a new framework. Trends Cogn Sci 11:211-218.

Schwarzbach J (2011) A simple framework (ASF) for behavioral and neuroimaging experiments based on the psychophysics toolbox for MATLAB. Behavior research methods 43:1194-1201.

Turken AU, Dronkers NF (2011) The neural architecture of the language comprehension network: converging evidence from lesion and connectivity analyses. Frontiers in systems neuroscience 5:1.

Vannuscorps G, Andres M, Pillon A (2013) When does action comprehension need motor involvement? Evidence from upper limb aplasia. Cognitive neuropsychology 30:253-283.

Watson CE, Cardillo ER, Ianni GR, Chatterjee A (2013) Action concepts in the brain: an activation likelihood estimation meta-analysis. J Cogn Neurosci 25:1191-1205.

Zaitsev M, Hennig J, Speck O (2004) Point spread function mapping with parallel imaging techniques and high acceleration factors: fast, robust, and flexible method for echo-planar imaging distortion correction. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 52:1156-1166.

Zeng H, Constable RT (2002) Image distortion correction in EPI: comparison of field mapping with point spread function mapping. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 48:137-146.

FIGURE CAPTIONS

Figure 1: Investigated levels of abstraction (A). The concrete level (red) describes actions based on perceptual stimulus properties like concrete kinematics and object exemplars involved in the action. The intermediate level (green) generalizes across kinematics and object exemplars. The abstract level (blue) generalizes across kinematics and object category. Decoding scheme (B). Different abstraction levels were isolated by training a classifier to discriminate the opening and closing of a specific bottle or box and tested it using actions involving either the same object (concrete), a different object from the same object category (intermediate), or an object from a different object category (abstract; see Methods for details of the procedure). Expected patterns of results for different regions coding actions at concrete but not intermediate and abstract levels (C), at concrete, intermediate and abstract levels (D), and at the abstract level only (E). Dotted line represents decoding accuracy at chance = 50% (for Details, see Methods, Section ROI MVPA).

Figure 2: ROI MVPA results. Mean classification accuracies for decoding at concrete (red), intermediate (green), and abstract (blue) levels. Error bars indicate standard error of mean, asterisks indicate statistical significance (different from 50% = chance, red = FDR corrected for the number of tests).

Figure 3: Mean accuracy maps of the searchlight MVPA at each abstraction level (concrete, intermediate, abstract). Individual accuracy maps were cortex-based aligned, averaged, and projected onto a common group surface (both flat maps and lateral views of inflated hemispheres). Decoding accuracy at chance is 50%. Abbreviations: CS: central sulcus, IFS: inferior frontal sulcus; IPS: intraparietal sulcus, ITS: inferior temporal sulcus, PrCS: precentral sulcus, PoCS: postcentral sulcus, SFS: superior frontal sulcus; STS: superior temporal sulcus.

Figure 4: Statistical maps of the searchlight MVPA. For intermediate and abstract levels, conjunctions (i.e., lowest common t value per vertex) of concrete/intermediate and concrete/intermediate/abstract levels, respectively, were used (see Methods for details). Alignment and projection procedures are the same as in Figure 3. Outlines around clusters indicate clusters surviving cluster size correction (dark red: concrete, dark green: intermediate, dark blue: abstract; thresholded at $p = 0.005$, corrected cluster threshold $p = 0.05$). Abbreviations: CS: central sulcus, IFS: inferior frontal sulcus; IPS: intraparietal sulcus, ITS: inferior temporal sulcus, PrCS: precentral sulcus, PoCS: postcentral sulcus, SFS: superior frontal sulcus; STS: superior temporal sulcus.

