

Sensitivity and specificity of the action observation network to kinematics, target object, and gesture meaning

Francesca Simonelli¹  | Giacomo Handjaras¹  | Francesca Benuzzi²  |
 Giulio Bernardi¹  | Andrea Leo¹  | Davide Duzzi² | Luca Cecchetti¹  |
 Paolo F. Nichelli²  | Carlo A. Porro²  | Pietro Pietrini¹  |
 Emiliano Ricciardi¹  | Fausta Lui² 

¹IMT School for Advanced Studies Lucca, Lucca, Italy

²Department of Biomedical, Metabolic and Neural Sciences and Center for Neuroscience and Neurotechnology, University of Modena and Reggio Emilia, Modena, Italy

Correspondence

Emiliano Ricciardi, IMT School for Advanced Studies Lucca, Piazza San Francesco 19, Lucca, Italy.

Email: emiliano.ricciardi@imtlucca.it

Funding information

European Commission, Grant/Award Number: ECS_00000017; Ministero dell'Università e della Ricerca, Grant/Award Numbers: 2017_55TKFE, 20223K8B3X, P20228PHN2

Abstract

Hierarchical models have been proposed to explain how the brain encodes actions, whereby different areas represent different features, such as gesture kinematics, target object, action goal, and meaning. The visual processing of action-related information is distributed over a well-known network of brain regions spanning separate anatomical areas, attuned to specific stimulus properties, and referred to as action observation network (AON). To determine the brain organization of these features, we measured representational geometries during the observation of a large set of transitive and intransitive gestures in two independent functional magnetic resonance imaging experiments. We provided evidence for a partial dissociation between kinematics, object characteristics, and action meaning in the occipito-parietal, ventro-temporal, and lateral occipito-temporal cortex, respectively. Importantly, most of the AON showed low specificity to all the explored features, and representational spaces sharing similar information content were spread across the cortex without being anatomically adjacent. Overall, our results support the notion that the AON relies on overlapping and distributed coding and may act as a unique representational space instead of mapping features in a modular and segregated manner.

KEYWORDS

action observation network, action representation, *d* prime, fMRI, representational similarity analysis

1 | INTRODUCTION

A fundamental problem in motor neuroscience is understanding how the brain represents distinct objects or motor features and combines

them to form a high-level representation of a finalized act. Several behavioral, neuropsychological, and brain-functional observations support a model of action processing based on a hierarchical organization, in which action features are processed over multiple consecutive stages and information is transferred and integrated from one level to the other (Filimon, 2010; Gallivan & Culham, 2015; Kilner, 2011;

Emiliano Ricciardi and Fausta Lui contributed equally to this work.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Author(s). *Human Brain Mapping* published by Wiley Periodicals LLC.

Tarhan et al., 2021). Within this framework, action-related information is thought to be represented across different levels of abstraction, with brain areas being differentially activated as a function of information complexity, from kinematics (e.g., effector, trajectory of the arm, grip configuration, the orientation of the hand to the object), to object-related features (i.e., identity and function of the object), to higher-level goal and action outcomes (Grafton & Hamilton, 2007). Toward this perspective of different levels of feature representation, several neuroimaging studies support the existence of a hierarchical organization during both observation and execution (Gallivan et al., 2013; Turella et al., 2020; Urgen et al., 2019; Wurm & Lingnau, 2015).

The visual processing of action-related information is organized over a well-known network of brain regions spanning separate anatomical areas and collectively referred to as action observation network (AON; Culham & Valyear, 2006; Caspers et al., 2010; Kilner, 2011; Filimon et al., 2015; Reynaud et al., 2019; Vannuscorps et al., 2019). The AON relies on distributed representations, where low-level visual features (e.g., biological motion; Giese & Poggio, 2003; Zeki, 2015) are processed in specific patches of the cortex, while more abstract properties (e.g., semantics) are encoded at a larger spatial scale (Giese & Rizzolatti, 2015; Handjaras et al., 2015; Tucciarelli et al., 2019). Consequently, distinct action representations overlap across a wide expanse of the cortex and subserve a topographically organized coding: single voxels encode multiple visual and action-related features and support a high-dimensional representational space (Filimon et al., 2015; Handjaras et al., 2015; Wurm & Caramazza, 2019).

Building upon this functional perspective of action representation, the present study aims to better characterize the complex hierarchy within the AON by studying action categorical organization at different levels of granularity (i.e., the level of detail characterizing action representations and, therefore, the extent to which these representations can be differentiated; see Desmet et al., 2021), ranging from elementary descriptors, such as kinematics, object categories, and their interactions, to higher level information, such as the meaning of gestures. Consistently with previous accounts of distributed representations, we propose a novel functional parcellation of the AON based on representational similarity, where the clustering of voxels is guided by their information content rather than their anatomical proximity. Our approach aims to better characterize the intrinsic dimensionality and organization of the AON by testing whether and to what extent similar categorical information is distributed in multiple functional parcels when the anatomical proximity constraint is deliberately neglected. For this purpose, representational similarity analysis (RSA; Kriegeskorte, Mur, & Bandettini, 2008) and clustering techniques (Maaten & Hinton, 2008) were employed in two fMRI experiments to describe the representational geometries of the AON by examining multiple sets of theoretical categorical models and using the sensitivity index d' (Macmillan & Creelman, 2004) to measure discriminability of each action model. In the first fMRI experiment, we identified functional clusters of voxels encoding observed object-directed actions and investigated their specific tuning to different dimensions characterizing

transitive actions (i.e., kinematic trajectory, animacy, and target object category), which have been previously shown to modulate brain responses during action observation (Kemmerer, 2021). In the second fMRI experiment, we tested the reliability of voxel responses to a different set of transitive gestures. Finally, by looking at the representation of intransitive gestures and quantifying the cortical overlap among action properties, we evaluated the sensitivity and specificity of the AON.

We expected the same brain regions to participate at multiple levels of action representation and the great majority of the information content to be spread across the cortex. This would favor the hypothesis of a unique, high-dimensional space for action representation, rather than a representational system with a modular organization based on segregated action features.

2 | MATERIALS AND METHODS

Two independent fMRI experiments were used to investigate the organization of action representation. Both experimental designs employed visual presentation of action stimuli in the form of video clips: in the first one, distinct clips depicting only transitive actions directed at different categories of objects were presented; in the second experimental design, transitive and intransitive actions were embedded in a continuous visual stimulation. The dimensions describing the stimuli feature space in the first experiment were: the kinematics of transitive actions (i.e., trajectory of the arm movement and dynamic interaction with the target object; see Grafton & Hamilton, 2007), the animacy and category of the target of the action. The features describing the second stimulus set were: the kinematics, the target object of transitive actions, and the presence/absence of meaning of non-object-directed, intransitive gestures.

Representational similarity analysis was conducted in both experiments. Initially, brain voxels responsive to actions were isolated in the first experiment to define brain regions belonging to the AON. Next, cortical voxels were clustered according to their similarity structures, and the resulting representational spaces were compared to multiple categorical models across the two experiments to highlight, in each cortical region, the dominant dimensions driving stimulus representation.

2.1 | Participants

Fourteen healthy right-handed subjects (4 males, mean age \pm SD: 37 ± 6 years) participated in the first experiment. Twenty-five healthy right-handed participants (11 males, mean age \pm SD: 26 ± 4 years) were enrolled in the second experiment. All subjects received a medical examination, including a brain structural MRI scan, to exclude any disorder affecting brain structure or function. All participants gave their written informed consent after the study procedures and potential risks had been explained. The experiments were conducted under protocols approved by the Modena Ethical Committee (Protocol number 134/14).

2.2 | Stimuli and procedure

2.2.1 | Experiment 1

Subjects were presented with movie clips depicting a hand-performed transitive action in a third-person perspective, with only the agent's right forearm and hand visible (Figure S1a). The performed action varied in different categorical features: kinematic-based ("grasping to lift," "pushing away," and "putting down"), animacy dimension of the distal object (10 animate and 10 inanimate exemplars for each kinematic-based dimension), and its category (i.e., actual living animals, body parts, natural objects, such as vegetables and fruits, artificial objects), which comprised a further subdivision within the animacy category. The choice of animate targets was mainly constrained by practicality issues, for example, they had to be small enough to allow for a comfortable grip and, in the case of animals, docile enough. For the inanimate targets, common everyday graspable objects were chosen that varied in shape and would allow for different grip configurations. For a complete list of the target objects used, refer to Table 1. The 60 clips were 3 s long, presented with an inter-stimulus interval of 7 s, and each clip was filmed twice. Thus, participants observed 120 clips, randomly alternating between different action types.

2.2.2 | Experiment 2

Participants were presented with a continuous video depicting an actor, whose arms and trunk only were visible, sitting at a table with everyday artificial objects placed in front of him; every 17.5 s, the actor performed a movement pertaining to one of the following four classes with the upper limb: transitive, either grasping an object or touching an object, and intransitive, either symbolic or nonsense (Figure S1b). For a list of target objects and symbolic gestures, please refer to Table 2.

Participants were instructed to observe and pay attention to the video; catch trials, lasting 20 s each, were interspersed among

experimental ones to ensure compliance with the task instructions. During these trials, a static image of the same actor in the same environment was presented, and a red cross appeared on the screen as a cue for the subjects to press a button. Participants completed a total of 81 trials divided into three runs; in each run, six examples of each of the four action categories were presented (24 experimental trials per run plus catch trials; 3 in the first run, 2 in the second, and 4 in the third). No stimulus was repeated within each session.

