Center for Mind/Brain Sciences, University of Trento

Doctoral School in Cognitive and Brain Sciences (XXXVII Cycle 2021 - 2025)

ACTION-RELATED ORGANIZATION OF OBJECT REPRESENTATIONS IN HUMAN VISUAL CORTEX AND TOPOGRAPHIC MODELS

Author

Davide Cortinovis

Supervisor

Stefania Bracci

Summary

Decades of research have shown that object representations in occipitotemporal cortex (OTC) are organized according to multiple spatial and functional principles, including category-selective areas and large-scale topographic maps structured by properties such as animacy, shape, and real-world size. In parallel, a growing body of work has proposed a functional dissociation within OTC, distinguishing a ventral pathway primarily supporting object recognition from a lateral pathway supporting action-related processing. These two lines of research have largely developed independently, leaving open the question of how functional pathway specialization relates to the spatial organization of object representations.

The central aim of this thesis is to bridge these frameworks by identifying an additional organizing dimension based on the action-related properties of objects that helps explain the spatial and functional arrangement of category-selective representations in human visual cortex. Across a series of neuroimaging and computational modelling studies, the thesis demonstrates that action-related information constitutes a fundamental organizing principle of lateral OTC, operating alongside, but independently from, classic dimensions such as animacy, shape, and real-world size.

The first study (Chapter 2) represents the core of the thesis. Here, I show that object representations in lateral OTC are systematically organized according to action-related properties, including graspability and effector properties. Using a stimulus set that orthogonally manipulates animacy and action properties, I demonstrate a smooth topographic gradient of object categories in lateral OTC corresponding to their amount of action-related information. In contrast, ventral OTC exhibits a markedly different organization dominated by animacy, with minimal sensitivity to action dimensions. These findings provide direct evidence that lateral OTC encodes object information in a manner aligned with action-related computational goals. The following studies represent further refinements of the core proposal.

In the second study (Chapter 3), I extend this framework to a recently identified category-selective response to food stimuli. I show that food-selective responses emerge in distinct cortical locations depending on which object properties dominate processing. In ventral OTC, food responses are primarily explained by surface-level features such as colour and ensemble statistics. In lateral OTC, food-selective responses align with action-related properties shared with other graspable objects. This dissociation provides a strong test of the action-based framework and demonstrates its predictive power beyond the canonical category distinctions.

In the third study (Chapter 4), I shift focus from spatial organization to functional selectivity. Using a combination of image-level functional selectivity analysis and encoding models, I show that body-, hand-, and tool-selective areas exhibit robust and dissociable functional tuning. Areas selective for the same category differ systematically in their feature sensitivity depending on their position within ventral versus lateral OTC and across hemispheres. These results indicate that category selectivity in OTC reflects multiple, partially dissociable computational roles rather than a single shared representational code.

Finally, throughout the thesis I evaluate whether current artificial neural networks (including recent topographic models) capture the same organizing principles observed in the brain. While these models successfully reproduce aspects of ventral OTC organization (i.e., related to animacy, shape, or other mid-level properties), they consistently fail to account for the action-based organization of lateral OTC. This modelling gap highlights current limitations of topographic models and motivates future work aimed at incorporating behaviorally relevant constraints and inductive biases.

Together, the findings of this thesis demonstrate that understanding object organization in visual cortex requires moving beyond purely visual dimensions and incorporating action-related object properties as a core organizing principle. By unifying spatial topography, functional selectivity, and computational modelling, this work provides a more complete

account of how object representations are structured in human visual cortex and offers concrete directions for narrowing the gap between biological and artificial systems.

Acknowledgements

To fully appreciate all the people that helped me throughout these years I need at least three languages:

First, I would like to thank my supervisor, Stefania Bracci, for these incredible and challenging years. You were the best supervisor I could have asked for, and your scientific and personal mentorship was what (literally) pushed me through the PhD.

Second, allow me to thank my family and friends in Italian: grazie ai miei genitori – Monica e Ivan – e alle mie sorelle – Federica e Michela – del vostro supporto, anche se spesso non capivate cosa stessi facendo (sicuramente per i miei limiti nello spiegare il mio progetto di ricerca). E un grazie anche alle mie amiche – Paola, Valentina, e Corinne; entrare nel mondo del lavoro (e degli adulti) sarebbe stato ancora più difficile senza di voi.

Finally, thank you Nhựt for the incredible patience you have shown in these years. Luận văn này được dành tặng cho anh.

Contents

Chapter 1 - Introduction

1.1 Background

Vision is our most powerful source of information about the external world, and the cortical systems that support it are both highly complex and remarkably structured. Indeed, since the earliest anatomical and electrophysiological studies, researchers have recognized that the visual cortex is not a homogeneous sheet but a highly organized mosaic of areas and functional clusters. One of the major organizational principles of visual cortex is the division between two streams of information, a ventral visual stream processing object information and a dorsal stream processing object location and action guidance (Baker & Kravitz, 2024; Goodale & Milner, 1992; Ungerleider & Mishkin, 1982). The focus of the current work is on the spatial and functional organization of the ventral visual stream.

The ventral visual stream comprises a series of hierarchically interconnected cortical areas, extending from primary visual cortex (V1) to high-level occipitotemporal cortex (OTC). Two hallmark properties of the ventral visual stream are hierarchical organization and the presence of functionally meaningful clusters (Grill-Spector & Malach, 2004; Felleman & van Essen, 1991). Specifically, regions along the stream encode progressively more complex visual features, from line orientations and edges in V1 to shapes, objects, and scenes in high-level visual cortex (Di Carlo et al., 2012). Moreover, neurons that respond to these functionally similar inputs are clustered together within the OTC (Op de Beeck et al., 2008; Tanaka, 1996). In high-level visual cortex, multiple spatial organizational principles coexist. For example, voxels selective for faces form a series of face-selective areas (Kanwisher et al., 1997; Grill-Spector et al., 2017). Within these areas, finer-grained micro-organization has been observed, such as clustering for specific facial features (Brants et al., 2011; de Haas et al., 2021; Henriksson et al., 2015; Sato et al., 2013). Broader organizational gradients also exist: for instance, voxels that prefer faces tend to include and lie near voxels preferring animate or

small objects in general, reflecting continuous representational transitions across cortex (Kriegeskorte et al., 2008; Konkle & Oliva, 2012). The nested structure of object information can be visualised in Figure 1.1a.

More recent models propose that the ventral visual stream can be further subdivided into multiple pathways: a ventral pathway (including the fusiform gyrus, collateral sulcus, and parahippocampal gyrus) extracts features that support object processing; a more lateral pathway (including the inferior temporal gyrus, lateral occipital sulcus, and superior temporal sulcus) is involved in the processing of (inter)actions (Haak & Beckmann, 2018; Lingnau & Downing, 2015; Pitcher & Ungerleider, 2021; Ritchie et al., 2024b; Weiner & Grill-Spector, 2013; Wurm & Caramazza, 2022). This lateral pathway is organized along two functional gradients: one that processes progressively more conceptual representations of actions (from posterior-to-anterior), and the other that distinguish actions involving inanimate objects from social actions, such as communicative interactions between biological agents (from inferior to superior LOTC; Wurm et al., 2017; Wurm & Lingnau, 2015). Figure 1.1b-c illustrates these anatomical landmarks and the subdivision of the ventral and lateral pathways.

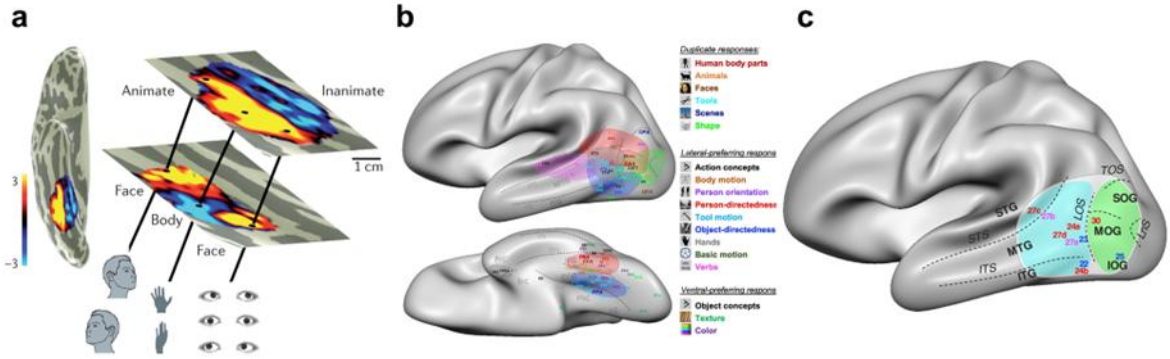


Figure 1.1. Spatial and functional organization of high-level visual cortex. a) Nested functional representations in ventral visual cortex. The top map depicts the broad animacy division; the middle map shows embedded body- and face-selective clusters within the animate sector; and the bottom map illustrates finer-grained specificity for face- and body-related features. Adapted from Grill-Spector & Weiner (2014). **b)** Lateral (top) and ventral (bottom) views of the visual cortex showing major anatomical landmarks and category-selective areas, together with schematic functional responses along both pathways. Adapted from Wurm & Caramazza (2022). **c)** Enlarged view of the lateral pathway

highlighting areas involved in perceptual (green) and conceptual (light blue) processing of actions. Adapted from Wurm & Caramazza (2022).