Table 1: Results of post hoc paired samples *t* tests between ROIs (mean decoding accuracies collapsed across hemispheres; two-tailed)

	PMv-IPL		PMv-LOTC		IPL-LOTC	
	t(21)	*P*	*t*(21)	*P*	*t*(21)	*P*
Concrete	-1.784	0.088	-4.450	<0.001*	-2.845	0.009*
Intermediate	-2.253	0.035*	-3.507	0.002*	-1.082	0.291
Abstract	-2.440	0.023*	-3.140	0.005*	-0.990	0.333

*Significant *p*-values (FDR corrected for number of tests).

Table 2: Results of post hoc paired samples *t* tests between abstraction levels (mean decoding accuracies collapsed across hemispheres; one-tailed)

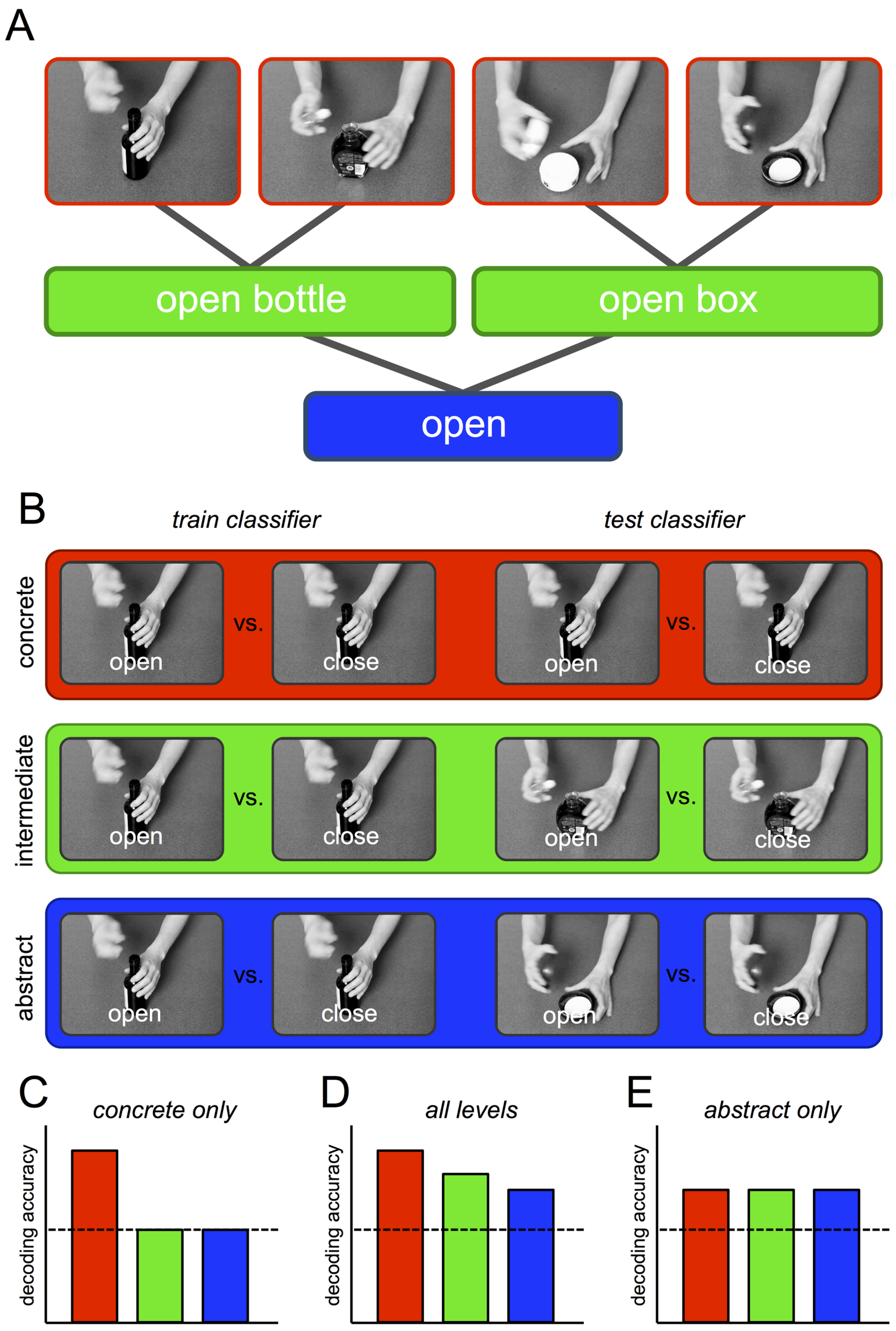
	concrete-intermediate		concrete-abstract		intermediate-abstract	
	t(21)	*P*	*t*(21)	*P*	*t*(21)	*P*
PMv	1.314	0.101	1.962	0.031*	0.599	0.277
IPL	1.392	0.089	1.671	0.054	0.051	0.479
LOTC	3.369	0.001*	3.517	0.001*	-0.012	0.504

