The cognitive and neural representations of actions

Inaugural-Dissertation
zur Erlangung der Doktorwürde
der Fakultät für Humanwissenschaften
der Universität Regensburg

vorlegt von
Zuzanna Kabulska
aus Warschau, Polen
2023

Regensburg 2023

Gutachterin (Betreuerin): Prof. Dr. Angelika Lingnau

Gutachter: Prof. Dr. Mark W. Greenlee

PREFACE

The aim of this thesis is to explore how people assign meaning to actions. For that reason, I investigated the cognitive and neural structures underlying action representations, taking into account the key features of actions and the categories that these actions form.

The thesis consists of five chapters. The first chapter provides a general background on the topic of action understanding and the methods I chose for analyzing the data. Chapters 2 - 4 incorporate the three main studies that I conducted throughout the PhD project. Chapter 5 includes a general discussion, the implications of the studies, their limitations, and ideas for future studies.

The article presented in Chapter 2 has been published in Behavior Research Methods (impact factor = 5.95) on July 5th 2022. The manuscript in Chapter 3 has been submitted to Human Brain Mapping (impact factor = 5.40) and is currently in revision. All references have been combined into one bibliography at the end of the thesis. Supplementary files and figures for all three manuscripts have been merged in the Appendix following Chapter 5. No further changes have been made to the text of the published article.

The PhD project has been funded by a Research Grant from the German Research Foundation (Li 2840/1-1). I was supported by the Research Grant from the German Research Foundation (Li 2840/1-1) and the stipend from *Finanzielles Anreizsystem zur Förderung der Gleichstellung*.

ACKNOWLEDGEMENT

Throughout my PhD journey, I was fortunate to be surrounded by many people who offered me both scientific guidance and genuine friendship. Completing this work would not have been possible without the invaluable support from them and I would like to express my deepest gratitude to each and every one of them.

First and foremost, I am profoundly thankful to my supervisor, Prof. Dr. Angelika Lingnau, for continuous guidance, offering invaluable insights for my projects, and providing environment for a constant growth. Dear Angelika, I greatly appreciate your trust, giving me freedom to explore, yet always steering me back on track when my path was not optimal. During the highs and lows of the past years, you always cared for my well-being and encouraged me which was a constant source of strength for me. I am extremely lucky that I had the opportunity to work under your supervision during my PhD, and for that, I am profoundly grateful.

I would also like to thank Prof. Dr. Mark Greenlee who kindly accepted the role of evaluating my PhD work. I sincerely value the time you devoted to reviewing my thesis and offering insightful feedback.

I am also very thankful to my dearest colleagues and fellow researchers, whom I had the pleasure of meeting during my PhD. A special thanks to my lab colleagues for fostering a positive environment in the office and consistently offered their support. Especially to Tonghe Zhuang, my office mate and good friend, as I could always rely on her support, both professionally and personally. Sharing an office with such a compassionate and understanding person who always made time for me, encouraged me and saw potential in me that I often overlooked, was truly a blessing. Equally, I am grateful to Marius Zimmermann,

Oleg Vrabie, Markus Becker, Federica Danaj, Max Reger, and André Bockes for their insightful feedback on my work and fruitful discussion. I would also like to thank Lucca Scheuermeyer for his tremendous help on data collection and analysis for Study 1. Last but not least, I would like to thank Robert Bosek, who consistently offered me support, particularly during the challenging final stages of my thesis – I am grateful for his patience and constant uplifting spirit.

I would also like to mention the funding sources for the course of my PhD: Research Grant from the German Research Foundation (Li 2840/1-1) and the stipend from the *Finanzielles Anreizsystem zur Förderung der Gleichstellung*. This work would not be possible without their financial support.

Finally, my deepest thanks go to my family. To my mom and grandparents, who fostered my love for science from a young age and made me stay always curious. To my father, whose unconventional and unique perspective on the world never ceases to amaze and inspire me. My deepest gratitude also extends to my dear sister, Maja, my best companion through both good and bad times. Without each of you, pursuing my dreams would not have been possible. Thank you for your endless belief in me, your encouragement, and for always being by my side.

TABLE OF CONTENTS

ABBREVIATIONS

AI	Artificial intelligence
aIPS	Anterior intraparietal sulcus
AON	Action observation network
ASF	A simple framework
BOLD	Blood-oxygenation level dependent
CG	Cingulate gyrus
CR-IV	Cerebral cortex IV
D	Dimension
DNN	Deep neural networks
EBA	Extrastriate body area
EEG	Electroencephalography
EPI	Echoplanar imaging
ERP	Event-related brain potential
FA	Flip angle
FDR	False discovery rate
FEAT	FMRI expert analysis tool
FFA	Fusiform face area
FLIRT	FMRIB's linear image registration tool
fMRI	Functional magnetic resonance imaging
FoV	Field of view
FSL	FMRIB's software library
FWHM	Full width at half maximum
GLM	General linear model
Hz	Hertz
IPL	Inferior parietal lobe
IFG	Inferior frontal gyrus
LG	Lingual gyrus
LIBSVM	Library for support vector machines
LOC	Lateral occipital cortex
LOTC	Lateral occipitotemporal cortex
MDS	Multidimensional scaling
MEG	Magnetoencephalography
min	Minutes
mm	Millimetre
MPRAGE	Magnetization prepared – rapid gradient echo
ms	Millisecond
MT	Middle temporal area

MTG	Middle temporal gyrus
MVPA	Multivariate pattern analysis
N	Number
OF	Occipital fusiform gyrus
p	p-value
PC	Principal component
PCA	Principal component analysis
PET	Positron emission tomography
PH	Parahippocampal gyrus
PMv	Ventral premotor cortex
POG	Postcentral gyrus
PPA	Parahippocampal place area
PrCG	Precentral gyrus
PRG	Precentral gyrus
RDM	Representational dissimilarity matrix
RF	Radiofrequency
RFX	Random-effects
ROI	Regions of interest
RSA	Representational similarity analysis
s	Second
SEM	Standard error of the mean
si	Silhouette index
SMG	Supramarginal gyrus
std	Standard deviation
STG	Superior temporal gyrus
STS	Superior temporal sulcus
SVM	Support vector machine
t	t-value
T	Tesla
T1	Longitudinal relaxation time
T2	Transverse relaxation time
TE	Echo time
TFCE	Threshold-free cluster enhancement
TMS	Transcranial magnetic stimulation
TR	Repetition time
UPGMA	Unweighted pair group method with arithmetic mean
VIF	Variance inflation factor
VOTC	Ventral occipitotemporal cortex
z	z-score
τ_A	Tau alpha

CONTRIBUTIONS

Study 1	The cognitive structure underlying the organization of observed actions
Authors	Zuzanna Kabulska[1], and Angelika Lingnau[1]
Authors contribution	Z.K. designed and conducted the experiments, analyzed and interpreted the data, and wrote the first draft of the manuscript. A.L. provided ideas for the study and analysis methods, interpreted the data, and proofread and revised the article.
Study supervision	A.L.
Funding	Open Access funding enabled and organized by Projekt DEAL. This project was funded by a Research Grant from the German Research Foundation (Li 2840/1-1). Angelika Lingnau was funded by a Heisenberg-Professorship (German Research Foundation, Li 2840/2-1).
Acknowledgement	We are thankful to Kate Storrs for providing MATLAB code for comparison of reweighted models, and to Lucca Scheuermeyer for help with collapsing duplicates of features. We also thank Seth Levine, Tonghe Zhuang, Oleg Vrabie, and Andre Bockes for thoughtful comments and feedback on the manuscript.
Journal (Impact Factor)	Behavior Research Methods (5.95)
Publication status	Submitted October 13, 2021; Accepted May 26, 2022; Published July 5, 2022

Study 2	Overlapping but distinct representations of observed actions and action-related features
Authors	Zuzanna Kabulska[1], Tonghe Zhuang[1], and Angelika Lingnau[1]
Authors contribution	Z.K. designed and conducted the experiments, analyzed and interpreted the data, and wrote the first draft of the manuscript. T.Z. helped in interpreting the data and preparing the manuscript. A.L. provided ideas for the study and analysis methods, interpreted the data, and proofread and revised the manuscript.
Study supervision	A.L.
Funding	This project was funded by a Research Grant from the German Research Foundation (Li 2840/1-1). Angelika Lingnau was funded by a Heisenberg-Professorship (German Research Foundation, Li 2840/2-1).

Acknowledgement	We are thankful to Leyla Tarhan for providing MATLAB code for computing the reliability maps and for conducting the winner-takes-all analysis. We also thank Lucca Scheuermeyer for help with MRI data collection, as well as Marius Zimmermann, Oleg Vrabie, Federica Danaj, Max Reger, Andre Bockes, and Robert Bosek for helpful discussions and comments on the manuscript.
Journal (Impact Factor)	Human Brain Mapping (5.40)
Publication status	Submitted August 30, 2023; Revision requested October 5, 2023

Study 3	Neural underpinnings of action categories
Authors	Zuzanna Kabulska[1], and Angelika Lingnau[1]
Authors contribution	Z.K. designed and conducted the experiments, analyzed and interpreted the data, and wrote the first draft of the manuscript. A.L. provided ideas for the study and analysis methods, interpreted the data, and proofread the manuscript.
Study supervision	A.L.
Funding	This project was funded by a Research Grant from the German Research Foundation (Li 2840/1-1). Angelika Lingnau was funded by a Heisenberg-Professorship (German Research Foundation, Li 2840/2-1).
Acknowledgement	We are thankful to Lucca Scheuermeyer for his help in acquiring the data. We also thank Tonghe Zhuang, Marius Zimmermann, Oleg Vrabie, Federica Danaj, Max Reger, and Andre Bockes for thoughtful comments and feedback on the manuscript.
Publication status	The manuscript has not been submitted

[1]Department of Psychology, Faculty of Human Sciences, University of Regensburg, Regensburg, Germany

CHAPTER 1: GENERAL INTRODUCTION

A world full of actions

Imagine walking down a busy street: a group of smiling teenagers talk and gesticulate vigorously; a businessman pushes through the crowd, nervously looking at his watch; a cyclist quickly passes by, forcing us to step to the side. The world around us is a complex and dynamic environment, full of various objects, landscapes, and events that we are constantly processing. As social beings, we are particularly attuned to the actions of others. Understanding others' actions is essential for successful social interaction and communication, as it allows us to predict the behavior of those around us and to respond accordingly. Moreover, our ability to perceive, interpret, and replicate actions is crucial to many of our daily activities, such as learning new skills, doing sports and performing complex tasks. Although understanding actions seems easy and comes effortlessly, the underlying mechanisms are still a matter of debate.

In my research, I conducted several behavioral and neuroimaging experiments to investigate the cognitive structure underlying action organization and its neural underpinnings. The central focus of this project revolves around understanding daily actions by considering both their features and the categories they form at both the cognitive and neural levels. In the Introduction, I provide readers with an overview of studies on action understanding. I delve into the underlying theories of action understanding and investigate specific brain regions believed to play a role in this process. I then identify gaps in the field and discuss how my research addresses them. Next, I provide a background on the core concept of how our minds structure the world around us, namely: Representations. In the last part of the Introduction, I give an overview of the methodological aspects and present two

primary analysis techniques that I employed in this research to investigate the neural structure of actions. Overall, in the Introduction I present the background of my research and highlight the existing gaps that I intend to explore. I then delve into the methods I employed and justify why they are appropriate for addressing the questions in this study.

Research on action understanding

The field of cognitive neuroscience emerged in the latter half of the 20th century, aimed to understand the neural basis of cognitive functions. A key method used was single-cell recordings, which involve measuring electrical activity of individual neurons. One of the pioneering studies on the early visual perception focused on the primary visual cortex of cats and monkeys (Hubel & Wiesel, 1959, 1968). In their experiments, they presented simple visual stimuli to the animals, like black bars with different orientations. By recording the electrical activity of individual neurons in the visual cortex, they discovered a neural specificity linked to different orientations of presented bars and furthermore concluded a hierarchical nature of the visual system (Hubel & Wiesel, 1962).

Single-cell studies provide detailed information about the activity of individual neurons but are not sufficient to understand the whole brain with its complex functions. Understanding higher-order cognitive processes often requires examining patterns of activity involving large numbers of neurons. Additionally, the brain's function comes from complex interactions across billions of neurons forming large brain networks, which is beyond the scope of single-cell studies. A significant leap in cognitive neuroscience came from the development of imaging techniques, such as functional Magnetic Resonance Imaging (fMRI) or Electroencephalography (EEG), which enable non-invasive whole-brain measurement. In the following section, I provide readers a comprehensive overview of the research in the field

of action understanding and show how the development of techniques and analysis methods shifted the focus to previously overlooked brain regions.

Motor theory – mirror neuron

In 1996, an influential study came up reporting the discovery of mirror neurons (see review Rizzolatti & Sinigaglia, 2016). The authors reported a class of neurons that fired both when the monkey executed an action as well as when it observed another individual (a human or another monkey) performing that action (di Pellegrino et al., 1992; Gallese et al., 1996; Rizzolatti et al., 1996). Due to the behavior of mirror neurons, the authors suggested that these neurons serve as a neural basis for action understanding by simulating the action in the observer's motor system. That idea places itself within the motor theory, already suggested for the speech perception (Liberman et al., 1967). The core hypothesis of the motor theory states that several aspects of our cognitive processes are linked to our own motor system, thereby connecting our understanding of them to the movement and gestures involved. The mirror neurons were originally discovered in the monkey's premotor cortex, area F5, however in the subsequent studies they were also found in the inferior parietal lobe (IPL, Rizzolatti et al., 2001; Fogassi et al., 2005). It has been also suggested that the mirror neuron system might exist in humans, constituting a part of the ventral premotor cortex (PMv), IPL, and the posterior part of the inferior frontal gyrus (IFG) (Rizzolatti & Craighero, 2004).

However, motor theory of action understanding has been challenged due to increasing amount of empirical evidence (see review papers: Caramazza et al, 2014; Hickok, 2009; Mahon & Caramazza, 2005). Numerous patient studies provide strong evidence of dissociation between action production and action recognition. Patients with apraxia (Buxbaum et al., 2005; Pazzaglia et al., 2008) and individuals with brain lesions (Halsband

et al., 1997; Pazzaglia et al., 2008) show dissociations between gesture production and recognition tasks, indicating that these abilities can be separate. Similar dissociations are observed in studies on stroke patients involving pantomime recognition, object use, and object recognition tasks (Negri et al., 2007). Thus, the ability to execute correct actions is not essential for recognizing them effectively, and vice versa. Moreover, the mirror system can function independently from action understanding, as humans can understand actions, that they have never performed themselves (Vannuscorps & Caramazza, 2016). An alternative explanation (e.g., Tucciarelli et al, 2015; Wurm & Lingnau, 2015) states that the mirror neurons do not play a causal role in decoding of action goals but are activated as a consequence of action understanding.

Current view: Beyond the motor system

The development of non-invasive whole-brain measurement techniques (fMRI, EEG/MEG, Positron Emission Tomography (PET)) pushed forward the research in cognitive neuroscience, including action understanding. A large meta-analysis revealed a set of regions involved in observation of actions, a so-called Action Observation Network (AON) consisting of frontal (BA 44, 45: Broca's area (Amunts et al., 1999); BA 6: lateral premotor cortex (Geyer, 2004)), parietal (inferior parietal and intraparietal areas), and occipito-temporal (posterior middle temporal gyrus (pMTG) and V5 in the extrastriate cortex) areas in both hemispheres (Caspers et al., 2010). Thus, the brain areas believed to constitute the AON encompass not only the previously identified mirror neuron regions but also the occipitotemporal areas.

While understanding of actions engages a network of occipitotemporal, parietal and frontal regions, their exact function is still debated. Although some researchers reported that

the premotor cortex might represent action meaning and the end-goal (e.g., Nelissen et al., 2005; Majdandić et al., 2009; Rizzolatti & Craighero, 2004; Rizzolatti et al., 2014), there is more and more evidence that this region may not show the level of generality as originally believed (Kilner, 2011; Cook & Bird, 2013) and instead carries information about perceptual properties of actions, such as kinematics (Wurm & Lingnau, 2015) or involved objects (Wurm & Lingnau, 2015; Wurm et al., 2015). Other studies highlighted the role of parietal regions in understanding the actions and their goals. Studies using Transcranial Magnetic Stimulation (TMS) and repetition suppression showed that the inferior parietal lobe processes high-level action understanding as well as goals and intentions of actions, generalizing across effectors (Cattaneo et al., 2010), the kinematic parameters (Hamilton & Grafton 2006, 2007), and trajectory of an action (Hamilton & Grafton, 2008). An fMRI study revealed that different subregions of the parietal cortex host information about different action classes (Abdollahi et al., 2013; Ferri et al., 2015; Corbo & Orban, 2017) and that the inferior parietal lobe transforms visual information about actions into more abstract representations (Urgen et al., 2019).

Recent studies showed that not only the IPL, but also the lateral occipitotemporal cortex (LOTC), represents actions at an abstract level, as both regions generalize across kinematics and used objects (Wurm et al., 2015; Wurm & Lingnau, 2015), viewpoints (Oosterhof et al., 2010, 2012), and stimulus format (Hafri et al., 2017). Numerous studies indicate that these regions have distinct roles, implicating that the parietal regions are involved in planning actions (especially the anterior IPL; Goldenberg & Spatt, 2009; Buxbaum & Kalénine, 2010) as well as inferring the goals and intentions (specifically the posterior IPL; Leshinskaya & Caramazza, 2014), whereas the LOTC represents high-level semantic action knowledge (Oosterhof et al., 2010, 2012, 2013; Wurm & Lingnau, 2015;

Wurm et al., 2015; 2017; Leshinskaya et al., 2020; see reviews Lingnau & Downing, 2015 and Wurm & Caramazza, 2022).

The existing literature indicates that the functions within the LOTC are not uniform. Studies have demonstrated that the ventral and dorsal portions of the LOTC differ in respect to whether the action is object-directed (ventral part) or social (dorsal part) (Isik et al., 2017; Wurm et al., 2017; Wurm & Caramazza, 2022). Additionally, researchers have identified clusters within the LOTC that process different types of information, such as body parts (Downing et al., 2001; Orlov et al., 2010), hands (Bracci et al., 2010; Grosbras et al., 2012), tools (Bracci et al., 2012), faces and limbs (Grosbras et al., 2012; Weiner & Grill-Spector, 2013). Moreover, evidence suggests a gradient of increasing abstraction from the posterior to the anterior parts (Watson et al., 2013; Papeo et al., 2019; Tarhan et al., 2021; for review, see Lingnau & Downing, 2015).

Aim of this study

As presented above, the neural mechanisms underlying action understanding have been extensively studied. While the brain regions involved in action observation have been identified, their exact roles remain debated. I believe that advancements in non-invasive neuroimaging techniques, which enable whole-brain measurements, coupled with computational methods like multivariate pattern analysis, can shed light on these roles.

The aim of this project was multifold. Given the ongoing debate about the role of AON regions, I aimed to investigate how the regions are engaged in processing a big number of naturalistic images of actions and action categories. While many studies emphasize the role of the parietal cortex in action understanding, the potential significance of the LOTC is often overlooked. Moreover, it has been shown that different types of action-related

information are represented in distributed clusters of the LOTC. Yet, most of the studies used features selected by the authors, likely capturing only a subset of potentially relevant action features. I thus adopted a data-driven approach to collect action features and then examined their neural representations.

To address these questions, I carried out a series of behavioral and neuroimaging experiments comprising three main studies:

- **Study 1.** The aim was to explore the cognitive structure underlying the organization of actions, adopting a data-driven approach. The study involved using an inverse multidimensional scaling technique (Kriegeskorte & Mur, 2012) to investigate the category-based organization of actions, and a free-feature listing experiment to obtain key action features. The findings, published in Kabulska & Lingnau (2022), laid the groundwork for the next two studies.
- **Study 2.** The focus of Study 2 centered on exploring the neural representations of the action features identified in Study 1. The aim was to investigate where the crucial action features are represented in the brain and whether specific regions have preference for certain features. The findings have been submitted to *Human Brain Mapping* journal.
- **Study 3.** The goal was to investigate whether different action categories evoke unique activity patterns within the brain, with a particular focus on the AON regions. Additionally, whether these categories exhibit unique connectivity patterns across different brain areas, including the AON as well as category-specific regions. Based on results of Study 1, I selected four action categories: *Communication*, *Grooming*, *Ingestion*, and *Locomotion*.

For the aim of Study 1, I performed several behavioral experiments including inverse multidimensional scaling (Kriegeskorte & Mur, 2012), free-feature listing, and feature-based ratings, which are explained in more details in Chapter 2. For Studies 2 and 3, I conducted two separate fMRI experiments. Since for the big part of neuroimaging data analysis I used multivariate methods, the following section elaborates on the *Representations* which is the background of these techniques.

Representations

One approach to understand how we grasp concepts and differentiate between them is through "Representations". Representations refer to the mental or neural codes that we use to encode and process information about the concepts around us. Let us take three objects as an example: an orange, a banana, and a carrot. The objects belong to different categories, namely *fruits* and *vegetables*. Each object is represented by a variety of features, such as *having a specific color* and a *specific shape* (Figure 1.1A). These features can be represented in a multidimensional space, so-called representational space, where each dimension might reflect a specific feature. Semantically similar objects, here an orange and a banana (fruits), share more semantic features and thus will be located closer to each other in this representational space compared to the other semantically unrelated object (carrot). However, when we consider another feature, like *shape*, bananas and carrots are more similar to each other and therefore will be positioned closer to each other than to an orange. Such multidimensional organizations between concepts can be mapped onto two-dimensional space and form a representation model (Figure 1.1B).

In the external world, objects are represented by sets of features. Inside the brain, however, objects are represented by activations within different neural populations. Since in

my work I investigate the brain using fMRI, I will focus on the level of voxels. To understand how certain concepts, in the case of this work - actions, are represented in the brain, it is beneficial to examine activity patterns across those voxels. In the following section I discuss methods of multivariate pattern analysis - techniques used for brain data analysis, that take into account that the information is encoded in the distributed activation patterns across multiple neurons.

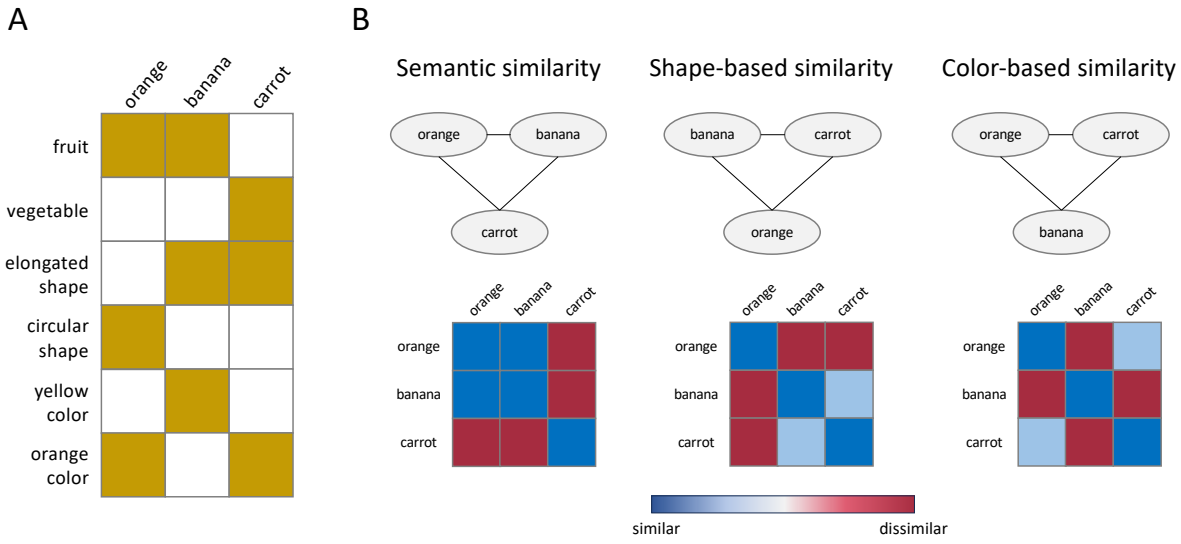


Figure 1.1. Representations. A Every object can be characterized by multiple features and classified into specific categories. For example, three objects - an orange, a banana and a carrot - are represented by features such as an *elongated shape* or a *yellow color,* and they belong to the categories of fruits or vegetables. **B** The degree of similarity between objects can be visualized using distances. The closer the objects are, the more similar they are, while greater distances indicate bigger differences. Depending on the chosen feature, the location of these objects relative to one another can shift. For instance, when comparing objects based on their semantic similarity, an orange and a banana (both fruits) will be located near each other, in contrast to a carrot (a vegetable). However, when considering shape, banana and carrot are more similar to each other (both have an elongated shape) than either is to an orange. These between-object similarities can also be visualized with dissimilarity matrices (representation models).

Multivariate Pattern Analysis (MVPA)

In 2001, Haxby and colleagues (Haxby et al., 2001) discovered that each of the examined stimulus category evoked a distinct brain activity pattern across the voxels. It was the first time when a multivariate pattern analysis was used for fMRI data and was a turning point that influenced other researchers to shift their research practices and adopt this approach. In contrast to univariate analysis, which is used to calculate an average activation level in a brain region, the multivariate pattern analysis relies on fine-grained activity patterns yielded by each experimental condition (Figure 1.2). As the univariate analysis is performed to tell where in the brain the information is represented, the multivariate pattern analysis gives additional details about the nature of those representations. The advantages over the univariate analysis include a greater sensitivity and resolution, as the multivariate pattern analysis takes into account the differences between voxels and the relationship between them (Davis & Poldrack, 2013; Popov et al., 2018). Two main tools employed in the multivariate pattern analysis focus on (1) the classification of conditions (i.e., MVP classification) and (2) the assessment of similarities between the conditions (i.e., representational similarity analysis, RSA) (Figure 1.2) based on the activity patterns. In the following section, I focus on these two crucial methods that I also used for neuroimaging data analysis.

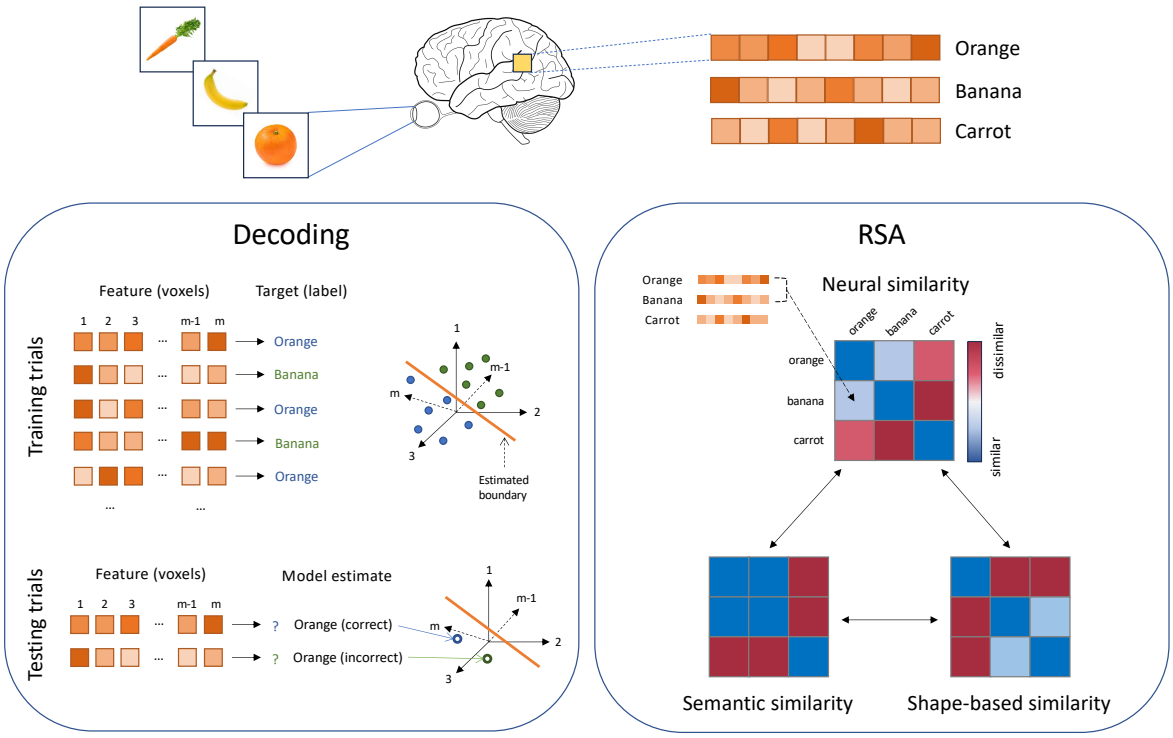


Figure 1.2. Methods of multivariate pattern analysis performed on data from fMRI. Stimuli (here a carrot, banana and orange) evoke neural activities. Each object can then be characterized by a specific pattern of activity within voxels, presented as a vector (see the upper part of the figure). **Decoding (Classification analysis).** Objects can be visualized as data points in a multi-dimensional space, with each dimension represented by a voxel. The dataset is divided into training and testing sets. During the training phase, the classifier learns to differentiate between objects (in this case, two object types) and establishes a decision boundary. In the subsequent testing phase, the classifier predicts the object labels using a dataset it has not seen before, resulting in the accuracy with which each object can be identified. This figure is based on a figure from Weaverdyck et al., 2020. **Representational similarity analysis (RSA)**. The method uses representational dissimilarity matrices (RDMs) to compare data from different sources. The *Neural similarity* RDM is constructed using activity patterns within voxels from the fMRI and then by computing dissimilarities between the object representations. This RDM can be correlated with others, for example, conceptual RDMs representing semantic and shape-based similarities (see Figure 1.1). The images of the vegetable and fruits are adapted from www.istockphoto.com, whereas the image of the brain is adapted from www.wikipedia.com.

Multivariate pattern classification (MVP classification)

The idea behind the multivariate pattern classification is to decode information about stimuli from the brain activity patterns they evoke. These "patterns" refer to spatial

distribution of neural activity across multiple voxels. More specifically, a dataset consists of brain activity patterns obtained in response to, for example, seeing images of oranges, bananas, and carrots. A classification algorithm is then trained to differentiate between these categories based on their activity patterns and later tested on a set of data consisting of patterns elicited by the same stimuli but not seen by the algorithm. Thus, by training the algorithm with activity patterns linked to specific conditions or stimuli, we can classify new patterns and infer the associated conditions or stimuli (see the box *Decoding* in Figure 1.2).

MVP classification employs various classification methods from the field of machine learning, with the Support Vector Machine (SVM) classifier being a popular choice in fMRI studies. This classifier obtains vectors from two categories in the training set, where each vector represents the stimulus in a high-dimensional space. The length of the vector corresponds to the number of dimensions. The objective of the classifier is to identify a decision boundary within that space that effectively separates the two categories. The decision boundary is determined by so-called 'support vectors', which are data points that determine the classification of other points; hence the name of the classifier, as it relies on these support vectors to achieve the optimal separation between different classes. In order to evaluate if the classifier can generalize to new data, researchers use a cross-validation method. It involves iterating the classification over different subsets of the available dataset. As an example, the training dataset can be divided into four parts and each part is to be used in the classification separately. Next, an average predictive performance of a classifier is calculated (Huettel, 2014), with the outcome usually being the accuracy values for each voxel.

In my work, I used the MVP classification with the SVM classifier for the purpose of

Study 3.

Representational similarity analysis (RSA)

In the "Representations" paragraph, I discussed how different objects are represented based on the sets of features. Representational models can be created based on information from different sources, such as previously mentioned cognitive evaluations (shape and color), and fMRI experiments, as well as single-cell recordings from monkey brains, and computational models. Due to the need of comparing information across different sources, investigation of how information is represented in the brain seems like a big challenge. A pivotal breakthrough in cognitive neuroscience came with the introduction of the representational similarity analysis (RSA; Kriegeskorte et al., 2008), as it has provided a means to establish connection between three key research domains: neural, behavioral, and computational. RSA employs a construct called a representational dissimilarity matrix (RDM), which enables to bridge information between these domains (see the box *RSA* in Figure 1.2). RDM is a square matrix consisting of similarities or dissimilarities between pairs of stimuli. Choosing a measure of similarity (e.g., Pearson's r) or dissimilarity (e.g., 1-Pearsons'r) does not have a statistical impact on the results, but using a dissimilarity measure is more commonly used, as it provides a more intuitive explanation of the spatial relation between the stimuli. That means, when representing the stimuli in a representational space, the stimuli that are similar will be located close to each other (small dissimilarity), whereas dissimilar stimuli will be placed apart (Popal et al., 2020).

There are different types of model RDMs and their choice depends on the researchers' needs. One type, called a "conceptual model", explores between-stimuli differences based on selected features of interest (e.g., animate-inanimate objects) (Kriegeskorte et al., 2008). A

"behavioral model", on the other hand, contains dissimilarities between stimuli collected from behavioral experiments, e.g., inverse MDS or ratings. A "computational model" can be created based on, e.g., a layer from a neural network model or a function that mimics the V1 brain regions (Nili et al., 2014). Lastly, a "neural model" contains dissimilarities between activity patterns from either a single region of interest (ROI) or a so-called searchlight sphere (Kriegeskorte et al., 2006). The obtained model RDMs can be compared by calculating correlations between them. For example, correlating a neural RDM from a specific ROI with a conceptual RDM carrying information about social and non-social actions will show how well social information is represented in the investigated brain region (Popal et al., 2020). Overall, RSA involves comparing the (dis)similarity patterns evoked by stimuli through different methods, enabling indirect comparison between the stimuli representations.

In my work, I employed the RSA in the Study 2. The analysis incorporated behavioral, computational and neural models.

CHAPTER 2: STUDY 1

„THE COGNITIVE STRUCTURE UNDERLYING THE ORGANIZATION OF OBSERVED ACTIONS”

This study has been published after peer-review in Behavior Research Methods on July 5th 2022. The online version with the supplementary materials is available at https://doi.org/10.3758/s13428-022-01894-5. The code and the full list of the obtained action features are publicly available at https://osf.io/73v58/.

Abstract

In daily life, we frequently encounter actions performed by other people. Here we aimed to examine the key categories and features underlying the organization of a wide range of actions in three behavioral experiments (N = 378 participants). In Experiment 1, we used a multi-arrangement task of 100 different actions. Inverse multidimensional scaling and hierarchical clustering revealed 11 action categories, including *Locomotion*, *Communication*, and *Aggressive actions*. In Experiment 2, we used a feature-listing paradigm to obtain a wide range of action features that were subsequently reduced to 59 key features and used in a rating study (Experiment 3). A direct comparison of the feature ratings obtained in Experiment 3 between actions belonging to the categories identified in Experiment 1 revealed a number of features that appear to be critical for the distinction between these categories, e.g., the features *Harm* and *Noise* for the category *Aggressive actions*, and the features *Targeting a person* and *Contact with others* for the category *Interaction*. Finally, we found that a part of the category-based organization is explained by a combination of weighted features, whereas a significant proportion of variability remained unexplained, suggesting that there are additional sources of information that contribute to the categorization of observed actions. The characterization of action categories and their associated features serves as an important extension of previous studies examining the cognitive structure of actions. Moreover, our results may serve as the basis for future behavioral, neuroimaging and computational modeling studies.

Introduction

Being able to tell whether we are greeted or attacked by another person is a crucial skill for our survival. What are the key categories underlying the organization of observed actions, and what kind of information do we exploit to quickly categorize and understand actions performed by other people? There is a long tradition in asking this question in the domain of object categories. Aristotle (Aristotle, 1995/350 BCE) argued that categories can be distinguished on the basis of the presence or absence of relevant features (such as a tail or a wing). More recent views emphasize the similarity of weighted features (e.g., Cree & McRae, 2003; Vinson et al., 2003). Some authors pointed out the importance of sensory, functional, motor, and manipulation features (e.g., Binder et al., 2016; Cree & McRae, 2003; McRae et al., 2005; Vigliocco et al., 2004; Vinson & Vigliocco, 2008). According to this view, a cat will be distinguished from other animals by visual features, such as its posture and whiskers, whereas a chair will be distinguished from other non-living objects by functional features, such as that it is something to sit on. Binder et al. (2016) emphasized the role of features with known corresponding neural representations, such as sensory, spatial, and temporal features. In contrast to views that emphasize the role of different types of features, some authors argued that categories differ with respect to the *distribution* and *correlation* of features across different categories (for review, see Mahon & Caramazza, 2009).

A vast number of neuroimaging studies have reported a preference for object categories such as faces, houses, tools, and animals in ventral stream regions (Downing et al., 2001; Kanwisher et al., 1997; Malach et al., 1995). These results are supported by corresponding category-selective deficits in patients (e.g., Humphreys & Rumiati, 1998;

Moscovitch et al., 1997). It has been argued that category selectivity reported in the ventral stream is at least partially due to visual features that systematically differ between object categories (for review, see e.g., Bracci et al., 2017). In line with this view, it has been proposed that object categories are represented by distributed feature maps rather than by functionally specific regions (see e.g., Haxby et al., 2001). Recent neuroimaging studies that directly compared the organization according to features and object categories revealed that features alone are not sufficient to account for the category-based structure (e.g., Jozwik et al., 2016).

To which degree do the principles regarding the cognitive organization of objects according to features and categories described above apply to the organization of observed actions? Vigliocco et al. (2004) reported that both for object and action words, feature-based similarities can predict human similarity judgments. However, as pointed out by Vinson & Vigliocco (2008), objects and actions differ with respect to a number of aspects. Importantly, objects typically can be understood in isolation and often can be identified on the basis of a small number of features that show a strong correspondence with specific categories (e.g., beak and wings would be typical features of a bird). By contrast, many actions can only be understood on the basis of their relation towards objects (e.g., opening a door) or other agents (e.g., hugging someone), and they can be performed in various different ways (e.g., eating with a fork, with chop sticks, or with both hands). Moreover, actions differ with respect to the desired goal state, which can be described along a concrete-abstract continuum, from a change of posture (e.g., getting up) or location (e.g., riding the bike) towards a change of an object configuration (e.g., opening a book) to a change of a mental state (e.g., listening to music; see also Hamilton & Grafton, 2006; Vallacher & Wegner, 1985; Wurm & Lingnau, 2015). Consequently, determining the principles underlying the organization of actions is

challenging and has only recently started to attract a growing level of attention in the literature. As an example, Watson and Buxbaum (2014) revealed two dimensions that underlie the organization of actions involving tools, namely, the amount of arm movement and the hand posture. Using standard univariate analyses of fMRI data, different types of information pertaining to observed actions have been reported to engage different brain regions, such as the ventral premotor cortex in response to the effector type used in an action (e.g., foot, hand, mouth) and the parietal cortex in response to movement direction (Jastorf et al., 2010) and different types of actions (Abdollahi et al., 2013; Ferri et al., 2015). Using multivariate pattern analysis of fMRI data, Wurm et al. (2017) reported a distinction between person-directed and object-directed actions in the lateral occipito-temporal cortex (LOTC). Using a multi-arrangement task, Tucciarelli et al. (2019) identified a number of categories according to which observed actions are organized behaviorally, including locomotion, communicative actions, food-related actions, and cleaning-related actions. Using multiple-regression representational similarity analysis (RSA) of fMRI data, the authors identified a region in the LOTC that reflected this category-based similarity structure while accounting for a number of other components such as the context, body parts, kinematics, and low-level visual features. Using a similar approach, Tarhan et al. (2021) revealed an action processing hierarchy in the visual system and emphasized the importance of actors' goals. Finally, Tarhan and Konkle (2020b) revealed brain networks recruited during the processing of videos of actions that carry information regarding body parts and the target of an action.

Understanding how objects and actions are perceived and which features are useful in this process is important also beyond the field of cognitive neuroscience. Advances in the knowledge about the anatomy and function of the visual system proved to be useful in computer science (Cichy et al., 2019; Hassabis et al., 2017; Wardle & Baker, 2020), first

allowing to create a mathematical model of an artificial neuron (McCulloch & Pitts, 1943) that later became a building block for a single-layer perceptron (Rosenblatt, 1958) and for currently widely used deep neural networks (DNNs) (Cios, 2018; Rumelhart et al., 1986). Likewise, methodological developments in the field of human object recognition contributed to creating better computational models. For instance, DNNs differ from humans in the source of information used for object classification, relying more on the texture of images rather than on the shape. It has been shown that teaching a DNN to classify objects based on shape improved the network's performance, resulting in a more accurate imitation of human-like judgments (Geirhos et al., 2019). Thus, understanding which features are important for humans to distinguish between objects and actions and to categorize them is crucial for building artificial models that can mimic cognitive processes.

In sum, whereas previous studies revealed a number of potential organizing principles of observed actions, we are lacking a thorough investigation of the categories and features that are used to identify and distinguish between them. The current study aims to address this gap in the literature. In Experiment 1, we used a multi-arrangement task of 100 actions depicted as static images in combination with inverse multidimensional scaling analysis (Kriegeskorte & Mur, 2012) to obtain the category-based structure that captures similarities between different actions. In short, participants were asked to arrange a set of action images on a computer screen in a way that the distances between the images reflect action similarity. In Experiment 2, we performed a free feature-listing experiment for the same actions as in Experiment 1 (using verbal material) that resulted in a wide range (approx. 6000) of action features. Subsequently, we reduced that list to 59 key features, for which we collected ratings in Experiment 3. By combining these ratings with the results of Experiment 1 we reveal

critical features that contribute towards the distinction between action categories, and we show how features may contribute to the category-based organization of observed actions.

Experiment 1

The aim of Experiment 1 was to determine the categories underlying the organization of observed actions.

Methods

Participants

Twenty participants took part in the experiment (ten females; mean age = 22 years, age range = 19–27). Participants were financially reimbursed for their participation. Experimental procedures were approved by the ethics committee at the University of Regensburg.

Materials

We used 100 images of a wide range of daily actions. The initial list of actions was chosen from a study of Vinson and Vigliocco (2008), which reported semantic feature production norms for a wide range of verbs (N = 216) referring to events from different semantic fields such as manner of motion, body motion, and communication. We selected verbs that present typical, well-known daily actions that are easy to depict as static images, e.g., *brushing hair*, *driving a car*, *eating*. Details regarding the selection of action word and images and the full list of actions are provided in the Supplementary Materials (Sections A.1.1, A.1.2, and Table A1).

Procedure

The multi-arrangement experiment (Kriegeskorte & Mur, 2012) was conducted at the University of Regensburg. Participants were asked to position images on the screen in such a way that the distance between the images reflected their perceived similarity in terms of their meaning, rather than the background (e.g., an outdoor scene or a kitchen) or the overall composition of the picture (see also Tucciarelli et al., 2019). As an example, actions with a very similar meaning (e.g., *running* and *walking*) should be positioned close to each other, while actions with very different meanings (e.g., *running* and *taking a shower*) should be positioned further apart (see Figure A2 for an illustration). In the first trial, all the 100 action images appeared on a so-called circular "arena". The arrangement was performed by drag-and-drop using the mouse and, when all the images were sorted inside the arena, the participant was asked to press a button, which started the next trial. In each trial, the program determined the dissimilarities between all the actions in Euclidean space on the basis of their pairwise distances on the screen. The program updates the estimates of the pairwise distances, such that the pairwise dissimilarity evidence increases progressively. In subsequent trials, images were sampled from the original stimulus set by picking those with the least amount of evidence. A detailed description of the multi-arrangement procedure can be found in Kriegeskorte and Mur (2012). The average duration of the experiment was 120 min.

Data analysis

Data analysis was carried out in MATLAB (The MathWorks Inc., Natick, MA, USA). Separately for each participant, Euclidean distances for all 4950 pairwise comparisons of the 100 actions were reshaped into a vector, and, subsequently, averaged across participants. The obtained vector was transformed into a 100 x 100 representational dissimilarity matrix

(RDM; see Figure A3), depicting the relation between actions for all possible pairs of actions. To visualize the results in 2D, we used non-metric multidimensional scaling (MDS; criterion: metric stress, stress value = 0.237, distance measure: Euclidean). To access the structure underlying the representation of observed actions, we conducted hierarchical clustering analysis (see also Tucciarelli et al. (2019) for a similar approach). First, to reveal the metric that is best suited for clustering the data, we calculated the cophenetic correlation coefficients (Sokal & Rohlf, 1962; function *cophenet* in MATLAB) for different metrics. This method allows computing the correlation between cluster distances (so-called cophenetic distances generated by the *linkage* function in MATLAB) and actual Euclidean distances between the clusters (generated by the pdist function), enabling to assess whether the chosen clustering method reflects the original distances accurately. We obtained the highest value (cophenetic correlation = 0.854) for the *average* method (unweighted pair group method with arithmetic mean (UPGMA), Sokal & Michener, 1958). The resulting method indicates which algorithm will be used to group the data points into clusters and compute between-cluster distances (*linkage* function). UPGMA is an agglomerative method for hierarchical clustering that starts with each data point being its own single cluster and, moving bottom-up, forms a cluster from two clusters for which the average distance is the smallest. The average distance is calculated as the mean distance between all the members of each group. Second, to determine the number of clusters which best describe the dataset, we computed the silhouette index (*si*) (Rousseeuw, 1987) in a range from 3 to 50 (which corresponds to half of the number of stimuli) (Figure A4). In brief, the silhouette index reveals how appropriate a clustering solution is, by taking one data point at a time and comparing its distances with all other data points within the cluster to the between-cluster distances of the nearest cluster. The silhouette index ranges from – 1 to 1, where 1 indicates the best clustering of the data, whereas 0

represents a random clustering. To obtain labels for the action categories revealed by hierarchical clustering, we conducted an online experiment in a separate group of participants. We only used clusters that contained at least two actions (which was the case for 11 out of 12 clusters). Participants were asked to provide category labels based on actions belonging to each category. We selected the final category labels on the basis of their frequency (see Section A.1.5 in the Supplementary Materials for a detailed description of the Category naming experiment).