Aside from neuroimaging, a core approach for understanding visual cortex organization is computational modelling (for reviews see Kriegeskorte, 2015; Kriegeskorte & Douglas, 2018). While neuroimaging has been instrumental in characterizing the spatial and functional layout of ventral visual cortex, it is inherently limited in explaining *why* this organization emerges. Computational models provide a complementary approach by allowing researchers to test how specific task demands, architectural constraints, and other inductive biases give rise to particular spatial and representational structures (Kay, 2018). In visual neuroscience, such modelling efforts are dominated by Artificial Neural Networks (ANNs; LeCun et al., 2015). Although there is ongoing debate about the extent to which ANNs can offer good mechanistic explanations of visual processing (Bowers et al., 2023; Doerig et al., 2023; Richards et al., 2019), their value lies in the possibility to systematically manipulate inductive biases shaping them and observe the resulting organization (Kietzmann et al., 2019). Importantly, abstraction in ANNs is not necessarily a limitation but often represents an advantage, as omitting biological details allow researchers to understand which constraints are important for particular visual tasks (Cao & Yamins, 2024a; Lindsay, 2021). More broadly, ANNs can help explain why the ventral visual stream is organized the way it is by linking the specific “ingredients” of an ANN (task, architecture, visual diet) with the specific representational space that emerges when modifying them (Cao & Yamins, 2024b; Kanwisher et al., 2023). For example, work with ANNs has shown that functional specialization for faces and their behavioral signatures are a consequence of optimizations to solve face-specific tasks (Dobs et al., 2022; 2023). Similarly, the division of visual processing into multiple pathways can spontaneously arise as a consequence of computational optimization for distinct functional demands (Jacobs et al., 1991). As I discuss in a later section, such modelling approaches have been extended to

investigate which biologically inspired inductive biases are required to reproduce the topographic organization observed in ventral visual cortex (Margalit et al., 2024).

In this thesis, I build on the growing body of neuroimaging and computational work focused on the ventral visual stream by proposing additional organizing principles that shape the spatial and functional architecture of high-level visual cortex, with a particular focus on the features that distinguish the ventral and lateral pathways. I also test if current state-of-the-art (topographic) computational models can account for this organization and discuss their potentialities and limitations. To provide the necessary background, I will first introduce the concept of cortical topography and briefly consider why topographic organization arises in the first place. I will then review neuroimaging evidence on OTC spatial organization, highlighting principles and frameworks proposed to explain it. After this, I will outline limitations of the existent neuroimaging evidence in explaining underexplored aspects of OTC organization, finally followed by an overview of computational models that attempt to capture OTC structure, thereby motivating the central proposal advanced in this work.

1.2 The spatial and functional organization of the ventral visual stream

The mammalian cerebral cortex contains a big variety of topographic maps, where neurons that are physically close to each other respond to similar functional inputs (Silver & Kastner, 2009). More formally, topography can be defined as a mapping between continuous stimulus dimensions and continuous cortical space, such that neighbouring neurons exhibit correlated response patterns (Durbin & Mitchinson, 1990; Mountcastle, 1997; Patel et al., 2014). Classic examples include the “sensory homunculus” in primary somatosensory cortex (Penfield &

Boldrey, 1937; Saadon-Grosman et al., 2020) and the tonotopic gradient in auditory cortex (Brewer & Barton, 2016; Merzenich et al., 1975; Saenz & Langers, 2014).

Primary visual cortex (V1) provides one of the clearest and most extensively studied examples of cortical topography. Its retinotopic map preserves the spatial layout of the visual field, while orientation selectivity is organized in a characteristic pinwheel-like structure (Bonhoeffer & Grinvald, 1991; Blasdel, 1992; Paik & Ringach, 2012). Additional organizational features include ocular dominance columns, spatial frequency maps, and colour-selective “blob” regions (Livingstone & Hubel, 1984; Lu & Roe, 2009; Nauhaus et al., 2012).

Understanding whether similar principles extend beyond early sensory cortex has been made possible by the advent of non-invasive neuroimaging, particularly functional magnetic resonance imaging (fMRI), alongside increasingly sophisticated computational models. fMRI has enabled the discovery of numerous topographic maps within the ventral visual stream. Parallel developments in computational neuroscience produced mechanistic accounts of how such organization may arise, both through biological constraints (e.g., wiring length minimization; Chklovskii & Koulakov, 2004) and through experience-dependent learning (Kohonen, 1982). This work has helped give us insights into the spatial and functional organization of the ventral visual stream. Indeed, representations within these regions support object perception and thus must map an extremely high-dimensional feature space onto the cortex, incorporating for instance features such as texture, color, shape, semantics, affordances, etc. (Weiner & Grill-Spector, 2014). This raises a fundamental question: which feature dimensions guide the mapping of object space onto the two-dimensional cortical sheet of OTC?

Researchers have proposed a wide range of dimensions that might structure this map, from low-level visual factors such as spatial frequency and eccentricity, to mid-level properties such as curvature and texture, to high-level dimensions like animacy and real-world size (Bracci & Op de Beeck, 2023; Jagadeesh & Gardner, 2022; Konkle & Caramazza, 2013; Konkle & Oliva,

2012; Yue et al., 2020). These larger-scale topographic maps are related to another fundamental type of organization: category selectivity. In this case, neurons (or voxels) that respond to similar functional inputs (a given category) cluster together in visual cortex (Downing et al., 2006; Kanwisher et al., 2010; Reddy & Kanwisher, 2006). Within each area, images belonging to the preferred category elicit much higher responses than images of other categories (Downing et al., 2006; Kanwisher, 2010; Mur et al., 2012; Tsao et al., 2006). Category-selective areas have been discovered only for a small number of categories, such as faces (Kanwisher, 1997), bodies (Downing & Kanwisher, 2001; Peelen & Downing, 2005), hands (Bracci et al., 2010), words (McCandliss et al. 2003), scenes (Epstein & Kanwisher, 1998), tools (Chao et al., 1999), and – most recently – food (Henderson et al., 2025). The spatial arrangement of these areas is far from arbitrary: for example, face-selective areas occupy regions of cortex that respond – to a certain extent – to stimuli that typically fall within the fovea, are curvilinear, small, and animate, all features that are characteristic of faces (Arcaro & Livingstone, 2024). This raises a long-standing chicken-and-egg question: do category-selective areas arise as a consequence of superimposition of multiple topographic maps (Arcaro & Livingstone, 2021; 2024; Op de Beeck et al., 2008), or sensitivity to these features are a by-product of category clustering itself (Kanwisher et al., 2010)? This debate has shaped visual neuroscience for years, but, while the directionality remains controversial, the interdependence between category selectivity and larger-scale topographic maps is now well established.

In what follows, I summarize evidence for four of the best-studied organizing dimensions, from low-level (eccentricity) to mid-level (shape) to higher-level (animacy and real-world size). I then look at broader theoretical frameworks that encompass these and other dimensions and that proposed more general explanations for the spatial organization of OTC. Figure 1.2 presents an overview of the main dimensions and organization found in ventral visual cortex.

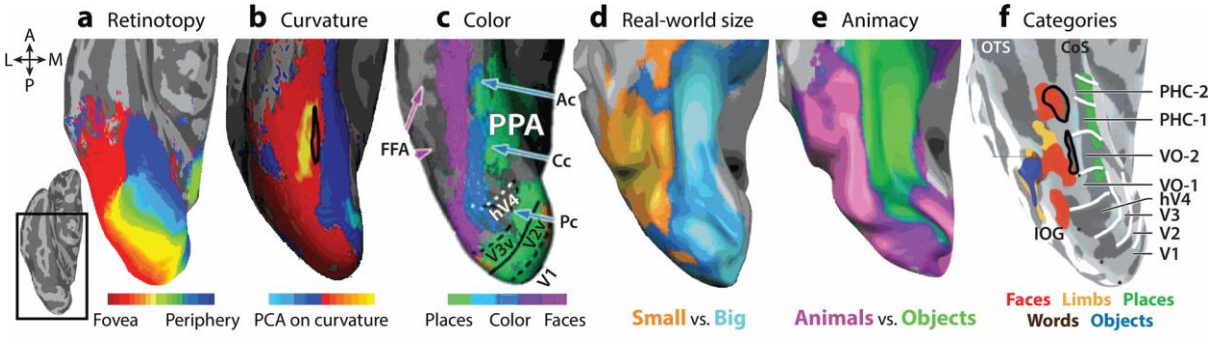


Figure 1.2. Topographic organization of ventral OTC. Several large-scale spatial organizations have been proposed for ventral OTC, including maps of (a) retinotopy, (b) curvature, (c) colour, (d) real-world size, (e) animacy, and (f) object category. Category-selective areas systematically intersect these broader topographic gradients: for example, face-selective areas (shown in red in panel f) lie within regions of cortex that preferentially process foveal, curvilinear, small, and animate stimuli. Notice that, with the exception of colour, these dimensions are mirrored on the lateral side of OTC (not shown). Adapted from Arcaro & Livingstone (2024).

Eccentricity

One of the most influential and empirically supported topographic constraints is eccentricity, the typical retinal distance at which a stimulus category is viewed. Early work proposed that the spatial layout of category-selective areas reflects these characteristic viewing patterns (Levy et al., 2001; Malach et al., 2002). For instance, faces, which we foveate, and scenes, which are typically viewed peripherally, activate distinct regions in ventral visual cortex, the lateral fusiform gyrus and the parahippocampal gyrus\collateral sulcus, respectively, that are associated with distinct eccentricity bands. This framework was extended to explain the emergence of other categories, such as for words and tools, whose position is consistent with the eccentricity framework (Hasson et al., 2002). Eccentricity also explains the fact that category-selective areas are mirrored across ventral and lateral OTC (Hasson et al., 2003; Silson et al., 2015) and can predict the emergence of selectivity for novel domains (Arcaro et al., 2017; Gomez et al., 2019; Srihasam et al., 2014).

However, the explanatory power of the eccentricity dimension is limited by some considerations. First, the typical position at which object categories fall within the retina might not explain other categories beyond faces, scenes, and words: for instance, it is unclear why body-parts – which fall at more peripheral regions on the retina compared to faces and words – elicit responses at more lateral locations than face-selective areas. This mismatch has motivated a revisiting of the classic medial-to-lateral eccentricity-band model (Daniel-Hertz et al., 2025), where the lateral fusiform gyrus is foveally biased, and peripheral eccentricity biases extend both medially and laterally, a model that however does not account for the fact that body-selective areas border word-selective areas in the fusiform gyrus \ occipitotemporal sulcus (e.g., Pillet et al., 2024a). A second limitation is revealed by studies of blindness: congenitally blind individuals exhibit a similar arrangement of category-selective areas despite the lack of visual experience with those categories (van den Hurk et al., 2017; Ratan-Murty et al., 2020). Such findings indicate that eccentricity – while representing a fundamental scaffolding principle (Groen et al., 2022; Kubota et al., 2025) – cannot be the sole determinant of category-related spatial organization and point instead to intrinsic constraints, such as long-range connectivity (see below), that may influence the large-scale organization of OTC (Mahon & Caramazza, 2011).

