The Hierarchical Representation of Observed Actions: Evidence From Behavioral and Neuroimaging Studies

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

vorgelegt von
Tonghe Zhuang
aus Liaoning, China

Regensburg 2023

Universität Regensburg

Gutachterin (Betreuerin): Prof. Dr. Angelika Lingnau

Gutachter: Prof. Dr. Gregor Volberg

Preface

The goal of this thesis, titled ‘The Hierarchical Representation of Observed Actions: Evidence From Behavioral and Neuroimaging Studies,’ is to investigate the behavioral and neural traits of actions at the subordinate, basic, and superordinate levels. The thesis comprises five chapters: Chapter 1 provides the general introduction, Chapters 2–4 encompass the three main studies, and Chapter 5 contains the general discussion.

An article based on the first study was published in Psychological Research (impact factor > 2.4) on 14th December 2021 and a second article, based on the second study, has been accepted for publication in the Journal of Neuroscience (impact factor >5.3) on 6th September 2023. Chapter 2 includes the printed versions of the first study and Chapter 3 contains the accepted version of the second study. Access to the data and codes of the second study is available online at https://osf.io/b6ea4/. Chapter 4 is formatted for the eLife journal and incorporates the third study, which is currently being prepared for submission. The figure, table, and equation numbers have been arranged in ascending order. Supplementary files and figures have been merged and organized in the Appendix section following Chapter 5. Additionally, a unified bibliography has been created, including all references at the end of the thesis. No additional revisions have been made to the text of the published articles.

The PhD project received support and funding from the Chinese Scholarship Council, and a stipend from the ‘*Finanzielles Anreizsystem zur Förderung der Gleichstellung.*’

Acknowledgments

I would like to express my deep gratitude to all those who have contributed to my journey towards completing this Ph.D. dissertation. This effort is the result of many years of hard work, determination, and the help of many people.

First, I would like to express my sincere appreciation to my **supervisor**, Prof. Dr. Angelika Lingnau. You gave me a great example of what it means to be a real mentor. I can vividly recall the day I first arrived in Regensburg, and the warm welcome you extended to me, hosting a welcome party that left a lasting impression. Your support and encouragement on that very first day set the tone for the incredible journey that followed. Choosing you as my supervisor was undoubtedly the best decision I have made in my academic career, and I consider myself incredibly fortunate to have had you by my side throughout this Ph.D. program. In fact, you healed me like a mental therapist. You have consistently supported and believed in me, and as a result, I have gained the confidence to trust myself. I always do not want to express my appreciation too much because I am afraid to make you feel uncomfortable. It is also difficult to convey the depth of my appreciation for your contributions sufficiently. Now, it is the last moment. Thank you. I really appreciate you. Many thanks for these years: the first time watching a show in the theater, the first-time city run, the first party for me and my first-conference talk. All in all, your mentorship has not only guided me but also rekindle my passion for science. Furthermore, your presence has been an inspiring model of leadership, demonstrating both your strength and kindness. Thanks again for your enduring guidance and inspiration.

Second, I would like to express my heartfelt gratitude to the amazing **lab members** who have not only been my colleagues but my friends during my time in Regensburg. The friendship and warmth that you all have extended to me have made my days in Regensburg truly special. I am deeply thankful for the valuable discussions, technical assistance, and the unwavering encouragement I have received from each one of you. You have all played unique and crucial roles in my research, and I am genuinely grateful for your friendship and contributions. I would like to offer special thanks to a few individuals who have left a particularly significant mark on my journey. Thank you, Gregor for your support on EEG data processing. I still remember the moments after EEG course when you patiently answered many of my questions. Andre, I am grateful for the times you revised my writing and the memories of our early years in the Ph.D. program. Oleg, your technical support is importance and I cherish our friendship. You are always very kind. Marius, I want to express my gratitude for your willingness to patiently answer my quick, door-knocking questions. With our interactions, I've learned a lot about working smartly. Max, I appreciate you revised my manuscript and the comfortable space you provided for cooking and gaming, which created countless enjoyable moments. Federica, thanks for your on-time feedback on my manuscript. It has been a true pleasure having you as part of our lab. Bobi, thanks to your mental support for my academic and personal life. I want to express my sincere gratitude for the delightful cocktail party we had – it was an occasion I thoroughly enjoyed and will always cherish as a wonderful memory. Additionally, my deep gratitude to Zuza, who has not only been a close colleague but a dear friend to me. Sharing the same office with you has brought me immense happiness throughout my journey. You have been a strong and reliable support, and I am grateful for all the times we have inspired and enjoyed each other's company. I will deeply miss the pictures that capture our time together in the office – they hold special places in my heart.

Last, to my **family**, who have been a constant source of love and support, I extend my deepest gratitude. Your belief in my abilities and encouragement made it possible for me to pursue my academic aspirations. Additionally, I want to express my heartfelt apologies for not being physically present with you all these years. I am sorry. Furthermore, your unconditional support and encouragement have never wavered, regardless of the miles that separated us. Mum -- Ms.

Wang (王苏), you are my greatest role model. Your determination and the motivating words you have shared have been the wind beneath my wings, propelling me further in my academic and personal endeavors. Dady -- Mr. Zhuang (庄来洪), you have been my unwavering source of support, a comforting reminder that I always have a place to go back, our home. To my lovely sister- Tinghe (庄婷贺), your companionship and support, especially when I was lonely. I hold deep affection for you and appreciate the love and care you have shown, not only to me but to our parents. I am deeply thankful for each of you, and I look forward to the day when I can repay your love and support in person.

Thanks, Tonghe. You did a great job.

In summary, this Ph.D. journey has been a remarkable chapter in my life, and I am deeply appreciative of the support, guidance, and friendships that have made it possible. Thank you to all who have been a part of this journey.

Best wishes,

Tonghe Zhuang 庄童贺

Table of Contents

Abstract

Objects can be classified at different hierarchical levels, ranging from the subordinate level (e.g., ‘golden delicious’), to the basic level (e.g., ‘apple’) and the superordinate level (e.g., ‘fruit’). Previous studies have shown that the basic level is crucial for object categorization, such as the number of features used for description and processing speed. Similar to objects, actions can be described at different hierarchical levels, ranging from very specific (subordinate level: e.g., ‘to swim breaststroke’) to very broad (superordinate level: e.g., ‘locomotion’) information, with a mid-level of generality (basic level: e.g., ‘to swim’). However, it is not entirely clear whether the same principles that apply to objects also hold for the categorization of actions. Thus, the goals of my PhD projects are 1) to investigate the characteristics of actions at different hierarchical levels; 2) to examine the neural representation of observed actions at different taxonomic levels; 3) to explore at what point the brain represents actions at three different levels.

Three studies have been performed to address these goals. In the first study, I first had a selection of verbs from a linguistics book (Levin 1996), which had been divided into two classes. Next, based on what I selected, I asked participants to rate the semantic relationship and obtained several clusters. Then I asked the participants to name the clusters with the superordinate names. I also generated the subordinate names for each action in experiment 2. Afterwards, I asked the participants to rate how well the subordinate names fit to the superordinate names in experiment 3. Then, using these selected stimuli, I performed a feature listing task in which participants generated as many features as possible for each action at each level in a limited time. Following the feature-listing experiment, I conducted a prime effect design in experiment 5 and a rapid category verification task in experiment 6 on the actions across different taxonomic levels. The results showed that in feature-listing experiment, the actions the basic level were provided the most common features, while the actions at the subordinate level were generated more shared features in comparison with the superordinate level. The actions at the superordinate level had the most distinctive features compared to the other two levels. In experiment 5 and 6, I found that actions at the subordinate and basic levels were recognized more quickly than actions at the superordinate level. These behavioral results suggest that actions at the basic level have the maximum information, indicating that the basic level plays an important role in action categorization.

In the second study, to further explore the relationship between behavioral and neural representations of actions at different taxonomic levels, I conducted an event-related fMRI study and established models for these three levels (subordinate, basic, and superordinate). Participants were instructed to judge the semantic similarity of actions for creating a behavioral space model. Using multiple regression representational similarity analysis (RSA) in regions of interest, I found that bilateral lateral occipitotemporal cortex (LOTC) best represented the three taxonomic levels, particularly aligning with the basic level model. A whole-brain multiple regression RSA revealed that unique information at the basic level was encoded in LOTC and parietal regions. In contrast, the unique information for the subordinate level was represented in the bilateral occipitotemporal cortex, with no specific cluster capturing unique data for the superordinate level. Additionally, the behavioral action space was best represented in the LOTC and superior parietal cortex, showing the highest similarity to the basic-level model. In conclusion, the results suggest that the occipitotemporal cortex has a preference for the basic-level information and can flexibly access to both the subordinate and basic levels.

In the third study, to examine the timing of representations for actions at the subordinate, basic, and superordinate levels, I conducted an EEG experiment. I found that peak latencies in the EEG response occurred around 170 ms for actions at all three taxonomic levels. Furthermore, the EEG-fMRI fusion results indicated that action representations in LOTC were the most similar to EEG data at approximately 220 ms, which is earlier than V1. In summary, our findings suggest that actions across three taxonomic levels occur simultaneously, with the LOTC showing a later response compared to V1. This enhances our understanding of the spatiotemporal aspects of the hierarchical organization of action representations.

Contributions

Study 1	**The characterization of actions at the superordinate, basic and subordinate level**
Authors	Tonghe Zhuang, Angelika Lingnau
Author Contributions	T. Z. and A. L. designed research; T. Z. performed research and analyzed data; T. Z. and A. L. wrote the paper.
Study supervision	A. L.
Acknowledgement	Angelika Lingnau was funded by a Heisenberg-Professorship (German Research Foundation, LI 2840/2-1). Tonghe Zhuang was funded by a PhD stipend from the Chinese Scholarship Council. We thank Nina Wunder, Lucca Scheuermeyer, Lisa Kaiser, and Veronika Marina Holler for the coding of the results in Experiments 4. Moreover, we are grateful to Andre Bockes, Caterina Pedersini, Lucca Scheuermeyer, Micah Adams, Isabell Griebl, Zuza Kabulska and Yih-Shiuan Lin for feedback on the manuscript.
Commercial relationships/conflict of interests	None
Declaration of Helsinki	Yes
Clinical trial	No
Journal (Impact Factor)	Psychological Research (2022-2023 Impact Factor: 2.424)
Publication status	Submitted: 2 November 2020 / Accepted: 26 November 2021 / Published: 14 December 2021

Study 2	**The representation of observed actions at the subordinate, basic and superordinate level**
Authors	Tonghe Zhuang, Zuzanna Kabulska, and Angelika Lingnau
Author Contributions	T.Z. and A.L. designed research; T.Z. and Z.K. performed research; T.Z. analyzed data; T.Z. and A.L. wrote the paper; Z.K. edited the paper.
Study supervision	A. L.
Acknowledgement	Angelika Lingnau was funded by a Heisenberg-Professorship (German Research Foundation, LI 2840/2-1; LI 2840/4-1). Tonghe Zhuang was funded by a PhD stipend from the Chinese Scholarship Council and a

	stipend from the *'Finanzielles Anreizsystem zur Förderung der Gleichstellung.'*
Commercial relationships/conflict of interests	None
Declaration of Helsinki	Yes
Clinical trial	No
Journal (Impact Factor)	Journal of Neuroscience (2022-2023 Impact Factor: 5.3)
Publication status	Submitted: 8 April 2022 / Accepted: 6 September 2023

Study 3	**Neural dynamics of hierarchical action representations**
Authors	Tonghe Zhuang, Gregor Volberg, Angelika Lingnau
Author Contributions	T. Z. and A. L. designed research; T. Z. performed research and analyzed data; T. Z. and A. L. wrote the paper; G. V. suggested data analysis and revised the paper.
Study supervision	A. L.
Acknowledgement	Angelika Lingnau was funded by a Heisenberg-Professorship (German Research Foundation, LI 2840/2-1; LI 2840/4-1). Tonghe Zhuang was funded by a PhD stipend from the Chinese Scholarship Council and a stipend from the *'Finanzielles Anreizsystem zur Förderung der Gleichstellung.'*
Commercial relationships/conflict of interests	None
Declaration of Helsinki	Yes
Clinical trial	No
Publication status	The manuscript is currently being prepared for submission to eLife.

Author Information

Chair of Cognitive Neuroscience, Institute of Psychology, Faculty of Human Sciences, University of Regensburg

Corresponding author: Prof. Dr. Angelika Lingnau, angelika.lingnau@ur.de, Chair of Cognitive Neuroscience, Institute of Psychology, Faculty of Human Sciences, University of Regensburg
Universitätsstrasse 31, 93053 Regensburg, Germany

Abbreviations

fMRI	Functional Magnetic Resonance Imaging
EEG	Electroencephalogram
STS	Superior temporal sulcus
PPC	Posterior parietal cortex
aIPL	Anterior inferior parietal lobe
EBA	Extrastriate body area
RSA	Representational similarity analysis
LOTC	Lateral occipitotemporal cortex
LOC	Lateral occipital cortex
DCM	Dynamic causal modeling
SD	Semantic dementia
IT	Inferior temporal cortex
VTC	Ventral temporal cortex
MEG	Magnetoencephalography
DNN	Deep neural network
LOA	Level of abstraction
MTG	Middle temporal gyrus
IFG	Inferior frontal gyrus
MVPA	Multi-voxel pattern analysis
PMv	Ventral premotor cortex
SOA	Stimulus onset asynchrony
RT	Response time
ASF	A Simple Framework
IPL	Inferior parietal lobule
MDS	Multidimensional scaling
MB	Multiband
EPI	Echo-planar imaging
TR	Repetition time
TE	Time echo
FoV	Field of view
MPRAGE	Magnetization Prepared Rapid Gradient Echo
DOF	Degrees of freedom
FWHM	Full width half maximum
MNI	Montreal Neurological Institute
ROI	Region of interest
SPL	Superior parietal lobule
RFX	Random effects
GLM	General linear model
TFCE	Threshold-Free Cluster Enhancement
RDM	Representational dissimilarity matrix
LH	Left hemisphere
RH	Right hemisphere
LDA	Linear discriminant analysis

OTC	Occipitotemporal cortex
REFS	RSA-based EEG-fMRI fusion
AIP	Anterior intraparietal area
pAIP	Posterior part of AIP
ms	Millisecond
EVC	Early visual cortex

Overview

Understanding how humans comprehend actions is a fundamental topic in cognitive neuroscience. Consider, for instance, a child engaged in motion by moving his arms and legs in the water. In this situation, we might spontaneously come up with multiple concepts in our minds to name this activity, such as 'to swim', 'to swim butterfly', or 'locomotion'. This raises a number of interesting questions: do humans exhibit differential cognitive responses when identifying actions through distinct concepts? Which type of concept tends to dominate our conceptualization? Are there significant differences in the patterns of neural activity in our brains during these naming processes? Previous studies divided objects into three taxonomic levels (the subordinate, basic, and superordinate levels) and have well-investigated different representations of objects across the three taxonomic levels (Rosch et al., 1976; Iordan et al., 2015). However, these questions have been unsolved and require further investigation for actions. In this thesis, I investigated the different representations of actions at different taxonomic levels. To address these questions, I used a series of behavioral experiments, a high-spatial resolution functional Magnetic Resonance Imaging (fMRI) study, and a high-temporal resolution Electroencephalogram (EEG) experiment.

In Chapter 1, I will provide a review of the theoretical foundations of action understanding and illustrate the central research questions addressed in the context of my PhD projects. In Chapter 2, I will introduce the findings related to the characterizations of observed actions at different levels, investigated from a behavioral perspective. Subsequently, in Chapter 3, I will present the investigation of the brain regions responsible for the representation of observed actions at three taxonomic levels. In particular, I focus on whether these actions are encoded in common or distinct neural regions. In Chapter 4, I will introduce the results from temporal representations of action representations across these taxonomic levels using EEG, with the primary aim of identifying the specific time points in which actions are represented at different levels. Finally, Chapter 5 will include a comprehensive discussion of our research findings, the limitations of our work, and an outline of potential directions for future research in this area.

1 Chapter 1: General introduction

In the following chapter, I will primarily focus on the critical brain regions that play a vital role in action understanding. Specifically, I will introduce the three distinct visual pathways crucial for comprehending actions: the ventral ‘what’ pathway, the dorsal ‘where’ pathway, and the third pathway dedicated to social information (Milner & Goodale, 2008; Pitcher & Ungerleider, 2021). I will focus on their individual contributions, interactions and roles in the perception and understanding of actions. Additionally, this chapter will offer a comprehensive examination of the hierarchical organization of objects. Determining the hierarchical structure of objects can provide profound insights and methodologies for understanding the cognitive processes of actions. Furthermore, I will conduct a review of previous studies on action hierarchies, offering context for the existing knowledge in this area. Finally, I will present my research questions and hypotheses.

1.1 Action Understanding

Action understanding is essential for our daily life, such as social interaction and communication. It refers to the cognitive process by which humans comprehend and explain behavior, which helps people predict others' actions, read their intentions and react appropriately in society. Over the past few decades, researchers have investigated the brain mechanisms that underpin action understanding. In this review, I will discuss two conventional pathways and introduce a recent proposed third pathway, leading to a deeper comprehension of actions. In particular, the dorsal pathway is mainly associated with the processing of spatial, motor, and action guidance information, addressing the ‘how’ question, which refers to the manner in which an action is executed (Milner & Goodale, 2008; Rizzolatti & Matelli, 2003). The ventral pathway is well-known for its role in object recognition, enabling us to address the ‘what’ question (Milner & Goodale, 2008), and researchers found the ventral pathway is able to represent action-related components such as faces, objects, and body parts, along with higher-level semantic knowledge related to actions (e.g., Bracci et al., 2019; Tucciarelli et al., 2019; Wurm & Caramazza, 2022). Additionally, Pitcher and Ungerleider (2021) proposed the third visual pathway, dedicated to the dynamic aspects of social perception, which runs from the early visual cortex over motion-selective areas, into the superior temporal sulcus (STS). Subsequently, I introduce each of these three pathways in detail.

1.1.1 Action Understanding in the Dorsal Pathway

The dorsal pathway, also called the ‘where/how’ pathway, extends from V1 forward to the parietal lobe (Milner & Goodale, 2008). The dorsal pathway is mainly involved in the processing of the spatial information of objects and action guidance (Binkofski & Buxbaum, 2013; Milner & Goodale, 2008). For actions, the dorsal pathway is crucial in the manipulating or grasping the object (Balint, 1909; Culham et al., 2006; Rizzolatti & Matelli, 2003). For example, when we observe an action, the dorsal pathway is involved in understanding the speed and direction of hand movement, which helps us predict where the hands will end up. Taking patients study as an example, Balint (1909) found that patients with dorsal pathway lesions, particularly in the posterior parietal cortex (PPC), perform poorly in grasping tasks.

Additionally, the dorsal pathway plays an important role in action guidance and motor control during action observation (Gallivan & Goodale, 2018; Hamilton & Grafton, 2006; Nelissen et al., 2011; Rizzolatti & Sinigaglia, 2010). For example, Hamilton and Grafton (2006) instructed participants to view a series of movies showing people doing actions with specific goals. Some goals were shown multiple times, while others were novel. When the same goal was repeated, the left intraparietal sulcus showed repeated suppression effects. This suggests that the anterior intraparietal sulcus is involved in representing the goal of an observed action.

Furthermore, dorsal pathway, especially the parietal cortex, is sensitive to action outcomes (Hamilton & Grafton, 2008; Krasovsky et al., 2014; Leshinskaya & Caramazza, 2015). For example, Leshinskaya and Caramazza (2015) used fMRI experiment to unravel the neural representations of abstract function concepts of object-related actions. They instructed participants to evaluate objects for different goals, such as decoration and protection. They found that anterior inferior parietal lobe (aIPL) plays a broader role in understanding how objects can be used, encompassing both physical and abstract functional criteria, indicating IPL represents concepts of actions, regardless of motor and visual properties.

To sum up, the dorsal pathway plays a vital role in processing spatial and action-related information, encompassing functions like guidance, comprehension, observation, and execution (Gallivan & Goodale, 2018; Hamilton & Grafton, 2006; Milner & Goodale, 2008; Nelissen et al., 2011; Rizzolatti & Sinigaglia, 2010).

1.1.2 Action Understanding in the Ventral Pathway

The ventral pathway, also called the ‘what’ stream, originates from the primary visual cortex (V1) and continues downward to the temporal lobe (Milner & Goodale, 2008). Although the ventral pathway is renowned for its role in object identification, it also plays a role in action recognition by processing action-related features (e.g., body-parts and tools) and action semantics (Bracci et al., 2019; Milner & Goodale, 2008; Tarhan & Konkle, 2020; Tucciarelli et al., 2019; Wurm & Caramazza, 2019).

Consider, for example, the scenario of recognizing an action, such as ‘playing’, in a scene where a woman is playing frisbee with a dog. In such a instance, we would typically require not only an understanding of the action itself but also of the other entities involved, including the ‘woman,’ the ‘frisbee,’ and the ‘dog.’ It is important to note that while not all actions involve objects, the interaction between objects and actions is a significant area of investigation for researchers (Bracci et al., 2012; Bracci & Peelen, 2013; Leshinskaya et al., 2020). Therefore, acquiring knowledge about objects is also crucial for understanding actions. In this context, the ventral pathway, being a significant component of object recognition, actively contributes to our comprehension of actions. In particular, the object-selective region has been found in the lateral occipital cortex (Haxby et al., 2001).

Apart from the object, some key components of actions, such as the body parts involved, are processed in the ventral pathway. The ventral part of extrastriate body area (EBA) is known to encode body parts (Downing et al., 2001). Furthermore, the ventral pathway plays a crucial role in representing the interaction between hands and objects (Baldassano et al., 2017; Bracci et al., 2012; Bracci & Peelen, 2013). For example, using fMRI experiments, Bracci and Peelen (2013) measured the action-related object representations in high-level visual cortex. They found that a hand-selective region in left LOTC selectively responded to object effectors, independently of general action-related properties, suggesting that the left LOTC is involved in body-object interactions. Considering these action-related components found within the ventral pathways, it becomes apparent that the ventral pathway contributes to action understanding, particularly some action components.

Furthermore, it is possible to elicit higher-level semantic action knowledge in the ventral pathway (Tarhan & Konkle, 2020; Tucciarelli et al., 2019; Wurm et al., 2017; Wurm & Caramazza, 2019). As an example, Tucciarelli et al. (2019) conducted an fMRI experiment on different types of actions to measure the representation of different key components with representational similarity analysis (RSA, Kriegeskorte, Mur, & Bandettini, 2008), in which representational dissimilarity matrix (RDM) is used to measure the similar representation between neural response patterns and human behavior. They found that action semantics are best represented in the lateral occipitotemporal cortex (LOTC). Moreover, Hafri et al., (2017) revealed that posterior LOTC represents action information across stimulus format (static and dynamic) with images and videos of actions. Similar, with different stimulus types (video and words), Wurm and Caramazza (2019) found that the left lateral posterior temporal cortex (LPTC) (also called anterior LOTC) is able to encode conceptual action information.

Additionally, Nastase et al., (2017) revealed that the LOTC is more sensitive to animal actions (action categories: eating, fighting, running, and swimming) rather than animal categories, whereas the ventral temporal cortex is more sensitive for animal categories (primates, ungulates, birds, reptiles, and insects). The LOTC is one key part of the ventral pathway, which reflect that the ventral pathway is associated with action understanding.

Together, these previous studies suggest that the 'what' pathway is involved not only in object recognition, but also in action understanding, ranging from feature-based information to general, conceptual action representation.

1.1.3 Action Understanding in the Third Pathway

In a recent review, Pitcher and Ungerleider (2021) provided an expansion of the traditional view in which humans have two visual systems -- ventral ('what') and dorsal ('where') pathways. They proposed a third pathway specialized for processing dynamic information of social perceptions, especially of faces. This pathway runs from the early visual cortex, along motion-preference regions (V5/ middle temporal cortex) to STS, where it encodes features from low-level visual features to high-level social information. The authors emphasized that this third pathway is particularly sensitive to the dynamic information of social perception, specifically in relation to moving faces and bodies. Overall, the proposed third visual pathway expands the action understanding into three distinct pathways, providing a new perspective on social perception. Additionally, McMahon et al., (2023) used an fMRI experiment to provide evidence supporting the existence of the third visual pathway for social information. Using naturalistic stimuli in the form of daily action videos, the study revealed that the early visual cortex represents low-level visual features. The EBA and lateral occipital cortex (LOC) were found to capture mid-level visual-social features, while the STS exhibited sensitivity to high-level social interaction information. These findings suggest that the primary function of third pathway is the recognition of social information.

1.1.4 Interaction Across Three Pathways

These three streams are able to interact with each other rather than being strictly independent (Milner, 2017; Pitcher & Ungerleider, 2021; van Polanen & Davare, 2015). Taking the famous patient D.F. with visual form agnosia as an example (James et al., 2003), the damage to the ventral pathway not only impair the processing of visual information but also damaged the processing of the dorsal stream (e.g., complex shapes task), which suggest that two streams are not independent but can also interact.

Additionally, some neuroimaging studies showed the interaction between ventral and dorsal pathways (Chen et al., 2018; Milner, 2017; van Polanen & Davare, 2015; Wurm & Caramazza, 2022). For instance, using two different shapes of tools (elongated vs. stubby) and two different objects (tool vs. non-tool), Chen et al., (2018) measured the communication between ventral and dorsal streams in tool-selective regions using dynamic causal modeling (DCM), a tool used to infer about the directed connectivity within brain regions. Their results found that during tool observation, the connectivity from the ventral to the dorsal pathway was increased, and that only elongated tools would increase interconnectivity between regions. These studies, from patient studies to neuroimaging experiments, suggest the necessity of interaction between ventral and dorsal pathways in some situations. Furthermore, the third pathway is involved in processing dynamic information from social perception, an important component of action understanding, indicating that these three pathways work together to comprehend dynamic actions.

1.2 Hierarchical organization

1.2.1 Categorization

To investigate the functions of regions involved in action understanding, researchers have explored various methodologies, one of which involves leveraging the concept of categorization. Given the immense volume of information in the world, the brain tends to recognize what it sees in an economical way - through categorization (Rosch, 1978). Previous studies have revealed a preference for object categories (e.g., faces, houses, tools and animals) in ventral pathway (e.g., Downing et al., 2001; Kanwisher, 2000). Additionally, as reviewed above in the last section, both ventral and dorsal pathway are involved in the processing of action categories, such as grasping, reaching (Hamilton & Grafton, 2006; Rizzolatti & Sinigaglia, 2010; Wurm et al., 2017). These studies of both objects and actions have consistently highlighted categorization as a common and essential cognitive mechanism of the human brain, in line with Rosch's findings in 1978.

Indeed, when humans recognize the objects and actions, brains encounter an overwhelming amount of sensory input, making it impractical to analyze each detail individually. By categorizing objects, events, and actions, the brain can streamline its information processing. This allows us to rapidly comprehend and react to our surroundings which refers to cognitive Economy (Rosch, 1978). Categorization significantly reduces cognitive load (Downey et al., 2008; Rosch et al., 1976). Instead of memorizing specific details about each individual item or event, we can store and retrieve information based on the category to which it belongs. The purpose of categorization involves not only considering stimuli within the same category as equivalent to each other, but also recognizing them as distinct from stimuli that do not belong to that category (Downey et al., 2008; Rosch, 1978; Rosch et al., 1976).

According to different classification fashions of structure, it is possible to separate items (such as objects or actions) into different categories and different taxonomic levels (also called hierarchies). In particular, a category is a group of items that share similar features and are able to discriminate from other groups with dissimilar features, such as ‘dogs’ vs. ‘cats’. Apart from categories, one object or action can be further assigned into different levels, providing a hierarchically structured framework to organize our perception and experiences. That is, the hierarchical organization refers to a system where each lower-level item is subordinate to a higher-level item, creating a top-to-bottom organizational structure (Rosch et al., 1976). Specifically, using a seminal series of experiments, Rosch et al., (1976) divided objects into three levels based on their abstraction. In particular, the most abstract level is the superordinate level, such as ‘the instrument’; the intermediate level, also called the basic level, such as ‘the guitar’; and the concrete level is the subordinate level, such as ‘electric guitar’. From Rosch and subsequent studies on object levels (I will review more details in next section object levels; Rosch et al., 1976, Carlson et al., 2013; Iordan et al., 2015), we have gained insights into the distinctive attributes of objects across various taxonomic levels. In particular, the basic level includes the maximum information and is represented in LOC (more details will be offered in the next section: object levels). However, it is unclear whether these principles governing object categorization extend to the hierarchical organization of actions. Hence, during my PhD period, I investigated the behavioral and neural representations of actions across the subordinate, basic and superordinate level.

In the following, I will review previous research on hierarchical organization, beginning with object levels and then action hierarchies. While research on object hierarchies has been well-documented across behavioral and neuroimaging studies (e.g., Rosch et al., 1976; Iordan et al., 2015; Carlson et al., 2013), comparatively, we know less about action hierarchies. Furthermore, the methods used to study object levels can provide valuable insights for investigating action hierarchies. Hence, before diving into action hierarchies, I will provide a comprehensive overview of the understanding of object levels. In the section action hierarchies, I will review the current state and point out the gaps in the literatures and my thesis aims are to fill these gaps.

1.2.2 Object levels

1.2.2.1 Behavioral research

Previous patient studies provided direct evidence that object recognition is impaired differently at different taxonomic levels (Humphreys & Forde, 2005; Rogers & Patterson, 2007). As an example, Humphreys and Forde (2005) reported a single-case study of a patient with difficulties naming objects only at the superordinate level as a consequence of damage to large portions of the occipital lobe due to carbon monoxide poisoning. In addition, Rogers and Patterson (2007) found that healthy participants responded faster and more accurately than patients with semantic dementia (SD) when asked to verify categories of objects at the basic level, whereas SD patients performed better at the superordinate level. These results suggest that the processing of objects at the superordinate level can be spared despite impairments in processing at the basic level (and vice versa).

Additionally, the evidences from behavioral experiments showed different characterizations of objects levels (Mack et al., 2008; Mack & Palmeri, 2011; Rosch et al., 1976). In particular, Rosch et al. (1976) argued that object categories can be organized into superordinate (e.g., vehicle, animal), basic (e.g., chair, dog) and subordinate level (e.g., kitchen chair, bulldog), depending on the degree of abstraction. Utilizing a feature-listing paradigm, the authors found that objects at the basic level were described with the maximum number of features. Using priming and object detection paradigms, they found that objects at the basic level were processed faster than actions at the superordinate level. Together, Rosch proposed that the basic level of objects plays a central role in categorization.

Furthermore, to investigate the dynamic processes of visual categorization, the duration of stimulus exposure was manipulated during tasks involving the identification of categories at different hierarchical levels (Grill-Spector & Kanwisher, 2005; Mack et al., 2008; Mack & Palmeri, 2015). Specifically, researchers (Grill-Spector & Kanwisher, 2005; Mack et al., 2008, Exp.1) instructed participants to complete a categorization verification task between objects within a category (e.g., discriminating German Shepherd or other dogs, referring to the subordinate level) and between categories (e.g., chairs, cars, referring to the basic level), and object detection task with identifying objects or not (referring to the superordinate level). The results showed that participants performed better and faster during between-category discrimination and object detection task, in comparison with the within-category verification task. Moreover, Mack et al. (2008) did an additional experiment to examine the visual processing of object categorization and found that participants were more accurate and responded faster during the categorization of objects (depicted as inverted stimuli) at the basic level compared to the superordinate level.

Additionally, Mack and Palmeri (2011) reviewed lots of papers related to the timing of object categorization. They concluded that objects at the basic level are detected before the subordinate and superordinate levels and these findings can change based on personal experience. Furthermore, with presenting the target name before or after the line-draw stimulus, Taniguchi et al., (2020) investigated the effect of the category level on the categorization process. They found that the categorization process depends on the category level and target-word position. Shape properties affect the categorization process differently for basic-level and superordinate-level categorizations. Higher levels of visual information (e.g., shape complexity) impact superordinate-level decisions.

In summary, the studies mentioned above suggest that that there are differences in the speed of object recognition and categorization processes at different taxonomic levels.

1.2.2.2 Neuroimaging research of object levels in space

Neuroimaging research has revealed the different representations of objects across levels in space (Carlson et al., 2013; Iordan et al., 2015; Jozwik et al., 2023). In particular, previous studies involving macaques have demonstrated that the high-level visual cortex represents various hierarchical category information (Dehaqani et al., 2016; Kriegeskorte et al., 2008; Yamins et al., 2014). Kriegeskorte et al. (2008) found that the inferior temporal cortex (IT) captures the hierarchical representations of objects (i.e., the high level -- animate and inanimate objects, the middle level -- categories and the low level -- within-category exemplars) for both monkeys and humans. Additionally, Iordan et al. (2015) investigated the vertical organization of objects in the ventral visual cortex using RSA (Kriegeskorte et al., 2008) and multivariate pattern analysis (MVPA, Haxby et al., 2001). They found that although all three levels were decoded in the whole ventral visual cortex, the subordinate level of objects was best reflected by patterns of activity in early visual cortex, whereas the basic level of objects showed the strongest similarity with patterns of activity in the LOC. These findings are consistent with the proposal by Grill-Spector and Weiner (2014) regarding the neural architecture underlying the organization of objects. That is, the ventral temporal cortex (VTC) has flexible access to the representations of objects at different hierarchical levels. Furthermore, Margalit et al., (2020) used ultra-high-resolution fMRI to measure the category representations in VTC partitions at submillimeter scale and different cortical depths. They found that lateral VTC primarily organized representations at the domain level (e.g., written characters, faces and places), while medial VTC also organized at the specific category level, and these structures decreased in strength with increasing cortical depth. In conclusion, different taxonomic levels of objects can be dissociated at the neural level.