2.3 | Image acquisition and preprocessing

In the first experiment, functional brain images were acquired during stimulus presentation using a 3 T Philips Intera scanner (gradient echo echo-planar images, 35 axial slices, TR 2.5 s, TE 35 ms, 3 mm isovoxel) and a six-run slow event-related design. T1-weighted spoiled gradient recall images (TR 9.9 ms, TE 4.6 ms, 170 sagittal slices, 1 mm isovoxel) were obtained for each participant to provide detailed brain anatomy. In the second experiment, functional brain images were acquired using the same scanner (3T Philips Intera; TR: 2.5 s, TE: 34 ms, 25 axial slices, voxel size: 3.8 × 3.8 × 4.6 mm). T1-weighted anatomical images (170 sagittal slices, 1 mm isovoxel) were acquired at the beginning of the session.

In both experiments, visual stimulation was provided through the same equipment, using an in-house developed software via the IFIS-SA fMRI System (Invivo Corp, FL; visual field: 15°, 7", 352 × 288 pixels, 60 Hz).

MRI and fMRI data were preprocessed following the same pipeline for both experiments. The anatomical volume was corrected for field inhomogeneity, skull-stripped using the Advanced Normalization Tools package (ANTs; Avants et al., 2009), co-registered to the functional image, and then aligned to the MNI template (Fonov et al., 2009). Functional data were processed using AFNI (Cox, 1996). Despiking (*3dDespike*) and slice timing correction (*3dTshift*) were applied to each functional volume; the resulting time series

TABLE 1 List of objects used as action targets in Experiment 1.

Animate targets		Inanimate targets	
Animals	Body parts	Natural	Artificial
1. Chick	1. Elbow	1. Apple	1. Cup
2. Fish	2. Foot	2. Banana	2. Drill
3. Lobster	3. Hand	3. Eggplant	3. Computer –Mouse
4. Rabbit		4. Fennel	4. Phone
5. Shrimp		5. Onion	5. Scissors
6. Snail			
7. Turtle			

Note: Twenty different objects were used as stimuli: 10 animate exemplars, which could be further categorized into two categories (living animals and body parts), and 10 inanimate objects, which could be further divided into natural and artificial objects.

TABLE 2 List of stimuli used in Experiment 2.

Transitive actions targets		Symbolic gestures	
1. Cup	10. Eyeglasses	1. Ok	10. Not-okay
2. Penholder	11. Keys	2. Waving	11. Come
3. Scissors	12. Wristwatch	3. More-or-less	12. Stop
4. Paperknife	13. Glass	4. No	13. Horns
5. Plate	14. Bottle	5. Blessing	14. Hello
6. Duct tape	15. Pen	6. Victory sign	15. Cut
7. Highlighter	16. Post-it	7. Hitchhiking	16. Fine
8. Eraser	17. Knife	8. Later	17. Tightening
9. Stapler	18. Spoon	9. Money	18. What?

Note: Eighteen distinct objects were used as targets of the transitive actions, and 36 symbolic and nonsense gestures (18 each) were used as intransitive stimuli. Transitive actions included grasping and touching performed on the same target object.

were motion corrected (*3dvolreg*) and spatially smoothed at full-width-at-half-maximum Gaussian kernel of 6 mm (*3dBlurToFWHM*). Single-subject images were entered into a General Linear Model (GLM, *3dDeconvolve*), with each unique stimulus modeled as a separate regressor and including motion parameters and outlier frames (Power et al., 2012) as noise regressors.

2.4 | Representational similarity analysis

T-score maps from the GLM were spatially normalized to the MNI template and resampled into 2 mm isovoxels. A volume-based searchlight approach (Kriegeskorte et al., 2006) was employed, restricting the analysis to grey matter cortical voxels: for each voxel, data was extracted from a searchlight of a 10 mm radius, normalized and de-noised with principal component analysis (PCA), retaining only components explaining 75% of the variance; median activity patterns were computed across subjects. Activity patterns were then compared to each other using a dissimilarity metric (1. Pearson's *r* correlation coefficient) to obtain a Representational Dissimilarity Matrix (RDM; Kriegeskorte, Mur, & Bandettini, 2008).

Within each searchlight, the ability to discriminate between categorical features was evaluated with sensitivity index d' (Macmillan & Creelman, 2004), which quantifies the separation between the average dissimilarity computed between stimuli of different classes and the average within-class dissimilarity. The index was calculated as:

$$d' = \frac{\overline{d_b} - \overline{d_w}}{\sqrt{\frac{1}{2}(\sigma_b^2 + \sigma_w^2)}},$$

where $\overline{d_b}$ is the average between-category dissimilarity, $\overline{d_w}$ is the average within-category dissimilarity, σ_b^2 is the variance of between-category dissimilarity, and σ_w^2 is the within-category dissimilarity variance. d' is equal to 0 if the average between-category and within-category distances are comparable, and becomes larger as the between-category dissimilarity increases (Figure 1a). We opted for d' as a measure of stimulus discriminability because it is an unbiased measure of effect size, allowing a direct comparison between categorical models with different numerosity of within- and between-category elements (Figure S2 for further details).

In the first dataset, we compared single voxel RDMs to four theoretical categorical models (Figure 1b). The kinematic model distinguished the three action types (“grasping to lift,” “pushing away,” and “putting down”), characterized by different arm trajectories and different manipulation of the target object. Two other models tested the discriminability of object-related features at different levels of granularity: the animacy dimension (animacy model, i.e., animate vs. inanimate), and the category of the action target (category model, i.e., animals, body parts, natural and artificial objects). Lastly, the granularity model was based on the combination of the kinematic and animacy dimensions, and encoded actions as gestures with specific kinematics and acted upon distinct targets. The resulting representational space was characterized by a higher level of differentiation in the action hierarchy.

The second experiment evaluated the following dimensions (Figure 1c): transitivity (i.e., object-directed actions considered as within category), kinematics (i.e., “grasping” and “touching” considered as distinct categories), object identity (i.e., every unique object considered as a category) and meaning of intransitive gestures (i.e., symbolic and nonsense considered as separate categories).

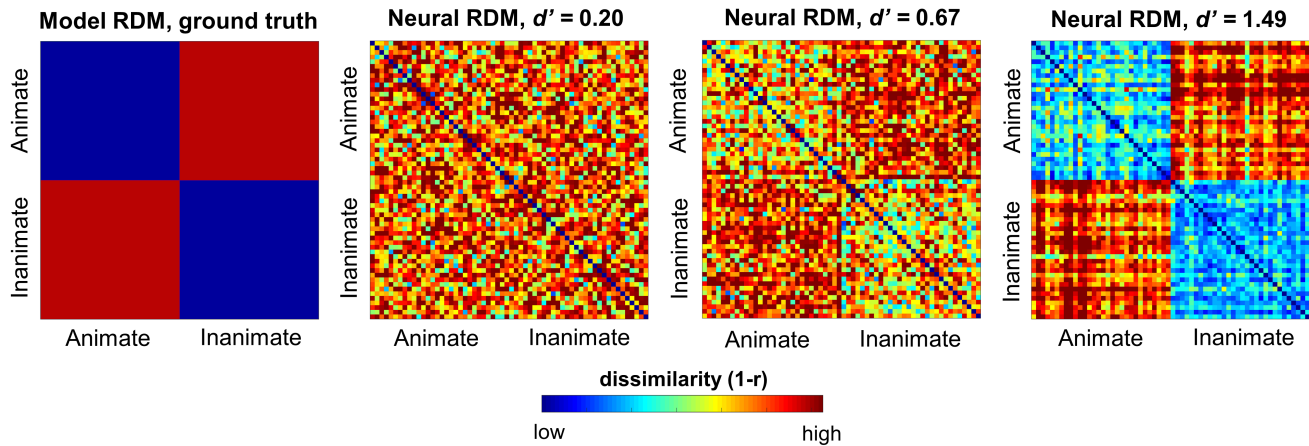
Statistical significance of the obtained d' value was tested through a permutation test: for each voxel, a null distribution of d' values was obtained by shuffling the stimuli labels 1000 times and computing dissimilarity and d' on the shuffled activity patterns. The exact *p*-value for each voxel was computed by approximating the tail of the null distribution to the Pareto distribution (Winkler et al., 2016).

2.5 | Clustering and voxel tuning

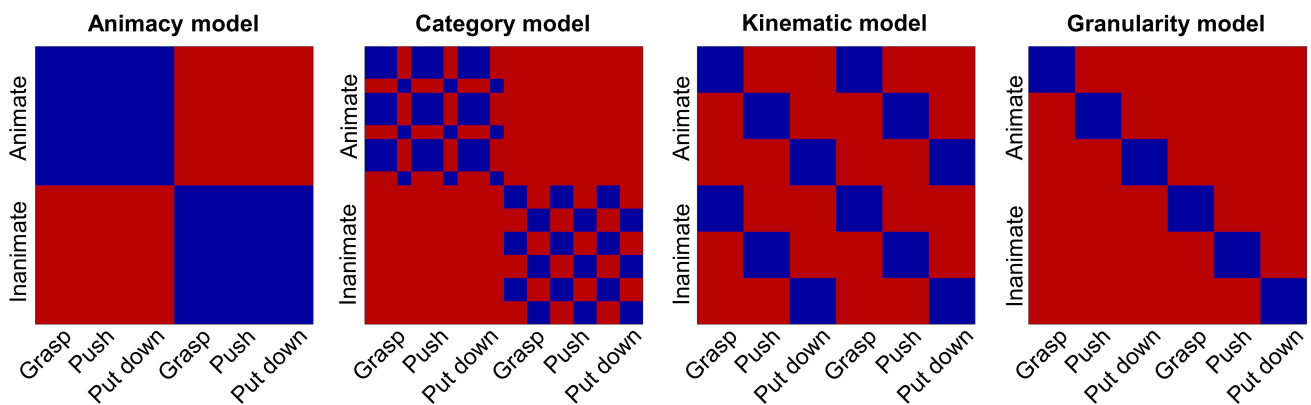
To identify the AON, we selected voxels with statistically significant d' values for the granularity model in the first dataset ($q < 0.01$, using the False Discovery Rate correction under dependency assumptions; Benjamini & Yekutieli, 2001). This procedure identified a set of voxels reliably encoding our stimuli while minimizing the potential number of false positives. T-distributed Stochastic Neighbor Embedding (t-SNE; Maaten & Hinton, 2008) was then applied on the voxels RDMs obtained during the searchlight analysis and projected onto a 2D embedding space so that spatially close points reflect voxels with similar representational geometry. The *k-means* algorithm (Lloyd, 1982) was then used to identify the main clusters based on the spatial distance between voxel projections.