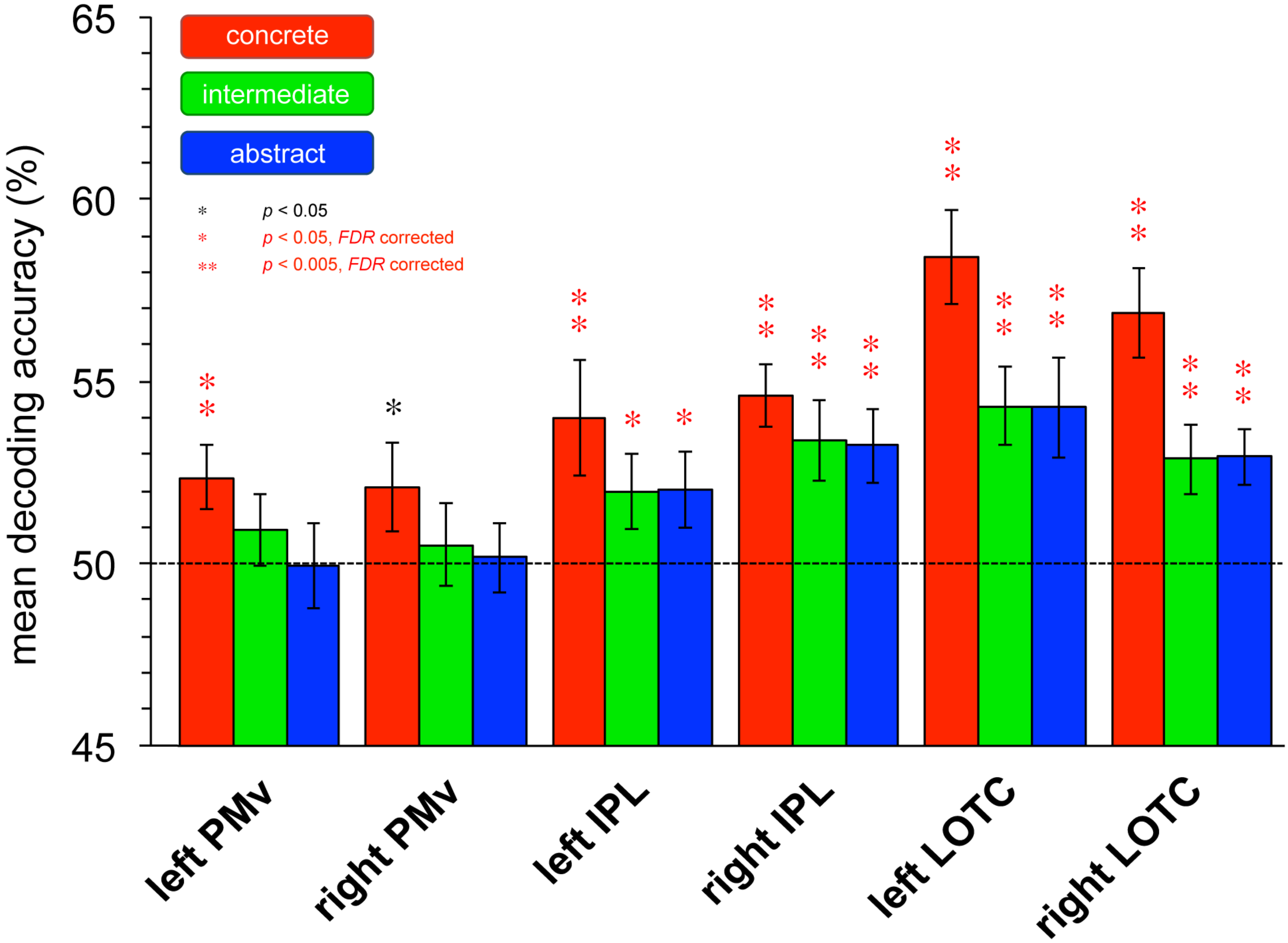
*Significant *p*-values (FDR corrected for number of tests).