Shape

Shape constitutes a rich mid-level feature that co-varies with object category. Two major operational definitions of shape have been found to play a role in establishing the emergence of category in OTC: curvature and aspect-ratio.

Electrophysiological recordings in macaque IT revealed patches preferring curved over rectilinear stimuli (Yue et al., 2014), a pattern mirrored in human OTC (Nasr et al., 2014; Yue et al., 2020). These curvature-sensitive patches intersect with category-selective areas: scene-selective cortex overlaps rectilinear-preferring zones, while face-selective areas lie adjacent to curvilinear-preferring zones. This suggests that curvature serves as a mid-level

feature that shapes where categories emerge, even if evidence already shows that category information cannot be reduced to its underlying shape (Bougou et al., 2024; Bracci et al., 2016; Proklova et al., 2016).

A seminal study found that another aspect of shape is fundamental in determining the organization of primate object space: aspect-ratio (Bao et al., 2020). In this study, the macaque IT cortex could be divided into a four-quadrant object space whose main dimensions represented animacy and stubbiness-spikiness; object categories fall into one of the quadrants of this object space: for instance, face-selective areas occupy the animate-stubby quadrant, and body-selective areas the animate-spiky quadrant. In humans, a coarser map of object space based on similar properties (animacy and aspect ratio) has been reported (Coggan & Tong, 2023). However, other work using shape-controlled image sets argues that animacy and the face–body distinction together explain more variance in OTC organization than aspect ratio per se (Ritchie et al., 2021; Yargholi & Op de Beeck, 2023).

Animacy and real-world size

Higher-level distinctions also shape OTC organization. The animacy division has emerged as one of the most replicable dimensions in both humans and macaques (Kriegeskorte et al., 2008; Sha et al., 2015; Thorat et al., 2019), with a medial-to-lateral organization of inanimate and animate preferring cortex in ventral OTC, mirrored on the lateral OTC. A separate proposed dimension distinguishes small vs. large inanimate objects (Konkle & Oliva, 2012). Based on this evidence, Konkle & Caramazza (2013) proposed a tripartite organizational structure in OTC, where distinct regions are sensitive to either animate or inanimate stimuli, with further divisions for inanimate objects-responsive regions based on their preference for either small or big objects. Once again, category-selective areas intersect these maps: face-selective areas are associated with regions selective to animate and small-preferring objects, and the opposite pattern is observed for scene-selective areas.

What is the content of animate-inanimate or big-small object-preferring regions? Do they compute a high-level representation of animacy and real-world size? Some evidence shows that this might not be the case, and that animacy and real-world size can be partially reduced to mid-level properties of the objects. Specifically, a study employed texform stimuli – unrecognizable images that preserve the coarse texture and shape of the objects – and found that OTC largely exhibits similar responses and spatial organization by animacy and real-world size even when objects are unrecognizable (Long et al., 2018). Behavioral analyses using the same stimuli showed that the preservation of responses and of the larger-scale organization by animacy and real-world size may be due to curvature features of the objects, that are sufficient to categorize objects as either animate or inanimate, or either big or small (Long et al., 2016; 2017). In a similar fashion, complementary studies using other forms of texture scrambling (Gatys et al., 2015) similarly suggest that texture-based (rather than shape-based) statistics may play a larger role in ventral visual representations than previously assumed (Jagadeesh & Gardner, 2022; Jagadeesh & Livingstone, 2024; see also Ayzenberg & Behrmann, 2022).

Overall, animacy and real-world size represent robust, replicable, and interacting topographic gradients in high-level visual cortex, intersecting with category-selective areas; however, their emergence may partially rely on mid-level visual features (such as curvature) aside from purely conceptual distinctions.

Connectivity-based constraints

An alternative but not mutually exclusive account stresses the role of top-down connectivity in shaping OTC spatial organization. According to this perspective, the location of category-selective areas in OTC is strongly influenced by long-range connections with downstream systems specialized for processing particular types of information (Mahon & Caramazza, 2011).

Findings from congenital blindness provide support for this view. Indeed, despite lacking visual experience, blind individuals show the typical organization of face-, body-, tool-, and word-selective areas, as well as preserved animate–inanimate and real-world size-based topographic patterns (He et al., 2013; Kitada et al., 2014; Mahon et al., 2009; Peelen et al., 2013; Ratan-Murty et al., 2020; Reich et al., 2011; Striem-Amit et al., 2012; van den Hurk et al., 2017).

Neuroimaging evidence in sighted individuals has confirmed that the location of some category-selective patches can be predicted from their structural and functional connectivity profiles. For instance, taking once again the example of faces, studies investigating anatomical connectivity found that the position of face-selective areas in an individual can be predicted from their anatomical connectivity patterns (Saygin et al., 2012), and, more specifically, face-selective areas show privileged connectivity with social cognition networks in medial prefrontal cortex (Powell et al., 2018; however, see Scott & Arcaro, 2023, for a different perspective). Similarly, the position of the visual word form area can be predicted from its connectivity to language regions even before the area develops during literacy acquisition (Saygin et al., 2016). Connectivity can also explain the distinct lateralization observed for specific categories, most notably the dissociation between words and faces, which occupy a similar (albeit differentiable) position in right and left ventral OTC: specifically, this dissociation has been explained by the distinct connectivity they form with systems involved in language (for words) and social processing (for faces; Blauch et al., 2025; Rajimehr et al., 2022). Finally, over and above categories, a study also found that the tripartite organization by animacy and real-world size is associated with a tripartite pattern of functional connectivity as measured with resting-state fMRI data (Konkle & Caramazza, 2017).

Behavioral goals

Other proposals emphasize the primacy of behavioral goals (Bracci & Op de Beeck, 2023). According to this framework, to understand visual cortex organization we should move beyond

mapping fixed object-categories to dedicated cortical regions, and instead ask: which behaviorally relevant information derived from those object categories supports our interactions with the world? In other words, the computational goal of an area selective for a specific category is not to simply recognize an object as belonging to a given category, but rather to extract features that support the specific behavioral goal that that category affords. For example, instead of assuming that the role of the face-selective areas is to categorise objects as faces or non-faces, we should consider that faces are often used for social functions, such as for recognizing identity, inferring gaze, emotion, intention, and social interaction. Thus, visual cortex might be organized around the behavioral relevance of stimuli (social, manipulation, navigation, etc.) rather than their categorical identity per se (Bracci & Op de Beeck, 2023; Peelen & Downing, 2017). In a similar way, a recent proposal argued for the necessity to move beyond the category-centric framework altogether and instead focus on mapping specific behavioral goals to distinct regions of cortex (Ritchie et al., 2025). A related but alternative proposal (Contier et al., 2024) argued that the entire ventral visual stream codes for behaviorally-relevant dimensions among which category-selective areas may represent a special case within the broader multidimensional landscape, with sparse responses to only a limited number of dimensions (that are associated with the preferred category).

In general, a dimensional view based on behavioral-relevant properties of objects and a category-focused approach are not necessarily in opposition, but rather categories contain important information to understand the type of behavioral goal supported by an area (van Dyck & Dobs, 2025; van Dyck et al., 2025).

Summary

In this section, I reviewed several major principles that have been proposed to shape the spatial and functional organization of high-level visual cortex. This was not intended as an exhaustive list of all dimensions or frameworks explaining high-level vision; rather, the ones reviewed are the most important to understand how cognitive neuroscientists have thought

about the spatial and functional organization of OTC, and they are the one that have been most used to develop and test artificial neural networks (see below). An important consideration emerging from this literature is that no single principle can account for the full richness of OTC organization. Instead, multiple constraints operate in parallel, and it is their interaction that produces the mosaic of category-selective responses observed across cortex (Op de Beeck et al., 2008; 2019).

In the next section, I evaluate the capacity of each organizing principle to account for the arrangement of another, often understudied, set of category-selective areas, extending the discussion beyond the canonical categories typically emphasized in the literature.

1.3 Hand and tool selectivity in visual cortex

Part of the content of this section has been published: Cortinovis, D., Peelen, M. V., & Bracci, S. (2025b). Tool representations in human visual cortex. *Journal of Cognitive Neuroscience, 37*(3), 515-531.

A particularly compelling test case for behavioral-goal accounts of visual cortex organization comes from categories whose visual form alone does not fully explain their cortical selectivity, most notably hands and tools. Unlike canonical categories such as faces or scenes, hands and tools are not easily characterized by simple visual dimensions such as shape or curvature, nor do they occupy a privileged position along known topographic axes (e.g., they differ based on animacy). Instead, these categories are defined by their action-related properties. As such, their representation and location in visual cortex provide a testbed to evaluate whether high-level visual cortex reflects the extraction of behaviorally relevant information that supports specific goals, such as grasping, tool use, and, more generally, object interaction. In the following section, I review evidence from hand- and tool-selective regions in visual cortex,

arguing that their organization and representational content are more naturally explained within a behavioral-goal framework than by category identity or visual dimensions alone.

Hands

Possibly the first discovered category-selective neuron in macaque IT cortex was the hand-selective neuron reported by Gross and colleagues (1969; see also Gross, 2008 for a historical view). However, after the advent of non-invasive neuroimaging, the field focused on identifying the cortical locations of faces first and bodies later (see Kanwisher, 2025 and Vogels, 2022 for recent reviews), with work investigating visual cortex responses to “bodies” and “body-parts” interchangeably (with a few partial exceptions, see for instance Astafiev et al., 2004 and Taylor et al., 2007). To simplify this considerable amount of work, results show two areas selective for bodies in ventral and lateral OTC, called the Extrastriate Body Area (EBA; Downing et al., 2001) and the Fusiform Body Area (FBA; Schwarzlose et al., 2005; Peelen & Downing, 2005) respectively (see Peelen & Downing, 2007 for a review).

Most importantly for the current work, a series of studies by Weiner & Grill-Spector (2010; 2011) found a topographic organization of body (or limb) selective patches and face selective patches in human ventral and lateral OTC. Specifically, they found multiple limb selective areas (and not a single EBA area) arranged in a ring-like fashion around the motion area, and a more ventral activation nearby the occipitotemporal sulcus (corresponding with FBA). These multiple body patches are always near face-selective areas, suggesting a new organizational principle linking the two categories (Weiner & Grill-Spector, 2013). The authors, however, used images of limbs and bodies interchangeably, thus not allowing them to disentangle potential differences in location between hands and other body-parts.