Tuning of selected voxels to different stimulus features was evaluated by color coding *p*-values associated with d' for different models and mapping them onto the space defined by the t-SNE to highlight the relative contribution of each voxel to a specific dimension. First, voxel sensitivity to kinematic, animacy, and object category dimensions was represented by associating each one with a different RGB color channel. Then, data from the second experiment was extracted at the same coordinates of selected voxels, and *p*-values associated with the discriminability of object-directed actions were mapped onto the t-SNE space; the contribution of object identity was also assessed. Furthermore, within the region identified by the granularity model ($q < 0.01$), we assessed the spatial distribution of cortical tunings (Venn diagrams) by applying a threshold to voxels responsive to each categorical model ($p < .05$) and quantifying their overlap as a percentage.

To test the reliability of tunings to animacy, category, and action dimensions in an independent set of voxels, we used data provided by Hardwick et al. (2018), who identified an extended network consistently recruited during action observation, execution, and imagery, across studies. We defined two Regions of Interest (ROIs) using the meta-analytic maps from Hardwick et al. (2018): one coinciding with identified Action Observation areas and one encompassing a broader set of cortex involved in either Observation, Motor Imagery, or Execution (Figure S6a). To isolate significantly responding voxels in a manner analogous to our definition of the AON, we first obtained a

(a) Modulation of d' as a function of simulated noise

(b) Categorical models tested in experiment 1



(c) Categorical models tested in experiment 2

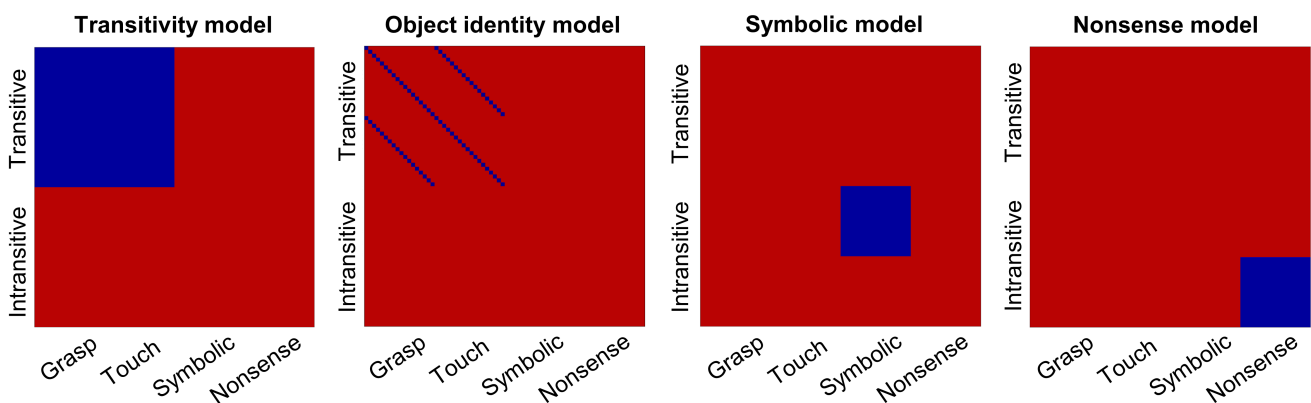


FIGURE 1 Categorical models used to obtain d' values. The sensitivity index d' is computed by comparing the average between-category dissimilarity (in red) with the average within-category dissimilarity (in blue). If a voxel is attuned to a categorical dimension, the average dissimilarity across blue squares should be lower than the average dissimilarity across red squares, resulting in a higher d' value. (a) Examples of d' computed by comparing simulated data with increasing noise levels (from left to right) to the animacy model; d' increases as a function of between-category dissimilarity. (b) Models tested in the first experiment: the animacy model assessed voxel tuning to animacy of the target object; the kinematic model tested voxel tuning to movement type, independently of the action target; the category model tested voxel tuning to target semantic domain; the granularity model is built on the specific combination of movement type and animacy dimension of the target, describing a higher level of specificity in the action hierarchy. (c) Models tested in the second experiment: the transitivity model assessed voxel tuning to transitive actions using a different set of stimuli to demonstrate the reliability of voxel response; the object identity model assessed generalizability of voxel tuning to object properties and the level of specificity in their representational content; symbolic and nonsense models tested discriminability of intransitive gestures to evaluate the specificity of voxel tuning.

statistical map by performing a one-sample t -test on median GLM t -score maps (each one representing a stimulus) and correcting for multiple comparisons ($q < 0.01$, FDR correction; Benjamini & Yekutieli, 2001). Tuning to different stimulus features in the surviving voxels was then assessed by thresholding p -values associated with d' for the animacy, kinematic, and category models ($p < .05$) and quantifying the number of significantly attuned voxels.

To explore the representational geometry at a larger scale, we selected the clusters with the highest d' values for each model and performed multidimensional scaling (Kruskal, 1964) on the clusters' RDMs. Multidimensional scaling transforms information about the dissimilarity between stimuli into Euclidean distances so that the spatial arrangement of individual data points reflects activity pattern similarity, and allows visualization of the dominant features that drive stimulus representations, without any a priori assumption about the representational organization (Kriegeskorte, Mur, Ruff, et al., 2008).

Lastly, we tested the specificity of selected voxels for transitive actions by evaluating the discriminability of intransitive gestures from the second experiment, that is, by mapping sensitivity to symbolic versus nonsense gestures.

3 | RESULTS

3.1 | Sensitivity of the AON to action features

Brain areas showing sensitivity to both kinematic- and object-related features and therefore encoding actions at a high granularity level

were identified by selecting voxels associated with a statistically significant d' (granularity model; $q < 0.01$, FDR corrected). Significant voxels were found in ventral occipito-temporal cortices (occipito-temporal sulcus -OTS-, fusiform gyrus -FusG-, and lingual gyrus -Ling-), sensory-motor cortices (precentral -PreCS- and postcentral -PostCS- sulci, postcentral gyrus -PostCG-, and parietal operculum -ParOper-), posterior temporal areas (superior temporal sulcus -pSTS- and middle temporal gyrus -pMTG-), occipital areas (middle occipital gyrus -MOG-, pericalcarine cortex -PeriCalc-, transverse occipital sulcus -TOS-), the middle cingulate cortex -MidCC-, precuneus -PreCun- and posterior inferior frontal sulcus -pIFS- (Figure 2a). These areas were consistent with previous literature, aligning with the well-known network of regions involved in action observation, namely the AON (Caspers et al., 2010; Hardwick et al., 2018; Kilner, 2011).

Next, a t-SNE dimensionality reduction was applied to perform a functional parcellation that grouped brain regions based on their similarity in the representational space and without considering anatomical constraints. The clustering algorithm identified 11 clusters spanning anatomically distant brain regions (Figure 2b).

To evaluate the tuning of the AON to different stimulus features, the relative contribution of the categorical dimensions was mapped by thresholding ($p < .05$) the voxels responsive to each categorical model and quantifying their overlap (Figure 3a). Of all AON voxels, 45.8%, 89.1%, and 86.2% were significantly attuned to the kinematic, animacy and category dimensions, respectively (Figure 3b), thus suggesting a partial dissociation between the kinematic-based and the object-based dimensions. The inferior temporal and premotor areas retained primarily object-related information, while precentral/