Results

Participants sorted actions according to several clusters; according to the silhouette index, the optimal number of clusters was 12 (*si* = 0.23; see Figure A4). As mentioned in Section *Experiment 1*, *Data analysis*, we removed one cluster since it only consisted of one single action. The remaining 11 action categories were labeled as follows: *Aggressive actions*, *Communication*, *Food-related actions, Gestures*, *Hand-related actions*, *Hobby*, *Household-related actions*, *Interaction*, *Locomotion*, *Morning routine*, and *Sport-related actions*. Information about the categories and the corresponding actions is provided in Table A3. Figure 2.1 presents a 2D MDS solution depicting the 11 action categories together with the corresponding labels. Clusters along the first dimension appear to be organized into pleasant (*Sport-related actions*, *Hobby*) and non-pleasant actions (*Household-related actions*, *Aggressive actions*). The second dimension might represent the presence (*Food-related actions*, *Morning routine*, *Household-related actions*) or absence (*Hand-related actions*, *Locomotion*, *Interaction*, *Gestures*, *Aggressive actions*) of a tool.

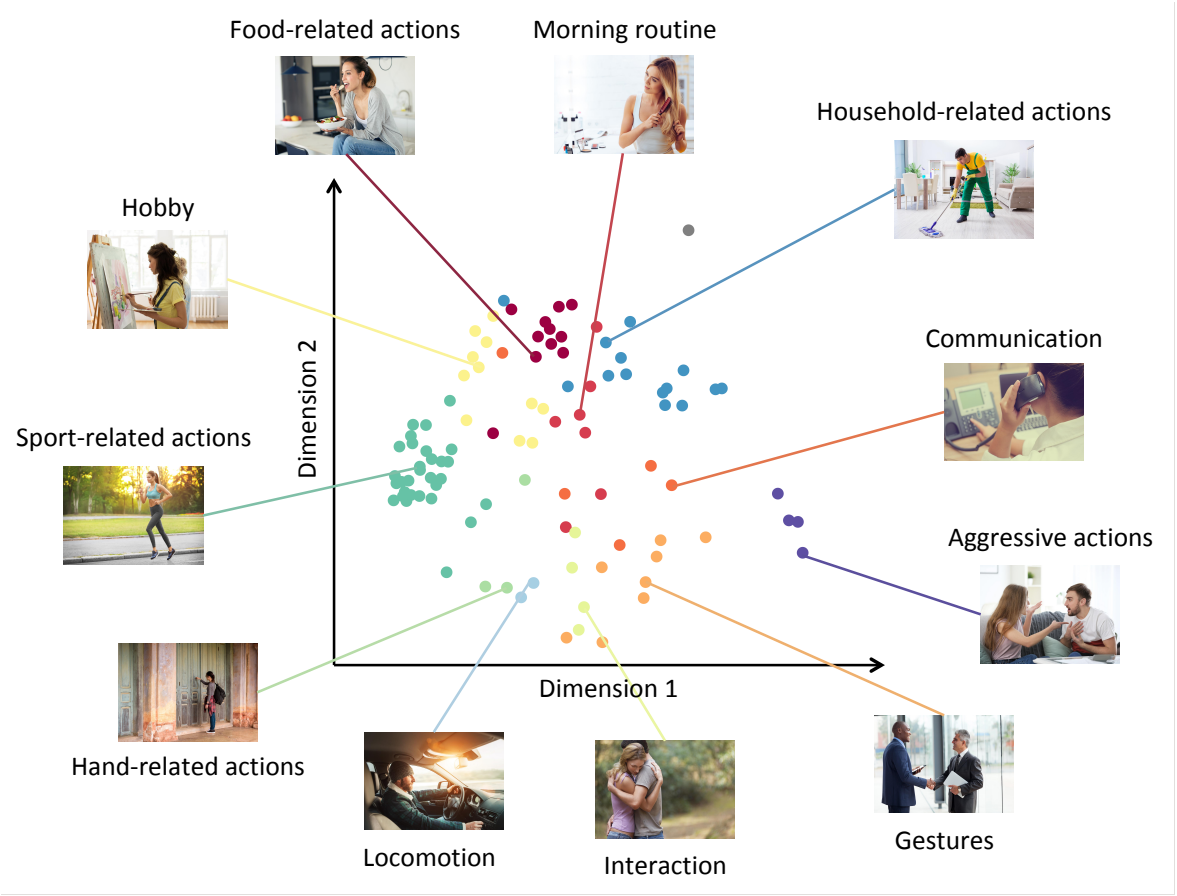


Figure 2.1 Action categories revealed by the multi-arrangement experiment and their arrangement in a two-dimensional space. Each dot represents one action, colors correspond to different action categories. For each category, a representative action with the corresponding category label is shown for ease of visualization. The gray dot indicates the action smoking belonging to the single-action cluster that was not considered in the subsequent experiment (see Section *Experiment 1, Data analysis*).

Discussion

Experiment 1 revealed 11 action categories, namely *Aggressive actions*, *Communication, Food-related actions*, *Gestures*, *Hand-related actions*, *Hobby*, *Household-related actions*, *Interaction*, *Locomotion*, *Morning routine*, and *Sport-related actions*. The obtained categories provide an extension of the five categories obtained from 28 daily life activities by Tucciarelli et al. (2019). We discuss these results in relation to the existing literature in the General discussion.

Experiment 2

In Experiment 2, we aimed to explore the feature-based organization of actions. It is not obvious which features should be used to best describe actions, and to distinguish them from other actions (see also Hebart et al., 2020; Zheng et al., 2019). One possibility would be to select a number of features that are either theoretically motivated, or that have been proposed in previous studies (e.g., Binder et al., 2016; Orlov et al., 2014; Tarhan & Konkle, 2020b; Vigliocco et al., 2004; Vinson & Vigliocco, 2008; Watson & Buxbaum, 2014; Wurm et al., 2017; Yang et al., 2017). We reasoned that a disadvantage of such an approach is the risk to miss relevant features. To minimize this risk, we decided not to base the selection of features on the basis of previous studies alone, but to support this step with an exploratory feature generation task. To this aim, we asked a separate group of participants to provide action features for each of the 100 actions used in Experiment 1 using a free feature-listing paradigm. Subsequently, we used the obtained features in combination with features proposed in previous studies to select a subset of features to be used in an explicit feature rating of actions (Experiment 3).

Methods

Participants

Forty participants (15 females; mean age = 23 years, age range = 18–36 years), recruited among students at the University of Regensburg, took part in the study and were financially reimbursed for their time. Experimental procedures were approved by the ethics committee at the University of Regensburg.

Apparatus

We collected action features for 100 actions using an online survey (https://www.soscisurvey.de/).

Materials

Stimuli consisted of the same 100 actions as used in Experiment 1, depicted as German verbs in their infinitival form (see Table A1).

Procedure

Free feature-listing experiment In each trial, participants were provided with an action word (e.g., "laufen – to run") and were asked to generate features that are typical for that action, which are relevant to understand it, and by which the action can be distinguished from others. Participants were instructed to provide at least five features per action, and they were provided with example features for two actions that were not part of the experiment (see Supplementary Materials, Section A.2.1, for the full instruction). Each participant was asked to provide features for 25 actions, such that we obtained features from ten participants for each of the 100 actions. The duration of the experiment was approximately 25 min.

Selection of themes and key features The obtained list ($N = 5683$ features) consisted of duplicate features as well as features that were phrased differently but carried the same or a very similar meaning (e.g., "Werkzeug – *tool*" and "Werkzeuge – *tools*"). Whereas the identification of distinct features specific for each action or action category can without doubt be useful as well (e.g., Zhuang & Lingnau, 2021), the focus of the current study was to identify more general features that are suitable for the collection of ratings across a wide range of actions. This was the reason to reduce the obtained set of features. Reduction of

features was performed in several steps. First, separately for each of the 100 actions, a native German speaker collapsed duplicates of features or features with similar meaning (e.g., singular or plural nouns), while keeping different grammatical forms (e.g., "Anspannen – *to tighten*" and "Anspannung – *tension*") separate. This resulted in $N = 4504$ features; the corresponding list is provided on osf [https:// osf.io/73v58/], table 'Action features – 100 actions'. Next, we collapsed duplicates of features across the whole dataset, which resulted in 3243 unique features (see table 'Action features – unique' on osf). This set of unique features consisted of single words and phrases that differed in terms of their grammatical class (e.g., nouns, verbs, adjectives), and the level of abstraction (ranging from concrete, specific features such as *lifting an arm* to abstract features, such as *communication*). Given this large variety and number of features, our next goal was to reduce the number of collected features to the most crucial ones, while at the same time keeping as much information from the collected dataset as possible. To this aim, we grouped the features into "themes" that keep conceptually related features together (see Table 2.1 and Figure A6). To identify themes according to which these features can be organized, we conducted an exploratory analysis of the dataset. To do so, the same native German speaker and one of the authors thoroughly went through the data set and, to avoid subjectivity, independently came up with main themes that could best describe the content of the dataset. In the next step, following previous studies, we selected themes that could be backed-up by the features provided by the participants in Experiment 2. We chose the following themes: *Body parts* (Orlov et al., 2014; Tarhan & Konkle, 2020b; Yang et al., 2017), *Object-directedness* (Tarhan & Konkle, 2020b; Wurm et al., 2017; Yang et al., 2017), *Type of limb movement* (Tranel et al., 2003; Watson & Buxbaum, 2014), *Posture* (Peelen & Downing, 2007), *Location* (Wurm & Schubotz, 2017), *Keeping balance* (Vaessen et al., 2018), *Harm* (Tranel et al., 2003; Binder et al., 2016),

Change of location (Vinson & Vigliocco, 2008), *Duration* (Tranel et al., 2003; Binder et al., 2016; Yang et al., 2017), *Contact with others* (Vinson & Vigliocco, 2008; Binder et al., 2016), *Use of force* (Watson & Buxbaum, 2014; Yang et al., 2017), *Goal-directedness* (Hamilton & Grafton, 2007; Wurm et al., 2016), *Pace* and *Concentration* (Binder et al., 2016), *Noise* (Tranel et al., 2003; Vinson & Vigliocco, 2008), and *Valence* (Tranel et al., 2003; Binder et al., 2016). The remaining themes (i.e., *Trajectory*, *Water*, and *Season-dependence*) were selected based on the collected features. We decided to include these themes because they were mentioned frequently.

In the next step, for each theme, we identified features belonging to each theme and selected those that were frequently mentioned by the participants. Based on these features, we came up with key features. This way, we aimed to preserve detailed features from participants' responses while keeping them organized within groups containing conceptually related information.

Because of the wide range of features obtained from the feature-listing experiment, we decided on two different types of key features: binary features that are reasonable to be judged with a yes/ no answer (e.g., *Arms*) and continuous features for which we considered it more useful to ask for a rating on a scale (e.g., *Pace*) (see Table 2.1 for a better understanding of the key features and the possible answers).

Results

From the free feature-listing experiment, we obtained 5683 features describing 100 daily actions. Due to the nature of the task, the features varied in a number of different ways, e.g., in terms of the grammatical class (verbs, nouns, adjectives), the phrase length (single words, phrases, sentences), and the level of abstraction. The complete list of features

separately for each of the 100 actions as well as a table with all the unique features across the 100 actions are available at https://osf. io/73v58.

In the second part of Experiment 2, we selected 59 key features (such as *Arms*, *Hands*, *Targeting a tool*, *Targeting a person*), which we divided into 19 broader themes: *Body parts*, *Object-directedness*, *Trajectory*, *Type of limb movement*, *Posture*, *Location*, *Keeping balance*, *Harm*, *Water*, *Season-dependence*, *Change of location*, *Duration*, *Contact with others*, *Pace*, *Use of force*, *Goal-directedness*, *Concentration*, *Noise*, and *Valence*. Thirteen of the themes contained binary features, i.e., those that can be either involved in an action or not (e.g., *Arms*, *Legs*), whereas the remaining themes contained features that could be described on a continuous scale (e.g., *Pace*, from slow – 1 – to fast – 7 –). Moreover, some of the themes contained multiple features (e.g., *Arms*, *Shoulders*, *Legs* etc. for the theme *Body parts*) whereas some contained one feature only (e.g., *Keeping balance*). For the latter, we refer to the themes and the features with the same name. The full list of themes and their corresponding features is provided in Table 2.1.

Table 2.1. List of action features grouped by the corresponding themes

		Theme	Description of the theme	Features	Possible answers
1		Body parts	Which body parts are involved?	Arms; Shoulders; Dominant Hand; Both hands; Fingers; Legs; Hips; Feet; Head; Mouth	Yes/No
2		Object-directedness	What is the target of the action?	Targeting a non-manipulable object; Targeting a manipulable object; Targeting a tool; Targeting a person; No object involved	Yes/No
3		Trajectory	In which direction does the body move during the action?	Horizontal; Vertical; No movement; Unspecified trajectory	Yes/No
4		Type of limb movement	How do the limbs move?	Circular arms; Circular legs; Rotating arms; Rotating legs; Abduction/adduction arms; Abduction/adduction legs; Sweeping arms; Sweeping legs; Up-down arms; Up-down legs	Yes/No
5		Posture	What posture is involved during the action?	Straight posture; Bent posture; Sitting; Laying; No specific posture	Yes/No
6		Location	Does the action take place indoor, outdoor, or can be both?	Indoor; Outdoor	Yes/No
7		Keeping balance	Does the action require keeping balance?	Keeping balance	Yes/No
8		Harm	Is it likely that the action can cause harm?	Harm	Yes/No
9		Water	Does the action require water?	Water	Yes/No
10		Season-dependence	Is the action season-dependent?	Season-dependence	Yes/No
11		Change of location	How much does the actor change location during the action?	Far away from the starting point; In proximity; No change of location	Yes/No
12		Duration	What is the duration of the action?	Up to a few seconds; A few seconds to a few minutes; A few minutes to half an hour; Half an hour to an hour; Several hours; A day	Yes/No
13		Contact with others	Does the action involve contact with another person?	Contact from a distance; Touching another person; Indirect contact; Does not require contact with a person	Yes/No

14	Pace	With which pace is the action performed?	Pace	[scale from 1 (slow) to 7 (fast)]
15	Use of force	How much force is required to perform the action?	Use of force	[scale from 1 (no force) to 7 (lots of force)]
16	Goal-directedness	To which degree is the action goal-directed?	Goal-directedness	[scale from 1 (not goal-directed) to 7 (goal-directed)]
17	Concentration	To which degree does the action require concentration?	Concentration	[scale from 1 (barely) to 7 (a lot)]
18	Noise	What is the degree of noise?	Noise	[scale from 1 (silent) to 7 (loud)]
19	Valence	To which degree does the action evoke positive or negative emotions?	Valence	[scale from 1 (negative) to 7 (positive)]

Themes and key features selected based on the features obtained from the free feature-listing experiment (see Section *Experiment 2*, *Procedure*, *Selection of themes and key features* for a detailed description of the procedure) and based on the existing literature. The selection resulted in 59 key features grouped into 19 themes. Each of the themes is marked by a different color. Thirteen out of the 19 themes consist of a set of binary features (possible answers: "Yes/ No"). The other six themes (e.g., *Pace*, *Use of force*) correspond to features that can be judged on a continuous scale (e.g., *Pace*, from 1 = slow to 7 = fast). Nine themes contain multiple features whereas the remaining ten themes consist of one single feature only. In the latter case, we refer to theme and feature with the same name (e.g., *Pace*). Questions describing each theme are provided to better illustrate the meaning of the theme. The last column contains possible answers (see *Experiment 3* for details).

Discussion

In Experiment 2, we obtained a wide range of different features (5683 in total) that human participants associate with different actions. We subsequently summarized that list to 59 features, organized by different themes, to be used in an explicit feature rating (Experiment 3). The reduced set of features covers varying levels of abstractness, ranging from concrete features, e.g., *Body parts*, *Type of limb movement*, *Posture* to more abstract features such as *Harm*, *Valence* and *Goal-directedness* (Wurm et al., 2016). Moreover, the features cover different semantic domains (see e.g., Binder et al., 2016). Specifically, features are part of the sensory (e.g., *Noise*), motor (e.g., *Type of limb movement*), space (e.g., *Location*, *Change of location*), time (*Duration*), social (*Contact with others*), emotion (*Valence*), and drive (*Goal-directedness*) domains. Both the complete set of features provided by the participants as well as the selected set of features may serve as a starting point for future studies concerned with the behavioral and neural correlates of specific features, and for computational models aimed at the recognition of human actions.

Experiment 3

The goal of Experiment 3 was (1) to identify critical features that are used to distinguish between different action categories, and (2) to directly compare the feature- and category-based organization of actions. To this aim, we collected ratings for 59 action-related features obtained from Experiment 2. Methods Participants A total of 273 participants took part in the rating experiment (231 females; mean age = 28 years, age range = 16–67 years) and were financially reimbursed for their time. Experimental procedures were approved by the ethics committee at the University of Regensburg.

Apparatus

The study was conducted as an online survey (www.sosci survey.de).

Materials

Stimuli consisted of the same 100 actions as used in Experiments 1 and 2, depicted as German verbs in their infinitival form (see Table A1 for a list of all actions). The actions were rated based on 59 key features selected in Experiment 2.

Procedure

Participants were asked to provide ratings for 59 features (see Table 2.1). Instructions provided to the raters can be found in Section A.3.1. For the features of 13 of the themes, binary answers ("Yes" or "No") were required, whereas features for the remaining six themes had to be judged on a scale from 1 to 7 (e.g., 1: "Not at all", 7: "Very much") (see Table 2.1 for details). In addition to the instruction, participants were provided with an example action (not used in the study) with corresponding example ratings.

For the features from the themes *Body parts*, *Object-directedness*, *Trajectory*, *Type of limb movement*, *Posture*, *Location*, *Keeping balance*, *Duration*, *Contact with others*, *Pace*, *Use of force*, and *Goal-directedness* 17 participants rated 25 action words each (425 ratings in total), which took approximately 45 min. To reduce the amount of time per participant, another set of 107 participants rated five action words each (535 ratings in total), which took about 10 min. For features from the remaining themes, we collected ratings from a separate group of 149 participants. Each participant rated five action words (745 ratings in total) and the full experimental session lasted approximately 10 min. The set of actions chosen for each participant was randomized.

Data analysis

All subsequent analyses, unless stated otherwise, were conducted using MATLAB (The MathWorks Inc., Natick, MA, USA). We obtained 1680 ratings in total, with the number of ratings per action ranging between seven and eleven. First, we reduced the redundancy within the features (see Supplementary Materials, Section A.3.2.1, for details). Since ratings differed depending on the theme (either "Yes/No" answers or ratings on a scale), we transformed "Yes/No"- answers to values of 1 or 0 and rescaled values of continuous ratings to a range from 0 to 1. To avoid multicollinearity, we removed features that were highly correlated (see Section A.3.2.2). The final set comprises of 44 features.

Multi-feature model To depict which features are important for different actions, we averaged ratings across participants and created a multi-feature model (Figure 2.2). Additionally, we selected four exemplary features and showed actions that received minimum and maximum ratings (Figure A9) for an additional visualization of the results of the feature rating.

Feature-based representations of action categories The aim of this analysis was to identify features that are most important to describe action categories and to distinguish between them. First, separately for each of the categories identified in Experiment 1, we collapsed the ratings across actions within a given category. Subsequently, we used radar charts together with 95% confidence intervals (across feature ratings of actions within a category) to visualize the mean action ratings within each action category (Figure 2.3, left panel, and Figure A10).

Following Binder et al. (2016), to depict quantitative differences between each action category and the remaining categories, we computed the difference between the rating for

each feature of a given category (same as depicted in Figure 2.3, left panel, and Figure A10) and the mean rating of that feature for all remaining categories (see Figure 2.3, right panel and Figure A12). The difference is presented as z-scores and the significance threshold was set at $p < 0.05$, corrected for multiple comparisons using false discovery rate (FDR) estimation (Benjamini & Hochberg, 1995).

Feature-based representations of single actions To depict feature-based ratings of single actions, we computed the mean ratings of each feature across participants and mapped them on individual radar charts (see Figure A11), separately for each action.

Correlation of category- and feature-based models To determine how features contribute to explaining the category-based structure revealed by Experiment 1, we compared the results from the multi-arrangement task (Experiment 1) with the feature-based ratings (Experiment 3), using 52 different feature models: two multi-feature models (weighted and unweighted), 44 single-feature models and six theme models. As a first step, we transformed the category-based model and the feature models into RDMs (see Sections *Experiment 1*, *Data analysis* and *Experiment 3, Data analysis*, *Creating feature RDMs*, for details). Next, we computed the correlations between the category-based RDM and the feature RDMs using the RSA toolbox (Nili et al., 2014) and MATLAB scripts available from Storrs et al. (2020, 2021): within each cross-validation fold, together with feature weights (see Section *Experiment 3, Data analysis, Cross-validated reweighting of features*), we calculated correlations between the feature RDMs and the category-based RDM (Kendall τ_A) as well as the lower and upper bounds of the noise ceiling (see also Storrs et al., 2020). We ran 50 cross-validation folds, and within each fold ten randomly selected actions and five randomly selected participants were assigned as test data. At the end of the 50 cross-validation folds,

the correlations, weights, and bounds of the noise ceiling were averaged. This procedure was repeated for 1000 bootstrap samples. Significance of the RDM correlations was determined by bootstrap resampling of the stimuli and controlled for multiple comparisons using FDR at 0.05.

Creating feature RDMs Feature RDMs were generated by computing the Euclidean distance for each pair of actions for (1) all the unweighted features together (resulting in one unweighted multi-feature RDM), (2) all the weighted features together (resulting in one weighted multi-feature RDM (see Section *Experiment 3, Data analysis, Cross-validated reweighting of features*), (3) each feature separately (resulting in 44 single-feature RDMs), and (4) each theme separately (resulting in six theme-based RDMs). The theme-based RDMs were computed for those themes that contained more than one feature, more precisely *Body parts*, *Object-directedness*, *Type of limb movement*, *Trajectory*, *Posture*, and *Location*. This allows investigating sets of related features together. Figure A13 shows all 52 feature RDMs (i.e., weighted and unweighted multi-feature RDMs, 44 single feature RDMs and six theme-based RDMs).

Cross-validated reweighting of features In addition to the multi-feature RDM with equal weights for each feature, we aimed to examine whether the category-based organization could be accounted for better by a weighted multi-feature RDM. Following Jozwik et al. (2016) and Storrs et al. (2021), we thus performed feature reweighting to fit the categorical action structure, while cross-validating over participants and actions. For that purpose, we used non-negative least squares fitting with the MATLAB function lsqnonneg. Weights were estimated for all the 44 single-feature models.

Results

Multi-feature model

The multi-feature model (Figure 2.2) depicts information regarding the rated contribution of features for different actions. Each row presents one action, whereas each column represents a feature. The grayscale represents the mean rating of a feature (black: “Yes/Very much”; white: “No/Not at all”). For an intuitive understanding of the model, Figure A9 shows example actions that received the minimum and maximum rating for some exemplary features.

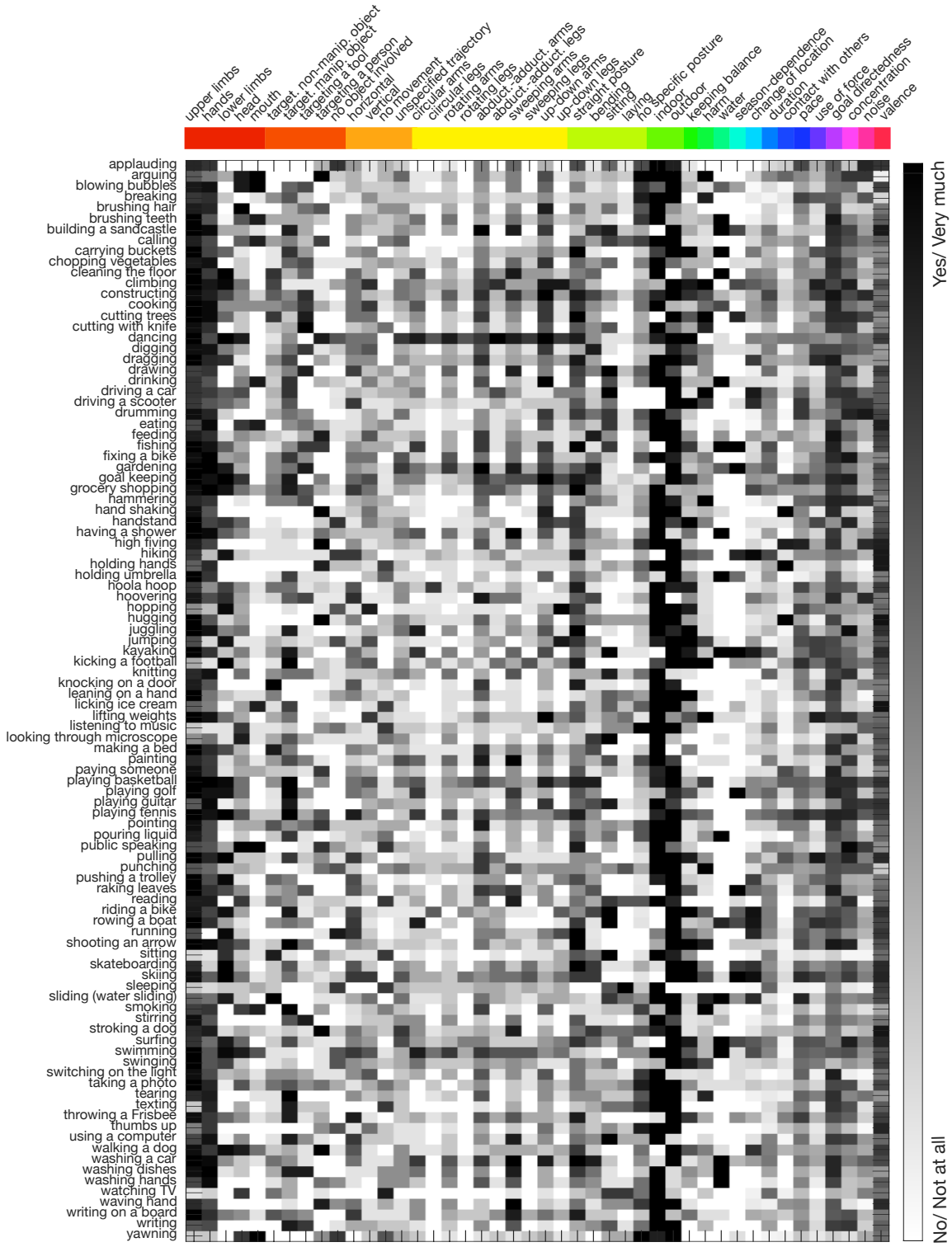


Figure 2.2. Multi-feature model. The model depicts ratings for 100 actions (*rows*) based on 44 features (*columns*). Different shades of gray indicate the mean rating of a feature (*black*: high rating, *white*: low rating). Features belonging to the same theme are depicted by the same color on the top of the figure (same color code as Table 2.1).

Feature-based representations of action categories

Next, we aimed to identify features that were judged as important for specific action categories. Separately for each category, we visualized feature-based ratings across actions as radar charts (see Section Experiment 3, Data analysis, for details). Figure 2.3 (left column) shows the results for four representative action categories; Figure A10 shows all 11 categories. As can be seen, radar charts reveal similarities as well as differences between the different categories. As an example, the features *Goal directedness*, *Upper Limbs* and *Hands* appeared to be judged as important for most categories. As can be seen in Figure A11, showing radar charts for single actions grouped by action categories, this observation appears to be rather consistent across individual actions within each category. By contrast, other features appeared to be more distinct, such as the features *Harm* and *Noise* for the category *Aggressive actions*, *Targeting a person* and *Contact with others* for the category *Interaction*, or the feature *Keeping balance* for the category *Sport-related actions*.

Quantitative differences of feature ratings between each action category and the remaining categories are depicted in Figure 2.3, right column (see Figure A12 for all 11 categories). This comparison allows identifying the most relevant features for a given category. For example, not surprisingly, *Aggressive actions* (Figure 2.3, 1st row) got significantly higher ratings on *Harm* and *Noise*, and significantly lower ratings on *Valence*, in comparison to the feature ratings obtained for the remaining categories.

For the category *Locomotion* (Figure 2.3, 2nd row), this analysis revealed that the features *Change of location*, *Noise*, and *Harm* received significantly higher ratings than the remaining categories. By contrast, the feature *Indoors* was rated lower in comparison to the ratings obtained in the remaining categories. While the importance of the features *Noise* and

Harm might be surprising at first, they are likely due to the specific actions that were part of this category (i.e., *Driving a car* and *Driving a scooter*).

For the category *Interaction* (Figure 2.3, 3rd row), the analysis highlighted the importance of the features *Targeting a person* and *Contact with others*, and lower ratings for the *Use of force*, *Targeting a manipulable object*, *Lower limbs*, and *Duration*.

Finally, for the category *Sport-related actions* (Figure 2.3, 4th row), pairwise comparisons revealed several movement-related features, such as *Keeping balance* and *Lower limbs*.

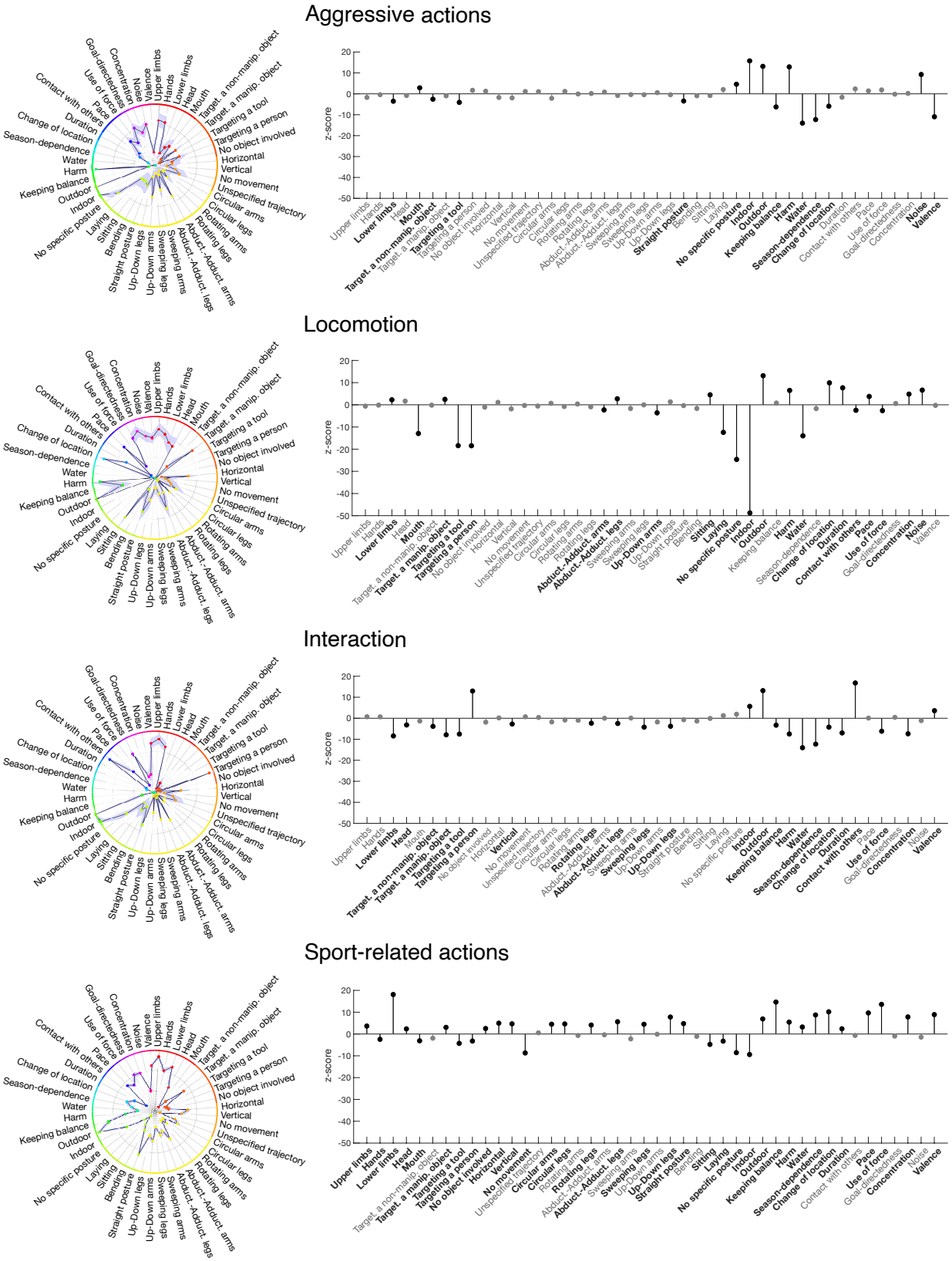


Figure 2.3 Visualization of feature vectors for selected action categories. Left column: Feature-based radar charts. *Colors* indicate features belonging to the same theme (same color code as Table 2.1). The length of each

spike is proportional to the mean rating of the corresponding feature. *Shaded area* indicates 95% confidence interval computed across feature-based ratings of actions within a category. Right column: Quantification of the difference of the mean rating for features of a given category (depicted in the left column) with the mean ratings of the remaining categories, presented as z-scores. Z-scores above zero indicate higher feature ratings for a given category compared to the remaining categories, whereas z-scores below zero indicate lower feature ratings for that category compared to the remaining categories. Features that differentiate the category from the remaining categories ($p < 0.05$, FDR corrected) are marked in *black* and their corresponding names are highlighted in *bold*.

Correlation of category- and feature-based models

In this analysis, we aimed to determine whether and to what extent the models based on the similarity of features account for the organization based on action categories. As a first step, we computed the correlation (Kendall τ_A) between the category RDM (obtained from Experiment 1) with the (unweighted) multi-feature RDM (obtained from Experiment 3) that consisted of all 44 features treated equally. As shown in Figure 2.4a (right bar), the correlation between the unweighted multi-feature RDM and the category RDM (corr = 0.085) is significantly different from zero but does not reach the noise ceiling.

So far, we treated each feature as contributing equally to the representation of observed actions. However, it is likely that some features are shared across categories, whereas other features or combinations of features are distinct for specific categories (Tyler & Moss, 2001). Similarly, to what has been proposed for object categories (Jozwik et al., 2016; Khaligh-Razavi & Kriegeskorte, 2014), action categories might differ with respect to the weights assigned to different features. To identify which features contribute the most to the categorical organization of actions revealed by Experiment 1, we thus used non-negative least-squares fitting for assigning weights to the features (see Section *Experiment 3, Data analysis*, for details) and created a weighted multi-feature model that consisted of 44

weighted features. It should be noted that feature ratings and feature weights are two separate values. In the case of feature ratings, the ratings are provided by participants and each feature is treated as contributing equally to the category-based model. However, the assumption that all features contribute equally to the category-based organization is likely to be too simplistic. The estimates of the weights of each feature, generated by non-negative linear least squares fitting, provide an insight into the importance of a given feature in explaining the category-based action representation.

As can be seen in Figure 2.4a (left bar), the correlation between the weighted multi-feature RDM and the category RDM (corr = 0.121) is significantly different from zero and significantly different from the unweighted multi-feature model (calculated with stimulus bootstrap test, $p < 0.05$, FDR corrected). An overview of the significant correlations between the category model and the different feature models (unweighted multi-feature model, weighted feature model, single-feature models, and theme models) is shown in Figure A14.

Features and the corresponding weights that formed the weighted multi-feature model are shown in Figure 2.4b. The highest weight was obtained for the feature Valence, followed by substantially lower weights for features related to noise, posture, tool-directedness, and head.

Discussion

The results of Experiment 3 allow narrowing down which features might contribute most to the recognition and distinction between different action categories. Additionally, a direct comparison between category- and feature-based RDMs indicated that the former can be best explained by a combination of weighted features, showing an importance of the rated

valence of the action. However, part of the structure remains unexplained as evidenced by the best model not reaching the noise ceiling.

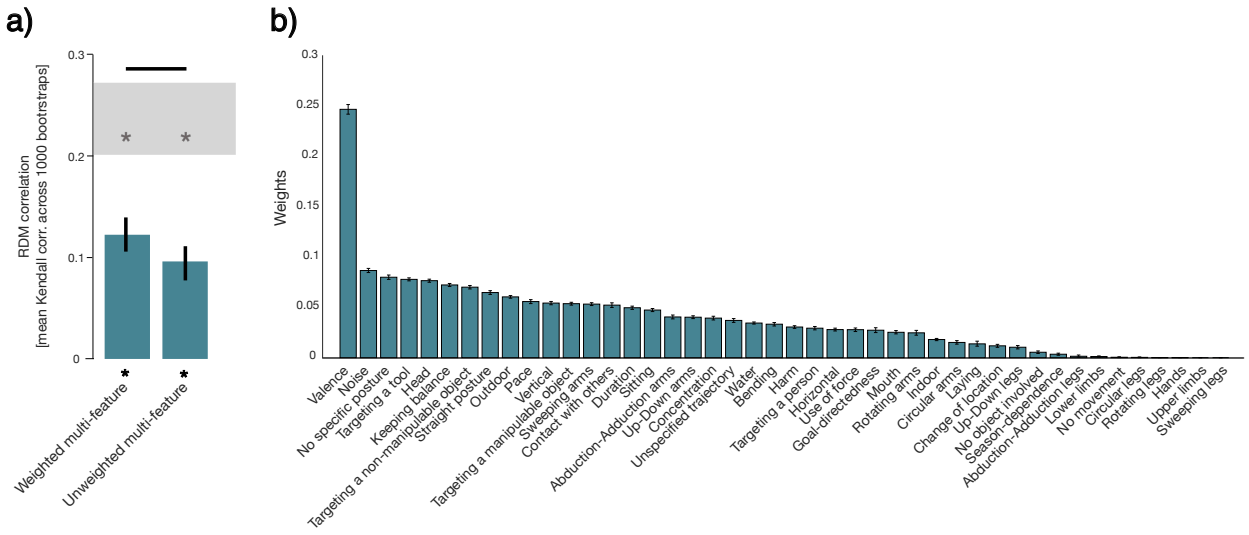


Figure 1.4. (**a**) Correlation between the weighted and unweighted multi-feature RDMs (obtained from Experiment 3) with the category RDM (obtained from Experiment 1). *Error bars* show the 95% confidence interval estimated by bootstrap resampling of the stimulus set. *Asterisks* at the bottom indicate that both multi-feature RDMs were significantly correlated with the category RDM (stimulus bootstrap test, $p < 0.05$, FDR corrected). The noise ceiling (*shaded area above the bars*) indicates the expected performance of the true model taking into account the noise in the data. None of the models reached the noise ceiling. The *horizontal line* above the noise ceiling indicates significant differences between the weighted and the unweighted multi-feature model (stimulus bootstrap test, $p < 0.05$, FDR corrected). Correlations between all feature RDMs and the category RDM are shown in Figure A14. (**b**) Feature weights contributing to the weighted multi-feature model, obtained with non-negative least squares in the cross-validation procedure (50 cross-validation folds across participants and stimuli, averaged across 1000 bootstrapping iterations). *Error bars* indicate the 95% confidence interval of the bootstrap distribution. Only features with non-zero weights are shown.

General discussion

We performed three behavioral experiments to characterize the category- and feature-based structure underlying the organization of observed actions. Moreover, we examined the degree to which the feature-based similarity between actions accounts for the category-based organization.

In Experiment 1, we identified 11 categories, including *Communication*, *Gestures*, *Locomotion*, and *Aggressive actions* (Figure 2.1). Subsets of these categories have been reported in previous studies focusing on actions depicted as pictures, videos or animations, ranging from *Hand-related actions* (Handjaras et al., 2015; Wurm et al., 2017; Wurm & Lingnau, 2015) over *Locomotion*, *Food-related actions*, *Morning routine*, and *Sport-related actions* (Abdollahi et al., 2013; Tarhan & Konkle, 2020b; Tucciarelli et al., 2019) to *Interaction* (Isik et al., 2017; Papeo, 2020; Tucciarelli et al., 2019; Wurm & Caramazza, 2019b). Not surprisingly, the obtained categories partly overlap with action verb categories, some of which formed the basis of stimulus selection for the current study (e.g., Vinson & Vigliocco, 2008). Although some of these category labels were similar (e.g., *Food-related actions* obtained in the current study and *Cooking* obtained from action verbs), the actions belonging to them differed: e.g., actions from the *Food-related actions* category belonged to action verb categories such as *Body-action* (e.g., *Drinking*, *Eating*, *Licking*) or *Change of state* (e.g., *Pouring*, *Stirring*). These slight differences in action categorization might depend on the stimulus material (visual vs. verbal) and the way these categories were determined (multi arrangement task of static images vs. semantic relations/ verb usage patterns). Our reason to use verbal material instead of visual stimuli to obtain features and feature ratings in Experiments 2 and 3 was to obtain features that are less dependent on the specific (visual)

exemplar presented to the participant. That said, we would like to stress that action categories revealed on the basis of visual material are unlikely to be identical to the action categories revealed on the basis of verbal material (see also Tucciarelli et al., 2019; Vinson & Vigliocco, 2008; Watson & Buxbaum, 2014).

Previous studies examining dimensions underlying the organization of observed actions revealed the importance of the hand configuration and the magnitude of the arm movement for actions related to a tool (Watson & Buxbaum, 2014). For 28 daily-life actions depicted as static images, Tucciarelli et al. (2019) highlighted the type of change induced by the action and the fulfillment of basic versus higher needs. Finally, using 152 everyday action videos and an odd-one-out task combined with a data-driven approach, Dima et al. (2020) reported a number of dimensions, ranging from visual information to more social and semantic aspects of actions. A study based on large-scale text analyses revealed that actions can be described by six dimensions, including Abstraction, Food, and Spiritualism (Thornton & Tamir, 2022). In the current study, Experiment 1 suggested that a pleasant/ non-pleasant and a tool-/ non-tool-related dimension might underlie the category-based structure. The assumption of a dimension related to pleasant/ non-pleasant actions is compatible with the results of Experiment 3 that revealed that the judged valence of an action strongly contributed to the category-based organization. It is likely that the specific dimensions revealed by different studies depend, among other things, on the type of actions included in the experiment and the methods used to reveal these dimensions.

In Experiment 2, we used a free feature-listing paradigm to identify features participants consider important for the description of actions, and for their distinction from other actions. According to the Theory of Action Identification (Vallacher & Wegner, 1985;

Wegner & Vallacher, 1986), actions can be identified at different hierarchical levels that differ based on whether the feature provides information about *how* or *why* a given action is performed. According to this theory, lower identity levels are expected to rely more on movement-related information, whereas higher levels of the hierarchy are expected to be associated with a more abstract understanding of actions. Likewise, Hamilton and Grafton (2007) proposed that actions can be identified at different hierarchical levels (specifically, muscles, kinematics, goals and intentions), and that these different levels engage different brain regions. To be able to cover features from different hierarchical levels, we explicitly instructed participants to consider both concrete and abstract features (see Supplementary Materials, section A.2.1, for details on the instructions). Moreover, the obtained features cover a range of different semantic domains (e.g., Binder et al., 2016), such as sensory, motor, spatial, and temporal information.

In contrast to a selection of features purely based on previous studies, our exploratory approach for the selection of features reduces the risk of missing potentially relevant features so far not covered in the literature. The obtained set of features thus is considered to be an important extension of previous studies (e.g., Binder et al., 2016; Tarhan & Konkle, 2020b; Tucciarelli et al., 2019; Vinson & Vigliocco, 2008) while serving as a basis for future experiments.

The goal of Experiment 3 was to better understand the feature-based structure of observed actions. To this aim, we collected explicit ratings for each of the 100 actions for 59 features selected on the basis of a free feature listing paradigm (Experiment 2) and features proposed in previous studies. This allowed us to compute a feature-based similarity structure of observed actions, which we used to determine (a) which features are most informative for

the description of action categories and their distinction from other categories and (b) to directly compare the feature-based similarity structure with the category-based similarity structure revealed in Experiment 1. We identified a number of plausible critical features for specific categories, e.g., *Targeting a person* for the category *Interactions*; *Harm*, *Noise* and *Negative valence* for the category *Aggressive actions*; and features such as *Lower limbs*, *Keeping balance*, and *Use of force* for the category *Sport-related actions* (see Figure 2.3, Figure A10, and Figure A12). Together, our results are in line with the view that each action category is characterized by some distinct combination of features. However, it is worth mentioning that a large proportion of the category-based structure remained unexplained. We will return to this point in the following sections.

A direct comparison of the category- and feature-based organization revealed that the weighted multi-feature model best explained the variability of the category-based structure and was significantly different from the remaining models (Figure 2.4a, Figure A14). The feature that contributed most to this organization was the valence (positive/negative) of the actions (Figure 2.4b), highlighting the role of valence-related information for the categorization of actions included in the current study. In Figure 2.1, we visualized the action category structure resulting from the multi-arrangement task (Experiment 1) in a two-dimensional space and referred to a pleasant/ non-pleasant dimension: Action categories associated with pleasure (e.g., *Sport-related actions*, *Hobby*, *Food-related actions*) are on the left side, whereas action categories with displeasure (e.g., *Aggressive actions*, *Household-related actions*) are on the right side. In Figure A15, we present the same arrangement of actions in a two-dimensional space, and color-code the rated valence of each action, ranging from low (red) to high (yellow) values. This figure highlights the gradual change of valence moving from the left (positive valence) to the right (negative valence) side of the figure.