The first study reporting hand-selective activations in humans was conducted by Bracci and colleagues (2010). The authors presented to participants under the fMRI scanner images of various body-parts, including hands, feet, and the whole body. They found a left-lateralized area within the lateral OTC responding selectively to images of hands, that could be

dissociated from an area preferring whole-body images. Other studies focused on identifying further parcellations of body-parts: for example, Orlov et al. (2010) found a topographic organization of body-parts in ventral and lateral OTC, with large portions dedicated to the processing of hands and upper limbs (vs. other body-parts). Multivariate analyses showed that the body-parts organization within this region are clustered according to semantic similarity over and above shape similarity (Bracci et al., 2015, see also Mazurchuk et al., 2024). Separable responses to hands and bodies were also observed using time-resolved methods (Moreau et al., 2018), most notably a left-lateralized component selective for hands and not bodies that was found using the same stimulus set as Bracci and colleagues (Santo et al., 2017). Recent evidence provided the highest possible resolution at present for dissociable hand- and body-selective responses: a 7 Tesla fMRI experiment reports the presence of multiple hand-selective clusters in ventral and lateral OTC (Pillet et al., 2024a), and hand selective neurons were recorded in the visual cortex of a patient with implanted electrodes in the vicinity of the body area (Ramirez et al., 2024).

Studying dissociations between hands and other body-parts does not only give us a better picture of the organization of the brain; rather, it can inform us on the representational content that distinguishes these parcellations of cortex. Indeed, subsequent studies found that hand-selective areas have distinct computational roles compared to nearby body and object areas. For instance, a study suggested that the hand areas (but not the nearby body or object preferring areas) are involved in representing view invariant hand postures, including those conveying communication- and action-related information (Bracci et al., 2018). Another study found a role of the hand area (but not the nearby tool or hand area) in representing hand postures needed to grasp tools (Knights et al., 2021; see also Moreau et al., 2023). Finally, hand- and body-selective areas can also be dissociated based on their distinct patterns of functional connectivity with regions outside of visual cortex (Bracci et al., 2012), and recent evidence shows that these areas even have distinct developmental trajectories (Nordt et al., 2021).

Tools

Tool-related activations have been reported throughout occipitotemporal, parietal, and frontal cortices, but for the present discussion the most relevant regions lie in lateral OTC, along the inferior temporal gyrus, and in ventral OTC around the medial fusiform gyrus \ collateral sulcus.

Univariate and multivariate analyses reveal a dissociation between these two areas in both their functional selectivity and their representations of tool-related properties. Studies have shown that the lateral tool area exhibits reliable category selectivity. Specifically, it responds more strongly to tools than to visually similar non-tool graspable objects (Mruczek et al., 2013), and its selectivity cannot be explained by mid-level visual features such as elongation or real-world size (Chen et al., 2018; Macdonald & Culham, 2015). These responses are present even in congenitally blind individuals (Peelen et al., 2013). However, the selectivity of the lateral tool area shows a lower degree of categorical step boundary than that of other high-level category-selective areas, such as those for bodies, scenes, or faces (Downing et al., 2006; Mur et al., 2012). From a functional point of view, the lateral tool-selective area appears to be tuned to both shape features and action-relevant features (e.g., the type of action a tool affords; Wu et al., 2020), including those that determine how an object functions as an end-effector (Bracci et al., 2013; see next section).

By contrast, the medial region does not show strong univariate tool selectivity (Mahon et al., 2007). Instead, it responds broadly to inanimate objects and is particularly engaged by large, non-manipulable objects, likely due to its proximity with the parahippocampal place area (Epstein & Kanwisher, 1998; He et al., 2013). Ventral OTC areas are most sensitive to surface and material properties (Hiramatsu et al., 2011; Komatsu & Goda, 2018), features shared across many inanimate object categories, rather than supporting directly action-related properties of objects (but see Mahon et al., 2007).

The controversy surrounding the presence of tool-selective areas in visual cortex (e.g., Downing et al., 2006; Khosla et al., 2022) may also be related to the difficulty of defining what

constitutes a “tool”. Unlike faces or bodies, categories whose membership is perceptually intuitive, tools do not lend themselves to a simple category boundary. Many everyday objects may intuitively appear as tools (e.g., smartphones or remote controls), yet may not be represented as tools by the brain when judged on the basis of their action-related properties or their relationships to the hand. For the present work, we define tools as “hand-held objects that physically or directly act on another object or surface”. This can give us an operational definition of tools that we can test experimentally and, eventually, refine (see Chapter 4). Moreover, it allows us to distinguish tools from similar objects: tools are end-effectors, “act-with” objects used to perform actions on other objects (e.g., hammers, knives, scissors, pliers…); this differentiates them from generally manipulable objects, that are “acted-upon” (books, mugs, bottles, smartphones…); finally, our definition is different compared to other definitions, such as that of Mahon et al. (2007) that classify tools based on their general function rather than purely on their action-related properties. For example, specific manipulable objects, such as a glass, would be considered tools following the definition of Mahon et al. (2007), whereas, as they are not effectors, our definition consider them only generally graspable objects, with the prediction that they would not elicit tool-related responses to the same extent as more canonical tools (e.g., hammers, scissors).

Hands and Tools

When we perform specific types of actions, for example when we use a hammer, hands and tools are perceived together as a single functional entity, whereby the hammer becomes an extension of the hand. This functional similarity between hands and tools is reflected in the organization of the brain.

A seminal study identified overlapping activations between hands and tools in left lateral OTC (Bracci et al., 2012). Participants were presented with images of various categories, including hands, whole-bodies, tools, and non-tool objects; the left lateral OTC exhibited clusters selective for hands and tools, and these clusters partially overlapped; no such overlap was

found between tools and other body-parts (including whole bodies). Moreover, functional connectivity analyses showed that the hand-tool overlap area, but not the nearby body area, was more strongly connected with regions in left parietal and premotor cortex involved in object-directed actions.

The proximity and partial overlap between hands and tools have been replicated across multiple experiments, including in studies of individuals born without hands (Striem-Amit et al., 2017), and, most recently, in a high-resolution 7T fMRI study on the native unsmoothed surface of single subjects (Pillet et al. 2024b). Data-driven analyses on a large-scale fMRI dataset similarly revealed a component selective to hands and hand actions, including those involving objects (Marvi et al., 2024).

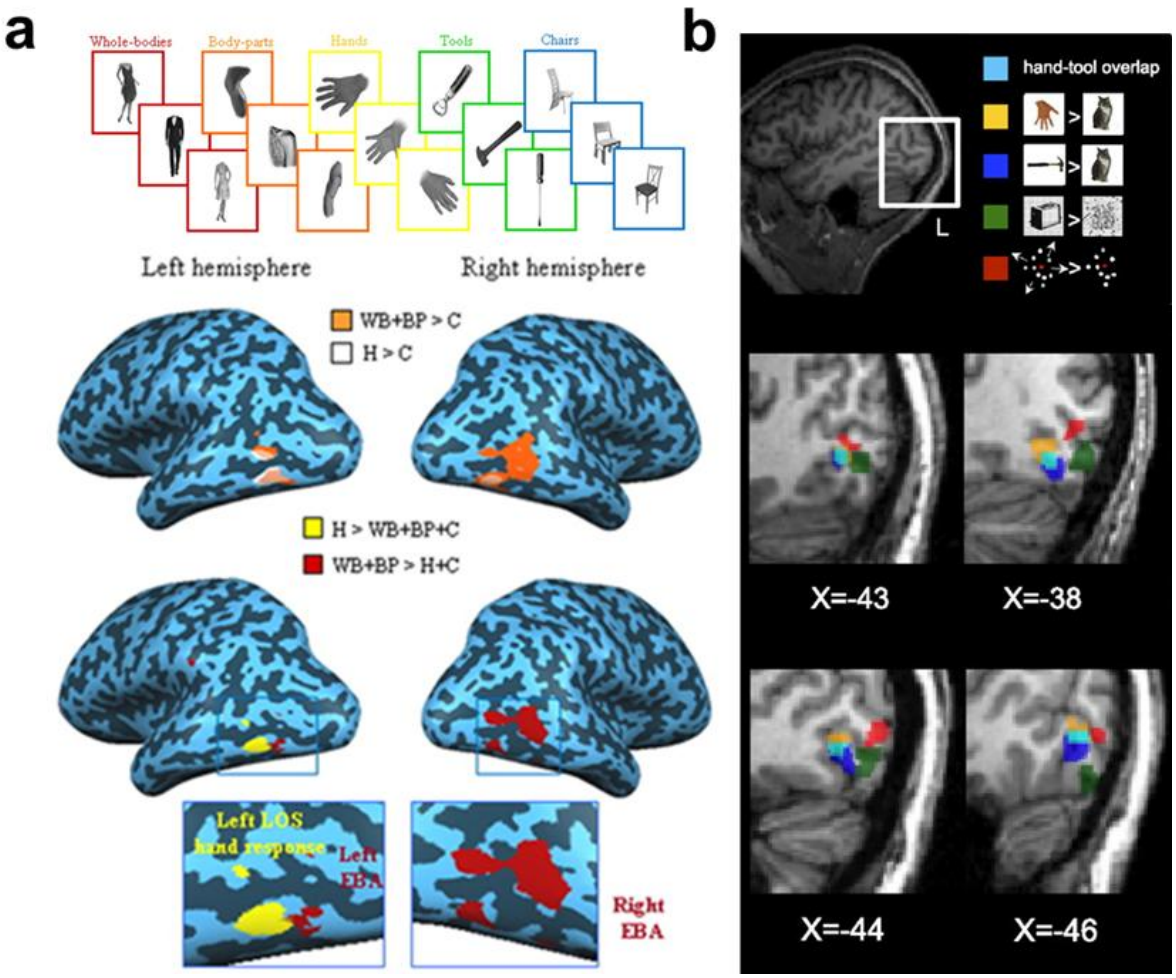


Figure 1.3. The topographic organization of hand- and tool-selective areas in visual cortex. a) Dissociable responses to hands (yellow) and whole-bodies and other body-parts (red) in left lateral OTC. Adapted from Bracci et al. (2010) **b)** Overlapping activations between hand- (yellow) and tool-responses (blue) in left lateral OTC of four representative participants. Adapted from Bracci et al. (2012).