(a) Brain areas significantly tuned to the granularity model

(b) 2D representation of voxels similarity

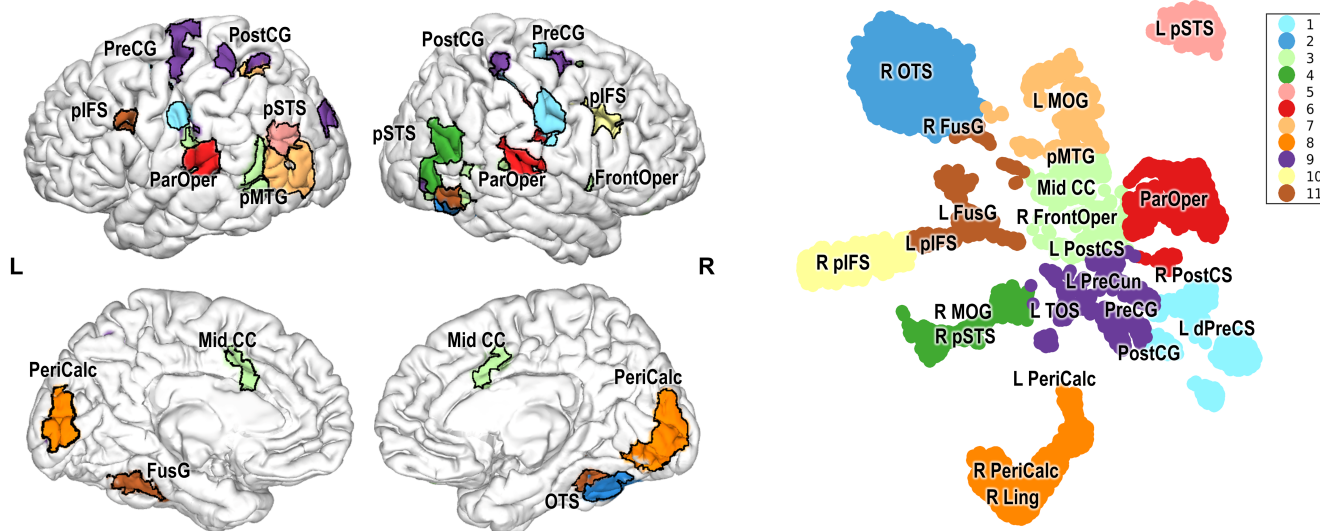
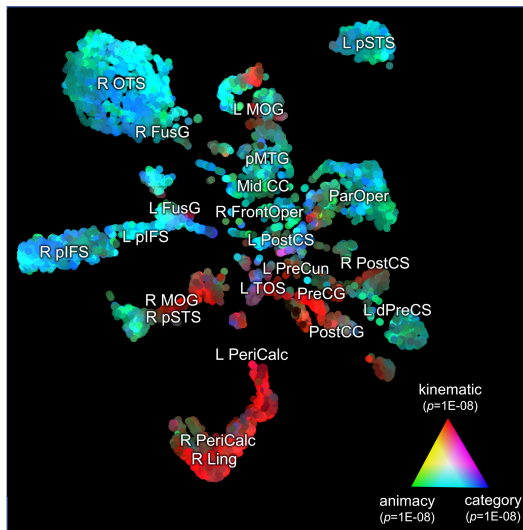


FIGURE 2 (a) Areas showing significant sensitivity to all action features (i.e., granularity model): voxels were selected by computing d' on the RDM ($q < 0.01$, FDR corrected). (b) Selected voxels were projected as points in a 2D embedding space defined on RDM similarities, using t-SNE: the relative distance of voxels/points reflects similarities in the representational space; based on the spatial distance between their projections in the embedding space, voxels were grouped into 11 clusters using the k -means algorithm. L, left; R, right; *FrontOper*, frontal operculum; *OTS*, occipito-temporal sulcus; *dPreCS*, dorsal precentral sulcus; *PostCS*, postcentral sulcus; *FusG*, fusiform gyrus; *Ling*, lingual gyrus; *Mid CC*, middle cingulate cortex; *MOG*, middle occipital gyrus; *ParOper*, parietal operculum; *PeriCalc*, pericalcarine cortex; *pIFS*, posterior inferior frontal sulcus; *pMTG*, posterior middle temporal gyrus; *PostCG*, postcentral gyrus; *PreCG*, precentral gyrus; *PreCun*, precuneus; *pSTS*, posterior superior temporal sulcus; *TOS*, transverse occipital sulcus.

(a) 2D representation of voxel tuning



(b) Shared tuning

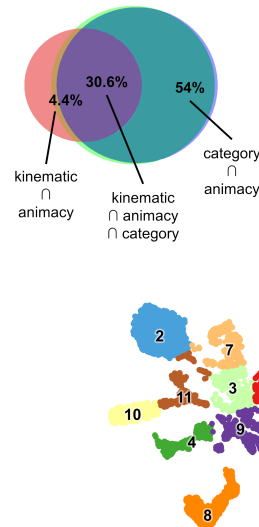
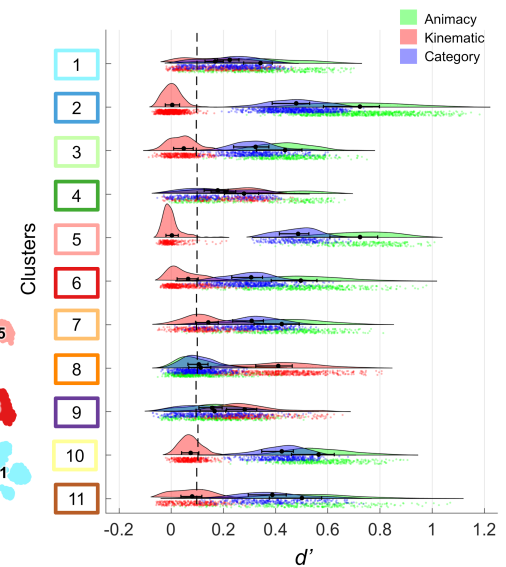
(c) Distributions of d' values

FIGURE 3 (a) Dimensions describing the first set of stimuli were mapped using the p -value associated with d' onto the space defined by the t-SNE; the kinematic dimension was coded in red, the animacy in green, and the object category in blue; maximal saturation of each channel reflects an uncorrected $p < 1E-08$. (b) The Venn diagram represents the proportion of voxels with a statistically significant d' ($p < .05$) for the kinematic, animacy, and category dimensions. (c) d' distributions and clusters averages (black dot) for the animacy, kinematic, and category dimensions; bootstrap confidence intervals (95%) were computed for each d' value and averaged across voxels (solid black line) within a cluster; the average critical value at $\alpha = 0.05$ (dotted line) was obtained from the null distribution.

postcentral and temporo-parieto-occipital areas exhibited higher sensitivity to kinematic features.

Notably, 30% of the AON voxels were significantly attuned to all three dimensions (Figure 3b), indicating the existence of overlapping representations of action features. Moreover, 4.4% of the voxels showed sensitivity to both kinematic and animacy dimensions, but not to object category, 54.0% were attuned to both animacy and object category and not to kinematic features, while 9.3% to kinematic features only. Considerable overlap for the category and animacy dimensions was found in most clusters (Figure 3c), with regions in the temporal cortex showing sensitivity to both features. To assess whether and to what extent similarities between the object-based and the kinematic models could account for the overlap measured at the brain level, we evaluated the collinearities between them. We found that kinematic- and object-related features are substantially independent (Figure S3) and that the specificity of voxel responses to each model was unlikely to be driven by their shared variance. However, it is important to note that the extent of shared representation between action features still relies on how object- and kinematic-based models are compared: direct contrasts or any orthogonalization procedure are likely to eliminate part of the shared tuning across the two models, as illustrated in Figure S4. Moreover, to ensure the reliability of these findings, we assessed the impact of the AON identification and confirmed our observations using a broader definition of AON based on an independent ROI (Figure S6).

In summary, our first experiment revealed that a significant portion of the cortex exhibits broad tuning toward both kinematic and

object-related features, with the largest extent of the AON encoding object-related features.

3.2 | Generalizability of the AON to different kinematics, objects, and action categories

To assess whether the identified clusters were consistently attuned to transitive gestures, we further tested the discriminability of transitive actions in the second experiment. Results demonstrated that 73.2% of the significant voxels in experiment 1 also reached statistical significance in experiment 2 ($p < .05$; Figure 4a, left plot), highlighting the generalizability of the effect to a new stimulus set. The majority of the voxels significantly encoded the type of movement (grasping vs. touching), showing tuning to a different feature space characterizing transitive actions. Specifically, frontal areas, such as the precentral gyrus, were primarily attuned to grasping, whereas tuning to touching-only actions was observed in parietal and occipito-temporal regions, such as the middle occipital gyrus, the occipito-temporal sulcus, and the posterior middle temporal gyrus (Figure S5).

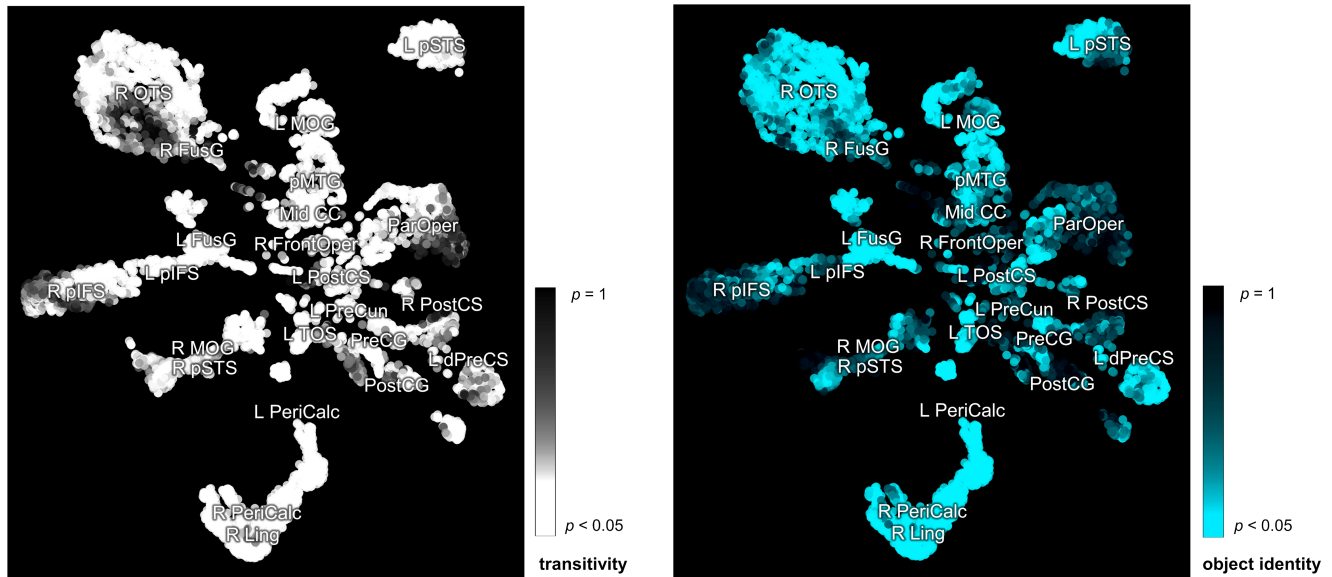
Sensitivity to object identity was also tested using stimuli from the second experiment to confirm the tuning to object-related features. Object identity discriminability was observed in 37.9% of the selected voxels ($p < .05$), indicating high specificity in their representational content (Figure 4a, right plot). Tuning to this feature was found in the left superior temporal cortex and bilaterally in the occipito-temporal cortices, fusiform gyri, and occipital regions,

including the pericalcarine cortex. A total of 34.1% of voxels were significantly attuned to both transitivity and object identity (Figure 4b).