1.2.2.3 Temporal representations of objects across different levels

Correspondingly, some researchers investigated the neural representation of object-level information over time (Carlson et al., 2013; Cichy et al., 2014; Grootswagers et al., 2022). For instance, Carlson et al. (2013) investigated the time course of object representations during the first 1000 ms of visual processing using magnetoencephalography (MEG). Their results showed that objects at the subordinate and basic level were decoded during the first 100 ms, while objects at the superordinate level were decoded beyond 100 ms after stimulus onset. Moreover, latencies for peaks of decoding accuracies across levels varied based on the level of abstraction. This investigation has shed light on the flexibility and efficiency of brain in organizing information.

Additionally, in another study conducted by Cichy and his colleagues (2014), they used a combination of spatially high-resolution fMRI and temporally high-resolution MEG to colleagues the spatial and temporal representations associated with object categorization. The study found an integrated space- and time-resolved pattern of object categorization within the first few hundred milliseconds. This study provided the following studies and our study with a novel method to explore neural dynamics. More recently, Jozwik et al., (2023) also used the same approach as Cichy did, the spatial-temporal combination method, to explore neural dynamics of object cognition in the human ventral-stream. They used both Deep neural network (DNN) and visuo-semantic models (i.e., category: subordinate-, basic and superordinate levels and features: object parts, color, shape, and texture). They found that DNN features can better explain the low-level visual areas at 66 ms after the stimulus onset, while visuo-semanic features were the best effective in caputuring high-level cortical dynamics at 146 ms after stimulus onset. This study implicates the hierarchy theory of visual perception.

1.2.2.4 Short summary on object levels

To summarize, previous studies have revealed distinct representations of object levels in both behavioral and neuroimaging research (Rosch et al., 1976; Iordan et al., 2015; Carlson et al., 2013). Some behavioral studies have demonstrated that the basic level is crucial in categorization (Hajibayova, 2013; Mack & Palmeri, 2011; Rosch et al., 1976). Additionally, it has been found that object representations at all three taxonomic levels can be decoded in the ventral visual cortex, while LOC captures basic information best (Iordan et al., 2015). Additionally, objects at three taxonomic levels are able to distinguish representations over time.

In particular, the latencies for peak decodability showed a systematic variation in relation to the degree of category abstraction, with more concrete categories showing an early emergence (Carlson et al., 2013). The principles of object levels provide a relatively clear insight into object categorization. That is, object categorization involves a hierarchical organization, with different levels of abstraction (i.e. subordinate, basic and superordinate levels) influencing how we perceive and categorize objects. This hierarchical structure helps us understand how we mentally group and distinguish objects based on their shared characteristics, making object categorization a more efficient cognitive process. However, it is unclear whether these principles of object levels also fit to action understanding. In the following sections, I will introduce some background on what has been researched for action hierarchy and what still remains unclear.

1.2.3 Action hierarchical organization

1.2.3.1 Prior action hierarchies

Regarding action hierarchies, several theories have been proposed from different perspectives. According to the Theory of Action Identification (Vallacher & Wegner, 1985; Wegner & Vallacher, 1986), action execution and understanding can be described at different levels. Lower levels are associated with concrete action-implementation, while higher levels focus on more abstract aspects such as reasons and implications of actions.

Another well-known theory is proposed by Hamilton and Grafton (2006, 2008). They suggested three hierarchical levels of the motor system: the muscle level (activity patterns of the muscles), the kinematic level (the shape and movement of hands and arms), and the goal level (the purpose and outcome of the action). When observing someone else’s actions, humans may process and understand them at these different levels, ranging from the specific muscle movements, the shape and movement of hands and arms, to the overall goals and intentions behind those actions.

Furthermore, Spunt et al., (2010, 2016) examined two key aspects of an action: how it is done and why it is done. More specifically, ‘how’ question refers to its implementation (e.g., tying words, low abstraction) and ‘why’ question indicates to the goal (e.g., sharing knowledge, high abstraction). Spunt et al. (2016) found that temporoparietal junction was one of the regions significantly associated with higher level of abstraction (LOA) of action representation. Lower LOAs were associated with the rostral inferior parietal lobule (IPL), posterior middle temporal gyrus (MTG) and inferior frontal gyrus (IFG).

Additionally, Wurm and Lingnau (2015) divided actions into three taxonomic levels from concrete, intermediate, to abstract level. They found that actions at the concrete level (e.g., to open vs. to close a specific object) are represented in premotor cortex. Actions at the intermediate level (e.g., to open different bottles with different tools) are associated with action understanding. Actions at the abstract level (e.g., to open different objects) decoded in inferior parietal and occipitotemporal cortex. Likewise, similar results were also shown on action execution in the study of Turella and colleagues (2020). They examined hierarchical encoding of action execution, specifically from concrete actions (within effector and orientation), to middle-level actions (effector-dependent goal), to more-abstract actions (effector-independent goal). Using MVPA, they found that the parietal cortex was able to represent executed actions at all three hierarchies, while the motor cortex encoded executed actions at the concrete and middle level. Additionally, Turella et al., (2020) found the effector-independent goal representation decoded in the LOTC, which indicated that the LOTC hosted a crucial role in abstract action representations.

These studies align with the perspective that action comprehension follows a hierarchical structure, and these distinctions can be observed at both behavioral and neural levels.

1.2.3.2 Temporal dynamics of action representations

Although there is less emphasis on the dynamics of action representations across taxonomic levels, prior studies have investigated the temporal representations of actions (Dima et al., 2022; Tucciarelli et al., 2015). For instance, researchers used event-related potentials (ERP) to measure action understanding (Amoruso et al., 2013; Proverbio & Riva, 2009) and found that N400 (a nagative wave around 400 ms) is related to action understanding. In particular, Proverbio and Riva (2009) used pseudo actions as stimuli and the ERP results showed that N400 amplitudes were associated with the pseudo-actions, indicating action comprehension elicited N400 effect.

Additionally, Tucciarelli et al. (2015) used MEG to examine abstract action representations during action observation and found that the LOTC gains earlier access to abstract action representations (around 200 ms) compared to precentral regions. Additionally, Kolasinski et al. (2020) identified distinct encoding of muscle and kinematic information, with kinematic information appearing first (around 220 ms) before muscle activity, indicating top-down control of action. In a study using EEG, Dima et al. (2022) explored the representations of social-affective aspects of observed actions through large-scale sets of naturalistic daily action videos. Their findings showed that visual features emerged first (around 100 ms), followed by the representation of action features (around 150-600 ms), while information about social-affective features emerged at a later stage (around 400 ms). Altogether, these studies reveal distinct temporal representations of actions across various dimensions.

1.2.4 Comparisons between prior hierarchical theories and object levels

Previous researchers have proposed several theories of action hierarchies that differ significantly from object levels proposed by Rosch. The key distinction is that action hierarchies are centered on understanding action goals and their implementation to achieve them. In contrast, Rosch's object levels emphasize cognitive efficiency and the hierarchical organization of categorization, including the subordinate, basic, and superordinate levels.

The basic level is the best level of categorization because objects at the basic level tend to possess attributes with the highest cue validity and are recognized more quickly than those at the superordinate and subordinate levels. It is worth investigating whether these principles of object levels also fit for the structure of observed actions. Therefore, my PhD projects focused on three main questions: 1) What are the specific characteristics associated with actions at each taxonomic level? 2) Which regions of the brain can decode actions at the three taxonomic level respectively? 3) At what point in time are these actions represented across various levels?

1.3 Goals, Hypothesis and Innovation of the study

1.3.1 Goals

The aim of my dissertation comprises three parts. First, I conducted a series of behavioral studies to examine the characterization of action hierarchies based on the degree of abstraction, and to investigate the speed of action recognition across different taxonomic levels.

Second, I used fMRI to examine where in the human brain actions are represented at three different levels. Additionally, I explored the convergence of these representations across all three levels. To address these questions, I conducted RSA in both regions of interest (ROIs) associated with action understanding and whole-brain searchlight analysis. Indeed, RSA is an advanced method in neuroscience. It involves measuring the similarity between patterns of brain activity in response to different stimuli or conditions and the hypothesized models. By examining these patterns and their similarities, researchers can gain insights into how the brain processes and represents information.

Finally, I used EEG to explore at what point the brain represents actions at three different levels. I also conducted RSA analysis on EEG data. Furthermore, to enhance our comprehension of how the brain represents actions, I integrated fMRI and EEG data, combining the insights from both methods.

1.3.2 Hypotheses

According to previous studies about action hierarchies, here I proposed several hypotheses for each research question in my projects.

1) Behavioral experiments: observed actions at the superordinate level are described with fewer features than the other two levels. Actions at the basic level contain the most features. Actions at the subordinate level have many features, but these features are shared with other actions at the same level. Additionally, participants perform faster and more accurately with actions at the basic level in comparison with those at the superordinate and subordinate levels. These assumptions are proposed based on studies of Rosch et al. (1976).

2) fMRI experiment: actions at the subordinate level are represented in early visual cortex, LOTC, IPL and vPM, whereas actions at the basic level are represented in high-level visual cortex, such as the LOTC. Additionally, actions at the superordinate recruit anterior LOTC and parietal cortex (Abdollahi et al., 2013; Wurm & Lingnau, 2015).

3) EEG experiment: actions at the three taxonomic levels are decoded in different time windows. In particular, actions at the subordinate and basic level are decoded at the early points (80 ms and 120 ms) in time, while actions at the superordinate level are decoded later at around 160 ms (see also Carlson et al., 2013).

1.3.3 Innovation

In the current study, I first established three taxonomic levels of actions with a series of behavioral experiments, in the same way that Rosch did for object categories (Figure 1.1), and then investigated the characterization of actions across three taxonomic levels, using both behavioral and neuroimaging experiments. The purpose of my work is to shed light on action hierarchical organization. These projects filled the gap in our understanding of how actions are organized and represented across three taxonomic levels in both the brain and behavior. This connection helps us better understand how the brain processes different hierarchies of actions, from concrete to abstract. It also forms a scientific foundation for our understanding of how actions are represented, connecting what we observe in behavior to what happens in the brain. Additionally, it provides valuable insights into how our brains work when we perceive and categorize actions. This insight shows that the brain can differentiate between actions at different levels of abstraction.

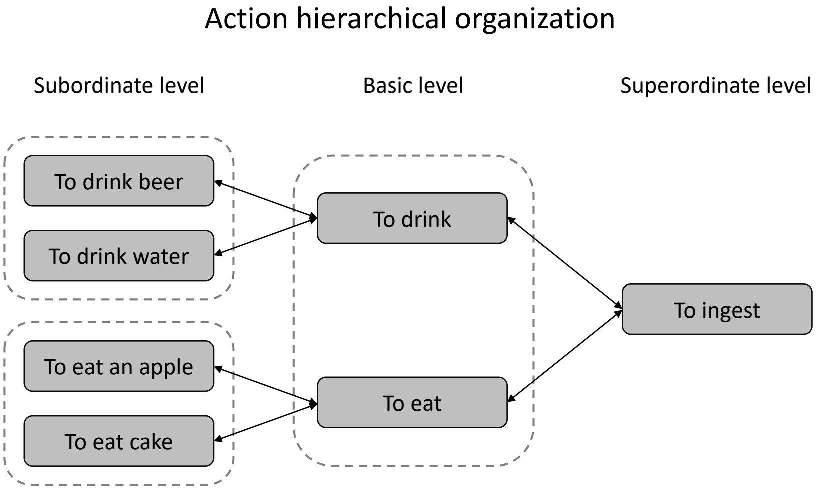


Figure 1.1. The hierarchical organization of action from the subordinate to the superordinate levels. Following the classification of objects (Rosch et al., 1976), actions are described at the subordinate (e.g. to drink beer, to eat an apple), the basic (e.g. to drink, to eat) and the superordinate level (e.g. to ingest).

2 Chapter 2: Study 1

‘The characterization of actions at the superordinate, basic and subordinate level’

This study was published on December 14, 2021, in Psychology Research and is available through an open-access link with the DOI: 10.1007/s00426-021-01624-0. Supplementary files can be found in Appendix.

2.1 Abstract

Objects can be categorized at different levels of abstraction, ranging from the superordinate (e.g., fruit) and the basic (e.g., apple) to the subordinate level (e.g., golden delicious). The basic level is assumed to play a key role in categorization, e.g., in terms of the number of features used to describe these objects and the speed of processing. To which degree do these principles also apply to the categorization of observed actions? To address this question, we first selected a range of actions at the superordinate (e.g., locomotion), basic (e.g., to swim) and subordinate level (e.g., to swim breaststroke), using verbal material (Experiments 1–3). Experiments 4–6 aimed to determine the characteristics of these actions across the three taxonomic levels. Using a feature listing paradigm (Experiment 4), we determined the number of features that were provided by at least six out of twenty participants (common features), separately for the three different levels. In addition, we examined the number of shared (i.e., provided for more than one category) and distinct (i.e., provided for one category only) features. Participants produced the highest number of common features for actions at the basic level. Actions at the subordinate level shared more features with other actions at the same level than those at the superordinate level. Actions at the superordinate and basic level were described with more distinct features compared to those provided at the subordinate level. Using an auditory priming paradigm (Experiment 5), we observed that participants responded faster to action images preceded by a matching auditory cue corresponding to the basic and subordinate level, but not for superordinate level cues, suggesting that the basic level is the most abstract level at which verbal cues facilitate the processing of an upcoming action. Using a category verification task (Experiment 6), we found that participants were faster and more accurate to verify action categories (depicted as images) at the basic and subordinate level in comparison to the superordinate level. Together, in line with the object categorization literature, our results suggest that information about action categories is maximized at the basic level.

2.2 Introduction

Categorization is a cognitively economical way for humans to get to know and refer to the world (Rosch et al., 1976). That is, we tend to divide objects into different groups (also called categories), such as animals, tools, fruits, and vehicles. Categorization distinguishes one category from others, and at the same time reduces differences among objects falling into the same category, thus supporting effective recognition. At the same time, objects can be classified into different levels (superordinate, basic, and subordinate) based on the degree of abstraction (Rosch et al., 1976). For example, depending on our knowledge and the situation, we might label the food in front of us as a ‘fruit’ (superordinate level), an ‘apple’ (basic level), or a ‘golden delicious’ (subordinate level). In other words, objects have been proposed to be organized according to categories (which we refer to as ‘horizontal organization’) as well as to a hierarchy (referred to as ‘vertical organization’ in the remainder of this article).

Rosch et al. (1976) suggested that the principles of categorization are cognitive economy and cue validity. Cognitive economy refers to the balance between the ability to categorize objects at a high level of precision (e.g., using a granny smith rather than a golden delicious for your favorite apple pie), while at the same time building in enough flexibility to limit the capability of this system to the distinctions required by a given task or circumstances (e.g., by knowing that an apple can have a range of sizes and different colors from red over yellow to green, while the shape is always round). Cue validity refers to the validity that a given cue (e.g., ‘round’, ‘green’, ‘slightly sour’) serves as a predictor of a certain category (e.g., ‘granny smith’). Cue validity increases the association between the cue and the category and decreases the association between the cue and other categories.

In a series of experiments, Rosch et al. (1976) investigated the informational capacity of objects at the superordinate, basic and subordinate level. Using a feature listing paradigm, participants were instructed to write down as many features as they could think of for objects at each level.

Features that were provided by at least six out of 20 participants were considered common features, i.e., features that reflect the cue validity of an object. The authors observed that participants provided more common features for objects at the basic level in comparison to the superordinate level. Participants provided the smallest number of common features for objects at the superordinate level. Assuming that lower taxonomic levels include all the features listed at the higher level, Rosch et al. (1976) determined the number of features added at the basic level (in comparison to the superordinate level) and the number of features added at the subordinate level (in comparison to the basic level). They observed that a smaller number of features was added at the subordinate level than at the basic level and concluded that the basic level is most inclusive in terms of the attributes objects at that level have in common. Moreover, Rosch (1978) argued that objects at the subordinate and superordinate level have a lower cue validity in comparison to the basic level, because (a) many common features were shared with other objects from other categories at the subordinate level, and (b) objects at the superordinate level were described with a large proportion of distinctive features.

In a separate experiment, Rosch et al. (1976) instructed participants to list motor (body and muscle) movements associated with specific objects. They found that participants listed fewer motor movements for objects at the superordinate level in comparison to objects at the basic and subordinate level. By contrast, participants listed a similar number of motor movements associated with objects at the basic and subordinate level. According to the authors, these results illustrated that the basic level was the most inclusive level at which many motor movements are associated with objects.

Using picture-word-matching and priming paradigms, Rosch et al. (1976) found that object categories were recognized faster at the basic and subordinate level in comparison to the superordinate level. Using an object recognition task, the authors reported that participants verified object images fastest for object names at the basic level. Likewise, Rosch et al. (1976)

reported that children learn to categorize objects at the basic level prior to objects at the superordinate or subordinate level. In sum, Rosch et al. (1976) concluded that objects at the basic level are associated with the highest cue validity.

More recent studies investigated the neural representation of objects at different hierarchical levels (Gauthier et al., 1997; Grill-Spector & Kanwisher, 2005; Grill-Spector et al., 2004; Iordan et al., 2015; Margalit et al., 2020). As an example, Gauthier et al. (1997) examined which brain areas are recruited during the processing of objects at the basic (e.g., bird) and the subordinate level (e.g., eagle) using a picture-word matching paradigm and a semantic judgement task. They observed a stronger recruitment of the fusiform gyrus, the inferior temporal gyrus and the occipital cortex during the judgement of objects at the subordinate level in comparison to judgements of objects at the basic level during both tasks. Likewise, using multi-voxel pattern analysis (MVPA), Iordan et al. (2015) reported representations of object categories at the subordinate in early visual cortex, whereas representations at the basic level were found both in early visual cortex and in object-selective regions. Representational similarity analysis (RSA, also a form of MVPA) revealed that the perirhinal cortex uniquely represents object categories at the superordinate level (animals, fruit, vegetables, tools, vehicles and musical instruments; Clarke & Tyler, 2014). Finally, different levels of the hierarchy have been argued to be represented at different spatial scales (e.g., Grill-Spector & Weiner, 2014; Margalit et al., 2020). Taken together, these studies demonstrate that the different hierarchical levels of objects proposed by Rosch et al. (1976) can also be distinguished at the neural level.

To which degree does the horizontal and vertical organization of objects described so far apply to the organization of actions? Note that this question is of theoretical interest both for the organization of actions we perform ourselves and actions we observe. Whereas there exist a number of interesting cross-links between these two lines of research that we will refer to, the current study focuses on the organization of observed actions. Similar to objects, observed actions have been suggested to be organized horizontally according to several superordinate categories, such as manipulation, locomotion and communication (Corbo & Orban, 2017; Tucciarelli et al., 2019; Wurm et al., 2017). Likewise, principles related (but not identical) to the vertical organization of objects can be found in the literature on the organization of actions. The importance of different levels of representations of actions has been explicitly spelled out in the Theory of Action Identification (Vallacher & Wegner, 1985; Wegner & Vallacher, 1986). This theory emphasizes the relationship between processes involved in comprehending and performing an action. Importantly, the theory proposes that actions can be identified at different levels, with lower levels related to the concrete implementation of an action, whereas higher levels provide a more abstract representation of an action reflecting the reasons and effects of the action. Likewise, Hamilton and Grafton (2006, 2008) and Grafton and Hamilton (2007) suggested three hierarchical levels of the motor system: the goal level (the purpose and outcome of action), the kinematic level (the shape and movement of hands and arms) and the muscle level (active patterns of the muscles). Using repetition suppression, they observed that the anterior intraparietal sulcus (aIPS) represents goals of observed actions (e.g., to grasp an object; Hamilton & Grafton, 2006). By contrast, kinematic aspects of the observed action have been reported to be represented in the lateral occipital cortex (LOC), superior parietal lobe, the fusiform cortex and the superior temporal sulcus (Hamilton & Grafton, 2006, 2008; for related studies, see also Grafton & Hamilton, 2007; Spunt et al., 2016). Using MVPA, Wurm and Lingnau (2015) distinguished between representations of observed actions (opening vs. closing an object) at a concrete level, referring to specific combinations of objects, kinematics and grip types, and an abstract level, which showed generalization across objects and kinematics. Their data revealed representations of actions at the concrete level in the lateral occipitotemporal cortex (LOTC), the inferior parietal lobe (IPL) and the ventral premotor cortex (PMv), whereas representations at the abstract level were restricted to the LOTC and the IPL. Finally, using

MVPA, several studies revealed spatially distinct representations of executed actions at the goal level and the kinematic level (Gallivan et al., 2013; Kadmon Harpaz et al., 2014; Turella et al., 2020). Together, these studies are in line with the view that the representation of observed and executed actions follow a hierarchical organization, and that these can be distinguished at the neural level.

What the proposed hierarchies of objects and actions have in common is that the three proposed hierarchical levels differ with respect to their degree of abstractness. The main point in which they differ, however, is that the action hierarchies described above focused on the goal of the action and the different means by which these goals can be achieved. Given an action goal (such as eating a piece of cake), what are the underlying essential elements (such as cutting the cake, placing a piece of cake on a plate, picking up a fork) that are required to achieve it? By contrast, the taxonomy of objects proposed by Rosch et al. (1976) emphasizes the importance of object representations at the basic level, given that attributes provided for objects at this level have the highest cue validity, and that objects at this level are recognized faster than objects at the other two levels. Do the characteristics of objects at the superordinate, basic, and subordinate level reported by Rosch et al. (1976) also hold for the organization of observed actions?

Experiments 1–3 were conducted to create a set of actions (using verbal material) with a clear taxonomic structure. Since we aimed to use this stimulus set for future behavioral and neuroimaging studies, we wished to arrive at a final set of three to four superordinate categories and at least two basic and subordinate level actions within each of these superordinate categories. Moreover, since we anticipated that we would have to remove some of the actions for practical reasons (e.g., due to imbalances regarding ratings of abstractness or complexity), we started out with a broader range of actions than the final numbers we aimed to reach. Experiments 4–6 aimed to characterize the actions selected in Experiments 1–3 at the three taxonomic levels. Experiment 4 used a feature listing paradigm to determine the numbers of common, distinct and shared features across taxonomic levels. In Experiment 5 and 6, using visual material, we examined whether the three taxonomic levels differ in terms of priming and speed of recognition.

2.3 Experiments (rationales, methods, results and interim discussion)

2.3.1 Experiment 1

(stimulus selection, part I): rationale

In Experiment 1 we aimed to lay the foundation for the definition of a range of different actions at the three taxonomic levels. As a starting point, we selected action verbs corresponding to the basic level from Levin (1993), who suggested a hierarchical organization of English verb classes based on similarities of verb meaning and syntactic expressions. Note that the classes of verbs from Levin are based on linguistic criteria rather than explicit human judgment. Thus, to determine whether explicit ratings of semantic similarity reveal similar clusters as those proposed by Levin (1993), we instructed participants to provide ratings for pairs of action verbs in terms of the similarity of their meaning. To select suitable actions for our stimulus set from at least three superordinate categories, we carried out hierarchical cluster analysis.

Experiment 1: methods

Participants

Twenty-one native German speakers (female: 12, age: 22.3 ± 4.2 years) took part in the experiment. All participants received written instructions about the experimental procedures and were reimbursed for their participation. Participants consented to participate in the study via button click. Procedures were approved by the local Ethics Committee at the University of Regensburg.

Stimuli

We selected German translations for 35 verbs from eight categories (communication, locomotion, ingesting, change of state, learning, grooming and body care, perception, creation and transformation). For the aim of future behavioral experiments (see, e.g., Experiments 5 and 6) for which we aimed to use images of actions, we focused on action verbs that we considered suitable to be depicted as images. The list of verbs in their infinitive form, together with their German translations and the corresponding categories is provided in Table S1.

Procedure

During the experiment, action phrases consisting of verbs in their infinitive form were presented in pairs on the screen. Participants were instructed to rate the semantic similarity of these pairs of action phrases on a scale from 1 (very dissimilar) to 7 (very similar). The rating procedure was carried out using an online survey (SoSciSurvey) and took approximately 20 min per participant. Details of the instruction are provided in the Supplementary Material (Experiment 1 section).

Data analysis

Data analysis was carried out in MATLAB. To examine taxonomies of verbs, we conducted hierarchical cluster analysis on the provided semantic similarity ratings and visualized the results as a dendrogram. To determine the optimal number of clusters, we computed the silhouette index (Rousseeuw, 1987). The silhouette index provides an estimate of the average distance between clusters—the higher the index, the better the items cluster.

Experiment 1: results

The silhouette analysis indicated that the optimal number of clusters was six (Fig. S1). The corresponding clusters are shown in Fig. 2.1 and consist of actions related to locomotion, ingestion, object manipulation, sensation/ leisure related actions, learning/studying, and communication (see also Tucciarelli et al., 2019). The results showed similarities with the classification proposed by Levin (1993), but we also obtained some differences. For example, we obtained a cluster consisting of 'to see', 'to observe', 'to write' and 'to read' that was distinct from another cluster that contained the verbs 'to feel', 'to touch', 'to paint' and 'to draw'. By contrast, in Levin's classification, 'to see' and 'to observe' were clustered together with 'to feel' and 'to touch'. These differences are not unexpected, given that Levin's classification was based on associations between syntactic properties of English verbs and their meanings, whereas the current study used explicit ratings of semantic similarities of verbs provided in German. We wish to point out that while it is interesting to notice these differences, a systematic comparison between the classification proposed by Levin and the results of the explicit ratings of semantic similarity was outside the scope of the current study.

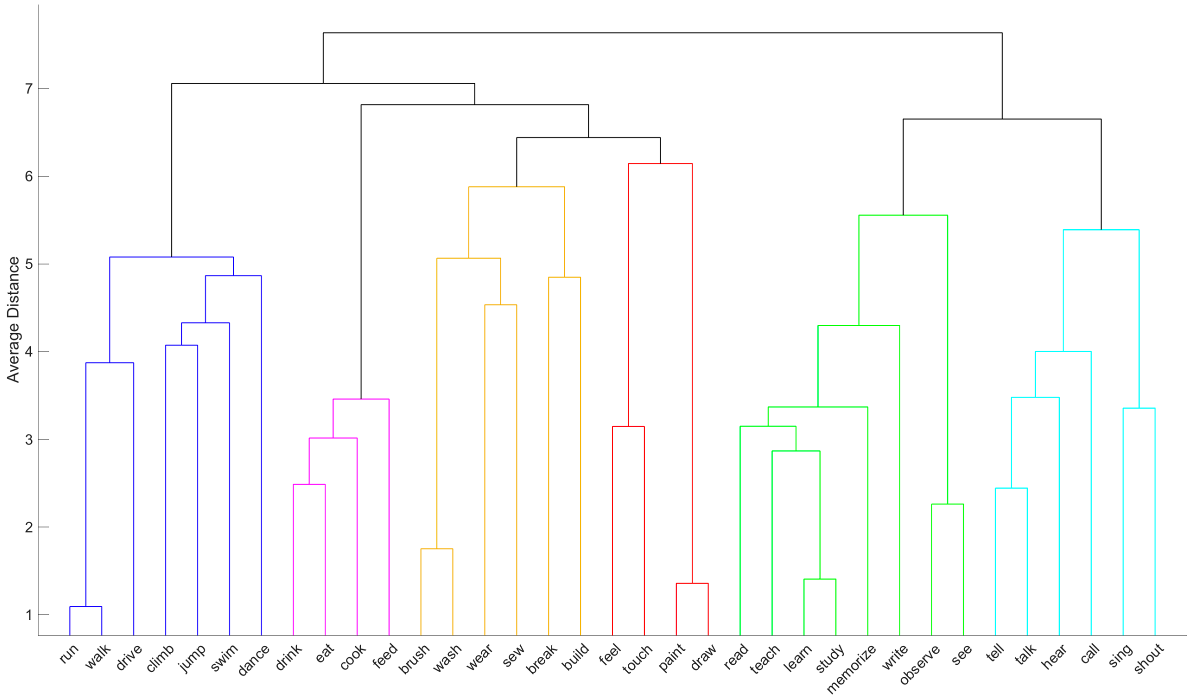


Fig. 2.1. Dendrogram illustrating the results of the hierarchical clustering analysis. Actions belonging to the same cluster are highlighted in the same color. Blue: locomotion, purple: ingestion, yellow: object manipulation, red: sensation/ leisure-related actions, green: learning/ studying, turquoise: communication.

2.3.2 Experiment 2

(stimulus selection, part II): rationale

From the results of Experiment 1, we obtained six clusters (which we considered to correspond to actions at the superordinate level) and their members (which we considered to correspond to actions at the basic level). The goal of Experiment 2 was to select and establish the superordinate, basic, and subordinate level structure of actions with an independent set of participants using a taxonomic depth task (see also Rifkin, 1985). To this aim, participants were instructed to write down labels for members of basic level actions (corresponding to subordinate level actions), and to provide labels (corresponding to superordinate categories) for the clusters revealed in Experiment 1.

Experiment 2: methods

Participants

Twenty native German speakers (female: 17, age: 21.5 ± 1.5 years) participated in Experiment 2. All participants were informed about the study procedures and received payment for participating. Participants consented to participate in the study via button click. Procedures were approved by the local Ethics Committee at the University of Regensburg.

Stimuli

We used a subset of the German action verbs used in Experiment 1. First, we excluded action verbs that might be ambiguous when depicted as a picture rather than a word (e.g., to learn—lernen, to teach—unterrichten, to read—lesen). Moreover, we excluded action verbs for which their subordinate actions could have been difficult to identify as a picture (e.g., to memorize—merken). Based on these criteria, we excluded two superordinate categories (sensation/ leisure related actions: to feel, to touch, to paint, to draw; learning/ studying: to read, to teach, to learn, to study, to memorize, to write, to observe, to see). Following the same criteria, we kept two to three basic level actions for each of the remaining superordinate categories for the taxonomic depth task (locomotion: to walk, to drive, to swim; ingestion: to drink, to eat, to cook; object manipulation: to brush, to wash; communication: to tell, to talk, to hear).

Procedure

The experiment consisted of a taxonomic depth task similar to the one used by Rifkin (1985). It was carried out with an online questionnaire (http://www.soscisurvey.com) and consisted of two parts. In the first part, participants were asked to provide names of actions at the subordinate level. In each trial, one of the action verbs at the basic level (e.g., to drive/ fahren) was displayed on the top of the screen. Participants were instructed to write down at least two subordinate names within 30 s, such as ‘to ride a bike’ (Fahrrad fahren) or ‘to drive a car’ (Auto fahren). The order of the action verbs was randomized.

In the second part, participants were asked to provide superordinate names for clusters of actions. During each trial, participants were provided with those two to three verbs belonging to individual clusters obtained from Experiment 1 (e.g., to walk, to drive, to swim), and were instructed to type the superordinate name these verbs belong to (e.g., locomotion) within 30 s. The order in which clusters were presented to participants was randomized. The whole experiment took approximately 10 min. Instructions for both parts are provided in the Supplementary Material.

Data analysis

After obtaining names of actions at the subordinate (first part) and superordinate (second part) level, two German native speakers combined similar answers into one answer and then labelled it. For instance, ‘Bewegung’ and ‘Fortbewegen’ was combined together and labelled ‘Bewegung’ (locomotion). For each cluster at the superordinate level, we chose one final label based on the most frequently used labels. For the subordinate level, we kept/ chose the two or three action verbs with the highest frequencies.

Experiment 2: results

The results of Experiment 2 are summarized in Table 2.1, showing labels obtained for superordinate (left column) and subordinate level actions (right column) next to the basic level actions (middle column) that served as the input to the taxonomic depth task. Note that due to the nature in which we generated the subordinate level actions, we did not control for the use of object versus non-object-directed actions.

Table 2.1. Labels for action categories at the superordinate and subordinate level resulting from the taxonomic depth task (Experiment 2).

Superordinate level	Basic level	Subordinate level
Locomotion (*sich fortbewegen*)	to go (*gehen*)	to walk (*spazieren gehen*)
		to hike (*wandern*)
		to walk a dog (*Gassi gehen*)
	to drive (*fahren*)	to drive a car (*Auto fahren*)
		to ride a bike (*Fahrrad fahren*)
		to take a bus (*Bus fahren*)
	to swim (*schwimmen*)	to swim front crawl *(Kraulschwimmen)*
		to swim breaststroke (*Brustschwimmen*)
Ingestion (*Nahrung aufnehmen*)	to drink (*trinken*)	to drink water (*Wasser trinken*)
		to drink beer (*Bier trinken*)
		to drink coffee (*Kaffee trinken*)
	to eat (*essen*)	to eat an apple (*einen Apfel essen*)
		to eat cake (*Kuchen essen*)
	to cook (*kochen*)	to cook noodles (*Nudeln kochen*)
		to cook soup (*Suppe kochen*)
Cleaning (*sauber machen*)	to brush (*putzen*)	to clean windows (*Fenster putzen*)
		to brush teeth (*Zähne putzen*)
		to clean the bathroom (*Bad putzen*)
	to wash (*waschen*)	to wash clothes (*Wäsche waschen*)
		to do the dishes (*Geschirr abwaschen*)
		to clean the face (*Gesicht waschen*)
Communication (*Kommunizieren*)	to talk (*sich unterhalten*)	to talk to friends (*sich mit Freunden unterhalten*)
		to talk on the phone (*sich am Telefon unterhalten*)
	to listen (*Hören*)	to listen to someone (*jemandem zuhören*)
		to listen to the radio (*Radio hören*)
	to tell (*erzählen*)	to tell a joke (*einen Witz erzählen*)
		to tell a story (*eine Geschichte erzählen*)

Basic level actions were selected on the basis of Experiment 1.