Importantly, while higher-resolution acquisitions (e.g., at the single-neuron level) may ultimately separate activations to hands and tools, their anatomical proximity remains important: first, anatomical proximity reflects functional similarity (as a consequence of the pressure to minimize wiring length and increase processing efficiency; Reddy & Kanwisher, 2006); thus, the spatial clustering of hand and tool responses likely reflects shared computational demands.

Classic dimensions proposed to organize object categories (e.g., animacy, shape, real-world size, eccentricity) cannot account for this arrangement: hands and tools differ in shape and animacy, and neither real-world size nor fixation patterns predict their overlap, as this overlap is not seen for hands and manipulable objects or tools and feet that share similar eccentricity and real-world size (Bracci et al., 2012; Striem-Amit et al., 2017). Instead, what might best explain this spatial organization is that hands and tools share action-related features: specifically, both function as end-effectors for interacting with the world.

Evidence for this comes from studies examining multivariate response patterns in this region. Bracci and Peelen (2013) tested several competing explanations and found that the best account of LOTC responses was that tools extend the hand. These results were not explained by general graspability nor by any other kind of interaction involving the hand; rather, the common representation of hands and tools as end-effectors drove the observed neural representations, thus paralleling the earlier univariate findings.

Complementing this work, Perini et al. (2014) proposed that the overlap reflects activation of hand-posture representations linked to tool use. Using a combination of fMRI and TMS, they showed that LOTC in general, and, above all, the overlap area in particular responded more

strongly when participants judged the typical hand action a tool requires (e.g., squeezing vs. rotating) than when they judged the tool's typical location (e.g., kitchen vs. garage). Critically, TMS disruption of this region selectively impaired action judgments while sparing contextual judgments, demonstrating a causal role for LOTC in processing tool-related action properties.

Another study has examined how a related motor-relevant dimension interacts with real-world size (Magri et al., 2021). Because small objects are often manipulable, these dimensions usually covary, but it is unclear if they can also be dissociated. Therefore, the authors developed a design that attempted to orthogonalize motor relevance and real-world size, and showed that regions of visual cortex, including left LOTC, are sensitive to both dimensions, with manipulability emerging as the most relevant factor. However, this study did not include hands and therefore could not test if either manipulable vs. end-effector properties are a better account of the representational space of LOTC.

Further evidence for an action-based organization of visual cortex comes from a recent study that investigated the representational space underlying a newly discovered category-selective area for food images. Large-scale fMRI analyses on the Natural Scenes Dataset (NSD; Allen et al., 2022) have identified a ventral OTC area preferring images of food, a preference not explained by food-related visual features such as curvature, colour, or texture (Jain et al., 2022; Khosla et al., 2022; Pennock et al., 2023; reviewed in Henderson et al., 2025). Ritchie et al. (2024a) argued that to understand this new category-selective area we need to refer to the graspable properties of food stimuli, properties that make them similar to categories such as tools and manipulable objects. Indeed, the authors tested response to food and tools and found largely overlapping activations in ventral and lateral visual cortex for these object categories. In light of these results, the authors predict that a hand-food overlap will appear in visual cortex, similar to the hand-tool overlap previously described. However, this graspability-based account differs from the end-effector framework proposed by Bracci and Peelen (2013): it focuses on general graspable properties shared by food, tools, and manipulable objects, without distinguishing objects based on their effector-like properties. Moreover, because

hands are not graspable per se, it remains unclear whether they would overlap with food-selective responses. Finally, although graspable features may influence processing in regions of OTC, ventral visual cortex (including food-selective areas) is strongly associated with processing surface features such as ensemble statistics (Cant & Xu, 2012), and colour (Lafer-Sousa et al., 2016; Pennock et al., 2023).

Overall, in the object perception literature, action-related properties have received less attention than other visual or semantic features, yet they may play a fundamental role in shaping the object space of visual cortex. However, it remains unclear whether an action-based topographic dimension organizes object categories more broadly, beyond the specific case of hands and tools. In the next section, I will review computational models that attempt to capture the organization of ventral visual cortex, highlighting their potential and limitations in accounting for its action-based organization.

1.4 Computational models of visual cortex

In parallel with neuroimaging studies investigating category selectivity and spatial organization in OTC, neuroscientists have increasingly turned to computational models to replicate and explain these patterns. Deep Artificial Neural Networks (ANNs), the most common computational framework currently used in neuroscience, are not a recent invention. Since the 1980s and 1990s, researchers have developed computational models that successfully captured the spatial and functional organization of early visual cortex. However, it is the advent of convolutional neural networks (CNNs) that changed the model landscape dramatically.

CNNs are image computable models that learn during training from raw images the features needed to solve specific tasks (LeCun et al., 2015). Seminal papers have found that, perhaps surprisingly, CNNs – which are only loosely inspired by the hierarchical organization of visual cortex – predict neural responses in primate inferotemporal cortex and in human OTC

(Khaligh-Razavi & Kriegeskorte, 2014; Yamins et al., 2014). Since then, a minor methodological revolution has initiated in the visual and cognitive neuroscience field, with numerous studies showing the ability of CNNs to capture various aspects of primate ventral stream, from its general hierarchical object organization (Eickenberg et al., 2017; Hong et al., 2016; Seeliger et al., 2018), to human similarity judgements about object categories (King et al., 2019; Kubilius et al., 2016; Jozwik et al., 2017; Tarigopula et al., 2023; Truong et al., 2025) to finer-grained category-specific organization, such as face processing (Dobs et al., 2022; Grill-Spector et al., 2018; Gupta & Dobs, 2025; van Dyck & Gruber, 2023).

CNNs also capture aspects of many of the object dimensions reviewed above that structure representational spaces in human OTC. For instance, they capture OTC representations related to animacy (Khaligh-Razavi & Kriegeskorte, 2014), real-world size (Huang et al., 2022), aspect-ratio (Bao et al., 2020), and overall shape (Zeman et al., 2020). These networks also mirror the functional selectivity exhibited by OTC for distinct categories (Dobs et al. 2022; Dwivedi et al., 2021; Ratan-Murty et al., 2021). For instance, work investigating real-world size in CNNs has shown that, similar to visual cortex, size is one of the principal axes shaping the object space across these models (Huang et al., 2022).

Despite these successes, ANNs are not perfect models of visual cortex, and a growing body of work has identified systematic divergences between network representations and human neural responses. For instance, ANNs over rely on texture information rather than shape information for object categorization (Geirhos et al., 2018; but see Jagadeesh & Gardner, 2022). While studies reveal spatiotemporal hierarchical correspondence between CNNs and the ventral visual stream (Cichy et al., 2016; Güçlü & Van Gerven, 2015), other work argued that the predictive power of ANNs do not come from its hierarchical organization, as this organization is not necessary to achieve high correspondence with neural data (St-Yves et al., 2023). CNNs are also limited in explaining the temporal dynamics of categorical representations in high-level visual cortex, particularly when contrasted with categorical

models (Jozwik et al., 2023). Moreover, they show limitations in generalizing to unseen stimuli, such as artificial, meaningless shapes (Xu & Vaziri-Pashkam, 2021).

Across these findings, a consistent pattern emerges: CNNs tend to better capture earlier stages of the ventral visual hierarchy, while gaps remain at higher levels of visual cortex. A possible explanation is that CNNs rely predominantly on visual dimensions, whereas humans abstract beyond visual features and incorporate semantic or behaviorally relevant dimensions (Mahner et al., 2025). Of particular relevance for the present work, recent evidence indicates that CNNs struggle to capture scene-related affordances (Bartnik et al., 2025). Specifically, CNNs fail to represent navigationally relevant properties of scenes that are represented in scene-selective cortical regions, highlighting more general limitations in modelling action-related (in this case, locomotion-related) processing in visual cortex. Thus, despite their promise, important gaps remain in explaining ventral visual cortex responses, and substantial effort has been devoted to identifying model ingredients that could close these gaps, a topic I return to in the final chapter.

Overall, regardless of their limitations, ANNs provide a powerful and flexible framework that can not only predict visual cortex representations to a high degree but also generate testable hypotheses about the inductive biases required to solve specific tasks and offer computational explanations for properties of the ventral visual stream. However, irrespective of their predictive or explanatory success, a fundamental limitation of standard ANNs is their lack of spatial organization. Unlike biological visual cortex, ANNs do not structure their representations topographically. For example, while it is possible to investigate which inductive biases give rise to face-selective units in a network, these units do not cluster spatially. In contrast, visual cortex does both: it exhibits feature selectivity *and* spatial clustering of that selectivity. In the next section, I review models that extend ANNs to explicitly address this issue, asking which inductive biases (or model ingredients) are necessary to generate clustering, and why such clustering emerges in the first place.

1.5 Topographic modelling of visual cortex

In this section, I first review earlier models of visual cortex topographic organization and, secondly, I introduce more recent models that aim to combine structure and function within a single framework.

V1's well-characterized spatial organization has served as a benchmark for evaluating whether computational models (and their related principles) can reproduce cortical topography (Erwin et al., 1995). Despite differences in the architectural and training procedures, most early computational models can be understood as mathematical implementations of Hebbian learning combined with local spatial competition or other forms of self-organizing principles (Bednar & Wilson, 2016; Miikkulainen et al., 2005; Swindale, 1996). Notable examples include classic self-organizing maps (SOMs; Kohonen, 1982; 2002) and elastic net models (Durbin & Mitchinson, 1990; Obermayer et al., 1990). These models demonstrated that constraining connectivity and enforcing local interactions naturally give rise to spatially organized feature maps (Jacobs & Jordan, 1992). These models typically consisted of units arranged on a two-dimensional grid, where learning rules encouraged locally correlated activity and penalized long-range connections. Such models successfully reproduced aspects of V1 organization, including retinotopy, orientation maps, colour "blobs", and ocular dominance columns (Barrow et al., 1996; Durbin & Mitchinson, 1990; Kohonen, 2002 or 1982; Obermayer et al., 1990; Sirosh & Miikkulainen, 1997).