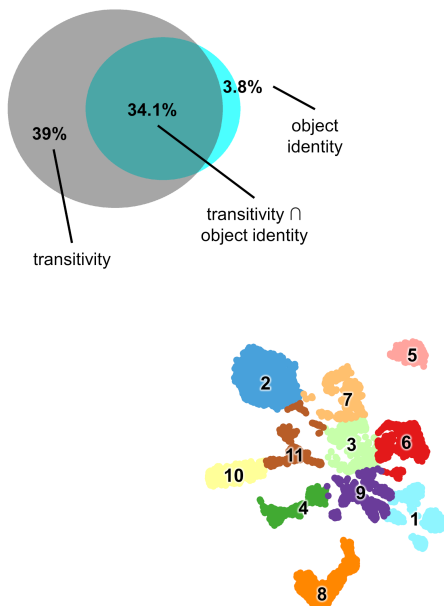
To further characterize the organization of action representations along different dimensions, clusters with the highest sensitivity for object-related features, aside from early visual areas, were selected

(Figure 3c). These clusters included: a set of cortical areas pertaining to the ventral visual stream (cluster 2, comprising the occipito-temporal sulcus and the fusiform gyrus), and the posterior portion of the left superior temporal cortex (cluster 5). We also identified two clusters for the kinematic model: the first encompassing the right

(a) 2D representation of voxel tuning



(b) Shared tuning



(c) Distributions of d' values

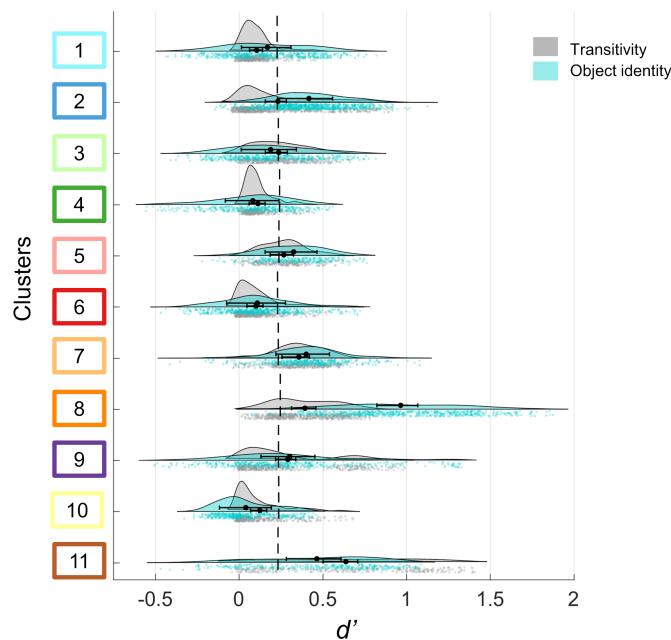


FIGURE 4 (a) Voxels tuning to transitive actions was further tested using an independent dataset, and p -values from the second set were mapped onto the same 2D space. Voxels associated with p -values < 0.05 were mapped in white. Sensitivity to the object-identity feature was also mapped, and voxels associated with p -values < 0.05 were colored in cyan. (b) The Venn diagram represents the proportion of voxels associated with a statistically significant d' ($p < .05$) for transitivity and object identity. (c) d' distributions and clusters averages (black dot) with 95% bootstrap confidence intervals for transitivity and object identity dimensions; dotted lines represent the average critical value at $\alpha = 0.05$ obtained from the null distribution.

middle occipital cortex and superior temporal cortex (cluster 4), and the second including regions of the dorsal visual stream, such as dorso-occipital, superior parietal, and sensorimotor cortex (cluster 9). Multidimensional scaling on these four clusters showed that the same dimension-based dissociation found at the voxel level was also present at the cluster level: stimulus representations were organized primarily around the animacy of the object in occipito-temporal and left posterior superior temporal cortex and around movement type in the middle occipital, superior temporal, dorsal stream, and sensorimotor clusters (Figure 5a). The same effect, although less pronounced, was found when performing multidimensional scaling on data obtained from the second experiment (Figure 5b): representations in clusters 4 and 9 were organized primarily around kinematic-based features, as revealed by the separation in the representational space between grasping and touching, while the same organization was not present in clusters 2 and 5.

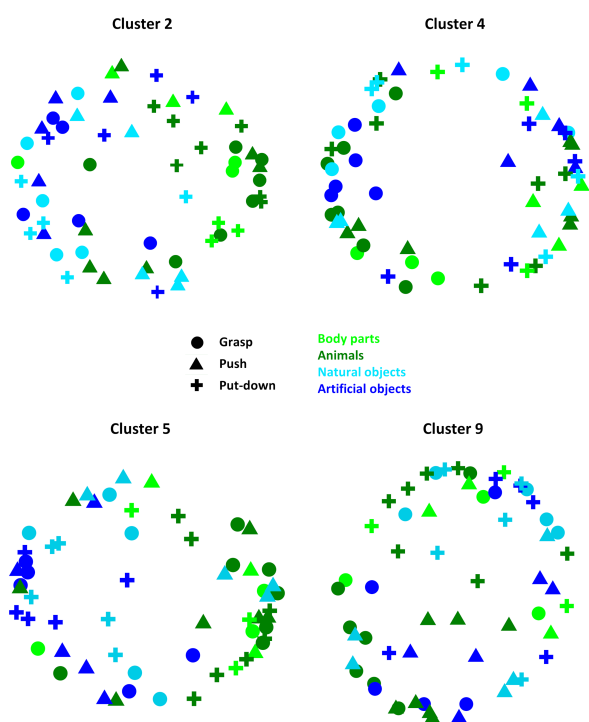
Lastly, we tested the sensitivity of the AON to a different category of actions, a set of intransitive gestures that do not involve objects and subserve symbolic and/or communicative functions. Specifically, we considered two classes of intransitive actions, meaningful symbolic (e.g., waving) and meaningless, nonsense gestures. Information related to intransitive gestures was represented in the AON:

results showed that areas coding transitive actions in the first experiment generalized to novel features, with 82.8% and 87.5% of voxels exhibiting significant tuning to symbolic and nonsense intransitive stimuli ($p < .05$), respectively, and 72.8% of voxels attuned to both dimensions (Figure 6).

4 | DISCUSSION

Using RSA and measuring the sensitivity index d' , we localized cortical areas responding to a high level of differentiation (granularity model), where actions were encoded as gestures with specific kinematics and distinct targets. These AON regions were projected onto a two-dimensional embedding space based on voxel similarity to extract a set of functional clusters representing the various action features. The generalizability of action representation was tested in a second experiment using transitive gestures involving different target objects and kinematics. Finally, voxel specificity to transitivity was evaluated using a set of intransitive gestures modulating their information content. We provided evidence for a partial dissociation between kinematics, object characteristics, and meaning in the occipito-parietal, ventro-temporal, and lateral occipito-temporal cortex (LOTc),

(a) Representational geometries in experiment 1



(b) Representational geometries in experiment 2

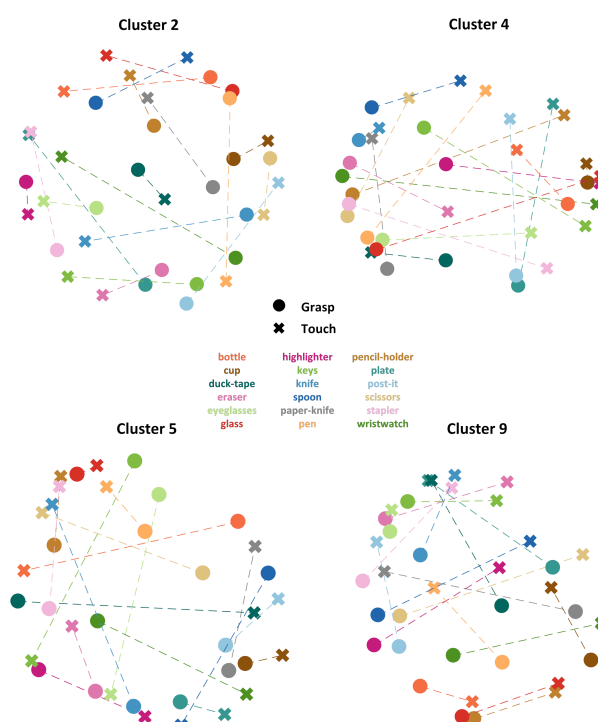


FIGURE 5 Multidimensional scaling for clusters 2 and 5 from the ventral stream, and 4 and 9 for the dorsal stream. Multidimensional scaling was performed on the RDM constructed using data from all the voxels of the cluster. Euclidean distances between markers reflect activity pattern similarity between stimuli. (a) First experiment: action type is coded by symbols, while animacy and object category are color-coded; in clusters 2 and 5, animate and inanimate objects are further apart, reflecting higher d' for the animacy model, while the effect for object category is weaker, but still present; in clusters 4 and 9, stimuli are grouped closer together based on kinematic features, rather than object-related dimensions. (b) Second experiment: action type is coded by symbols, object identity is color-coded, and dashed lines connect markers representing the same object.