Thus, the organization of actions shown in Figure 2.1 and Figure A15 might reflect whether the participants associated the actions with pleasure/ positive valence or with displeasure/ negative valence. These results are in line with previous studies examining the relationship between the processing of emotions and actions (e.g., Kroczek et al., 2021; Poyo Solanas et al., 2020). The importance of valence has also been proposed by Tamir and Thornton (2018) who suggested that valence is involved in the process of prediction other person's actions. As suggested by the authors, understanding another person's action relies upon understanding another person's mental states and traits. Persons with traits of a high valence value (e.g., agreeable) most likely exhibit positive mental states (e.g., content) that lead to performing positive actions (e.g., cooperation). The authors explored dimensionalities of mental states and traits revealing the importance of valence, but they did not explore the corresponding organization of actions. In our work, we aimed to fill this gap. Based on the results from Tamir and Thornton (2018), it seems plausible that valence serves as one of the crucial dimensions underling the organization of observed action.

As mentioned above, whereas the weighted multi-feature RDM showed the highest similarity with the category RDM, a substantial part of this structure remained unexplained. In other words, even the combination of intermediate- (e.g., *Body parts*, *Type of limb movement*) and high-level features (e.g., *Valence*, *Goal-directedness*) cannot fully account for the categorical model of actions. This suggests that there must be additional organizing principles underlying the cognitive architecture of actions (see also Vinson & Vigliocco, 2008). Likely candidates that may contribute towards the organization of actions that were not specifically examined in the current study is the context or scene in which an action typically takes place (see also Tucciarelli et al., 2019; Wurm et al., 2012; Wurm & Schubotz, 2012), as well as specific objects or tools used for a given action (Bach et al., 2014; Wurm

et al., 2012). We expect that more fine-grained ratings for specific scenes (e.g., kitchen, office) and objects (e.g., knife, pen), which were beyond the scope of the current study, will allow a more accurate representation of the category-based action structure.

Future directions

Characterizing and establishing a set of action categories and features obtained from a wide range of actions, as well as better understanding of how features contribute to category-based action representation, might be useful for future behavioral and neuroimaging studies. Given the data-driven nature of Experiment 1, the individual actions that formed the categories revealed by the multi-arrangement task varied between 2 and 30, making it hard to draw firm conclusions regarding a final set of critical features for some of the categories (e.g., *Locomotion*). To be able to examine these categories in more detail, future studies should choose a more balanced number of actions per category. With these limitations in mind, we believe that our results will serve as a useful basis to stimulate future studies. As an example, future behavioral and neuroimaging studies might address how human participants process and respond to different action categories, how these capabilities develop, and under which circumstances they may become impaired. Moreover, our results might be useful in the field of computer vision and human–robot interactions. As mentioned by Wardle & Baker (2020), object recognition relies on different types of information, such as object's appearance, task context, function of the object, and its possible interactions with other objects. The more we understand how the human brain tackles the problem of object recognition, the more accurate and human-like artificial models can be built. The same can be applied to actions: a better understanding of the way humans represent and categorize actions may lead to the creation of more accurate and efficient computational models of

action recognition with higher resemblance to human's understanding of the world. In the future, such models could be applied in automated action recognition that has become a growing demand in the industry, including autonomous vehicles and driver-assistance systems, smart surveillance, and health care systems (Serpush & Rezaei, 2021).

Conclusions

We identified a set of action categories and showed that each of them is represented by a unique combination of action features. The reported action features can be grouped into more general themes, such as *Body parts* and *Posture*, partly overlapping with features already reported in the existing literature. Whereas the weighted multi-feature model performed best among all the examined models in explaining the category-based structure, a significant proportion of variability remained unexplained, suggesting that there are additional sources of information that contribute to the categorization of observed actions, beyond the features examined in the current study. Together, our results provide important insights into the cognitive architecture underlying our ability to distinguish between different actions and serve as an extension of the existing literature (e.g., Tucciarelli et al., 2019; Vigliocco et al., 2004; Vinson & Vigliocco, 2008). Additionally, the obtained features might be applied in computational science and help improving neural network models that could lead to more accurate computer-based action recognition in the future.

CHAPTER 3: STUDY 2

„OVERLAPPING BUT DISTINCT REPRESENTATIONS OF OBSERVED ACTIONS AND ACTION-RELATED FEATURES”

The manuscript has been submitted to Human Brain Mapping and is currently in revision.

Abstract

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

Introduction

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

The dimensions spanning the space of actions have been investigated by several behavioral studies. For instance, in the realm of tool usage, Watson & Buxbaum (2014) demonstrated that tools can be sorted into distinct groups based on two dimensions: one associated with the hand configuration and the other with the magnitude of the arm movement. Tucciarelli et al. (2019) showed that daily-life actions can be mapped onto dimensions reflecting the type of change induced by the action, and the type of need to be fulfilled by the actions (ranging from basic, physiological needs to higher social needs). Furthermore, social importance has emerged as a prominent factor in various other studies, either as the main factor in judgement of action similarity (Dima et al., 2022) or as one of the factors, together with semantic dimensions (e.g., food, work, home life) and visual information (scene setting; Dima et al., 2023). A recent study of Vinton et al. (2023)

suggested that actions might be projected onto four dimensions: two related to facial traits and emotions (e.g., friendly – unfriendly) and two others unique to actions (e.g., planned – unplanned). Another important dimension that emerged is the actor's goals (Tarhan et al., 2021). Additionally, using large text data, Thornton & Tamir (2022) revealed six abstract dimensions including *Abstraction*, *Creation*, and *Spiritualism*. Lastly, Kabulska & Lingnau (2022) highlighted the importance of the valence of an action, i.e. the differentiation between pleasant (e.g., sport-related) and unpleasant (e.g., aggressive) actions.

A number of previous studies examined the neural representation of an action space established behaviorally as well as the underlying action dimensions. Tucciarelli et al. (2019) demonstrated that the behaviorally obtained organization of 27 different actions is captured by patterns of activation in the lateral occipitotemporal cortex (LOTC). Regarding the neural representation of action dimensions, Tarhan & Konkle (2020b) revealed five large-scale brain networks associated with action processing: one dedicated to social aspects of actions (such as targeting an agent), and four pertained to a "scale of space" (i.e. near space / far space). Tarhan et al. (2021) proposed a hierarchy in processing actions along the posterior-to-anterior lateral surface of the visual cortex, ranging from information about visual aspects of actions, followed by movement-related information and, lastly, the goals of actions, in line with the results of a recent EEG study by Dima et al. (2023). Furthermore, superior and inferior portions of the lateral occipitotemporal cortex (LOTC) have been shown to carry information about actions along the dimensions sociality and transitivity, respectively (Wurm et al., 2017). Overall, these findings contribute to our understanding of the neural substrates underlying the representation of visually presented actions in the human brain. However, most previous studies either used a small set of pre-selected dimensions, or a rather small stimulus set, which might restrict our understanding of action representation in a real-world

environment (for an exception, see the study by Thornton and Tamir, 2022, which however was based on large-scale text corpora).

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

Materials & Methods

Participants

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

consent was obtained from all participants before the experiment. Participants were rewarded for taking part in the study.

Stimuli

Stimuli consisted of 400 colored images of daily actions that portrayed 100 different daily actions in front of a naturalistic background, such as *running*, *biking*, and *eating* (same as in Kabulska & Lingnau, 2022; see Figure 3.1 for examples), with four different exemplars per action. Stimuli were carefully chosen on the basis of the following criteria: (a) actions were clearly visible, (b) no other distracting actions were depicted in the image; (c) the action was embedded in a natural background. The stimulus set was collected from www.shutterstock.de. All selected images were in landscape orientation and were cropped to 600 x 400 pixels. The full set of images used in the study is shown in Figure B1.

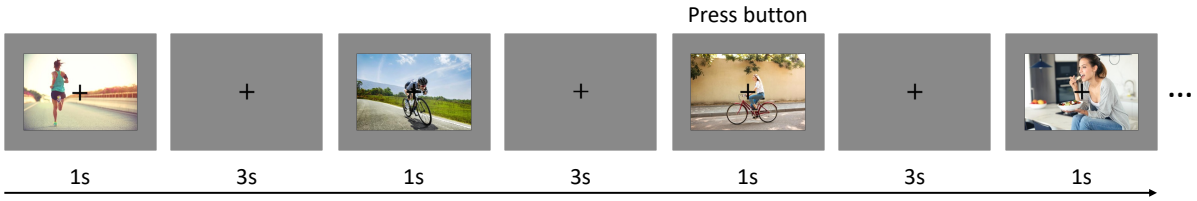


Figure 3.1. Example trial sequence and experimental design. We conducted an fMRI experiment using a rapid event-related design. Each trial consisted of the presentation of an image depicting an action (e.g., running, biking, eating; 1 s) followed by a gray screen (3 s). Throughout the experiment, a central fixation cross was presented on the screen. Participants were instructed to attentively observe the actions while keeping their eyes at fixation and to press a button with their right index finger whenever they saw a repetition of the same action in two subsequent trials (here: biking). Each functional run lasted approx. 9 min and included 100 experimental trials, seven catch trials and 20 null trials (see Methods for details). The whole fMRI session consisted of eight functional runs.

Experimental design and task

We used a rapid event-related design (see Figure 3.1) adopting the design used by Tucciarelli et al. (2019). There were 8 functional runs in total (approx. 9 min each). Each

run started and ended with 12 s fixation period. Each functional run consisted of 100 experimental trials, 20 null trials (4 s long each), and 7 catch trials during which the same action (but not the same exemplar) was shown as during the previous trial. The order of experimental trials was randomized, whereas null trials and catch trials were pseudorandomly interspersed between experimental trials, preventing two consecutive null trials and two consecutive catch trials.

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

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

Post-session questionnaire

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

Data acquisition

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

Data analysis

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

and custom written scripts (available on https://osf.io/efn3w/?view_only=c2b87331de8b45aab23bf182b9921a57).

Preprocessing

The preprocessing of functional data included: (1) removal of the first four volumes; (2) slice time correction; (3) head motion correction (trilinear interpolation) with respect to the first volume of the first run for each participant (using MCFLIRT); (4) BET brain extraction; (5) spatial smoothing with a Gaussian kernel of 5 mm FWHM, (6) high-pass filtering (cutoff frequency of 100 mHz). Note that step (5) was carried out for reliability-based voxel selection (following Magri et al., 2021; Park et al., 2022; Thornton & Tamir, 2023), whereas this step was omitted for representational similarity analysis. Data were linearly registered using FMRIB's Linear Image Registration Tool (FLIRT, Jenkinson et al., 2002; Jenkinson & Smith, 2001, first to each participant's 3D T1-weighted image (7 degrees of freedom) and then to the MNI152 standard brain (12 degrees of freedom).

First-level univariate fMRI analysis

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

Reliability-based voxel selection

To ensure that analyses are performed within a set of voxels that systematically respond during the processing of observed actions, we selected voxels based on their reliability following Tarhan & Konkle (2020a). With this approach, the voxels are considered as reliable when they (a) show systematic differences in activation across the different experimental conditions (in our case, actions), and that (b) show similar activation levels across conditions in different sets (i.e., different exemplars) of the stimuli. To implement this method, we performed the second- and group-level univariate analysis on spatially smoothed data split into odd and even runs (averaged across runs within each split). Separately for each voxel, both at the subject- and group-level, we correlated the obtained beta weights of each condition between the two halves. Based on a group item reliability plot (Figure B2) we decided on a voxel-reliability threshold equal to 0.25. All subsequent analyses were performed within voxels exceeding this threshold.

Representational Similarity Analysis (RSA)

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

RDM was correlated (Pearson correlation) with a selected target RDM (see *RSA: Model RDMs* for details) and the correlation value was assigned to the center voxel of each sphere, resulting in a correlation map.

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

The obtained beta maps were subsequently spatially smoothed with a 5 mm FWHM kernel and entered into a one-sample t test. Statistical significance for the group-level analyses was determined by correcting the beta maps for multiple comparisons using threshold-free cluster enhancement (TFCE, Smith & Nichols (2009) in combination with cluster level correction ($p = 0.05$, one-tailed, $z = 1.65$, 5000 iterations).

We carried out multiple regression RSA for (1) the behavioral action space model, regressing out the low level visual control model and the GIST model and (2) the behavioral action space model, regressing out the low level visual model, the GIST model and the action feature model. In addition, to be able to compare the topography of the areas capturing the

behavioral action space model and the action feature model, we computed a multiple regression RSA for (3) the action feature model (regressing out the low-level visual model and the GIST model.

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

RSA: Model RDMs

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

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

Action feature model We established this model in order to examine to which degree the behavioral action space can be accounted for on the basis of the similarity of a wide range of features. First, using a free feature-listing experiment, we asked N = 40 participants to list at least 5 features per action which resulted in approx. 6000 collected responses describing a set of 100 daily actions (same set as used in the current study). Second, we reduced that list

of features to 44 key action features (e.g., *Upper/Lower limbs*, *Targeting a person/tool*, *Pace*, *Duration*, *Valence*) and, from another set of $N = 273$ participants, obtained feature-based ratings for the same set of 100 actions. The averaged and rescaled ratings were subsequently used to create a feature model by computing pairwise distances between actions (Euclidean distance).

Low-level visual control model We constructed this model to be able to account for low-level visual features. Since representations of objects in early layers of artificial neural networks have been shown to resemble neural activity within early visual cortex (Güçlü & van Gerven, 2015; Lindsay, 2021) we decided to use the first convolutional layer from ResNet50, a deep convolutional network with 50 layers, pretrained on object categories (He et al., 2016) and fine-tuned on 339 action categories from the Moments in Time dataset (Monfort et al., 2020). We fed the ResNet50 model with the 400 action images (100 actions with 4 exemplars each) which we used in the fMRI experiment. Next, we (1) determined the activations within the first convolutional layer and stored them as vectors and (2) averaged the resulting vectors across action exemplars, resulting in 100 activation vectors (one vector per action). (3) Next, we computed 1-Pearson's R correlation for each pairwise combination of vectors resulting in a 100 x 100 RDM. We also created an RDM based on the first layer of AlexNet (Krizhevsky et al., 2017), pretrained on the ImageNet dataset (Russakovsky et al., 2015), which is another frequently used convolutional neural network (e.g., Kietzmann et al., 2019; Lee Masson & Isik, 2021, see Figures B3 and B4 for the RSA results).

GIST model To account for the spatial structure of the scenes presented in the stimuli, we employed the GIST model (Torralba & Oliva, 2001). This computational model extracts information about scenes based on several dimensions, such as naturalness,

openness, and roughness, and defines stimuli as similar if the semantic categories of the scenes share similarities (e.g., highways and streets). We generated GIST descriptors for all 400 action images (100 actions with 4 exemplars each) using the default parameters for the number of orientations at which the Gabor filters are applied, and the filter used to reduce illumination effects of input images. Subsequently, we averaged the descriptors across action exemplars, resulting in a set of 100 GIST descriptors, one for each action. To construct the GIST RDM, we computed pairwise distances between the actions using the Euclidean distance metric.

Principal Component Analysis (PCA)

The action feature model contained information about all 44 action features reported by Kabulska & Lingnau (2022) (see *section RSA: Model RDMs* for details). To be able to reduce this large number of features to a smaller set of dimensions, we conducted a principal component analysis (PCA) on the 44 feature-based ratings of 100 actions (same ratings as used to create the feature RDM, see *RSA: Behavioral RDMs*). The components were derived using varimax rotation which maintains orthogonality between them. We identified 11 components with eigenvalues greater than one (Table B1). Based on the scree plot combined with the "elbow method" (Figure B5), we chose eight dimensions, accounting for approx. 64.1% in total of the variability in the feature ratings.

RSA with principal components (PCs)

Subsequently we wanted to determine which brain regions best capture these dimensions. To address this question, we performed a regression-based RSA, separately for each of the eight dimensions while regressing out the low-level visual control model and the GIST model (see *section RSA: Model RDMs* for details). In order to construct the dimension-

based RDMs, the first step involved multiplying the action feature ratings (obtained in Kabulska & Lingnau, 2022) by loadings on a given dimension. Subsequently, we took the resulting 100 vectors (one per action) of weighted feature ratings and computed pairwise distances (*Euclidean* distance) between them. Prior to conducting multiple regression RSA, we computed the Variance Inflation Factor. The VIFs were below 4 for all the models indicating low multicollinearity between them (PC1: 2.34; PC2: 2.91; PC3: 3.16, PC4: 3.46, PC5: 3.16, PC6: 2,40, PC7: 3.33, PC8: 2.30, low-level visual model: 1.04, GIST model: 1.15).

Winner-takes-all with PCs

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

Results

Behavioral results

We performed an fMRI experiment with 100 daily actions (four exemplars per action; see Figure B1 for a complete overview of all stimuli) using a rapid event-related design (see *Methods,* section *Experimental design and task* for details). Mean reaction time for correct responses was 959.84 ms (± 43.60 ms SEM). Participants identified catch trials with a mean

error rate of 24.73% (±2.74% SEM), corresponding to approx. 14 out of 56 catch trials per participant.

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

Individual error rates and answers provided in the post-session questionnaire are provided in Table B2.

Reliability map

Following Tarhan & Konkle (2020a), we used a reliability-based voxel selection (see Methods section for details). This analysis revealed voxels with high reliability in occipital brain areas, covering both ventral and dorsal visual streams, and part of the parietal lobe (see Figure B2), whereas voxel reliability was lower in frontal areas. All subsequent analyses were performed within the reliability map.

Searchlight-based RSA

To determine which brain areas reflect the behavioral action space, corresponding to the categorical organization obtained from the multi-arrangement task while accounting for variability due to low- level visual features, mid-level scene-related features, and high-level

action features, we performed a multiple regression searchlight-based RSA (see Methods, section *Representational similarity analysis* for details). Moreover, to determine the spatial correspondence between the regions capturing the behavioral action space and the regions capturing a feature-based organization, we carried out an additional searchlight-based RSA on the action feature model.

The resulting searchlight maps for the behavioral action space model (while regressing out the low-level visual control model and the GIST model) revealed significant correlations between neural patterns of activation and the action space model in bilateral occipitotemporal and fusiform cortex as well as in small portions of the superior parietal lobe (Figure 3.2A). Additionally regressing out the action feature model resulted in a qualitatively similar, but less widespread map (Figure 3.2B) that was limited to bilateral occipitotemporal and temporal occipital fusiform cortex.

The action feature model was captured by patterns of activation in a comparable, but slightly more widespread set of regions, including the bilateral occipitotemporal and fusiform cortex as well as the superior parietal lobe (see Figure 3.2C).

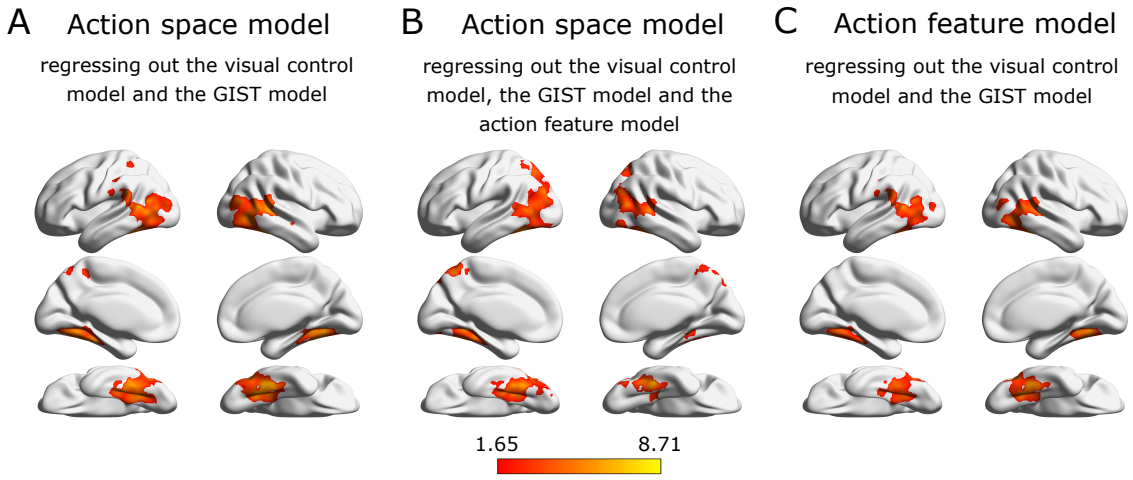


Figure 3.2. Results of the group-level searchlight-based RSA for: **(A)** the behavioral action space model (regressing out the low- level visual control model and the GIST model); **(B)** the behavioral action space model (regressing out the low-level visual control model, the GIST model and the action feature model; **(C)** the action feature model (regressing out the low-level visual control model and the GIST model). Statistical maps show t-values thresholded at a z-score of 1.65, corresponding to $p < 0.05$ (one-tailed), corrected for multiple comparisons (TFCE, $p < 0.05$, 5000 Monte Carlo permutations).

Principle Component Analysis (PCA)

PCA on the 44 feature ratings for the 100 different actions revealed eight components that explained 64.1% of the variance (see *Methods* section for details). We labeled these components on the basis of the features belonging to each component (see Table B1). The first component that explained most of the variance (21.6%) we labeled *General movements* due to high positive loadings for a variety of features, such as *lower limb movements*, *change of location*, *use of force* and negative loadings on *no movement* and *sitting* features. The second component was mostly related to different arm movement kinematics (e.g., rotating, sweeping, circular) and therefore labeled Arm movement kinematics. The third component was related to features related to the *goal/ target object* and features related to the *arm and hand* and therefore labeled *Goal-directedness*. The subsequent components were associated with features related to the *Context* of the actions (*Indoor, outdoor, season-dependence*), the *Posture* of the agents performing the actions, *Contact with others* (i.e., whether or not the action involved a direct or indirect contact with another person), and Object-directedness (i.e., whether or not the action targeted a manipulable object). The last component referred to the features *noise, harm* and *negative valence* and thus was labeled *Negative Emotions*.

RSA on dimensions resulting from principal component analysis

To determine which brain areas represent the information captured by each of the dimensions resulting from PCA on the feature ratings, we conducted a searchlight-based

RSA, separately for each of the eight dimensions, regressing out the low-level visual control model and the GIST model. The results of this analysis are shown in Figure 3.3. For the dimension *General movements*, that explains the largest amount of variance (21.63%), we identified significant clusters in bilateral temporal occipital fusiform cortices and lateral occipital cortices, extending towards the superior parietal lobules. The dimension *Goal-directedness* was captured by clusters in the left inferior temporal gyrus, bilateral temporal fusiform cortices and lateral occipital cortices. For the dimension *Context*, we obtained clusters in bilateral lateral occipital cortices and the left temporal occipital fusiform cortex. Clusters in bilateral temporal occipital fusiform cortices and superior parietal lobules corresponded to the dimension *Posture.* The dimension *Contact with others* was associated with clusters in the left lateral occipital cortex (superior and inferior division) and a smaller cluster in the right lateral occipital cortex (inferior division) as well as a cluster in the left temporal occipital fusiform cortex. The dimension *Object-directedness* showed a significant correlation with activation patterns within clusters in bilateral temporal occipital fusiform cortices and lateral occipital cortices (superior division). In sum, this analysis revealed a substantial degree of overlap between the different dimensions along the ventral visual stream and the superior parietal lobe, in particular for the dimensions *General movements*, *Context*, *Posture*, and *Object-directedness*. By contrast, the dimensions *Goal-directedness* and *Contact with others* were associated with more circumscribed clusters of voxels. To explore the spatial arrangement of these dimensions, we carried out a Winner-Takes-All analysis (see next section).

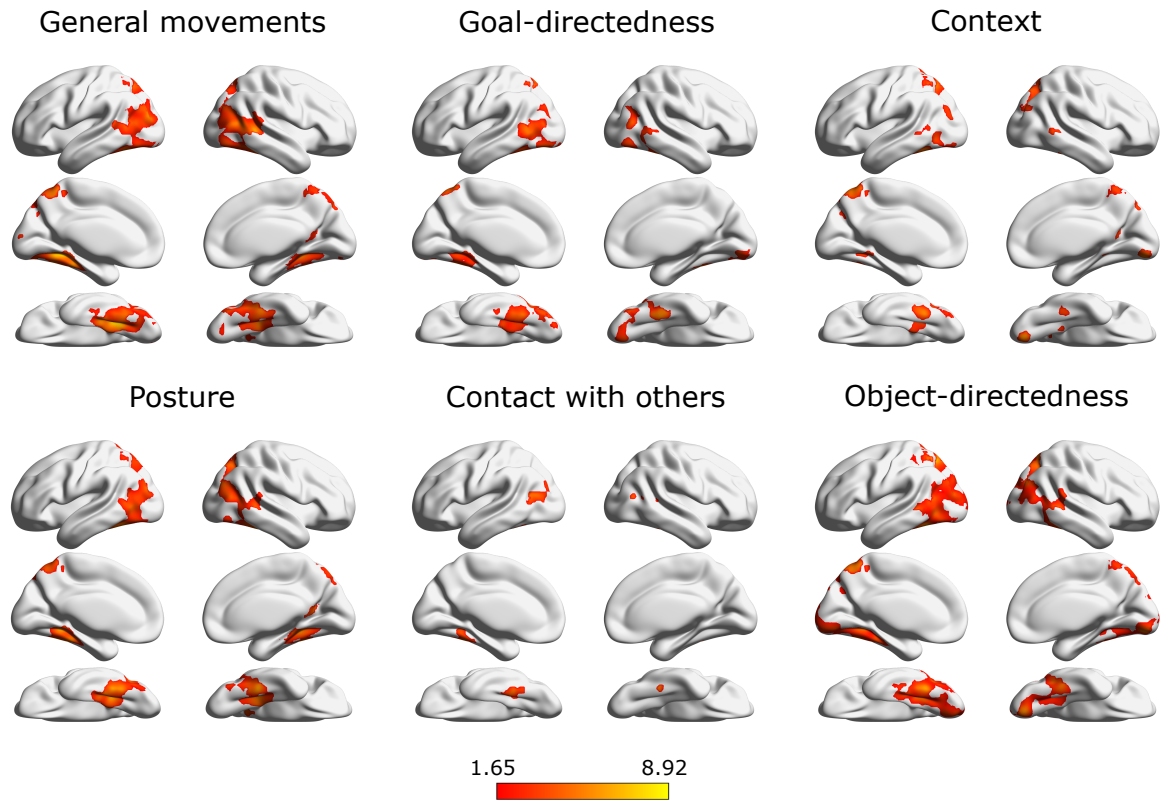


Figure 3.3. Results of the searchlight RSA, carried out separately for each of the eight dimensions (regressing out the low-level visual control model and the GIST model). Six out of eight dimensions showed a significant correlation with neural activation patterns after correction for multiple comparisons (TFCE, $p < 0.05$, 5000 Monte Carlo permutations). Statistical maps show t-maps thresholded using TFCE at z-score of 1.65. The remaining dimensions, namely Arm movement kinematics and Negative Emotions, did not survive the correction.

Winner-takes-all map

To explore clusters of voxels with a preference for individual principal components, we calculated a winner-takes-all map (see Methods for details). Note that since we provide no additional statistics for these maps, this analysis merely serves as an additional visualization of the results shown in Figure 3.3. That said, the winner-takes-all analysis highlights multiple clusters displaying the highest correlation with the *General movements* dimension in a prominent portion of the right LOTC as well as the left dorsal LOTC (Figure 3.4, blue). The *Goal-directedness* dimension showed the highest correlations with patterns

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

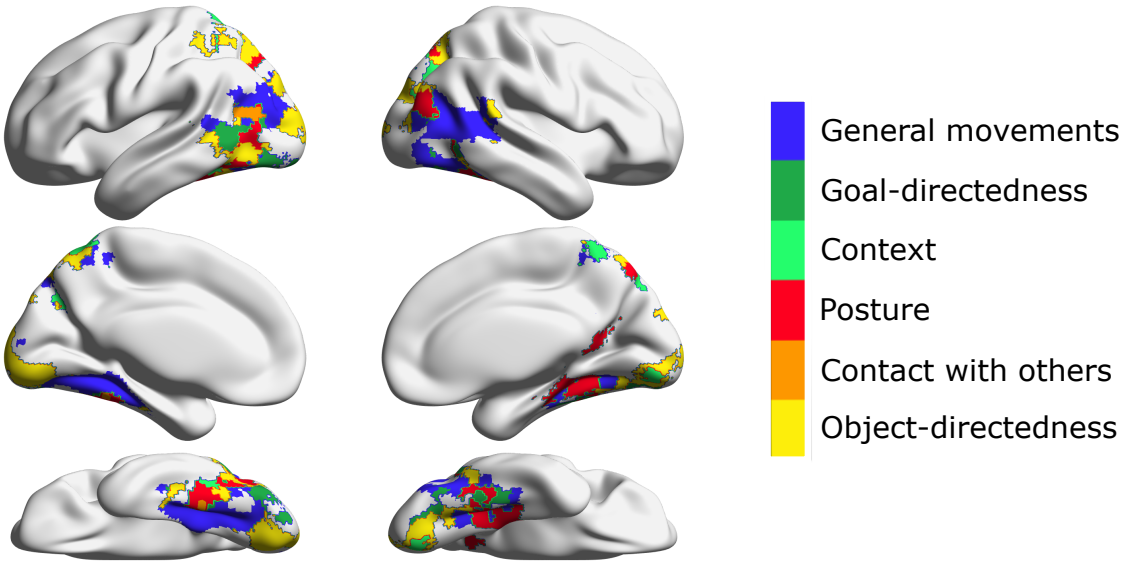


Figure 3.4. Results of the winner-takes-all analysis with maps for six different dimensions obtained from the searchlight-based RSA (see Figure 3.3; see also Tarhan et al., 2021). Each voxel was assigned a color corresponding to the dimension that showed the strongest correlation (see legend on the right for the assignment of colors to each dimension).

Discussion

In this study, we investigated the neural architecture underlying the organization of a wide range of observed actions. For that purpose, we conducted an fMRI experiment with

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

Neural representation of the behavioral action space

Our results are in line with the results by Tucciarelli et al. (2019) who reported that a behaviorally determined action space assumed to capture the semantic similarity of a set of 27 actions is reflected by patterns of activation in the LOTC. To account for additional action components that might covary with the behavioral action space, Tucciarelli et al. (2019) regressed out nine additional models capturing diverse aspects, including the similarity of objects, body parts and the distance between the observer and the actor. These additional action components partially overlapped with the cluster capturing the behavioral action space. The current study advances the findings of Tucciarelli et al. (2019) in two important ways. First, we demonstrated that the results of Tucciarelli et al. (2019) generalizes to a significantly wider range of actions (i.e., 100 instead of 27 actions). Second, our whole-brain searchlight RSA revealed the highest similarity between the behavioral action space model

and patterns of activation in dorsal and ventral portions of the LOTC, even after regressing out (a) a low-level visual control model derived from the first convolutional layer of a neural network (ResNet50), (b) mid-level spatial information of the scenes captured by the GIST model and (3) high-level information related to 44 different action features. Together, our results provide an important extension of a growing number of studies suggesting that the LOTC gathers not only perceptual evidence on the basis of action features, but also more conceptual action aspects (Hafri et al., 2017; Oosterhof et al., 2010, 2012; Wurm et al., 2015; Zhuang et al., 2023; for reviews, see Wurm & Caramazza, 2022 and Lingnau & Downing, 2015).

Dimensions underlying the organization of high-level action features

The action feature model used in the current study is based on ratings obtained for 44 high-level action features carried out for 100 different actions (Kabulska & Lingnau, 2022). PCA revealed eight dimensions underlying the organization of these features (*General movements*, *Arm movement kinematics*, *Goal-directedness*, *Context*, *Posture*, *Contact with others*, *Object-directedness*, and *Negative Emotions*). These dimensions align remarkably well with those previously proposed and examined. Movements, Posture and Goal-Directedness are undeniably crucial aspects of actions, as they are sufficient to identify a wide range of actions (Johansson, 1973, see e.g., Beauchamp et al., 2003; Grossman et al., 2000; Papeo et al., 2017 for studies with point-light displays). Moreover, numerous actions involve the use of tools (e.g., Buxbaum, 2001; Chao & Martin, 2000; Watson & Buxbaum, 2014) or are directed towards specific objects (e.g., Bach et al., 2014; Wurm et al., 2017). Additionally, contact with other people and social actions play a vital role in our daily lives, enabling successful communication with others, while comprehending and interpreting

emotions is a crucial part in this process (e.g., Isik et al., 2017; Papeo, 2020; Poyo Solanas et al., 2020). Moreover, in contrast to objects that can be understood in isolation, understanding actions involves information that extends beyond the body itself (e.g., information regarding the scene; see Wurm & Schubotz, 2012; Wurm & Schubotz, 2017). Hence, using a wide range of actions and action features, our study revealed a set of dimensions that have been proposed in previous studies but, to the best of our knowledge, have not been collectively investigated before. Our approach allowed us to determine the degree to which the neural representation of these high-level action features contributes to the neural representation of the behavioral action space model (see section *Neural representation of the behavioral action space*), and to examine to which degree the topographies corresponding to the neural representation of the different action dimensions spatially overlap with the neural territory capturing the behavioral action space, which we discuss in more detail in the following paragraphs.

Neural representation of action dimensions

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

We found that activation patterns in the dorsal LOTC showed the highest similarity with the dimension labeled *Contact with others*, while activation patterns in the ventral

LOTC showed the highest similarity with the dimension labeled *Object-directedness*. These results are in line with several recent studies indicating an animate-inanimate organization of dorsal and ventral portions of the LOTC (e.g., Lingnau & Downing, 2015; Wurm et al., 2017; Wurm & Caramazza, 2022). More precisely, it has been shown that dorsal portions of the LOTC have a preference for animate things (Chao et al., 1999; He et al., 2020), body parts (e.g., Downing et al., 2001), movements (Beauchamp et al., 2003), and person-directed actions (Wurm & Caramazza, 2019a; Wurm & Caramazza, 2019b), while ventral portions have a preference for inanimate things (Chao et al., 1999; He et al., 2020), action-specific tool motion (Beauchamp et al., 2002; Beauchamp et al., 2003), and actions involving objects (e.g., Wurm et al., 2017; Wurm & Caramazza, 2019a; see Wurm & Caramazza, 2022 for a recent review on the animate-inanimate organization). Note that the clusters representing the dimensions *Contact with others* and *Object-directedness* obtained in the current study (see Figures 3.3 and 3.4) are well aligned with the clusters showing a high similarity with the *Sociality and the Transitivity model* reported by Wurm et al. (2017).

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

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

High-dimensional spaces in the LOTC

The regions capturing the higher-level action feature model and the underlying dimensions strongly overlapped with those capturing the behavioral action space model. The overlap encompassed the LOTC, indicating the pivotal role of this region in representing diverse information about actions (see also Lingnau & Downing, 2015; Wurm & Caramazza, 2022). This raises the question according to which principles this diverse information is represented along the LOTC. One option, though speculative, is the idea put forward by the work by Graziano & Aflalo (2007), indicating that the motor cortex is organized along multiple dimensions – such as somatotopic information and information about different types of limb movements. As suggested by Graziano & Aflalo (2007), this structure is not limited to the motor cortex and may extend to any region that processes multidimensional and complex knowledge. Whereas future studies are required to test these predictions more systematically, the results obtained in the current study are compatible with the view that this principle also holds for the LOTC.

Conclusion

Our results provide an important extension of previous studies, suggesting that action representations within the LOTC generalize to a wide range of actions. Moreover, our results suggest and that the areas capturing this representational space overlap with dimensions corresponding to high-level action features, in line with the idea that the LOTC, like other areas of the brain such as the motor cortex and parietal cortex, is organized along multiple dimensions (see also Graziano & Aflalo, 2007; Lingnau & Downing, 2015; Wurm & Caramazza, 2022).

CHAPTER 4: STUDY 3

„NEURAL UNDERPINNINGS OF ACTION CATEGORIES”

The study has been pre-registered on OSF (https://osf.io/9d4sm)

Abstract

Throughout our lives, we encounter a diverse range of actions we instinctively categorize to help us make sense of the world. Previous research has revealed that human participants tend to sort actions into clusters corresponding to distinct categories such as Food-related, Social/Communicative, and Locomotion actions. In this study, we aimed to investigate the neural basis of action category processing using fMRI. During MRI sessions, participants viewed static images of actions from four categories: *Communication, Grooming, Ingestion*, and *Locomotion.* We obtained significant decoding accuracies of action categories within regions of the Action Observation Network (AON), with the highest decoding accuracies in bilateral lateral occipitotemporal cortex (LOTC). Moreover, we investigated regions outside of the AON, unique for each action category. Finally, functional connectivity analysis revealed that some of the categories, e.g., *Communication* and *Grooming*, have distinct connectivity patterns between regions belonging to both the AON and the category-specific brain maps. The weights obtained from the SVM classifier shed light on the key regions crucial for decoding these category pairs. Overall, our findings provide insights into the neural underpinnings of action categories, capturing both the underlying neural activity and connectivity patterns.

Introduction

Humans tend to categorize the surrounding to make sense of the world. For example, we differentiate between animate and inanimate objects or, on a finer level, between different animal species (Connolly et al., 2012). Such distinctions can be based on a range of properties, including semantic information, such as animacy, as well as visual characteristics such as shape and size (Bracci et al., 2019). Taking inspiration from research on object recognition, there has been a growing interest in the categorization of actions. Studies employing multi-arrangement tasks, combined with dimensionality reduction methods such as clustering analysis and principal component analysis revealed, have identified several semantic action categories, including Locomotion, Social/Communicative actions, and Food-related actions (Tucciarelli et al, 2019; Kabulska & Lingnau, 2022). However, the neural underpinnings of action categories are not well understood yet.

Studies on monkeys have provided insights into the neural correlations of action observation. A class of neurons, termed “mirror neurons”, was discovered that activates both when a monkey executes an action and when it observes another individual executing the action (di Pellegrino et al., 1992; Gallese et al., 1996; Rizzolatti et al., 1996). The mirror neuron system is primarly located in the premotor cortex, more precisely area F5, and the inferior parietal lobe (IPL) (Fogassi et al., 2005; Rizzolatti et al., 2001). Additionally, the superior temporal sulcus (STS) has been shown to play a role in observing biological actions (Oram & Perrett, 1994; Rizzolatti et al., 2001). Connectivity analysis based on monkey tracer studies showed that these brain regions are connected, i.e., the STS is reciprocally connected to the IPL (Seltzer & Pandya, 1994) and to the F5 via the IPL (Petrides & Pandya, 1984; Rizzolatti et al., 2001). An fMRI study confirmed these results, showing connectivity

between the STS, parietal areas and the ventral premotor cortex while monkeys observe actions (Nelissen et al., 2011). In humans, a similar network, the Action Observation Network (AON), has been identified, encompassing occipito-temporal, parietal, and premotor areas (Caspers et al., 2010; Hardwick et al., 2018). As has been reported in monkey studies, research has shown anatomical connections between the premotor and parietal areas (Rushworth et al., 2006) as well as throughout all three regions of the AON (Caspers et al., 2010; Urgen, 2020). Furthermore, studies have demonstrated functional connectivity between the AON regions during movement observation (Kilner et al., 2007b, 2007a; Nishitani & Hari, 2002).

However, the specific roles of these regions remain a topic of debate. Some studies suggests that the inferior parietal cortex is central to high-level action understanding, as it generalizes across effectors (Cattaneo et al., 2010), the kinematic parameters (Hamilton & Grafton 2006, 2007) and trajectory of an action (Hamilton & Grafton, 2008). When it comes to action categories, several recent studies support the idea that regions of the parietal cortex are involved in observation of different classes of actions. For example, it has been shown that the IPL was activated when observing goal-directed actions such as dragging and dropping, regardless of the body part used (Jastorff et al., 2010). Another study showed that observation of three different classes of actions, namely manipulation, locomotion, and climbing, evoked activity in parts of the superior parietal lobule (SPL) (Abdollahi et al., 2013). The parietal cortex was also involved in observing upper limb actions, such as grasping a person and rubbing a skin (Ferri et al., 2015) as well as actions of vocal communication and oral manipulation (Corbo & Orban, 2017), and indirect communication, such as writing and drawing (Urgen & Orban, 2021).

Yet, recent findings indicate that not only IPL but also the lateral occipitotemporal cortex (LOTC) contribute to action understanding (Lingnau & Downing, 2015). Research has shown that the LOTC can generalize across viewpoints (Oosterhof et al., 2010, 2012), target objects (Wurm et al., 2015), involved kinematics (Wurm & Lingnau, 2015), and it carries semantic information about actions (Tucciarelli et al., 2019). When it comes to action categories, the LOTC can differentiate between social and object-directed action categories (Wurm et al., 2017).

Despite extensive research on action processing and the established role of the AON in action observation, there is still no consensus regarding the regions crucial for high-level understanding of observed action categories. Previous studies examining the action categories have mostly focused on the parietal cortex and often overlooked the role of high visual areas in understanding action categories. Given the increasing evidence of the importance of the LOTC in action understanding, we aimed to investigate its role in processing information about action categories and compare with neural activation in the other AON regions. Moreover, we wanted to identify brain regions outside the AON that are involved in processing unique information to specific action categories. Based on recent findings from our behavioral study, which showed that each of the investigated action categories engages a unique combination of features (Kabulska & Lingnau, 2022), we expected to find regions that carry information about category-specific features. Finally, given the evidence of anatomical and functional connectivity between the AON regions, we were interested in the functional connectivity between the set of brain regions comprising the AON and the category-specific regions engaged in observing action categories. More precisely, we wanted to investigate whether the action categories could be differentiated based on their distinct connectivity patterns between these regions, and, if so, which pairs of

regions contribute to successful decoding. While we anticipated that the AON regions alone might not show differences in decoding, as they are activated for all action categories, we believed that distinctions might emerge when considering regions outside the AON that are specific to certain categories.

To test these hypotheses, we conducted an fMRI study using static images of actions from four categories: *Communication, Grooming, Ingestion*, and *Locomotion.* As a first step, we performed a multivariate pattern analysis (MVPA) based on activity patterns in the AON brain regions to investigate whether the action categories could be decoded within these areas. Additionally, we explored whether there were any statistically significant differences between the AON regions in terms of decoding action category-related information. Subsequently, we carried out a conjunction analysis to identify unique activity maps evoked by specific action categories. We further investigated the connectivity within the identified regions and the AON regions. Specifically, we analyzed time series within these regions and conducted a functional connectivity analysis for pairwise action category decoding. Using Support Vector Machine (SVM) weights, we determined which pairs of brain regions played a pivotal role in successful decoding of categories based on connectivity patterns. The hypotheses and analysis methods have been pre-registered on OSF (https://osf.io/9d4sm).

Materials and methods

Participants

Thirty right-handed healthy adults (19 females; mean age: 25; age range: 19 - 39) participated in the study. Data of one subject were not included in the data analysis due to excessive head motion (translation/rotation bigger than 3 mm). All participants had normal or corrected-to-normal vision and no history of neurological or psychiatric disease.

Participants gave written informed consent before participation in the experiment. The experimental procedures were approved by the ethics committee at the University of Regensburg. Participants were rewarded for taking part in the study.

Stimuli

Stimuli were 160 colored images of actions belonging to four action categories: *Communication*, *Grooming*, *Ingestion*, and *Locomotion*. Each action category consisted of five basic-level actions (e.g., *arguing*, *pointing*, *talking*, *thumbs up*, and *waving* for the category *Communication*, see Table C1 for the full list of action stimuli and Figure C1 for the corresponding images). The four action categories were chosen from the sets of action categories reported in Kabulska & Lingnau (2022) and Tucciarelli et al., (2019), both identified empirically through an inverse multidimensional scaling experiment (Kriegeskorte & Mur, 2012). The reason for selecting these four categories is that they address key aspects of daily life and have been also used in previous neuroimaging studies (Abdollahi et al., 2013; Corbo & Orban, 2017; Hafri et al., 2017; Tarhan & Konkle, 2020b). For each category, we selected actions by taking into account actions that (1) are typical for a given category, (2) can be depictable, and (3) are easily recognizable. Moreover, we tried to diversify the actions within each category by varying the objects used, the targeted body parts, and the actors' postures, among other factors, to ensure that understanding the action categories would require generalizing across these aspects. Additionally, we ensured that the background is natural and does not interfere with the actions (i.e., the background should not catch the participants' attention).

For each of the twenty basic-level actions, we chose eight different pictures (referred to as 'action examples' throughout the remainder of this paper). For each action, four of the

pictures showed a single person performing an action, whereas the other four images showed either multiple people performing that action or, in case we could not find a proper picture, a single person performing that action with other people sitting or standing nearby or in the background. The reason for that was to ensure that the differences between the categories (e.g., *Ingestion* vs *Communication*) are not due to the number of people shown on the images, given that the category *Communication* by definition involves more than one person, whereas the remaining three categories can be performed in the absence of another person. The stimulus set was collected from http://www.shutterstock.de. All selected images were in landscape orientation and were cropped to 480 x 320 pixels.

Experimental design

Stimuli were presented in a mixed design, with the action category blocked, and basic level actions randomized within a block (see Figure 4.1). Each block consisted of 10 trials of five different basic level actions from the same category. Each trial within a block consisted of an action image (1 s) followed by a fixation period (2 s). Half of the trials within a block consisted of images showing a single person performing an action, while the other half of trials used images of several people. The fixation period between blocks lasted 10 s.

Separately for each participant, the whole set of 160 images was divided into two sets (each containing an equal number of single-person and multiple-people images). One of the sets was presented in runs with an odd number, and the other set was presented in runs with an even number. This way we wanted to ensure that the same action examples do not appear in consecutive runs, and that they are spread evenly across the runs. Each of the four action category was presented twice within a run, for a total of 8 blocks per run.