2.3.3 Experiment 3

(stimulus selection, part III): rationale

The purpose of Experiment 3 was twofold. First, to further validate the relationship between actions at the subordinate and the superordinate level resulting from Experiment 2, we instructed a new set of participants to explicitly rate the relationship between actions at these two levels. Second, for the purpose of future behavioral and neuroimaging studies, we aimed to balance actions at the subordinate level with respect to their degree of abstraction and complexity. The reason for this is that both complexity and the level of abstraction are known to have an impact on behavioral measures as well as on the corresponding neuronal signatures (e.g., Breedin et al., 1998; Gennari & Poeppel, 2003; Moseley & Pulvermüller, 2014; Schwanenflugel, 1991; Wang et al., 2010). To this aim, we asked participants to rate the degree of abstraction and complexity of each action at the subordinate level. We used these ratings to identify and remove 'unsuitable actions' that we treated as outliers, since they differed from the other actions at the subordinate level in terms of the level of abstractness and the level of complexity, and since the judged relationship between the subordinate and the superordinate level for these actions was either too low within a category or too high between categories.

Experiment 3: methods

Participants

Twenty-five native German speakers (female: 16, age: 22.0 ± 6.0 years) were recruited to take part in this experiment. All participants were informed about the study procedures and received payment for participating. Participants consented to participate in the study via button click. Procedures were approved by the local Ethics Committee at the University of Regensburg.

Stimuli

We used the German action category labels at the subordinate and superordinate level resulting from Experiment 2 (see Table 2.1).

Procedure

Both parts of the experiment were performed using an online platform (http://www.soscisurvey.com). For the rating of the relationship of actions at the superordinate and subordinate level, separately for each superordinate category, the superordinate category label of an action (e.g., locomotion) was shown at the top of the screen and all 27 subordinate level actions (e.g., to walk a dog, to hike, …, to tell a story; see Table 2.1) were shown underneath. Participants were instructed to assess the relation of each action at the subordinate level with the action at the superordinate level on a scale from 1 (weak relationship) to 7 (strong relationship).

For the rating of abstraction and complexity, participants were instructed to rate the degree of abstraction and the degree of complexity on a scale from 1 to 7 (1: very concrete, 7: very abstract; 1: very simple, 7: very complex). Further details about the instruction are provided in the Supplementary Material.

Data analysis

First, to remove outliers within a category for ratings of relatedness, we used the median absolute deviation (MAD; Leys et al., 2013). Second, after removing outliers we used the Mann–Whitney U test to compare ratings of relatedness of actions at the superordinate and superordinate level between within-category (e.g., locomotion—to walk a dog) and between-category (e.g., locomotion—to eat an apple) pairs of actions. Third, we removed outliers based on ratings of abstraction and complexity using ± 1.5 MAD.

Experiment 3: results

The ratings of the relationship between actions at the superordinate and the subordinate level within and across categories are shown in Table 2.2. Based on the MAD, we removed the subordinate level action ‘to listen to the radio’, because its relation to the superordinate category ‘communication’ was considered too low (rating = 2.96, median = 5.82, lower bound = 4.12).

As expected, the median rating for the relationship between actions at the superordinate and the subordinate level was higher for subordinate level actions belonging to the same superordinate category (e.g., ingestion—to drink water; median = 5.80) than for subordinate level actions belonging to a different superordinate category (e.g., ingestion—to hike; median = 2.16). This observation was supported by the Mann–Whitney U test [$U_{(N \text{ within category} = 27, N \text{ between categories} = 81)} = 6.00$, $z = -7.72$, $p < 0.001$, (N: sample size)].

Ratings of abstraction and complexity are provided in Table 2.3. The actions ‘to tell a joke’ and ‘to tell a story’ were removed, since their ratings of abstraction were detected as outliers by the MAD. Finally, we excluded the category ‘communication’ from further experiments, because only one action at the basic level was left after these actions were excluded.

Table 2.2. The relationship between actions at the superordinate (columns) and subordinate (rows) level within (highlighted in grey, yellow, green, and blue) and across categories (1: very weak relationship; 7: very strong relationship).

	locomotion (*sich fortbewegen*)	ingestion (*Nahrung aufnehmen*)	cleaning (*sauber machen*)	communication (*kommunizieren*)
to walk (*spazieren gehen*)	5.92	2.16	1.64	3.52
to hike (*wandern*)	5.84	2.64	1.80	2.84
to walk a dog (*Gassi gehen*)	5.56	1.76	2.24	2.96
to drive a car (*Auto fahren*)	6.16	2.32	2.16	3.20
to ride a bike (*Fahrrad fahren*)	6.12	2.04	1.92	2.24
to take a bus (*Bus fahren*)	6.36	2.20	1.72	2.92
to swim front crawl (*Kraulschwimmen*)	4.52	1.88	1.72	2.56
to swim breaststroke (*Brustschwimmen*)	5.40	1.96	1.68	2.24
to drink water (*Wasser trinken*)	2.36	5.68	2.20	1.84
to drink beer (*Bier trinken*)	2.60	4.96	1.84	3.60
to drink coffee (*Kaffee trinken*)	2.08	5.80	1.92	3.24
to eat an apple (*einen Apfel essen*)	2.12	6.00	1.92	1.80
to eat cake (*Kuchen essen*)	2.20	6.24	2.32	2.96
to cook noodles (*Nudeln kochen*)	2.00	5.04	2.16	1.88
to cook soup (*Suppe kochen*)	2.28	5.44	2.36	1.92
to clean windows (*Fenster putzen*)	2.40	1.88	6.44	1.84
to brush teeth (*Zähne putzen*)	2.20	2.64	5.20	1.76
to clean the bathroom (*Bad putzen*)	2.68	1.84	6.24	1.72
to wash clothes (*Wäsche waschen*)	2.36	2.00	6.40	1.92
to do the dishes (*Geschirr abwaschen*)	1.88	3.12	5.80	1.76
to clean the face (*Gesicht waschen*)	1.72	2.08	5.80	1.68
to talk to friends (*sich mit Freunden unterhalten*)	2.68	2.68	1.88	6.44
to talk on the phone (*sich am Telefon unterhalten*)	2.24	2.04	1.72	6.28
to listen to someone (*jemandem zuhören*)	2.28	2.60	1.92	5.88
to listen to the radio (*Radio hören*)	2.52	2.12	2.20	**2.96**
to tell a joke (*einen Witz erzählen*)	2.44	1.92	2.00	5.24
to tell a story (*eine Geschichte erzählen*)	2.44	2.36	1.60	5.76

Note: Actions that were rated as outliers using MAD analysis (Leys et al., 2013) are marked in bold and were removed from further analyses.

Table 2.3. Ratings of abstraction (1: very concrete, 7: very abstract) and complexity (1: very simple, 7: very complex).

	abstractness	complexity
to walk (*spazieren gehen*)	2.76	2.68
to hike (*wandern*)	2.64	3.68
to walk a dog (*Gassi gehen*)	3.72	3.04
to drive a car (*Auto fahren*)	2.36	4.36
to ride a bike (*Fahrrad fahren*)	2.20	3.72
to take a bus (*Bus fahren*)	2.56	2.96
to swim front crawl (*Kraulschwimmen*)	2.76	4.08
to swim breaststroke (*Brustschwimmen*)	2.44	4.08
to drink water (*Wasser trinken*)	1.92	1.80
to drink beer (*Bier trinken*)	2.56	2.08
to drink coffee (*Kaffee trinken*)	2.60	2.04
to eat an apple (*einen Apfel essen*)	2.28	2.52
to eat cake (*Kuchen essen*)	2.24	1.92
to cook noodles (*Nudeln kochen*)	2.20	2.64
to cook soup (*Suppe kochen*)	2.84	3.56
to clean windows (*Fenster putzen*)	2.92	3.68
to brush teeth (*Zähne putzen*)	1.92	2.60
to clean the bathroom (*Bad putzen*)	3.08	3.80
to wash clothes (*Wäsche waschen*)	2.60	3.36
to do the dishes (*Geschirr abwaschen*)	2.04	3.12
to clean face (*Gesicht waschen*)	1.96	2.20
to talk to friends (*sich mit Freunden unterhalten*)	3.36	3.20
to talk on the phone (*sich am Telefon unterhalten*)	3.16	2.88
to listen to someone (*jemandem zuhören*)	3.52	2.92
to listen to the radio (*Radio hören*)	3.36	2.28
to tell a joke (*einen Witz erzählen*)	**4.88**	4.00
to tell a story (*eine Geschichte erzählen*)	**4.48**	3.92

Note: Mean ratings of abstraction and complexity for actions provided at the subordinate level belonging to one of four different superordinate levels (grey: 'locomotion'; yellow: 'ingestion'; green: 'cleaning' and blue: 'communication'). Actions that were determined as outliers using MAD analysis (Leys et al., 2013) are marked in bold and were removed.

2.3.4 Experiment 4

(feature listing): rationale

The purpose of Experiment 4 was to examine the characteristics of actions at the three different taxonomic levels. To this aim, separately for each action at the superordinate, basic and subordinate level, participants were instructed to list features. As an example, the features of the action 'to write' could be 'hand', 'fingers', 'paper,' and 'type'. Following Rosch et al. (1976), we determined the number of common features (see "Data analysis" for details). Moreover, we further divided common features into shared and distinct features (see also

Rosch, 1978). Following Rosch et al. (1976), shared features are the common features that are shared with other categories, whereas distinct features refer to those features that belong to one category only. We hypothesized that participants provide more common features for actions at the basic level in comparison to the other two levels. Moreover, we hypothesized that participants provide the fewest common features for actions at the superordinate level, and that actions at the subordinate level share most features with actions from other categories at the same level.

Experiment 4: methods

Participants

To prevent that providing features at one level influences the ratings of features provided at another level, three separate groups of participants took part in the experiment for the listing of features at the superordinate, basic and subordinate level. Each group consisted of N = 20 native German speakers (group 1: 17 females, age: 22.9 ± 3.2 years; group 2: 15 females, age: 28.1 ± 10.5 years; group 3: 12 females, age: 21.1 ± 4.5 years). All participants were informed about the study procedures and were reimbursed for participating in the study. Participants consented to participate in the study via button click. Procedures were approved by the local Ethics Committee at the University of Regensburg.

Stimuli

As in Experiments 1–3, all action phrases were provided in German. Based on the results of Experiment 3, we excluded the german translations for ‘to listen to the radio’, ‘to tell a story’ and ‘to tell a joke’. Due to the small number of remaining actions, we decided to remove all actions corresponding to the cluster ‘communication’. Thus, we selected the superordinate level categories ‘locomotion’, ‘ingestion’ and ‘cleaning’ as well as their members at the subordinate level. Note that we did not control for co-occurences between different subordinate actions on the basis of the nouns included in the action phrases (e.g., to drink beer— Bier trinken; to clean the face—das Gesicht waschen).

To balance the number of actions within each category for the final set of actions, we chose two basic action categories for each action at the superordinate level and two subordinate action categories for each action at the basic level (see Table 2.4). Finally, to verify whether action phrases at the three taxonomic levels differ with respect to their degree of abstraction, we asked a separate group of N = 20 participants to rate the degree of abstraction (1: very concrete; 7: very abstract), separately for each of the action phrases provided in Table 2.4 (see Supplementary Material, Experiment 4 section, for details). As expected, ratings for abstraction were highest for the superordinate level (mean rating: 5.10), intermediate for the basic level (mean rating: 4.03), and lowest for the subordinate level (mean rating: 1.45). These observations are supported by the corresponding statistics ($H_{(2)} = 288.75$, $p < 0.0001$, $\eta^2 = 0.69$).

Table 2.4. Actions at the superordinate, basic and subordinate level used in Experiments 4 - 6.

Superordinate level	Basic level	Subordinate level
locomotion (*sich fortbewegen*)	to drive (*fahren*)	to drive a car (*Auto fahren*)
		to take a bus (*Bus fahren*)
	to swim (*schwimmen*)	to swim front crawl (*Kraulschwimmen*)
		to swim breaststroke (*Brustschwimmen*)
ingestion (*Nahrung aufnehmen*)	to drink (*trinken*)	to drink water (*Wasser trinken*)
		to drink beer (*Bier trinken*)
	to eat (*essen*)	to eat an apple (*einen Apfel essen*)
		to eat cake (*Kuchen essen*)
cleaning (*sauber machen*)	to brush (*putzen*)	to clean the windows (*Fenster putzen*)
		to brush teeth (*Zähne putzen*)
	to wash *(waschen)*	to do the dishes (*Geschirr abwaschen*)
		to clean the face (*Gesicht waschen*)

Procedure

Following the original procedures by Rosch et al. (1976), we carried out a free feature listing paradigm implemented as an online survey (http://www.soscisurvey.com) to investigate characteristics of actions at the taxonomic different levels. Three different groups of participants were presented with phrases describing actions at either the superordinate, basic or subordinate level. During each trial, a German action phrase consisting of a verb in its infinitive form (e.g., to drink water—Wasser trinken) was shown at the top of the page. Participants were instructed to write down as many features as possible related to that action within 2 min. The features could be related to body parts involved in the action, the target of the action, the type of movements involved, specific postures, the duration, the use of force, the pace etc. For example, features for ‘to drink water’ could be ‘glass’, ‘water’, ‘mouth’, ‘hand’, ‘bend’, etc. Features of ‘to drink’ could be ‘hand movement’, ‘water’, ‘beer’, ‘swallow’ and so on. Features of ‘ingestion’ could be ‘food’, ‘swallow’, ‘mouth’, ‘pour’, ‘hand’, etc. Details regarding the instructions are provided in the Supplementary Material.

Data analysis

To identify reliable features of each action per level, four German speakers judged these features. Features referring to the same meaning (e.g., ‘rag’ and ‘cleaning rag’) were merged (‘rag’). Answers that contained several features (e.g., ‘rotating arm’) were separated into individual features (e.g., ‘to rotate’, ‘arm’). By contrast, if only the combination of two features (e.g., ‘frische Luft’—‘fresh air’) but not the two individual features (‘frische’—‘fresh’, ‘Luft’—‘air’) referred to the action, we counted the combined feature rather than the individual features. The four judges were provided with a detailed coding scheme explaining these rules, together with concrete examples. The judge-amended tallies of each feature for each action were determined for further statistical comparisons.

To investigate the characteristics of actions at each taxonomic level, we computed the number of common features (the number of features provided by at least six out of 20 participants), separately for each of the three levels. Next, we separated common features into shared and distinct features at the three different levels to determine the distinctions of features. On the basis of the common features, distinct features were defined as features that were provided for one category only whereas shared features were determined as those that were provided for more than one category. Moreover, to examine distinctions between feature types, we distinguished common features into movement features, body-part features and object features.

To examine differences between the three hierarchical levels in terms of the mean number of common, distinct and shared features, we used the Kruskal–Wallis H test, which is a rank-based nonparametric test for the comparison of two or more independent samples with equal or unequal sample sizes. Next, we used Dunn's post hoc tests and corrected for multiple tests using Bonferroni correction. All statistical analyses were implemented in SPSS (https://www.ibm. com/analytics/spss-statistics-software). Effect sizes for the Kruskal–Wallis H test are reported as η^2.

Experiment 4: results

As can be seen in Fig. 2.2, actions at the basic level were described with more common features in comparison to the other two levels (panel A). At the same time, actions were described with more distinct features at the superordinate and the basic level in comparison to the subordinate level (panel B). Actions at the subordinate level were described with more shared features than actions at the superordinate level (panel C). These observations are supported by the corresponding statistics (see following section). To compare different types of features between the three taxonomic levels, we used the Kruskal–Wallis H test. The mean rank of common features (Fig. 2.2a) differed significantly between the three different levels ($H_{(2)} = 11.68$, $p = 0.003$, $\eta^2 = 0.58$; see Table 2.5 for details). Furthermore, the mean rank of common features of actions was 7.00 (numbers of actions (n) = 3) at the superordinate level, 18.17 (n = 6) at the basic level, and 8.42 (n = 12) at the subordinate level. Dunn's post-hoc tests indicated that participants provided more common features at the basic level in comparison to the superordinate (basic vs. superordinate level: $p = 0.03$) and the subordinate level (basic vs.

subordinate level: $p = 0.004$). However, we obtained no significant difference of common features between actions at the superordinate and subordinate level ($p > 0.99$).

The mean rank of distinct features (Fig. 2.2b) differed significantly between the three taxonomic levels ($H_{(2)} = 11.81$, $p = 0.003$, $\eta^2 = 0.59$). The mean rank of distinct features of actions provided by participants was 17.67 ($n = 3$) at the superordinate level, 15.58 ($n = 6$) at the basic level, and 7.04 ($n = 12$) at the subordinate level. Dunn's post-hoc test revealed that participants described actions at the superordinate and basic level with more distinct features than actions at the subordinate level (superordinate vs. subordinate level: $p = 0.02$; basic vs. subordinate level: $p = 0.02$). However, there was no significant difference in distinct features between the superordinate and basic level ($p > 0.99$).

The mean rank of shared features (Fig. 2.2c) differed significantly between the three levels ($H_{(2)} = 7.77$, $p = 0.02$, $\eta^2 = 0.39$). The mean rank of shared features was 2.00 ($n = 3$) at the superordinate level, 11.50 ($n = 6$) at the basic level, and 13.00 ($n = 12$) at the subordinate level. Dunn's post-hoc tests showed that participants provided more common features that were shared with other categories at the subordinate level compared to actions at the superordinate levels ($p = 0.02$). The number of shared features did not differ between actions at the basic and subordinate level ($p > 0.99$), and between actions at the basic and superordinate level ($p = 0.09$).

To explore differences between different types of features, we computed the number of common features separately for movement features, body-part features, and object features. The Kruskal–Wallis H test showed that the three levels differed in terms of the number of movement features ($H_{(2)} = 7.67$, $p = 0.02$, $\eta^2 = 0.38$, Table 2.5). The mean rank of movement features at the different levels were 5.00 ($n = 3$) at the superordinate level, 16.17 ($n = 6$) at the basic level, and 9.92 ($n = 12$) at the subordinate level. Post hoc tests revealed that participants provided more movement-related features at the basic in comparison to the superordinate level (p = 0.03). The number of movement features did not differ between actions at the subordinate and superordinate level (p = 0.63), or between actions at the basic and subordinate level (p = 0.12). In contrast to movement features, the number of body-part features ($H_{(2)}$ = 1.03, p = 0.317, η^2 = 0.05) and object features ($H_{(2)}$ = 4.39, p = 0.314, η^2 = 0.22) was not modulated by the taxonomic level.

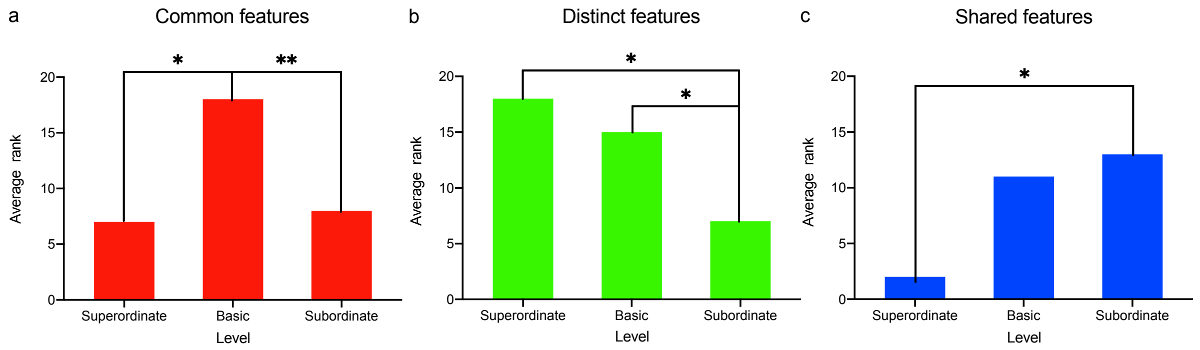


Fig. 2.2. Mean number of common, distinct and shared features of actions at the superordinate, basic and subordinate level. A: Actions at the basic level were described with more common features than actions at the other two levels. B: Actions at the superordinate and basic level were described with more distinct features than actions at the subordinate level. C: Actions at the subordinate level were described with more shared features than actions at the superordinate level.

Table 2.5. Results of the Kruskal-Wallis H test for common, distinct and shared features (upper part), and for common features, separately for movement, body part and object features (lower part).

	Test Statistic	Total number	df.	Sig.	η^2
Common feature	11.68	21	2	0.003***	0.58
Distinct feature	11.81	21	2	0.003***	0.59
Shared feature	7.77	21	2	0.021*	0.39
Common feature					
Movement feature	7.67	21	2	0.022*	0.38
Body part feature	1.03	21	2	0.317	0.05
Object feature	4.39	21	2	0.314	0.22

Note: p < .05*; p < .01**; p < .005***; p < .001****.

Experiment 4: interim discussion

The results from Experiment 4 revealed that the types of features provided by participants differed between the taxonomic levels, with the largest number of common features obtained at the basic level, the largest number of distinct features at the superordinate level and the largest number of shared features at the subordinate level. Regarding the number of common features, Rosch et al., (1976, experiment 1) reported more common features at the basic in comparison to the superordinate level, in line with our results. Rosch et al. (1976) argued that this may illustrate the distinctiveness of objects at the superordinate level. The larger number of distinct features at the superordinate in comparison to the subordinate level obtained in the current study is compatible with this interpretation.

Rosch et al. (1976) furthermore reported that the number of features added at the subordinate level (in comparison to the basic level) was smaller than the number of features added at the basic level (in comparison to the superordinate level). They argued that less information was added at the subordinate level, because at this level, many features were shared with other categories, and concluded that objects at the subordinate level have lower cue validity in comparison to objects at the basic level (Rosch, 1978). Note that in the current study, participants provided fewer common features at the subordinate in comparison to the basic level. Whereas Rosch et al. (1976) did not report the statistics between the number of common features beetween the basic and the subordinate level, it is obvious from their Table 2.2 that the number of common features provided at the subordinate level was larger in comparison to the basic level. We are uncertain regarding the reasons for this discrepancy between the results of Rosch et al. (1976) and our results. That said, the larger number of shared features for actions at the subordinate in comparison to the superordinate level obtained in the current study suggests that actions at the subordinate level have lower cue validity in comparison to actions at the superordinate level.

In sum, the results of Experiment 4 reveal systematic differences regarding the types of features provided for actions at the three taxonomic levels that are broadly consistent with the results reported by Rosch et al. (1976). To examine whether these differences are reflected in terms of the speed of processing, we followed up these results with an auditory priming experiment (Experiment 5).

2.3.5 Experiment 5

(priming): rationale

This experiment had two purposes. First, we aimed to examine at which taxonomic level actions are recognized first when depicted as pictures. Second, we aimed to determine whether participants are faster in processing the picture of an action if it is preceded by a matching (in comparison to a non-matching) action label (presented as an auditory cue), and if so, to which degree this priming effect is affected by the taxonomic level. To address these questions, we carried out a category verification task in which participants were presented with action labels (presented as auditory cues) at the subordinate, basic or superordinate level, followed by the image of an action. The task was to judge via button press whether the image corresponded to the action phrase. We hypothesized that participants verified actions at the basic level more rapidly than actions at the other two levels, and that participants responded faster and more accurately in matched in comparison to non-matched trials.

Experiment 5: methods

Participants

Twenty-three native German speakers (female: 20; age: 21 ± 3 years) took part in this experiment via the online platform lab.js (https://labjs.readthedocs.io/en/latest/). Participants were reimbursed for participating and were informed about the experimental procedures prior to consenting to take part in the experiment via button click. Procedures were approved by the local Ethics Committee at the University of Regensburg.

Stimuli

Action stimuli corresponded to the action category labels used in Experiment 4. Specifically, category labels at the superordinate, basic and subordinate level consisted of auditory recordings of a male native German speaker reading the category labels depicted in Table 2.4. Visual stimuli consisted of colour photographs depicting twelve actions at the subordinate level (six exemplars per action, for a total of 72 images).

Design and procedure

During each trial, participants were presented with an auditory cue corresponding to a category label at the superordinate, basic or subordinate level, followed by a static image of an action that disappeared as soon as participants provided a response (for a maximum of 2 s in case no response was provided). The stimulus onset asynchrony (SOA) between the onset of the auditory cue and the onset of the image was 1100 ms. In half of the trials, the auditory cue corresponded to the action image ('matched trials'), whereas the auditory cue and the action image belonged to different categories in the other half of the trials ('non-matched trials'). Participants were instructed to judge whether the action depicted in the image corresponded to the auditory cue. In case of a match, participants were asked to press the key 'f' with the right index finger (e.g., the auditory label 'to swim', followed by a picture of a person swimming backstroke), and the key 'j' with the left index finger in case of a non-match (e.g., the auditory label 'to ingest' followed by a picture of a person swimming backstroke). Participants were instructed to respond as fast and accurate as possible. The experiment consisted of 216 matched trials (72 trials for each of the three taxonomic levels), and 216 non-matched trials. The order of conditions was randomized.

Data analysis

Data from one participant were excluded, because he did not finish the experiment. Data from two participants were excluded, because the mean accuracy was more than two standard deviations below the group mean. Data from one participant were removed, since her mean RT was more than two standard deviations above the group mean. Data from 19 participants thus were used for further analysis. Next, we calculated the mean and standard deviation of response time (RT) and accuracy, separately for the three taxonomic levels and matched/non-matched auditory cues. Mean RT was computed on the basis of correct trials. To compare mean accuracy and RT between the three taxonomic levels and matched/non-matched trials, we used a 2-factorial repeated-measures ANOVA with the factors taxonomic level (superordinate, basic, subordinate) and type of auditory cue (matched/non-matched). We used paired samples t test to examine pairwise comparisons, and we used Bonferroni correction to correct for multiple comparisons. Effect sizes for the results of the ANOVA are reported as partial η^2.

Experiment 5: results

Figure 2.3 shows the mean RT and accuracy for matched and non-matched trials as a function of the taxonomic level of the auditory cue. RT was modulated by the taxonomic level of the auditory cue [main effect taxonomic level: $F_{(2, 36)} = 80.59$, $p < 0.001$, partial $\eta^2 = 0.82$]. Specifically, in comparison to actions preceded by an auditory cue at the superordinate level [mean RT = 661.24 ms, SEM: 18.20], participants responded faster to actions preceded by an auditory cue at the basic [(mean RT = 612.49 ms, SEM: 15.63, $t_{(37)} = 8.11$, $p < 0.001$)] and subordinate level [mean RT: 598.45, SEM: 17.01, $t_{(37)} = 11.06$, $p < 0.001$]. RT did not differ between actions preceded by an auditory cue at the basic and subordinate level ($t_{(37)} = 2.64$, $p = 0.07$). Participants responded faster when the auditory cue matched the action depicted in the image [main effect for factor type of auditory cue: $F_{(1, 18)} = 7.72$, $p = 0.01$, partial $\eta^2 = 0.30$], and this effect was modulated by the taxonomic level [interaction taxonomic level and type of auditory cue [$F_{(2, 36)} = 9.36$, $p = 0.001$, partial $\eta^2 = 0.34$]. Specifically, in comparison to non-matched trials, participants recognized actions faster if the auditory labels matched the action, both for auditory cues at the basic ($t_{(18)} = -4.40$, $p < 0.001$) and the subordinate ($t_{(18)} = -2.73$, $p = 0.01$) level, but not for the superordinate level ($t_{(18)} = -0.65$, $p = 0.52$).

As can be seen in Fig. 2.3b, accuracy was modulated by the taxonomic level of the auditory cue preceding the action image [main effect taxonomic level: $F_{(2, 36)} = 6.97$, $p = 0.003$, partial

$\eta^2 = 0.28$]. Specifically, in comparison to the subordinate level (mean accuracy: 0.962, SEM: 0.006), participants responded less accurate at the superordinate (mean accuracy: 0.941, SEM: 0.07; $t_{(37)} = -2.45$, $p = 0.004$) and the basic level (mean accuracy: 0.945, SEM: 0.07, $t_{(37)} = 2.87$, $p = 0.03$). The effect of the taxonomic level of the auditory cue was modulated by the match between the auditory cue and the action image [interaction taxonomic level × type of auditory cue: $F_{(2,\ 36)} = 15.70$, $p < 0.001$, partial $\eta^2 = 0.47$]. Pairwise t test showed that participants responded less accurate in matched in comparison to non-matched trials for auditory cues at the superordinate level ($t_{(18)} = -4.67$, $p < 0.001$). By contrast, they responded more accurately in matched in comparison to non-matched trials for auditory cues at the subordinate level ($t_{(18)} = 2.77$, $p = 0.01$). There was no difference in accuracy between matched and non-matched trials for auditory cues at the basic level ($t_{(18)} = 1.58$, $p = 0.13$).

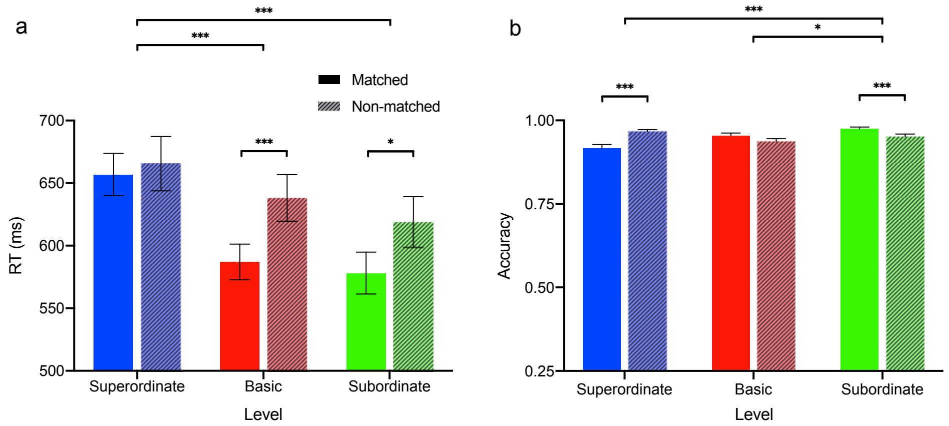


Fig. 2.3. Mean RT and accuracy in the auditory priming experiment (Experiment 5). A: Auditory cues at the basic and subordinate level led to faster responses in comparison to auditory cues at the superordinate level, and this effect was stronger if the auditory cue matched the action. B: Auditory cues at the subordinate level led to more accurate responses in comparison to auditory cues at the basic and superordinate level. Auditory cues that matched the following action picture led to more accurate responses at the subordinate level, whereas they led to less accurate responses at the superordinate level. Error bars show SEM.

Experiment 5: interim discussion

Experiment 5 showed that verbal cues at the basic and subordinate level, but not at the superordinate level, facilitate the speed of processing of observed actions, suggesting that the basic level is the most abstract level at which participants profit from an auditory cue. It is worth pointing out that these results were not fully matched by the corresponding accuracy data. Whereas participants responded more accurately in matched in comparison to non-matched trials for verbal cues at the subordinate level, in line with the RT results, this difference was not significant for verbal cues at the basic level. Moreover, unexpectedly, participants responded more accurately during non-matched in comparison to matched trials for verbal cues at the superordinate level. It is possible that for verbal cues at the subordinate level, participants profited from the fact that subordinate labels of observed actions always included the basic label action names (e.g., the subordinate level action ‘to swim breaststroke—‘Brustschwimmen’ in german includes the basic level action name ‘to swim’—‘schwimmen’ in German). Regarding the lack of a priming effect at the superordinate level, it seems likely that the superordinate action labels (locomotion, ingestion, cleaning) were simply too abstract to lead to a facilitatory effect on the processing of upcoming action images. We will return to this point in the General Discussion. We are less certain about the reasons underlying the observation that participants responded more accurately during non-matched in comparison to matched trials for verbal cues at the superordinate level.

To examine whether the three taxomomic levels also differ with respect to the speed of processing, we used a category verification task in Experiment 6 in which we systematically varied the taxonomic level of the category and the exposure duration of the action images.

Experiment 6
(recognition): rationale

The aim of Experiment 6 was to examine whether the speed to recognize an action depends on the taxonomic level. To examine the time course of action categorization, we used a rapid category verification task in which we varied both the taxonomic level of written action labels and the exposure duration of action images (see also de la Rosa et al., 2014; Hafri et al., 2013;

Mack et al., 2008). Experiments 4 and 5 revealed that the basic level includes the most common features, and that matched auditory cues speed up the processing of actions at the basic and the subordinate level, but not at the superordinate level. We thus hypothesized that participants required less time to verify the category of an action at the basic and subordinate level in comparison to the superordinate level.

Experiment 6: methods

Participants

Twenty native German speakers (female: 12; age: 26 ± 5 years) joined the experiment in a behavioral lab at the Institute of Psychology at the University of Regensburg. All participants consented to take part in the experiment. They either received course credits or money as a reward for their participation. Procedures were approved by the local Ethics Committee at the University of Regensburg.