However, while successful in replicating some forms of spatial organization observed in V1, early computational work exhibited limitations in capturing the functionality of visual cortex. Indeed, these models were not image computable: their input features had to be hand-engineered (and, hence, previously known), explaining why, with some exceptions (Aflalo & Graziano, 2011; Cowell & Cottrell, 2013), they were rarely extended to account for the emergence of the spatial organization in high-level visual cortex.

Indeed, as mentioned above, standard CNNs predict functional responses but cannot reproduce the spatial organization of visual cortex. To give a concrete example: while CNNs exhibit face-selective units whose functional properties are similar (to a certain degree) to those of ventral face-selective areas, they do not have any knowledge on the way face-selective neurons or voxels are physically clustered in the cortical sheet, thus limiting their biological plausibility.

To address this limitation, two main modelling approaches have emerged in recent years. One approach combines the strengths of CNNs and SOMs: CNNs serve as feature extractors, features that are sent to the SOM that organizes them spatially. Depending on which CNN layer provides the inputs, distinct spatial patterns emerge. This approach has been successful in replicating topographic organization across the entire ventral visual stream, from the pinwheel-like organization of V1 to the category selectivity and topographic gradients of high-level visual cortex (Doshi & Konkle, 2023; Zhang et al., 2024). For instance, this approach was able to replicate the emergence of selective areas for faces and scenes, their location within the broader topographic gradients of eccentricity, aspect-ratio, animacy, and real-world size (Zhang et al., 2024), and the sufficiency of mid-level features (using texform stimuli) for these organizational patterns to emerge (Doshi & Konkle, 2023). However, this framework relies on two separate systems, one modelling function (CNN) and the other modelling structure (SOM), and, therefore, does not provide a single unified account.

A second approach aims to induce topography directly within the ANN itself, yielding a single model with a single architecture capable of capturing both functional selectivity and spatial organization. This strategy was introduced by Lee et al. (2020) and has since been expanded through several architectures (Blauch et al., 2022; 2025; Deb et al., 2024; Margalit et al., 2024; Lu et al., 2025; Qian et al., 2024). Given its relevance to the present work, here I focus more extensively on the architecture of Margalit et al. (2024), which, in my view, represents the current state of the art in topographic modelling of category selectivity in high-level visual cortex.

The Topographic Deep Artificial Neural Network (TDANN)

The Topographic Deep Artificial Neural Network (TDANN) is designed to capture the organization of the entire ventral visual stream within a single model, with a simple set of principles. Built on a ResNet-18 backbone, the model incorporates nine topographic layers, each consisting of a grid of units with pre-assigned coordinates based on preset optimizations such as a coarse retinotopic structure. During training, the TDANN jointly minimizes two loss functions: a standard task loss (either supervised object categorization or self-supervised contrastive learning) and a spatial loss that encourages nearby units to exhibit correlated responses. The strength of this constraint is governed by a critical hyperparameter, α, which controls how strongly neighbouring units must respond to similar functional inputs. Across experiments, intermediate values of α yielded the most biologically-realistic outcomes, capturing diverse aspects of cortical organization and leading to the minimization of wiring length, even though this was not explicitly optimized. Notably, wiring length was minimized most effectively at the same intermediate α levels that best matched cortical topography, and the self-supervised variant of the model provided the closest correspondence to biological data.

Armed with this set of principles, the TDANN reproduces characteristic properties across the ventral visual stream: its V1-like layer exhibits macaque V1 features such as orientation-selective pinwheels, spatial-frequency tuning, and colour-selective blobs, whereas its VTC-like layer develops category-selective clusters for faces, bodies, characters, and scenes, including similar patterns of overlap and proximity between face- and body-selective patches as found in human visual cortex (Kliger & Yovel, 2024; Peelen & Downing, 2005; Schwarzlose et al., 2005). However, the model also produces some non-biological features, such as a cluster selective for certain object categories (e.g., cars) not typically observed in visual cortex (though see previous sections).

Generally, TDANNs offer a set of principles that can potentially explain any kind of organization within and outside ventral visual cortex. For example, when trained with self-supervision and spatial constraints, it captures the organization of visual cortex into ventral, lateral, and dorsal streams more effectively than three separate models optimized for distinct tasks such as categorization, action recognition, and object localization (Finzi et al., 2023).

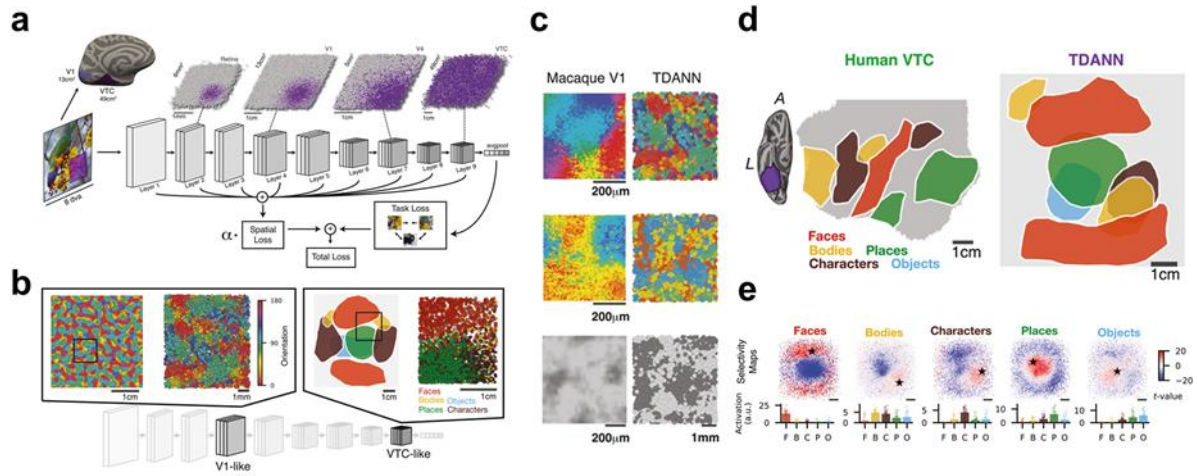


Figure 1.4. The Topographic Deep Artificial Neural Network (TDANN). a) *Architecture.* The TDANN is built on a standard ResNet-18 backbone. Eight topographic layers are created by assigning each unit a fixed position on a 2-D cortical sheet prior to training, using a procedure that includes the replication of coarse retinotopic patterns. During training, the model jointly optimizes a task loss and a spatial loss, the latter encouraging nearby units to develop correlated response profiles; the strength of this constraint is controlled by a hyperparameter called alpha. **b)** *Emergent organization.* The network develops hierarchical representations that parallel those in the biological ventral visual stream: early layers acquire V1-like spatial organization, whereas later layers exhibit VTC-like structure. **c)** *V1-like layer.* The model reproduces common spatial motifs of macaque V1. Left: maps of orientation selectivity, spatial frequency preference, and colour "blob" organization from recordings in the macaque. Right: analogous patterns emerging in the TDANN's V1-like layer. **d)** *VTC-like layer.* Higher-level layers recapitulate category-selective clusters observed in human ventral temporal cortex. Left: schematic organization of major category-selective areas. Right: comparable category-selective clusters in the TDANN's VTC-like layer. **e)** *Functional selectivity*. Top: spatial distribution of selectivity for each category separately; the star represents the most selective unit for that category. Bottom: functional selectivity analysis (i.e., responses to all stimuli and all categories in the image set) of the most selective unit for each category. Adapted from Margalit et al. (2024).

Other architectures

Following the study by Margalit et al. (2024), several subsequent models have expanded on specific aspects of organizational structure in ANNs by introducing new architectures or loss functions. For example, Lu et al. (2025) developed a network that avoids convolutions and weight sharing – two non-biological features of CNNs. Their model successfully reproduces spatial properties of the ventral visual stream, including orientation-selective maps arranged in pinwheel-like patterns with a higher density of features at the "fovea" (the model's centre) relative to the periphery, mirroring cortical magnification in V1. Despite a typical drop in task accuracy (shared by many topographic models), this topographic variant was also substantially more energy-efficient than its non-topographic counterpart, indicating a computational benefit of topographic organization.

Another line of work removed the need for an explicit spatial loss function altogether. Qian et al. (2024) demonstrated that adding local lateral connectivity is by itself sufficient to produce ventral-stream-like spatial organization, both in terms of category-selective clusters and larger-scale topographic principles (i.e., animacy and real-world size). This is notable because it shows that smoothness does not need to be hard-wired into the model for category-selective patterns to emerge. Their model also displayed improved robustness to noise, again pointing to a computational advantage of topography.

Finally, because the TDANN exhibits reduced accuracy on object categorization task relative to its non-topographic variant, several studies have focused on reducing this performance gap, some by modifying the spatial loss function (Deb et al., 2025; Dehghani et al., 2025), and others by adopting more biologically plausible architectures such as spiking neural networks (Zhou et al., 2025).

Overall, in recent years an explosion of models meant to capture the spatial organization of the ventral visual stream has been developed; however, no topographic model has been applied to explain the more complex spatial organization in OTC, such as the presence of

further parcellations (e.g., hands and tools and their overlap) or the dissociation between the organization of ventral and lateral OTC.

1.6 Rationale of the current thesis

The current work was motivated by the body of evidence reviewed above. We start from the observation that lateral occipitotemporal cortex (OTC) – the pathway proposed to support action processing – responds to distinct object categories that are characterized, to varying degrees, by action-related properties, including bodies, hands, and tools. Critically, responses to hands and tools show both anatomical proximity and overlapping representations. Based on this evidence, we ask whether there exists a more general topographic principle that, together with dimensions such as animacy, shape, and real-world size, explains the arrangement of object representations in lateral OTC, beyond the specific case of hands and tools and, similarly, we ask whether the object space in visual cortex reflects action-related properties of objects. In other words, our central hypothesis is that understanding object organization in visual cortex requires going beyond the dimensions traditionally proposed (shape, animacy, real-world size) and incorporating action as an additional organizing principle. This hypothesis forms the core of the thesis (Chapter 2).

Second, we ask whether this proposed action-related topographic dimension can predict the emergence and cortical location of a newly discovered category-selective area responding to food. Food stimuli can be defined by multiple properties, including surface properties such as colour and texture, as well as action-related properties such as graspability. We therefore ask whether food-selective responses fall within meaningful regions of cortex based on the differential sensitivity of ventral and lateral pathways to these properties. Our hypothesis is that the ventral pathway will respond to food stimuli primarily on the basis of surface properties (e.g., colour) that characterize food images, whereas the lateral pathway will respond to food

based on action-related properties shared with other graspable objects, such as manipulable objects and tools (but not hands). We investigate this question in Chapter 3.