There were eight runs in total. Each run started and ended with a 13 s fixation period. Throughout the scanning session, each of the 20 different basic level actions was presented 32 times (four trials per run x eight runs), and each action example was presented four times. The stimuli were back-projected onto a screen (resolution 1024 x 768 at 60 Hz; viewing distance 106 cm; 10.38 x 6.82 degree of visual angle) and viewed via a mirror mounted on the radiofrequency (RF) coil. Stimulus presentation and response collection was controlled via *A Simple Framework* (Schwarzbach, 2011), a toolbox based on the MATLAB Psychtoolbox-3 for Windows (Brainard, 1997)

The order of blocks was counterbalanced across participants and within each scanning session. Blocks within runs were ordered based on the Latin square design: starting from action category 1 (i.e., 1, 2, 3, 4), or from category 4 (i.e., 4, 3, 2, 1) for participants with odd or even participant numbers, respectively. Trials within blocks were ordered randomly.

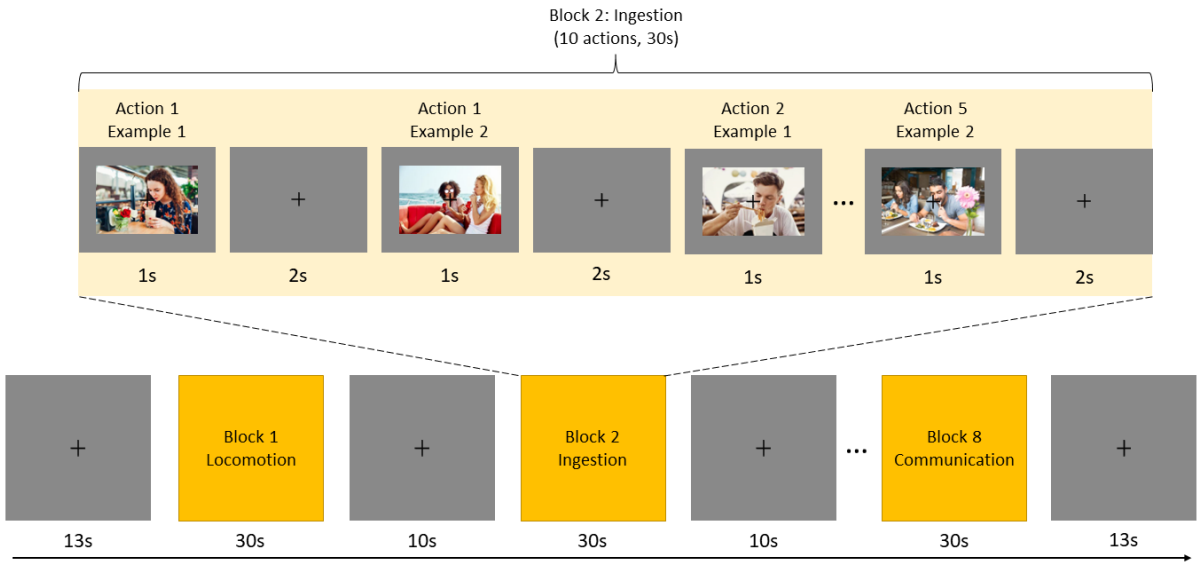


Figure 4.1. Experimental design and an example run of the fMRI experiment. The experiment consisted of eight runs that started and ended with a 13 s fixation period. Within each run there were eight blocks belonging to one of the four action categories (*Communication*, *Grooming*, *Ingestion*, *Locomotion*) separated by 10 s

fixation. Blocks consisted of 10 trials. During each trial, one of the five basic level actions belonging to a given action category (in the example shown above: *Ingestion*) was depicted via a static image. The type of basic level action and the example was randomly assigned within a block. See section 'Experimental design' for details.

Each run contained either one or two catch blocks that contained catch trials, i.e., images of actions belonging to another category, which served as targets in the catch trial task (see next paragraph). The number of catch trials within a catch block varied between one and two. The first catch trial in a catch block was always preceded by at least three images of actions belonging to the same category such that the participant knew which category is being presented. In total, there were twelve catch blocks within the whole scanning session (half of the runs contained one catch block, whereas the other half contained two catch blocks). In case of two catch blocks in a run, the catch blocks were interspersed with normal blocks. Otherwise, regardless of the number of catch blocks per run, the position of catch blocks in relation to normal blocks was not restricted. Catch blocks were later used to calculate the accuracy of identifying a catch trial and the response time (see Results: Behavioral data analysis) but were not included in neuroimaging data analysis.

Task

To ensure that participants paid attention to the actions, they were instructed to attentively watch the images, keep their eyes at the fixation cross, and focus on action categories within each block. They were asked to press a button with their right index finger whenever they spotted an action belonging to another category (i.e., a catch trial within a catch block). Before the experiment, participants completed a 3-minute-long practice version of the experiment to ensure that they understand the task.

Data acquisition

Functional and structural data were collected using a 3T Siemens Prisma MRI scanner at the University of Regensburg and a 64-channel RF head coil. Functional images were acquired with a T2*-weighted gradient echoplanar imaging (EPI) sequence (voxel volume: 2,5 x 2,5 x 2,5 mm^3; 60 axial slices that cover the whole brain; repetition time (TR): 2 s, echo time (TE): 30 s, flip angle (FA): 75°, field of view (FoV): 192 mm, matrix size: 96 x 96, 188 or 208 volumes per run). Structural T1-weighted images were acquired in the middle of the scanning session with an MPRAGE sequence (voxel volume: 1 x 1 x 1 mm^3, 160 axial slices, TR: 1910 ms, TE: 3.67 s, FA: 9°, matrix size: 256 x 256).

Data analysis

Data preprocessing and univariate analyses were performed using FEAT (FMRI Expert Analysis Tool; (Woolrich et al., 2001, 2004) which is a part of FSL (FMRIB's Software Library, Jenkinson et al., 2012). FSL was also used for the extraction of information about the clusters of the statistical maps (command: *cluster*), creating ROIs, smoothing the maps, performing high-pass filtering (command: *fslmaths*) and extracting time series (command: *fslmeants*). Data de-noising (see section 'Preprocessing' for details) was performed using Python scripts. All further analyses were conducted in MATLAB (The MathWorks Inc.) using specific toolboxes mentioned below and custom written scripts.

Overview of the analyses

The questions posed in our study are multifaceted. Firstly, we anticipated that observing action categories would evoke activation within the regions of the Action Observation Network (AON; Casper et al., 2010). To explore this, we performed a univariate analysis using separate categories as regressors. Secondly, we hypothesized that observing

action categories would engage additional brain regions specific for processing each particular category. To investigate it, we carried out a conjunction analysis. Next, we expected that each action category would be represented by unique patterns of brain activation. To examine this, we performed a multivariate pattern analysis (MVPA), decoding action categories based on patterns of neural activations evoked by each action separately. Lastly, we expected that each action category would be represented by unique patterns of functional connectivity between AON areas, and potentially additional areas specific to each category. For this analysis, we performed de-noising of fMRI data and used the time series within the selected regions of interest (ROIs).

Preprocessing

The preprocessing of functional data included: (a) removal of the first four volumes; (b) slice time correction; (c) head motion correction (trilinear interpolation) with respect to the first volume of the first run for each participant (using MCFLIRT); (d) BET brain extraction; and (e) high-pass filtering (cutoff frequency of 100 mHz). Data were linearly registered using FMRIB's Linear Image Registration Tool (FLIRT, (Jenkinson et al., 2002; Jenkinson & Smith, 2001)), first to each participant's 3D T1-weighted image (7 degrees of freedom) and then to the MNI152 standard brain (12 degrees of freedom). We applied spatial smoothing with a Gaussian kernel of 5 mm FWHM for the data used for identifying ROI peaks and for the conjunction analysis, however not for the data used for the MVPA.

Preprocessing of the data used for the functional connectivity analysis included performing (a-d) points mentioned above, subsequently co-registration of the data to participant's 3D T1-weighted image (FLIRT), and then to the MNI152 standard space using both FLIRT and non-linear registration (FNIRT (Andersson et al., 2010), Warp resolution of

10 mm). In the next step, we performed non-aggressive denoising with ICA-AROMA (Independent Component Analysis-based Automatic Removal Of Motion Artifacts; Pruim et al., 2015 a,b) ran via Python. In short, ICA-AROMA allows identifying motion-related components and removing them from the data. Output images from ICA-AROMA were subsequently high-pass filtered using *fslmaths* command (cutoff frequency of 100 mHz, same as e.g., Agosta et al. (2018) and Yang et al. (2017). Similar as Heinzle et al. (2012), we did not apply spatial smoothing, which, although commonly used to increase the signal-to-noise ratio in standard GLM analysis, has been shown not to be necessary for ROI-level functional brain network analysis (Alakörkkö et al., 2017).

Output maps from the preprocessing analyses were subsequently fit to a general linear model (GLM) in the univariate analyses.

First-level univariate fMRI analyses

We performed three first-level univariate analyses: one using separate categories as regressors of interest on data that were not de-noised, second using separate categories as regressors of interest on data after de-noising, and the third using single actions as regressors of interest on data that were not de-noised. Data from the first analysis were used to identify peaks of ROIs and for the conjunction analysis. Data from the second analysis were used in the functional connectivity analysis. Data from the third analysis were used for the MVPA – by using 20 actions as regressors, instead of four categories, we wanted to ensure to have a sufficient number of training and testing data.

Separate categories as regressors of interest:

There were four regressors of interests (corresponding to the four action categories) and seven regressors of no interest (six resulting from 3D motion correction - x, y, z

translation and rotation - obtained during preprocessing and one regressor corresponding to the catch blocks). Each regressor of interest was convolved with a standard dual gamma hemodynamic response function (Friston et al., 1998). Each action category was modeled as a block lasting 30 s, starting from the onset of the first trial to the offset of the last trial of that block.

To identify ROIs and to conduct the conjunction analysis, we subsequently performed the second- and group-level univariate analyses using the obtained β maps. For the functional connectivity analysis, we used time series extracted from individual runs.

Single actions as regressors of interest:

There were 20 regressors of interest (corresponding to 4 categories x 5 basic-level actions each = 20 single actions) and seven regressors of no interest (same as in the analysis described above). Each single action was modeled as a trial lasting 1 s (equal to the duration of the shown image). Each regressor of interest was convolved with a standard dual gamma hemodynamic response function. The first level analysis resulted in β estimates for each action separately. The β estimates were subsequently transformed to t-values that were used as input into the MVPA. We applied spatial smoothing with a Gaussian kernel of 5 mm FWHM on accuracy maps obtained from the MVP analysis (before conducting multiple comparison correction).

Second- and group-level univariate analyses

The second and the group-level univariate analyses were performed only for the data used for identifying ROIs and for performing the conjunction analysis. The group-level GLM-based analysis resulted in statistical maps corrected for multiple comparisons at $p <$

0.05 (thresholded non-parametrically at $Z > 3.1$) using cluster-based correction (Worsley, 2001).

ROI definition

For the MVPA and the functional connectivity, we focused on regions that are known to be recruited during action observation, specifically, the LOTC, IPL, and ventral premotor cortex (PMv). Following previous studies (e.g., Oosterhof et al., 2010, 2012; Wurm et al., 2015; Wurm & Lingnau, 2015), we used a combination of anatomical and functional criteria. In a first step, we defined anatomical ROIs based on two brain atlases: Harvard-Oxford probabilistic atlas (LOTC) and Jülich atlas (IPL, PMv) within FSL with a Maximal Probability Threshold of 30%. Second, the group peak for each ROI was chosen by taking the group level statistical map (all categories vs. baseline, see section *Separate categories as regressors of interest* above) and finding the voxel with the highest t-score within the anatomical ROI. Since we failed to obtain significant β estimates within the anatomically defined IPL and PMv, we selected the group peaks within anatomical ROIs in their vicinity, i.e., anterior intraparietal sulcus (aIPS) (from the Jülich atlas) and inferior frontal gyrus (IFG) (from the Harvard-Oxford atlas). In the third step, separately for each participant we identified individual ROIs as 10 mm radius spheres centered around the activation peak of the second-level GLM map that lie within a circle of 10 mm radius centered around the group peak (see also Oosterhof et al., 2010; Wurm et al., 2015).

Conjunction analysis

The conjunction analysis was performed using statistical maps at the group level (see section *Second- and group-level univariate analysis* above). First, we investigated which brain areas are engaged during the processing of all the four action categories. To this aim,

we computed the conjunction between the four contrasts corresponding to the four action categories (i.e., [*Communication* – baseline] ∧ [*Grooming* – baseline] ∧ [*Ingestion* – baseline] ∧ [*Locomotion* – baseline]). The resulting cluster peaks are reported in the Supplementary Materials (Table C3).

Next, we wanted to determine whether there are unique brain areas that are recruited during the processing of specific action categories. To examine this, we used group-level statistical maps of contrasts of each category versus each other category. For each action category separately, we computed the conjunction of three contrasts (e.g., [*Communication – Grooming*] ∧ [*Communication – Ingestion*] ∧ [*Communication – Locomotion*]) (see also Urgen & Orban, 2021, for a similar approach).

Conjunctions were computed by taking the minimum t-value for each voxel across the overlapping maps (Nichols et al., 2005). The resulted t-map was then thresholded with t = 1.65 and projected on a standard brain.

Multivariate pattern analysis (MVPA)

In order to investigate whether action categories can be distinguished between each other based on activation patterns, we performed multivariate pattern analysis (MVPA) using a linear support vector machine (SVM) classifier, as implemented in the CoSMoMVPA Toolbox (Oosterhof et al., 2016) and LIBSVM (Chang & Lin, 2011). We performed both ROI-based and searchlight-based (Kriegeskorte et al., 2006) decoding analyses. As input to the classifier, we used t-values from 20 regressors (i.e., each basic action vs baseline) calculated from the β estimates obtained from the first level univariate analysis (see section *Single actions as regressors of interest a*bove). At each cross-validation fold, training and testing data were z-normalized prior to classification. Decoding was performed for all pairs

of categories (e.g., decoding of actions from Category 1 versus actions from Category 2; Category 1 versus Category 3, Category 1 vs Category 4) and subsequently averaged across the three pairwise comparisons for each of four categories separately.

ROI MVPA

To investigate representations of action categories within the core regions involved in action observation (see *ROI definition*), we performed ROI-based MVPA. We selected 6 ROIs: LOTC, aIPS, and IFG, all bilateral (coordinates of the peaks are reported in Table C2). Each individual ROI consisted of 515 voxels. Subject-specific t-maps of 20 individual actions were used as input to the analysis, thus for each subject and each ROI separately there were 160 t-maps (20 (number of regressors) x 8 (number of runs)).

In order to compute classification accuracies, we used a leave-one-run-out cross-validation method. In our approach, we used pairwise decoding to compare each category against every other category (e.g., Category 1 vs Category 2, Category 1 vs Category 3, Category 1 vs Category 4) and then averaged the results across the three accuracy maps. For each subject and for each action category an SVM classifier was trained on 70 t-maps (35 from one category and 35 from another category) and tested on 10 t-maps (5 from one category and 5 from another category). This procedure was performed in 8 iterations. The classification accuracies were then averaged across the iterations, resulting in one accuracy value per subject and per ROI. To obtain the decoding accuracies of one category against all others, we averaged the results from the three pairwise decoding analyses. The mean classification accuracy was subsequently entered into a one-tailed one-sample t-test against chance level (50%). We report corrected results, where the correction was performed using a False Discovery Rate (FDR) at $q = 0.05$ (Benjamini & Hochberg, 1995) accounting for

multiple comparisons of the number of one sample t tests (i.e., 6 ROIs x 4 categories = 24 tests).

To investigate possible differences between ROIs and action categories, we conducted a repeated-measures three-way ANOVA [ROI (LOTC, aIPS, IFG) x CATEGORY (*Communication*, *Grooming*, *Ingestion*, *Locomotion*) x HEMISPHERE (left, right)] on the mean decoding accuracies. Subsequently, we performed post hoc two-tailed paired samples t tests between ROIs and between Categories. Statistical results were FDR corrected (at q = 0.05) for the number of tested models (i.e., 12 for analysis between ROIs and 18 for analysis between Categories)

Searchlight-based MVPA

Searchlight-based MVPA was performed to reveal regions outside of the AON regions that can distinguish between the four different action categories on the basis of patterns of neural activation. The decoding was performed the same way as for the ROI-based MVPA (see *ROI MVPA*), however within searchlight spheres. For each voxel in the brain, the decoding accuracy was assigned based on t-values of 100 voxels located in a sphere around that center voxel. Mean classification accuracies were assigned to the center voxel of the searchlight sphere, yielding maps of classification accuracy values for decoding a given category for a given subject. The resulting individual accuracy maps were subsequently spatially smoothed with a 5 mm FWHM kernel and entered into one-sample t-tests. Statistical significance for the group-level analyses was determined by correcting the accuracy maps for multiple comparisons using Threshold-Free Cluster Enhancement (TFCE, Smith & Nichols, 2009) in combination with cluster level correction (p = 0.05, one-tailed, z = 1.65, 5000 iterations). The decoding analysis resulted in one group-level statistical map per

category. For visualization purpose, we projected the resulting thresholded t-maps onto an inflated standard surface map provided by BrainNet Viewer (Xia et al., 2013). The results are reported in the Supplementary materials (Figure C2 and Table C4).

Functional connectivity analysis

To investigate whether different action categories are represented by unique connectivity patterns we performed a functional connectivity analysis. The procedure was adapted from the work of Heinzle et al. (2012). Based on the so far obtained results, we selected 18 ROIs: six regions of the AON obtained from the group-level univariate analysis, i.e., bilateral LOTC, aIPS, and IFG and 12 regions chosen based on the peaks obtained from the conjunction analyses. We selected three peaks with the highest t-values per category (see Table C3). The peaks encompassed several brain regions, mainly located in the frontal, anterior lateral occipitotemporal, and ventral cortices.

From the de-noised unsmoothed filtered functional images obtained from first-level univariate analysis (see section *Separate categories as regressors of interest* above), we extracted time series within each ROI. The sphere size of each ROI was 10 mm, encompassing 515 voxels. Time series were extracted for all the 515 voxels within a given ROI and subsequently averaged across these voxels. That resulted in one array of time series per run per ROI for each subject where the length of the array was equal the number of volumes within a run (184 (run with one catch block) or 204 (run with two catch blocks)).

Next, for each category separately, we extracted the relevant time points from the time series by selecting volumes covering all the trials within a block (30 s) additionally shifted by 6 s to account for the hemodynamic delay (e.g., Aguirre et al., 1998; Ekman et al., 2012). This resulted in two subsets of time series per run per category, consisting of 15

volumes each. Thus, for each subject, we obtained 16 subsets per category (2 (number of subsets in one run) x 8 (number of runs)).

Functional connectivity analysis: Correlations

In the first part of the functional connectivity analysis, we aimed to visualize how the time courses are correlated across the selected ROIs depending on the viewed action category. For each action category separately, we calculated Pearson's correlation coefficients by correlating every time series with every other time series across all the ROIs. Separately for each subject and each category, this resulted in 288 x 288 correlation matrices (18 ROIs multiplied by 16 subsets). Next, for each subject and category, we computed the mean correlation across the 16 subsets, resulting in 18 x 18 matrices. Finally, we computed the mean across participants, resulting in four 18 x 18 correlation matrices in total (one matrix per category). It is worth noting that the obtained matrices were used for visualizing the between-ROIs correlations, however, were not the input to the MVPA analysis.

Functional connectivity analysis: Category decoding

In the second part of the functional connectivity analysis, we aimed to examine how well action categories can be decoded based on the connectivity patterns. We investigated this question by taking into account (a) only 6 key regions of the AON, and (b) a larger set of 18 ROIs, including the key regions of the AON and 12 additional functionally determined ROIs (see section *Functional connectivity analysis* for details). We used a linear SVM classifier and performed leave-one-subject-out classification analysis. The process of obtaining the input data was similar to the one described above (*Functional connectivity analysis: Correlations*) but the specific procedure was as follows. Separately for each category and each of the 16 category subsets, we calculated correlation coefficients between

time series across the ROIs. In total, for each subject there were 64 correlation matrices (16 matrices per category), each matrix of a size [number ROIs x number ROIs], i.e. (a) 6 x 6 and (b) 18 x 18. These matrices were used as an input for pairwise category decoding. With four categories, there are six possible pairwise combination. The decoding for each category pair was performed in 29 iterations, equal the number of subjects, such that correlation matrices from each subject were used once in a test dataset. In each fold of the cross-validation, we trained the classifier on 896 correlation matrices (32 (16 from one category and 16 from another category) x 28 (number of subjects except one)) and tested on 32 correlation matrices (16 from one category and 16 from another category for the left-out subject).

Classifier performance was then tested against the chance level (50%) with a one-tailed one-sample t-test. Correction was performed using the FDR at $q = 0.05$ (Benjamini & Hochberg, 1995), accounting for multiple comparisons for the number of one-sample t-tests (i.e., 6 tests) within both types of analysis. We report both uncorrected and corrected results.

Functional connectivity analysis: Analysis of SVM weights

Classification analysis with the SVM classifier results in a set of feature weights that provide information regarding the contribution of a given feature for the between-class separation (Guyon & Elisseeff, 2003; Sato et al., 2008). Applied to our case, where the SVM features are correlations between the ROIs, the SVM weights can reveal pairs of ROIs whose functional connectivity is crucial for decoding action categories. Thus, for each pairwise comparison between two categories, we extracted feature weights obtained for each subject. The sign of weights indicates whether the connection between ROIs was stronger or weaker for one or the other category (Guyon et al., 2002).

Results

Behavioral results

Participants identified catch trials with a mean accuracy rate of 84.86% (± 2.40% SEM). Mean reaction time for correct responses was 1206 ms (± 27.16 ms SEM).

Univariate fMRI Analyses

To determine ROIs for the MVPA and the functional connectivity analysis, we computed a group contrast of all four categories versus baseline (cluster-based corrected, p = 0.05). The contrast revealed recruitment of regions in the occipital pole, lateral occipital cortex and occipital temporal fusiform cortex. Then, we extracted peak coordinates with the highest t-value within the LOTC, aIPS, and IFG. Peak Talairach coordinates are as follows: -40/-74/-8 (left LOTC), 38/-75/-10 (right LOTC), -24/-56/36 (left aIPS), 28/-52/34 (right aIPS), -42/8/24 (left IFG), 42/12/24 (right IFG).

Conjunction analysis

We expected that similar brain regions of the AON are engaged during the processing of the action categories. To examine the overlap in brain activation across these categories, we performed a conjunction analysis between the four statistical maps resulting from GLM contrasts of each category versus baseline. This analysis revealed a large significant cluster in the bilateral occipital pole, lateral occipitotemporal cortex, central occipitotemporal cortex, inferior frontal gyrus, and in paracingulate gyrus (Figure 4.2).

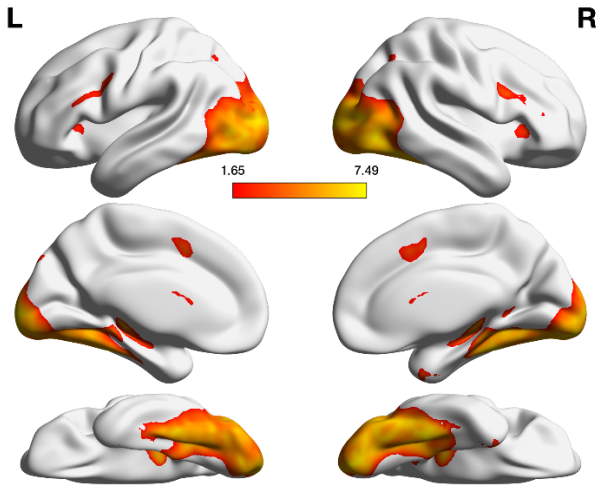


Figure 4.2. Conjunction of four RFX GLM univariate contrasts of each of the four action categories versus baseline. Cluster-based thresholding was applied to each of the four statistical maps that were entered into the conjunction analysis (corrected cluster threshold of p = 0.05).

To determine significant clusters that show a preference for one specific category, we performed a conjunction analysis on the basis of RFX GLM contrasts between a given category and each other category. For example, to determine a conjunction map for clusters that show a significant univariate difference between *Communication* and each of the three other categories, we computed a conjunction across statistical maps resulting from the RFX contrasts '*Communication* vs *Locomotion'*, '*Communication* vs *Ingestion'*, and '*Communication* vs *Grooming'*. As shown in Figure 4.3, the processing of actions belonging to the category *Communication* specifically engages clusters in the bilateral supramarginal gyrus as well as the superior and middle temporal gyrus, and bilateral inferior frontal gyrus (Figure 4.3). Processing of actions belonging to the category *Grooming* engages bilateral early visual cortex (V1, V2, and left V4), and a small cluster in the superior parietal lobe. For actions belonging to the category *Ingestion*, we obtained several smaller clusters in bilateral precentral gyrus, insular cortex, small clusters in frontal gyrus, and in the anterior division of

cingulate gyrus. Lastly, actions belonging to the category *Locomotion* engage clusters in bilateral parahippocampal gyrus, and the precuneus. See Table C3 for details.

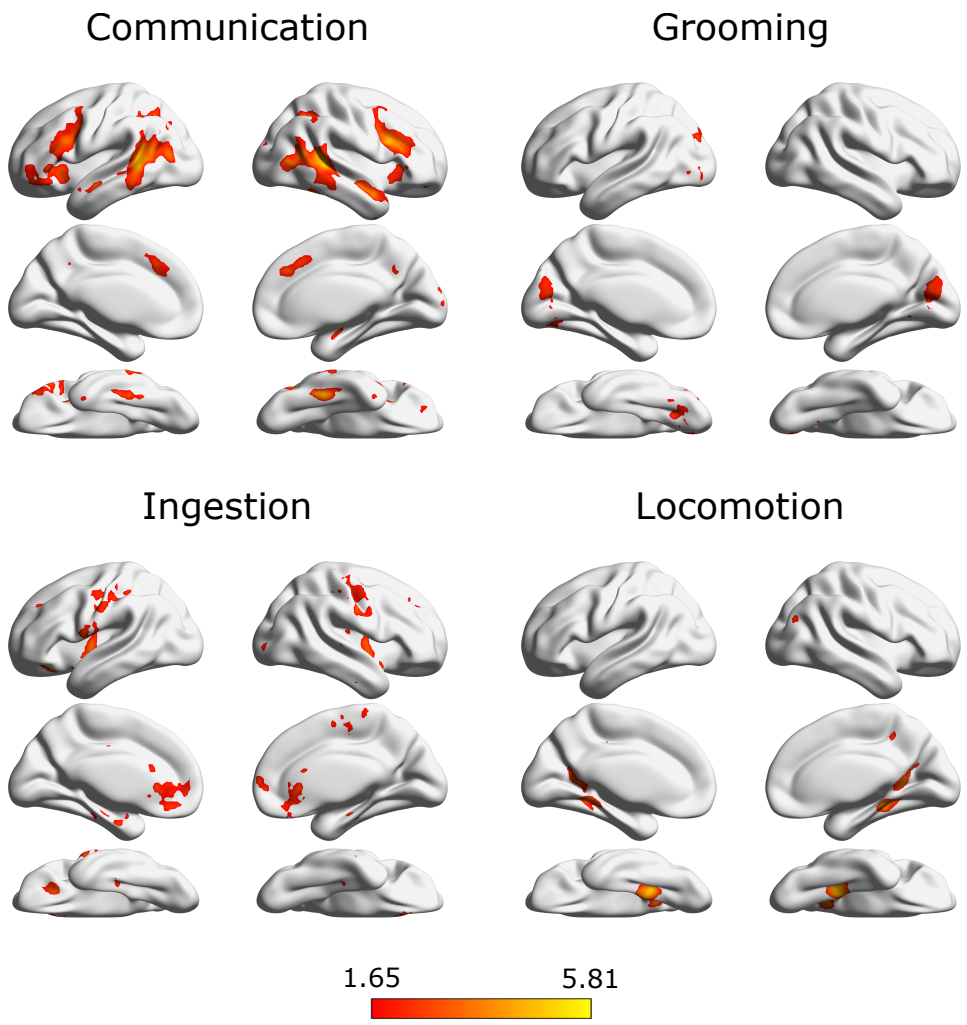


Figure 4.3. Results of the conjunction analysis showing maps unique for each category. Each conjunction map was calculated from three univariate contrasts of a given category versus each of the other three categories (e.g., conjunction of *Communication* vs *Grooming*, *Communication* vs *Ingestion*, and *Communication* vs *Locomotion*). The input contrast maps are group-level maps after cluster-based thresholding (corrected cluster threshold of p = 0.05).

ROI-based MVPA

In the ROI-based MVPA we investigated whether the four action categories can be decoded within ROIs of the AON (i.e., bilateral LOTC, aIPS, and IFG) on the basis of patterns of activation (Figure 4.4). In the analysis, we compared each action category against

every other category. Subsequently, for each category and each subject separately, we averaged the results across the three decoding accuracies corresponding to three categories, which resulted in a single accuracy value representing the decoding of one category versus the combined other three.

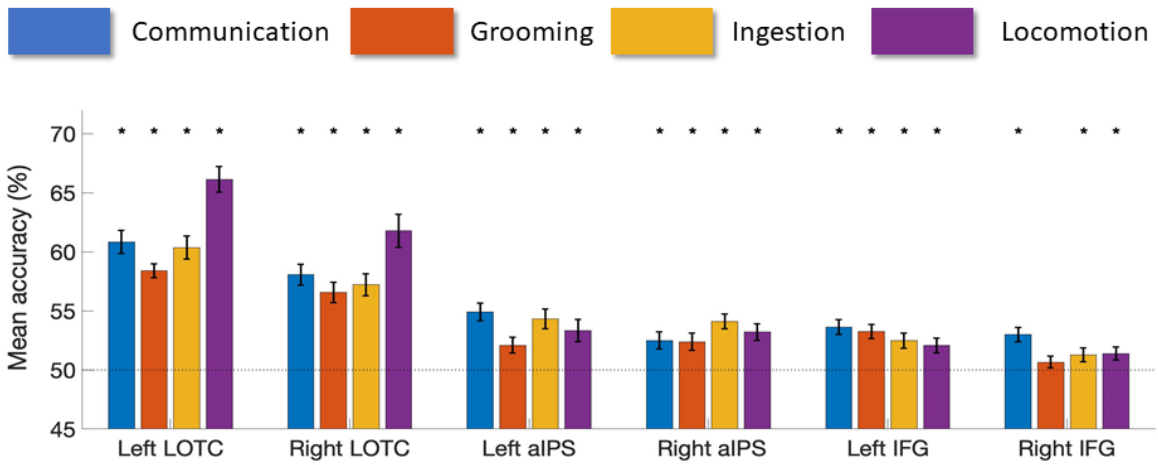


Figure 4.4. ROI-based MVPA results. Mean classification accuracies for decoding four action categories: *Communication* (blue), *Grooming* (red), *Ingestion* (yellow), and *Locomotion* (purple). Error bars indicate SEM across subjects, asterisks indicate statistical significance with one-tailed t-tests against chance-level (FDR-corrected, $q < .05$; chance level indicated by the dotted line).

Except for the category *Grooming* in the right IFG, decoding accuracies for the categories was above chance level (50%) in all the selected ROIs.

A three-way repeated-measures ANOVA [ROI x CATEGORY x HEMISPHERE] revealed a main effect of ROI ($F_{(2,56)} = 55.055$, $p < 0.001$), CATEGORY ($F_{(3,84)} = 26.919$, $p < 0.001$), HEMISPHERE ($F_{(1,28)} = 18.014$, $p < 0.001$) and a significant interaction of ROI and CATEGORY ($F_{(6,168)} = 20.030$, $p < 0.001$). No other interactions were significant (all $p > 0.05$). Since the factor HEMISPHERE did not interact with the factor ROI nor the factor

CATEGORY, we collapsed decoding accuracies across the two hemispheres in each of the ROIs.

Post hoc two-tailed paired samples t-tests revealed that decoding accuracies for each action category, when contrasted with the other three, were significantly higher in the LOTC compared to both the aIPS and IFG (Table 4.1). Moreover, decoding accuracy for the category *Ingestion* was significantly higher in aIPS compared to IFG ($t_{(28)} = 3.226$, $p = 0.003$, $q = 0.004$). We observed significant differences within the LOTC between the categories *Communication* and *Grooming* ($t_{(28)} = 3.498$, $p = 0.002$, $q = 0.005$), *Communication* and *Locomotion* ($t_{(28)} = -6.786$, $p = 0.001$, $q = 0.003$), *Grooming* and *Locomotion* ($t_{(28)} = -8.364$, $p = 0.001$, $q = 0.003$), and *Ingestion* and *Locomotion* ($t_{(28)} = -6.798$, $p = 0.001$, $q = 0.003$) (Table 4.2). Moreover, there was a significant difference within the aIPS for the categories *Communication* and *Grooming* ($t_{(28)} = 4.170$, $p = 0.001$, $q = 0.003$) as well as *Grooming* and *Ingestion* ($t_{(28)} = -4.245$, $p = 0.001$, $q = 0.003$). We also observed significant differences within the IFG for categories *Communication* and *Ingestion* ($t_{(28)} = 3.556$, $p = 0.001$, $q = 0.003$) and the categories *Communication* and *Locomotion* ($t_{(28)} = 3.945$, $p = 0.001$, $q = 0.003$).

Table 4.1. Results of the post-hoc paired samples t-tests between ROIs, computed on the basis of decoding accuracies of action categories (separately for each subject), collapsed across hemispheres. Asterisks indicate significant q values (FDR corrected for the number of tests, i.e., 12).

	LOTC - aIPS			**LOTC - IFG**			**aIPS - IFG**		
	t	p	q	t	p	q	t	p	q
Communication	7.432	0.001	0.002*	7.249	0.001	0.002*	0.663	0.513	0.560
Grooming	8.771	0.001	0.002*	7.963	0.001	0.002*	0.408	0.686	0.686
Ingestion	5.568	0.001	0.002*	8.048	0.001	0.002*	3.226	0.003	0.004*
Locomotion	9.030	0.001	0.002*	11.105	0.001	0.002*	1.953	0.061	0.073

Table 4.2. Results of the post-hoc paired samples t-tests between Categories, computed on the basis of decoding accuracies of action categories (separately for each subject), collapsed across hemispheres. Asterisks indicate significant q values (FDR corrected for number of tests, i.e., 18).

	Communication - Grooming			Communication - Ingestion			Communication - Locomotion			Grooming - Ingestion			Grooming - Locomotion			Ingestion - Locomotion		
	t	p	q	t	p	q	t	p	q	t	p	q	t	p	q	t	p	q
LOTC	3.498	0.002	0.005*	1.347	0.189	0.262	-6.786	0.001	0.003*	-2.133	0.042	0.076	-8.364	0.001	0.003*	-6.798	0.001	0.003*
aIPS	4.170	0.001	0.003*	-1.116	0.274	0.352	0.853	0.401	0.481	-4.245	0.001	0.003*	-2.046	0.050	0.082	1.847	0.075	0.113
IFG	3.081	0.005	0.010	3.556	0.001	0.003*	3.945	0.001	0.003*	0.211	0.834	0.834	0.569	0.574	0.646	0.301	0.766	0.811

Functional connectivity analysis: Correlations

Correlation matrices were created to visualize the functional connectivity between the selected ROIs evoked while observing actions from four different categories (Figure 4.5). The selected ROIs included 6 ROIs from the action observation network and 12 ROIs obtained from the conjunction analysis.

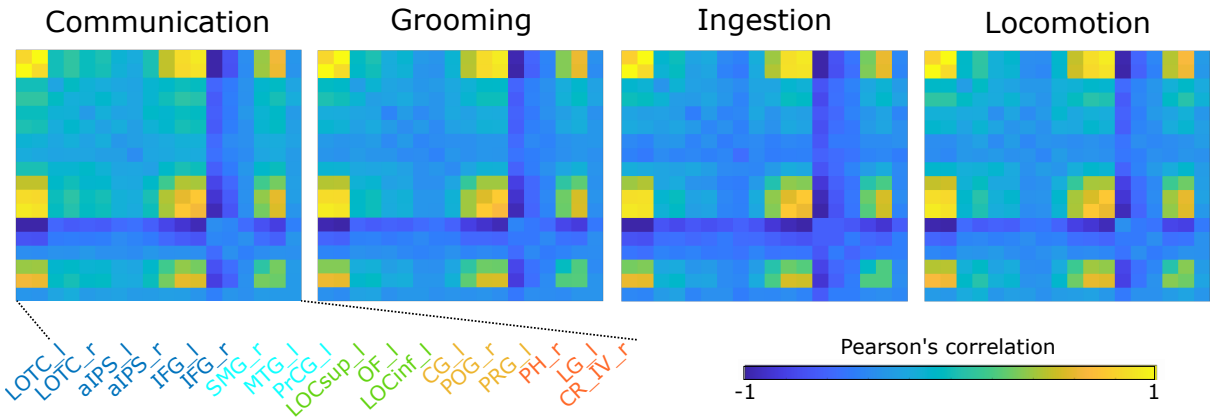


Figure 4.5. The matrices visualize between-ROIs correlations calculated based on time series obtained within each region. The ROI labels are provided below the matrices: 6 ROIs of the AON (dark blue) and 12 ROIs obtained from the conjunction analysis, i.e. *Communication* (ligth blue), *Grooming* (green), *Ingestion* (yellow), and *Locomotion* (orange) (see section *Methods: Functional connectivity analysis* and Table C3 for details). The abbreviations stand for the following brain regions: lateral occipitotemporal cortex (LOTC), anterior intraparietal sulcus (aIPS), inferior frontal gyrus (IFG), supramarginal gyrus (SMG), middle temporal gyrus (MTG), precentral gyrus (PrCG), lateral occipital cortex - superior division (LOCsup), occipital fusiform gyrus (OF), lateral occipital cortex - inferior division (LOCinf), cingulate gyrus (CG), postcentral gyrus (POG), precentral gyrus (PRG), parahippocampal gyrus (PH), lingual gyrus (LG), cerebral cortex IV (CR-IV), either in the left hemisphere (l) or the right hemisphere (r).

Functional connectivity analysis: Category decoding

In the second part of the functional connectivity analysis, we aimed to assess whether action categories can be distinguished between each other based on the functional connectivity patterns. We conducted a leave-one-subject-out cross-validation analysis, training and testing the SVM classifier to distinguish between two action categories on the basis of matrices consisting of between-region correlations. We performed the classification analysis based on (a) 6 ROIs belonging to the AON as well as (b) all 18 ROIs including 6 AON regions and 12 regions reported in the conjunction analysis. While we anticipated minimal differences in decoding of action categories using only the 6 AON regions, we expected that including the category-selective regions might enhance decoding of action categories, as these regions provide more category-specific information. The mean decoding accuracies for all between-categories comparisons are shown in Figure 4.6. As expected, when taking into account only the six regions of action observation network (Figure 4.6A), we observed an above-chance decoding between the category *Grooming* and *Ingestion.* However, that effect did not survive the FDR-based correction for multiple comparisons. When taking into account all the 18 regions (Figure 4.6B), we were able to distinguish between the following categories on the basis of the functional connectivity patterns: *Communication* and *Grooming* (decoding accuracy = 55.71%), *Communication* and *Ingestion* (decoding accuracy = 54.31%), and *Grooming* and *Ingestion* (decoding accuracy = 55.39%).

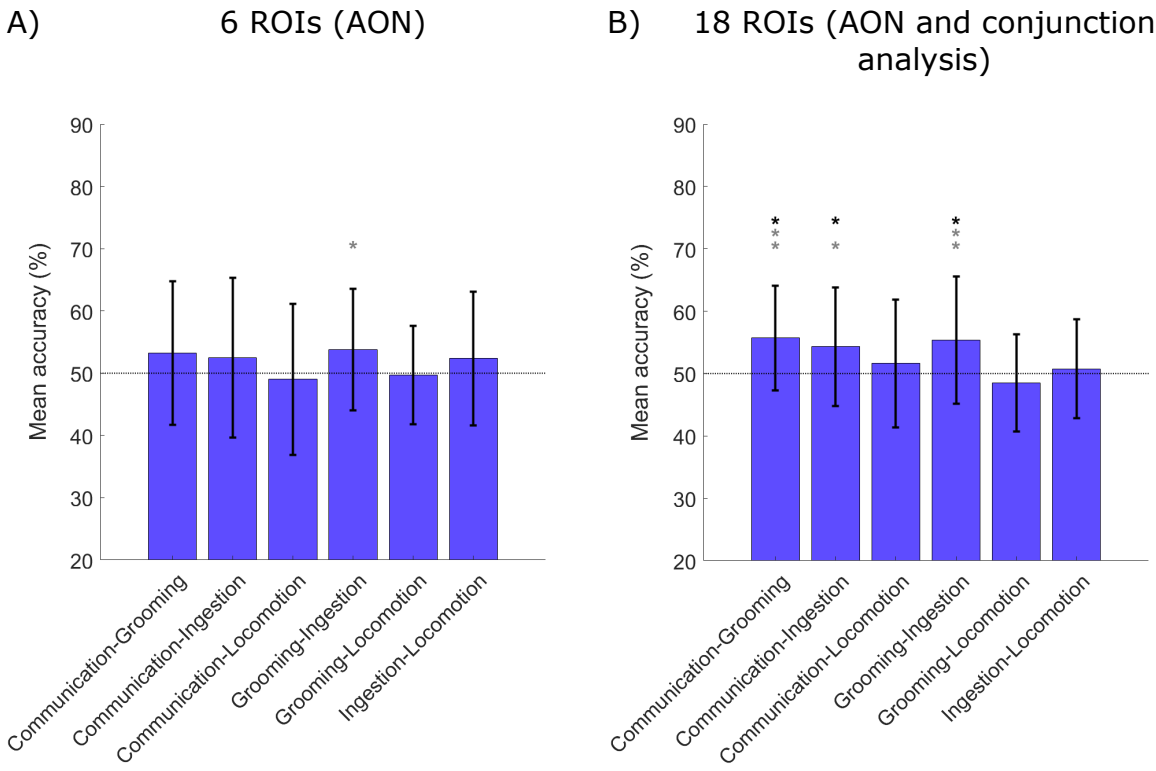


Figure 4.6. Results of the pairwise category decoding analysis on the basis of the functional connectivity between **(A)** 6 ROIs belonging to the AON and **(B)** 18 ROIs including 6 AON regions and 12 regions obtained from the conjunction analysis. Bars represent mean decoding accuracies. Error bars indicate SEM across participants, asterisks indicate statistical significance with one-tailed t-tests against chance level (i.e., 50%), indicated by the dotted line. * uncorrected $p < 0.05$, ** uncorrected $p < 0.005$, black stars $q < 0.05$ FDR corrected for number of tests (i.e., 6)).

Subsequently, we investigated which specific between-ROIs connections drive these differences. The importance of functional connectivity within pairs of ROIs was assessed by the SVM weights obtained from the classification analysis (Figure 4.7).