Stimuli

We used the same images of action stimuli as in Experiment 5 (i.e., 12 subordinate actions × 6 exemplars each, for a total of 72 images). Scrambled images were created by randomly selecting and shuffling 10 × 10 pixels squares from all action images. Written category labels (font type: Calibri) corresponded to the german action words depicted in Table 2.4. Stimulus presentation and data collection was implemented with A Simple Framework (ASF, Schwarzbach, 2011), which is built around the Psychophysics Psychtoolbox (Brainard, 1997).

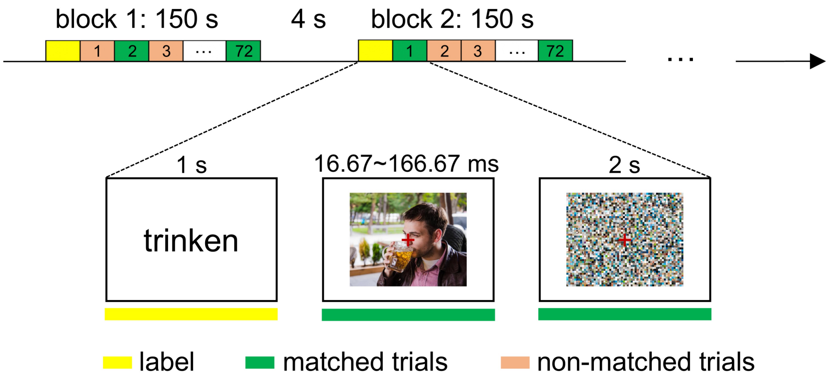


Fig. 2.4. Design and procedure used in Experiment 6. Upper panel: Each block consisted of 72 trials and lasted 150 s. Lower panel: At the beginning of each block, participants were presented with a written label (in german) corresponding to an action at one of the three taxonomic levels (e.g. 'trinken' – 'to drink'; see Table 2.4) for 1 s. This label was followed by a block of 72 trials. In each trial, participants were presented with an image of an action (duration: 16.67 to 166.67 ms in steps of 16.67 ms), followed by a scrambled mask (2 s). In each trial, participants were instructed to judge whether the action image (e.g. a picture of a person drinking a glass of beer) corresponded to the label provided at the beginning of the block (e.g. 'trinken' – 'to drink'). In the case of a match between the action depicted in the action image and the label ('matched trials'), participants were instructed to click the left mouse button, whereas they were asked to press the right button in the case of a non-match ('non-matched trials').

Design and procedure

At the beginning of a block of trials, participants were presented with a written label of an action at one of the three taxonomic levels (e.g., "locomotion", "to swim" or "to swim backstroke") at the centre of the screen for 1 s, which was immediately followed by the first trial of a block (Fig. 2.4). Each trial consisted of an image of an action (duration: 16.67, 33.33, 50, 66.67, 83.33 or 166.67 ms), immediately followed by a scrambled mask (2 s). Exposure duration was chosen on the basis of previous studies (de la Rosa et al., 2014; Hafri, et al., 2013; Mack et al., 2008).

Participants were instructed to decide as quickly as possible whether the image just shown matched the action label. If the image (e.g., a person driving a car) matched the written label (e.g., "locomotion"), participants should click the left mouse button. If image and action label did not match, (e.g., an image of a person talking on the phone and the label “locomotion”), participants should click the right mouse button. Participants were instructed to answer after each image of an action, even if they could not recognize the action.

Each block consisted of an equal proportion of matched and non-matched trials. In total, the experiment consisted of nine blocks (three for each taxonomic level). Note that using all the category labels shown in Table 2.4 would have led to an imbalance regarding the number of trials at the three taxonomic levels. To avoid this, we selected three category labels at the basic level and three category labels at the subordinate level for each participant, whereas we chose all three superordinate category labels for each participant. The selected category labels were balanced across participants. Regarding the action of the selected category label, action images of all six exemplars were presented with an equal proportion within a block. That is, each combination of the selected action (six exemplars) and exposure duration (six levels) was shown exactly once within each block, for a total of 36 matched trials per block. To balance the number of matched and non-matched trials, we randomly selected the same number of non-matched trials from all possible combinations. Thus, in total, each block contained 72 trials (36 matched, 36 non-matched). The order of conditions within each block was randomized. Each block lasted approximately 2.5 min with a 4 s break between blocks. The whole experiment lasted approximately 23 min.

Data analysis

Data from two participants were removed, since one participant’s mean RT was more than two standard deviations above the group mean, and another participant’s mean accuracy was more than two standard deviations above the group mean. Using the data from the remaining N = 18 participants, we analyzed mean RT (based on correct trials) and accuracy for matched trials using a two-factorial repeated- measures ANOVA with the factors taxonomic level (superordinate, basic and subordinate level) and exposure duration of the action image (16.67, 33.33, 50, 66.67, 83.33 and 166.67 ms). Significant interactions were followed up with pairwise comparisons (Bonferroni-corrected for multiple comparisons). Effect sizes for the results of the ANOVA are reported as partial η^2.

Experiment 6: results

Figure 2.5 shows mean RT (panel A) and accuracy (panel B) during matched trials as a function of the exposure duration of the action image, separately for the three taxonomic levels. As can be seen, participants responded faster and more accurately with increasing exposure duration [main effect exposure duration: $F_{(5,80)} = 13.51$, $p < 0.001$, partial $\eta^2 = 0.46$]. Importantly, participants responded faster to action images preceded by category labels at the basic and subordinate level in comparison to category labels at the superordinate level [main effect taxonomic level: $F_{(2,32)} = 9.99$, $p < 0.001$, partial $\eta^2 = 0.38$; pairwise comparison basic vs. superordinate level: $t_{(107)} = 4.36$, $p < 0.001$; subordinate versus superordinate level: $t_{(107)} = 5.24$, $p < 0.001$]. RT did not differ between actions preceded by category labels at the basic and subordinate level ($t_{(107)} = - 0.35$, $p > 0.99$). The effect of taxonomic level was not modulated by exposure duration [$F_{(10,160)} = 0.51$, $p = 0.88$, partial $\eta^2 = 0.03$].

As can be seen in Fig. 2.5b, accuracy was affected by taxonomic level at the longer exposure duration, but not at the shorter exposure duration [interaction taxonomic level × exposure duration: $F_{(10,170)} = 2.00$, $p = 0.04$, partial $\eta^2 = 0.11$]. At longer exposure durations (83.33 ms and 166.67 ms), participants responded more accurately to action images preceded by category labels at the basic and subordinate level in comparison to labels at the superordinate level [pairwise comparisons superordinate vs. basic at exposure duration = 83.33 ms: $t_{(17)} = - 4.68$, $p < 0.001$; at exposure duration = 166.67 ms: $t_{(17)} = - 3.69$, $p = 0.005$; superordinate vs. subordinate at exposure duration = 83.33 ms: $t_{(17)} = - 3.84$, $p = 0.004$; at exposure duration = 166.67 ms: $t_{(17)} = - 4.08$, $p = 0.002$]. In addition, at exposure duration = 50 ms, participants responded more accurately to action images preceded by category labels at the subordinate in comparison to the superordinate level [superordinate vs. subordinate at exposure duration = 50 ms: $t_{(17)} = 5.25$, p = 0.02]. Accuracy did not differ between actions preceded by category labels at the basic and subordinate level [$t_{(17)} = -1.54$, p = 0.64], or between the superordinate and basic level [$t_{(17)} = -1.29$, p = 0.43] at exposure duration = 50 ms.

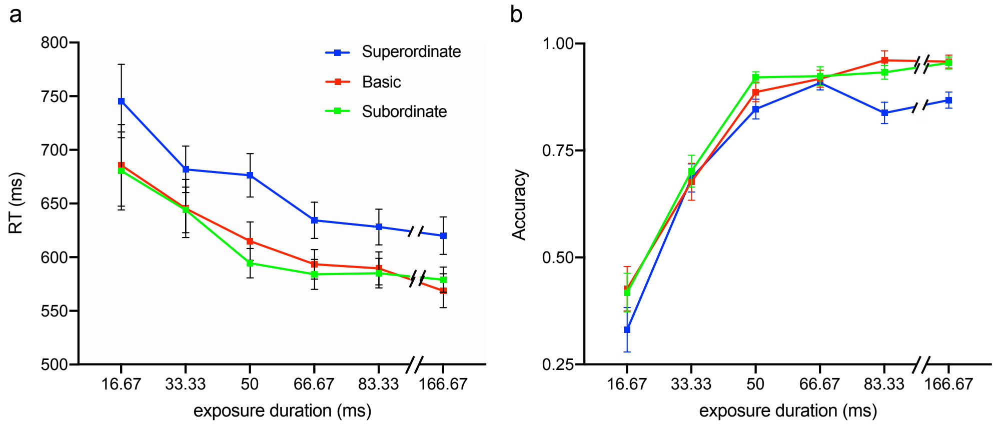


Fig. 2.5. RT and accuracy for matched trials as a function of the exposure duration of the action image, separately for the three taxonomic levels. A: Participants were faster to verify the category of actions at the basic and the subordinate level in comparison to the superordinate level across all examined exposure duration. B: For short exposure durations, the accuracy to verify the category of actions was not affected by the taxonomic level. For long exposure durations, participants were more accurate to verify the category of actions at the basic and subordinate level in comparison to the superordinate level. At exposure duration = 50 ms, participants responded more accurately to action images preceded by category labels at the subordinate in comparison to the superordinate level.

Experiment 6: interim discussion

Experiment 6 revealed that participants were faster and more accurate to verify actions at the subordinate and basic level in comparison to actions at the superordinate level. Performance did not differ between the subordinate and basic level. Whereas these results are broadly consistent with the results of Experiment 5, in particular with respect to RT, they differ from the results on the speed of object recognition reported by Rosch et al., (1976, Experiment 7), where participants were faster to verify the category of an object at the basic level in comparison to the superordinate and the subordinate level. We will return to this observation in the “General discussion”.

2.4 General discussion

Here we aimed to investigate the characteristics of actions at different hierarchical levels. The purpose of Experiments 1–3 was to select and characterize actions at the superordinate, basic and subordinate level to be used in the following experiments. The final set of actions we selected in Experiments 1–3 for the three taxonomic levels differed with respect to their degree of abstraction (see Rating of abstraction, Supplementary Material). In Experiment 4, using a feature listing paradigm, we found that participants provided the most common features for actions at the basic level. Actions at the basic and the superordinate level were described with more distinct features than actions at the subordinate level, while actions at the subordinate level shared more information with actions from different categories at the same level than actions at the superordinate level. In Experiment 5, we found that participants are faster to respond to images of actions preceded by a matching auditory cue at the basic and subordinate level, but not for matching cues at the superordinate level. In Experiment 6, we observed that participants are faster and more accurate to verify the category of an action depicted as an image at the basic and subordinate level in comparison to the superordinate level. In sum, basic level actions were described with the largest number of common features, with more distinct features in comparison to the subordinate level, and the basic level was the most abstract level at which a verbal cue facilitated the processing of an upcoming visual action. Together, these results are in line with the view that information about action categories is maximized at the basic level. In the following sections, we are going to discuss these results in more detail in relation to previous findings.

2.4.1 Comparison of taxonomic levels of objects and actions

Our feature listing paradigm (Experiment 4) revealed that participants provided more common features for actions at the basic level in comparison to the superordinate level, in line with the results reported by Rosch et al., (1976, Experiment 1). Likewise, using a similar paradigm for event categories, Rifkin (1985) and Morris and Murphy (1990) reported that participants provided more features for basic level events in comparison to events at the superordinate level, whereas they obtained no differences between the basic and the subordinate level. The results of our feature rating experiment thus suggest that, as for objects, the basic level contains more inclusive information about actions than the superordinate level, and that information at the basic level is best suited to determine similar items within a category and distinctions between other categories. Moreover, the larger number of distinct features at the superordinate level in comparison to the subordinate level suggests a higher distinctiveness between actions at this level. By contrast, the higher number of shared features at the subordinate level suggests that actions at this level are more cohesive and thus less distinguishable (see also Rosch, 1978).

The results of our auditory priming paradigm (Experiment 5), with faster responses in matching in comparison to non-matching trials for the basic and subordinate level, and the absence of a difference between matching and non- matching trials for the superordinate level, are in line with the results of the priming experiment reported by Rosch et al. (1976) for objects.

Our category verification task (Experiment 6) revealed that participants recognized actions faster at the basic level in comparison to the superordinate level, while we obtained similar results for the basic and the subordinate level, in line with the results of Experiment 5. Note that the absence of a difference between the basic and the subordinate level we obtained both in Experiments 5 and 6 is in line with the results of the feature-listing paradigm used by Rifkin (1985) and Morris and Murphy (1990) for events. Likewise, Rosch et al. (1976) reported faster responses in matched in comparison to non-matched trials for primes at the basic and the subordinate level, but no difference between matched and non-matched trials for primes at the superordinate level in a priming task. The priming effect did not differ between the basic and the subordinate level. By contrast, Rosch et al. (1976) obtained faster responses at the basic level in comparison to the superordinate and the subordinate level in an object recognition task (Experiment 7 in their study).

In summary, the difference between the basic and the superordinate level has been reported consistently across paradigms and stimulus domains (objects, actions). By contrast, the difference between the basic and the subordinate level is less consistent, with some studies reporting a difference, while other studies obtained no such difference. Finally, de la Rosa et al. (2014) directly compared the average recognition time for objects (e.g., car) and social interactions (e.g., to hug) at the basic and the subordinate level as a function of exposure duration (similar to the paradigm used in Experiment 5 in the current study). They found that both objects and social interactions were recognized faster and more accurately at the basic than at the subordinate level. However, this difference was substantially larger for objects than for social interactions.

What might determine under which circumstances categories at the basic level and the subordinate level are processed in a similar or in a different way? First, prior studies found no difference between the basic and the subordinate level in terms of speed and accuracy for the recognition of objects in experts from the corresponding object fields (Johnson & Mervis, 1997; Tanaka, 2001; Tanaka & Taylor, 1991). This raises the possibility that some of the differences between the subordinate and the basic level obtained in previous studies might be due to differences in terms of familiarity or expertise with the objects at the two taxonomic levels, whereas participants were likely to be highly familiar with the actions at the basic and subordinate level in the current study. Second, a possible reason for the lack of a difference in terms of priming effects (Experiment 5) and the speed of recognition (Experiment 6) between actions at the basic and the subordinate level observed in the current study lies in the fact that labels of actions at the subordinate level (e.g., ‘to swim breaststroke’—‘Brustschwimmen’) always included the label of the basic level (e.g., ‘to swim’—‘Schwimmen’). Note that the same holds for some of the object categories examined by Rosch et al., (1976; e.g., ‘desk lamp’—subordinate level/‘lamp’—basic level), but not for all of them (e.g., ‘Levis’—subordinate level/‘pants’—basic level). Third, and not mutually exclusive with respect to the previous two points, participants in the current study may have profited from the presence of objects in the majority of the subordinate action names, which might have abolished any differences between the basic and the subordinate level. Future studies are required to examine the exact circumstances under which the number of common features obtained in feature listing paradigms as well as the speed and accuracy for the recognition of objects and actions differs between the basic and the subordinate level.

In sum, while we noticed some differences with respect to previous studies, most of our results are in line with the results previously reported for the basic level advantage of objects and events. Together, the results of Experiments 4–6 suggest that actions at the basic level had maximum cue validity. Moreover, they reflect the effect of cognitive economy, with a trade-off between distinctiveness and informativeness.

2.4.2 The role of stimulus format

In comparison to visual representations, verbal descriptions are more specific and informative regarding the taxonomic level of a category (Morris & Murphy, 1990; Rosch et al., 1976). As an example, we can refer to the subordinate (‘to eat an apple’), basic (‘to eat’) and superordinate level (‘to ingest’) verbally, whereas it is difficult to depict the three different levels in the visual format (unless the visual format is embedded in a task, such as a category verification task or a priming paradigm). Another important difference between the verbal and the visual format lies in the fact that a number of studies on actions in the visual domain emphasized the way in which these actions are performed and the goal that one aims to achieve with these actions (e.g., Spunt et al., 2016; Hamilton & Grafton, 2006; Wurm & Lingnau, 2015). Consequently, it has been proposed that one important principle underlying the representation of observed actions is an organization according to their goals (Hamilton & Grafton, 2006; Tunik et al., 2005).

Regarding verbal material, Schank and Abelson (1977) described the internal structure of scripts, such as ‘going to a restaurant’, with a specific emphasis on action primitives, such as move, speak, or ingest. However, as mentioned also by Morris and Murphy (1990), this line of research was concerned about the relationship between parts of scripts (such as ‘waiting to be seated’, ‘ordering food’) rather than the relationship between different semantic categories (such as ‘going to a restaurant’ and ‘visiting a museum’). By contrast, other studies on actions in the verbal format focused on the role of the grammatical class (in particular, verbs vs. nouns; see, e.g., Peelen et al., 2012), the distinction between action versus non-action verbs (e.g., Papeo & Lingnau, 2015; Papeo et al., 2015), and semantic categories or semantic fields (e.g., Pinker, 1989; Talmy, 1985). Regarding the latter, a number of studies focused on the (horizontal) organization of semantic categories, such as change of location, communication and change of state (e.g., Vinson & Vigliocco, 2008), corresponding to the superordinate level used in the current study.

2.4.3 The role of different types of features

In their experiment 2, Rosch et al. (1976) instructed participants to list motor (i.e., body and muscle) movements associated with specific objects. They found that participants provided fewer motor movements for objects at the superordinate level in comparison to objects at the basic and subordinate level. By contrast, they observed no difference for motor movements associated with objects at the basic and subordinate level. These results illustrated that the basic level of objects was the most inclusive level at which many motor movements interacted with objects. Experiment 2 by Rosch et al. (1976) provided a good foundation for investigating motor features of objects across taxonomic levels. In line with this view, a number of studies highlighted the importance of motor- and body-related features, such as movement kinematics (Cavallo, et al., 2016), movement force (Casiraghi et al., 2019) and the amount of arm movement and hand posture (Watson & Buxbaum, 2014) for the processing of observed actions (see also de Gelder & Poyo Solanas, 2021, for a recent discussion of the importance of midlevel features).

Inspired by these previous studies, we further subdivided the common features into movement, body-part and object features in an exploratory analysis. Features related to body parts and objects did not differ between taxonomic levels. By contrast, participants used more movement features to describe actions at the basic level in comparison to the superordinate level, in line with the idea that the basic level is most inclusive also with respect to movement-related information (see also Rosch et al., 1976, Experiment 2). However, features related to motor movements are not the only features that play a role in the categorization of actions. As an example, several recent studies emphasized the role of high level features, such as the target (e.g., a person or an object; Tarhan & Konkle, 2020; Wurm et al., 2017) or the emotional valence of an action (e.g., Kroczek, et al., 2021; Portugal et al., 2020).

2.4.4 Future directions

The current set of experiments lays the foundation for a number of interesting lines for future research. As an example, under which conditions is there a behavioral advantage for actions at the basic in comparison to the subordinate level, and under which circumstances do the two levels lead to similar behavioral effects? To which degree are the behavioral effects modulated by the typicality of an observed action (see also Murphy & Brownell, 1985, for typicality constraints on the basic object advantage), or by the presence or absence of a concrete target object? Moreover, it will be interesting to examine which taxonomic level is learned first by children, and whether there are differences in the processing of observed actions at the different taxonomic levels for young and elderly adults. Finally, it will be important to establish a link between the hierarchical organization of actions examined in the current study and the underlying neural representation in space and time.

2.4.5 Conclusions

Understanding whether a hierarchical structure may be an emergent property underlying the organization of observed actions is key to answering the broader question about how the human brain extracts and organizes information from the surrounding world in a flexible way. The current study extends previous studies focusing on the horizontal organization of observed actions by examining the vertical organization of actions (see also Vallacher & Wegner, 1985; Rifkin, 1985; Morris & Murphy, 1990). Our findings are in line with the view that there is a basic level advantage not only for objects, but also for actions.

3 Chapter 3: Study2

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

This study was accepted on September 6, 2023 and first published on October 5, 2023, in Journal of Neuroscience and is available through an open-access link with the DOI: 10.1523/JNEUROSCI.0700-22.2023.

3.1 Abstract

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

Keywords: action categorization; action observation; action recognition.

Significance statement

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

3.2 Introduction

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

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

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

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

Zhuang and Lingnau (2022) examined the characteristics of observed actions at the three taxonomic levels. Actions at the three levels differed with respect to the number and type of features participants used to describe them, and in their ratings of abstraction. Moreover, participants verified the action category faster at the basic and subordinate level in comparison to the superordinate level. Together, these results suggest that the basic level holds the maximized information, consistent with the basic level advantage reported for objects (Rosch et al., 1976). Given these behavioral results, here we aimed to determine which brain regions (a) represent observed actions at the three taxonomic levels, and (b) which brain regions host a joint representation across these levels.

3.3 Methods

Overall rationale and hypotheses

To reveal which brain areas represent actions at the three taxonomic levels, we separated twelve daily actions into three action categories at the superordinate level (see also Zhuang and Lingnau, 2022). Each superordinate action category consisted of two types of actions at the basic level, and each basic level action encompassed two actions at the subordinate level (Figure 3.1). To verify this hierarchy, we used a multi-arrangement experiment (Kriegeskorte and Mur, 2012) combined with inverse multidimensional scaling (MDS) and hierarchical cluster analysis. Next, to determine which brain areas represent observed actions at the three different hierarchical levels, we conducted an fMRI experiment and carried out ROI-based and whole-brain searchlight-based representational similarity analysis (RSA; Kriegeskorte et al., 2008). Specifically, we examined the representation of observed actions at the subordinate, basic and superordinate level, and the representation of the behavioral similarity structure resulting from the multi-arrangement experiment.

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

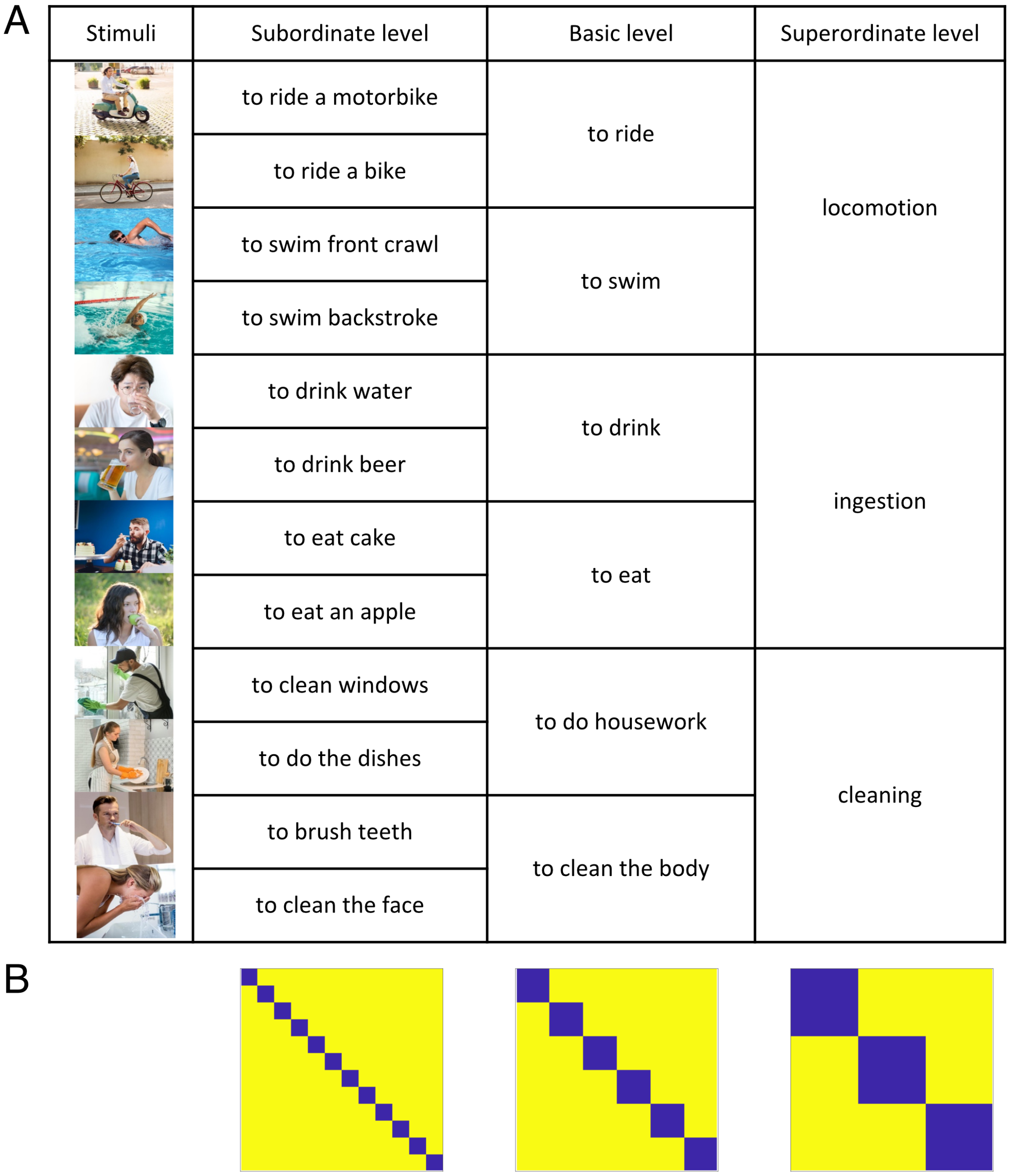
Stimulus selection and validation

Stimuli consisted of static images of twelve different actions (600 x 480 pixels, 14.36 x 11.07 degree of visual angle; six exemplars each; see Figure 3.1 for an overview of stimulus exemplars and corresponding action words). The twelve actions were chosen on the basis of a series of rating and behavioral studies (Zhuang and Lingnau, 2022) that we briefly summarize here. First, we selected action verbs corresponding to the basic level from Levin (1993). Using these action verbs, we carried out a semantic similarity rating, followed by hierarchical cluster analysis. Based on the resulting clusters, we selected a subset of basic level actions, excluding actions that might be hard to portray as a picture (e.g. to learn, to memorize). To select labels for the superordinate level, a new set of participants was provided with the basic level labels of actions belonging to a given cluster revealed by the hierarchical cluster analysis. To select actions belonging to the subordinate level, participants were provided with different action verbs corresponding to the basic level and were asked to generate action verbs corresponding to the subordinate level. Next, another group of participants was asked to rate (a) the relationship between actions at the subordinate and the superordinate level (e.g. between ‘swim front crawl’ and ‘locomotion’, or between ‘swim front crawl’ ‘and ingestion’) and (b) the degree of abstraction and complexity of each action at the subordinate level. Actions were only included in the final set if they were consistently rated to belong to a given superordinate category, and not to other superordinate categories.

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

Since the rating studies were based on written words, we first wanted to verify how human participants categorize these actions when presented as static images. To this aim, we carried out a multi-arrangement experiment (Kriegeskorte et al., 2008) as implemented in the online platform MEADOWS (https://meadows-research.com) in a group of N=18 participants (12 female; mean age: 27 years; range: 22-31 years) that did not take part in the fMRI experiment. Participants were instructed to judge the degree of similarity between the twelve actions depicted in these static images, to arrange them accordingly (i.e., the more similar in meaning, the closer they should be positioned on the screen), and to press a button when they were satisfied with the arrangement of the stimuli. In the first trial, all stimuli appeared on the screen (stimulus size: 48 x 38 pixel). In all subsequent trials, an adaptive algorithm chose a subset of all stimuli in order to provide the optimal evidence for pairwise dissimilarity estimates (see Kriegeskorte and Mur, 2012, for details). The experiment continued until the adaptive algorithm reached the required evidence level for pairwise dissimilarities. The full stimulus set contained 72 static images (12 actions x 6 exemplars). Each participant was provided with 12 different actions. Action exemplars were counterbalanced across participants.

Results of the multi-arrangement experiment were collapsed across participants and averaged across image exemplars. We visualized the results by creating a 12 x 12 representational dissimilarity matrix (RDM) where each cell contains a value corresponding to the Euclidean distance between two actions (Figure 3.2A). Figure 3.2B shows a 2-dimensional (2D) arrangement derived from multidimensional scaling (metric stress), averaged across participants. Inverse MDS revealed three larger clusters corresponding to the three superordinate action categories. Subsequently, to reveal the corresponding hierarchical structure, we carried out average-linkage hierarchical cluster analysis (using the Matlab function *linkage*) on the results obtained from the multi-arrangement experiment. The results are shown in Figure 3.2C. This analysis confirmed the hierarchical structure of the selected actions, with three action categories corresponding to the superordinate level (locomotion, ingestion and cleaning), six action categories at the basic level (to swim, to ride, to eat, to drink, to clean the body and to do housework), and twelve actions at the subordinate level (to ride a motorbike, to ride a bike, to swim front crawl, to swim backstroke, to drink water, to drink beer, to eat cake, to eat an apple, to clean windows, to do the dishes, to brush teeth, and to clean the face).



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

Figure 3.1. Top panel: Stimulus set and corresponding hierarchical structure (based on Zhuang and Lingnau, 2021). Left column: example stimuli (one out of six exemplars per subordinate action). 2nd-4th column: English labels of actions at the subordinate, basic and superordinate level. Bottom panel: model RDMs for observed actions at the subordinate (left), basic (middle), and superordinate (right) level. Each model RDM consists of a 72 x 72 matrix (12 actions, with 6 exemplars per action), where each cell in the matrix corresponds to the dissimilarity between a pair of actions. Yellow: high dissimilarity, blue: low dissimilarity.

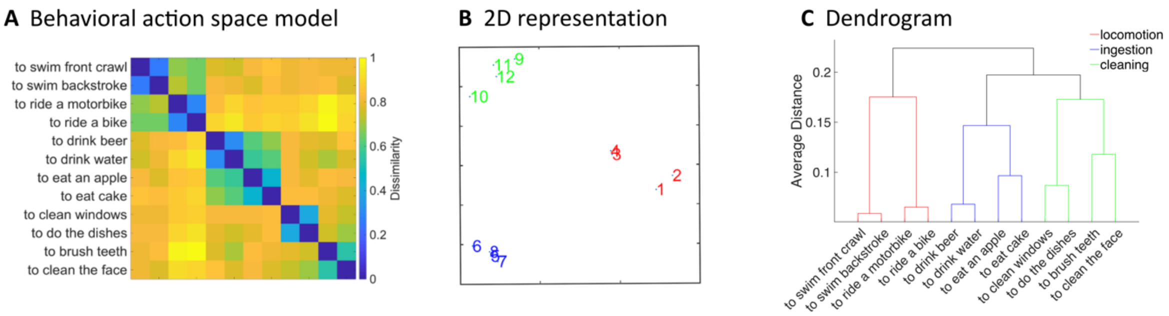


Figure 3.2. Stimulus selection and validation. A. Behavioral action space model, averaged across N=18 participants and across six exemplars per action. B. 2-dimensional representation of the results shown in panel A, resulting from multidimensional scaling analysis (Borg and Groenen, 2005). For color code see legend in panel C. The numbers refer to the twelve actions at the subordinate level (1: 'to ride a motorbike', 2: 'to ride a bike', 3: 'to swim front crawl', 4: 'to swim backstroke', 5: 'to drink beer', 6: 'to drink water', 7: 'to eat an apple', 8: 'to eat cake', 9: 'to clean windows', 10: 'to brush teeth', 11: 'to do the dishes', and 12: 'to clean the face'. C. Dendrogram resulting from hierarchical cluster analysis, confirming the clusters corresponding to the three superordinate categories (locomotion, ingestion, cleaning) and the six basic categories (see also Figure 3.1).

Participants

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

fMRI experimental design and task

We used a rapid event-related fMRI design (see Figure 3.3), programmed in ASF (Schwarzbach, 2011), adopting the design used by Tucciarelli et al., (2019). During each trial, participants were provided with a static image of an action and a central fixation cross superimposed on the image (1 s), followed by a blank screen and a central fixation cross (3 s). Participants were instructed to observe the action while keeping their eyes at fixation. During occasional catch trials (11% of all trials), participants were presented with a phrase depicting an action (e.g. ‘to swim?’, 1 s) followed by 3 s fixation. During catch trials, participants had to perform a category verification task. Specifically, they were instructed to indicate by button press with the index or middle finger of the right hand whether or not the action shown in the previous trial corresponded to the phrase shown during the catch trial. To make sure that participants were not biased towards answering questions at one of the three different levels, phrases presented during catch trials had an equal probability to address the superordinate (e.g. ‘locomotion?’), basic (e.g. ‘to swim?’) or subordinate (‘to swim breaststroke?’) level. That is, there was the same number of catch trials for each of the three taxonomic levels (four catch trials per taxonomic level in each run). Additionally, to improve design efficiency, we included null events (22.2% of all trials) that consisted of 4 s fixation and were presented pseudo-randomly (no consecutive null events and catch trials).

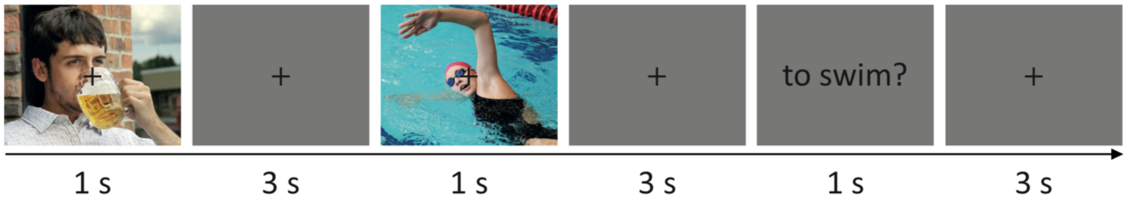


Figure 3.3. Experimental procedure used in the fMRI experiment. We used a rapid event-related design, where each trial consisted of a static image (1 s) depicting one of the twelve actions (see Table 3.1, 2nd column), followed by a fixation cross (3 s). During occasional catch trials (11 % of all trials), participants had to perform a category verification task, targeting the action presented in the previous trial. Questions to be answered during catch trials targeted the superordinate (locomotion?), basic (to swim?) or subordinate (to swim breaststroke?) level with an equal probability. Moreover, null events (22.2% of all trials) were included to enhance design efficiency (see text for details).