In parallel with these neuroimaging investigations, we leverage recent advances in computational modelling to ask whether topographic models capture the same organizing dimensions observed in visual cortex. Specifically, we examine whether these models represent shape, animacy, and action-related properties – and, more broadly, behaviorally relevant features – or whether they remain biased toward purely visual properties, thereby failing to account for the action-based organization of lateral OTC (a question we turn to in both Chapter 2 and Chapter 3).

Finally, we consider another class of computational models that has been successfully used to investigate category-selective responses in visual cortex: ANN-based encoding models (Chapter 4). Using this approach, we ask whether strong selectivity can be identified for whole bodies, hands, and tools, and whether encoding models can reveal the features that distinguish areas selective for the same category but located in different hemispheres (right vs. left) or different pathways (ventral vs. lateral). We hypothesize that this approach will reveal robust image-level selectivity for bodies, hands, and tools, and, critically, will dissociate areas responding to the same category based on their cortical location: areas selective for hands and tools in left lateral OTC will be sensitive to objects containing high levels of action-related information, whereas ventral and right-hemisphere areas will be more sensitive to non-action-related features.

Chapter 2 - Investigating action topography in visual cortex and deep artificial neural networks

This chapter has been published in Nature Communications: Cortinovis, D., Truong, N., Op de Beeck, H., & Bracci, S. (2025). Investigating action topography in visual cortex and deep artificial neural networks. *Nature Communications*.

Abstract

High-level visual cortex contains category-selective areas embedded within larger-scale topographic maps like animacy and real-world size. Here, we propose action as a key organizing factor shaping visual cortex topography and assess the ability of topographic deep artificial neural networks (DANNs) in capturing this organization. Using fMRI, we examined responses to images of body-parts and objects with different degrees of action properties. In left lateral occipitotemporal cortex, we identified a topographically-organized action gradient, with overlapping activations for bodies, hands, tools, and manipulable objects along a dorsal-posterior to ventral-anterior axis, culminating at the intersection of body parts and objects exhibiting higher action properties. Multivariate analyses confirmed action as a crucial organizing principle, while shape and animacy dominated ventral occipitotemporal cortex and DANNs, which exhibited no action-based organization. Our proposed action dimension serves as a further organizing principle of object categories, advancing understanding of visual cortex organization and its divergence from DANN-based models.

Introduction

Topography – the systematic, spatial organization in which neurons (or voxels) with similar functional properties are located near one another in the cortex (Durbin & Mitchinson, 1990) – is ubiquitous throughout the cortex, from the retinotopy and pinwheels of primary visual cortex (Wandell et al., 2007) to the complex somatotopic organization of body parts in the so-called motor homunculus in M1 (Penfield et al., 1937) In occipitotemporal cortex (OTC), a topographic organization of functionally selective areas has been shown, with areas responding preferentially to ethologically-relevant categories such as faces, body parts, words, and scenes (Kanwisher, 2010; Op de Beeck et al., 2008), mirrored along the ventral and lateral OTC (Taylor & Downing, 2011), and forming a consistent spatial arrangement across participants (Grill-Spector & Weiner, 2014).

Several accounts have tried to explain this organization by highlighting the role of different features that map object space onto the two-dimensional cortical sheet, leading to the emergence of functionally selective areas. These features span from low-level principles like eccentricity (Gomez et al., 2019; Levy et al., 2001; Malach et al., 2002), to mid-level properties (e.g., curvature [Yue et al., 2020], aspect-ratio [Bao et al., 2020; Coggan & Tong, 2023], texture [Jagadeesh & Gardner, 2022]), and to semantic principles like animacy (Kriegeskorte et al., 2008) and real-world size (Konkle & Oliva, 2012). Some of these dimensions appear to be repeated across ventral and lateral OTC, explaining the mirrored organization of category-selective areas (Hasson et al., 2003; Konkle & Caramazza, 2013; Silson et al., 2015). Remarkably, the representational space of higher-level layers in DANNs trained on object recognition captures the same object dimensions observed in the visual cortex (e.g., animacy [Khaligh-Razavi & Kriegeskorte, 2014], aspect-ratio [Bao et al., 2020] – but see [Yargholi & Op de Beeck, 2023] – shape [Zeman et al., 2020], real-world size [Huang et al., 2022]). Moreover, topographic DANNs – architectures that incorporate biologically inspired spatial constraints (Blauch et al., 2022; Lu et al., 2025; Margalit et al., 2024) – develop category-

selective responses (e.g., for faces, bodies, and scenes) that mirror the topographic organization found in the visual cortex.

Notably, accumulating evidence suggests that despite the fact that lateral and ventral OTC show a similar mirrored object topography, their underlying representational space might be better explained by different object dimensions (Lingnau & Downing, 2015; Wurm & Caramazza, 2022). For instance, the left lateral OTC shows sensitivity to categories characterized by their action-related properties such as hands and tools (Bracci et al., 2010; Mahon et al., 2007; Weiner & Grill-Spector, 2010), whose underlying selectivity is spatially adjacent to, and partially overlaps with, one another (Bracci et al., 2012). Hands and tools differ in many visual and semantic properties, such as their shape and animacy; eccentricity and real-world size accounts also cannot explain this pattern of results as this effect does not extend to other object categories sharing similar eccentricity or real-world size (Bracci & Peelen, 2013; Striem-Amit et al., 2017). Instead, this evidence suggests that another dimension plays a role in shaping the topographic organization of visual cortex object space: action (Bracci & Peelen, 2013).

The present study aims to investigate the principles underlying the organization of functionally selective areas, with a focus on how behaviorally relevant action properties of objects shape the spatial organization and content of representations in ventral and lateral OTC. We conducted an fMRI experiment where participants viewed images (Matić et al., 2020) of body parts and objects varying in their degree of action properties.

Using univariate and multivariate analyses on fMRI data, along with representational predictions based on human similarity judgments, we tested how action dimensions interact with other proposed dimensions and compared results in human visual cortex with DANNs. Our results show a dissociation between ventral and lateral OTC in both topography and representational space. Action—alongside shape and animacy—emerged as a key principle explaining the arrangement of categories in lateral OTC, while animacy best explained

topography and representational content in ventral OTC and in DANNs, which in turn did not show any action-related organization. These results demonstrate that action is a fundamental organizing dimension of OTC, and that further developments are necessary for current computational models to fully capture both topography and function of high-level visual cortex.

Methods

fMRI experiment and analyses

Participants

19 participants took part in the fMRI experiment (11 females, sex self-reported, mean age 25.6 years, standard deviation 6.06). One male participant was excluded due to head motion exceeding one voxel. All participants were right-handed except one, all had normal or corrected-to-normal vision, and no history of neurological disorder. All participants gave informed consent and were financially compensated. The Ethics Committee of the University of Trento approved the procedure.

Stimuli

The stimulus set included 6 categories (Figure 2.1). Part of the images were used in (Matić et al., 2020). The set comprised 3 body-parts (hands, headless bodies, and faces), 3 inanimate object categories (tools, manipulable objects, and non-manipulable objects), and chairs as a control category. Each object category was associated with a different degree of action-related properties. Tools were defined as hand-held objects that are typically used to physically and directly act on another object or surface (e.g., hammer); therefore, tools are not only graspable and manipulable, but also serve as action-effectors, akin to our hands (Bracci & Peelen, 2013). Manipulable objects are objects that can be grasped, lifted, and manipulated but are not usually used as action-effectors (e.g., glass). Finally, non-manipulable objects were defined as large objects that cannot be grasped nor manipulated (e.g., bed). To control for low- and mid-level visual features, the object categories were matched for their perceived shape and orientation (Figure 2.1). In addition, tools and manipulable objects were matched for real-world size, ensuring that any difference between the two categories cannot be attributed to their actual size. Three additional categories (monkey faces, headless monkey bodies, monkey hands) were part of the experimental design but are not analysed for this report. Each category included 12 grey-scale images with a white background of 400x400 pixels. Behavioral ratings

confirmed that hands and tools were perceived as carrying the most action-related information, with mean scores of 6.3 and 5.7, respectively, on a 1–7 Likert scale. Specifically, hands were rated as conveying a higher level of action-related information than both bodies (4.5) and faces (3.4). Similarly, tools received higher ratings than both manipulable (3.3) and non-manipulable objects (2.9).

Scanning procedure

In the fMRI experiment we collected 8 runs per participant. Each run lasted 400 sec (200 volumes). Each image was presented for 0.4 s, with an ISI of 0.266 s, in blocks of 8 s (i.e., 12 images per block). For each subject and for each run, a fully randomized sequence of all conditions was repeated 4 times, with a fixation block of 16 seconds at the beginning, in the middle (between sequences), and at the end of each run.

Stimuli were presented with the Psychophysics Toolbox package (Brainard, 1997) in MATLAB (2021b) (The MathWorks). Images were projected onto a screen (8 x 8 degrees of visual angle) and shown to the participants through a mirror mounted on the head coil. Participants were instructed to fixate their gaze on the fixation cross in the middle of the screen and press a button whenever the same image was repeated twice in a row within each block. The repeating image appeared once per block. Behavioral performance during the task was quantified by calculating response accuracy (mean = 93%, SD = 2.7%) and reaction times (mean = 0.6 s, SD = 0.02 s) for hits. Accuracy was defined as the proportion of correctly identified target stimuli, with responses considered correct if made within two trials following the targets, taking into account the fast presentation of the stimuli (0.4 s) and the reaction time of participants.

Imaging parameters

The fMRI data was collected using a 3T Siemens scanner with a 64-channel head coil in the Center for Mind/Brain Sciences at the University of Trento. MRI volumes were collected using echo planar (EPI) T2*-weighted sequence, with repetition time (TR) of 2 s, echo time (TE) of 28 ms, flip angle (FA) of 75°, and field of view of 220 mm. Each volume contained 50 axial slices, covering the whole brain, with matrix size 200 x 200 mm and 3x3x3 mm voxel size. Slices were acquired with a multiband (multi-slice) sequence, with slice acceleration factor = 3. Anatomical images were acquired using the T1-weighted acquisition and MP-RAGE sequence, with a resolution of 1x1x1 mm.