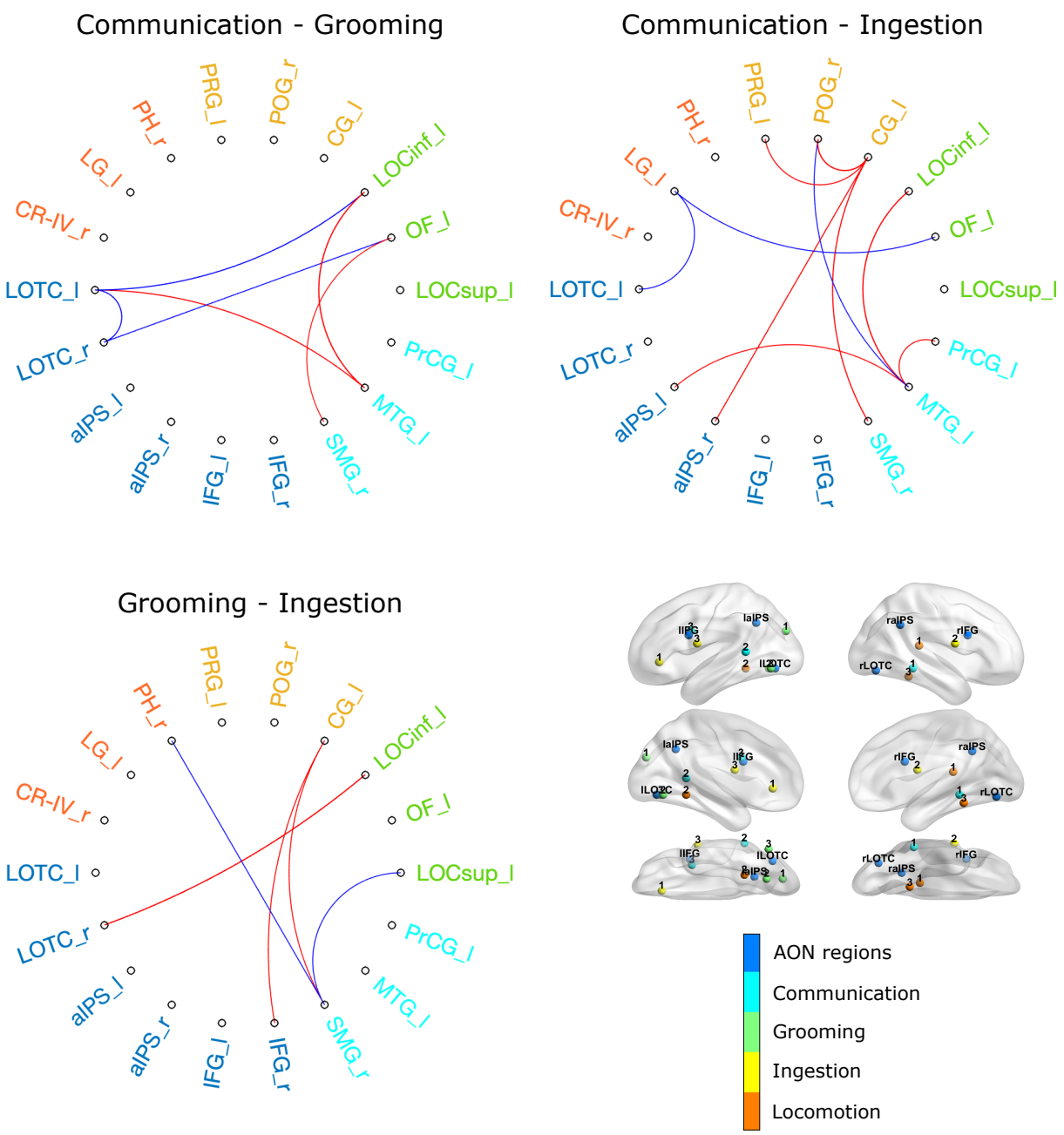


Figure 4.7. Representation of SVM weights corresponding to the importance of connections between pairs of ROIs for classifying two selected action categories. Importance of weights is visualized only for those pairs of categories that could be distinguished from each other with an accuracy significantly higher than chance level, i.e., *Communication* vs *Grooming*, *Communication* vs *Ingestion*, and *Grooming* vs *Ingestion*, see Figure 4.6). The graph represents weights of pairwise connections between the regions averaged across 29 cross folds, after applying a threshold of 0.75 (for the positive weights) and -0.75 (for the negative weights). Positive weights are shown in red and indicate that these pairs of ROIs enabled a successful decoding of the firstly mentioned category, whereas blue color – negative weights – indicate that these pairs of ROIs enabled a successful decoding of the secondly mentioned category. The abbreviations stand for the following brain regions: lateral

occipitotemporal cortex (LOTC), anterior intraparietal sulcus (aIPS), inferior frontal gyrus (IFG), supramarginal gyrus (SMG), middle temporal gyrus (MTG), precentral gyrus (PrCG), lateral occipital cortex - superior division (LOCsup), occipital fusiform gyrus (OF), lateral occipital cortex - inferior division (LOCinf), cingulate gyrus (CG), postcentral gyrus (POG), precentral gyrus (PRG), parahippocampal gyrus (PH), lingual gyrus (LG), cerebral cortex IV (CR-IV), either in the left hemisphere (l) or the right hemisphere (r). To facilitate understanding of the connections, all the peaks used in the analysis are visualized in the right bottom part of the figure. The colorbar indicates the origin of these ROI (ROIs of the AON (dark blue) and 12 ROIs obtained from the conjunction analysis, i.e., *Communication* (ligth blue), *Grooming* (green), *Ingestion* (yellow), and *Locomotion* (orange).

Discussion

In this study, we sought to determine what are the neural bases on which human participants understand different action categories and distinguish between these categories. For that purpose, we conducted a functional Magnetic Resonance Imaging (fMRI) study using actions belonging to four categories, namely *Communication*, *Grooming*, *Ingestion* and *Locomotion*. First, a conjunction analysis across the whole brain revealed that individual action categories evoke different activation maps and engage unique brain regions. ROI-based decoding analysis demonstrated that all action categories can be decoded on the basis of unique activity patterns within regions of the action observation network, with the highest decoding accuracy in the bilateral LOTC. The subsequent functional connectivity analysis revealed that some categories could be distinguished from one another on the basis of the functional connectivity patterns between regions engaged in understanding action categories. The analysis of SVM weights further highlighted pairs of regions which functional connectivity patterns contributed to the distinction between these categories. Overall, we showed that the AON regions, specifically the LOTC, enable understanding action categories, however, a broader network of selective regions carrying category-specific information is crucial for distinguishing between categories.

Role of the AON in understanding action categories

There is strong evidence that observing actions engages a set of regions that form the action observation network (Casper et al., 2010). Our group-level univariate analysis extends the current knowledge of the neural basis of action observation, demonstrating that the AON is also involved in observing actions at the level of action categories. Results of the ROI-based MVPA (Figure 4.4) revealed that all the regions within the AON evoke category-specific patterns of neural activations, allowing all the four categories to be successfully distinguished from one another (with the exception of *Grooming* in the right IFG). The successful decoding of action categories within the AON regions suggest that these regions collectively contribute to understanding of observed action categories.

Moreover, the existing literature provides evidence that, among the AON regions, the LOTC plays a crucial role in understanding actions at an abstract level (Oosterhof et al., 2010; Lingnau & Petris, 2013; Wurm & Lingnau, 2015; Wurm et al., 2015; Hafri et al., 2017; Tucciarelli et al., 2019; Yargholi et al., 2021; see review by Lingnau & Downing, 2015). Our findings support this notion, as we observed that the accuracy of action category decoding within the LOTC was significantly higher than in the other AON regions. This might indicate that the LOTC plays an important role in understanding actions at the level of categories.

The functional connectivity analysis revealed that it was not possible to decode action categories when taking into account information about time series exclusively from the AON regions. This suggests that, while these regions are crucial for action understanding, the functional connectivity between them may not be sufficient for distinguishing between different types of action categories. This could imply that the AON acts jointly with other brain regions for successful categorization. We tested this hypothesis using additional ROIs

obtained from the conjunction analysis. We will discuss these results in the subsequent paragraphs.

Category-selective brain regions

Drawing from the existing studies on objects (Haxby et al., 2001) and actions (Abdollahi et al., 2013; Ferri et al., 2015; Corbo & Orban, 2017), we hypothesized that each action category might evoke activation in different brain areas. To investigate this, we conducted a conjunction analysis that confirmed our hypothesis and revealed unique activity maps for each category. Building on our recent behavioral study, which showed that action categories can be defined and distinguished by specific set of features (Kabulska & Lingnau, 2022), we expected that category-specific brain regions (e.g., Downing et al., 2001; Bracci et al., 2010; Orlov et al., 2010; Bracci et al., 2012; Isik et al., 2017) would be involved in decoding these key features. In the following sections, we delve further into the neural maps obtained from the conjunction analysis.

Communication. As communication involves interaction with other people, we expected that observing communicative actions will activate regions typically involved in social interactions. As anticipated, the conjunction analysis revealed unique activation maps in the bilateral pSTS/STG and the middle temporal gyrus (MTG). Posterior STS is considered a hub for biological motion (Calder & Young, 2005; Deen et al., 2015; Grossman et al., 2005; Ishai et al., 2000; van Kemenade et al., 2012) such as gaze shifts (Hoffman & Haxby, 2000; Puce et al., 1998; Wicker et al., 1997), body movement (Bonda et al., 1996; Grossman et al., 2000; Kourtzi & Kanwisher, 2000; Senior et al., 2000), mouth movement (Puce at al., 1998), and lip reading (Calvert et al., 1997) (see Allison et al. (2000) for review). It also belongs to the so-called „social brain network“ (Frith, 2007) and plays a role in social perception,

including perception of faces (Haxby et al. 2000; Pitcher et al., 2011), voices (Belin et al., 2000), social actions (Isiki et al., 2017; Wurm et al., 2017) and modality-invariant representation of person identity (Anzellotti & Caramazza, 2017). Its activation increases especially in response to dynamic faces (e.g., Allison et al., 2000; Haxby et al., 2000; Pitcher et al., 2011). Moreover, the obtained clusters in the MTG extends from the posterior to anterior division and covers the visual pathway specialized for social perception (Pitcher & Ungerleider, 2021), that begins in primary visual cortex (V1), goes through motion-selective area V5/middle temporal (MT), posterior STS, and ends in the anterior STS. A meta-analysis by Grosbras et al. (2012) showed that the right pSTS and bilateral MTG are also involved in hand and face movements. These regions, along with the right extrastriate body area (EBA) and MT/V5 areas that also emerged in the analysis, have been reported to play a role in the perception of body, hands and face movements as well as in recognizing dynamic human faces (Grosbras et al., 2012; Sato et al., 2004). Studies on anatomical connectivity have shown that the STS receives input from the visual stream through the motion-sensitive area MT, both in monkeys (Seltzer & Pandya, 1994) and humans (Pitcher & Ungerleider, 2021). Human functional connectivity studies have identified connections between the pSTS and the fusiform face area (FFA) (von Kriegstein et al., 2003). Furthermore, the STS has been shown to be functionally connected to both the premotor and primary motor cortices (Deen et al., 2015). Overall, our results showed that observing communicative actions evoked neural activation in regions playing a role in biological motion, social perception, and dynamic facial stimuli.

Grooming. We observed that the neural activity associated with grooming actions was primarily located in the visual pathway, more precisely in bilateral V4 regions and in a portion of the dorsal visual stream. Area V4 is known for its role in visual object recognition,

contributing to the processing of object-related features, such as color, shape, and contour, as evidenced by both primate (Mountcastle et al., 1987; Roe et al., 2012; Zeki, 1973), and human (Bracci & Op de Beeck, 2016; James et al., 2003) studies. Furthermore, the lateral occipital cortex (LOC) has been shown to play a crucial role in object recognition (Grill-Spector et al., 1999, 2001). As has been demonstrated in studies on monkeys, there are anatomical connections linking V4 and the inferotemporal cortex, a region functionally similar to the human LOC (Felleman et al., 1997). Based on Goodale and Milner's model (1992), activation in the dorsal visual stream may imply that participants were also processing information related to potential actions they could perform with these objects. This interpretation aligns with the nature of grooming actions, which often involve the use of specific objects like toothbrushes or makeup applicators. Therefore, our findings indicate that the brain regions activated in response to grooming actions are engaged in both recognizing objects and potential actions involving these objects.

Ingestion. We expected to find unique activation within regions engaged in processing objects, especially food-related information as well as hands and mouth. The cluster in the postcentral gyrus obtained in our study cluster overlapped with a peak reported by Cornier et al., (2009) that showed higher activation to images containing food compared to non-food images. We also reported small clusters in bilateral ventral visual cortex, that has been recently shown to have high selectivity to visual images of food (Khosla et al., 2022). The map also revealed a cluster in the precentral area, a region which has been shown from the ALE meta-analysis to carry information about movement of hands, faces, and an object-directed hand movement (Grosbras et al., 2012).

Locomotion. Locomotion requires change of location (Kabulska & Lingnau, 2022) and thus often has to be performed outdoor. We then expected that one of the brain area unique for this category might be related to processing the spatial context. Results of the conjunction analysis revealed a peak in the parahippocampal gyrus that overlaps with the parahippocampal place area (PPA). The PPA has been associated with scene recognition, such as viewing landscapes and indoor layouts (Epstein & Kanwisher, 1998) as well as scene perception and spatial navigation (Aguirre & D'Esposito, 1999; Epstein et al., 2001, 2003). The parahippocampal gyrus, together with the lingual gyrus and the posterior cingulate cortex, i.e., the regions that we also reported from the conjunction analysis, have been shown to increase activation in spatial retrieval task (when participants attended spatial information) in contrast to temporal retrieval task (when participants attended to temporal information) (Ekstrom et al., 2011). Our results also revealed a cluster in the right precuneous. It has been reported that the precuneous plays a role in visuo-spatial processing, including coordination of motor behavior, directing attention towards moving targets, and imagining movements. It is also connected to motor areas, i.e., dorsal premotor area and supplementary motor area (see review by Cavanna & Trimble (2006)). In summary, our results show that observing locomotive actions in natural environment engages regions that are related to spatial processing and scene perception.

Extended network for action categories

We analyzed time series data from regions within the AON as well as from regions identified through the conjunction analysis. We were able to successfully decode certain pairs of categories using time series from these ROIs. The results show that, as expected, action understanding is a distributed process extending beyond just the AON. Additionally, we

show which regions might play the most crucial role in this decoding. As an example, decoding of the categories *Communication* and *Grooming* relied on both bilateral LOTC as well as peaks specific to each category. This implies that specialized regions contribute unique information that enables the categorization of specific types of actions.

Limitations

However, it is important to interpret the functional connectivity results with caution. First, we did not pre-register the exact number of category-selective ROIs used for pairwise category decoding. Instead, we selected it in a way that the number of ROIs is not too high (three ROIs per category) and the results can still be interpretable (see Figure 4.7). Second, the threshold for visualizing SVM weights was also not defined during the pre-registration. We chose it such that we could display only the strongest connections (Figure 4.7). While we have provided rationales for these choices, it is essential to note that these parameters were determined subjectively. Therefore, a different selection of these parameters might yield different results.

Conclusion

Our research revealed that observing action categories engages distinct brain areas, with each category evoking activation within specific regions that process information about unique features of that category. The findings further showed that, while the AON serves as foundation for action understanding, effective action categorization requires involvement of a broad network of specialized regions. Overall, our results provide insights into the neural underpinnings of action categories, encompassing information about neural activation as well as between-ROIs connections.

CHAPTER 5: GENERAL DISCUSSION

In my dissertation, the goal was to explore the cognitive and neural bases of action understanding. The first part of this work sought to identify the key action features, that are crucial for action recognition, as well as more general action categories. Subsequently, I aimed to examine the underlying neural structures of action organization. Specifically, the whole PhD project consisted of three main studies:

Study 1. This study included a series of behavioral experiments performed in order to better understand the cognitive organization of actions. The findings of Study 1 revealed a list of key action features, such as *arms, legs, action targeting a tool, pace of action, and duration* as well as eleven action categories, like *Locomotion*, *Communication*, and *Food-related actions*.

Study 2. The goal of this study was to explore the neural representations of action features and the underlying dimensions. The methodology included conducting an fMRI experiment and using the features identified in Study 1. The findings revealed distinct clusters within the lateral and ventral occipitotemporal cortices that corresponded to different feature dimensions, such as *Posture*, *Contact with others* and *Object-directedness*.

Study 3. The goal of the subsequent fMRI experiment was to examine the neural underpinnings of action categories derived from Study 1, namely *Communication, Ingestion, Grooming,* and *Locomotion*. The results showed that the LOTC has the highest decoding accuracy compared to other AON areas. Furthermore, some of the action categories could be decoded from each other based on unique connectivity patterns between regions encompassing the AON and the category-selective brain areas.

In the following sections, I will discuss the findings in a broader context and explore their impact across various research fields. Lastly, I will discuss the limitations of the studies and present ideas for potential future studies.

Cognitive principles underlying action organization

To explore the underlying structure of action organization, I first examined the action features at the cognitive level. Actions are complex; among other factors, they evolve over time, they can be performed in different environments, and involve a variety of objects and tools. For instance, a single action can be performed using various objects, while a single object can be involved in multiple actions. Given these complexities, the exploration of action features poses challenges for their identification that require a suitable method.

To tackle this problem, I adopted a data-driven approach and collected features from naïve participants. While I could have proposed my own set of features – a method that has been proven effective in numerous studies (e.g., Gainotti et al., 2009; Klatzky et al., 1993; Magri et al., 2021; Proklova et al., 2016) – this method would have potentially restricted the research to a limited scope. Instead, my goal was to obtain an objective set of features that reflects the view of a wider group of people. This way, I could explore a wide range of potential features and possibly uncover ones not previously discussed in the literature. The free-feature listing experiment in Study 1 resulted in an extensive set of action features, which can serve as feature norms for future experiments. Such norms have been already proposed in studies on objects, providing featural descriptions for living and nonliving objects (Garrard et al., 2001; Lynott & Connell, 2009; McRae et al., 2005) as well as actions and events (Vinson & Vigliocco, 2008). Feature norms provide insights into the structure of semantic representations (Rogers et al., 2004) and their importance has been shown in a variety of

experiments, including investigations of the structure underlying conceptual categories (Gainotti et al., 2013), knowledge (Hoffman & Lambon Ralph, 2013), and language comprehension (Cree et al., 2006), as well as in studies on patients with semantic deficits (Cree & McRae, 2003).

The next step of Study 1 involved identifying the key features within this extensive feature set that could be subsequently tested at the neural level. Taking the obtained features into account and considering the existing literature (e.g., Tarhan & Konkle, 2020b; Tucciarelli et al., 2019; Watson & Buxbaum, 2014; Wurm et al., 2017), I narrowed down the list to 44 key features. These cover a broad range of action-related information and vary in levels of abstraction. Through a principal component analysis, I mapped these features into a multidimensional space, comprising dimensions related to *General movements*, *Arm movement kinematics*, *Goal-directedness*, *Context*, *Posture*, *Contact with others, Object-directedness*, and *Negative emotions*. The neural underpinnings associated with these key features and feature dimensions were further investigated in Study 2.

Neural representations of action features

It has been shown that objects and actions (Huth et al., 2012) as well as mental states (Tamir et al., 2016) can be mapped onto dimensions which are reflected in dedicated neural systems. Following these studies, I investigated whether such neural systems also exist for the dimensions underlying action features that were obtained in Study 1. The results showed that all these dimensions except two, namely *Arm movement kinematics* and *Negative emotions*, were represented in the brain. The obtained clusters encompassed several brain regions, including the ventral and dorsal visual stream.

In the literature on action understanding, there is still an ongoing debate regarding the role of the AON regions. The first step of Study 2 involved computing a reliability map which identifies voxels consistently activated in response to actions. The obtained reliability map encompassed the parietal and occipitotemporal areas, but not the premotor regions, meaning that in the premotor regions the neural activation in response to action stimuli was not consistent across imaging runs. Given that I used naturalistic action images as stimuli, this could indicate that the premotor cortex does not generalize across aspects that vary in these stimuli, such as backgrounds, agents involved in actions, and their postures. This challenges theories that emphasize the critical role of the premotor cortex in representing action meaning (Majdandić et al., 2009; Nelissen et al., 2005; Rizzolatti et al., 2014) and aligns with the findings, which suggest that the premotor cortex rather codes low-level, perceptual action features (e.g., Wurm et al., 2015; Wurm & Lingnau, 2015).

Given the growing evidence highlighting the crucial role of the LOTC in action understanding (Tucciarelli et al., 2019; Wurm & Caramazza, 2019a), I explored the idea previously proposed by Lingnau & Downing (2015), that the LOTC might host and integrate different types of action-related information. More precisely, I examined which dimensions underlying action organization are represented within the LOTC and how these representations relate to one another. The results have shown that distinct subregions of the LOTC held information about different dimensions, specifically *General movements*, *Goal-directedness*, *Context*, *Posture*, *Contact with others*, and *Object-directedness*. I also observed that the representations of these dimensions overlapped. To identify the most dominant dimension in each region, I conducted a winner-takes-all analysis. This allowed examining the selectivity of different brain areas, revealing which dimension was most strongly represented in each voxel. The findings were consistent with earlier studies on actions that

explored action dimensions, such as social interactions and object-related information (Isik et al., 2017; Wurm et al., 2017; Wurm & Caramazza, 2022), and the representation of body parts (Downing et al., 2001). The results also support the findings of Wurm & Caramazza (2022), who demonstrated that the LOTC can differentiate between animate and inanimate action-related information. Overall, the findings indicate that the LOTC has a multifunctional role, with its subregions processing various action-related information. It is worth noting that the studies referenced above examined dimensions separately, e.g., focusing on just one type of action-related information (e.g., social interaction) at a time. My study provides additional insights as it allowed to explore the feature dimensions collectively. This approach not only revealed where these dimensions are represented but also how they relate to one another.

The presence of overlapping representations of certain dimensions within the LOTC raises the possibility that this region might not have domain-specific subregions. A similar conclusion has been drawn when examining brain regions previously believed to exhibit selectivity for specific object categories. Specifically, the fusiform gyrus demonstrated strong responses not only to faces but also to bodies (Peelen & Downing, 2005), while the left EBA exhibited strong responses to bodies as well as to mammals (Downing et al., 2005). Building upon the discussion presented by these authors, I propose a hypothesis that the LOTC contains multiple distinct feature-selective neural representations, instead of domain-specific regions that process single feature dimensions. It is plausible that the LOTC is composed of intertwined populations of feature-selective neurons (Downing et al., 2005; Quiroga et al., 2005). In conclusion, the investigated dimensions might share a common representational space within the LOTC, resulting in their coexistence in the same neural areas.

Neural underpinnings of action categories

As it is well established that object categories, like chairs and houses, evoke unique brain activity patterns (Haxby et al., 2001), I proposed that a similar phenomenon might be true for action categories. The findings revealed that while in all AON regions it was possible to decode action categories, the LOTC showed the highest decoding accuracy, significantly higher than in the aIPS. These results contribute to the discussion regarding the potential role of the parietal- and occipitotemporal regions in processing the meaning of actions. Previous studies have shown that actions (e.g., *dragging*, *grasping*) and action categories (e.g., *climbing*, *running, performing manipulative actions*) evoke neural activity patterns within the parietal cortex (Abdollahi et al., 2013; Corbo & Orban, 2017; Ferri et al., 2015; Jastorff et al., 2010; Urgen et al., 2019), concluding that the parietal cortex is central to the abstract understanding of actions. However, these studies primarily focused on the parietal cortex, overlooking other brain regions, such as the occipitotemporal regions. The findings of my study provide evidence that both regions are important for the understanding of action categories, with significantly better decoding of the categories in the LOTC.

In the second part of Study 3, I investigated the functional connectivity between multiple brain regions including those belonging to the AON and those activated in response to specific action categories. The analysis showed that the LOTC was consistently involved in decoding action categories, as its connections with category-specific regions appeared to be crucial in this process. It was the only AON region consistently engaged in category decoding.

LOTC as a hub for action understanding

The results from Studies 2 and 3 showed a multifunctional role of the LOTC: it carries representational maps of dimensions underlying action features, it holds representations of feature-based and category-based organizations of actions, it can accurately decode action categories, and it is consistently connected with regions hosting category-specific information. Taking all this into account, I considered a hypothesis previously proposed by Lingnau & Downing (2015) that the LOTC might act as a hub for action understanding.

Typically, a region is considered a "hub" when it integrates information from multiple modality-specific brain regions and puts them into multi-modal representations (Anzellotti, 2017). For instance, the anterior temporal lobe, believed to be a hub for semantic knowledge, connects different kinds of knowledge represented in specialized brain regions. This includes color information from color regions, shape information from visual-form regions, names of the objects in language regions, etc. (Anzellotti, 2017; Hoffman et al., 2014; Patterson & Lambon Ralph, 2016).

Action recognition relies on combining diverse types of information, which include identification of specific effectors involved in an action, recognizing the usage of tools and objects within the action context, as well as understanding possible changes over time and actions' end-goals. The recognition of an action should be independent of the effector (e.g., a cup can be gripped in several ways) or the environmental context (e.g., jogging can be performed in an indoor gym as well as in a park). The previous literature has provided evidence that the LOTC carries information about various types of information, such as motion (Papeo & Lingnau, 2015; Tootell et al., 1995; Zeki et al., 1991), biological motion (Grosbras et al., 2012; Lingnau & Petris, 2013; Papeo & Lingnau, 2015), tool viewing (Bracci

et al., 2012; Chao et al., 1999), body parts (Downing et al., 2001; Orlov et al., 2010), action observation (Caspers et al., 2010), action planning (Astafiev et al., 2004; Johnson-Frey et al., 2005), and verbs (Bedny et al., 2008; Papeo & Lingnau, 2015; Peelen et al., 2012), see Lingnau & Downing (2015) for a review. As mentioned earlier, the LOTC represents actions at an abstract level, generalizing across different target objects (Wurm et al., 2015) as well as across objects and kinematics involved in performing an action (Wurm & Lingnau, 2015). In the LOTC, actions can be decoded across various formats, such as static images versus dynamic videos (Hafri et al., 2017) as well as videos and written descriptions (Wurm & Caramazza, 2019a).

Both the existing literature (see the review by Lingnau & Downing, 2015) as well as the findings from my studies suggest that the role of the LOTC might be integration of action-related information, thereby enabling action understanding. Consequently, the obtained results provide additional evidence for the theory that the LOTC could indeed function as a hub for action understanding.

Implications

Implications for cognitive neuroscience

First, the findings of my projects revealed the cognitive structure underlying observed daily actions, including a wide set of action features as well as a set of action categories. This feature set might be used as action feature norms for future studies. Second, I investigated the neural mechanisms underlying action understanding, considering both the representation of feature dimensions as well as the brain regions carrying activity patterns unique to specific action categories. The results provide new insights into the neural underpinnings of understanding daily actions. Moreover, the results highlight the role of the LOTC in

understanding actions at an abstract level, which is in line with a vast body of research (see the review by Lingnau & Downing, 2015) and contributes to the ongoing debate about the functions of the AON regions (Wurm & Caramazza, 2019a).

Implications for computational science

The findings of my project might also have implications for the field of computational science. Through a series of behavioral experiments, I have identified action features that could describe 100 daily actions depicted as static images. These action features do not only provide valuable insights into human behavior but also might be used in the development of more human-like neural models. For instance, convolutional neural networks can be designed with a bias towards features that are important for humans, an approach that has been applied in object recognition (Geirhos et al., 2019). Incorporating action features derived from human behavior into neural networks can enable machines to understand and mimic human actions more effectively, leading to more realistic and context-aware AI systems (Wichmann & Geirhos, 2023).

Limitations

Stimuli

In all three studies, I used naturalistic, static images. This choice was influenced by previous research, though it might introduce certain limitations. Below, I outline both the advantages and disadvantages of this approach.

First, as demonstrated with objects, using naturalistic stimuli provides a better understanding of real-world scenarios (Haxby et al., 2020). However, the usage of highly controlled stimuli enables controlling for other factors, such as the surrounding objects and

scenes (e.g., Wurm et al., 2017). While naturalistic stimuli provide a more realistic experience as they present actions in their true context, they might also introduce variability and complexity. As an example, the stimuli might depict people in the background who are unrelated to the main action, or present other objects that could capture the participant's attention. This can pose challenges in isolating specific factors of interest and might introduce noise unrelated to the actions themselves.

Second, we used static stimuli. Static images are easier to design and manipulate, allowing for precise control of the timing of stimulus presentation and better experimental control. As has been shown, even a brief display of a static snapshot of an action is sufficient for accurate action recognition (Hafri et al., 2013). However, real-world actions are dynamic and unfold over time, whereas static images capture only a single moment. Consequently, one of the primary limitations is the reduced ecological validity, as static images may not fully capture the nuances of how the brain responds to dynamic actions. Moreover, statics images lack crucial kinematic information, such as motion trajectories and velocity, which, as I reported in Study 1, are important features of action understanding.

Functional Magnetic Resonance Imaging

Within the scope of my research, I conducted all neuroimaging experiments using fMRI. This technique has been successfully used in neuroscience for the past 30 years as it allows for non-invasive mapping of the human brain function with a good spatial resolution. However, the technique also comes with several limitations (Logothetis, 2008; Logothetis & Wandell, 2004; Turner, 2016). It has a limited temporal resolution - in my projects, the MRI scanner was set up to capture neural signals every 2 seconds - which means that it might miss rapid neural processes. Although it provides relatively good spatial resolution in comparison

to other neuroimaging techniques (e.g., EEG, MEG, PET (Sejnowski et al., 2014)), for my scans, I set up the voxel size to 3x3x3 mm, which encompasses activity within large groups of neurons. Moreover, fMRI indirectly reflects neural activity by measuring changes in blood flow and oxygenation (the blood-oxygenation level dependent (BOLD) response) which introduces a delay and leads to missing the real-time neural activation. Additionally, the noisy and confined environment in the MRI scanner can be uncomfortable and tiring for participants, potentially affecting their cognitive performance during tasks.

Future studies

In my study, I used fMRI to investigate the neural underpinnings of action representations, which gave insights into the spatial information of these neural systems. However, the processing of action-related information might vary in time which cannot be captured using fMRI methods. Given that actions are dynamic by nature, it is essential to view action recognition as a continuous process and to investigate it over time. Therefore, the next step would be to gain a better understanding of the temporal dynamics of action representations.

First, as a follow-up to Study 2, I would explore the temporal dynamics associated with various action features and dimensions. This idea draws inspiration from studies on objects that examined how information about object-related features evolves over time. For instance, Carlson et al. (2013) demonstrated that the time of decoding varies based on the level of abstraction, with concrete object examples being identified earlier than the more abstract ones. A recent study highlighted that the hierarchical representations of objects unfold over time: low-level visual features emerge as early as 70 ms, while more conceptual object representations appear around 150 ms (Hebart et al., 2018). Similarly, research on

actions has shown hierarchical information processing, where the earliest detected features pertain to visual information (e.g., captured by early layers of a convolutional neural network), followed by action information (e.g., effectors, transitivity), and finally, by social-affective information (e.g., valence, arousal) (Dima et al., 2022). Interestingly, these findings contrast with prior research, which showed that action goals are processed first (Hafri et al., 2013), and motor properties such as grip force and movement speed are adjusted based on these goals (Gentilucci et al., 1997; Rosenbaum et al., 2001). Such discoveries underscore the potential impact of high-level action understanding on low-level visual perception (Kilner, 2011). I believe that using a wider set of features could add to the discussion on the temporal hierarchy of action features in the human brain. As the features identified in Study 1 vary in the level of abstraction, for instance *Body parts* are more concrete than *Concentration*, I believe that they would be suitable for answering this question.

Second, following Study 3, I would explore the temporal dynamics of the four action categories. The primary goal would be to determine the time windows when each category is perceived and when these categories become distinguishable. Past research involving eye saccades in object recognition has demonstrated that differentiation between certain object categories can occur already after 120 ms (Kirchner & Thorpe, 2006), whereas various objects, such as faces and vehicles, are detected with varying reaction times (Crouzet et al., 2010). Studies based on electroencephalography (EEG) have revealed that different object categories, including buildings, cars, faces, animals, and tools, evoke unique event-related brain potentials (ERPs) for each category and thus can be differentiated based on their unique time courses (Murphy et al., 2011; Simanova et al., 2010; Wang et al., 2012). Drawing from these studies on objects, I anticipate that each action category might also evoke distinct temporal patterns.

For both experiments, I would use an EEG. In comparison to MRI, EEG offers a good temporal resolution, which allows the investigation of rapidly changing brain patterns. To investigate the temporal information of action features and dimensions, I would apply the design and dataset I used in Study 2, whereas to investigate the action categories I would use the design and dataset of Study 3. This approach would allow for a direct comparison of the results from both neuroimaging methods. Combining neuroimaging data from fMRI and M/EEG has already been applied by some of the leading research groups in the field of human object recognition (Cichy & Oliva, 2020).

Overall, these further studies would help to reveal the sequence in which different types of information (feature- and category-based) arise and would provide insights into how action understanding unfolds in time and space.

APPENDIX

A. Study 1 Supplementary materials

A.1 Experiment 1

A.1.1 Selection of action words

Actions were chosen from a study by Vinson & Vigliocco (2008). As a first step, we discarded verbs that are difficult to depict as static images (e.g., animal sounds such as "oink", "chirp"), with the aim to arrive at a final set of 100 actions we considered suitable for our experiments. Subsequently, we adjusted several action words towards more common actions. As an example, we chose "feeding" instead of "feeding a horse", and "hugging" instead of "tree hugging". Additionally, we tried to avoid keeping actions with very similar meaning in the dataset, e.g., "playing tennis" and "hitting a tennis ball". Thus, three other actions (i.e. "hitting a tennis ball", "playing piano", and "fist bumping") were removed and replaced by the actions "reading", "riding on a bike", and "writing on a board" from the "Stanford 40 Actions" dataset (Yao et al., 2011).

Table A1. Actions used in the experiments

1. applauding
2. arguing
3. blowing bubbles
4. breaking
5. brushing hair
6. brushing teeth
7. building a sandcastle
8. calling (phone)
9. carrying buckets
10. chopping vegetables
11. cleaning the floor
12. climbing
13. constructing
14. cooking
15. cutting trees
16. cutting with knife
17. dancing
18. digging
19. dragging
20. drawing
21. drinking
22. driving a car
23. driving a scooter
24. drumming
25. eating
26. feeding
27. fishing
28. fixing a bike
29. gardening
30. goal keeping
31. grocery shopping
32. hammering
33. hand shaking
34. handstand
35. having a shower
36. high fiving

37. hiking
38. holding hands
39. holding umbrella
40. hula hoop
41. hoovering
42. hopping
43. hugging
44. juggling
45. jumping
46. kayaking
47. kicking a football
48. knitting
49. knocking on a door
50. leaning on a hand
51. licking ice cream
52. lifting weights
53. listening to music
54. looking through microscope
55. making a bed
56. painting
57. paying someone
58. playing basketball
59. playing golf
60. playing guitar
61. playing tennis
62. pointing
63. pouring liquid
64. public speaking
65. pulling (tug of war)
66. punching
67. pushing a trolley
68. raking leaves
69. reading
70. riding a bike
71. rowing a boat
72. running
73. shooting an arrow
74. sitting
75. skateboarding
76. skiing
77. sleeping
78. sliding (water slide)
79. smoking
80. stirring
81. stroking a dog
82. surfing
83. swimming
84. swinging
85. switching on the light
86. taking a photo
87. tearing
88. texting
89. throwing a Frisbee
90. thumbs up
91. using a computer
92. walking a dog
93. washing a car
94. washing dishes
95. washing hands
96. watching TV
97. waving hand
98. writing on a board
99. writing
100. yawning

A.1.2 Stimulus selection

A.1.2.1 Participants

Nineteen healthy participants took part in the study (15 females; mean age = 22 years, age range = 18-26 years). Experimental procedures were approved by the ethics committee at the University of Regensburg.

A.1.2.2 Methods

We aimed for a final set of 100 action images depicting the actions listed in Table A1. First, we selected 160 images from Shutterstock (www.shutterstock.com). We chose images according to the following criteria: (1) the depicted action is the main aspect of the image, (2) the action is depicted in front of a natural (rather than a uniform) background, (3) the body of the person performing the action is fully visible, (4) there is only one person on the image (unless the action is directed at another person), and (5) the image is taken in landscape (rather than portrait) orientation. For three of the actions (*tearing*, *switching on the light* and *breaking*), we could not find suitable pictures showing the full body and thus chose images showing the upper body/arms only. To ensure that the actions were recognized as the actions we had in mind during stimulus selection, we carried out an online survey (https://www.soscisurvey.de/) with the initial set of 160 images. For actions that might be more difficult to depict as an image (e.g., arguing), we chose more than one exemplar (on average, 1.6 images per action). Participants were presented with a set of action pictures on a screen, one after the other, and were asked to type the name of the depicted action on the keyboard. The final set of stimuli was chosen by selecting, out of the set of used pictures, those for which the "correct" action label was mentioned most frequently. Based on the naming agreement, we chose a set of 100 images (see Figure A1) for the multi-arrangement task (Experiment 1).

Figure A1. Stimuli used for Experiment 1. Actions are sorted alphabetically (from left to right, row by row). For corresponding labels, see Table A1.

A.1.3 Procedure

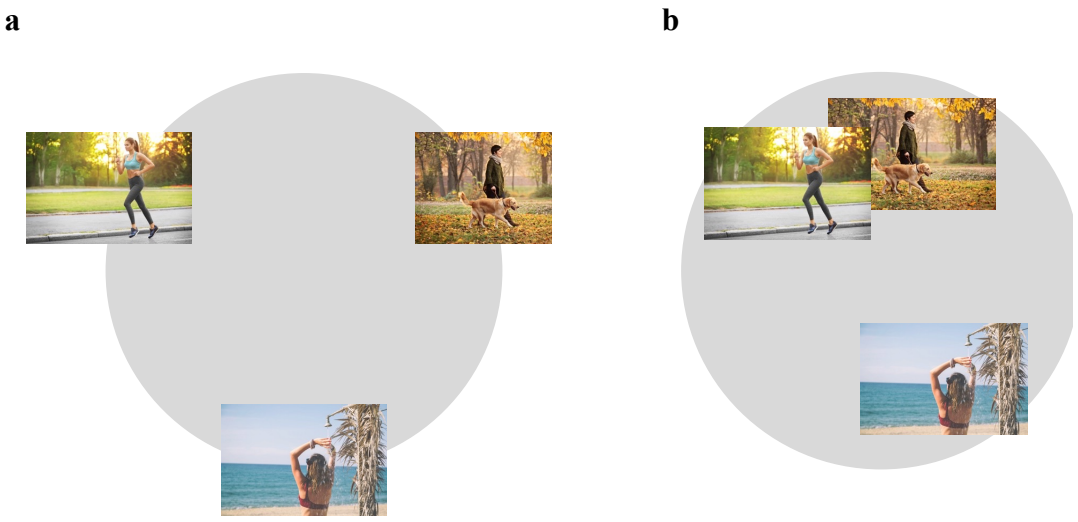


Figure A2. Exemplary trial of the multi-arrangement experiment (Kriegeskorte & Mur, 2012). Participants were asked to arrange the images by mouse drag-and-drop such that the physical distance between the images on the screen reflects the perceived similarity in terms of the meaning of the actions. In the first trial, all the 100 actions appeared on the screen around the arena. After arranging all the images within the arena, the next trial started. With each consecutive trial, a subset of images was presented, depending on pairwise dissimilarity evidence between the images (see Section Experiment 1, Procedure, for details). (**a)** Example trial with a subset of three action images (size of the images enlarged for ease of visualization). (**b)** Example arrangement at the end of a single trial.

A.1.4 Results

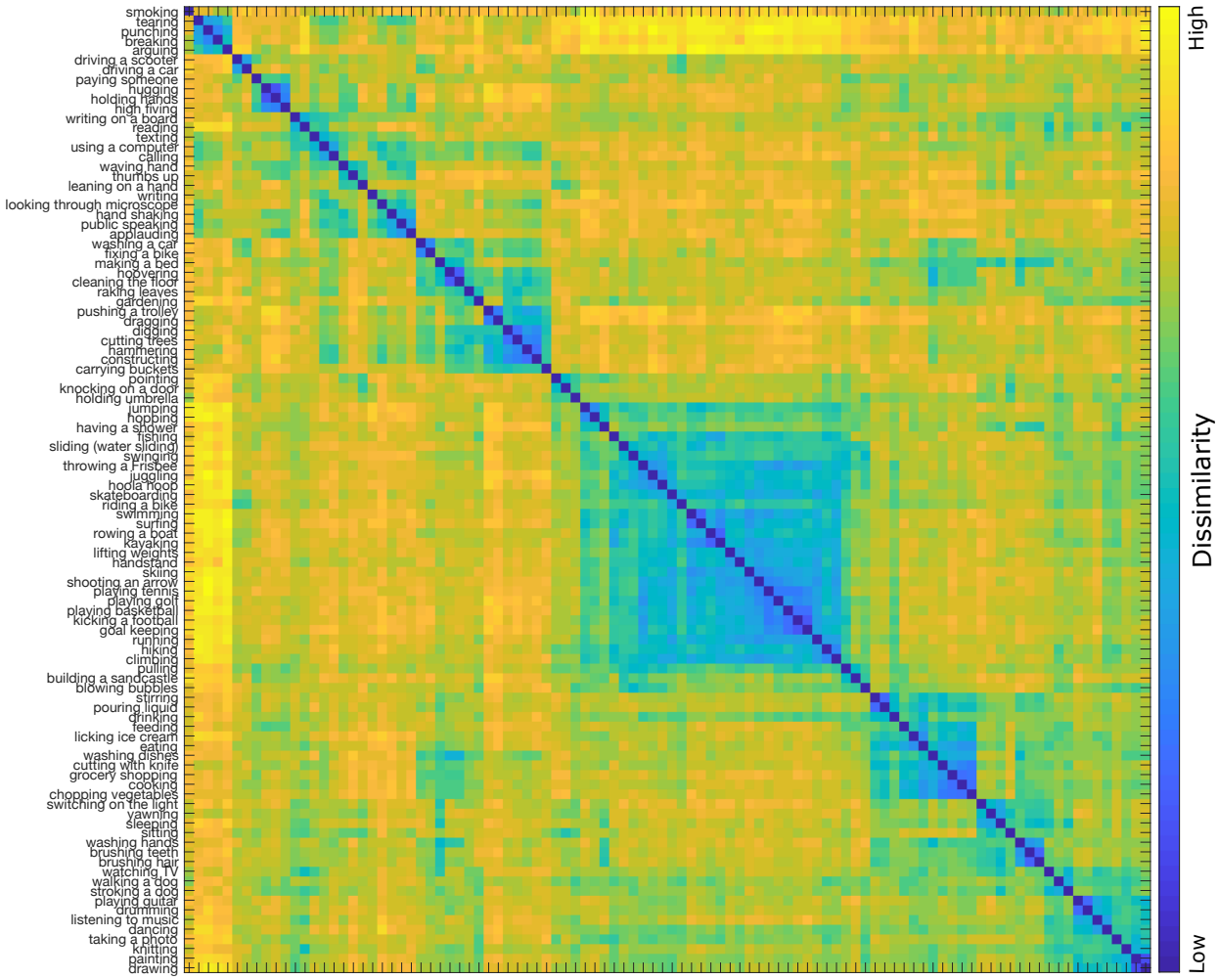


Figure A3. Results from the multi-arrangement experiment. Blue indicates low dissimilarity (high similarity), whereas yellow indicates high dissimilarity (low similarity). Labels on the x-axis are identical to labels on the y-axis.

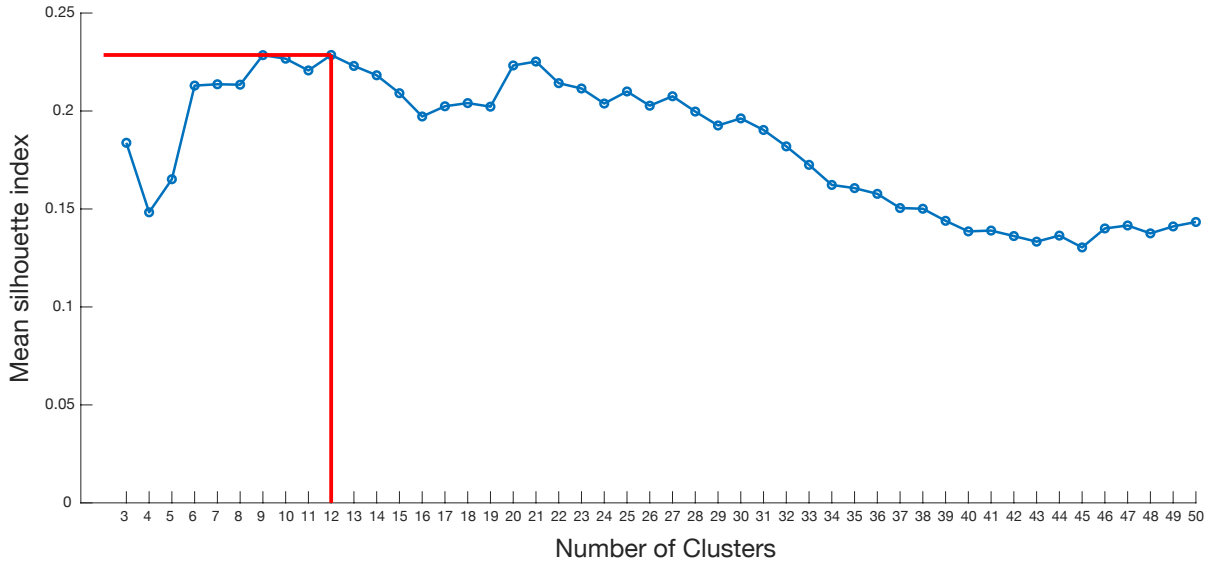


Figure A4. Mean silhouette index (*si*) as a function of the number of clusters obtained from hierarchical clustering in a range from 3 to 50. For each number of clusters, the mean silhouette index was computed by averaging across 100 iterations. The red line shows the index for the number of clusters chosen in the study (*si* = 0.23).

A.1.5 Category naming

A.1.5.1 Participants

Twenty-six participants took part in the study (21 females; mean age = 26 years, age range = 18–51 years). Experimental procedures were approved by the ethics committee at the University of Regensburg.

A.1.5.2 Instruction

Participants received the following written instruction in German (for convenience, we provide the English translation):

In this study we will ask you to find labels for sets of words. Please think about what the words have in common and try to come up with a "heading" that subsumes all of the words equally. You can give us multiple examples. Please write the label that fits best at first. There is no time limit so you can take as much time as you want. Below, you can find two examples.

1. Example

- *to brake*

- *to accelerate*
- *to blink*
- *to turn*

Possible examples for category labels could be: driving a car, riding a motorcycle, etc.

2. Example

- *Karate*
- *to kick someone*
- *to hit with a fist*
- *to shout*
- *to use a weapon*
- *to break a window*

Possible examples for category labels could be: fighting, destroying, etc.

A.1.5.3 Apparatus

The study was conducted using an online survey (https://www.soscisurvey.de/).

A.1.5.4 Procedure

We only asked participants to provide names for clusters that contained at least two different actions. This was the case for 11 out of 12 clusters. Separately for each of the 11 clusters, participants were provided with a list of all action words (in German) belonging to a given cluster. The order of the words was randomized for each participant. Participants were instructed to provide a label that could best describe a given cluster by considering what the words had in common (see Section A.*1.5.2* for the exact wording of the instruction). Participants were allowed to provide more than one label per category.

A.1.5.5 Data analysis

In total, participants produced 493 labels (44.82 per category). To choose labels best describing each action category, we took into account the frequency of provided labels. In Figure S5 we visualized the collected category labels using word clouds, where more frequently provided labels are shown with a bigger font (in order to keep the original labels, we provide them in German). Table A2 contains all the labels provided by the participants (in German), together with the frequency of mentions. Taking into account the most frequently mentioned labels for each cluster, we selected subsequent category labels: *Aggressive actions*, *Communication*, *Food-related actions*, *Gestures*, *Hand-related actions*,

Hobby, *Household-related actions*, *Interaction*, *Locomotion*, *Morning routine*, and *Sport-related actions*. The final list of action categories and the corresponding actions is provided in Table A3.