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

Data acquisition

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

fMRI data preprocessing

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

Region of interest (ROI) definition

ROI definition was carried out using a combination of functional data and anatomical masks obtained from the Harvard-Oxford Cortical structural and the Jülich Histological atlas (see Table 3.1). We focused on key areas of the action observation network described in previous studies (e.g., Grill-Spector and Kanwisher, 2005; Hamilton and Grafton, 2008; Kilner, 2011; Binkofski and Buxbaum, 2013; Hétu et al., 2013; Hoffman et al., 2015; Wurm et al., 2017b). Specifically, we selected the bilateral LOTC, IPL, superior parietal lobe (SPL), dorsal premotor cortex (dPM), and inferior frontal gyrus (IFG). In addition, to be able to compare results with an area concerned with low-level visual analysis, we identified bilateral V1.

ROI definition consisted of several steps. First, we computed the random-effects (RFX) general linear model (GLM) contrast 'all actions versus baseline' with spatially smoothed data (5mm FWHM). The baseline consisted of all events not explicitly modelled in the GLM. The RFX GLM included seventy-two regressors (12 actions x 6 exemplars), one regressor for the catch trials, and six regressors for head motion. The statistical map resulting from the RFX GLM contrast was corrected for multiple comparisons using Threshold-Free Cluster Enhancement (TFCE, Smith and Nichols, 2009, $p<0.05$, $|z|>1.96$, two tailed, 5000 permutations) as implemented in CoSMoMVPA (Oosterhof et al., 2016; http://cosmomvpa.org/index.html). Second, we defined anatomical masks from the Harvard-Oxford Cortical structural and the Jülich Histological atlas (threshold: 20%; see Table 3.1 for details). The mask for the LOTC was created by merging the anatomical masks for the inferior LOC and the occipito-temporal cortex. Third, within each resulting mask, we extracted the peak coordinate resulting from the RFX GLM contrast 'all actions vs baseline'. ROIs were defined as spheres (radius: 10 mm) centered around these peaks.

Table 3.1. Overview Regions of Interest.

ROIs	Peaks in MNI 152			Label	Atlas
	X	Y	Z		
V1, LH	-14	-92	-4	visual cortex (V1 BA 17L)	Jülich Histological atlas
V1, RH	12	-90	0	visual cortex (V1 BA 17R)	Jülich Histological atlas
LOTC, LH	-42	-66	0	combination of inferior lateral occipital cortex and temporooccipital part	Harvard-Oxford Cortical structural atlas
LOTC, RH	46	-74	0	combination of inferior lateral occipital cortex and temporooccipital part	Harvard-Oxford Cortical structural atlas
SPL, LH	-30	-52	52	superior parietal lobule	Harvard-Oxford Cortical structural atlas
SPL, RH	20	-56	52	superior parietal lobule	Harvard-Oxford Cortical structural atlas
IPL, LH	-58	-44	18	inferior parietal lobule (PFL)	Jülich Histological atlas
IPL, RH	62	-38	12	inferior parietal lobule (PFR)	Jülich Histological atlas
dPM, LH	-6	10	44	premotor cortex (BA6)	Jülich Histological atlas
dPM, RH	2	10	46	premotor cortex (BA6)	Jülich Histological atlas
IFG, LH	-40	10	24	inferior frontal gyrus pars opercularis	Harvard-Oxford Cortical structural atlas
IFG, RH	44	6	24	inferior frontal gyrus pars opercularis	Harvard-Oxford Cortical structural atlas

First column: abbreviations used in the following figures (LH: left hemisphere, RH: right hemisphere). Second column: peak MNI coordinates (identified based on the RFX GLM contrast 'all actions vs. baseline'; see text for details). Third and forth column: label and anatomical atlas used for ROI selection.

Representational similarity analysis

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

ROI-based RSA (partial correlations)

To examine the representation of observed actions in predefined ROIs, we used RSA with partial correlations (Kriegeskorte et al., 2008). First, we created a model RDM for each of the three taxonomic levels (Figure 3.1, bottom panel), corresponding to the hierarchical structure shown in Figure 3.1 (top panel; see also Zhuang and Lingnau, 2021). Since we used 6 exemplars for each of the 12 actions, each model RDM consisted of a 72 x 72 matrix, where each cell in the matrix corresponds to the dissimilarity between a pair of actions. The subordinate level model consists of twelve clusters along the diagonal, with each cluster comprising six exemplars of the same type of action. The basic level model consists of six clusters along the diagonal, while the superordinate model consists of three clusters along the diagonal.

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

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

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

Searchlight-based RSA (multiple regression)

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

To examine risks of multicollinearity, we determined the Variance Inflation Factor (VIF) for the three taxonomic models depicted in Figure 3.1 (bottom panel), the low-level visual control model and the scene control model. We obtained small to moderate VIFs (subordinate model: 1.90; basic model: 2.55; superordinate model: 1.78; low-level visual control model: 1.31, scene control model: 1.77), suggesting a low risk of multicollinearity (Mason et al., 2003).

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

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

Finally, in order to examine the relationship between neural RDMs and the behavioral dissimilarity matrix obtained from the multi-arrangement task (see Figure 3.2B), we carried out another searchlight-based multiple regression RSA in which we included (a) the behavioral model, the low-level visual control model and the scene control model, and (b) the behavioral model, each of the taxonomic models separately, the low-level visual control model and the scene control model. VIFs were also computed and the results were small to moderate, justifying the use of multiple regression RSA. In particular, the VIFs for the low-level visual control model, the scene control model and the behavioral model were small (low-level visual control model, scene control model and behavioral action space model: VIF=1.31, 1.85, 1.53, respectively). Additionally, the VIFs for regression models including the behavioral action space model, each of the taxonomic level models, the low visual control model and the scene control model were small to moderate (behavioral action space model, subordinate level model, low-level visual control model and scene control model: VIF= 3.84, 3.18, 1.33, 1.88, respectively; behavioral action space model, basic level model, low-level visual control model and scene control model: VIF=6.10, 5.34, 1.32, 1.88, respectively; behavioral action space model, superordinate level model, low-level visual control model and scene control model: VIF=2.63, 2.24, 1.31, 1.87, respectively).

Conjunction analysis

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

Statistics

For the ROI-based partial correlation analysis, we first used one-sample t-tests to compare the Fisher-transformed partial correlation values against zero. Second, to determine whether the similarity between neural and model RDMs depicted in Figure 3.4 was modulated by the taxonomic level and ROI, we carried out a two-way repeated measures ANOVA with the factors *taxonomic level* (superordinate, basic, subordinate) and *ROI* (see Table 3.1), followed by pairwise comparisons. Corrections for multiple comparisons for all factorial combinations (taxonomic level x ROI) was carried out using the false discovery rate (FDR; Benjamini and Hochberg, 1995). Since the assumption of sphericity was violated for the 2-way ANOVA, we report Greenhouse-Geisser-corrected p values (indicated as p_{GG}).

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

3.4 Results

Behavioral results

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

ROI-based results

Figure 3.4 shows the results from ROI-based RSA for the three taxonomic levels, regressing out low-level visual properties by means of a control model obtained from the first convolutional layer of a Deep Neural Network (ResNet50_conv1; see section ROI-based RSA for details) and the scene control model (based on the last second layer of ResNet50 pretrained with places; see section ROI-based RSA for details). We obtained significant partial correlations between neural RDMs and all three taxonomic level models in most of the examined ROIs, with the exception of the right IFG, where no significant partial correlation was observed between neural RDMs and the subordinate level. Similarities (partial correlation values) between neural RDMs and model RDMs differed between the three taxonomic levels in bilateral LOTC and SPL. In bilateral LOTC, we obtained the highest partial correlations for actions at the basic level in comparison to the other two taxonomic levels. In bilateral SPL, we observed higher partial correlations between the neural RDM and the basic level model in comparison to the superordinate level model.

These observations are supported by the corresponding statistics. In particular, a two-way (taxonomic level x ROI) repeated measures ANOVA revealed a significant main effect of taxonomic level ($F_{(2,44)} = 8.88$, $p_{GG} = .002$, partial $\eta^2 = .29$) and ROI ($F_{(11, 242)} = 32.93$, $p_{GG} < .001$, partial $\eta^2 = .60$), and a significant interaction between taxonomic level and ROI ($F_{(22, 484)} = 3.46$, $p_{GG} < .001$, partial $\eta^2 = .14$). Pairwise comparisons within each ROI (using FDR to correct for multiple comparisons) revealed that neural RDMs in the bilateral LOTC showed the highest similarity with the basic level model in comparison to the other two models (left: basic vs subordinate level, $t_{(1,22)} =6.11$, $q_{FDR} < .001$; basic vs superordinate level, $t_{(1,22)} = 3.57$, $q_{FDR} =. 003$; right: basic vs subordinate level, $t_{(1,22)} = 5.04$, $q_{FDR} < .001$; basic vs superordinate level, $t_{(1,22)} = 4.04$, $q_{FDR}< .001$). Additionally, neural RDMs in bilateral SPL showed a higher partial correlation with the basic level model in comparison to the superordinate level (left:

basic vs superordinate level, $t_{(1,22)} = 3.53$, $q_{FDR} = .006$; right: basic vs superordinate level, $t_{(1,22)} = 3.86$, $q_{FDR} = .003$).

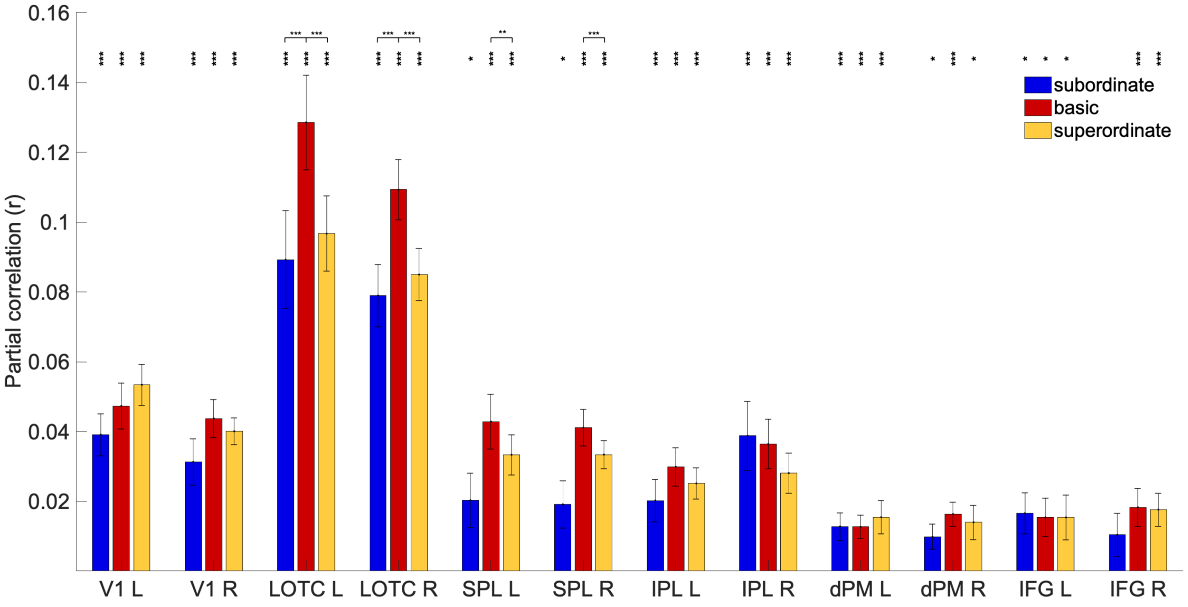


Figure 3.4. Partial correlations between neural RDMs and the models corresponding to the three taxonomic levels (see also Figure 3.1), regressing out a low-level visual control model, based on the first convolutional layer of a deep neural network (ResNet50, He et al., 2015), trained on a large scale image data set (ImageNet dataset) to distinguish image classes and a scene control model, based on the second-to last layer of ResNet50, trained to distinguish between different scene categories; Zhou et al., 2017 and 2018; see Section ROI-based RSA for details). Error bars show the standard error of the mean. Asterisks illustrate statistical significance (*: qFDR < 0.05, **: qFDR < 0.01, ***: qFDR < 0.005) of one-sample t-tests against zero with FDR correction for all possible combinations of taxonomic level and ROI. Since a 2-factorial (level x ROI) repeated measures ANOVA revealed a significant interaction between taxonomic level and ROI, partial correlation values between the three taxonomic levels within each ROI were compared using pairwise comparisons (corrected for multiple comparisons using FDR).

To illustrate the results obtained in bilateral LOTC in comparison to bilateral early visual cortex (V1), Figure 3.5 visualizes neural RDMs (left panel) and the corresponding 2D-arrangements obtained from multidimensional scaling analysis (right panel) in bilateral LOTC (top row) and bilateral V1 (bottom row). Using the same approach described for the searchlight-based RSA, we created a 72 X 72 neural RDM in LOTC and V1, and then collapsed the neural RDM across hemispheres and participants. Next, to account for low-level visual features and scene-related properties, we extracted values from the lower triangular part of the neural RDM and regressed out the corresponding lower triangular part of the low-level visual control model and the scene control model. Finally, the resulting residuals were rescaled to values from 0 to 100 (Nili et al., 2014). In the 2D visualization of the MDS (Figure 3.5B, D), the three superordinate action categories are highlighted in different colors (blue, red, and green), the basic action categories within a superordinate category are shown by open and filled symbols, and different symbols (circles, squares) indicate the different actions at the subordinate level within each basic level category. As can be seen from Figure 3.5, the neural RDM in bilateral LOTC (Figure 3.5A) and the corresponding 2-D arrangement resulting from multidimensional scaling (Figure 3.5B) reveals a broad distinction into actions at the basic level, with open symbols mostly on the right side and filled symbols mostly on the left side, and an additional broad clustering into the three different superordinate categories highlighted in red, green and blue (Figure 3.5B). By contrast, this pattern is less obvious in early visual cortex (Figure 3.5C, D), in line with the results of the ROI-based analysis shown in Figure 3.4.

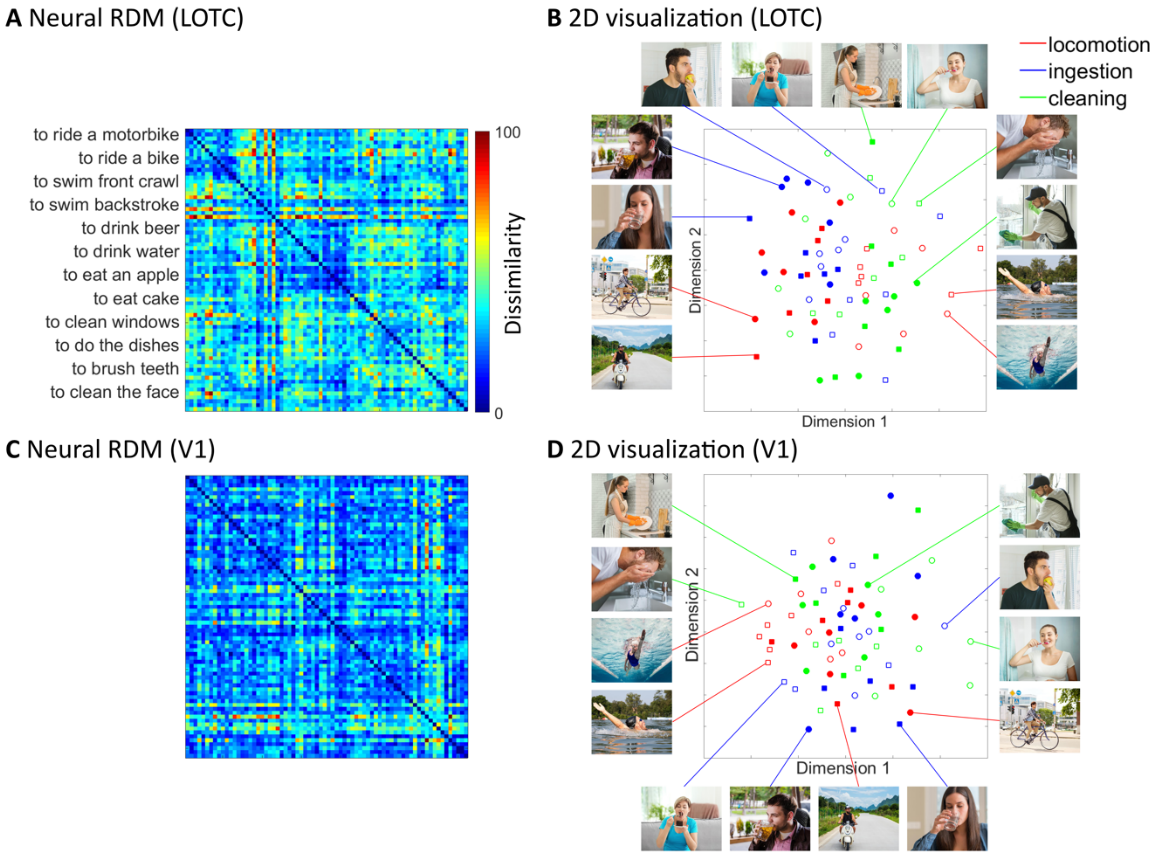


Figure 3.5. Neural RDM (left panel) and the corresponding 2D visualization of the MDS results (right panel) obtained in bilateral LOTC (A and B) and bilateral V1 (C and D) after regressing out the low-level visual control model (based on the first convolutional layer of ResNet50, trained to distinguish image classes) and the scene control model (based on the second-to last layer of ResNet50, trained to distinguish between scene categories; Zhou et al., 2017 and 2018; see Section ROI-based RSA for details), using the same ROIs as those shown in Figure 3.4. Colors to indicate actions belonging to the three superordinate level categories (red-locomotion, blue-ingestion, and green-cleaning) are the same as in Figure 3.2. Opened versus filled symbols indicate different actions at the basic level, whereas circles versus squares are used to distinguish actions at the subordinate level.

Whole-brain searchlight results

Figure 3.6A reveals areas that capture the similarity between actions at the three different taxonomic levels when low-level visual features and scene-related properties are regressed out via a low-level visual control model and a scene control model (see Section *ROI-based RSA* for details). This analysis reveals a wide set of regions for the subordinate and basic level, with peaks in the right fusiform cortex (Table 3.2). By contrast, the representation of observed actions at the superordinate level is restricted to a more circumscribed region in occipitotemporal cortex extending into the anterior IPL, with a peak in right occipital pole (Table 3.2).

Figure 3.6B shows which areas capture the similarity structure of observed actions that is unique to each of the taxonomic levels (accounting for the other two taxonomic models as well as for the low-level visual features and the scene control model). This analysis reveals that the dissimilarity structure for actions that is unique to the subordinate level is restricted to bilateral occipito-temporal cortex, with a peak in right fusiform cortex, while information that is unique to the basic level is associated with bilateral occipito-temporal cortex (OTC), IPL and SPL, with a peak in the right occipital fusiform cortex (see Table 3.2). By contrast, we obtained no region that represented information that is unique to the superordinate level. We will return to this observation in the discussion.

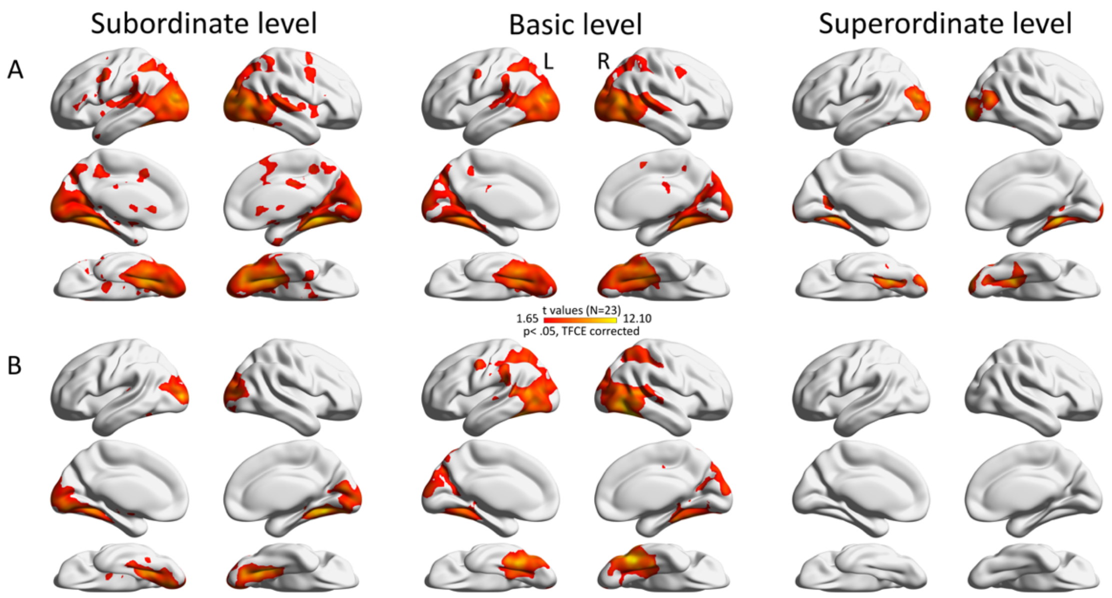


Figure 3.6. Results of the searchlight-based multiple regression RSA for the subordinate (left panel), basic (middle panel) and superordinate (right panel) model. **A**: Multiple regression RSA, regressing out the low-level visual control model and the scene control model (see section Searchlight-based RSA for details). **B**: Multiple regression RSA, regressing out the low-level visual control model, the scene control model and the remaining two taxonomic models. For group statistics (N=23), beta weights resulting from the multiple regression RSA were

entered into one-sample t-tests against zero. The statistical t-value maps were corrected for multiple comparisons using TFCE (z>1.65, TFCE-corrected p <.05, one-tailed) as implemented in CoSMoMVPA (Oosterhof et al., 2016).

Table 3.2. Peak locations for the three taxonomic levels in the whole-brain searchlight

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

To identify regions in which actions are jointly represented at the subordinate and basic level, we computed a conjunction analysis on the basis of the statistical maps corresponding to the unique representation of the subordinate and basic level shown in Figure 3.6B (see Methods section for details). The results are shown in Figure 3.7. The convergence of observed actions at the subordinate and basic level was located in bilateral OTC.

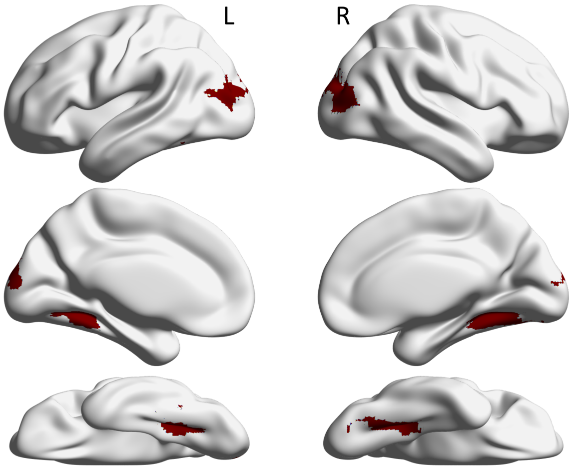


Figure 3.7. Conjunction map for the unique representation of observed actions at the subordinate and basic level (see Figure 3.6B). See text for details.

Relationship between neural RDMs and behavioral dissimilarity structure

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

First, we ran a multiple-regression RSA between neural RDMs and the behavioral dissimilarity structure shown in Figure 3.2A, regressing out the low-level visual control model and the scene control model. As shown in Figure 3.8A, this analysis revealed large clusters in bilateral occipitotemporal cortex, IPL, and SPL that were associated with the behavioral action space model.

Second, separately for each of the taxonomic levels, we carried out another multiple-regression RSA in which we included the behavioral action space model, the low-level visual control model, the scene control model and the respective taxonomic level (e.g., the subordinate level). A map of the beta weights corresponding to the behavioral action space model, after regressing out the low-level visual control model, the scene control model and each of the taxonomic level models are shown in Figure 3.8B-D. As can be seen, after regressing out the subordinate level, the low-level visual control model and scene control model, the behavioral action space model is associated with neural patterns in bilateral occipitotemporal cortex (Figure 3.8B). After regressing out the superordinate model (Figure 3.8D), the low-level visual control model and the scene control model, large regions of bilateral occipitotemporal cortex, IPL and SPL remained significantly associated with the behavioral action space model. By contrast, after regressing out the basic level model, the low-level visual control model and the scene control model, a small cluster in the temporal occipital fusiform cortex (bilaterally) and left superior LOC remained that captured the behavioral action space model (Figure 3.8C). Whereas these latter results need to be interpreted with caution, they are in line with the view that the cortical representations of the behavioral action space model best reflect information at the basic level (see also Figure 3.4 and 5). We will return to this observation in the discussion.

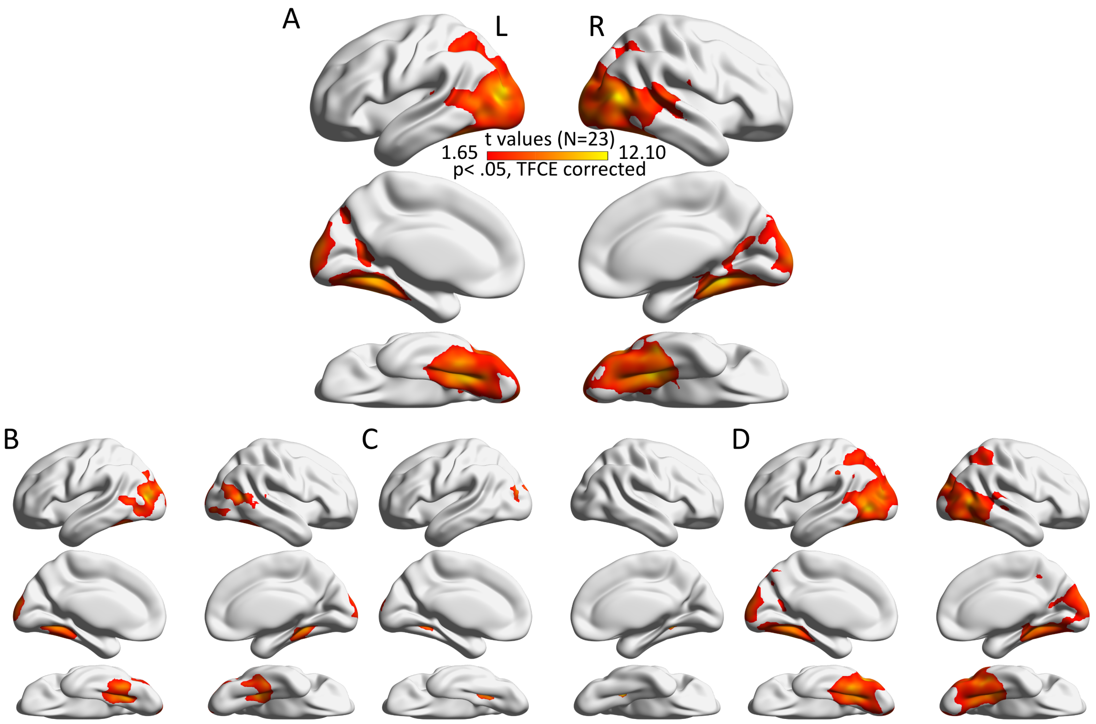


Figure 3.8. Brain areas capturing the representational space of actions obtained from behavioral ratings (see Figure 3.2A). A: Brain areas capturing the behavioral dissimilarity structure after regressing out the low-level visual control model and the scene control model. B-D: Brain areas capturing the behavioral dissimilarity structure after regressing out the low-level visual control model, the scene control model and models for the (B) subordinate, (C) basic and (D) superordinate level.

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

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

3.5 Discussion

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

The representational space of observed actions in the LOTC

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

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

Tucciarelli et al., 2019) at varying levels of abstraction (see also Wurm and Lingnau, 2015; Wurm and Caramazza, 2021).

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

Previous behavioral studies suggested that the basic level holds the maximized information for objects (Rosch et al., 1976; Murphy, 2004) and actions (Zhuang and Lingnau, 2022). The results of the current study provide insights into the neural basis underlying these behavioral findings. First, we found that patterns of activation in bilateral LOTC and SPL showed a higher similarity with the basic level model in comparison to the subordinate and the superordinate level (Figure 3.4). Second, information that is unique to the basic level was captured in more widespread regions, including OTC, IPL and SPL (Figure 3.6B), in comparison to information that is unique to the subordinate level (Figure 3.6A), while we did not obtain any area that captured information that is unique to the superordinate level. Third, the behavioral action space model showed the highest similarity with neural patterns capturing the basic level model (see Figure 8C). Together, the results of the current study on observed actions are consistent with results by Iordan et al. (2015) demonstrating that objects at the basic level show the strongest similarity with patterns of activation in the LOC.

The representation of actions at the superordinate level

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

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

Since we did not compare the processing of visually presented actions with the processing of the corresponding action verbs or phrases, we cannot clearly separate between the contribution of higher-level visual features (or perceptual action precursors, see Wurm and Caramazza, 2021) and semantics. As pointed out above, a study that focused on this direct comparison revealed shared representations between patterns of activations for different actions depicted as videos and as written sentences found this to be the case exclusively in the lateral posterior temporal cortex (Wurm and Caramazza, 2019). That said, to be able to tolerate varying degrees of variability, we assume that the categorization of visually presented actions requires access to semantic knowledge for all three taxonomic levels, while the amount of variability is likely to differ between the levels. Specifically, members of the same subordinate level are assumed to have a higher number of shared features that can be exploited to categorize actions at the subordinate level (e.g. the body posture to distinguish between front crawl and backstroke; see also Zhuang and Lingnau, 2022). At the basic level, which is considered to be the most informative level (see e.g. Rosch et al., 1976), categorization has been proposed to rely on more common features that are shared across members of a category (e.g. applying some liquid to the body with hands or a tool to categorize an image to belong to the basic level category ‘cleaning the body’). Finally, to distinguish actions at the superordinate level, an even higher degree of generalization across high-level visual features is required (e.g. bringing an object to the mouth for the category ingestion).

Comparison of object and action representations

There exists a long tradition both in human (see e.g. Bach et al., 2014, Wurm et al., 2017b; Livi et al., 2019; Wurm and Caramazza, 2021) and monkey studies (e.g. Bonini et al., 2014) demonstrating that objects provide strong clues regarding the actions they afford, and that objects play an important role during action observation. Consequently, the representation of objects and actions have been proposed to follow similar, though not identical, principles of organization (see e.g. Pillon and d'Honincthun, 2011; Wurm et al., 2017b). As an example, Wurm, Caramazza, and Lingnau (2017) demonstrated that features corresponding to actions directed towards persons and objects are preferentially represented in dorsal and ventral portions of the LOTC, parallel to the organization of information related to persons versus inanimate objects, respectively. Given these assumed similar principles of organizations, it is notoriously difficult to dissociate the representation of actions and objects. That said, Tucciarelli et al. (2019) showed that patterns of activation in the LOTC capture the behavioral similarity structure of actions, over and above variability due to action components such as objects involved in the action. Likewise, several previous studies demonstrated representations of observed actions in the LOTC that generalized across the object (Wurm and Lingnau, 2015; Wurm et al., 2016). More specifically, Wurm et al. (2016) demonstrated that it is possible to distinguish between observed opening and closing actions, irrespective of the object, in the LOTC, IPL and PMv, whereas decoding between objects across actions was restricted to clusters in the ventral stream (Wurm et al., 2016, Figures 3, 4 and Supplementary Figure 5). While we cannot clearly dissociate between the representation of action functions and object states/ object affordances in the current study, these previous studies indicate that action representations in the LOTC can be dissociated from representations of objects that are being manipulated.

The impact of scene-related information

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

Limitations and future directions

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

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

Conclusion

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

4 Chapter 4: Study 3

‘Neural dynamics of hierarchical action representations’

Chapter 4 is formatted for the eLife journal and includes the third study, which is currently being prepared for submission.

4.1 Abstract

Previous studies indicate that observed actions are represented at different hierarchical levels in both lateral occipitotemporal and parietal areas. However, the corresponding timing of these representations is not well understood. To address this question, we used time-resolved representational similarity analysis (RSA) of EEG data of human participants presented with static images depicting twelve different actions. We established models corresponding to three different taxonomic levels, i.e., the subordinate (e.g., swimming backstroke), basic (e.g., swimming), and superordinate (e.g., locomotion) level. Time-resolved RSA showed the highest peak latencies around 170 ms for actions at all three taxonomic models. Furthermore, EEG-fMRI fusion revealed that action representations obtained in the lateral occipitotemporal cortex (LOTC) showed the highest similarity with the EEG data around 220 ms, marking later time point compared to bilateral V1. Taken together, our findings indicate actions across three taxonomic levels occur simultaneously, with the LOTC showing a later response compared to V1. This contributes to a deeper comprehension of the spatiotemporal aspects of hierarchical organization of action representations.

Keywords: action, dynamic, subordinate, basic, superordinate level, RSA.