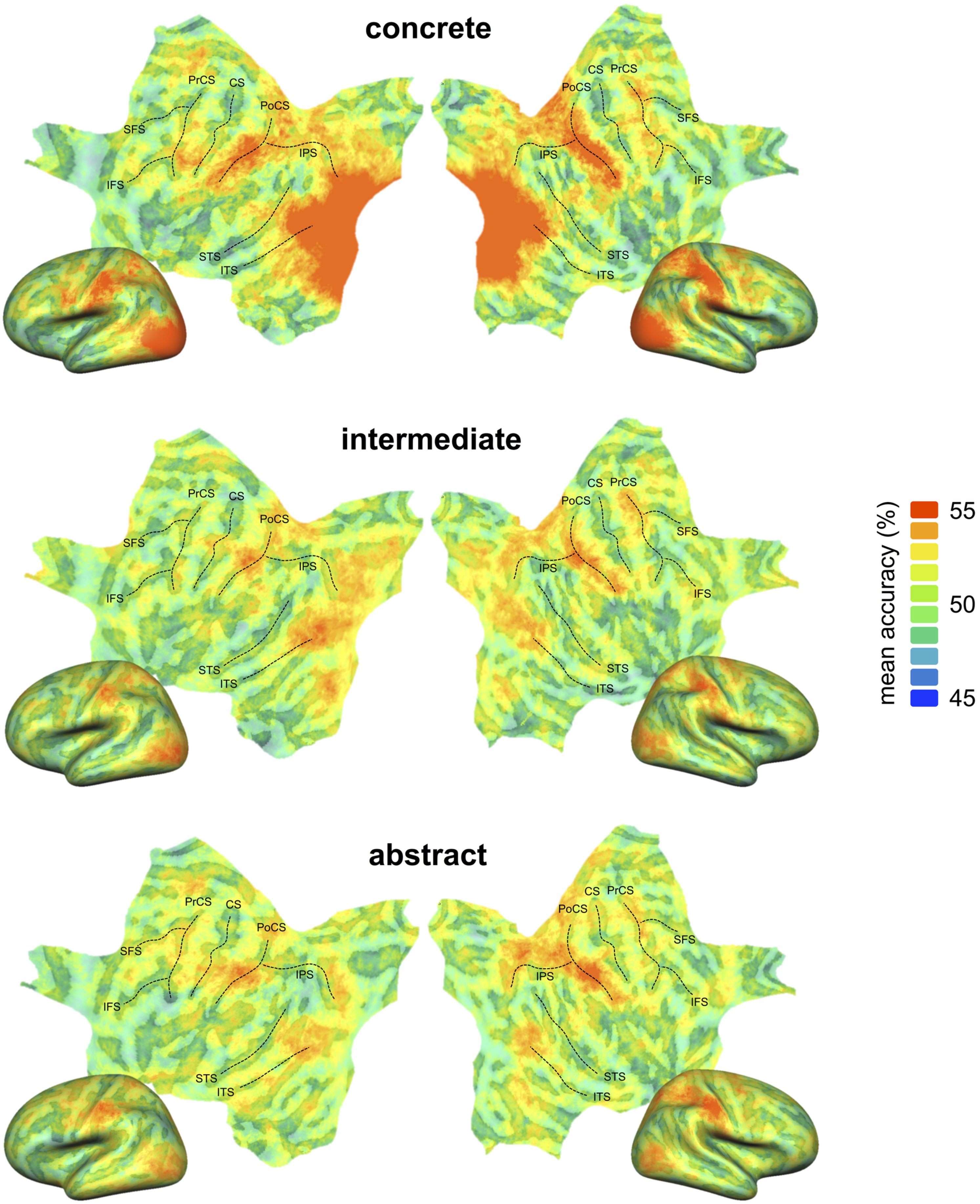
Table 3: Clusters identified in the searchlight MVP analysis for action decoding at concrete, intermediate, and abstract levels