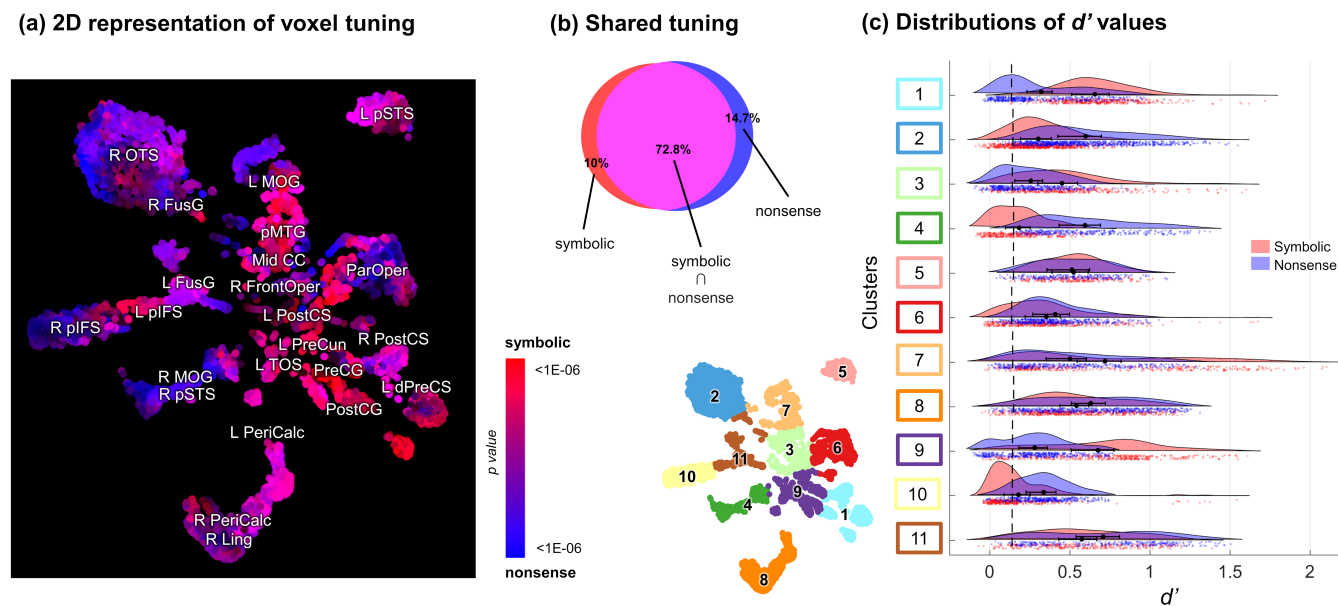


FIGURE 6 (a) The same voxels identified through their tuning to transitive action features were tested for sensitivity to other action dimensions (meaning of intransitive actions), and the relative contribution of voxels' sensitivity from symbolic to nonsense gestures was mapped with a color scale ranging from red to blue; maximal saturation of each channel reflects an uncorrected $p < 1E-06$. (b) The Venn diagram represents the proportion of voxels significantly attuned ($p < .05$) to symbolic and nonsense gestures. (c) d' distributions and clusters averages (black dot) for the symbolic and nonsense features; bootstrap confidence intervals (95%) were computed for each d' value and averaged across voxels (solid black line) within a cluster; the dotted line represents the average critical value at $\alpha = 0.05$ obtained from the null distribution.

respectively. However, the great majority of the AON exhibited low specificity to all the explored action dimensions, and its representational content did not depend on proximity constraints, as anatomically distant cortical patches shared similar tunings.

4.1 | A common representational space for action recognition

In order to explore action representation across the action hierarchy, we first selected brain areas responding to action observation at a high level of granularity, that is, attuned to the specific combination of different action features (e.g., grasping an inanimate target). Then, we assessed voxel tunings to action dimensions and across different sets of transitive and intransitive gestures. To ensure a fair comparison between our models, we relied on the d' measure, which is an estimate of the effect size robust to the intrinsic dimensionality and complexity of the models (Figure S2). We found considerable overlap between the representation of action features in the AON. Specifically, of the voxels identified through the first experiment, 30% were attuned to all three dimensions (i.e., kinematic, animacy, and target category), 34% responded to both transitivity and object identity in the second experiment, and 72% of voxels encoded intransitive gestures, regardless of their meaning. Finally, to ensure that our results did not depend on the definition of the AON, we measured voxel tuning to kinematic- and object-related features in an independent ROI (Figures S6B and S7), corresponding to an extended definition of the AON (Hardwick et al., 2018). Results suggested that, regardless of

being defined according to our granularity model or meta-analytic maps, the majority of the AON encoded multiple features, with object-related properties consistently having a dominant representation compared to kinematic ones.

Here, we purposely neglected anatomical constraints in defining our clusters, which were identified based solely on their representational content. The resulting functional parcellation grouped anatomically distant areas in the same clusters, such as the inferior temporal cortex and the inferior frontal sulcus (cluster 11), which shared related tunings. This functional parcellation, together with the finding of substantial overlap in feature tuning, supports the notion that the organization of the AON relies on overlapping and distributed representations, where, on the one hand, the same brain region participates at multiple levels of the representation, and, on the other hand, the same information is spread across the cortex without being necessarily constrained by anatomical proximity.

The idea of a unified space relying on distributed coding and integration of multiple action features contrasts with the description of the AON as a multiplicity of modules specifically attuned to different features. The gap between these two frameworks is analogous in other fields of cognitive neuroscience, such as the functional organization of the ventral visual stream (Haxby et al., 2001, 2011; Kanwisher, 2010), or language processing (Connolly et al., 2012; Handjaras et al., 2016; Huth et al., 2012, 2016), where the purported functional specificity of brain regions is the subject of intense debate. Action observation entails a cascade of processes ranging from the analysis of spatial frequencies to infer the effector and the object of the action, to the integration over time of biological motion, to the

transcoding of the visual percept into a sensorimotor representation, up to the extraction of the semantic components of the action (Kilner, 2011; Urgen et al., 2019). Although there is a general consensus regarding this hierarchy of action processing, neither lesional nor functional studies have provided clear evidence on the association between specific features or processes and brain regions (Cattaneo et al., 2010; Kemmerer, 2021; Mah et al., 2014; Urgesi et al., 2014). Specifically, impairments in recognition of both transitive and intransitive actions have been reported following lesions to the *IFG* (Pazzaglia et al., 2008) or by applying TMS on the frontal cortex (Ward et al., 2022). Left hemisphere stroke patients with damage to *pMTG* showed lower performance on action recognition tasks based on either semantic or kinematic features (Kalénine et al., 2010). Moreover, performance in goal recognition was affected by stimulation on both the frontal and somatosensory cortices (Jacquet & Avenanti, 2015), while impairments in stroke patients did not seem to be associated with selective brain lesions (Kalénine et al., 2013), suggesting that the coding of action goals may depend on a more distributed network.

While our results support multiple voxel tunings toward coarse high-order descriptions of action features, they do not provide a detailed mapping of the computational characteristics of these models. Each action exemplar differed at a low/mid-level of visual representation, and our models shared different degrees of collinearity with these descriptors (Figure S8). Therefore, the multiple voxel tunings observed here might be influenced by sets of low/mid-level features, which are likely spread and represented differently across the AON. While the decomposition of actions in a space defined either by taxonomical (Kemmerer, 2021) or by computational features (Lahner et al., 2023) is promising, addressing dimensionality, complexity, and collinearity issues in the hierarchical sets of descriptors is still an open challenge (Kriegeskorte & Diedrichsen, 2019). This impedes our comprehension of how various models of action representation interact and are processed within the brain.

In summary, our results are in line with evidence from other cognitive domains, such as vision (Haxby et al., 2001), language (Huth et al., 2016), and the semantics of object categories (Connolly et al., 2012; Huth et al., 2012), and suggest the existence of a unique, high-dimensional space (Haxby et al., 2011) for action observation.

4.2 | The cortical representation of transitive actions

Despite overlapping responses, mapping of single voxel tuning still revealed segregation between movement types and object-related properties, reflecting the classical dichotomy between ventral and dorsal streams (Goodale & Milner, 1992). In particular, we found that inferior temporal areas, that is, right *OTS* and bilateral *FusG*, retained primarily object-related information, consistent with previous literature indicating the involvement of ventro-temporal areas in object processing and encoding of object categories (Grill-Spector & Weiner, 2014; Haxby et al., 2001; Haxby et al., 2011). On the other

hand, dorsal occipital, precentral and postcentral regions (*PreCS*, *PostCS*, and *PostCG*) exhibited higher sensitivity for kinematic-related features: these areas have been shown to process kinematic-based components of movement and trajectory (Cavina-Pratesi et al., 2018; Grafton & Hamilton, 2007; Turella et al., 2020) and are part of the dorsal pathway, which supports visually guided actions (Kravitz et al., 2011). In our work, the dissociation between ventral and dorsal pathways generalizes across experiments, as representational geometries in the ventral stream were organized along animacy in experiment 1 and object-identity in experiment 2, and representations in the dorsal stream were organized primarily around kinematic-based features in both experiments.