4.2 Introduction

Our ability to quickly and effectively recognize and categorize others' actions is remarkable. We can solve this task at different levels of abstraction, ranging from very concrete actions, such as recognizing someone swimming backstroke, to more general concepts, such as swimming, and even broader ones, such as locomotion. What remains unclear is how neural hierarchical organization of actions evolves in time.

In the past, using behavioral experiments, researchers have observed differences in human performance when responding to actions at different taxonomic levels (de la Rosa et al., 2014; Zhuang & Lingnau, 2022). Utilizing methods like using auditory stimuli as a prime cue to visual stimuli or executing a rapid category verification task, participants can be directed to focus on the processing of an action at a specific taxonomic level, such as the basic level, like 'to swim'. de la Rosa et al., (2014) used the rapid category verification task on social interaction actions (e.g., arguing, shaking hands). Participants were presented with a label representing an action that belonged to one level (e.g., ‘shaking hands’). Next, participants were asked to respond quickly to identify whether the following images depicting different actions with various durations matched the action label. They found that participants responded faster and more accurately when actions were preceded by a cue at the basic level in comparison to a cue at the subordinate level. Utilizing both priming and rapid category verification, Zhuang and Lingnau (2022) found that actions at the subordinate and basic levels were recognized faster in comparison to the superordinate level, with no significant difference between the subordinate and basic levels. These findings from two studies showed some inconsistency in action recognition between actions at the subordinate and basic levels. It is worth to investigating the neural basis for temporal recognition of actions at the subordinate and basic levels.

Indeed, there are relatively few studies that have examined the temporal evolution of neural responses to observed actions. Earlier studies using event-related potentials (ERPs) of EEG data, found that N400 was associated with action-related comprehension, such as responses to the incongruent-actions (Proverbio & Riva, 2009). Utilizing MEG in combination with multivariate pattern analysis, Tucciarelli et al., (2015) found that the lateral occipitotemporal cortex (LOTC) has the access to abstract action (i.e., to generalize across effectors and reach direction) representations earlier (roughly from 150 ms until 550 ms) in comparison with the precentral regions (approximately from 550 ms to 1200 ms). Another MEG study by Isik et al. (2018) examined the dynamics of invariant action representations, which refer to neural encodings of actions that remain consistent and robust regardless of variations in factors such as viewpoint. They found that action categories are able be classified as early as 200 ms. More recently, Dima et al. (2022) used EEG in combination with time-resolved representational similarity analysis (RSA) and a large dataset of naturalistic stimuli to investigate how the representation of social action features, such as visual, social-affective, and action-related information, unfolds in time. Dima et al., (2022) observed a sequential representation pattern where visual, action-related, and social-affective features were processed from early to late stages. To sum up, these findings revealed diverse patterns of processing times and brain regions involved in action comprehension and representation, encompassing early access to abstract action representations and sequential processing of various features.

Neuroimaging studies have revealed different brain regions that capture the representation of actions at different levels of abstraction (Grafton & Hamilton, 2007; Hamilton & Grafton, 2006, 2008; Spunt et al., 2016; Wurm & Lingnau, 2015; Zhuang et al., 2023). For example, Spunt et al. (2016) varied the type of question that participants had to respond to, focusing either on how to perform the action (kinematic level; low level of abstraction) or why to do the action (goal level; high level of abstraction). The authors reported that the medial prefrontal cortex, anterior superior temporal sulcus, temporal pole and precuneus were significantly associated with a higher level of abstraction, whereas the posterior medial temporal gyrus, rostral inferior parietal lobe and inferior frontal gyrus were associated with lower level of abstraction during the processing of observed actions. Hamilton and Grafton (2006, 2008) proposed that actions can be organized into three hierarchical levels: the muscle level, the kinematic level and the goal level. The authors suggested that the muscle level pertains the muscles of the arm and hand, involving the activity of over thirty hand. They found that the observed kinematic patterns are associated with the lateral occipital cortex (LOC) and the goal information is represented in anterior intraparietal sulcus. Additionally, Wurm and Angelika (2015) investigated action-specific representations, ranging from concrete actions (e.g., to open vs. to close a specific bottle) to abstract action concepts (e.g., to open an object). The authors observed that actions at the concrete levels were represented in the LOTC, IPL, and ventral premotor cortex (PMv), while the abstract-level actions were represented in LOTC and IPL. Additionally, McMahon et al., (2023) examined the social action features along the lateral visual pathway. They found hierarchical representations of social actions, with the low-level visual features encoded in early visual cortex (EVC), mid-level visual-social features in extrastriate body area (EBA), and high-level social interaction information in superior temporal sulcus (STS). Using RSA of fMRI data, we recently demonstrated that it is possible to decode between different observed actions at all three taxonomic levels in the LOTC, with a preference for the basic level (Zhuang et al., 2023).

Taken together, the studies collectively demonstrate that humans recognize actions in a hierarchical manner. While a growing number of studies revealed specific brain regions involved in the processing of observed actions, the temporal evolution underlying the representation of actions at different levels of abstraction is not well understood. The current study aimed to shed light on this question. Specifically, we addressed the following questions: 1) At what time are observed actions represented at different taxonomic levels? 2) When do action exemplars group together to form categories in our minds, allowing us to explore the neural mechanisms involved in categorization? 3) what is the time-point representations of observed actions in the low and high visual cortex using fMRI-EEG fusion? To address these questions, we used high temporal resolution EEG to record brain activities during participants viewing action images, and then performed RSA and pairwise-exemplar decoding to capture representational structure of observed actions across three taxonomic levels.

4.3 Results

Behavioral results. During catch trials, participants were instructed to determine if the previous image aligned with the catch trial label. As a result, 22 participants (23.5 ± 4.7 years old, female: 20) performed well with the accuracy of 0.84 ± 0.09 and the reaction time of 833.1 ± 73.8 ms.

RSA results

Standard RSA. To determine how action representations at different taxonomic levels are represented unfold in time, we conducted standard RSA, separately for each of the three taxonomic level RDMs, separately for each time point. In addition, to examine the time course underlying the representation of the behaviorally established action space, we computed the correlation between EEG data and the behavioral RDM, separately for each time point. As can be seen from Figure 4.1A, the correlation between neural data and all three taxonomic level models peaked around 174 ms and 235 ms, with the significant onset at 90-100 ms. The behavioral model showed the highest similarities with the EEG RDM, reflecting the strongest representations for the behavioral model, followed by the subordinate level.

Multiple-regression RSA. To account for differences between different actions due to low visual features and features related to scene and background information, we performed multiple-regression RSA, regressing out a visual control RDM and a scene control RDM (see Methods for details). As shown in Figure 4.1 B, we found that after regressing out two control models, the highest correlation with EEG-RDM is still the behavioral model followed by subordinate level and represented at the same time window (~172 ms) as RSA results, in line with standard RSA results.

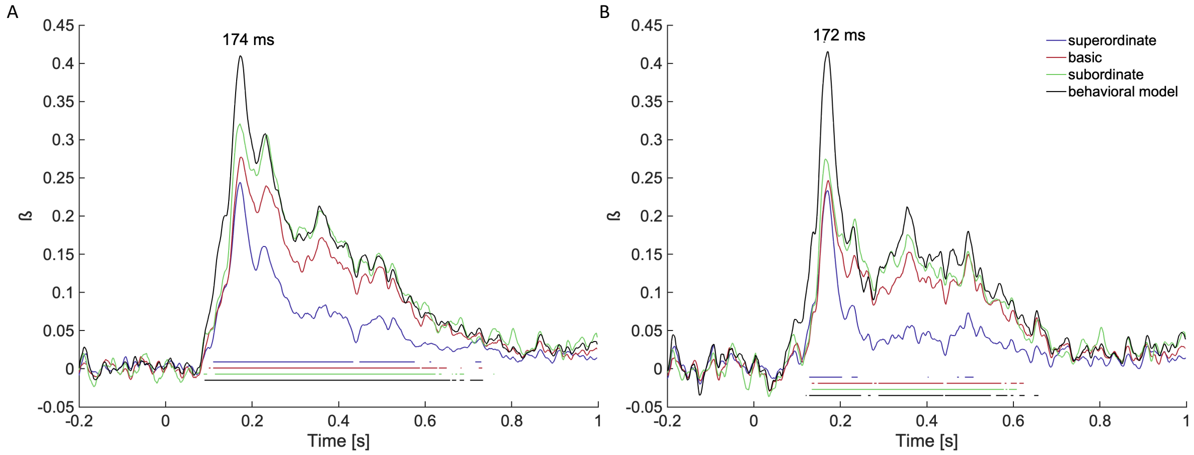


Figure 4.1. Standard RSA and multiple-regression RSA results. A. Standard RSA results between EEG-based RDMs and three taxonomic level models respectively. B. multiple-regression RSA of all models after regressing out the visual control model and the scene control model. Blue: superordinate model; red: basic model, green: subordinate model, black: behavioral model. The horizontal lines reflected the onset significance.

Pairwise-exemplar decoding and Multi-dimensional scaling (MDS)

To explore the neural mechanisms involved in categorization, we first established EEG-based Representational Dissimilarity Matrices (RDMs) and then visualized MDS results. To create EEG-based RDMs, we conducted pairwise exemplar decoding over time. Furthermore, to understand the neural temporal processes underlying categorization, we determined when action exemplars come together to form category-based clusters. We used MDS to visualize EEG-based RDMs across time (Figure 4.2). In the early time points, all exemplars are close to each other (Figure 4.2A), which is consistent with MDS results of Carlson et al., (2013). Then, the dots of MDS started to spread out from 90-100 ms (Figure 4.2B), which corresponds to the time point of decodability of exemplars. This finding is in line with our RSA, multiple-regression analysis, which showed the significant time points of three levels start from around

90 ms to 100 ms. Then, we obtained a very clear MDS from 170-180 ms (Figure 4.2C). The same color dots were clustered together, while different color dots were far away from each other. These results are consistent with the peak latency results determined by RSA analysis, suggesting that these peaks reflect the strongest decodability of action recognition. These clear MDSs showed again from 230 ms to 240 ms (Figure 2D), which corresponding to the second peak latency obtained from RSA.

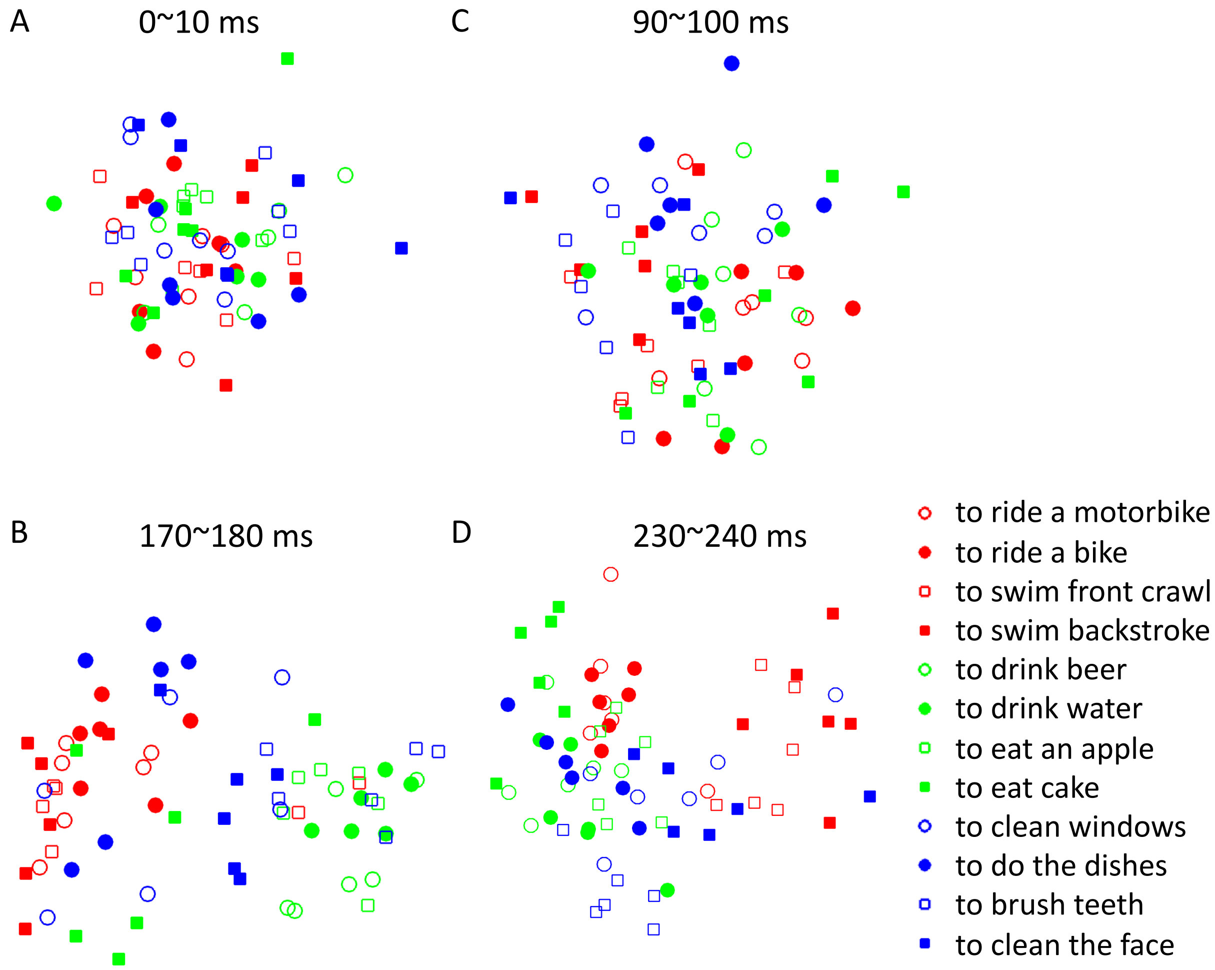


Figure 4.2. MDS results from 0-240 ms indicate the time span from before exemplars can be decoded until to the point of the best decodability at the first peak. The colors represent three action categories at the superordinate level (e.g., ‘locomotion’). The squares and dots represent two basic action categories at the three superordinate levels (e.g., ‘to ride’ and ‘to swim’ for upper-level - ‘locomotion’). The filled and unfilled dots represent two subordinate action categories at six basic levels (e.g., ‘to ride motorbike’ and ‘to ride bike’ for upper-level - ‘to ride’).

Spatiotemporal representation with EEG-fMRI fusion

To provide an integrative view of both spatial and temporal dynamics representations of observed actions across three taxonomic levels, we combined fMRI results with EEG findings. From our previous fMRI results, we found that occipitotemporal cortex was able to decode actions at all three levels. Therefore, using bilateral LOTC as ROIs, we compared RDMs of bilateral LOTC and EEG-based RDMs across time. Additionally, as a comparison to LOTC, we used bilateral V1 as ROIs to correlate with EEG-based RDM over time. As shown in Figure 4.3 A, B, we observed the most prominent peak of the correlation between the fMRI-based RDMs obtained in bilateral LOTC and the EEG-based RDM around 230 ms, with a second peak at 100 ms (left) and 168 ms. The correlation between neural RDMs obtained in bilateral V1 and EEG-based RDMs was overall weaker, without any prominent peak (Figure 4.3 C, D). These findings indicate a whole picture of the spatiotemporal representations of action understanding.

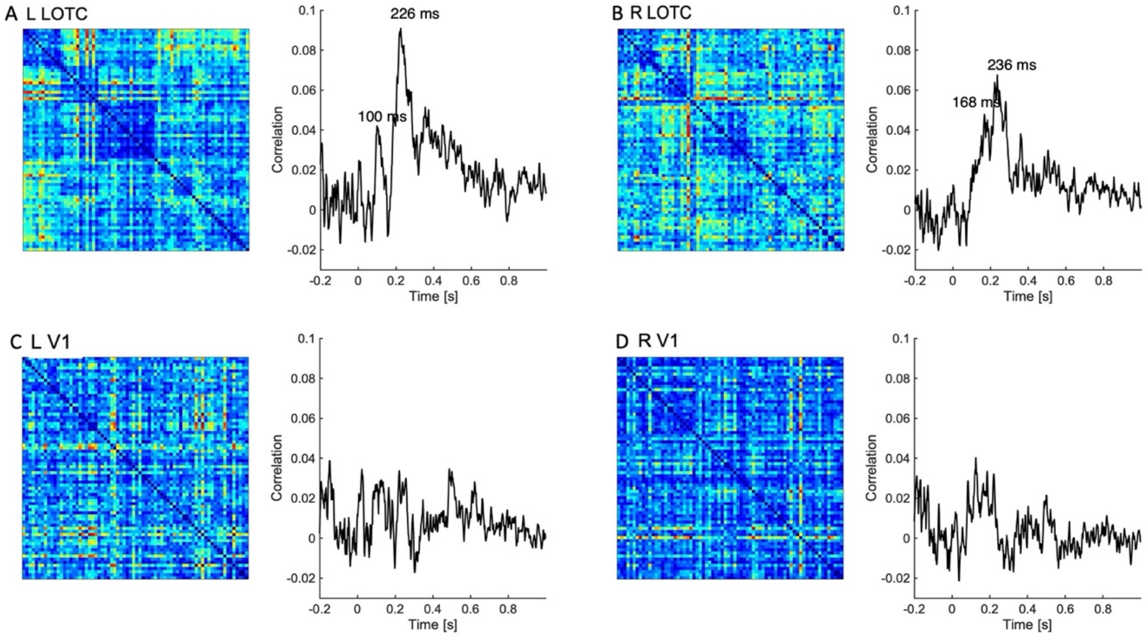


Figure 4.3. Neural RDMs and correlation between fMRI- and EEG-based RDMs. A, B: The similarity between the RDMs obtained in bilateral LOTC and the EEG-based RDMs was highest at around 230 ms, with the second peak at around 100 ms (left) and 168 ms (right). C, D: In comparison, the similarity between RDMs obtained in bilateral V1 and EEG-based RDMs were lower and more noise, with no consistent peak.

4.4 Discussion

In this study, we aimed to examine how representations of observed actions at different taxonomic levels evolve in time. To address this question, we used time-resolved RSA of EEG data. To be able to relate the temporal evolution of the dissimilarity structures to specific brain regions implicated in the processing of observed actions, we used RSA-based EEG-fMRI fusion (REFS).' We observed an above-chance correlation between the three taxonomic level models and the EEG-based neural data from about 90-100 ms onwards, with peaks around 170 and 235 ms. Neither the onset nor the peak latencies differed between the three taxonomic levels. Furthermore, EEG-fMRI fusion with data from the LOTC showed the higest latency peak at around 230 ms, with a second peak at 100 ms in the left hemisphere and 168 ms in the right hemisphere, along with peaks and onset time of RSA findings. These results provide a coherent complete picture of action representations in temporal and spatial domains, revealing the different taxonomic levels of actions are processed at the same time.

The time course of action representations across three taxonomical levels

We observed a significant onset around 90-100 ms, indicating that actions at all three taxonomic levels can be represented in neural activity as early as 100 ms after stimulus onset. This onset latency signifies the earliest point at which these action levels are encoded in neural representations. These results are in line with observed onset of the object levels (ranging from 80 to 100 ms; Carlson et al., 2013). Furthermore, we identified a peak latency of 170 ms for actions at all three taxonomic levels, suggesting that this is the optimal time to capture action representations across these levels. Notably, these peak latencies for actions are somewhat consistent with the peak latencies observed for object processing (ranging from 120 to 240 ms). The key difference lies in the timing sequence, as peak latencies for object processing occur in a sequence from lower-tier to higher-tier representations (Carlson et al., 2013). Further discussion will be found in the section of the representations between object levels and action hierarchies.

Additionally, comparing our current EEG findings with previous behavioral studies reveals differences, but they are explainable. In the rapid category verification task conducted by de la Rosa et al. (2014), participants were tasked with making judgments about action images based on cues presented at the beginning. Their findings indicated that actions were recognized more quickly at the basic level compared to the subordinate level. Zhuang and Lingnau (2022) observed faster responses at the subordinate and basic level in comparison to the superordinate level, but no significant difference between the subordinate and the basic level. In the current study, we obtained similar onsets and peak latencies for the correlation between the three taxonomic level models and neural data. This absence of an effect needs to be interpreted with caution. We reason that two factors might have contributed to this observation. One possible reason for not obtaining differences between the three taxonomic levels could be that we did not vary the task, as was done in the studies by de la Rosa et al. (2014) and Zhuang and Lingnau (2022). Instead, since we did not want to bias participants to focus on one specific taxonomic level, we instructed participants to attentively process the images, and to respond to occasional catch trials. Catch trials referred to the image presented at the previous trial and had an equal probability to require a judgment at the subordinate, basic and superordinate level. In the absence of an explicit task regarding the specific taxonomic level, we obtained that the neural representations at the three taxonomic levels occur simultaneously. Another reason might be that there is a difference, but we did not obtain it due to limited power in the current study. Further experiments are required to distinguish between these alternatives.

Furthermore, the temporal intervals corresponding to the two peaks of action representations identified in our study are consistent with the findings of Tucciarelli et al. (2015), who reported abstract action representations between 150-550 ms, and Isik et al. (2018), who described the dynamics of invariant action decoding at 200 ms. Further studies might investigate which kinds of information is important to which level of action understanding, and these processes occur at what time points.

Resolving hierarchical action representations in time and space

By combining high-temporal resolution-EEG and high-spatial resolution-fMRI, our EEG-fMRI fusion results provide a spatial-temporal picture of hierarchical action representations. We were able to measure representations within V1, associated with low-level visual features, and LOTC, associated with more complex visual features. The representations of V1 associated with EEG neural signatures showed a bit noise results, which might be due to small number of participants. While the structural representations of LOTC associated with EEG-based RDM showed the highest peaks at the 230 ms, which is line with the second peak of standard RSA results.

Additionally, Dima et al. (2022) found the sequential processing of visual, social-affective, and action-related information. Notably, our classification of actions at the subordinate, basic, and superordinate levels is based on the abstractness of the concepts, whereas Dima et al. (2022) classified actions based on feature-level characteristics, marking a difference from our current approach. Their findings align with our results in terms of the sequence from V1 (low-level features) to LOTC (high-level features).

The representations between object levels and action hierarchies

Our RSA results and MDS visualizations showed that the onset significant occurred early (before 90-100 ms) and two dominate peaks appeared later, in line with the corresponding representations of object levels (Carlson et al., 2013). While a slight difference between object and actions across these different levels is that the first peak of object representations showed a bit earlier (first onset significant: 80ms, first peak: 120ms) in comparison with action representations. However, a major distinction between object and action categorization lies in the sequential process observed in object levels, where the concrete level is processed first and the abstract levels tend to be processed later. In contrast, our current study demonstrated that actions across different levels were represented simultaneously. One possible reason for the observed differences is the use of distinct analysis methods in these two studies. Carlson et al., (2013) used a classification method to determine the time point at which objects can be distinguished. However, this approach blended visual information, potentially affecting the outcomes, particularly given the variations in distinctions between objects at different levels. In contrast, our research conducted multiple-RSA analysis, which can remove the effect of low-level visual models and scene information. Additionally, it is plausible that action categorization undergoes different processes compared to object categorization, but this requires further investigation.

Additionally, using EEG-fMRI fusion analysis, we observed that V1 was earlier associated with action representations in comparison with LOTC. These findings align with those of Cichy et al. (2014) on objects, indicating that the similarity between MEG responses and V1 occurred at an earlier stage compared to the corresponding association with the inferior temporal (IT) cortex.

Limitations and future directions

Exploring the different types of information that contribute to our understanding of actions represents an intriguing and promising avenue for future research. In the future, researchers can investigate the core features of actions at each specific level. For instance, what are the core features to actions at the subordinate level? Additionally, researchers can change the EEG experiments to make them more similar to previous behavioral studies. For example, research can use a block design task to instruct participants to focus on actions at a specific level. This exciting direction holds great potential for advancing our understanding of action perception.

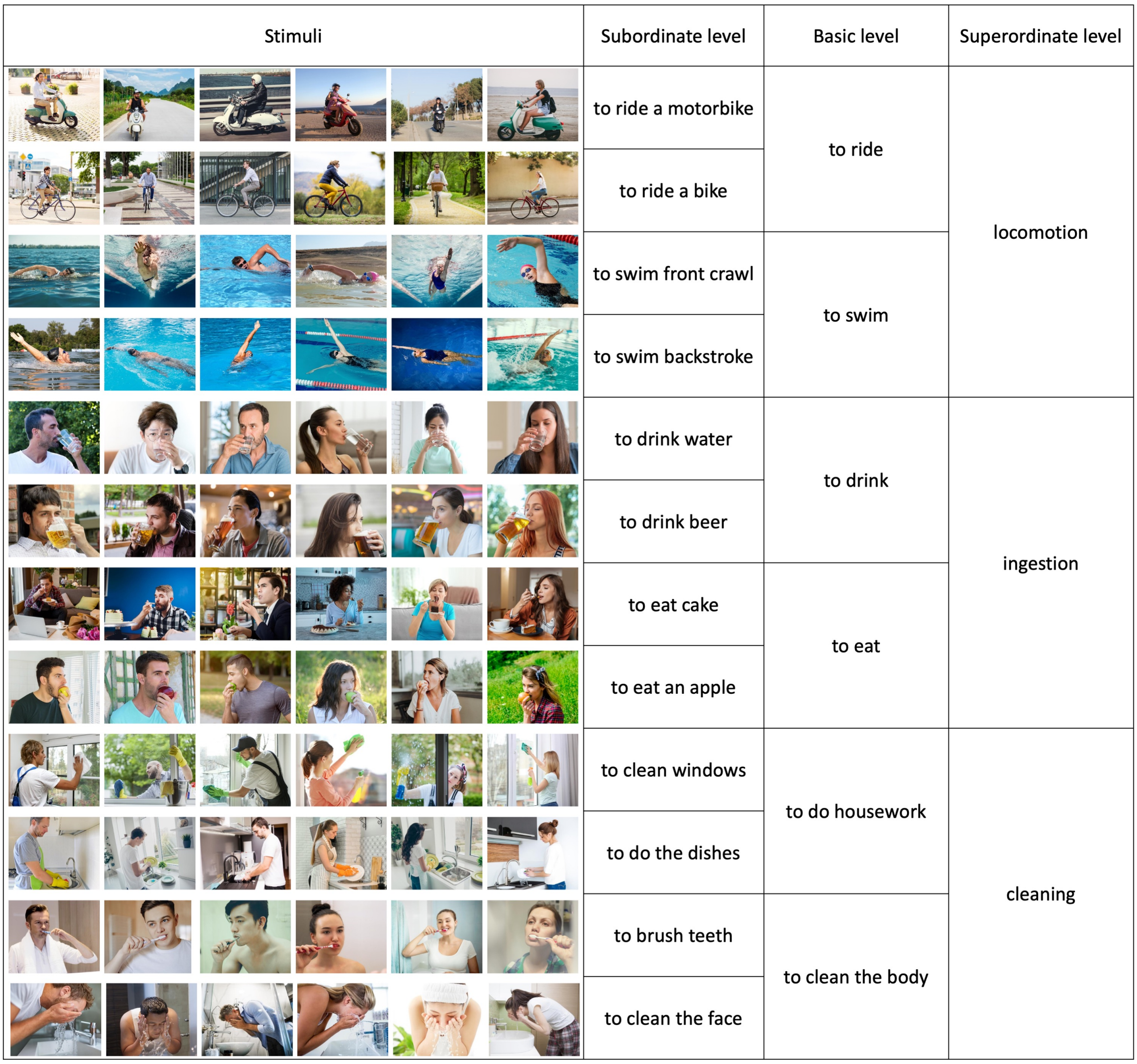
4.5 Methods

Participants

Twenty-five subjects (age: mean± SD = 23.8 ± 4.7 years old, female: 21) from the University of Regensburg participated in the EEG experiment. Three participants were removed since their behavioral performance during catch trials did not differ from chance level (all $p > .1$). Thus, 22 participants (23.5 ± 4.7 years old, female: 20) were entered into the analysis. All participants gave written informed consent and were rewarded with 10 Euro/hour or course credit. All participants were right-handed with normal or corrected-to-normal vision. The experiment was approved by the local Ethics Committee at the University of Regensburg.

Stimuli

We used a stimulus set with hierarchical structure, comprising of three taxonomic levels. In particular, the stimulus set contained 72 static images of twelve actions, with six exemplars per action (600 x 480 pixels). These twelve actions were organized into three hierarchical taxonomic levels (Figure 4.4): twelve actions at the subordinate level (to ride a bike, to ride a motorbike, to swim front crawl, to swim backstroke, to drink water, to drink beer, to eat cake, to eat an apple, to clean windows, to wash dishes, to brush teeth, and to clean the face); six action categories at the basic level (to ride, to swim, to drink, to eat, to do housework and to clean the body); and three action categories at the superordinate level (locomotion-related action, ingestion-related action, and cleaning-related action). The hierarchical organization of the stimulus set was examined to match the three taxonomic levels used in a multi-arrangement experiment (Kriegeskorte et al., 2008) of our previous work (Zhuang et al., 2023).



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

Figure 4.4. Stimulus material. 1st column: the exemplars of one action; 2nd ~4th column: labels of actions at the subordinate, basic and superordinate level.

Design and task

We used an event-related design (Figure 4.5) for both EEG and fMRI experiments. On each trial, participants first viewed a fixation cross (with a varied duration for EEG: 1000--1500 ms; a fixed duration for fMRI: 3000 ms), followed by a static image of an action (for EEG: 200 ms; fMRI: 1000 ms). Prior to each trial for the EEG experiment, a green fixation cross was presented for 800 ms, during which participants were encouraged to blink with their eyes if desired. A fixation cross was presented at the end of each trial for 1000 ms. Occasional catch trials, accounting for 11% of all trials, were used for the task. During the catch trial, participants were given an action label at one of the three taxonomic levels and instructed to judge whether the image of the previous trial was consistent with the label by using their right hand (index finger: yes; middle finger: no). To not bias participants, the labels of the three taxonomic levels were presented equally often. Additionally, null events representing 22% of all trials were pseudo-randomly provided to enhance design efficiency (no subsequent null events and catch trials). Finally, each exemplar was repeated twelve times during the EEG experiment, while each exemplar of the fMRI experiment was repeated six times. The EEG experiment lasted approximately 55 minutes, whereas the fMRI experiment lasted approximately 50 minutes, including an additional 5-min anatomical (T1) scan. Notably, all participants joined a short training-experiment before the main experiment.

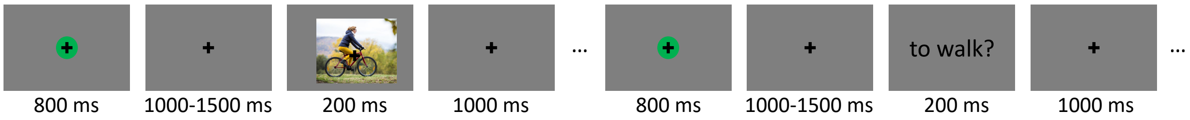


Figure 4.5. Example trial sequence. At the beginning of each trial, a green dot with an 800 ms duration is presented for participants' blinking, if they would like to blink their eyes. Following a 1000-1500 ms fixation, an action image is displayed for 200 ms, followed by another 1000 ms fixation.

EEG Data acquisition and preprocessing

We collected EEG data with an Easycap 64-channel system. An elastic cap (EasyCap, Herrsching-Breitbrunn, Germany) was used to situate the electrodes, which were adjusted for each subject due to anatomical landmarks and Cz as a reference during recording. Electrodes were evenly placed and positioned on five equidistant concentric circles around FCz. All of the points on the vertical (Fpz, AFz, Fz, FCz, Cz, CPz, Pz, Iz) and horizontal (C5, C2, C1 and C6) were in the same places as in the extended international 10–20 system (Oostenveld and

Praamstra, 2001). We did not conduct VEOG and HEOG. The impedances were maintained under 10 Ohm. The EEG signal was captured at a rate of 500 Hz (BrainAmp MR plus, Brain Products, Gilching, Germany) and amplified between 0.1 and 100 Hz.

We performed data preprocessing analysis with Matlab (https://www.mathworks.com/products/matlab.html) and the fieldtrip toolbox (https://www.fieldtriptoolbox.org/, Oostenveld et al., 2011). To run the analysis with RSA, we first constructed the EEG-based RDMs over time which is based on the time-lock decoding analysis. Hence, for time-locked decoding analysis, continuous EEG data were divided into epochs of -0.2 s to 1 s relative to stimulus onset. The average reference value was used as the reference for all the data. To exclude very slow frequencies, epochs were high-pass filtered with a 0.1 Hz. Using independent component analysis (ICA, Delorme and Makeig, 2004), we manually removed the components of eye blink and movement artifacts. Additionally, data was also checked to eliminate trials with eye blink or movement and other artifacts (e.g., tonic muscle activity) by visual inspection. No sensors were excluded by any participants.

Representational Similarity Analysis

RSA is a method for investigating how (e.g. where and when) the tested model matches activity patterns in brain. In this study, RSA has been used to capture the representational information of each of the three taxonomic levels. In addition, to avoid the influence of low-level features, we used multiple regression RSA, which regresses the visual control model and the scene control model (Krizhevsky et al., 2012, see DNN section for more details).