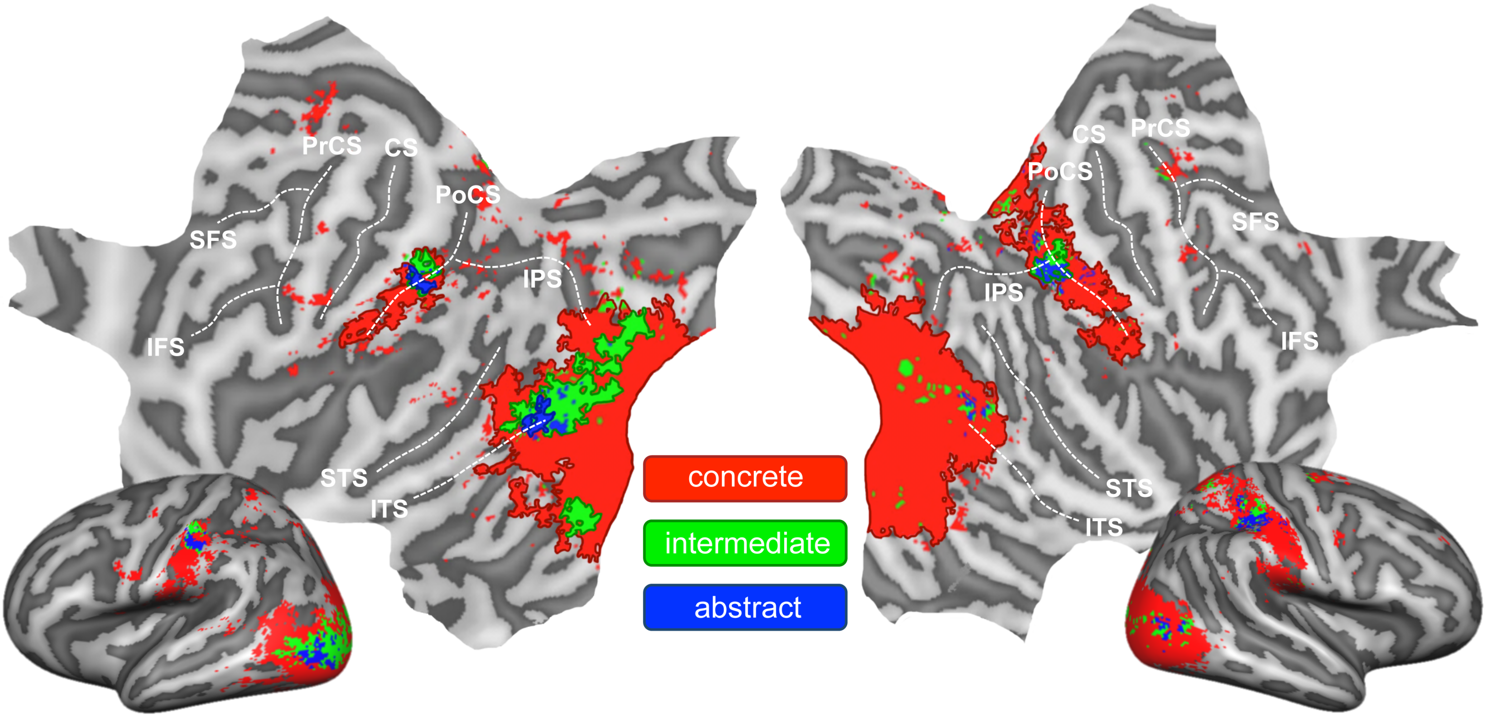
	Cluster				Peak					
Region	*t*	*P*	Accuracy	size	*t*	*p*	Accuracy	x	y	z
Concrete										
left pMTG/LOTC	4.978	0.0007	56.5	25017	12.027	<1.0E-07	61.9	-45	-69	-1
left PoCS/SMG	4.038	0.0013	54.7	2932	6.808	1.0E-06	56.8	-54	-20	30
right pMTG/LOTC	5.256	0.0005	56.6	24802	10.381	<1.0E-07	62.6	43	-71	5
right IPS	4.096	0.0014	55.1	5417	8.167	<1.0E-07	58.3	48	-25	37
right PoCS/SMG	3.929	0.0017	54.5	1791	6.382	3.0E-06	56.3	52	-27	24
right SPL	3.816	0.0021	54.2	1515	5.479	2.0E-05	55.8	27	-50	59
Intermediate										
left pMTG/ITS	4.163	0.0031	53.8	3558	6.946	1.0E-06	55.8	-42	-79	1
left LO	3.607	0.0096	53.0	1040	4.877	8.0E-05	54.2	-18	-91	14
left PoCS	3.617	0.0062	53.9	732	5.132	4.4E-05	55.7	-51	-26	34
left lingual gyrus	3.507	0.0146	53.6	647	4.477	2.1E-04	54.9	-14	-72	-14
right poCS/aIPS	3.595	0.0028	54.3	752	5.012	5.8E-05	56.0	37	-34	43
Abstract										
left pMTG/ITS	3.475	0.0130	53.4	547	4.222	3.8E-04	54.5	-41	-76	-4
left PoCS	3.433	0.0110	54.3	265	4.233	3.7E-04	55.4	-51	-29	36