Interestingly, we also found a dissociation between left and right posterior temporo-occipital areas: voxels in left *MOG*, *pSTS*, and *pMTG* were mostly attuned to animacy and category, while right *MOG* and *pSTS* showed higher discriminability for kinematic features. These regions are considered part of a third pathway, separate from ventral and dorsal streams (Pitcher & Ungerleider, 2021), and involved in the perception of biological motion (Peuskens et al., 2005; Zeki, 2015). Wurm and Caramazza (2022) proposed a conceptualization of this lateral pathway, and of *LOTc* in particular, as an analogous of the ventral stream, subserving recognition of actions rather than objects. This third pathway is characterized by a posterior-to-anterior gradient processing action-related representations from the more concrete and perceptual features (Han et al., 2013) to the more conceptual semantic aspects that are mainly left-lateralized (Tucciarelli et al., 2019).

Other clusters responding to action observation were identified in bilateral *pIFS*, which showed high sensitivity to object-related properties, indistinguishable from inferior temporal representations. The prefrontal cortex has been suggested to play an important role in spatial- and feature-based attention, in coordination with early visual and inferior temporal areas, to maintain the attentional template on the relevant location or object (Baldauf & Desimone, 2014). Thus, the involvement of *pIFS* may be specifically ascribed to information coding in a feature- or object-based representational format (Bedini & Baldauf, 2021).

4.3 | The specificity of the AON: Intransitive gestures

To test the specificity of the AON areas, identified solely based on their response to features characterizing transitive actions, we assessed sensitivity to a different category of actions, that is, intransitive gestures, which do not involve objects and subserve symbolic and/or communicative functions. Research on non-object-directed actions described stronger contributions of left-lateralized brain regions in the representation of these gestures (Króliczak & Frey, 2009; Kubiak & Króliczak, 2016), with particular involvement of posterior temporal areas (Caspers et al., 2010; Handjaras et al., 2015; Kubiak & Króliczak, 2016; Papeo et al., 2019). Our results showed that intransitive gestures, either symbolic or nonsense, were encoded across the entire AON. In the lateral pathway, we reported a

dissociation between the left and right hemispheres for intransitive gestures: left *pSTS* and *pMTG* primarily encoded symbolic gestures, consistently with reports of semantic processing of actions (Tucciarelli et al., 2019; Wurm & Caramazza, 2019), while right posterior temporal areas showed higher sensitivity to nonsense stimuli, probably reflecting salience of kinematic information. This additional evidence from our second experiment further supports the hypothesis that the representation of actions in the left hemisphere is connected to language (Hodgson et al., 2023; Pulvermüller, 2005), whereas, in the right hemisphere, neural activity during action observation was more strongly associated with visual properties, akin to biological motion (Han et al., 2013; Sokolov et al., 2018).

4.4 | Action representation in early visual areas

In both experiments, we highlighted the role of early visual cortices (i.e., cluster 8) in representing action features (Nishimoto et al., 2011). Specifically, voxels pertaining to the visual cortex were attuned solely to kinematics in the first experiment; in contrast, in the second experiment, they mapped kinematics, objects, and even classes of intransitive gestures. This complete representation may be related to the naturalistic set of visual stimuli, where objects were always located in the lower part of the visual field, and hand trajectories were consistently present in the upper part. This bias was particularly evident in the second experiment, where stimuli were administered to participants as a continuous video, and objects remained at the same location in space both when grasped and when touched, linking the specific identity of that object to its position in the visual field.

A possible ambiguity in the interpretation of our results is the definition of the kinematic model. In accordance with the framework delineated by Grafton and Hamilton (2007), the kinematic model identifies motor acts characterized by different arm trajectories and dynamic interactions with the target objects. In the literature, a similar categorization has also been referred to as “action type” (e.g., Urgen et al., 2019; Zabicki et al., 2017), “movement type” (Tucciarelli et al., 2015) or “action category” (Hafri et al., 2017). Such labels have been used to identify distinct “physical manners of action” (Hafri et al., 2017), characterized by different kinematic patterns (e.g., different hand configurations for pointing vs. grasping; Tucciarelli et al., 2015), or even incorporate higher-level and semantic aspects of the observed acts (Urgen et al., 2019). Therefore, since the “action type” classification is still ill-defined across previous studies and considering that our stimuli have different kinematic properties, we referred to our action model as “kinematic.”

In conclusion, by testing the sensitivity of the AON, we demonstrated a generally low specificity to action properties, with anatomically distant voxels sharing the great majority of their information content. Overall, the present results support the view that the AON should be considered a unique representational space instead of a system with a rigid modular organization based on the segregation of action features.

ACKNOWLEDGMENTS

This work was supported by the PRIN grants (2017_55TKFE and 20223K8B3X, P20228PHN2) by the Italian Ministry of University and Research to Pietro Pietrini and Emiliano Ricciardi and by the “Tuscan Health Ecosystem—THE” Project, Spoke 8, granted by Next Generation EU—National Recovery and Resilience Plan (Piano Nazionale di Ripresa e Resilienza, NRRP)—Mission 4 Component 2 Investment 1.4—Ministry of University and Research (MUR) Call N. 3277, Project Code ECS_00000017 to Emiliano Ricciardi, Pietro Pietrini, and Giacomo Handjaras.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

fMRI data and the Matlab code to replicate these findings are available at <https://osf.io/7ew46/>. Only t-score maps from the GLM analysis, spatially normalized to the MNI template, were shared. Raw structural and functional MRI data are available from the corresponding author upon reasonable request to comply with the European General Data Protection Regulation (GDPR).

ORCID

Francesca Simonelli  <https://orcid.org/0000-0001-7806-2226>

Giacomo Handjaras  <https://orcid.org/0000-0001-5677-8434>

Francesca Benuzzi  <https://orcid.org/0000-0002-3498-3303>

Giulio Bernardi  <https://orcid.org/0000-0003-4045-7075>

Andrea Leo  <https://orcid.org/0000-0002-1948-6608>

Luca Cecchetti  <https://orcid.org/0000-0001-5184-6477>

Paolo F. Nichelli  <https://orcid.org/0000-0001-9756-6796>

Carlo A. Porro  <https://orcid.org/0000-0003-2345-5852>

Pietro Pietrini  <https://orcid.org/0000-0002-6768-5556>

Emiliano Ricciardi  <https://orcid.org/0000-0002-7178-9534>

Fausta Lui  <https://orcid.org/0000-0001-6577-059X>

REFERENCES

- Avants, B. B., Tustison, N., & Song, G. (2009). Advanced normalisation tools (ANTS). *Insight Journal*, 2(365), 1–35.
- Baldauf, D., & Desimone, R. (2014). Neural mechanisms of object-based attention. *Science*, 344(6182), 424–427.
- Bedini, M., & Baldauf, D. (2021). Structure, function and connectivity fingerprints of the frontal eye field versus the inferior frontal junction: A comprehensive comparison. *European Journal of Neuroscience*, 54(4), 5462–5506.
- Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29, 1165–1188.
- Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *NeuroImage*, 50(3), 1148–1167.
- Cattaneo, L., Sandrini, M., & Schwarzbach, J. (2010). State-dependent TMS reveals a hierarchical representation of observed acts in the temporal, parietal, and premotor cortices. *Cerebral Cortex*, 20(9), 2252–2258.
- Cavina-Pratesi, C., Connolly, J. D., Monaco, S., Figley, T. D., Milner, A. D., Schenk, T., & Culham, J. C. (2018). Human neuroimaging reveals the