EEG-based RDMs construction

To establish EEG-based RDMs, we carried out pairwise exemplar decoding across times. In particular, we extracted 62 channel activations from the time-locked EEG data at each time point, and then organized these activations into pattern vectors for each of 72 exemplars. We conducted the linear discriminant analysis (LDA) as the classifier on all pairwise exemplars. Each classifier was trained and tested using 2/3 for training and the remaining 1/3 for testing in each pairwise classification. Each pairwise-exemplar classification was repeated 66 times with a random assignment of trials to training and testing bins. This resulted in a 50% accuracy of the chancel level. The accuracy of each pairwise exemplar was obtained by averaging these repetitions. The accuracy for each pairwise decoding was the value of each cell in the RDM. In this way, we made a 72 x 72 RDM at each time point. Additionally, to better visualize the EEG-based RDMs, we conducted MDS (Figure 4.2) with Matlab (using the function cmdscale).

Representational dissimilarity models (RDMs)

Taxonomic level RDMs. Based on our hierarchical stimuli set, we have twelve actions at the subordinate level, six action categories at the basic level and three action categories at the superordinate level. Hence, we constructed the subordinate level model using a 72 x 72 binary RDM with twelve clusters filled with zero in the diagonal, the basic level model by the same size of RDM (72 x72) with six clusters in the diagonal, and a 72 x 72 superordinate level model with three clusters in the diagonal (Figure 4.6 A).

Behavioral RDM. In addition to the taxonomic level RDMs, we used a behavioral RDM (see Figure 6B) based on a multi-arrangement task (Kriegeskorte & Mur, 2012), collected for the same stimulus set in a separate group of N=18 participants (see Zhuang, et al, 2023, for details). In brief, participants were presented with the action images on a screen and were asked to arrange them according to the perceived similarity of their meaning.

Low-level visual control RDM. To account for differences between the different stimuli with respect to low-level visual features, we used the first convolutional layer from ResNet 50 (He et al., 2015). Specifically, ResNet 50 is a deep neural network that has been pretrained on a large-scale image dataset (ImageNet dataset). To build the RDM from layer 1, we first extracted features of each image. The process involves identifying the activations of each unit in the first convolutional layer and then transforming these values into activation vectors. Finally, based on these activation vectors from each image, a 72 x 72 visual control RDM was constructed by 1-Pearson correlation (Kriegeskorte, et al., 2008). The visual control model can be seen in Figure 4.6 C.

Scene control RDM. To account for contextual information related to the scene or background in which the action occurred, we established a 72 x 72 scene control model. The process for creating this model closely followed the methodology outlined for constructing the low-level visual control model. The key distinction was that we generated activations in response to each image by extracting features from the second-to-last layer of the ResNet50 model trained on a large image dataset, specifically for the purpose of categorizing scenes (as detailed in Zhou et al., 2017 and 2018). We can see the scene control model in Figure 4.6D.

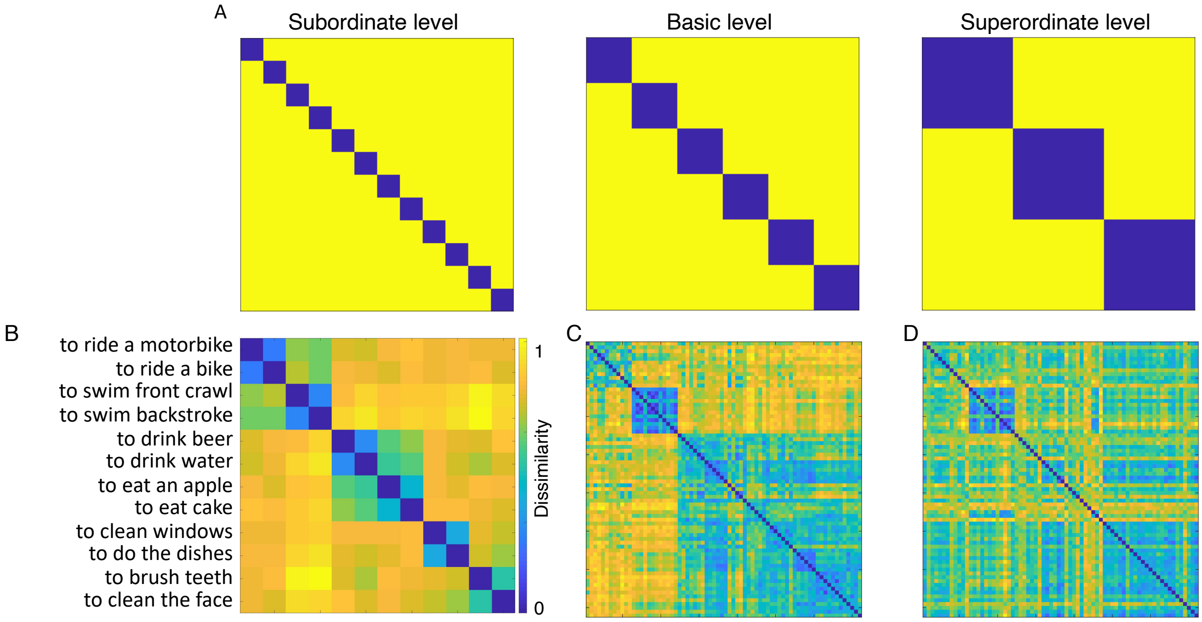


Figure 4.6. Representational dissimilarity models (RDMs). A: RDMs corresponding to the subordinate, basic and superordinate level. B: Behavioral RDM (averaged across N = 18 participants) resulting from multi-arrangement task. C, D: Low-level visual control RDM and scene control model obtained from a deep neural network (ResNet 50; see text for details).

EEG-fMRI fusion

To reveal hierarchical action representations in both time and space between the low-level and high-level visual cortex, we correlated ROI-specific fMRI RDMs with EEG-based RDMs over time (Cichy et al., 2016; Hebart et al., 2018).

For the low-level visual cortex, we selected V1 for its generating the low-level visual information. For the high-level visual cortex, bilateral LOTC was chosen due to our previous fMRI work (Zhuang et al., 2023). Specially, in the fMRI study (Zhuang et al., 2023), we examined action representations at three taxonomic levels in different regions related to action understanding, including bilateral V1, SPL, LOTC, premotor cortex and IFG. We found that actions at all three levels can be decoded in all these regions, and in LOTC, the basic level model showed the highest similarity in comparison with the subordinate and superordinate levels. Additionally, based on the whole-brain searchlight RSA analysis, we conducted the conjunction analysis and the results revealed that the bilateral occipitotemporal cortex (OTC) represented actions at all three taxonomic levels. Additionally, the OTC is the jointly regions for the unique information to subordinate and basic levels, which the unique information was obtained by regressing out the other two level models and two control models during the RSA analysis (see more details in Zhuang et al., 2023). Hence, we selected bilateral V1 and LOTC as ROIs to figure out at which time point the neural signature of fMRI RDM matches with that of EEG-based RDMs. Notably, all fMRI data was obtained from our previous fMRI data. Additionally, we used the same methods for preprocessing the fMRI data and creating the selective ROIs, as well as the corresponding ROI-based RDMs (see method section, Zhuang et al., 2023).

Finally, to make EEG-fMRI fusion, we computed the Spearman correlation between EEG-based RDMs and fMRI-based RDMs across times. In particular, for fMRI-based RDMs (Bilateral V1 and LOTC), the beta estimates for each voxel in a given ROI were extracted and these beta estimates were transformed into t-values, resulting in a vector of t-values for each exemplar for each participant. Then we created RDM based on these vectors and averaged the neural RDMs across participants. For EEG-based RDM, we also obtained RDMs for per participant at per time point and then average RDMs at each time point across participants. With these EEG- and fMRI- RDMs, we made the Spearman correlation over time and plotted in Figure 4.3.

Statistics

To determine the significant effect of RSA and multiple-regression RSA, we performed Threshold Free Cluster Enhancement (TFCE, Smith and Nichols, 2009) with the CosMoMVPA toolbox (Oosterhof et al., 2016; http://cosmomvpa.org/index.html). A sign-permutation test with 5000 booststrapping iterations was used for multiple-comparison correction. Thresholded statistical values were conducted at $Z > 1.64$ (corrected $p < .05$, one tailed, see results from Figure 4.4).

5 Chapter 5: Discussion

5.1 Thesis recap

The human brain possesses a remarkable ability to quickly recognize what it perceives at a single glance, a common occurrence in daily life (Greene & Oliva, 2009; Murphy, 2004). For example, when we observe physical movements that evoke multiple labels in our minds. According to the degree of abstraction, these labels of actions could be classified into three hierarchical levels (Rosch et al., 1976; Zhuang & Lingnau, 2022): subordinate (e.g. ‘to swim backstroke’), basic (e.g. ‘to swim’), and superordinate (e.g. ‘locomotion’). As the distinct representations have not been well investigated in the hierarchical organization of actions. Therefore, in current work, my goals were to investigate the distinct representations of observed actions at the subordinate, basic and superordinate levels in behavioral and neuroimaging experiments.

In the first study (Chapter 2), I conducted a series of behavioral experiments to determine the different characteristics of actions across different levels of abstraction. I chose a set of verbs with two classes from a linguistics book (Levin, 1993). Participants rated semantic relationships, clustered the verbs, and named these clusters leading to action labels at the superordinate level. Participants also generated subordinate names for each action, and then participants evaluated the compatibility of these subordinate names with superordinate names. Subsequent experiments (feature listing, priming task and the rapid category verification task) showed that basic-level actions had the most common features in comparison to the subordinate and superordinate levels, and subordinate- and basic-level actions were recognized more quickly than actions at the superordinate level. These findings illustrate that basic level representations play a central role in action categorization.

In the second study (Chapter 3), I used high spatial resolution fMRI to measure the neural representations of observed action across different taxonomic levels. The results showed that the unique information at the subordinate and basic levels was observed in the occipitotemporal cortex, and that actions at the basic level captured the most information in LOTC. Additionally, the behavioral action space, established through human ratings, was best represented by patterns of activation in the LOTC and SPL, showing the highest similarity to the basic-level model. These results suggest the salience of the basic level in action categorization, providing the neural evidence for basic-level advantage.

In the third study (Chapter 4), I used an EEG to examine the dynamical representations of observed actions at the three taxonomic levels. The findings illustrated that actions at all three taxonomic levels were represented at the same time of 170 ms, and the highest similarity between EEG-based RDMs. These findings indicate an integrated, space- and time-resolved view of the hierarchical organization of actions.

To sum up, the current projects examined how action understanding is hierarchically organized. The findings of behavior and fMRI experiments consistently highlighted the central role of the basic level in action categorization. Specifically, actions at the basic level had the most common features and was recognized faster. Additionally, the basic level model captured the most information in LOTC than the subordinate and superordinate models, while it also showed the highest similarity to the behavioral action space. These findings provide the neural evidence for the basic-level advantage. Finally, EEG results revealed that actions across different taxonomic levels are processed concurrently at approximately 170 ms, indicating the rapid and simultaneous processing of action across different levels of abstraction.

5.2 Different hierarchical action representations in behavioral studies (Chapter 2)

My behavioral studies aimed to explore the characterization of actions at three hierarchical levels: subordinate, basic, and superordinate levels. In line with findings regarding hierarchical levels of object representations (Rosch et al., 1976), both Experiment 2 (Rosch et al., 1976) and Experiment 4 (Chapter 2) revealed that the basic level showed the most features compared to the subordinate and superordinate levels. This highlighting its capacity to contain the maximum information. This finding suggests that the principles proposed in object categorization can be effectively applied to action understanding, highlighting the central and privileged role of the basic level in categorizing both actions and objects. Thus, my research adds to the establishment of a hierarchical theoretical framework for categorization in human cognition. My findings also indicate that the cognitive processes involved in categorization share common principles across different domains (objects and actions). This provides a deeper understanding of how the human brain efficiently organizes and recognizes various stimuli in the world.

However, it is worth noting that certain results in my behavioral study differed from those observed in object categorization studies. For instance, I did not detect significant differences between the subordinate and basic levels in the priming task (Experiment 5) and the rapid category verification task (Experiment 6). This differs to Rosch et al.'s (1976) finding that participants responded faster to objects at the basic level compared to the subordinate level (as observed in Experiment 7 of Rosch et al., 1976). To examine the differences between objects and actions at the subordinate and basic levels, de la Rosa et al. (2014) conducted a direct comparative experiment. Their findings revealed a greater difference between the subordinate and basic levels for objects in contrast to actions. Several factors could potentially account for these inconsistencies between objects and actions at the subordinate and basic level. One possible explanation for the lack of differences in priming effects (Experiment 5) and recognition speed (Experiment 6) between actions at the basic and subordinate level, as observed in our current study, can be due to the fact that labels for subordinate level actions (e.g., 'to swim breaststroke'— 'Brustschwimmen') always include the basic level label (e.g., 'to swim' — 'Schwimmen'). It is worth noting that this phenomenon is similar to some of the object categories studied by Rosch et al. (1976), such as 'desk lamp' at the subordinate level including 'lamp' at the basic level, although it is not always the same (e.g. 'Levis' at the subordinate level and 'pants' at the basic level). The other possible reason is that previous studies (Johnson & Mervis, 1997; Tanaka, 2001; Tanaka & Taylor, 1991) have found no difference in speed and accuracy between the basic and subordinate levels of object recognition among experts in the respective object fields. This suggests that some of the differences observed between the subordinate and basic levels in previous studies may be due to differences in familiarity or expertise with objects at these two taxonomic levels. In our current study, however, the participants were likely to have a high level of familiarity with actions at both the basic level and the subordinate level. For example, actions the subordinate and basic levels are frequently used in the daily life, such as ‘to drink beer’ (subordinate level) and ‘to swim’ (basic level). Additionally, participants in our current study may have profited from the inclusion of objects in the majority of the names of the subordinate actions. This inclusion may have eliminated some differences between the basic and subordinate levels. Further research is required to investigate the specific conditions under which variations in the number of shared features in feature-listing paradigms, as well as the speed and accuracy of object and action recognition, may differ between the basic and subordinate levels.

In summary, while there are some distinctions between subordinate and basic level for actions and objects, the majority of our results align with previous studies that support the view of a basic-level advantage. This correspondence implies that the hierarchical frameworks proposed for object categorization (Rosch et al., 1976) can be applied analogously to the understanding of action categories.

5.3 Resolving hierarchical action processing in space (Chapter 3)

As a natural progression from my behavioral study, the subsequent fMRI experiment further investigated the neural underpinnings of actions across three taxonomic levels. The fMRI results showed that the lateral and ventral portions of the LOTC appeared to capture the unique similarities of the observed actions at the subordinate and basic levels, with a preference for the basic level. These findings supported the perspective by Wurm and Caramazza (2022), as well as Lingnau and Downing (2015), who suggested that the LOTC is associated with conceptual action representations at different levels of abstraction and varying modalities. Previous studies have found that LOTC plays an important role in processing action components (Tarhan & Konkle, 2020; Tucciarelli et al., 2019; Wurm et al., 2017). For instance, using action videos, Tarhan and Konkle (2020) measured the neural representations of five large-scale networks (i.e., social and four interaction envelopes), and found that LOTC is sensitive to the features of person and object-directedness. When combined with the insights from my study and the propositions by Wurm and Caramazza (2022) and Lingnau and Downing (2015), a compelling conclusion emerges: LOTC possesses the abilities for flexible accessing actions across different levels and across modalities.

However, some researchers proposed that it is anterior intraparietal area (AIP), rather than LOTC or ventral visual cortex, plays an important role in action understanding (e.g., Lanzilotto, et al., 2020; Urgen & Orban, 2021). In particular, Lanzilotto et al., (2019) holds the view that AIP is a hub in manipulative action network. In particular, they conducted chronic neuronal recordings in monkeys' AIP, combined with tracer injections and found robust neural selectivity for seven distinct observed manipulative actions, particularly in the posterior part of AIP (pAIP). This selectivity was associated with motor coding of grip type and own-hand visual feedback. The findings indicate that pAIP, with its connections to temporal and prefrontal regions, plays a key role in processing observed manipulative actions. Additionally, Lanzilotto et al., (2020) recorded neuronal activity in AIP of macaques viewing different actions in different visual formats. They found that no invariant action-selective neurons were identified, and that action exemplars were classifiable across different visual formats. These findings suggest that the AIP integrates format-dependent information and visual features to achieve a stable readout of observed manipulative action identity. Furthermore, a recent proposition by Orban et al. (2021) suggests that the parietal region AIP, possesses the capability to integrate diverse visual features (e.g., body movement, object characteristics, and body-object interaction) and encode observed actions. While some studies emphasize the role of AIP in action manipulation, my research suggests that the LOTC, in contrast, is involved in comprehending actions across various hierarchical levels. Further research can explore the functions of both LOTC and AIP, providing a more comprehensive understanding of how the brain understands actions.

Another noteworthy aspect is the absence of any significant brain region representing unique information to actions at the superordinate level. Several factors may contribute to this phenomenon. One reason might be the lack of power in my current study. Alternatively, it is possible that no specific brain region encodes unique information for superordinate actions, as their information may primarily originate from lower levels, namely the subordinate and basic levels. Additionally, it might be the representation of actions at the superordinate level is more distributed in comparison to the subordinate and basic level. This interpretation can be supported by the findings of Abdollahi et al., (2013). They found that different categories of actions at the superordinate were represented different parts of parietal cortex.

Taken together with the findings from the behavioral experiments, my work might provide a plausible explanation for the privileged role of the basic level information. Specifically, using a hierarchical stimulus set and multiple-regression RSA, we observed the strongest similarity between neural patterns and the basic level model in bilateral LOTC. Unique information for actions at the basic and subordinate levels was jointly represented in the bilateral OTC. Furthermore, the behavioral action space model was represented in patterns in occipitotemporal and SPL regions, showing a high similarity with the basic level model. These findings may imply that the basic level advantage is a result of cortical processes. Additionally, the results from the current study on observed actions are consistent with those of Iordan et al. (2015), who found that objects at the basic level showed the strongest similarity with patterns of activation in the LOC. Together, these findings, along with the results of our study, support the perspectives of Wurm and Caramazza (2022) and Lingnau and Downing (2015), suggesting that the LOTC is associated with conceptual action representations at various levels of abstraction and modalities. To sum, the high-level visual cortex encodes basic-level information across various domains, including actions and objects.

5.4 Resolving hierarchical action processing in time (Chapter 4)

In Chapter 4, I investigated dynamic representations of hierarchical action processing by using EEG. Results from both RSA and multiple regression RSA revealed that we could decode actions at these different taxonomic levels at approximately 170 ms. The highest peak similarity between the RDMs of bilateral LOTC and EEG-based RDMs was identified at around 230 ms. These findings bridge the temporal and spatial dimensions of processing.

It is intriguing why actions at these three taxonomic levels exhibit temporal synchrony in their processing, particularly in contrast to previous research in the field of behavior studies (de la Rosa et al., 2014; Zhuang & Lingnau, 2022). One potential explanation for these results could be the nature of the task. Previous behavioral studies have identified variations in how participants respond to actions across different taxonomic levels (Zhuang & Lingnau, 2022; de la Rosa et al., 2015). These studies all instructed participants to concentrate on actions at a specific level within the hierarchy. In contrast, the design of the current EEG experiment did not specifically guide participants to focus solely on one hierarchical level. Instead, it aimed to minimize any potential bias towards a particular level. Therefore, the neural data recorded in this experiment reflected the natural progression of actions without imposing a strict focus on any single level.

Another plausible explanation for the varying results might be that the current study used different analysis methods compared to previous research (e.g., Carlson et al., 2013; Cichy et al., 2014). Prior studies used cross-validation classification to identify the time point at which two categories could be distinguished, whereas we used RSA method. Classification has the advantage of identifying categories within a specific level, such as distinguishing between 'face' and 'body,' or 'animals' and 'humans.' However, it can also include lower-level information when decoding higher-level categories. For instance, when a classifier is trained to distinguish between 'face' and 'body', it may unintentionally mix information from distinguishing between 'animal faces' and 'animal bodies' that comes from lower-level distinctions. Additionally, it is essential to consider the influence of low-level visual features for each category at different hierarchical levels, which could affect the interpretation of the results. For instance, the disparities in visual information between subordinate and basic levels might not be the same. In this context, the visual distinctions between two subordinate-level actions, like 'to drink coffee' and 'to drink water,' could be less pronounced compared to the basic level, where actions like 'to drink' and 'to ride' exhibit more significant differences in low-level visual features. Thus, although we could not divide category representations at a certain hierarchical level by using RSA, we were able to regress out the confounding of the low-level features, and the scene-related information. Further study could continue working on the concurrent processing of actions across levels.

Additionally, the findings of the current EEG study suggesting that action understanding processes occurre at approximately 170 ms, are in line with earlier research on the timing of action understanding. For instance, Tucciarelli et al. (2015) identified abstract action representations at around 200 ms, and Dima et al. (2022) observed action feature representations between 150 and 600 ms. This temporal alignment suggests consistency in the temporal dynamics of action understanding across various studies.

In summary, my EEG study (Chapter 4) on actions at different taxonomic levels uncovered a synchrony in their temporal processing, occurring at around 170 ms. The highest peak between the RDMs of bilateral LOTC and EEG-based RDMs was identified at around 230 ms. Together, our findings offer an integrated, space- and time-resolved perspective on the hierarchical organization of actions.

5.5 Integration of findings

My research project focused on unraveling the hierarchical organization of action understanding. The evidence gathered thus far strongly supports the idea that the basic level of actions may take precedence, aligning with findings from behavioral experiments on object levels by Rosch et al., (1976) and neuroimaging study by Iordan et al., (2015). Overall, my study offers a hierarchical perspective on action comprehension similar to the research on object levels (Rosch et al., 1976; Iordan et al., 2015). It implies that the basic level holds a central role not only in categorizing objects but also in categorizing actions, contributing significantly to the development of a comprehensive theoretical framework for understanding human cognitive categorization processes. This suggests an exciting parallel: the concept of basic-level categories, initially proposed for objects, may extend to actions, highlighting a common organizational principle in human cognition.

It is important to emphasize the significance of the basic level in categorization. This raises the question: why does the basic level hold such importance? One compelling explanation can be found in the observations from Experiment 2 of Rosch et al., (1976) and Experiment 4 (Chapter 2), which both indicate that the basic level provides the richest information compared to the subordinate and superordinate levels. This suggests that the basic level strikes a balance between informativeness and distinctiveness, consistent with the findings of Mervis and Crisafi (1982). They constructed a differentiation scale based on participant ratings of concept similarities at different levels and found that the basic level ranked higher in differentiation than the subordinate and superordinate levels. Another explanation comes from the fMRI experiment in Chapter 3, where we observed that the LOTC captures basic-level information best, and that the way our brains process behavioral space is most similar to the way they represent basic-level actions. This may imply that the basic-level advantage is underpinned by cortical processes.

Additionally, other factors might also influence the basic-level advantage. These include the frequency of basic-level information or the early age at which it is learned for objects (Rosch et al., 1976; Wisniewski & Murphy, 1989). Compared to the other two levels, basic-level concepts are not only more frequently encountered but also learned at a very young age (Rosch et al., 1976). Taken together, these factors may explain the privileged role of the basic level in human categorization processes.

In addition to the basic level, it is interesting to discuss the subordinate and superordinate levels. I will begin with the subordinate level, which represents the lowest and the concrete level in the hierarchical organization of action understanding. It offers a fine-grained categorization that enables a more detailed and specific distinction among actions. In contrast to the basic level, which captures common features most, the subordinate level contained more shared features in comparison to the superordinate level and less distinct features than the other two levels (Experiment 4, Chapter 2). Additionally, in several experiments (Experiments 5 and 6 in the behavioral study, Chapter 2; multiple regression RSA results in the fMRI experiment, Chapter 3; and temporal dynamics in Chapter 4), I consistently observed that the differences between actions at the subordinate and basic levels were sometimes not large. In particular, in experiments 5 and 6 of the behavioral study, both subordinate and basic level actions were recognized more quickly by participants than the superordinate level. The representation of unique information specific to the subordinate level was observed in the OTC. This is the same region in which we found basic level representations at slightly bigger sizes. Additionally, the

EEG study showed that actions at all three levels were represented simultaneously. One possible reason for the limited differences observed in my studies between subordinate and basic actions is the familiarity and frequency of use of subordinate actions in everyday life. Actions like 'to drink coffee' or 'to wash dishes' are everyday activities that people perform frequently, leading to a high level of familiarity. As a result, participants in the experiments recognized actions at the subordinate level as fast as actions at the basic level. Despite the challenges of differentiation, it is important to note that there are significant differences between these two levels, as evidenced by key findings such as feature capacity, representations within the LOTC, and their contributions to neural representations of action behavioral space (see Figure 8 in Chapter 4 for more details). These distinctions, although slight, highlight the unique characteristics of actions at the basic level compared to those at the subordinate level. In the future, researchers can further investigate the difference between actions at the subordinate and basic level. For example, researchers can explore the different core dimensions of actions at the subordinate and basic levels.

In contrast to the subordinate level, the superordinate-level concepts represent the highest level in the hierarchy of action categorization. This level includes actions at a broader, more abstract level of classification. Actions at the superordinate level tend to have more distinct and less shared features compared to actions at the subordinate level (see Experiment 4, Chapter 2 for details). It is worth noting that I did not identify any region uniquely representing actions at the superordinate level in the fMRI experiment. While this absence of distinct activation may partly be attributed to limitations in stimulus power, another plausible explanation is that the information at the superordinate level draws upon and integrates features from the lower levels. Future research could examine the representation of actions at the superordinate level and the extent to which it relies on information from lower hierarchical level. As an example, participants within an fMRI scanner can be guided to identify actions at superordinate, basic or subordinate level. This enables researchers to explore cross-decoding between actions at different levels to investigate whether lower-level neural representations can represent information related to superordinate-level actions.

Finally, I want to highlight some insight gained from my fMRI study. I found that OTC is able to represent actions across all three taxonomical levels, and I could identify joint regions capturing unique information regarding the subordinate and basic levels. These findings are consistent with the proposal made by Wurm and Caramazza (2022) regarding the object and action recognition. They propose that OTC divides into lateral and ventral pathways for recognizing actions and objects, offering a clearer understanding of their functional organization, from features to concepts, from concrete to abstract.

In summary, the current study provides a hierarchical perspective on action understanding and categorization. In particular, our findings regarding the hierarchical structure of action recognition imply that the principles proposed by object-level categorization also extend to the categorization of actions (Iordan et al., 2015; Rosch et al., 1976; Wurm & Caramazza, 2022; Zhuang & Lingnau, 2022). This bridges the gap between object and action categorization and contributes to a more comprehensive understanding of categorization processes in the human brain. Furthermore, the insights gained from the LOTC contribute to a deeper understanding of its functions, particularly its remarkable flexibility in generating representations at various levels of abstraction.

5.6 Limitations

There are several limitations of the studies presented here that need to be discussed and considered in future research. First, we used daily action images, the naturalistic stimuli, in order to increase the ecological validity of our findings (Haxby et al., 2020; Nastase et al., 2020) and to investigate actions at different taxonomic levels. However, one limitation revolves around the control of objects and scenes. In the existing design, I did not control for objects or scenes involved in the action. I used analysis method to avoid the confounding between objects/ scenes and actions. In future studies, it is recommended to control for scene and objects. Possible control conditions could control for object properties such as their size, shape, color, or position relative to the actions. Further researchers can also vary scenes and objects as a controlled approach and measure their effect on action representations. By doing so, researchers can gain a clearer understanding of how objects influence the representation of actions.

Another limitation of the current study is that, in order to examine the natural representation of actions, I did not instruct participants to focus on any particular level in the fMRI and EEG experiments. While the behavioral experiments measured the identified speed of actions at each individual level, these two designs refer to two different processes in the brain. Because of the current purposes of neuroimaging studies, I believe that the current design is much better because it is the first time that these effects of actions have been investigated. Based on what I have investigated, the direction for further studies can keep the same design as the behavioral experiments to instruct participants to pay attention at each level. I will discuss this in more detail in the next section.

5.7 Future Directions

Future researchers should investigate differences between the hierarchical organization of objects and actions. Although de la Rosa et al., (2014) have revealed that the differences between the subordinate and basic level labels are larger for objects than for actions, they focused on one category of actions, namely social interactions. It would be interesting to also examine the differences between the subordinate and basic level of other action categories.

Additionally, the current experimental design of fMRI and EEG studies is intentionally structured to reflect the natural states of participants when recognizing actions. As a result, we observed that actions at all three hierarchical levels were simultaneously represented. This observation presents an interesting contrast to previous behavioral experiments, such as those conducted by Zhuang and Lingnau (2022) and de la Rosa et al. (2014), where participants were specifically instructed to concentrate on a single, particular hierarchical level during the task. Given this difference, one potential way for future research might be to instruct participants to focus on only one specific level when collecting EEG data. By directing attention of participants to a single taxonomic level during neural data collection, researchers can examine how neural representations change as participants engage in specific levels of processing. This approach could help researchers gaining insight into the neural dynamics of actions at different levels when focusing on a certain level and how these dynamics differ from the simultaneous representations observed in the current study. Based on behavioral studies and prior object-level research (Zhuang & Lingnau, 2022; Carlson et al., 2013), my prediction is that when processing actions at different taxonomic levels, the subordinate level will appear first in neural representations, followed by the basic level, and finally, the most abstract superordinate level.

Another option for future research would be to include short action video clips instead of static pictures as stimuli in an experiment. Action movies provide a more realistic representation of dynamic movements, allowing for a deeper understanding of the nature of action (Haxby et al., 2020; Dima et al., 2022).

Furthermore, identifying the core attributes that define each taxonomic level is should be aimed in future research. What exactly distinguishes the representations of actions at each level, and are there fundamental features that are essential for each level? It would be interesting to measure the specific features that are unique to each level, such as objects and scenes (Jozwik et al., 2023; Loschky & Larson, 2010). By examining the distinctive features of each level, we can gain deeper insights into the hierarchical organization of action knowledge and potentially uncover key elements that drive the differentiation of action representations at different taxonomic levels.

In summary, the current study suggests several potential research directions. First, researchers can investigate the distinctions between subordinate and basic levels for objects compared to actions. Second, researchers can investigate the neural dynamics of actions at different levels when participants focus on a specific level. Third, future studies could use dynamic action video clips to examine specific levels of focused attention during EEG data collection. Lastly, researchers may explore the core dimensions of actions at different levels. These future studies hold the potential to enhance our understanding of action representation and cognitive processing.

5.8 Conclusion

In conclusion, the three studies conducted during my PhD shed light on the role of taxonomic levels in the action categorization. The first study revealed that basic-level representations are central to action categorization, as they showed the most common features and faster recognition times compared to the superordinate level. The second study, using fMRI, showed that the unique information to the subordinate and basic levels were represented in OTC and basic level is the most similar to the neural representations of the behavioral space. These findings highlighted the salience of the basic level during action categorization. The third study, using EEG, revealed the temporal representation of actions at all taxonomic levels at the same time and the peak similarity between the RDMs of bilateral LOTC and EEG-based RDMs was around 230 ms. Together, these findings provide a comprehensive understanding of the mechanisms underlying action categorization, from behavioral characteristics to neural representations. The basic-level information plays a key role in behavior and neural space representations, enhancing our knowledge of how the human brain classifies actions across different levels of abstraction. These insights contribute to a deeper understanding of action representation.

Appendix

Study 1 Supplementary materials

Supplementary Material

Experiment 1: Rating of semantic similarity

Table S1. Basic level action verbs in their infinitive forms (and their German translation) used in Experiment 1 (right column), together with the corresponding superordinate category labels (left column).

Category labels (superordinate level)	**Action verbs (basic level)**
communication (*Kommunikation*)	to read (*lesen*)
	to write (*schreiben*)
	to tell (*erzählen*)
	to sing (*singen*)
	to talk (*unterhalten*)
	to shout (*schreien*)
	to call (*rufen/ anrufen*)
locomotion (*Bewegung*)	to climb (*klettern*)
	to run (*laufen*)
	to swim (*schwimmen*)
	to drive (*fahren*)
	to walk (*gehen*)
	to jump (*springen*)
	to dance (*tanzen*)
to ingest (*Nahrungsaufnahme*)	to eat (*essen*)
	to drink (*trinken*)
	to feed (*füttern*)
change of state (*Veränderung des Zustandes*)	to cook (*kochen*)
	to brush (*putzen*)
	to break (*brechen*)
to learn (*Lernen*)	to teach (*lernen*)
	to learn (*lesen*)
	to study (*studieren*)
	to memorize (*merken*)
grooming & body care (*Pflege & Hygiene*)	to wash (*waschen*)
	to wear (*anziehen*)
perception (*Wahrnehmung*)	to observe (*beobachten*)
	to feel (*fühlen*)
	to touch (*anfassen*)
	to see (*ansehen*)
	to hear (*hören*)
creation & transformation (*Erzeugung & Herstellung*)	to build (*bauen*)
	to sew (*nähen*)
	to paint (*malen*)
	to draw (*zeichnen*)

Instructions

Scale: 1= completely dissimilar, 4=moderate, 7= completely similar

Below are pairs of action phrases. Please rate how dissimilar versus similar the meaning of each pair of phrases is on a scale from 1 to 7. E.g., the meaning of the words ‘taking a shower - jumping’ can be considered completely dissimilar in meaning, so we could rate their similarity as 1. The meaning of the words ‘taking a shower – taking a bath’ is very similar in meaning, so we could rate it as 7. Please consider only the meaning of these two phrases.