Size in mm^3. Thresholded at $p = 0.005$, corrected cluster threshold $p = 0.05$. Abbreviations: IPS, intraparietal sulcus; ITS, inferior temporal sulcus; LO, lateral occipital cortex; LOTC, lateral occipitotemporal cortex; pIPS, posterior intraparietal sulcus; pMTG, posterior middle temporal gyrus; PoCS, postcentral sulcus; SMG, supramarginal gyrus; SPL, superior parietal lobe.


A
open bottle
open box
open
B
train classifier
test classifier
concrete
intermediate
abstract
open
vs.
close
C
concrete only
D
all levels
E
abstract only
decoding accuracy


65
60
55
50
45
mean decoding accuracy (%)
concrete
intermediate
abstract
* $p < 0.05$
* $p < 0.05$, *FDR* corrected
** $p < 0.005$, *FDR* corrected
left PMv
right PMv
left IPL
right IPL
left LOTC
right LOTC


concrete
PrCS
CS
PoCS
SFS
IPS
IFS
STS
ITS
CS
PrCS
PoCS
SFS
IPS
IFS
STS
ITS
intermediate
PrCS
CS
PoCS
SFS
IPS
IFS
STS
ITS
CS
PrCS
PoCS
SFS
IPS
IFS
STS
ITS
abstract
PrCS
CS
PoCS
SFS
IPS
IFS
STS
ITS
CS
PrCS
PoCS
SFS
IPS
IFS
STS
ITS
mean accuracy (%)
55
50
45


PrCS
CS
PoCS
SFS
IFS
IPS
STS
ITS
concrete
intermediate
abstract
CS
PrCS
PoCS
SFS
IPS
IFS
STS
ITS