Figure A5. Word cloud forms of category labels obtained from the Category naming experiment (carried out in German). Font size is proportional to the frequency of the labels. Above each word cloud, the English translation of the label chosen for the action category is shown. For a detailed list of labels and their frequency see Table A2.

Table A2. Labels obtained in the Category naming experiment. Numbers on the left indicate the frequency of each label. Labels are organized by frequencies; labels with similar frequencies are sorted alphabetically. Labels above each list are the final action category names.

Aggressive actions

9	Aggression
9	Wut
7	Gewalt
5	Konflikt
4	Zerstörung
3	Aggressivität
2	Ärger
2	Zerstören
1	Aggressionsbewältigung
1	Aggressive Handlungen
1	Auseinandersetzung
1	Eskalation
1	Gefühlsausbruch
1	Gewalt ausüben
1	Kämpfen
1	Kaputt machen
1	negative Emotionen
1	physischen oder psychischen Druck ausüben
1	Probleme
1	sich aggressiv verhalten
1	Streit
1	Verletzen
1	Wutausbruch

Communication

11	Kommunikation
4	Kommunizieren
4	Sprache
2	Mediennutzung
1	Alltag
1	Alltagsaufgaben
1	Arbeit
1	Arbeiten/ Schule
1	Beruf
1	Büro
1	Ideen austauschen
1	Informationen austauschen
1	Informationen generieren
1	Interaktion mit einer anderen Person
1	Kommunikationsmedien nutzen
1	Lehrer:in
1	Lesen & Schreiben
1	Medien benutzen
1	mitteilen
1	Office
1	Schreibskills
1	Sprachnutzung
1	Wissen austauschen
1	Wörter

Food-related actions

4	Ernährung
4	Kochen
4	Küche
4	Nahrung
3	Lebensmittel
3	Haushalt
2	Alltag
2	Essen
1	Essen zubereiten
1	Essensbezogene Handlungen
1	Essensvorbereitungen
1	Essenszubereitung
1	Hausarbeit
1	Hausarbeiten erledigen
1	Hunger
1	Küchenarbeit
1	Küchenarbeit verrichten / In der Küche arbeiten
1	Küchentätigkeiten
1	Leben
1	Lebensmittel zubereiten
1	Mahlzeit
1	Nahrungsaufnahme
1	Nahrungsmittelverwendung
1	Nahrungszubereitung und -verarbeitung
1	Versorgung
1	Zuhause

Gestures

4	Gesten
3	Kommunizieren
2	Handbewegung
2	Hände
2	Mitteilen
2	Tätigkeiten
1	Aktivitäten mit der Hand
1	Arbeit
1	Forschungsvortrag
1	Gestik
1	Hände bewegen / benutzen
1	Handmotoriken
1	Handnutzung
1	Interaktion
1	jemanden begrüßen
1	jemanden schätzen
1	Kommunikation
1	Körperliche Tätigkeiten
1	Präsentation
1	Rede halten
1	Referat
1	Schule
1	Sprechen
1	Studieren
1	Tagung
1	Uni
1	USA Wahlen
1	Versammlung
1	Vorlesung
1	Vortrag
1	vortragen
1	Wissenschaftlicher Vortrag
1	Wissenschaftlicher Vortrag/Ausflug

Hand-related actions

2	draußen sein
2	Hände
1	Aktionen, die mit der Hand ausgeführt werde
1	Aktive Darbietung
1	Aktivitäten mit Händen
1	Ankommen
1	Anstand
1	Besuch
1	Fremdenführer
1	Geburtstagsfeier
1	Gegenstände benutzen
1	Handbewegungen im Alltag
1	Hände bewegen / benutzen
1	Handeln
1	händisch
1	Handlungen mit einer Person
1	Handmotorik
1	Handnutzung
1	Haptik
1	Hausbesuch
1	Interaktion mit Gegenständen
1	Outdoor Aktivitäten
1	Tätigkeiten
1	Tätigkeiten mit einer Hand
1	Vertreter
1	Vertreterbesuch
1	zielgerichtete Handlungen

Hobby

16	Hobby
8	Freizeit
8	Freizeitaktivitäten
2	Freizeitgestaltung
2	Kreativität
2	Spaß
1	Alltag
1	Erholung
1	Freizeitaktivitäten ausführen
1	Freizeitbeschäftigungen
1	kreativ
1	Unterhalten
1	Zurückzug

Household-related actions

6	Haushalt
3	Hausarbeit
3	Heimwerken
2	Handwerk
2	Instandhaltung
2	Ordnung
2	Tätigkeiten am Haus
1	Alltagstätigkeiten
1	Arbeit
1	außerberufliche Tätigkeiten
1	Bauernhof
1	Besitz pflegen
1	den Haushalt machen
1	Handwerken
1	Haus und Garten
1	Haus und Gartenarbeit
1	Haus- & Gartenarbeit verrichten
1	Hausarbeiten
1	Haushaltsaktivitäten
1	Haushaltsarbeit
1	Haushaltsarbeiten
1	Haushaltstätigkeiten
1	Heimwerkeln
1	Hobbys und Aktivitäten mit Händen
1	Körperliche Aktivität
1	Körperliche Arbeit am Haus und Grundstück
1	Körperliche Tätigkeiten
1	Kraft ausüben
1	Ordnung schaffen
1	Putzen
1	Routinearbeiten
1	Zuhause

Interaction

4	Interaktion
3	Kontakt
2	Geste
2	Interagieren
2	Zwischenmenschliche Aktionen
1	Austausch
1	berühren
1	Berührung
1	Beziehung
1	Date
1	Freundschaft
1	Handaktionen
1	Handbewegungen
1	händische Tätigkeiten
1	Handkontakt
1	Interaktion mit Menschen
1	Interaktion zwischen menschen
1	jemanden treffen
1	Kontakt mit Menschen
1	Körperkontakt
1	sich berühren
1	sich mit anderen Personen austauschen
1	soziale Interaktion
1	Tätigkeiten mit den Händen und Armen
1	Verabschiedung
1	Verbundenheit
1	Wertschätzung zeigen
1	Zusammenarbeiten
1	Zusammenhalt
1	Zwischenmenschlicher Austausch

Locomotion

9	Fortbewegung
5	Fahren
3	Transportmittel
3	Verkehrsmittel
2	Fahrzeug fahren
2	Fortbewegen
2	Lenken
2	Mobilität
2	sich fortbewegen
2	Straßenverkehr
2	unterwegs sein
1	am Straßenverkehr teilnehmen
1	Fahrzeug benutzen
1	Fortbewegungsmöglichkeiten
1	Führerschein
1	Lokomotion
1	Maschinelle Fortbewegung
1	Motorisierte Transportmittel
1	Personentransport
1	Reise
1	Straße benutzen
1	Transport
1	Transportieren
1	Transportmittel benutzen
1	Verkehr

Morning routine

8	Morgenroutine
5	Aufstehen
3	Abend
3	Abendroutine
1	Aktionen
1	Aktivitäten
1	Alltag
1	Alltagshandlungen ausführen
1	aufwachen
1	Bad
1	Badezimmer
1	Bettfertig machen
1	Fertig machen für den Tag
1	Handeln
1	Handlungen
1	Hygiene
1	Ins Bett gehen
1	Körperliche Bewegung
1	Körperliche Tätigkeiten
1	Morgen
1	Morgens
1	morgens aufstehen
1	Nachtschlafvorbereitung
1	Regeneration
1	runterfahren
1	Schlafen gehen
1	Schlafzimmer
1	sich fertig machen für das Schlafen
1	tägliche Routine
1	Tätigkeiten
1	Wach werden

Sport-related actions

15	Sport
8	Hobby
4	Freizeit
3	Aktivitäten
3	Bewegung
2	Freizeitaktivitäten
2	Körperliche Aktivität
2	Urlaub
1	Abenteuer
1	Bewegen
1	Fitness
1	Freizeitaktivitäten ausführen
1	Körperbezogene Aktivitäten
1	Körpereinsatz Bewegung
1	Körperliche Betätigung
1	Leibesübungen
1	Outdooraktivitäten
1	sich betätigen
1	sich bewegen
1	Spaß
1	Spiel
1	Sport machen
1	Sport treiben
1	Sportaktivitäten
1	Sportarten
1	sportliche Aktivitäten
1	Sportliche Tätigkeiten

Table A3. List of categories and corresponding actions obtained in Experiment 1.

Category	Actions
Aggressive actions	Arguing; Breaking; Punching; Tearing
Communication	Calling; Reading; Texting; Using a computer; Writing on a board
Food-related actions	Chopping vegetables; Cooking; Cutting with knife; Drinking; Eating; Feeding; Grocery shopping; Licking ice cream; Pouring liquid; Stirring; Washing dishes
Gestures	Applauding; Hand shaking; Leaning on a hand; Looking through microscope; Public speaking; Thumbs up; Waving hand; Writing
Hand-related actions	Holding umbrella; Knocking on a door; Pointing
Hobby	Dancing; Drawing; Drumming; Knitting; Listening to music; Painting; Playing guitar; Stroking a dog; Taking a photo; Walking a dog; Watching TV
Household-related actions	Carrying buckets; Cleaning the floor; Constructing; Cutting trees; Digging; Dragging; Fixing a bike; Gardening; Hammering; Hoovering; Making a bed; Pushing a trolley; Raking leaves; Washing a car
Interaction	High-fiving; Holding hands; Hugging; Paying someone
Locomotion	Driving a car; Driving a scooter
Morning routine	Brushing hair; Brushing teeth; Sitting; Sleeping; Switching on the light; Washing hands; Yawning
Sport-related actions	Blowing bubbles; Building a sandcastle; Climbing; Fishing; Goal keeping; Handstand; Having a shower; Hiking; Hula-hoop; Hopping; Juggling; Jumping; Kayaking; Kicking a football; Lifting weights; Playing basketball; Playing golf; Playing tennis; Pulling; Riding a bike; Rowing a boat; Running; Shooting an arrow; Skateboarding; Skiing; Sliding (water sliding); Surfing; Swimming; Swinging; Throwing a Frisbee
Cluster containing one action	Smoking

Hierarchical clustering resulted in 12 action categories (see left column). Actions belonging to a given category are provided in the right column. Categories containing at least two actions (11 categories) were used in the Category naming experiment, which was performed to generate category labels (corresponding names are provided in the left column, sorted alphabetically). One cluster that consisted of one action only (*smoking*) was discarded from further analyses (highlighted in gray, shown at the bottom of the table).

A.2 Experiment 2

A.2.1 Instructions

Participants taking part in the feature generation task received the following instructions (in German):

During the experiment, you will see a set of 25 action words and you will be asked to write down features of each action. You should write all the features which you think are relevant to describe a given action and to distinguish this action from the others. Please consider both abstract features (such as "sociality", "transitivity") and more detailed, concrete features (such as "moving fingers", "lifting up arms"). Please try to imagine a given action in many different scenes and write down features that are common.

You can type the features in boxes placed next to the action word. Please type as many features as possible, with a minimum of 5 per word. See the examples to get an idea of how to describe the action words in terms of their features.

***Playing piano**: Music related, Body-sense, Finger movement, Rapid change, Being focused*

***Talking**: Sociality, Contact, Communication, Mouth movement, Eye contact*

A.2.2 Procedure

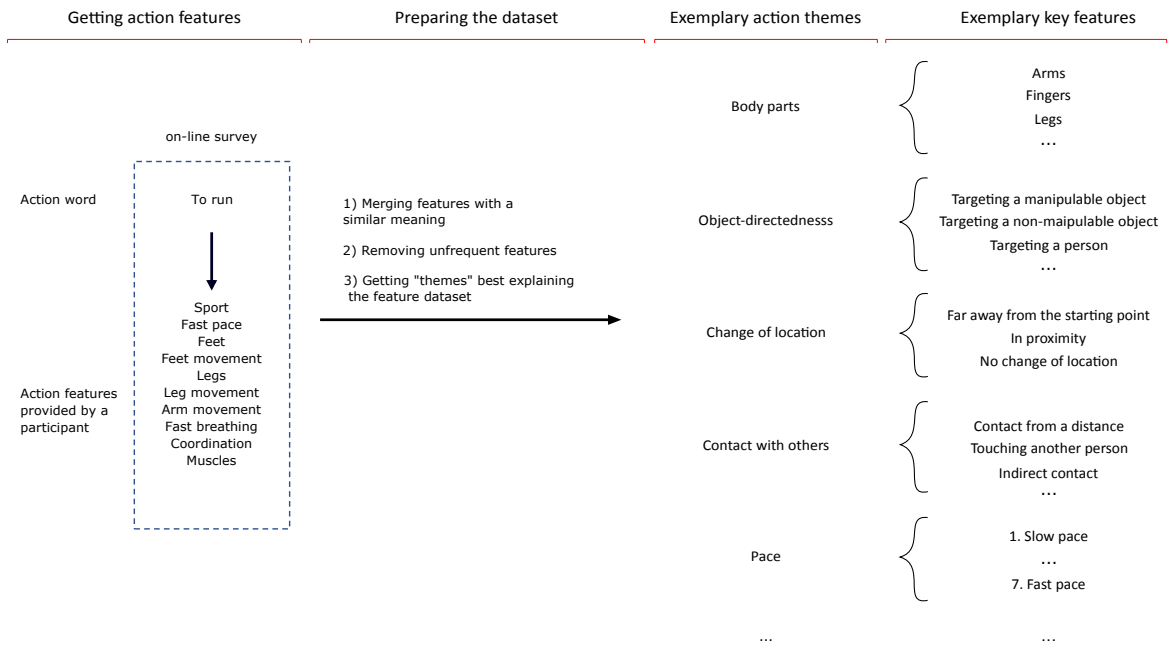


Figure A6. Generation of action themes and selection of key features. First, participants listed at least five features per action word ("Getting action features"), resulting in 5683 features in total. Next ("Preparing the dataset"), for each action, features with a similar meaning were merged, resulting in 4505 unique features (3243 unique features within the whole dataset). Guided by features examined in previous studies and the frequency of unique features, we selected a total of 59 features (see fourth column for examples), which we organized according to 19 broad action themes (see third column for examples).

A.3 Experiment 3

A.3.1 Instructions

Participants received the following written instructions (in German):

In this study, you will be asked to describe different actions based on ratings in different themes. These themes include various aspects of actions, such as: involved body parts, duration of the action, its speed, etc.

The experiment takes approximately 45 minutes and consists of 25 actions. In the course of the experiment, you may be asked to rate the same actions more than once.

Below you will find an example of the action "applying make-up" with sample answers. Please take a look at this example and the explanations of the themes to better understand what each theme means.

Unless otherwise stated, you can select more than one option for each theme.

A.3.2 Data analysis

A.3.2.1 Feature redundancy removal

During the feature rating, features for some of the themes (e.g., *Far away*, *In proximity*, *No change of location* for the theme *Change of location*; see Table A4) required binary judgments. To reduce the redundancy within the dataset, we merged features that could be expressed on a single scale. As a first step, we assigned a number to each of the features within a theme (e.g., 3, 2, 1 for the features *Far away*, *In proximity*, *No change*; see Figure A7) and subsequently transformed the ratings to one rating depictable on a scale. Figure A7 provides two example ratings. In the first example (for the action *Driving a car*), the participant indicated "Yes" for the feature *Far away*, and a "No" for the features *In proximity* and *No change of location*. This answer was transformed to the new rating "3". In

the second example (for the action *Arguing*), the participant gave the answer "No" for the feature *Far away*, and the answer "Yes" for the features *In proximity* and *No change of location*. In this case, the answer was transformed to the new rating "1.5" (i.e., the mean of the values corresponding to the two features judged with "Yes"). As a result, the number of features was reduced to 49, and nine of them could be depicted on a discrete scale. In the next step, we re-scaled ratings of all 49 features to a range of 0-1.

Table A4. List of merged binary features.

Original features	Location on the new scale	Merged features
Far away	3	Change of location
In proximity	2	
No change of location	1	
A day	6	Duration
Several hours	5	
Half an hour to an hour	4	
A few minutes to half an hour	3	
A few seconds to a few minutes	2	
Up to a few seconds	1	
Touching another person	4	Contact with others
Contact from a distance	3	
Indirect contact	2	
Does not require contact	1	

Numbers in the middle column indicate location of each "old" feature on the scale of the new theme. New feature labels are provided in the right column.

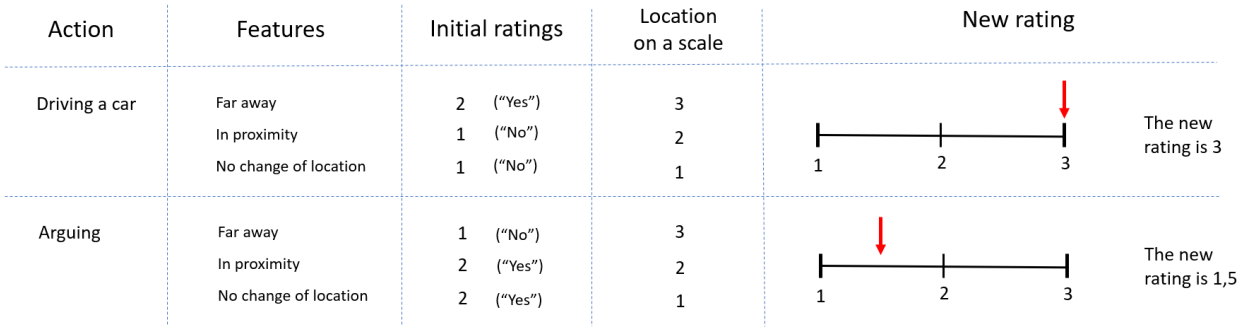


Figure A7. Feature redundancy removal. See text above for details.

A.3.2.2 Multicollinearity/Variance Inflation Factor

To estimate the amount of multicollinearity between feature ratings we computed the Variance Inflation Factor (VIF) (Table A5a) and between-feature correlations using Pearson correlation (Figure A8a). For features with a VIF > 10 and significant between-feature correlations ($p < 0.05$), we collapsed ratings. The ratings were collapsed at the level of individual participants. This resulted in the features *Upper limbs* (by collapsing *Arms* and *Shoulders*), *Hands* (by collapsing *Dominant hand*, *Both hands*, and *Fingers*) and *Lower limbs* (by collapsing *Legs*, *Hips*, and *Feet*), for a final set of 44 features in total. Results of the VIF for these merged features are shown in Table A5b and results for the between-feature correlations are provided in Figure A8b.

Table A5. Variance Inflation Factor. (a) Before and (b) after collapsing features.

a

Feature	VIF
Arms	6.365
Shoulder	6.441
Dominant hand	11.152
Both hands	4.894
Fingers	11.696
Legs	58.000
Hips	45.948
Feet	33.369
Head	5.310
Mouth	3.590
Targeting a non-manip. object	3.182
Targeting a manip. object	4.610
Targeting a tool	2.694
Targeting a person	6.570
No object involved	6.098
Horizontal	6.533
Vertical	4.979
No movement	12.164
Unspecified trajectory	3.295
Circular arms	4.746
Circular legs	4.392
Rotating arms	5.637
Rotating legs	5.745
Abduction-Adduction arms	3.404
Abduction-Adduction legs	6.892
Sweeping arms	4.706
Sweeping legs	5.572
Up-Down arms	4.512
Up-Down legs	5.566
Straight posture	4.122
Bending	3.789
Sitting	3.371
Laying	3.214
No specific posture	4.436
Indoor	10.497
Outdoor	2.762
Keeping balance	4.706
Harm	3.393
Water	2.591
Season-dependence	6.964
Change of location	6.827
Duration	6.451
Contact with others	5.168
Pace	4.880
Use of force	6.600
Goal-directedness	3.269
Concentration	3.676
Noise	3.942
Valence	2.577

b

Feature	VIF
Upper limbs	5.301
Hands	5.339
Lower limbs	10.663
Head	4.665
Mouth	2.984
Targeting a non-manip. object	2.379
Targeting a manip. object	3.863
Targeting a tool	2.513
Targeting a person	6.103
No object involved	4.461
Horizontal	5.944
Vertical	4.411
No movement	11.169
Unspecified trajectory	3.077
Circular arms	4.242
Circular legs	3.802
Rotating arms	5.082
Rotating legs	5.208
Abduction-Adduction arms	2.945
Abduction-Adduction legs	5.623
Sweeping arms	4.260
Sweeping legs	4.248
Up-Down arms	3.805
Up-Down legs	5.316
Straight posture	3.715
Bending	2.928
Sitting	3.214
Laying	2.980
No specific posture	4.962
Indoor	8.321
Outdoor	2.703
Keeping balance	4.426
Harm	3.268
Water	2.316
Season-dependence	5.286
Change of location	6.661
Duration	5.881
Contact with others	4.874
Pace	3.861
Use of force	5.568
Goal-directedness	3.016
Concentration	3.262
Noise	2.949
Valence	2.411

Different colors (i.e. yellow, green, and blue) indicate which features were merged together (column **a**), and which final features were formed based on them (column **b**).

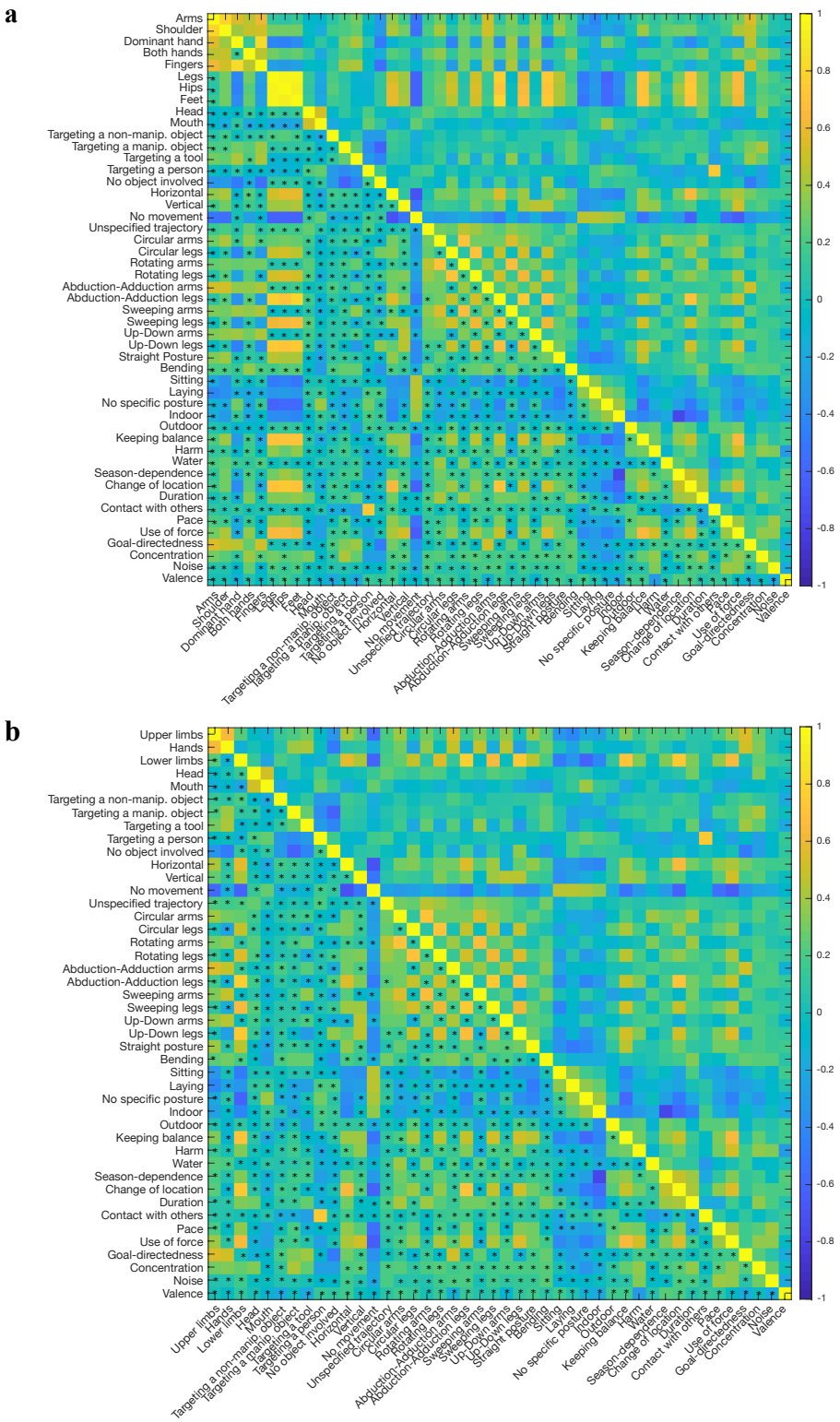


Figure A8. Between-feature correlations. Stars indicate significant correlation between features, corrected for multiple comparisons (FDR, $p < 0.05$). **(a)** Before averaging across highly correlated features (see Section A.3.2.2). **(b)** After averaging across highly correlated features.

A.3.3 Results

A.3.3.1 Multi-feature model

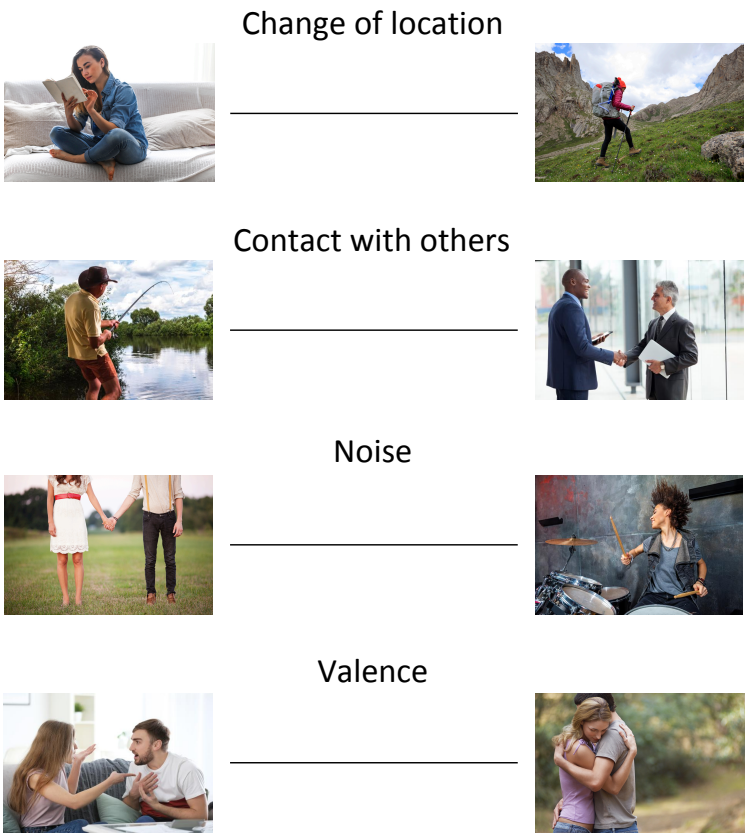


Figure A9. Actions that received minimum (left column) and maximum (right column) ratings for some exemplary features (*Change of location*, *Contact with others*, *Noise*, *Valence*).

A.3.3.2 Feature-based representations of all 11 categories

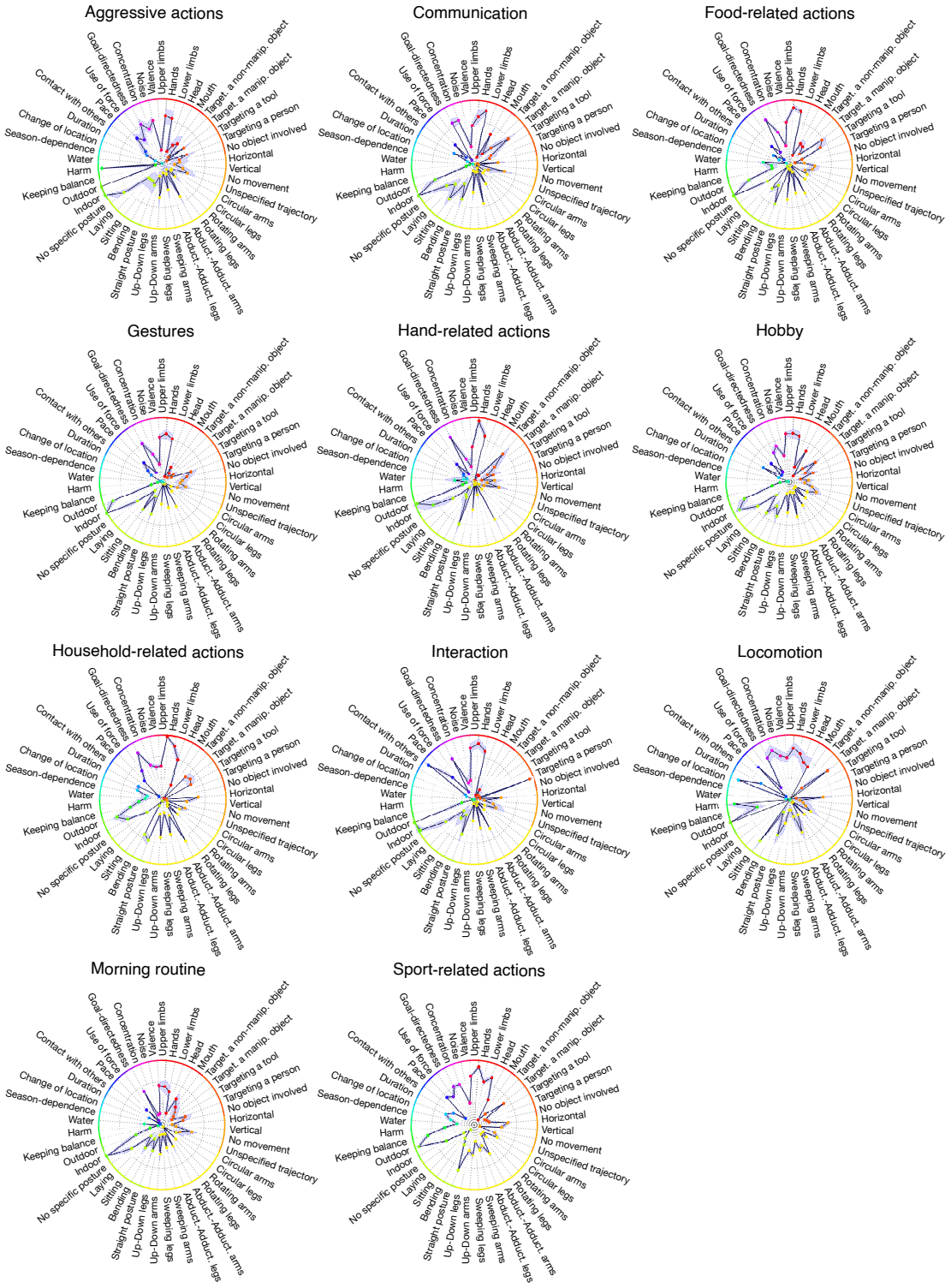
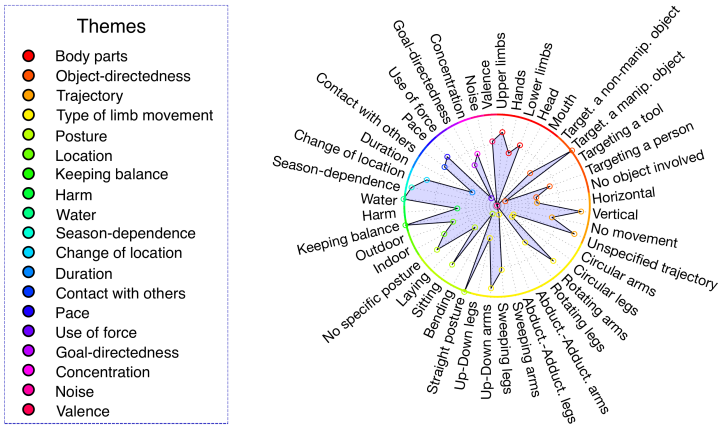
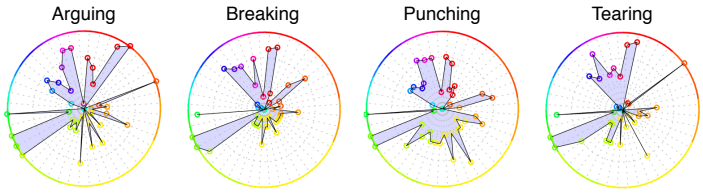


Figure A10. Feature-based representations of all 11 action categories. Different colors indicate features belonging to the same theme. The length of the spikes corresponds to the averaged rating of the corresponding feature for that category. Shaded area indicates a 95% confidence interval across actions within each category.

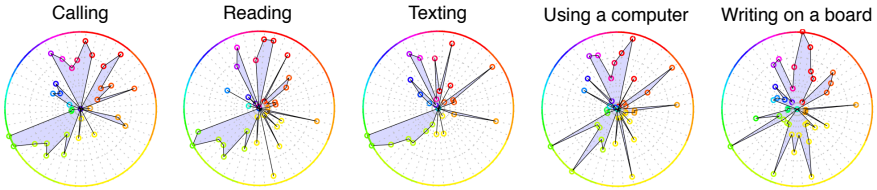
A.3.3.3 Feature-based representations of all 100 actions



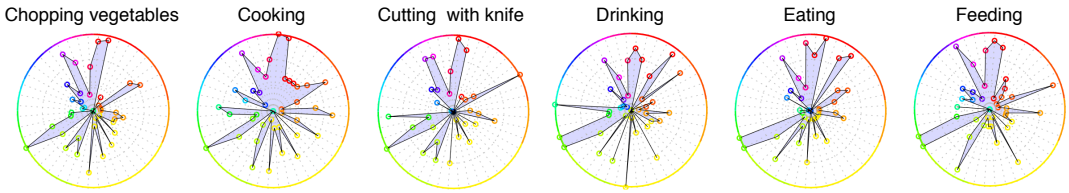
Aggressive actions



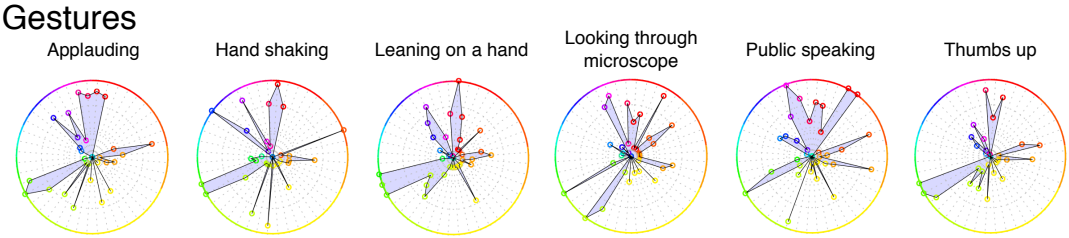
Communication

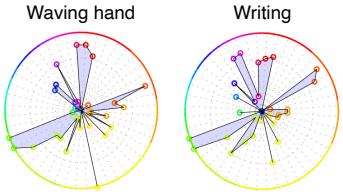


Food-related actions



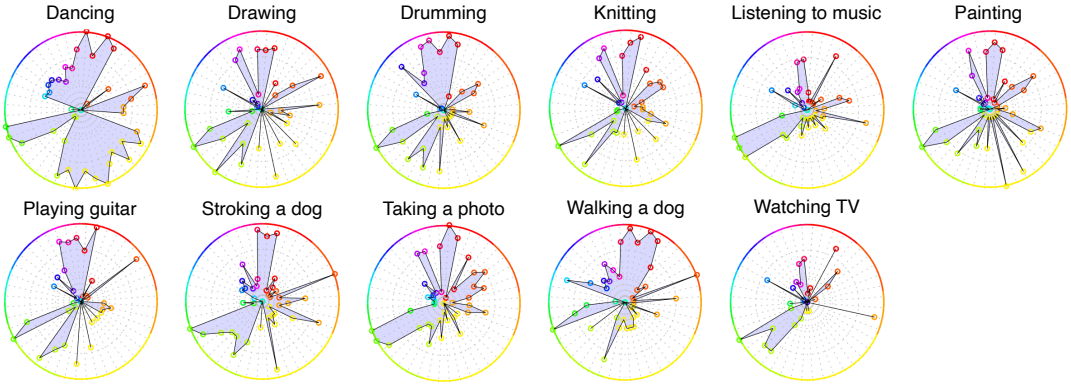
Gestures



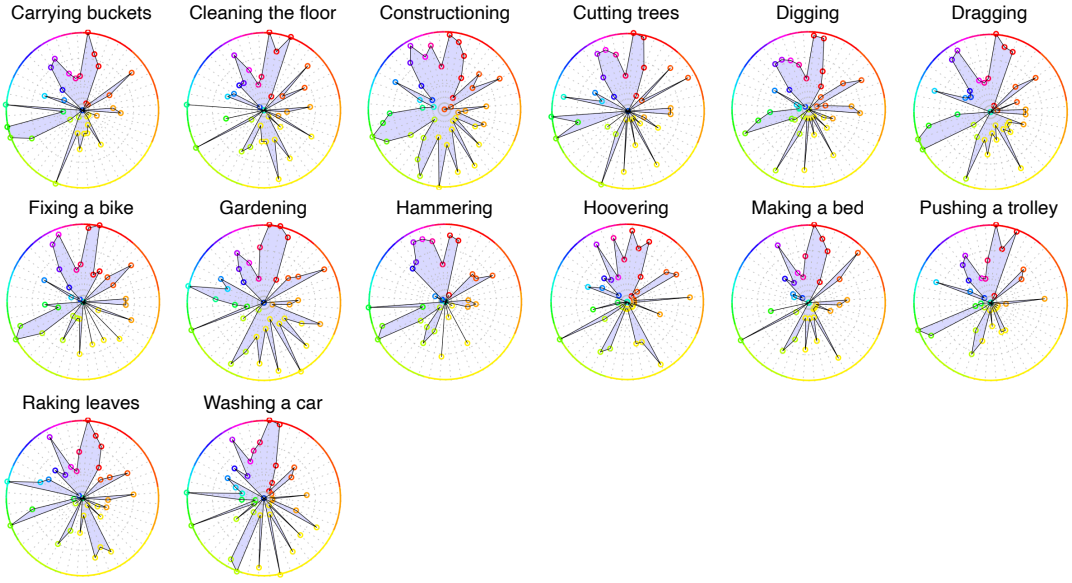


Hand-related actions

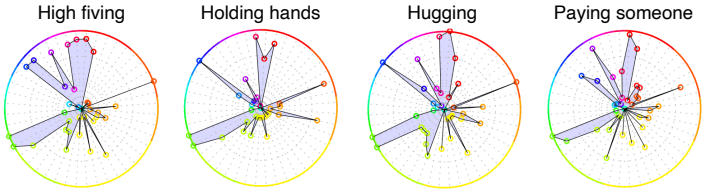
Hobby



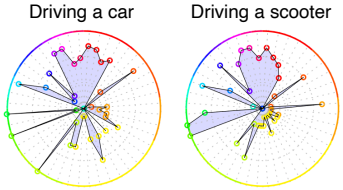
Household-related actions



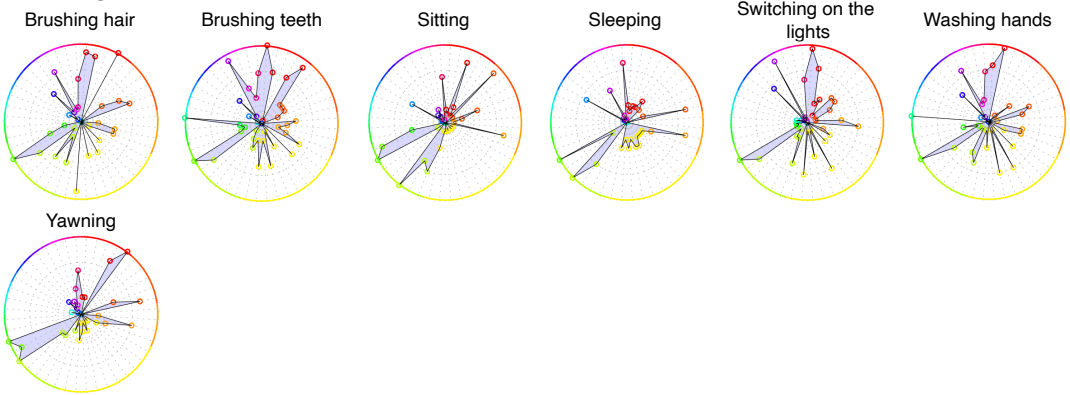
Interaction



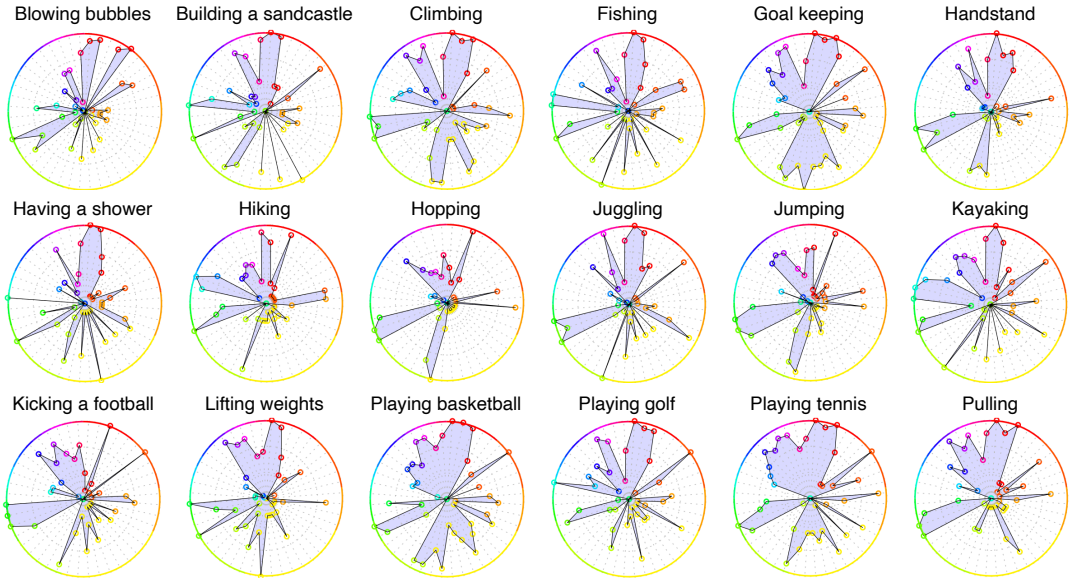
Locomotion



Morning routine



Sport-related actions



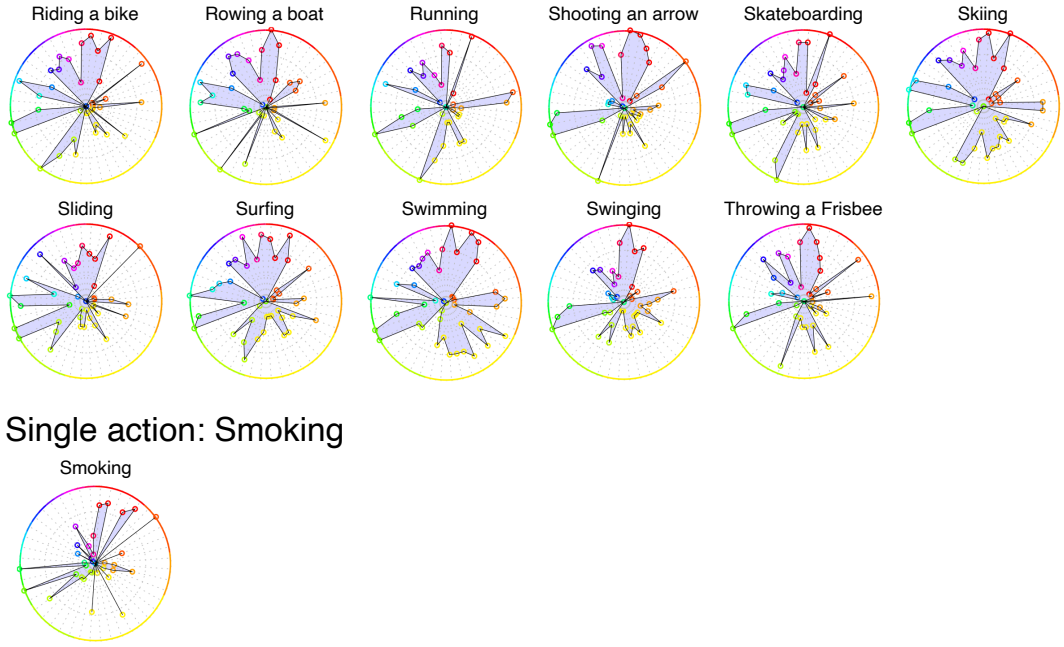
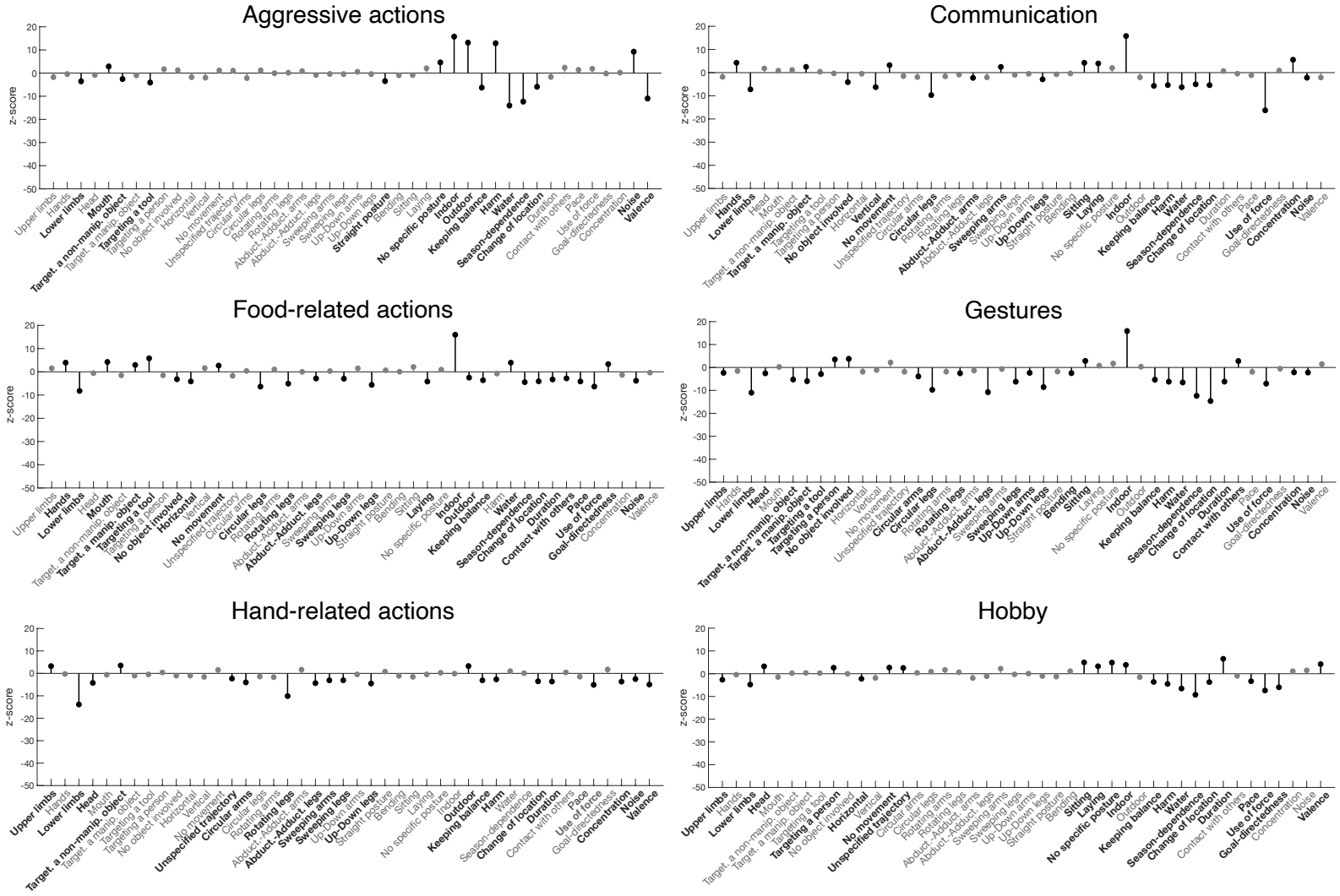


Figure A11. Feature-based representations of individual actions grouped by action categories. One hundred actions are depicted as radial plots and grouped by the categories obtained from Experiment 1. Radial plots show the importance of various features (obtained from Experiment 2) ranging from 0 to 1. Features are color-coded, with features belonging to one theme indicated by the same color (same color code as Figure 2.2). A list of the themes and corresponding colors is provided in the legend of the exemplary radial plot on the top of the figure.