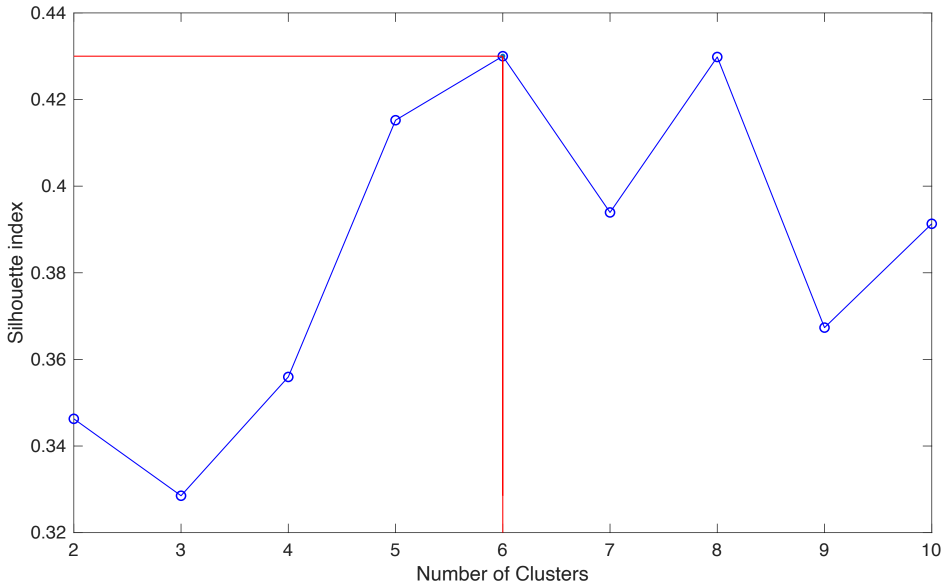


Fig. S1. Silhouette index. The silhouette index provides an estimate of the averaged distance between clusters as a function of the number of clusters. In a range of 2 to 10 clusters, we obtained the highest silhouette index for six clusters.

Experiment 2: Taxonomic depth task

Instructions

First part:

In this part, you will see a series of single phrase presented at the top of the screen. Please type down which kind of superordinate category this phrase belongs to. E.g. when ‘watching TV' shows up at the top of the screen, you could type down ‘relaxation-related action’. You have

30 s to type down the phrase. Please do your best in this study, your performance is important to our study.

Second part:

In this part, you will see a series of single words presented at the top of the screen. Please type down which types of subordinate categories belong to this word. For example, when the 'relaxation-related action' shows up at the top of the screen, you could type down 'listening to music', 'painting'… You have 30s to type your answer down. Please focus on this study, your performance is important to our study.

Experiment 3: Ratings of Relatedness, Complexity, and Abstraction

Instructions

Relationship (weak, strong)

In the following you will be provided with a generic term (e.g. 'Sport', 'Celebration') describing an action at the top of each page. Underneath, you will be provided with several action words. Please read each action word and judge the relationship with the generic term presented at the top of the page. For example, if you could judge the relationship between *watching TV* and *Sport* to be very weak, you could rate it as 1. By contrast, if you judge the relationship between *playing a board game* and *celebration* as very strong, so you could rate it as 7.

Scale: 1 = weak relationship, 4 = neutral, 7= strong relationship

Complexity (simple/ complex)

In the following, you will be provided with a number of brief German phrases. Please read each phrase and judge how easy or complex it is to perform the described action. For example, *making a fist* is very easy to do, so you should assign a 1 (easy) to this action. In contrast, *playing the violin* is a complex action, which is why this action should be assigned a 7 (complex). Please consider all levels of the scale when making your assessment.

Abstraction (concrete/ abstract)

In the following you are provided with a number of phrases describing human actions. Your task is to rate the level of abstraction of each phrase on a scale from 1 to 7. To make sure you understand what we mean by “concrete” and "abstract", please keep the following dictionary definitions in mind when making your ratings:

- Concrete: existing in a material or physical form; real or solid; not abstract
- Abstract: existing in thought or as an idea but not having a physical or concrete existence

For example, *peeling a potato* is very concrete, so you could rate the level of abstraction as 1; *buying a house* is very abstract, so you could rate it as 7.

If you think that a phrase is completely concrete, rate it as 1. If instead you think it is completely abstract, rate it as 7. If you think it is just as concrete as it is abstract, rate it as a 4. The other numbers represent intermediate points on the scale. Please use the entire range of the scale when making your ratings.

Experiment 4: Feature Listing Paradigm

Instructions

In this experiment you will be presented with phrases describing different actions (e.g. "to write an SMS") at the top of the page. Your task is to write down features of this action. Such features could be body parts involved in the action, the target of the action, the type of movements involved, specific postures, the duration, the force required, the speed etc. Of course, not all of these features will always be relevant to a particular action.

Please type down as many features as possible in two minutes. Here are some examples:

to write

Action features:

hold the pen between thumb and index finger
move hand
swinging movement from left to right
pen
paper
communication
small, fast movements
precision

To write an SMS
<u>*Action features:*</u>
type
mobile phone
write
communication
message
use thumb and index finger
contact someone
fast movements

Communication

<u>*Action features:*</u>
targeting another person
exchange of knowledge
mouth
emotional expression
eye contact
body language
speak
letter

Experiment 4: Rating of abstraction

Participants

Twenty native or fluent German speakers (female: 14, age: 27± 4 years old) took part in this experiment. All participants consented to participate in the study via button click, after they were informed about the instruction of the questionnaire.

Stimuli. We used the german action category labels at the three taxonomic levels as shown in Table 4.

Procedure. The questionnaire was conducted with an online platform (www.soscisurvey.com). After being provided with a written instruction (see details below), participants were provided with each of the action labels in a randomized order, one after the other. Participants were asked to judge the abstractness of action labels on the scale of 1 to 7 (1: very concrete, 7: very abstract).

Data analysis. We used the Kruskal-Wallis H test to compare ratings of abstraction between the three taxonomic levels. We used Dunn's post hoc test and corrected multiple tests using Bonferroni correction. All statistical analysis was implemented in SPSS.

Results: The mean rank of abstractedness differed between taxonomic levels (mean rating superordinate level: 5.10, basic level: 4.03, subordinate level: 1.45; $H_{(2)} = 288.75$, $p < .001$; effect size: $\eta^2 = 0.69$). Dunn's post-hoc test with multiple Bonferroni multiple corrections showed that participants regarded action category labels at the superordinate level as the highest abstract level, while they rated action category labels at the basic level as the intermediate level and action category labels at the subordinate level as the most concrete level [superordinate vs subordinate: $p < .001$; superordinate vs basic: $p < .05$; basic vs subordinate level: $p < .001$].

Instruction for rating of abstraction: In the following, you will be provided with a number of phrases describing an action (e.g., to walk; to eat an apple). Your task is to indicate on a scale from 1 to 7 how abstract these actions are (1: very concrete; e.g., 'to peel potatoes'; 7: very abstract, e.g., 'to communicate'). If a term is equally abstract and concrete, please assign it a 4 (equally abstract and concrete). All other digits on the scale fill in the spaces meaningfully. Please consider all levels of the scale when making your assessment.

You may refer to the following definitions:

- Concrete: existing in a material or physical form; real or solid; not abstract
- Abstract: existing in thought or as an idea but not having a physical or concrete existence

1=perfectly concrete;

4=equally abstract and concrete;

7=perfectly abstract

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

Aflalo, T., Zhang, C. Y., Rosario, E. R., Pouratian, N., Orban, G. A., & Andersen, R. A. (2020). A shared neural substrate for action verbs and observed actions in human posterior parietal cortex. Science Advances, 6(43). https://doi.org/10.1126/sciadv.abb3984

Amoruso, L., Gelormini, C., Aboitiz, F., Alvarez González, M., Manes, F., Cardona, J., & Ibanez, A. (2013). N400 ERPs for actions: Building meaning in context. *Frontiers in Human Neuroscience*, *7*(FEB), 1–16. https://doi.org/10.3389/fnhum.2013.00057

Ariani, G., Wurm, M. F., & Lingnau, A. (2015). Decoding internally and externally driven movement plans. *Journal of Neuroscience*, *35*(42), 14160–14171. https://doi.org/10.1523/JNEUROSCI.0596-15.2015

Bach, P., Nicholson, T., & Hudsons, M. (2014). The affordance-matching hypothesis: How objects guide action understanding and prediction. *Frontiers in Human Neuroscience*, *8*(MAY), 1–13. https://doi.org/10.3389/fnhum.2014.00254

Baldassano, C., Beck, D. M., & Fei-Fei, L. (2017). Human-object interactions are more than the sum of their parts. *Cerebral Cortex*, *27*(3), 2276–2288. https://doi.org/10.1093/cercor/bhw077

Balint. (1909). Seelenhammung des 'Schauens', optische Ataxie, raümliche Störungen des Aufmersamkeit. *Monastchrift Fur Psychiatrie Und Neurologie*, *25*, 51–81.

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.

Binkofski, F., & Buxbaum, L. J. (2013). Two action systems in the human brain. *Brain and Language*, *127*(2), 222–229. https://doi.org/10.1016/j.bandl.2012.07.007

Bonini, L., Maranesi, M., Livi, A., Fogassi, L., & Rizzolatti, G. (2014). Space-dependent representation of objects and other's action in monkey ventral premotor grasping neurons. *Journal of Neuroscience*, *34*(11), 4108–4119. https://doi.org/10.1523/JNEUROSCI.4187-13.2014

Bonner, M. F., & Epstein, R. A. (2018). Computational mechanisms underlying cortical responses to the affordance properties of visual scenes. *PLoS Computational Biology*, *14*(4), 1–31.

Borg, I., & Groenen, P. J. F. (2005). Modern Multidimensional Scaling -- Theory and Applications. In *Soviet journal of communications technology & electronics*. Springer.

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., & Peelen, M. V. (2013). Body and object effectors: The organization of object representations in high-level visual cortex reflects body-object interactions. *Journal of Neuroscience*, *33*(46), 18247–18258. https://doi.org/10.1523/JNEUROSCI.1322-13.2013

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. *The 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

Breedin, S. D., Saffran, E. M., & Schwartz, M. F. (1998). Semantic factors in verb retrieval: An effect of complexity. *Brain and Language*, *63*(1), 1–31. https://doi.org/10.1006/brln.1997.1923

Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision*, *13*(10), 1–19. https://doi.org/10.1167/13.10.1

Casiraghi, L., Alahmadi, A. A. S., Monteverdi, A., Palesi, F., Castellazzi, G., Savini, G., Friston, K., Gandini Wheeler-Kingshott, C. A. M., & D'Angelo, E. (2019). I see your effort: Force-Related BOLD effects in an extended action execution-observation network involving the cerebellum. *Cerebral Cortex*, *29*(3), 1351–1368. https://doi.org/10.1093/cercor/bhy322

Cavallo, A., Koul, A., Ansuini, C., Capozzi, F., & Becchio, C. (2016). Decoding intentions from movement kinematics. *Scientific Reports*, *6*(November), 1–8. https://doi.org/10.1038/srep37036

Chen, J., Snow, J. C., Culham, J. C., & Goodale, M. A. (2018). What Role Does "Elongation" Play in "Tool-Specific" Activation and Connectivity in the Dorsal and Ventral Visual Streams? *Cerebral Cortex*, *28*(4), 1117–1131. https://doi.org/10.1093/cercor/bhx017

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3), 455–462. https://doi.org/10.1038/nn.3635

Clarke, A., & Tyler, L. K. (2014). Object-Specific Semantic Coding in Human Perirhinal Cortex. *Journal of Neuroscience*, *34*(14), 4766–4775. https://doi.org/10.1523/JNEUROSCI.2828-13.2014

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

Culham, J. C., Cavina-Pratesi, C., & Singhal, A. (2006). The role of parietal cortex in visuomotor control: What have we learned from neuroimaging? *Neuropsychologia*, *44*(13), 2668–2684. https://doi.org/10.1016/j.neuropsychologia.2005.11.003

de Gelder, B., & Poyo Solanas, M. (2021). A computational neuroethology perspective on body and expression perception. *Trends in Cognitive Sciences*, *25*(9), 744–756. https://doi.org/10.1016/j.tics.2021.05.010

de la Rosa, S., Choudhery, R. N., Curio, C., Ullman, S., Assif, L., & Bülthoff, H. H. (2014). Visual categorization of social interactions. *Visual Cognition*, *22*(9–10), 1233–1271. https://doi.org/10.1080/13506285.2014.991368

Dehaqani, M. R. A., Vahabie, A. H., Kiani, R., Ahmadabadi, M. N., Araabi, B. N., & Esteky, H. (2016). Temporal dynamics of visual category representation in the macaque inferior temporal cortex. *Journal of Neurophysiology*, *116*(2), 587–601. https://doi.org/10.1152/jn.00018.2016

Diedrichsen, J., & Kriegeskorte, N. (2017). Representational models : A common framework for understanding encoding ,. *PLoS Computational Biology*, *13*(4), e1005508.

Dima, D. C., Tomita, T. M., Honey, C. J., & Isik, L. (2022). Social-affective features drive human representations of observed actions. *ELife*, *11:e75027*. https://doi.org/10.7554/eLife.75027

Downey, G., Zaki, J., Berenson, K. R., & Gillig, P. M. (2008). Cognitive Psychology. In *Psychiatry: Third Edition* (Vol. 1). https://doi.org/10.1002/9780470515167.ch26

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

Gallivan, J. P., & Culham, J. C. (2015). Neural coding within human brain areas involved in actions. *Current Opinion in Neurobiology*, *33*, 141–149. https://doi.org/10.1016/j.conb.2015.03.012

Gallivan, J. P., & Goodale, M. A. (2018). The dorsal “action” pathway. In H. B. C. Giuseppe Vallar (Ed.), *Handbook of Clinical Neurology* (Vol. 151, pp. 449–466). Elsevier. https://doi.org/10.1016/B978-0-444-63622-5.00023-1

Gallivan, J. P., McLean, D. A., Flanagan, J. R., & Culham, J. C. (2013). Where one hand meets the other: Limb-specific and action-dependent movement plans decoded from preparatory signals in single human frontoparietal brain areas. *Journal of Neuroscience*, *33*(5), 1991–2008. https://doi.org/10.1523/JNEUROSCI.0541-12.2013

Gallivan, J. P., McLean, D. A., Valyear, K. F., & Culham, J. C. (2013). Decoding the neural mechanisms of human tool use. *ELife*, *2*(e00425), 1–29. https://doi.org/10.7554/eLife.00425

Gauthier, I., Anderson, A. W., Tarr, M. J., Skudlarski, P., & Gore, J. C. (1997). Levels of categorization in visual recognition studied using functional magnetic resonance imaging. *Current Biology*, *7*(9), 645–651. https://doi.org/10.1016/S0960-9822(06)00291-0

Gennari, S., & Poeppel, D. (2003). Processing correlates of lexical semantic complexity. *Cognition*, *89*(1). https://doi.org/10.1016/S0010-0277(03)00069-6

Grafton, S. T., & Hamilton, A. F. d. C. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Human Movement Science*, *26*(4), 590–616. https://doi.org/10.1016/j.humov.2007.05.009

Greene, M., & Oliva, A. (2009). The briefest of glances: the time course of natural scene understanding. *Psychological Science*, *20*(4), 464–472.

Grill-Spector, K., & Kanwisher, N. (2005). Visual RecognitionVisual recognition: as soon as you know it is there, you know what it is. *Psychological Science*, *16*(2), 152–160. https://doi.org/10.1126/science.263.5154.1780

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, *7*(5), 555–562. https://doi.org/10.1038/nn1224

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, *15*(8), 536–548. https://doi.org/10.1038/nrn3747

Grootswagers, T., Zhou, I., Robinson, A. K., Hebart, M. N., & Carlson, T. A. (2022). Human EEG recordings for 1,854 concepts presented in rapid serial visual presentation streams. *Scientific Data*, *9*(1), 1–8. https://doi.org/10.1038/s41597-021-01102-7

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. *Journal of Neuroscience*, *37*(11), 3056–3071. https://doi.org/10.1523/JNEUROSCI.2496-16.2017

Hajibayova, L. (2013). Basic-level categories: A review. *Journal of Information Science*, *39*(5), 676–687. https://doi.org/10.1177/0165551513481443

Hamilton, A. F. d. C., & Grafton, S. T. (2006). Goal representation in human anterior intraparietal sulcus. *Journal of Neuroscience*, *26*(4), 1133–1137. https://doi.org/10.1523/JNEUROSCI.4551-05.2006

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

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425–2430. 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*(August 2019), 116561. https://doi.org/10.1016/j.neuroimage.2020.116561

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

Hétu, S., Grégoire, M., Saimpont, A., Coll, M. P., Eugène, F., Michon, P. E., & Jackson, P. L. (2013). The neural network of motor imagery: An ALE meta-analysis. *Neuroscience and Biobehavioral Reviews*, *37*(5), 930–949. https://doi.org/10.1016/j.neubiorev.2013.03.017

Hoffman, P., Binney, R. J., & Lambon Ralph, M. A. (2015). Differing contributions of inferior prefrontal and anterior temporal cortex to concrete and abstract conceptual knowledge. *Cortex*, *63*, 250–266. https://doi.org/10.1016/j.cortex.2014.09.001

Humphreys, G. W., & Forde, E. M. E. (2005). Naming a giraffe but not an animal: Base-level but not superordinate naming in a patient with impaired semantics. *Cognitive Neuropsychology*, *22*(5), 539–558. https://doi.org/10.1080/02643290442000176

Iordan, M. C., Greene, M. R., Beck, D. M., & Li, F.-F. (2015). Basic Level Category Structure Emerges Gradually across Human Ventral Visual Cortex. *Journal of Cognitive Neuroscience*, *27*(7), 1426–1446. https://doi.org/10.1162/jocn_a_00790

Isik, L., Tacchetti, A., & Poggio, T. (2018). A fast, invariant representation for human action in the visual system. *Journal of Neurophysiology*, *119*(2), 631–640. https://doi.org/10.1152/jn.00642.2017

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

Johnson, K. E., & Mervis, C. B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General*, *126*(3), 248–277. https://doi.org/10.1037//0096-3445.126.3.248

Jozwik, K. M., Kietzmann, T. C., Cichy, R. M., Kriegeskorte, N., & Mur, M. (2023). Deep Neural Networks and Visuo-Semantic Models Explain Complementary Components of Human Ventral-Stream Representational Dynamics. *Journal of Neuroscience*, *43*(10), 1731–1741.

Kadmon Harpaz, N., Flash, T., & Dinstein, I. (2014). Scale-invariant movement encoding in the human motor system. *Neuron*, *81*(2), 452–462. https://doi.org/10.1016/j.neuron.2013.10.058

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, *3*(8), 759–763. https://doi.org/10.1038/77664

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

Krasovsky, A., Gilron, R., Yeshurun, Y., & Mukamel, R. (2014). Differentiating intended sensory outcome from underlying motor actions in the human brain. *Journal of Neuroscience*, *34*(46), 15446–15454. https://doi.org/10.1523/JNEUROSCI.5435-13.2014

Kriegeskorte, N. (2015). Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. *Annual Review of Vision Science*, *1*(1), 417–446. https://doi.org/10.1146/annurev-vision-082114-035447

Kriegeskorte, N., & Mur, M. (2012). Inverse MDS: Inferring dissimilarity structure from multiple item arrangements. *Frontiers in Psychology*, *3*(JUL), 1–13. https://doi.org/10.3389/fpsyg.2012.00245

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(November), 1–28. https://doi.org/10.3389/neuro.06.004.2008

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126–1141. https://doi.org/10.1038/jid.2014.371

Kroczek, L. O. H., Lingnau, A., Schwind, V., Wolff, C., & Mühlberger, A. (2021). Angry facial expressions bias towards aversive actions. *PLoS ONE*, *16*(9 September), 1–13. https://doi.org/10.1371/journal.pone.0256912

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*(11), 571–579. https://doi.org/10.1016/s0166-2236(00)01657-x

Lanzilotto, M., Maranesi, M., Livi, A., Giulia Ferroni, C., Orban, G. A., & Bonini, L. (2020). Stable readout of observed actions from format-dependent activity of monkey's anterior intraparietal neurons. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(28), 16596–16605. https://doi.org/10.1073/pnas.2007018117/-/DCSupplemental

Leshinskaya, A., & Caramazza, A. (2015). Abstract categories of functions in anterior parietal lobe. *Neuropsychologia*, *76*, 27–40. https://doi.org/10.1016/j.neuropsychologia.2015.01.014

Leshinskaya, A., Wurm, M. F., & Caramazza, A. (2020). Concepts of Actions and Their Objects. In *The Cognitive Neurosciences*. https://doi.org/10.7551/mitpress/11442.003.0083

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

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

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

Livi, A., Lanzilotto, M., Maranesi, M., Fogassi, L., Rizzolatti, G., & Bonini, L. (2019). Agent-based representations of objects and actions in the monkey pre-supplementary motor area. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(7), 2691–2700. https://doi.org/10.1073/pnas.1810890116

Loschky, L. C., & Larson, A. M. (2010). The natural/man-made distinction is made before basic-level distinctions in scene gist processing. *Visual Cognition*, *18*(4), 513–536. https://doi.org/10.1080/13506280902937606

Macé, M. J. M., Joubert, O. R., Nespoulous, J. L., & Fabre-Thorpe, M. (2009). The time-course of visual categorizations: You spot the animal faster than the bird. *PLoS ONE*, *4*(6), e5927. https://doi.org/10.1371/journal.pone.0005927

Mack, M. L., Gauthier, I., Sadr, J., & Palmeri, T. J. (2008). Object detection and basic-level categorization: Sometimes you know it is there before you know what it is. *Psychonomic Bulletin and Review*, *15*(1), 28–35. https://doi.org/10.3758/PBR.15.1.28

Mack, M. L., & Palmeri, T. J. (2011). The timing of visual object categorization. *Frontiers in Psychology*, *2*(JUL). https://doi.org/10.3389/fpsyg.2011.00165

Mack, M. L., & Palmeri, T. J. (2015). The dynamics of categorization: Unraveling rapid categorization. *Journal of Experimental Psychology: General*, *144*(3), 551–569. https://doi.org/10.1037/a0039184

Mahendran, A., & Vedaldi, A. (2015). Understanding deep image representations by inverting them. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 5188–5196. https://doi.org/10.1109/CVPR.2015.7299155

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

Margalit, E., Jamison, K. W., Weiner, K. S., Vizioli, L., Zhang, R. Y., Kay, K. N., & Grill-Spector, K. (2020). Ultra-high-resolution fMRI of Human Ventral Temporal Cortex Reveals Differential Representation of Categories and Domains. *Journal of Neuroscience*, *40*(15), 3008–3024. https://doi.org/10.1523/JNEUROSCI.2106-19.2020

Mason, R. L., Gunst, R. F., & Hess, J. L. (2003). Statistical Design and Analysis of Experiments: With Applications to Engineering and Science. In *Journal of the American Statistical Association* (Vol. 85, Issue 412). https://doi.org/10.2307/2289624

Mcmahon, E., Bonner, M. F., & Isik, L. (2023). Hierarchical organization of social action features along the lateral visual pathway. *Current Biology*.

Mervis, C. B., & Crisafi, M. A. (1982). Order of Acquisition of Subordinate-, Basic-, and Superordinate-Level Categories. *Child Development*, *53*(1), 258–266.

Milner, A. D. (2017). How do the two visual streams interact with each other? *Experimental Brain Research*, *235*(5), 1297–1308. https://doi.org/10.1007/s00221-017-4917-4

Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, *46*(3), 774–785. https://doi.org/10.1016/j.neuropsychologia.2007.10.005

Morris, M. W., & Murphy, G. L. (1990). Converging operations on a basic level in event taxonomies. *Memory & Cognition*, *18*(4), 407–418. https://doi.org/10.3758/BF03197129

Moseley, R. L., & Pulvermüller, F. (2014). Nouns, verbs, objects, actions, and abstractions: Local fMRI activity indexes semantics, not lexical categories. *Brain and Language*, *132*, 28–42. https://doi.org/10.1016/j.bandl.2014.03.001

Murphy, G. (2004). The big book of concepts. Boston Review. https://doi.org/10.7551/mitpress/1602.001.0001

Murphy, G. L., & Brownell, H. H. (1985). Category Differentiation in Object Recognition. Typicality Constraints on the Basic Category Advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(1), 70–84. https://doi.org/10.1037/0278-7393.11.1.70

Nastase, S. A., Connolly, A. C., Oosterhof, N. N., Halchenko, Y. O., Guntupalli, J. S., Di Oleggio Castello, M. V., Gors, J., Gobbini, M. I., & Haxby, J. V. (2017). Attention selectively reshapes the geometry of distributed semantic representation. *Cerebral Cortex*, *27*(8), 4277–4291. https://doi.org/10.1093/cercor/bhx138

Nastase, S. A., Goldstein, A., & Hasson, U. (2020). Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage*, *222*(August), 117254. https://doi.org/10.1016/j.neuroimage.2020.117254

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

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *NeuroImage*, *25*(3), 653–660. https://doi.org/10.1016/j.neuroimage.2004.12.005

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

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*(JUL), 1–27. 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.

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*, 1077–1089. https://doi.org/10.1152/jn.00326.2010.

Papeo, L., Agostini, B., & Lingnau, A. (2019). The large-scale organization of gestures and words in the middle temporal gyrus. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *39*(30), 5966–5974.

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., Lingnau, A., Agosta, S., Pascual-Leone, A., Battelli, L., & Caramazza, A. (2015). The origin of word-related motor activity. *Cerebral Cortex*, *25*(6), 1668–1675. https://doi.org/10.1093/cercor/bht423

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

Pillon, A., & d'Honincthun, P. (2011). A common processing system for the concepts of artifacts and actions? evidence from a case of a disproportionate conceptual impairment for living things. *Cognitive Neuropsychology*, *28*(1), 1–43. https://doi.org/10.1080/02643294.2011.615828

Pinker, S. (1989). Learnability and cognition: the acquisition of argument structure. In *The MIT Press* (Vol. 19, Issue 1). https://doi.org/10.1017/s0305000900013726

Pitcher, D., & Ungerleider, L. G. (2021). Evidence for a Third Visual Pathway Specialized for Social Perception. *Trends in Cognitive Sciences*, *25*(2). https://doi.org/10.1016/j.tics.2020.11.006

Portugal, L. C. L., Alves, R. de C. S., Junior, O. F., Sanchez, T. A., Mocaiber, I., Volchan, E., Smith Erthal, F., David, I. A., Kim, J., Oliveira, L., Padmala, S., Chen, G., Pessoa, L., & Pereira, M. G. (2020). Interactions between emotion and action in the brain. *NeuroImage*, *214*(August). https://doi.org/10.1016/j.neuroimage.2020.116728

Proverbio, A. M., & Riva, F. (2009). RP and N400 ERP components reflect semantic violations in visual processing of human actions. *Neuroscience Letters*, *459*(3), 142–146. https://doi.org/10.1016/j.neulet.2009.05.012

Rifkin, A. (1985). Evidence for a basic level in event taxonomies. *Memory & Cognition*, *13*(6), 538–556. https://doi.org/10.3758/BF03198325

Rizzolatti, G., & Matelli, M. (2003). Two different streams form the dorsal visual system: Anatomy and functions. *Experimental Brain Research*, *153*(2), 146–157. https://doi.org/10.1007/s00221-003-1588-0

Rizzolatti, G., & Sinigaglia, C. (2010). *The functional role of the parieto-frontal mirror circuit : interpretations and misinterpretations*. *11*(ApRIl), 264–274. https://doi.org/10.1038/nrn2805

Rogers, T. T., & Patterson, K. (2007). Object Categorization: Reversals and Explanations of the Basic-Level Advantage. *Journal of Experimental Psychology: General*, *136*(3), 451–469. https://doi.org/10.1037/0096-3445.136.3.451

Rosch, E. (1975). Cognitive representations of semantic. *Journal of Experimental Psychology: General*, *104*(3), 192–233.

Rosch, E. (1978). Principles of Categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-braem, P. (1976). Basic Objects in Natural Categories. *Cognitive Psychology*, *8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*(1), 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: an inquiry into human knowledge structures*. Hillsdale, NJ: Erlbaum Lawrence.

Schwanenflugel, P. J. (1991). Why are Abstract Concepts Hard to Understand? In Schwanenflugel P. J. (Ed.), *The Psychology of Word Meanings* (Issue October, pp. 223–248). Lawrence Erlbaum Associates.

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

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

Spunt, R. P., Falk, E. B., & Lieberman, M. D. (2010). Dissociable Neural Systems Support Retrieval of How and Why Action Knowledge. *Psychological Science*, *21*(11), 1593–1598. https://doi.org/10.1177/0956797610386618

Spunt, R. P., Kemmerer, D., & Adolphs, R. (2016). The neural basis of conceptualizing the same action at different levels of abstraction. *Social Cognitive and Affective Neuroscience*, *11*(7), 1141–1151. https://doi.org/10.1093/scan/nsv084

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. *Language Typology and Syntactic Description*, *3*(47), 57–149.

Tanaka, J. W. (2001). The entry point of face recogniton. *Journal of Experimental Psychology. General*, *130*(3), 534–543.

Tanaka, J. W., & Taylor, M. (1991). Object Categories and Expertise: Is the Basic Level in the Eye of the Beholder? *Cognitive Psychology*, *23*(3), 457–482. https://doi.org/10.1016/0010-0285(91)90016-H

Taniguchi, K., Kuraguchi, K., Takano, Y., & Itakura, S. (2020). Object Categorization Processing Differs According to Category Level: Comparing Visual Information Between the Basic and Superordinate Levels. *Frontiers in Psychology*, *11*(March), 1–10. https://doi.org/10.3389/fpsyg.2020.00501

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

Tarhan, L., & Konkle, T. (2020). Sociality and interaction envelope organize visual action representations. *Nature Communications*, *11*(3002), 1–11. https://doi.org/10.1038/s41467-020-16846-w

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

Tunik, E., Frey, S. H., & Grafton, S. T. (2005). Virtual lesions of the anterior intraparietal area disrupt goal-dependent on-line adjustments of grasp. *Nature Neuroscience*, *8*(4), 505–511. https://doi.org/10.1038/nn1430

Turella, L., Rumiati, R., & Lingnau, A. (2020). Hierarchical Action Encoding Within the Human Brain. *Cerebral Cortex*, *30*(5), 2924–2938. https://doi.org/10.1093/cercor/bhz284

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*(May), 118220. https://doi.org/10.1016/j.neuroimage.2021.118220

Vallacher, R. R., & Wegner, D. M. (1985). *A theory of action identification*. Hillsdale, NJ: Lawrence Erlbaum Associates.

van Polanen, V., & Davare, M. (2015). Interactions between dorsal and ventral streams for controlling skilled grasp. *Neuropsychologia*, *79*, 186–191. https://doi.org/10.1016/j.neuropsychologia.2015.07.010

Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*, *40*(1), 183–190. https://doi.org/10.3758/BRM.40.1.183

Wang, J., Conder, J. A., Blitzer, D. N., & Shinkareva, S. V. (2010). Neural representation of abstract and concrete concepts: A meta-analysis of neuroimaging studies. *Human Brain Mapping*, *31*(10), 1459–1468. https://doi.org/10.1002/hbm.20950

Wang, X., Wu, W., Ling, Z., Xu, Y., Fang, Y., Wang, X., Binder, J. R., Men, W., Gao, J.-H., & Bi, Y. (2018). Organizational principles of abstract words in the human brain. *Cerebral Cortex*, *28*(12), 4305–4318. https://doi.org/10.1093/cercor/bhx283

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

Wegner, D. M., & Vallacher, R. R. (1986). Action identification. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (pp. 550–582). Guilford.

Wisniewski, E. J., & Murphy, G. L. (1989). Superordinate and Basic Category Names in Discourse: A Textual Analysis. *Discourse Processes*, *12*(2), 245–261. https://doi.org/10.1080/01638538909544728

Wurm, M. F., Ariani, G., Greenlee, M. W., & Lingnau, A. (2016). Decoding Concrete and Abstract Action Representations During Explicit and Implicit Conceptual Processing. *Cerebral Cortex*, *26*(8), 3390–3401. https://doi.org/10.1093/cercor/bhv169

Wurm, M. F., Artemenko, C., Giuliani, D., & Schubotz, R. I. (2017). Action at its place: Contextual settings enhance action recognition in 4- to 8-year-old children. *Developmental Psychology*, *53*(4), 662–670. https://doi.org/10.1037/dev0000273

Wurm, M. F., & Caramazza, A. (2019). 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. (2022). Two 'what' pathways for action and object recognition. *Trends in Cognitive Sciences*, 1–14. 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

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8619–8624. https://doi.org/10.1073/pnas.1403112111

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. *In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13*, 818–833. https://doi.org/10.1007/978-3-319-10590-1_53

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2018). Places: A 10 Million Image Database for Scene Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(6), 1452–1464. https://doi.org/10.1109/TPAMI.2017.2723009

Zhou, B., Lapedriza, A., Torralba, A., & Oliva, A. (2017). Places: An Image Database for Deep Scene Understanding. *Journal of Vision*, *17*(10), 296. https://doi.org/10.1167/17.10.296

Zhuang, T., Kabulska, Z., & Lingnau, A. (2023). The representation of observed actions at the subordinate, basic and superordinate level. *Journal of Neuroscience*, *October*(5). https://doi.org/10.1523/JNEUROSCI.0700-22.2023

Zhuang, T., & Lingnau, A. (2022). The characterization of actions at the superordinate, basic and subordinate level. *Psychological Research*, *86*(6), 1871–1891. https://doi.org/10.1007/s00426-021-01624-0