- subcomponents of grasping, reaching and pointing actions. *Cortex*, 98, 128–148.
- Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y. C., Abdi, H., & Haxby, J. V. (2012). The representation of biological classes in the human brain. *Journal of Neuroscience*, 32(8), 2608–2618.
- Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29(3), 162–173.
- Culham, J. C., & Valyear, K. F. (2006). Human parietal cortex in action. *Current Opinion in Neurobiology*, 16(2), 205–212.
- Desmet, P. M., Sauter, D. A., & Shiota, M. N. (2021). Apples and oranges: Three criteria for positive emotion typologies. *Current Opinion in Behavioral Sciences*, 39, 119–124.
- Filimon, F. (2010). Human cortical control of hand movements: Parietofrontal networks for reaching, grasping, and pointing. *The Neuroscientist*, 16(4), 388–407.
- Filimon, F., Rieth, C. A., Sereno, M. I., & Cottrell, G. W. (2015). Observed, executed, and imagined action representations can be decoded from ventral and dorsal areas. *Cerebral Cortex*, 25(9), 3144–3158.
- Fonov, V. S., Evans, A. C., McKinstry, R. C., Almlí, C. R., & Collins, D. L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, 47, S102.
- Gallivan, J. P., & Culham, J. C. (2015). Neural coding within human brain areas involved in actions. *Current Opinion in Neurobiology*, 33, 141–149.
- Gallivan, J. P., McLean, D. A., Valyear, K. F., & Culham, J. C. (2013). Decoding the neural mechanisms of human tool use. *eLife*, 2, e00425.
- Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*, 4(3), 179–192.
- Giese, M. A., & Rizzolatti, G. (2015). Neural and computational mechanisms of action processing: Interaction between visual and motor representations. *Neuron*, 88(1), 167–180.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1), 20–25.
- Grafton, S. T., & Hamilton, A. F. D. C. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Human Movement Science*, 26(4), 590–616.
- Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorisation. *Nature Reviews Neuroscience*, 15(8), 536–548.
- Hafri, A., Trueswell, J. C., & Epstein, R. A. (2017). Neural representations of observed actions generalize across static and dynamic visual input. *Journal of Neuroscience*, 37(11), 3056–3071.
- Han, Z., Bi, Y., Chen, J., Chen, Q., He, Y., & Caramazza, A. (2013). Distinct regions of right temporal cortex are associated with biological and human-agent motion: Functional magnetic resonance imaging and neuropsychological evidence. *Journal of Neuroscience*, 33(39), 15442–15453.
- Handjaras, G., Bernardi, G., Benuzzi, F., Nichelli, P. F., Pietrini, P., & Ricciardi, E. (2015). A topographical organisation for action representation in the human brain. *Human Brain Mapping*, 36(10), 3832–3844.
- Handjaras, G., Ricciardi, E., Leo, A., Lenci, A., Cecchetti, L., Cosottini, M., Marotta, G., & Pietrini, P. (2016). How concepts are encoded in the human brain: A modality independent, category-based cortical organization of semantic knowledge. *NeuroImage*, 135, 232–242.
- Hardwick, R. M., Caspers, S., Eickhoff, S. B., & Swinnen, S. P. (2018). Neural correlates of action: Comparing meta-analyses of imagery, observation, and execution. *Neuroscience & Biobehavioral Reviews*, 94, 31–44.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539), 2425–2430.
- Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., Hanke, M., & Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, 72(2), 404–416.
- Hodgson, V. J., Lambon Ralph, M. A., & Jackson, R. L. (2023). The cross-domain functional organization of posterior lateral temporal cortex: Insights from ALE meta-analyses of 7 cognitive domains spanning 12,000 participants. *Cerebral Cortex*, 33(8), 4990–5006.
- Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532(7600), 453–458.
- Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*, 76(6), 1210–1224.
- Jacquet, P. O., & Avenanti, A. (2015). Perturbing the action observation network during perception and categorization of actions' goals and grips: State-dependency and virtual lesion TMS effects. *Cerebral Cortex*, 25(3), 598–608.
- Kalénine, S., Buxbaum, L. J., & Coslett, H. B. (2010). Critical brain regions for action recognition: Lesion symptom mapping in left hemisphere stroke. *Brain*, 133(11), 3269–3280.
- Kalénine, S., Shapiro, A. D., & Buxbaum, L. J. (2013). Dissociations of action means and outcome processing in left-hemisphere stroke. *Neuropsychologia*, 51(7), 1224–1233.
- Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences*, 107(25), 11163–11170.
- Kemmerer, D. (2021). What modulates the mirror neuron system during action observation?: Multiple factors involving the action, the actor, the observer, the relationship between actor and observer, and the context. *Progress in Neurobiology*, 205, 102128.
- Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in Cognitive Sciences*, 15(8), 352–357.
- Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience*, 12(4), 217–230.
- Kriegeskorte, N., & Diedrichsen, J. (2019). Peeling the onion of brain representations. *Annual Review of Neuroscience*, 42, 407–432.
- Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences*, 103(10), 3863–3868.
- Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis – Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2, 4.
- Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60(6), 1126–1141.
- Króliczak, G., & Frey, S. H. (2009). A common network in the left cerebral hemisphere represents planning of tool use pantomimes and familiar intransitive gestures at the hand-independent level. *Cerebral Cortex*, 19(10), 2396–2410.
- Kruskal, J. B. (1964). Multidimensional scaling by optimising goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1), 1–27.
- Kubiak, A., & Króliczak, G. (2016). Left extrastriate body area is sensitive to the meaning of symbolic gesture: Evidence from fMRI repetition suppression. *Scientific Reports*, 6(1), 1–13.
- Lahner, B., Dwivedi, K., Iamshchinina, P., Graumann, M., Lascelles, A., Roig, G., Gifford, A., Pan, B., Jin, S., Murty, N. A. R., Kay, K., Oliva, A., & Cichy, R. (2023). BOLD moments: Modeling short visual events through a video fMRI dataset and metadata. *bioRxiv*, 2023–03. <https://doi.org/10.1101/2023.03.12.530887>
- Lloyd, S. (1982). Least squares quantisation in PCM. *IEEE Transactions on Information Theory*, 28(2), 129–137.
- Maaten, L. V. D., & Hinton, G. (2008). Visualising data using t-SNE. *Journal of Machine Learning Research*, 9(Nov), 2579–2605.

- Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology Press.
- Mah, Y. H., Husain, M., Rees, G., & Nachev, P. (2014). Human brain lesion-deficit inference remapped. *Brain*, *137*(9), 2522–2531.
- Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology*, *21*(19), 1641–1646.
- Papeo, L., Agostini, B., & Lingnau, A. (2019). The large-scale organization of gestures and words in the middle temporal gyrus. *Journal of Neuroscience*, *39*(30), 5966–5974.
- Pazzaglia, M., Pizzamiglio, L., Pes, E., & Aglioti, S. M. (2008). The sound of actions in apraxia. *Current Biology*, *18*(22), 1766–1772.
- Peuskens, H., Vanrie, J., Verfaillie, K., & Orban, G. A. (2005). Specificity of regions processing biological motion. *European Journal of Neuroscience*, *21*(10), 2864–2875.
- Pitcher, D., & Ungerleider, L. G. (2021). Evidence for a third visual pathway specialized for social perception. *Trends in Cognitive Sciences*, *25*(2), 100–110.
- Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, *59*(3), 2142–2154.
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*(7), 576–582.
- Reynaud, E., Navarro, J., Lesourd, M., & Osiurak, F. (2019). To watch is to work: A review of neuroimaging data on tool use observation network. *Neuropsychology Review*, *29*(4), 484–497.
- Sokolov, A. A., Zeidman, P., Erb, M., Ryvlin, P., Friston, K. J., & Pavlova, M. A. (2018). Structural and effective brain connectivity underlying biological motion detection. *Proceedings of the National Academy of Sciences*, *115*(51), E12034–E12042.
- Tarhan, L., De Freitas, J., & Konkle, T. (2021). Behavioral and neural representations en route to intuitive action understanding. *Neuropsychologia*, *163*, 108048.
- Tucciarelli, R., Turella, L., Oosterhof, N. N., Weisz, N., & Lingnau, A. (2015). MEG multivariate analysis reveals early abstract action representations in the lateral occipitotemporal cortex. *Journal of Neuroscience*, *35*(49), 16034–16045.
- Tucciarelli, R., Wurm, M., Baccolo, E., & Lingnau, A. (2019). The representational space of observed actions. *eLife*, *8*, e47686.
- Turella, L., Rumiati, R., & Lingnau, A. (2020). Hierarchical action encoding within the human brain. *Cerebral Cortex*, *30*, 2924–2938.
- Urgen, B. A., Pehlivan, S., & Saygin, A. P. (2019). Distinct representations in occipito-temporal, parietal, and premotor cortex during action perception revealed by fMRI and computational modeling. *Neuropsychologia*, *127*, 35–47.
- Urgesi, C., Candidi, M., & Avenanti, A. (2014). Neuroanatomical substrates of action perception and understanding: An anatomic likelihood estimation meta-analysis of lesion-symptom mapping studies in brain injured patients. *Frontiers in Human Neuroscience*, *8*, 344.
- Vannuscorps, G., Wurm, F. M., Striem-Amit, E., & Caramazza, A. (2019). Large-scale organisation of the hand action observation network in individuals born without hands. *Cerebral Cortex*, *29*(8), 3434–3444.
- Ward, E., Brownsett, S. L., McMahon, K. L., Hartwigsen, G., Mascelloni, M., & de Zubicaray, G. I. (2022). Online transcranial magnetic stimulation reveals differential effects of transitivity in left inferior parietal cortex but not premotor cortex during action naming. *Neuropsychologia*, *174*, 108339.
- Winkler, A. M., Ridgway, G. R., Douaud, G., Nichols, T. E., & Smith, S. M. (2016). Faster permutation inference in brain imaging. *NeuroImage*, *141*, 502–516.
- Wurm, M. F., & Caramazza, A. (2019). Distinct roles of temporal and frontoparietal cortex in representing actions across vision and language. *Nature Communications*, *10*(1), 289.
- Wurm, M. F., & Caramazza, A. (2022). Two ‘what’ pathways for action and object recognition. *Trends in Cognitive Sciences*, *26*(2), 103–116.
- Wurm, M. F., & Lingnau, A. (2015). Decoding actions at different levels of abstraction. *Journal of Neuroscience*, *35*(20), 7727–7735.
- Zabicki, A., De Haas, B., Zentgraf, K., Stark, R., Munzert, J., & Krüger, B. (2017). Imagined and executed actions in the human motor system: Testing neural similarity between execution and imagery of actions with a multivariate approach. *Cerebral Cortex*, *27*(9), 4523–4536.
- Zeki, S. (2015). Area V5—A microcosm of the visual brain. *Frontiers in Integrative Neuroscience*, *9*, 21.

SUPPORTING INFORMATION

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

How to cite this article: Simonelli, F., Handjaras, G., Benuzzi, F., Bernardi, G., Leo, A., Duzzi, D., Cecchetti, L., Nichelli, P. F., Porro, C. A., Pietrini, P., Ricciardi, E., & Lui, F. (2024). Sensitivity and specificity of the action observation network to kinematics, target object, and gesture meaning. *Human Brain Mapping*, *45*(11), e26762. <https://doi.org/10.1002/hbm.26762>