A.3.3.4 Quantitative differences between action categories



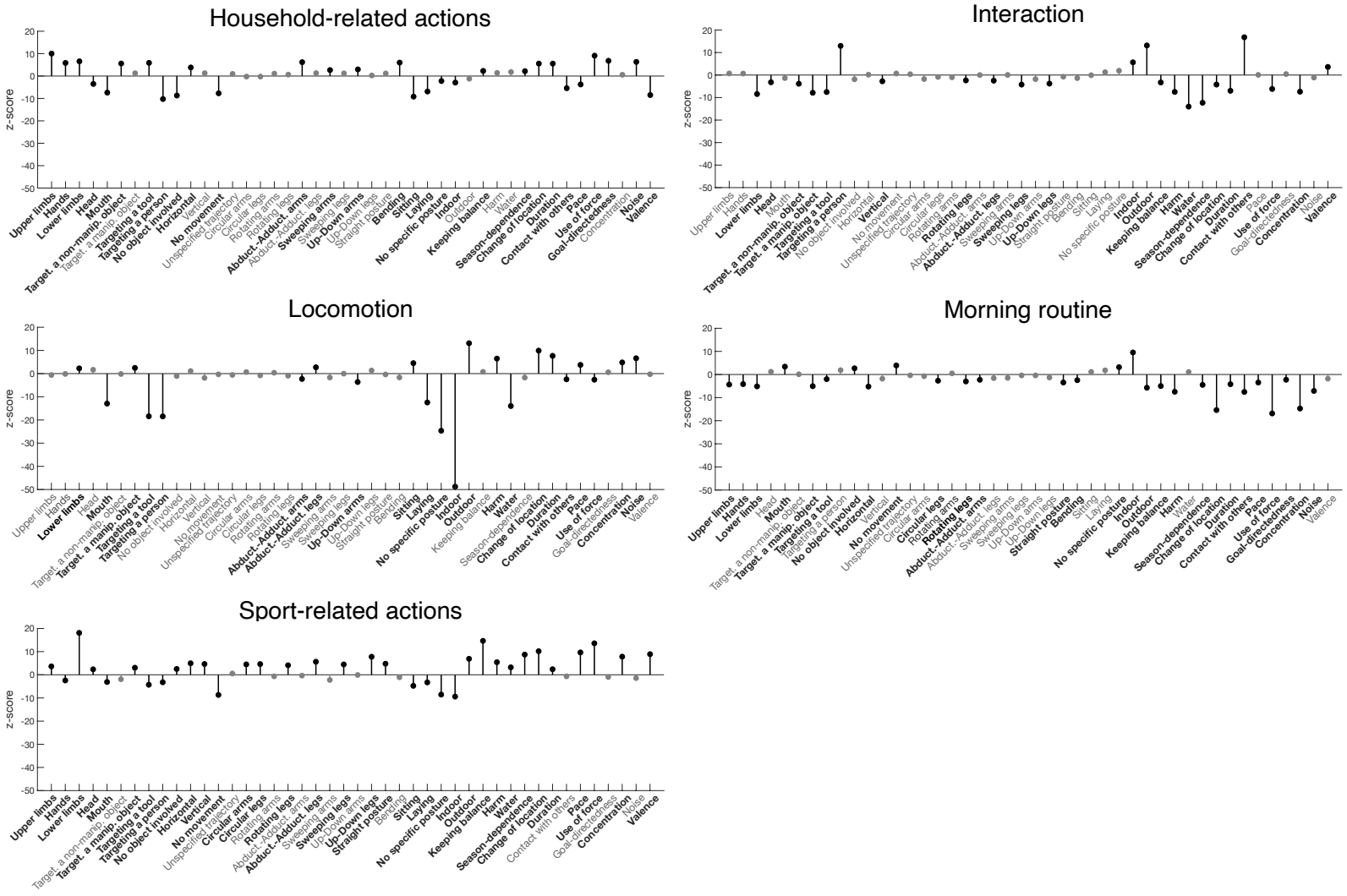


Figure A12. Quantitative differences for feature ratings of individual action categories in comparison to mean ratings obtained for the remaining categories, expressed as z-scores. This comparison reveals crucial features that distinguish between action categories. Significant pairwise comparisons are indicated by bold feature labels on the x axes ($p < 0.05$, FDR corrected).

A.3.3.5 Feature RDMs

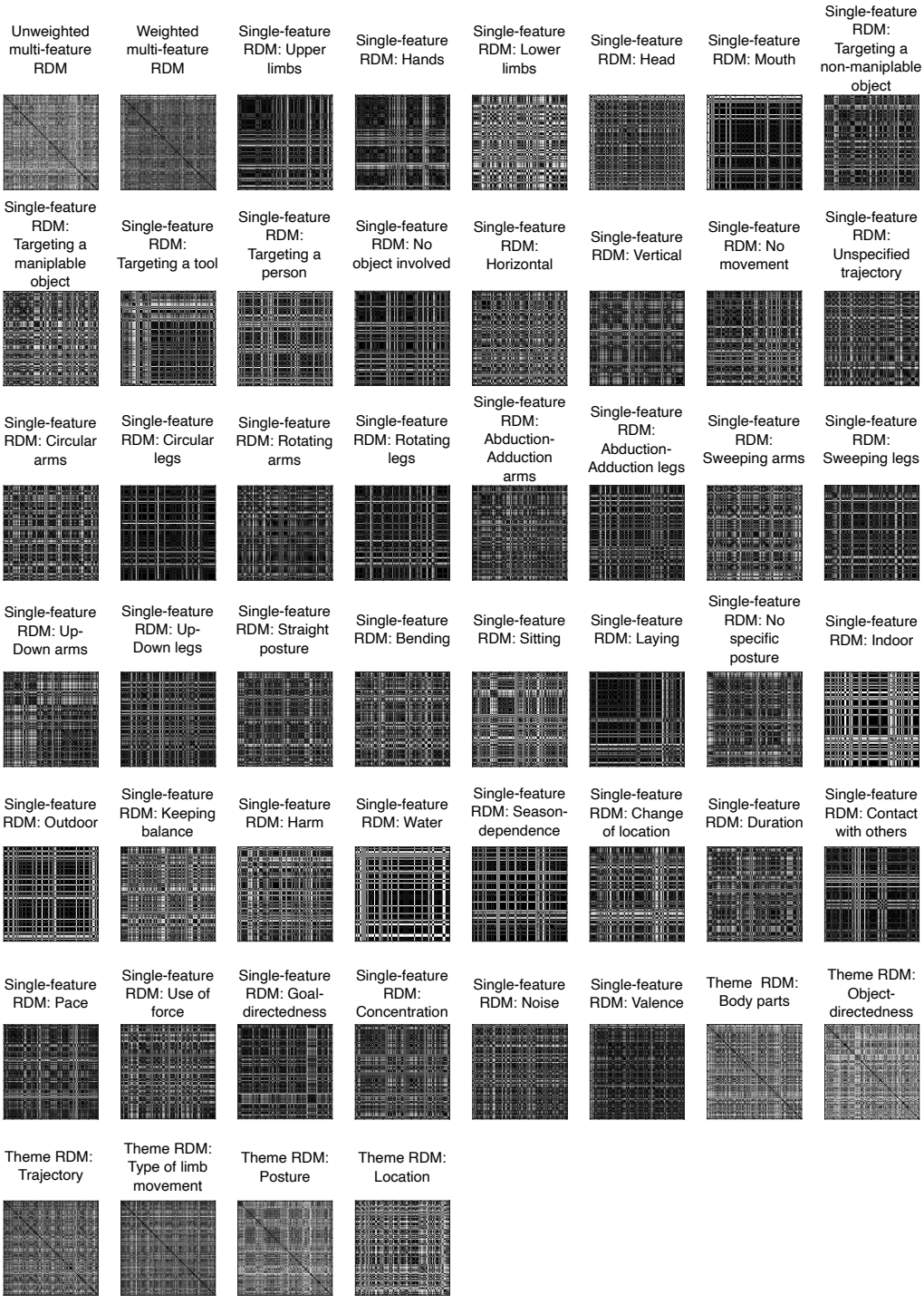


Figure A13. Feature RDMs used to correlate with the category model: unweighted multi-feature RDM, weighted multi-feature RDM, 44 single-feature RDMs, and six theme RDMs. Black squares indicate high similarity between the actions whereas white squares indicate low similarity. The RDMs were obtained by computing the Euclidean distance between pairs of actions.

A.3.3.6 Correlation between category- and feature-based representations

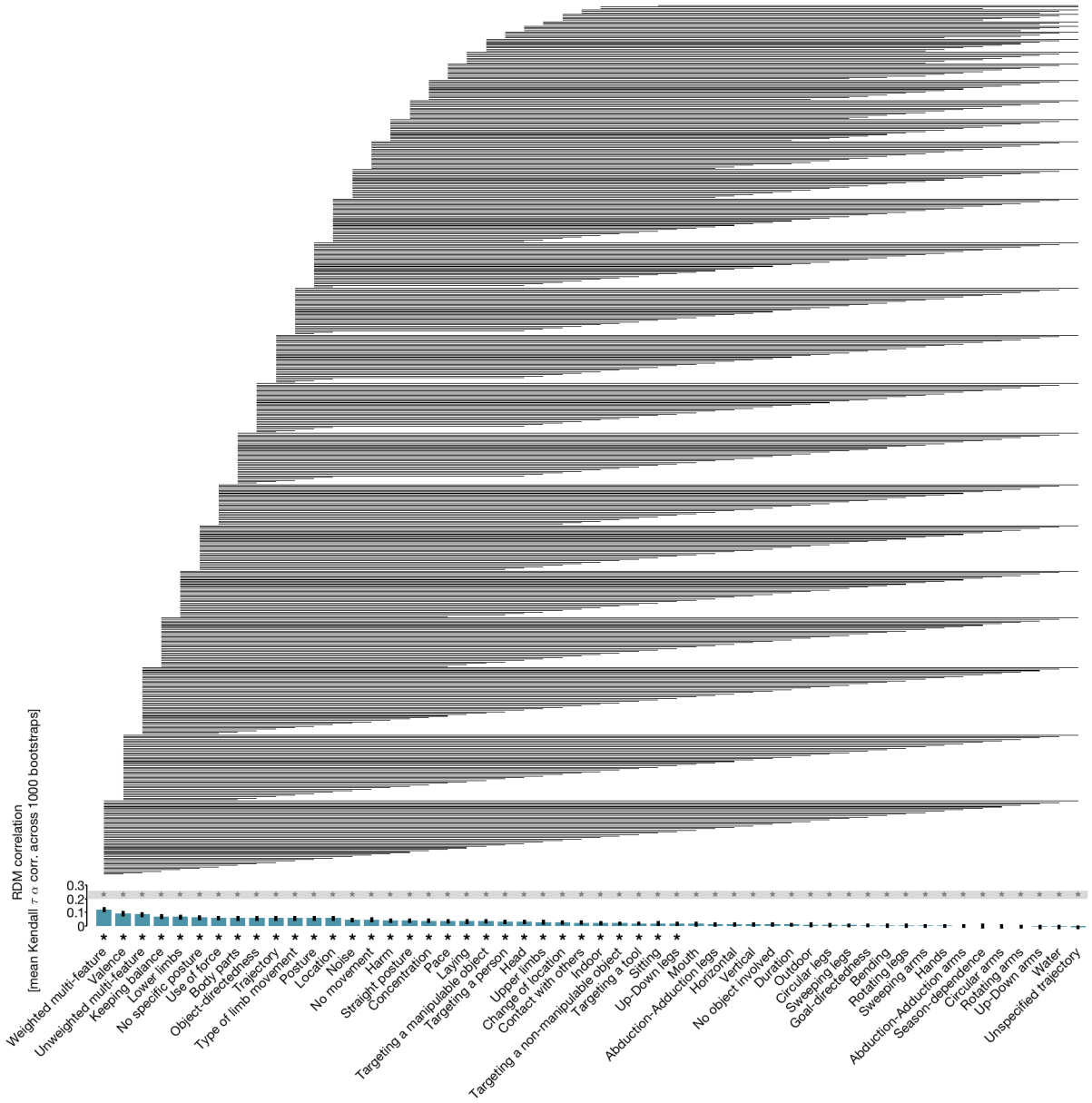


Figure A14. Correlation between category RDM (resulting from Experiment 1) and different feature RDMs (see Section *Experiment 3, Data analysis, Correlation of category- and feature-based models*, for details). The figure is an extension of Figure 2.4a. Significant differences between the feature RDMs are indicated by horizontal lines above the bars (stimulus bootstrap test, $p < 0.05$, FDR corrected).

Table A6. Correlations between the category RDM (resulting from Experiment 1) and feature RDMs (resulting from Experiment 3). Feature RDMs are organized in a descending manner (same order as in Figure A14).

Feature RDM correlated with category RDM	Kendall's τ_A correlation
Weighted multi-feature	0.1206
Valence	0.0926
Unweighted multi-feature	0.0850
Keeping balance	0.0701
Lower limbs	0.0671
No specific posture	0.0621
Use of force	0.0588
Body parts	0.0573
Object-directedness	0.0573
Trajectory	0.0573
Type of limb movement	0.0573
Posture	0.0573
Location	0.0573
Noise	0.0474
No movement	0.0467
Harm	0.0421
Straight posture	0.0417
Concentration	0.0386
Pace	0.0367
Laying	0.0354
Targeting a manipulable object	0.0348
Targeting a person	0.0318
Head	0.0314
Upper limbs	0.0281
Change of location	0.0279
Contact with others	0.0249
Indoor	0.0238
Targeting a non-manipulable object	0.0203
Targeting a tool	0.0198
Sitting	0.0197
Up-Down legs	0.0193
Mouth	0.0157
Abduction-Adduction legs	0.0135

Horizontal	0.0135
Vertical	0.0134
No object involved	0.0118
Duration	0.0116
Outdoor	0.0100
Circular legs	0.0086
Sweeping legs	0.0069
Goal-directedness	0.0060
Bending	0.0045
Rotating legs	0.0042
Sweeping arms	0.0021
Hands	0.0020
Abduction-Adduction arms	0.0017
Season-dependence	0.0017
Circular arms	0.0003
Rotating arms	-0.0015
Up-down arms	-0.0051
Water	-0.0058
Unspecified trajectory	-0.0091

A.3.3.7 Valence-based representation of actions

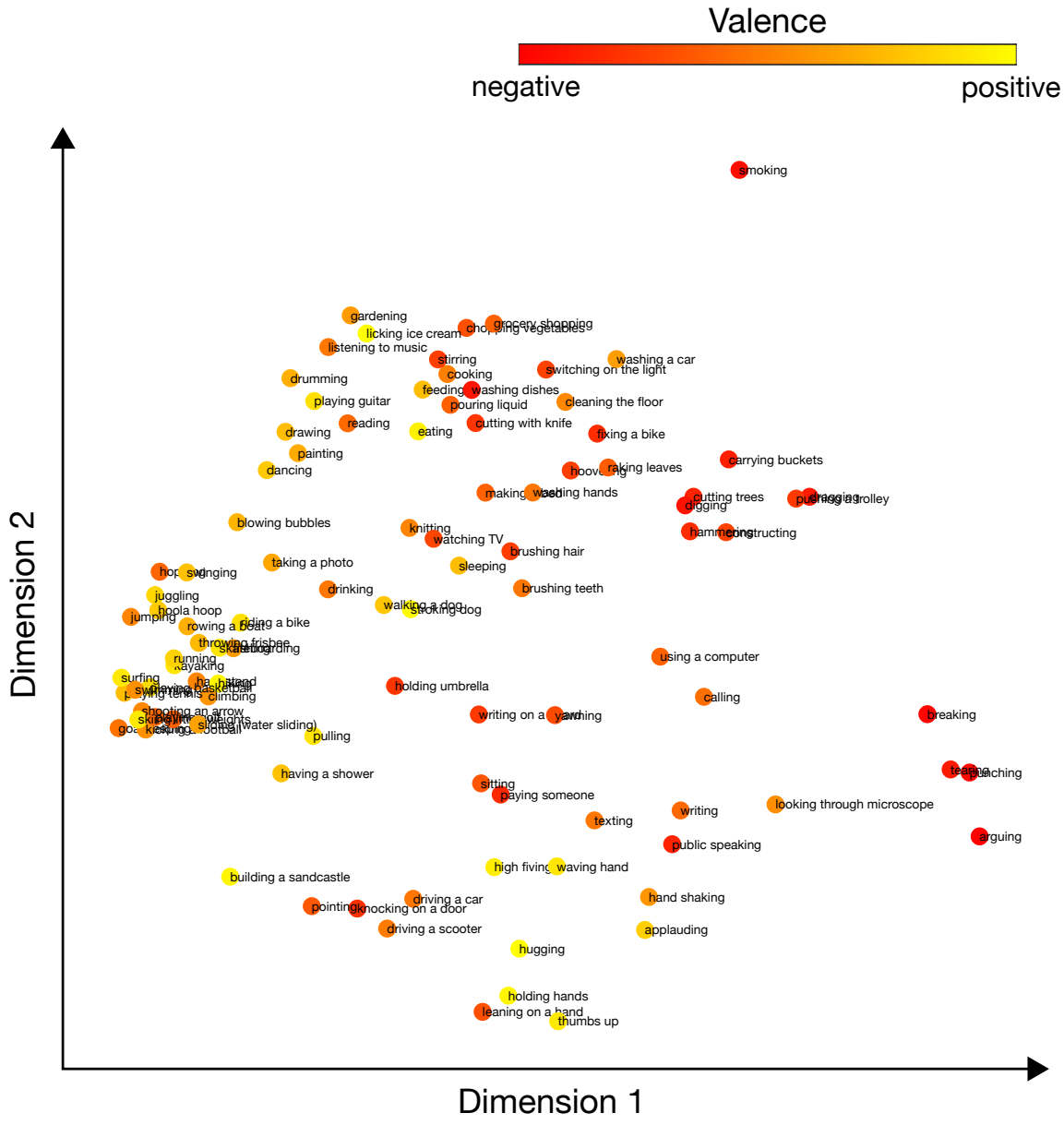


Figure A15. Valence-based representation of actions. 2-dimensional arrangement of actions obtained from the multi-arrangement experiment (Experiment 1; same representation of actions as shown in Figure 2.1), color-coded with respect to the valence ratings obtained from Experiment 3 (red: *negative valence*, yellow: *positive valence*). Based on the importance of the feature valence revealed by the results shown in Figure 2.4b, this visualization aims to better understand the organization of observed actions according to this feature.

B. Study 2 Supplementary materials

Figure B1. The complete set of stimuli used in the study. The stimulus set consisted of 100 actions, with four different exemplars per action. Here, the stimuli are organized alphabetically, presented in a grid format, progressing row-by-row. All four exemplars per action are included.

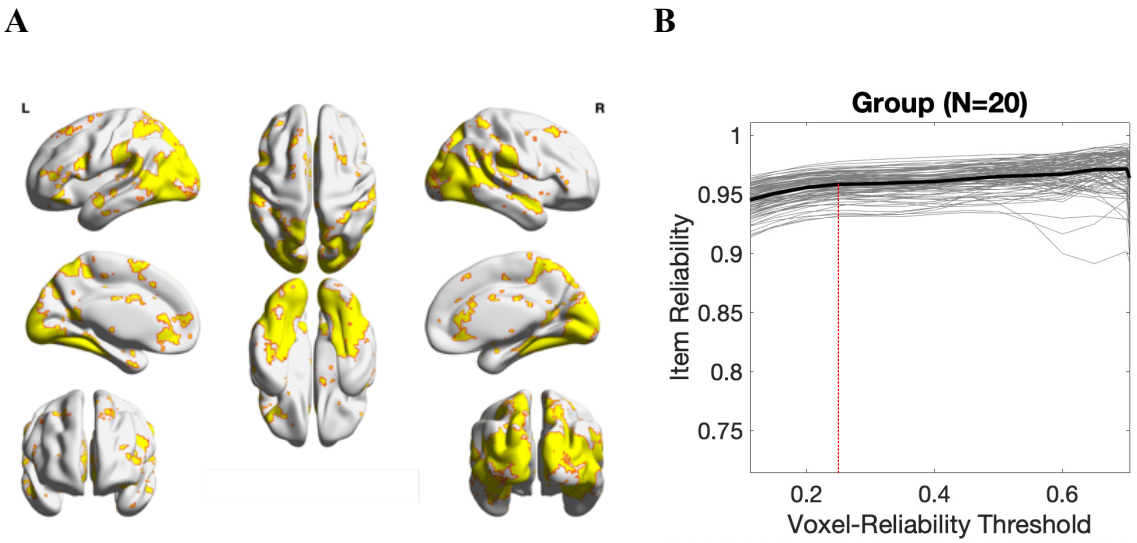


Figure B2. **A.** Reliability map including voxels above the threshold of 0.25. The voxel threshold was selected based on **B. B** Threshold selection. The red vertical line is positioned at voxel-reliability threshold equal 0.25. At this value, the curve reaches a plateau, indicating that increasing the threshold will only minimally increase the item reliability, with a risk of discarding informative voxels.

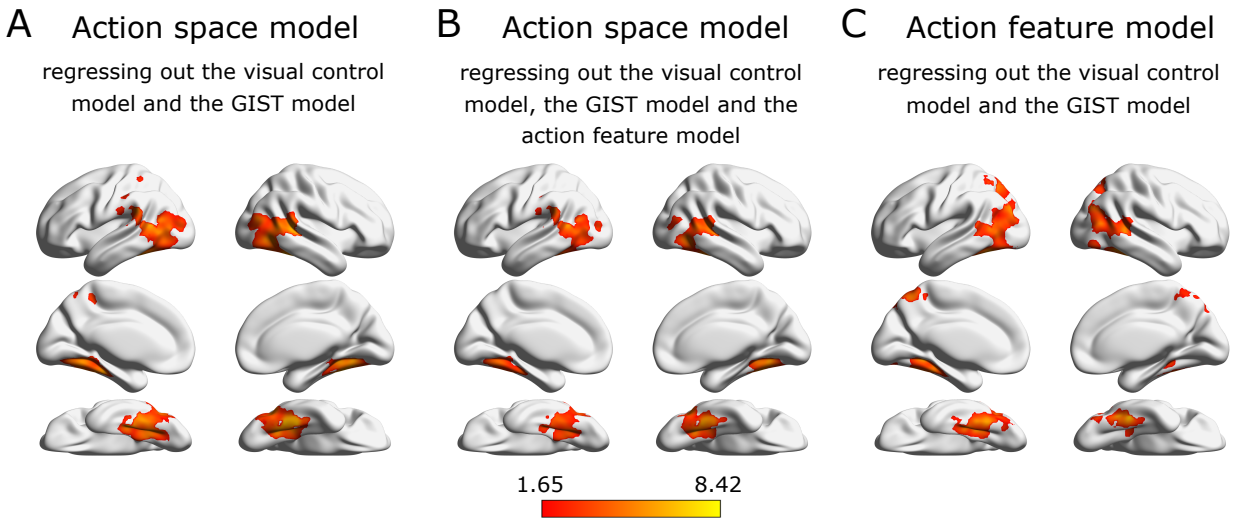


Figure B3. (see Figure 3.2 for direct comparison)**.** Results of the group-level searchlight-based RSA for: (A) the behavioral action space model (regressing out the low level visual control model and the GIST model); (B) the behavioral action space model (regressing out the low level visual control model, the GIST model and the action feature model; (C) the action feature model (regressing out the low level visual control model and the GIST model). Statistical maps show t-values thresholded at a z-score of 1.65, corresponding to $p < 0.05$ (one-tailed), corrected for multiple comparisons (TFCE, $p < 0.05$, 5000 Monte Carlo permutations). This figure shows the same analysis as shown in Figure 3.2, with the only difference that the low-level visual control model was constructed on the basis of the first layer of AlexNet.

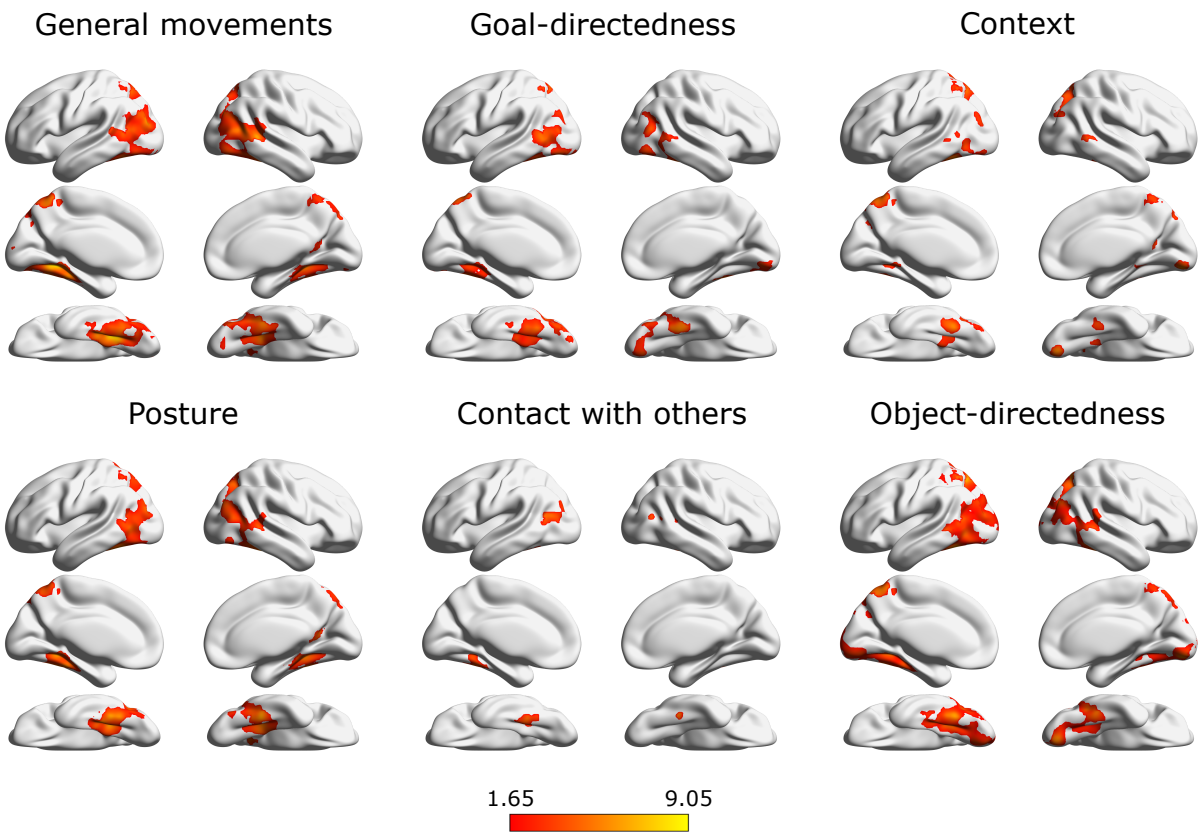


Figure B4 (see Figure 3.3 for direct comparison). Results of the searchlight RSA, carried out separately for each of the eight dimensions (regressing out the low-level visual control model and the GIST model). Six out of eight dimensions showed significant correlation with neural data after correction for multiple comparisons (TFCE, $p < 0.05$, 5000 Monte Carlo permutations). Statistical maps show t-maps thresholded using TFCE at z-score of 1.65. The remaining dimensions, namely *Arm movement* kinematics and *Negative Emotions*, did not survive the correction. This figure shows the same analysis as shown in Figure 3.3, with the only difference that the low-level visual control model was constructed on the basis of the first layer of AlexNet.

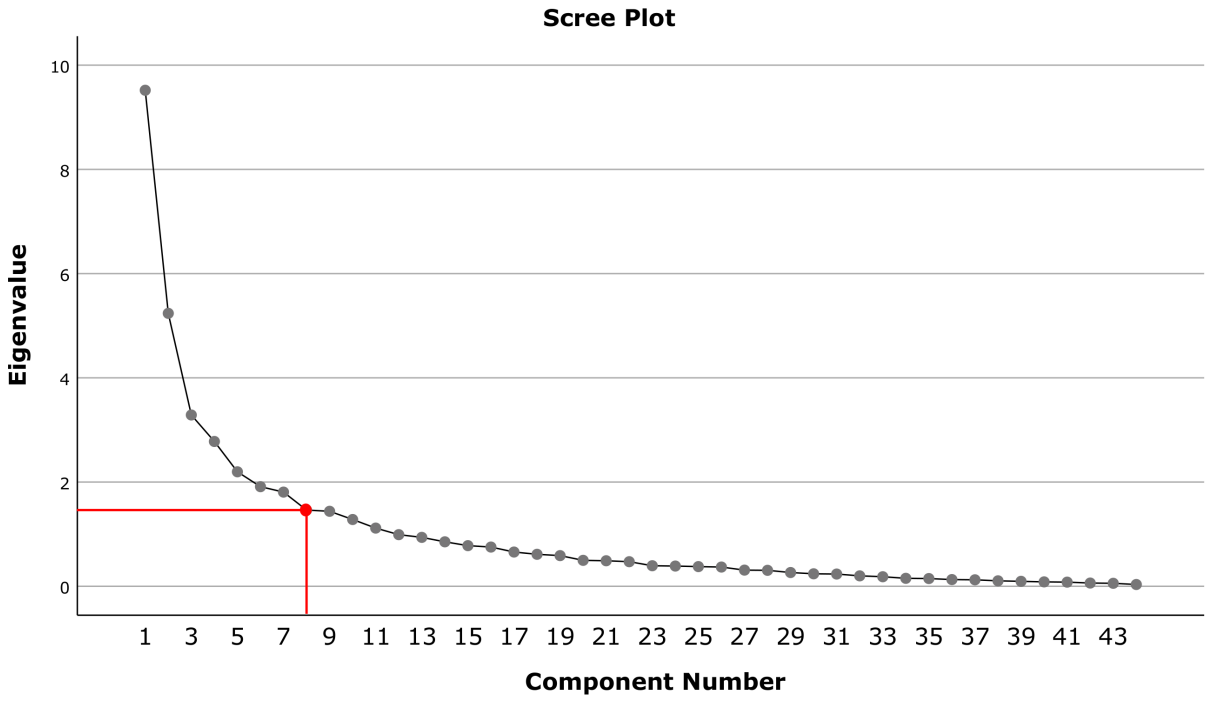


Figure B5. The scree plot illustrates the eigenvalues of principal components derived from the Principal Component Analysis (PCA). The plot displays a downward curve, indicating the distribution of variance explained by each component. The first components explain a large portion of the variability within the dataset, hence their eigenvalues are high, whereas the latter components contribute only a small fraction to the overall variance. Our focus was to identify components with high eigenvalues, as they provide insights into the most important aspects of the dataset. We used the "elbow method" to locate the point on the plot where the drop in eigenvalues reaches a plateau, indicating that the next components add relatively little to the information already extracted. Based on this approach, we selected eight components (see red line).

Table B1. Results of the Principal component analysis for components with eigenvalues greater than 1. Based on the scree plot (Figure B5), we selected eight components (marked with a gray background). In the *Features* column, we listed the features that had the most positive and negative loadings for each component, indicating the importance of the respective features for a given component. The column *Dimension label* contains labels assigned to the components by the authors of the paper, based on the features associated with each component. Dimensions that survived corrections for multiple comparisons in the searchlight-based RSA (see Figure 3.3) are highlighted in bold font.

#	% of variance	Features	Dimension label
1	21.63	Lower limbs Sweeping – legs Abduction/Adduction – legs Up-down – legs Rotating – legs Circular - legs Keeping balance Change of location Use of force Sitting (*negative loading*) No movement (*negative loading*) Horizontal trajectory	**General movements**
2	11.90	Rotating – arms Sweeping – arms Circular - arms Unspecified trajectory Water	Arm movement kinematics
3	7.47	Goal-directedness Abduction/Adduction - arms Upper limbs Hands Targeting a non-manipulable object Targeting a tool	**Goal-directedness**
4	6.32	Indoor (*negative loading*) Outdoor Season-dependence	**Context**
5	4.99	Straight posture Laying (*negative loading*) No specific posture (*negative loading*)	**Posture**
6	4.34	Contact with others Targeting a person	**Contact with others**
7	4.11	Targeting a manipulable object No object involved (*negative loading*) Concentration	**Object-directedness**
8	3.32	Noise Harm Valence (*negative loading*)	Negative emotions
9	3.27	Up-down – arms Vertical trajectory	Vertical movements
10	2.91	Mouth Head	Head
11	2.54	Duration Pace Bending	Dynamic posture

Table B2. Results of (**A**) the post-session questionnaire, provided to the participants after the MRI session, and (**B**) error rates for identifying catch trials within the MRI session, individual for each participant. Data from the participants marked in red were not included in the group level analysis due to excessive head motion (two participants) and due to stopping the scan after five runs (see *Materials & Methods, Participants* for details). Data from these three participants were not included in the analysis of the behavioral or the neuroimaging data.

A					**B**
PARTICIPANT	Comfort	Tiredness	Verbalization	Task	Error rate (%)
#1	5	3	4	2	44.64
#2	4	2	3	5	35.71
#3	3	3	5	3	23.21
#4	5	2	2	6	19.64
#5	5	3	4	5	8.93
#6	3	3	6	5	33.93
#7	3	4	6	2	33.93
#8	5	4	5	4	26.79
#9	4	2	6	4	7.14
#10	5	4	5	5	42.86
#11	3	3	5	4	26.79
#12	5	3	5	5	17.86
#13	4	2	5	5	17.86
#14	3	4	5	4	16.07
#15	5	6	6	4	21.43
#16	6	3	6	5	3.57
#17	4	5	4	2	17.86
#18	5	4	3	4	7.14
#19	5	4	6	3	19.64
#20	4	3	4	5	37.50
#21	4	5	4	3	26.79
#22	5	3	5	5	28.57
#23	4	4	4	4	44.64

C. Study 3 Supplementary materials

Figure C1. Overview of the stimuli used in the fMRI experiment. The stimulus set consisted of 160 images of actions. There were 20 actions belonging to four action categories (Communication, Grooming, Ingestion, Locomotion) and each action was presented with eight exemplary images. Here, the stimuli are organized based on categories. Each row represents one action (e.g., arguing), while each column represents an example of a given action.

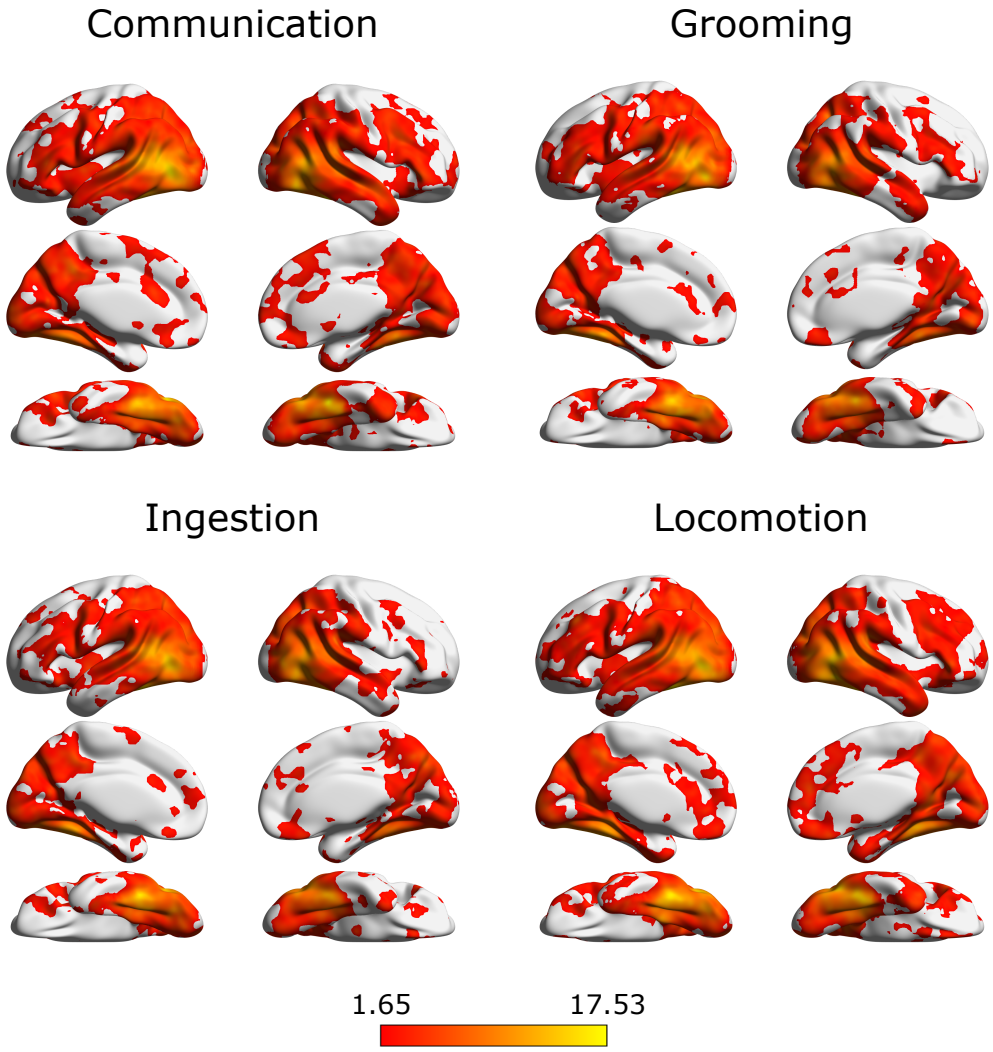


Figure C2. Mean decoding accuracy maps for the whole-brain MVPA. We performed a whole-brain searchlight analysis to investigate brain regions outside of the AON which would evoke unique for each category neural activity patterns. Decoding was performed on each category versus each other category (svm classifier) and subsequently averaged across these pairwise classifications. All the maps survived multiple comparison correction (TFCE corrected, $p < 0.05$, 5000 Monte Carlo permutations). The decoding was significant in widespread, qualitatively similar maps, consisting of the occipitotemporal, parietal and frontal regions. Due to high similarity between the maps, we did not include them in the functional connectivity analysis. See Table C2 for details on their peak coordinates.

Table C1. Stimuli used for the fMRI experiment. The stimulus set consisted of 20 actions belonging to four action categories.

Action category	Communication	Grooming	Ingestion	Locomotion
Actions	Arguing Pointing Talking Thumbs up Waving	Applying cream Applying makeup Brushing teeth Taking shower Washing hands	Drinking directly Drinking using straw Eating with chopsticks Eating with fork Eating with hands	Driving by bike Driving by scooter Roller skating Running Walking

Table C2. Coordinates of the peaks within the AON regions. The peaks were obtained from the group-level statistical maps with a contrast All Categories vs Baseline.

Hemisphere	**Region**	**t**	**Voxel location**	**MNI152**
Left	LOTC	12.17	[65 26 32]	[-40 -74 -8]
Right	LOTC	13.18	[26 26 31]	[38 -75 -10]
Left	aIPS	4.07	[57 35 54]	[-24 -56 36]
Right	aIPS	5.02	[31 37 53]	[28 -52 34]
Left	IFG	5.52	[66 67 48]	[-42 8 24]
Right	IFG	5.78	[24 69 48]	[42 12 24]

Table C3. Clusters identified in the conjunction analysis. Peaks marked with red color are those that were subsequently used in the functional connectivity analysis (three peaks with the highest t-values per category). L – left hemisphere. R – right hemisphere.

	Region	size	t	Voxel location	MNI152
All categories					
	L Occipital pole (45%)	29603	7.49	[53 13 37]	[-16 -100 2]
	L Insular cortex	1220	4.04	[58 76 36]	[-26 26 0]
	R Inferior frontal gyrus	753	3.92	[24 69 48]	[42 12 24]
	L Paracingulate gyrus	516	3.78	[47 68 59]	[-4 10 46]
	R Frontal orbital cortex	335	3.25	[29 77 36]	[32 28 0]
	L Callosal body	204	3.56	[48 65 46]	[-6 4 20]
	R Cerebellum	110	3.26	[31 30 14]	[28 -66 -44]
	L Cerebellum	101	3.55	[60 31 11]	[-30 -64 -50]
Communication					
	R Supramarginal gyrus	6028	5.32	[18 43 40]	[54 -40 8]
	L Middle temporal gyrus	5276	5.81	[74 40 40]	[-58 -46 8]
	L Precentral gyrus	3652	4.02	[63 66 50]	[-36 6 28]
	R Inferior frontal gyrus	3231	4.83	[26 70 47]	[38 14 22]
	L Cerebellum	2716	4.8	[52 25 17]	[-14 -76 -38]
	R Paracingulate gyrus	721	3.03	[41 80 53]	[8 34 34]
	R Occipital pole	314	2.9	[39 17 45]	[12 -92 18]
	R Amygdala	191	2.97	[31 59 27]	[28 -8 -18]
	R Insular cortex	121	2.84	[26 62 16]	[38 -2 -18]
Grooming					
	L Lateral occipital cortex, superior division	1634	3.41	[56 21 50]	[-22 -84 28]
	L Occipital fusiform gyrus	310	2.67	[56 29 32]	[-22 -68 -8]
	L Lateral occipital cortex, inferior division	212	2.71	[71 28 32]	[-52 -70 -8]
Ingestion					
	L Cingulate gyrus	1985	3.09	[50 81 35]	[-10 36 -2]
	R Postcentral gyrus	1831	3.25	[16 57 53]	[58 -12 34]
	L Precentral gyrus	1197	3.12	[73 63 44]	[-56 0 16]
	L Insular cortex	653	3.59	[64 60 37]	[-38 -6 2]
	R Cingulate gyrus	379	2.54	[45 49 57]	[0 -28 42]
	R Insular cortex	322	3.39	[27 61 39]	[36 -4 6]
	L Frontal orbital cortex	211	3.74	[57 76 27]	[-24 26 -18]
	L Frontal pole	164	2.68	[56 83 54]	[-22 40 36]
	R Frontal pole	125	2.31	[33 84 52]	[24 42 32]
	R Cingulate gyrus	105	2.51	[42 48 71]	[6 -30 32]
	R Middle temporal gyrus	100	2.78	[13 57 28]	[64 -12 -16]

Locomotion					
	R Parahippocampal gyrus	1187	5.07	[36 46 29]	[18 -34 14]
	L Lingual gyrus	868	5.13	[58 40 32]	[-26 -46 -8]
	Cerebellum	147	3.06	[38 41 59]	[14 -44 -16]

Table C4. Clusters identified in the searchlight MVP analysis for decoding action categories (svm classifier, z-normalized). The maps are visualized in Figure C2. L – left hemisphere. R – right hemisphere.

	Region	**size**	**t**	**Voxel location**	**MNI152**
Communication					
	R Temporal occipital fusiform cortex	96553	17.3	[23 41 26]	[44 -44 -20]
Grooming					
	L Temporal fusiform cortex	79205	17.5	[58 41 29]	[-26 -44 -14]
Ingestion					
	L Temporal occipital fusiform cortex	77803	16.9	[66 40 26]	[-42 -46 -20]
Locomotion					
	L Temporal occipital fusiform cortex	97734	17.2	[66 39 27]	[-42 -48 -18]

REFERENCES

Abdollahi, R. O., Jastorff, J., & Orban, G. A. (2013). Common and segregated processing of observed actions in human SPL. *Cerebral Cortex, 23*(11), 2734–2753. https://doi.org/10.1093/cercor/bhs264

Agosta, F., Mandic-Stojmenovic, G., Canu, E., Stojkovic, T., Imperiale, F., Caso, F., Stefanova, E., Copetti, M., Kostic, V. S., & Filippi, M. (2018). Functional and structural brain networks in posterior cortical atrophy: A two-centre multiparametric MRI study. *NeuroImage: Clinical*, *19*, 901–910. https://doi.org/10.1016/j.nicl.2018.06.013

Aguirre, G. K., & D'Esposito, M. (1999). Topographical disorientation: A synthesis and taxonomy. *Brain*, *122*(9), 1613–1628. https://doi.org/10.1093/brain/122.9.1613

Aguirre, G. K., Zarahn, E., & Esposito, M. D. (1998). An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. *Neuron*, *21*(2), 373-83. https://doi.org/10.1016/s0896-6273(00)80546-2

Alakörkkö, T., Saarimäki, H., Glerean, E., Saramäki, J., & Korhonen, O. (2017). Effects of spatial smoothing on functional brain networks. *European Journal of Neuroscience*, *46*(9), 2471–2480. https://doi.org/10.1111/ejn.13717

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends in Cognitive Sciences*, *4*(7), 267–278. https://doi.org/10.1016/s1364-6613(00)01501-1

Amunts, K., Schleicher, A., Burgel, U., Mohlberg, H., Uylings, H.B., & Zilles, K. (1999) Broca's region revisited: cytoarchitecture and inter-subject variability. *Journal of Comparative Neurology, 412*(2), 319-341. https://doi.org/10.1002/(SICI)1096-9861(19990920)412:2<319::AID-CNE10>3.0.CO;2-7

Andersson, J. L. R., Jenkinson, M., & Smith, S. (2010). Non-linear registration, aka spatial normalization (FMRIB technical report TR07JA2). June.

Anzellotti, S. (2017). Anterior temporal lobe and the representation of knowledge about people. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(16), 4042–4044. https://doi.org/10.1073/pnas.1703438114

Anzellotti, S., & Caramazza, A. (2017). Multimodal representations of person identity individuated with fMRI. *Cortex*, *8*, 85–97. https://doi.org/10.1016/j.cortex.2017.01.013

Aristotle. (1995). Categories (J. L. Ackrill, Trans.). In J. Barnes (Ed.), The Complete Works of Aristotle (pp. 3–24). Princeton: Princeton University Press.

Astafiev, S. V., Stanley, C. M., Shulman, G. L., & Corbetta, M. (2004). Extrastriate body area in human occipital cortex responds to the performance of motor actions. *Nature Neuroscience*, *7*(5), 542–548. https://doi.org/10.1038/nn1241

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

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

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2003). fMRI Responses to Video and Point-Light Displays of Moving Humans and Manipulable Objects. *Journal of Cognitive Neuroscience*, *15*(7), 991–1001. https://doi.org/10.1162/089892903770007380

Bedny, M., Caramazza, A., Grossman, E., Pascual-Leone, A., & Saxe, R. (2008). Concepts are more than percepts: The case of action verbs. *Journal of Neuroscience*, *28*(44), 11347–11353. https://doi.org/10.1523/JNEUROSCI.3039-08.2008

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-Selective Areas in Human Auditory Cortex. *Foundations in Social Neuroscience*, *403*(6767), 309–312. https://doi.org/10.1038/35002078

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. http://www.jstor.org/stable/2346101

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

Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cognitive Neuropsychology*, *33*(3–4), 130–174. https://doi.org/10.1080/02643294.2016.1147426

Bonda, E., Petrides, M., Ostry, D., & Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *Journal of Neuroscience*, *16*(11), 3737–3744. https://doi.org/10.1523/jneurosci.16-11-03737.1996

Bracci, S., Cavina-Pratesi, C., Ietswaart, M., Caramazza, A., & Peelen, M. V. (2012). Closely overlapping responses to tools and hands in left lateral occipitotemporal cortex. *Journal of Neurophysiology*, *107*(5), 1443–1446. https://doi.org/10.1152/jn.00619.2011

Bracci, S., Ietswaart, M., Peelen, M. V., & Cavina-Pratesi, C. (2010). Dissociable neural responses to hands and non-hand body parts in human left extrastriate visual cortex. *Journal of Neurophysiology*, *103*(6), 3389–3397. https://doi.org/10.1152/jn.00215.2010

Bracci, S., & Op de Beeck, H. (2016). Dissociations and associations between shape and category representations in the two visual pathways. *Journal of Neuroscience*, *36*(2), 432–444. https://doi.org/10.1523/JNEUROSCI.2314-15.2016

Bracci, S., Ritchie, J. B., & de Beeck, H. O. (2017). On the partnership between neural representations of object categories and visual features in the ventral visual pathway. *Neuropsychologia*, *105*, 153–164. https://doi.org/10.1016/j.neuropsychologia.2017.06.010

Bracci, S., Ritchie, J. B., Kalfas, I., & Op de Beeck, H. P. (2019). The ventral visual pathway represents animal appearance over animacy, unlike human behavior and deep neural

networks. *Journal of Neuroscience, 39*(33), 6513–6525. https://doi.org/10.1523/JNEUROSCI.1714-18.2019

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

Buxbaum, L. J. (2001). Ideomotor Apraxia: a Call to Action. *Neurocase*, *7*(6), 445–458. https://doi.org/10.1093/neucas/7.6.445

Buxbaum, L.J. and Kalénine, S. (2010), Action knowledge, visuomotor activation, and embodiment in the two action systems. *Annals of the New York Academy of Sciences, 1191*, 201-218. https://doi.org/10.1111/j.1749-6632.2010.05447.x

Buxbaum, L. J., Kyle, K. M., & Menon, R. (2005). On beyond mirror neurons: Internal representations subserving imitation and recognition of skilled object-related actions in humans. *Cognitive Brain Research*, *25*(1), 226–239. https://doi.org/10.1016/j.cogbrainres.2005.05.014

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, *6*(8), 641–651. https://doi.org/10.1038/nrn1724

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., Woodruff, P. W. R., Iversen, S. D., & David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science*, *276*(5312), 593–596. https://doi.org/10.1126/science.276.5312.593

Caramazza, A., Anzellotti, S., Strnad, L., & Lingnau, A. (2014). Embodied Cognition and Mirror Neurons: A Critical Assessment. *Annual Review of Neuroscience*, *37*, 1–15. https://doi.org/10.1146/annurev-neuro-071013-013950

Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision, 13*. https://doi.org/10.1167/13.10.1

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. https://doi.org/10.1016/j.neuroimage.2009.12.112

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. https://doi.org/10.1093/cercor/bhp291

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: A review of its functional anatomy and behavioural correlates. *Brain, 129*(3), 564–583. https://doi.org/10.1093/brain/awl004

Chang, C. C., & Lin, C. J. (2011). LIBSVM: A Library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2*(3). https://doi.org/10.1145/1961189.1961199

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

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

Cichy, R. M., & Oliva, A. (2020). A M/EEG-fMRI Fusion Primer: Resolving Human Brain Responses in Space and Time. *Neuron*, *107*(5), 772–781. https://doi.org/10.1016/j.neuron.2020.07.001

Cichy, R. M., Roig, G., Andonian, A., Dwivedi, K., Lahner, B., Lascelles, A., Mohsenzadeh, Y., Ramakrishnan, K., & Oliva, A. (2019). *The Algonauts Project: A Platform for Communication between the Sciences of Biological and Artificial Intelligence*. https://doi.org/10.32470/ccn.2019.1018-0

Cios, K. J. (2018). Deep neural networks—A brief history. In *Studies in Computational Intelligence* (pp. 183–200). https://doi.org/10.1007/978-3-319-67946-4_7

Connolly, A. C., Swaroop Guntupalli, J., 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. https://doi.org/10.1523/JNEUROSCI.5547-11.2012

Cook, R., & Bird, G. (2013). Do mirror neurons really mirror and do they really code for action goals? *Cortex*, *49*(10), 2944–2945. https://doi.org/10.1016/j.cortex.2013.05.006

Corbo, D., & Orban, G. A. (2017). Observing Others Speak or Sing Activates Spt and Neighboring Parietal Cortex. *Jounral of Cognitive Neuroscience*, *29*(6), 1002–1021. https://doi.org/10.1162/jocn

Cornier, M. A., Salzberg, A. K., Endly, D. C., Bessesen, D. H., Rojas, D. C., & Tregellas, J. R. (2009). The effects of overfeeding on the neuronal response to visual food cues in thin and reduced-obese individuals. *PLoS ONE*, *4*(7), 1–7. https://doi.org/10.1371/journal.pone.0006310

Cree, G. S., McNorgan, C., & Mcrae, K. (2006). Distinctive Features Hold a Privileged Status in the Computation of Word Meaning: Implications for Theories of Semantic Memory. *J Exp Psychol Learn Mem Cogn*, *32*(4), 643–658. https://doi.org/10.1037/0278-7393.32.4.643.Distinctive

Cree, G. S., & McRae, K. (2003). Analyzing the Factors Underlying the Structure and Computation of the Meaning of Chipmunk, Cherry, Chisel, Cheese, and Cello (and many Other Such Concrete Nouns). *Journal of Experimental Psychology: General*, *132*(2), 163–201. https://doi.org/10.1037/0096-3445.132.2.163

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, *10*(4), 1–17. https://doi.org/10.1167/10.4.16

Davis, T., & Poldrack, R. A. (2013). Measuring neural representations with fMRI: practices and pitfalls. *Annals of the New York Academy of Sciences*, *1296*, 108-134. https://doi.org/10.1111/nyas.12156

Deen, B., Koldewyn, K., Kanwisher, N., & Saxe, R. (2015). Functional organization of social perception and cognition in the superior temporal sulcus. *Cerebral Cortex*, *25*(11), 4596–4609. https://doi.org/10.1093/cercor/bhv111

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, *91*, 176–180. https://doi.org/10.1007/BF00230027

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

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

Downing, P. E., Chan, A. W. Y., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2005). Domain specificity in visual cortex. *Cerebral Cortex*, *16*(10), 1453–1461. https://doi.org/10.1093/cercor/bhj086

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

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

Ekman, M., Derrfuss, J., Tittgemeyer, M., & Fiebach, C. J. (2012). Predicting errors from reconfiguration patterns in human brain networks. *Proceedings of the National Academy of Sciences of the United States of America, 109*(41), 16714–16719. https://doi.org/10.1073/pnas.1207523109

Ekstrom, A. D., Copara, M. S., Isham, E. A., Wang, W.-c., & Yonelinas, A. P. (2011). Dissociable networks involved in spatial and temporal order source retrieval. *NeuroImage, 56*(3), 1803–1813. https://doi.org/10.1016/j.neuroimage.2011.02.033

Epstein, R., DeYoe, E. A., Press, D. Z., Rosen, A. C., & Kanwisher, N. (2001). Neuropsychological evidence for a topographical learning mechanism in parahippocampal cortex. *Cognitive Neuropsychology*, *18*(6), 481–508. https://doi.org/10.1080/02643290042000215

Epstein, R., Graham, K. S., & Downing, P. E. (2003). Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron*, *37*(5), 865–876. https://doi.org/10.1016/S0896-6273(03)00117-X

Epstein, R., & Kanwisher, N. (1998). A cortical representation the local visual environment. *Nature, 9*(392), 598-601. https://doi.org/10.1038/33402

Felleman, D. J., Xiao, Y., & McClendon, E. (1997). Modular organization of occipito-temporal pathways: Cortical connections between visual area 4 and visual area 2 and posterior inferotemporal ventral area in macaque monkeys. *Journal of Neuroscience*, *17*(9), 3185–3200. https://doi.org/10.1523/jneurosci.17-09-03185.1997

Ferri, S., Rizzolatti, G., & Orban, G. A. (2015). The organization of the posterior parietal cortex devoted to upper limb actions: An fMRI study. *Human Brain Mapping, 36*(10), 3845–3866. https://doi.org/10.1002/hbm.22882

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolotti, G. (2005). Parietal lobe: From action organization to intention understanding. *Science*, *308*(5722), 662–667. https://doi.org/10.1126/science.1106138

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

Frith, C. D. (2007). The social brain? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 671–678. https://doi.org/10.1098/rstb.2006.2003

Gainotti, G., Ciaraffa, F., Silveri, M. C., & Marra, C. (2009). Mental Representation of Normal Subjects About the Sources of Knowledge in Different Semantic Categories and Unique Entities. *Neuropsychology*, *23*(6), 803–812. https://doi.org/10.1037/a0016352

Gainotti, G., Spinelli, P., Scaricamazza, E., & Marra, C. (2013). The evaluation of sources of knowledge underlying different conceptual categories. *Frontiers in Human Neuroscience, 7*. https://doi.org/10.3389/fnhum.2013.00040

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain : A Journal of Neurology*, *119*(2), 593–609. https://doi.org/10.1093/brain/119.2.593

Garrard, P., Lambon Ralph, M. A., Hodges, J. R., & Patterson, K. (2001). Prototypicality , distinctiveness, and intercorrelation: Analyses of the semantic attributes of living and nonliving concepts. *Cognitive Neuropsychology, 18*(2), 125–174. https://doi.org/10.1080/02643290125857

Geirhos, R., Michaelis, C., Wichmann, F. A., Rubisch, P., Bethge, M., & Brendel, W. (2019). Imagenet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. 7th International Conference on Learning Representations, ICLR 2019.

Gentilucci, M., Negrotti, A., & Gangitano, M. (1997). Planning an action. *Experimental Brain Research, 115*, 116-128. https://doi.org/10.1007/PL00005671

Geyer, S. (2004). The microstructural border between the motor and the cognitive domain in the human cerebral cortex. *Advances in Anatomy, Embryology and Cell Biology, 174*:I-VIII, 1-89. https://doi.org/10.1007/978-3-642-18910-4

Goldenberg, G., & Spatt, J. (2009). The neural basis of tool use. *Brain*, *132*(6), 1645–1655. https://doi.org/10.1093/brain/awp080

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25. https://doi.org/10.1016/0166-2236(92)90344-8

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

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, *41*(10–11), 1409–1422. https://doi.org/10.1016/S0042-6989(01)00073-6

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, *24*(1), 187–203. https://doi.org/10.1016/S0896-6273(00)80832-6

Grosbras, M. H., Beaton, S., & Eickhoff, S. B. (2012). Brain regions involved in human movement perception: A quantitative voxel-based meta-analysis. *Human Brain Mapping*, *33*(2), 431–454. https://doi.org/10.1002/hbm.21222

Grossman, E. D., Battelli, L., & Pascual-Leone, A. (2005). Repetitive TMS over posterior STS disrupts perception of biological motion. *Vision Research*, *45*(22), 2847–2853. https://doi.org/10.1016/j.visres.2005.05.027

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

Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, *3*, 1157–1182. https://doi.org/10.1162/153244303322753616

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning*, *46*, 389–422. https://doi.org/10.1002/bit.24634

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

Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General, 142*(3), 880-905. https://doi.org/10.1037/a0030045

Hafri, A., Trueswell, J. C., & Epstein, R. A. (2017). Neural Representations of Observed Actions Generalize across Static and Dynamic Visual Input. *The Journal of Neuroscience*, *37*(11), 3056–3071. https://doi.org/10.1523/JNEUROSCI.2496-16.2017

Halsband, U., Schmitt, J., Weyers, M., Binkofski, F., Grützner, G., & Freund, H. J. (2001). Recognition and imitation of pantomimed motor acts after unilateral parietal and premotor lesions: A perspective on apraxia. *Neuropsychologia*, *39*(2), 200–216. https://doi.org/10.1016/S0028-3932(00)00088-9

Hamilton, A. F. d. C., & Grafton, S. T. (2006). Goal representation in human anterior intraparietal sulcus. *Journal of Neuroscience, 26*(4), 1133-7. https://doi.org/10.1523/JNEUROSCI.4551-05.2006

Hamilton, A. F. d. C., & Grafton, S. T. (2007). The motor hierarchy: from kinematics to goals and intentions. In *Sensorimotor Foundations of Higher Cognition*, (pp. 381–407). https://doi.org/10.1093/acprof:oso/9780199231447.003.0018

Hamilton, A. F. d. C., & Grafton, S. T. (2008). Action outcomes are represented in human inferior frontoparietal cortex. *Cerebral Cortex*, *18*(5), 1160–1168. https://doi.org/10.1093/cercor/bhm150

Handjaras, G., Bernardi, G., Benuzzi, F., Nichelli, P. F., Pietrini, P., & Ricciardi, E. (2015). A topographical organization for action representation in the human brain. *Human Brain Mapping, 36*(10), 3832–3844. https://doi.org/10.1002/hbm.22881

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 and Biobehavioral Reviews*, *94*(August), 31–44. https://doi.org/10.1016/j.neubiorev.2018.08.003

Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-Inspired Artificial Intelligence. *Neuron*, *95*(2), 245–258. https://doi.org/10.1016/j.neuron.2017.06.011

Haxby, J. V., Gobbini, I. M., Furey, M. L., Ishai, A., Schouten, J., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*, *293*(5539), 2425–2430. https://doi.org/10.1126/science.1063736

Haxby, J. V., Gobbini, M. I., & Nastase, S. A. (2020). Naturalistic stimuli reveal a dominant role for agentic action in visual representation. *NeuroImage*, *216*(January), 116561. https://doi.org/10.1016/j.neuroimage.2020.116561

Haxby, J. V, Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Science*, *4*(6), 223–233. https://doi.org/10.1016/s1364-6613(00)01482-0

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

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

Hebart, M. N., Bankson, B. B., Harel, A., Baker, C. I., & Cichy, R. M. (2018). The representational dynamics of task and object processing in humans. *ELife*. https://doi.org/10.7554/eLife.32816

Hebart, M. N., Zheng, C. Y., Pereira, F., & Baker, C. I. (2020). Revealing the multidimensional mental representations of natural objects underlying human similarity judgements. *Nature Human Behaviour*, *4*, 1173–1185. https://doi.org/10.1038/s41562-020-00951-3

Heinzle, J., Wenzel, M. A., & Haynes, J.-D. (2012). Visuomotor Functional Network Topology Predicts Upcoming Tasks. *Journal of Neuroscience, 32*(29), 9960–9968. https://doi.org/10.1523/JNEUROSCI.1604-12.2012

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience*, *21*(7), 1229–1243. https://doi.org/10.1162/jocn.2009.21189

Hoffman, P., Evans, G. A. L., & Lambon Ralph, M. A. (2014). The anterior temporal lobes are critically involved in acquiring new conceptual knowledge: Evidence for impaired feature integration in semantic dementia. *Cortex, 50*(1), 19–31. https://doi.org/10.1016/j.cortex.2013.10.006

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience, 3*(1), 80–84. https://doi.org/10.1038/71152

Hoffman, P., & Lambon Ralph, M. A. (2013). Shapes, scents and sounds: Quantifying the full multi-sensory basis of conceptual knowledge. *Neuropsychologia, 51*(1), 14–25. https://doi.org/10.1016/j.neuropsychologia.2012.11.009

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, *148*, 574–591. https://doi.org/10.1113/jphysiol.1959.sp006308

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*(1), 106–154. https://doi.org/10.1113/jphysiol.1962.sp006837

Hubel, D., & Wiesel, T. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*, 215–243. https://doi.org/10.1113/jphysiol.1968.sp008455

Humphreys, G. W., & Rumiati, R. I. (1998). Agnosia without prosopagnosia or Alexia: Evidence for stored visual memories specific to objects. *Cognitive Neuropsychology, 243–277*. https://doi.org/10.1080/026432998381177

Huettel, S.A., Song, A.W., & McCarthy, G. Functional magnetic resonance imaging. Third edition. Sunderland: Sinauer Associates, Inc; 2014

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. https://doi.org/10.1016/j.neuron.2012.10.014.A

Ishai, A., Ungerleider, L. G., & Haxby, J. V. (2000). Distributed neural systems for the generation of visual images. *Neuron, 28*(3), 979–990. https://doi.org/10.1016/S0896-6273(00)00168-9

Isik, L., Koldewyn, K., Beeler, D., & Kanwisher, N. (2017). Perceiving social interactions in the posterior superior temporal sulcus. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(43), 9145–9152. https://doi.org/10.1073/pnas.1714471114

James, T. W., Culham, J., Humphrey, G. K., Milner, A. D., & Goodale, M. A. (2003). Ventral occipital lesions impair object recognition but not object-directed grasping: An fMRI study. *Brain*, *126*(11), 2463–2475. https://doi.org/10.1093/brain/awg248

Jastorff, J., Begliomini, C., Fabbri-Destro, M., Rizzolatti, G., & Orban, G. A. (2010). Coding observed motor acts: Different organizational principles in the parietal and premotor cortex of humans. *Journal of Neurophysiology*, *104*(1), 128–140. https://doi.org/10.1152/jn.00254.2010

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

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). FSL 1. *Neuroimage.*

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

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

Johnson-Frey, S. H., Newman-Norlund, R., & Grafton, S. T. (2005). A distributed left hemisphere network active during planning of everyday tool use skills. *Cerebral Cortex, 15*(6), 681–695. https://doi.org/10.1093/cercor/bhh169

Jozwik, K. M., Kriegeskorte, N., & Mur, M. (2016). Visual features as stepping stones toward semantics: Explaining object similarity in IT and perception with non-negative least squares. *Neuropsychologia*, *83*, 201–226. https://doi.org/10.1016/j.neuropsychologia.2015.10.023

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

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*(11), 4302–4311. https://doi.org/10.1523/jneurosci.17-11-04302.1997

Khaligh-Razavi, S. M., & Kriegeskorte, N. (2014). Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. *PLoS Computational Biology, 10*(11). https://doi.org/10.1371/journal.pcbi.1003915

Khosla, M., Ratan Murty, N. A., & Kanwisher, N. (2022). A highly selective response to food in human visual cortex revealed by hypothesis-free voxel decomposition. *Current Biology*, *32*(19), 4159–4171.e9. https://doi.org/10.1016/j.cub.2022.08.009

Kietzmann, T. C., Spoerer, C. J., Sörensen, L. K. A., Cichy, R. M., Hauk, O., & Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. *Proceedings of the National Academy of Sciences of the United States of America, 116*(430, 21854–21863. https://doi.org/10.1073/pnas.1905544116

Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in Cognitive Sciences, 15*(8), 352–357. https://doi.org/10.1016/j.tics.2011.06.005

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007a). Predictive coding: An account of the mirror neuron system. *Cognitive Processing*, *8*(3), 159–166. https://doi.org/10.1007/s10339-007-0170-2

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007b). The mirror-neuron system: a Bayesian perspective. *Neuroreport, 18*(6):619-23. https://doi.org/10.1097/WNR.0b013e3281139ed0

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research, 46*(11), 1762–1776. https://doi.org/10.1016/j.visres.2005.10.002

Klatzky, R. L., Lederman, S. J., & Matula, D. E. (1993). Haptic Exploration in the Presence of Vision Haptic Exploratory Procedures and Associated Properties. *Journal of Experimental Psychology: Human Perception and Performance, 19*(4), 726–743. https://doi.org/10.1037/0096-1523.19.4.726

Kourtzi, Z., & Kanwisher, N. (2000). Activation in human MT/MST by static images with implied motion. *Journal of Cognitive Neuroscience*, *12*(1), 48–55. https://doi.org/10.1162/08989290051137594

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

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

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

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

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

Leshinskaya, A., & Caramazza, A. (2014). Nonmotor aspects of action concepts. *Journal of Cognitive Neuroscience*, *26*(12), 2863–2879. https://doi.org/10.1162/jocn

Leshinskaya, A., Wurm, M. F., & Caramazza, A. (2020). Concepts of Actions and their Objects. In *The Cognitive Neurosciences* (pp. 757–765).

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461. https://doi.org/10.1037/h0020279

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

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

Lingnau, A., & Petris, S. (2013). Action understanding within and outside the motor system: The role of task difficulty. *Cerebral Cortex, 23*(6), 1342–1350. https://doi.org/10.1093/cercor/bhs112

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature, 453*(7197), 869–878. https://doi.org/10.1038/nature06976

Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the BOLD signal. *Annual Review of Physiology, 66*, 735–769. https://doi.org/10.1146/annurev.physiol.66.082602.092845

Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods, 41*(2), 558–564. https://doi.org/10.3758/BRM.41.2.558

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

Mahon, B. Z., & Caramazza, A. (2005). The orchestration of the sensory-motor systems: Clues from neuropsychology. *Cognitive Neuropsychology*, *22*(3–4), 480–494. https://doi.org/10.1080/02643290442000446

Mahon, B. Z., & Caramazza, A. (2009). Concepts and Categories: A Cognitive Neuropsychological Perspective. *Annual Review of Psychology, 60*, 27–51. https://doi.org/10.1146/annurev.psych.60.110707.163532

Majdandić, J., Bekkering, H., Van Schie, H. T., & Toni, I. (2009). Movement-specific repetition suppression in ventral and dorsal premotor cortex during action observation. *Cerebral Cortex, 19(*11), 2736–2745. https://doi.org/10.1093/cercor/bhp049

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

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics, 5*, 115–133. https://doi.org/10.1007/BF02478259

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, 37*, 547–559. https://doi.org/10.3758/BF03192726

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

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience*, *9*(5), 555–604. https://doi.org/10.1162/jocn.1997.9.5.555

Mountcastle, V. B., Motter, B. C., Steinmetz, M. A., & Sestokas, A. K. (1987). Common and differential effects of attentive fixation on the excitability of parietal and prestriate (V4) cortical visual neurons in the macaque monkey. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *7*(7), 2239–2255. https://doi.org/10.1523/jneurosci.07-07-02239.1987

Murphy, B., Poesio, M., Bovolo, F., Bruzzone, L., Dalponte, M., & Lakany, H. (2011). EEG decoding of semantic category reveals distributed representations for single concepts. *Brain and Language, 117*(1), 12–22. https://doi.org/10.1016/j.bandl.2010.09.013

Negri, G. A. L., Rumiati, R., Zadini, A., Ukmar, M., Mahon, B., & Caramazza, A. (2007). What is the role of motor simulation in action and object recognition? Evidence from

apraxia. *Cognitive Neuropsychology*, *24*(8), 795–816. https://doi.org/10.1080/02643290701707412

Nelissen, K., Borra, E., Gerbella, M., Rozzi, S., Luppino, G., Vanduffel, W., Rizzolatti, G., & Orban, G. A. (2011). Action observation circuits in the macaque monkey cortex. *Journal of Neuroscience*, *31*(10), 3743–3756. https://doi.org/10.1523/JNEUROSCI.4803-10.2011

Nelissen, K., Luppino, G., Vanduffel, W., Rizzolatti, G., & Orban, G. A. (2005). Neuroscience: Observing others: Multiple action representation in the frontal lobe. *Science, 310*(5746), 332–336. https://doi.org/10.1126/science.1115593

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A Toolbox for Representational Similarity Analysis. *PLoS Computational Biology*, *10*(4). https://doi.org/10.1371/journal.pcbi.1003553

Nishitani, N., & Hari, R. (2002). Viewing lip forms: Cortical dynamics. *Neuron*, *36*(6), 1211–1220. https://doi.org/10.1016/S0896-6273(02)01089-9

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-modal multivariate pattern analysis of neuroimaging data in matlab/GNU octave. *Frontiers in Neuroinformatics*, *10*. https://doi.org/10.3389/fninf.2016.00027

Oosterhof, N. N., Tipper, S. P., & Downing, P. E. (2012). Viewpoint (in)dependence of action representations: An MVPA study. *Journal of Cognitive Neuroscience*, *24*(4), 975–989. https://doi.org/10.1162/jocn_a_00195

Oosterhof, N. N., Tipper, S. P., & Downing, P. E. (2013). Crossmodal and action-specific: Neuroimaging the human mirror neuron system. *Trends in Cognitive Sciences*, *17*(7), 311–318. https://doi.org/10.1016/j.tics.2013.04.012

Oosterhof, N. N., Wiggett, A. J., Diedrichsen, J., Tipper, S. P., & Downing, P. E. (2010). Surface-Based Information Mapping Reveals Crossmodal Vision-Action Representations in Human Parietal and Occipitotemporal Cortex. *Journal of Neurophysiology, 104*(2), 1077–1089. https://doi.org/10.1152/jn.00326.2010

Oram, M. W., & Perrett, D. I. (1994). Responses of anterior superior temporal polysensory (STPa) neurons to "biological motion" stimuli. *Journal of Cognitive Neuroscience*, *6*(2), 99–116. https://doi.org/10.1162/jocn.1994.6.2.99

Orlov, T., Makin, T. R., & Zohary, E. (2010). Topographic Representation of the Human Body in the Occipitotemporal Cortex. *Neuron, 68*(3), 586–600. https://doi.org/10.1016/j.neuron.2010.09.032

Orlov, T., Porat, Y., Makin, T. R., & Zohary, E. (2014). Hands in motion: An upper-limb-selective area in the occipitotemporal cortex shows sensitivity to viewed hand kinematics. *Journal of Neuroscience*, *34*(14), 4882–4895. https://doi.org/10.1523/JNEUROSCI.3352-13.2014

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

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. https://doi.org/10.1523/JNEUROSCI.2668-18.2019

Papeo, L., & Lingnau, A. (2015). First-person and third-person verbs in visual motion-perception regions. *Brain and Language, 141*, 135–141. https://doi.org/10.1016/j.bandl.2014.11.011

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

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

Patterson, K., & Lambon Ralph, M. A. (2016). The Hub-and-Spoke hypothesis of semantic memory. In *Neurobiology of Language* (pp. 765–775). https://doi.org/10.1016/C2011-0-07351-9

Pazzaglia, M., Smania, N., Corato, E., & Aglioti, S. M. (2008). Neural underpinnings of gesture discrimination in patients with limb apraxia. *Journal of Neuroscience*, *28*(12), 3030–3041. https://doi.org/10.1523/JNEUROSCI.5748-07.2008

Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *Journal of Neurophysiology, 93*(1), 603–608. https://doi.org/10.1152/jn.00513.2004

Peelen, M. V., Romagno, D., & Caramazza, A. (2012). Independent representations of verbs and actions in left lateral temporal cortex. *Journal of Cognitive Neuroscience, 24*(10), 2096–2107. https://doi.org/10.1162/jocn_a_00257

Petrides, M., & Pandya, D. N. (1984). Projections to the frontal cortex from the posterior parietal region in the rhesus monkey. *The Journal of Comparative Neurology, 228*(1), 105–116. https://doi.org/10.1002/cne.902280110

Pitcher, D., & Ungerleider, L. G. (2021). *Evidence for a third visual pathway specialized for social perception*. *25*(2), 100–110. https://doi.org/10.1016/j.tics.2020.11.006.Evidence

Pitcher, D., Walsh, V., & Duchaine, B. (2011). The role of the occipital face area in the cortical face perception network. *Experimental Brain Research*, *209*(4), 481–493. https://doi.org/10.1007/s00221-011-2579-1

Popal, H., Wang, Y., & Olson, I. R. (2019). A Guide to Representational Similarity Analysis for Social Neuroscience. *Social Cognitive and Affective Neuroscience*, *14*(11), 1243–1253. https://doi.org/10.1093/scan/nsz099

Popov, V., Ostarek, M., & Tenison, C. (2018). Practices and pitfalls in inferring neural representations. *NeuroImage*, *174*, 340–351. https://doi.org/10.1016/j.neuroimage.2018.03.041

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

Proklova, D., Kaiser, D., & Peelen, M. V. (2016). Disentangling Representations of Object Shape and Object Category in Human Visual Cortex: The Animate–Inanimate Distinction. *Journal of Cognitive Neuroscience, 28*(5), 680–692. https://doi.org/10.1162/jocn

Pruim, R. H. R., Mennes, M., Buitelaar, J. K., & Beckmann, C. F. (2015). Evaluation of ICA-AROMA and alternative strategies for motion artifact removal in resting state fMRI. *NeuroImage*, *112*, 278–287. https://doi.org/10.1016/j.neuroimage.2015.02.063

Pruim, R. H. R., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J. K., & Beckmann, C. F. (2015). ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *NeuroImage*, *112,* 267–277. https://doi.org/10.1016/j.neuroimage.2015.02.064

Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *Journal of Neuroscience, 18*(6), 2188–2199. https://doi.org/10.1523/jneurosci.18-06-02188.1998

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, *435*(7045), 1102–1107. https://doi.org/10.1038/nature03687

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*(2), 655–706. https://doi.org/10.1152/physrev.00009.2013

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192. https://doi.org/10.1146/annurev.neuro.27.070203.144230

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*(2), 131–141. https://doi.org/10.1016/0926-6410(95)00038-0

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, *2*, 661–670. https://doi.org/10.1038/35090060

Rizzolatti, G., & Sinigaglia, C. (2016). The mirror mechanism: A basic principle of brain function. *Nature Reviews Neuroscience*, *17*, 757–765. https://doi.org/10.1038/nrn.2016.135

Roe, A. W., Chelazzi, L., Connor, C. E., Conway, B. R., Fujita, I., Gallant, J. L., Lu, H., & Vanduffel, W. (2012). Toward a Unified Theory of Visual Area V4. *Neuron*, *74*(1), 12–29. https://doi.org/10.1016/j.neuron.2012.03.011

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and Deterioration of Semantic Memory: A Neuropsychological and Computational Investigation. *Psychological Review, 111*(1), 205–235. https://doi.org/10.1037/0033-295X.111.1.205

Rosenbaum, D. A., Vaughan, J., Meulenbroek, R. J., & Jansen, C. (2001). Posture-based motion planning: Applications to grasping. *Psychological Review*, *108*(4), 709–734. https://doi.org/10.1037/0033-295X.108.4.709

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386-408. https://doi.org/10.1037/h0042519

Rushworth, M. F. S., Behrens, T. E. J., & Johansen-Berg, H. (2006). Connection patterns distinguish 3 regions of human parietal cortex. *Cerebral Cortex*, *16*(10), 1418–1430. https://doi.org/10.1093/cercor/bhj079

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

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*, 533–536. https://doi.org/10.1038/323533a0

Sato, W., Kochiyama, T., Yoshikawa, S., Naito, E., & Matsumura, M. (2004). Enhanced neural activity in response to dynamic facial expressions of emotion: An fMRI study. *Cognitive Brain Research*, *20*(1), 81–91. https://doi.org/10.1016/j.cogbrainres.2004.01.008

Sato, J. R., Mourão-Miranda, J., Morais Martin, M. da G., Amaro, E., Morettin, P. A., & Brammer, M. J. (2008). The impact of functional connectivity changes on support vector machines mapping of fMRI data. *Journal of Neuroscience Methods*, *172*(1), 94–104. https://doi.org/10.1016/j.jneumeth.2008.04.008

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

Sejnowski, T. J., Churchland, P. S., & Movshon, J. A. (2014). Putting big data to good use in neuroscience. *Nature Neuroscience*, *17*(11), 1440–1441. https://doi.org/10.1038/nn.3839

Seltzer, B., & Pandya, D. N. (1994). Parietal, temporal, and occipita projections to cortex of the superior temporal sulcus in the rhesus monkey: A retrograde tracer study. *Journal of Comparative Neurology, 343*(3), 445–463. https://doi.org/10.1002/cne.903430308

Senior, C., Barnes, J., Giampietroc, V., Simmons, A., Bullmore, E. T., Brammer, M., & David, A. S. (2000). The functional neuroanatomy of implicit-motion perception or "representational momentum." *Current Biology*, *10*(1), 16–22. https://doi.org/10.1016/S0960-9822(99)00259-6

Serpush, F., & Rezaei, M. (2021). Complex Human Action Recognition Using a Hierarchical Feature Reduction and Deep Learning-Based Method. *SN Computer Science, 2*(94). https://doi.org/10.1007/s42979-021-00484-0

Simanova, I., van Gerven, M., Oostenveld, R., & Hagoort, P. (2010). Identifying object categories from event-related EEG: Toward decoding of conceptual representations. *PLoS ONE*, *5*(12). https://doi.org/10.1371/journal.pone.0014465

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

Sokal, R., & Michener, C. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, *2*, 1409–1438. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:A+statistical+method+for+evaluating+systematic+relationships#0

Sokal, R. R., & Rohlf, F. J. (1962). THE COMPARISON OF DENDROGRAMS BY OBJECTIVE METHODS. *TAXON*, *11*(2), 33049. https://doi.org/10.2307/1217208

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

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

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

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

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

Thornton, M. A., & Tamir, D. (2023). The brain represents situations and mental states as sums of their action affordances. *PsyArXiv*, 1–47.

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

Tootell, R. B. H., Reppas, J. B., Kwong, K. K., Malach, R., Born, R. T., Brady, T. J., Rosen, B. R., & Belliveau, J. W. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *Journal of Neuroscience, 15*(4), 3215–3230. https://doi.org/10.1523/jneurosci.15-04-03215.1995

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

Tucciarelli, R., 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. https://doi.org/10.1523/JNEUROSCI.1422-15.2015

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

Turner, R. (2016). Uses, misuses, new uses and fundamental limitations of magnetic resonance imaging in cognitive science. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1705). https://doi.org/10.1098/rstb.2015.0349

Tyler, L. K., & Moss, H. E. (2001). Towards a distributed account of conceptual knowledge. *Trends in Cognitive Sciences*, *5*(6), 244–252. https://doi.org/10.1016/S1364-6613(00)01651-X

Urgen, B. A. (2020). Predictive processing account of action perception: Evidence from effective connectivity in the Action Observation Network. *Cortex*, *128*, 132–142. https://10.0.3.248/j.cortex.2020.03.014

Urgen, B. A., & Orban, G. A. (2021). The unique role of parietal cortex in action observation: Functional organization for communicative and manipulative actions. *NeuroImage*, *237*, 118220. https://doi.org/10.1016/j.neuroimage.2021.118220

Urgen, B. A., Pehlivan, S., & Saygin, A. P. (2019). Distinct representations in occipito-temporal, parietal, and premotor cortex during action perception revealed by fMRI and computational modeling. *Neuropsychologia, 127*, 35–47. https://doi.org/10.1016/j.neuropsychologia.2019.02.006

Vallacher, R. R., & Wegner, D. M. (1985). A Theory of Action Identification. In A Theory of Action Identification. https://doi.org/10.4324/9781315802213

van Kemenade, B. M., Muggleton, N., Walsh, V., & Saygin, A. P. (2012). Effects of TMS over premotor and superior temporal cortices on biological motion perception. *Journal of Cognitive Neuroscience, 24*(4), 896–904. https://doi.org/10.1162/jocn_a_00194

Vannuscorps, G., & Caramazza, A. (2016). Typical action perception and interpretation without motor simulation. *Proceedings of the National Academy of Sciences*, *113*(1), 86–91. https://doi.org/10.1073/pnas.1516978112

von Kriegstein, K., Eger, E., Kleinschmidt, A., & Giraud, A. L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Cognitive Brain Research, 17*(1), 48–55. https://doi.org/10.1016/S0926-6410(03)00079-X

Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods, 40*, 183–190. https://doi.org/10.3758/BRM.40.1.183

Vinson, D. P., Vigliocco, G., Cappa, S., & Siri, S. (2003). The breakdown of semantic knowledge: Insights from a statistical model of meaning representation. *Brain and Language*, *86*(3), 347–365. https://doi.org/10.1016/S0093-934X(03)00144-5

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

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

Wang, C., Xiong, S., Hu, X., Yao, L., & Zhang, J. (2012). Combining features from ERP components in single-trial EEG for discriminating four-category visual objects. *Journal of Neural Engineering, 9*(5). https://doi.org/10.1088/1741-2560/9/5/056013

Wardle, S. G., & Baker, C. (2020). Recent advances in understanding object recognition in the human brain: Deep neural networks, temporal dynamics, and context. *F1000Research*, *9*(590). https://doi.org/10.12688/f1000research.22296.1

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

Watson, C. E., Cardillo, E. R., Ianni, G. R., & Chatterjee, A. (2013). Action Concepts in the Brain: An Activation Likelihood Estimation Meta-analysis. *Journal of Cognitive Neuroscience*, *25*(8), 1191–1205.

Weaverdyck, M. E., Lieberman, M. D., & Parkinson, C. (2020). Tools of the trade multivoxel pattern analysis in fMRI: A practical introduction for social and affective neuroscientists. *Social Cognitive and Affective Neuroscience, 15*(4), 487–509. https://doi.org/10.1093/scan/nsaa057

Wegner, D. M., & Vallacher, R. R. (1986). Action Identification. In *Handbook of motivation and cognition: Foundations of social behavior* (pp. 550–582).

Wichmann, F. A., & Geirhos, R. (2023). Are Deep Neural Networks Adequate Behavioral Models of Human Visual Perception? *Annual Review of Vision Science, 9*(1), 501–524. https://doi.org/10.1146/annurev-vision-120522-031739

Wicker, B., Michel, F., Henaff, M. A., & Decety, J. (1997). Brain regions involved in the perception of gaze:A PETstudy. *Neuroimage, 8*(8), 221–227. https://ac.els-cdn.com/S1053811998903573/1-s2.0-S1053811998903573-main.pdf?_tid=a990630a-5b8d-4924-8f38-cb8233b129eb&acdnat=1523474089_7c8dbfe747d132f395b40f6f3dd04e1b

Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. *NeuroImage, 21*(4), 1732-1747. https://doi.org/10.1016/j.neuroimage.2003.12.023

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage, 14*(6), 1370-1386. https://doi.org/10.1006/nimg.2001.0931

Worsley, K.J., Statistical analysis of activation images. Ch 14, in Functional MRI: An Introduction to Methods, eds. P. Jezzard, P.M. Matthews and S.M. Smith. OUP, 2001

Wurm, M. F., Ariani, G., Greenlee, M. W., & Lingnau, A. (2015). Decoding Concrete and Abstract Action Representations During Explicit and Implicit Conceptual Processing. *Cerebral Cortex*, *26*, 3390–3401. https://doi.org/10.1093/cercor/bhv169

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

Wurm, M. F., & Caramazza, A. (2019b). Lateral occipitotemporal cortex encodes perceptual components of social actions rather than abstract representations of sociality. *NeuroImage*, *202*. https://doi.org/10.1016/j.neuroimage.2019.116153

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

Wurm, M. F., Caramazza, A., & Lingnau, A. (2017). Action Categories in Lateral Occipitotemporal Cortex Are Organized Along Sociality and Transitivity. *The Journal of Neuroscience*, *37*(3), 562–575. https://doi.org/10.1523/JNEUROSCI.1717-16.2017

Wurm, M. F., & Lingnau, A. (2015). Decoding Actions at Different Levels of Abstraction. *Journal of Neuroscience*, *35*(20), 7727–7735. https://doi.org/10.1523/JNEUROSCI.0188-15.2015

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

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

Xia, M., Wang, J., & He, Y. (2013). BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics. *PLoS ONE*, *8*(7). https://doi.org/10.1371/journal.pone.0068910

Yang, Y. J. D., Allen, T., Abdullahi, S. M., Pelphrey, K. A., Volkmar, F. R., & Chapman, S. B. (2017). Brain responses to biological motion predict treatment outcome in young adults with autism receiving Virtual Reality Social Cognition Training: Preliminary findings. *Behaviour Research and Therapy*, *93*, 55–66. https://doi.org/10.1016/j.brat.2017.03.014

Yang, Y., Dickey, M. W., Fiez, J., Murphy, B., Mitchell, T., Collinger, J., Tyler-Kabara, E., Boninger, M., & Wang, W. (2017). Sensorimotor experience and verb-category mapping in human sensory, motor and parietal neurons. *Cortex*, *92*, 304–319. https://doi.org/10.1016/j.cortex.2017.04.021

Yao, B., Jiang, X., Khosla, A., Lin, A. L., Guibas, L., & Fei-Fei, L. (2011). Human action recognition by learning bases of action attributes and parts. *Proceedings of the IEEE International Conference on Computer Vision*. https://doi.org/10.1109/ICCV.2011.6126386

Yargholi, E., Hossein-Zadeh, G. A., & Vaziri-Pashkam, M. (2023). Two distinct networks containing position-tolerant representations of actions in the human brain. *Cerebral Cortex*, *33*(4), 1462–1475. https://doi.org/10.1093/cercor/bhac149

Zeki, S. M. (1973). Colour coding in rhesus monkey prestriate cortex. *Brain Research*, *53*(2), 422–427. https://doi.org/10.1016/0006-8993(73)90227-8

Zeki, S., Watson, J. D. G., Lueck, C. J., Friston, K. J., Kennard, C., & Frackowiak, R. S. J. (1991). A direct demonstration of functional specialization in human visual cortex.

Journal of Neuroscience, 11(3), 641–649. https://doi.org/10.1523/jneurosci.11-03-00641.1991

Zheng, C. Y., Baker, C. I., Pereira, F., & Hebart, M. N. (2019). Revealing interpretable object representations from human behavior. *7th International Conference on Learning Representations, ICLR 2019*

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