Preprocessing

The preprocessing was conducted using the Statistical Parametric Mapping software package (SPM12, Wellcome Trust Centre for Neuroimaging London) and MATLAB (R2021b, The MathWorks). The following standard preprocessing steps were applied to functional images: spatial realignment (to the first image) to correct for head motion; slice-timing correction; coregistration of functional and anatomical images; normalization to a Montreal Neurological Institute's ICMB152 template; and spatial smoothing by convolution with a Gaussian kernel of 4 mm FWHM (Op de Beeck, 2010). Following exclusion criteria defined prior to preprocessing, runs in which the head movement exceeded the size of one voxel (in either translation or rotation) were excluded from subsequent analysis. Based on this criterion, we excluded one participant; additionally, we excluded five runs in total in three participants (two runs in two participants and one in another participant).

The preprocessed signal was then modelled for each voxel, in each participant, and for each condition using a general linear model (GLM). The GLM included 7 regressors of interest, one for each experimental condition, and 6 nuisance regressors corresponding to the 6 motion correction parameters (x, y, z for translation and rotation). Convolution of the haemodynamic response function with the boxcar function was used to model the predictors' time course.

Vector-of-ROIs

To gain insights into the topographic organization of body parts and objects with different degree of action properties in left ventral and lateral occipitotemporal cortex (OTC), we used a vector-of-ROIs approach (Konkle & Caramazza, 2013; Chiou et al., 2018). This analysis allows exploring, in an unbiased way, how the topographic organization of objects, characterized by different properties, changes along a large swath of cortex from lateral to ventral OTC. We focused on the left hemisphere, as tool selectivity is strongly left-lateralized and the hand-tool overlap is larger and more robust in the left hemisphere (Bracci et al., 2012; Pillet et al., 2024b) (see Supplementary Fig. 1, Fig. 2, and Fig.3 for results in the right hemisphere). The vector-of-ROIs approach consists of the following steps: first, we defined two reference points (coordinates from Konkle & Caramazza, 2013), located in a medial region in left ventral OTC (around the parahippocampal cortex [PHC]) and in a superior and posterior region in left lateral OTC (around the transverse occipital sulcus [TOS]). Then, we build a vector connecting the two points by fitting a spline. To make sure that the vector passes through anatomical landmarks relevant for their selectivity profile, we defined 6 anchor points based on coordinates from previous studies. Three were in the left ventral OTC: the medial fusiform gyrus previously shown to respond to tools (mFG; Mahon et al., 2007), the fusiform face area in the lateral fusiform gyrus (lFG; Julian et al., 2012), and a region that responds to small objects around the occipitotemporal sulcus (OTS; Konkle & Oliva, 2012); the other three were in the left lateral OTC: the anterior portion of the inferior temporal gyrus previously known to respond to small objects (aITG; Konkle & Oliva, 2012), the hand-selective inferior temporal gyrus (pITG; Bracci et al., 2012), and the body-selective extrastriate body area within the lateral occipital sulcus (LOS; Julian et al., 2012). After fitting the spline, along the vector, we generated a series of partially overlapping spheres of 6 mm with a distance radius of 3 mm. The beta values extracted from each sphere were employed to perform univariate and multivariate analyses. Furthermore, to investigate how each category-selective peak represents all object categories, we selected the activation peak in the vector-of-ROIs for all

categories separately for ventral and lateral OTC and analysed their functional profile. Results were tested with paired two-tailed t-tests and corrected for multiple comparisons.

Category overlap analysis

We measured the amount of voxel overlap between the activation clusters for each condition, separately for ventral and lateral OTC. To do that, we selected two masks using a combination of functional and anatomical criteria; specifically, we used the Neuromorphometrics atlas (Neuromorphometrics, Inc.) to define regions within ventral and lateral OTC; ventral OTC included the fusiform gyrus and the parahippocampal gyrus, whereas lateral OTC included the inferior and middle occipital gyri and the inferior and middle temporal gyri; within these anatomical regions we selected all the active voxels with a contrast of all conditions vs. baseline with a liberal threshold ($p < .05$ uncorrected); these masks, which contain only the voxels modulated by visual information, were used for the subsequent analysis. To compute the overlap analysis, we calculated the number of active voxels within each of the two masks for each condition vs. all remaining conditions (e.g., hands vs. all others) with a more conservative threshold ($p < .001$ uncorrected at the voxel level and $p < .05$ FDR-corrected at the cluster level). Applying a cluster correction ensures that only contiguous voxels with a meaningful minimum size are considered for the analysis. The resulting active voxels were employed to compute the overlap index which was calculated pairwise for all possible combination of categories by taking the number of voxels common to two clusters (for instance, the voxels that are active for both hands and tools) and dividing it by the number of voxels of the smaller of the two clusters. An index of 0 indicated no overlap between two categories, whereas an index of 1 indicates that the smaller cluster of a category falls completely within the bigger cluster of the other category. Following previously adopted approaches (e.g., Luo et al., 2024), we calculated the overlap at the group level. Overlap analysis at the group level may introduce smoothing that overestimate the amount of overlap between categories; however, previous comparisons of overlap analyses based on single subjects vs. group analyses revealed little differences in the results between the two (Cant &

Xu, 2012); moreover, the use of relatively conservative thresholds and the use of selective contrasts ensure the control of overestimation of overlap effects.

Representational similarity analysis

From each sphere along the vector, we extracted the patterns of activation for each condition and correlated pairwise the patterns with each other to obtain a 6x6 correlational matrix. Values in the resulting correlation matrices represent how the pattern of activity for each category/stimulus correlates with the remaining categories/stimuli, allowing us to investigate how the representational space for the conditions changes from ventral to lateral OTC along the vector of ROIs. Representational similarity analysis (RSA; Kriegeskorte et al., 2008) was used to correlate (via Pearson) the matrix generated from each sphere along the vector-of-ROIs with three models capturing different properties of the stimuli: action, animacy, and aspect-ratio.

The action and the animacy models were generated based on ratings provided by an independent group of participants (n = 22; 13 females, sex self-reported, mean age 23.3 years, SD = 1.96; all participants gave informed consent and were financially compensated) that judged a subset of 36 stimuli, chosen randomly among the entire subset, using the inverse MDS procedure (Kriegeskorte & Mur, 2012). Specifically, to test action-effector properties, we asked participants to arrange the objects according to the degree to which an object or a body-part is typically used to physically/directly act on another object or surface similar to the definition used in (Bracci & Peelen, 2013). To test animacy, we asked participants to arrange the stimuli according to their animacy properties. To measure the overall shape of objects, a formula that captures aspect-ratio was used to test the influence of visual features in explaining patterns of activations for the inanimate objects, as most tools are elongated objects as they must be grasped to fulfil their function. The model was generated by calculating the aspect-ratio for all 72 stimuli using the following formula (as in Bao et al., 2020):

$$\text{Aspect ratio} = \frac{P^2}{4\pi A} \quad (1)$$

Where *P* is the perimeter of the object within the image and *A* is its area.

We generated the dissimilarity matrices for the models by computing pairwise the Euclidean distance between each value for each stimulus along the three dimensions. The three models are orthogonal to each other (see Results), indicating that they are independent and do not overlap in their predictions or dimensions. We calculated the lower bound of the noise ceiling by iteratively correlating each subject matrix with all the remaining subjects' group-average matrix, leading to a final score that indicates the best possible fit to the neural data that the model can achieve given the noise in the data (Nili et al., 2014). Confirming the high reliability of the data, the lower bound of the noise ceiling across lateral and ventral OTC ranged from 0.8-0.9 in VOTC and from 0.7-0.8 in LOTC (Figure 2.5b), indicating a strong correspondence across participants' activity patterns.

Index analysis

The values of correlation matrices (as generated above) were used to calculate two indices: the grasp index and the action-effector index. These indices capture the degree to which the representational content of each body part's activity pattern is correlated with the representational content capturing the action-effector and graspability properties of objects. The action-effector index measures the degree to which each body part relates to objects that are characterized as being action effectors, a property that is specific to tools (e.g., hammer) and not shared with other manipulable objects (e.g., we can grasp and manipulate a glass, but we do not typically use it to act on something else). The grasp index represents the degree to which each body part relates to objects that can be grasped and held in hands, a property common to both manipulable objects and tools (e.g., a glass and a hammer are both graspable), but not to large non-manipulable objects. To calculate the action-effector index, for each participant we took the correlation between each body-part with tools and from that

we subtracted the correlation between each body-part and manipulable objects (e.g., body-tool minus body-manipulable). To calculate the grasp index, for each participant, we took the correlation between each body-part with manipulable objects and from that we subtracted the correlation between each body-part and non-manipulable objects (e.g., body-manipulable minus body-non-manipulable). All results were corrected for multiple comparisons using Bonferroni correction.

Deep Artificial Neural Networks

We tested a series of deep artificial neural networks (DANNs) to test the possible convergence or divergence in the topographic organization and representational profile between visual cortex and DANNs. We selected three different models varying in architecture and training task which are described in detail below.

Non-topographic networks

We selected two non-topographic networks based on the ResNet-50 architecture (He et al., 2016) trained either in object recognition or action recognition. ResNet-object, trained in object recognition with ImageNet (Deng et al., 2009), has been shown to effectively capture representations within category-selective areas in visual cortex (Murty et al., 2021). ResNet-action, trained in action recognition with Moments-in-Time (Monfort et al., 2019), was chosen to test the influence of a training task focused on action recognition in capturing neural responses for action-related categories.

Topographic networks

As these standard networks do not have topographic constraints, we selected a further recently developed family of models that implement some constraints within their architecture to mimic the topographic organization of visual cortex (Margalit et al., 2024). These models – called Topographic Deep Artificial Neural Networks or TDANNs – were based on a ResNet-18 architecture and were trained with a self-supervised contrastive learning task (Chen et al., 2020) on the ImageNet dataset. Prior to training, a mapping of units is implemented within

each layer of the network, so that each unit has a corresponding 2D coordinate that maps them into a 2D grid that represents their physical distance. During training, a spatial loss function (together with the self-supervised task loss) is introduced: this function constraints nearby units to have correlated firing patterns to the same features within the dataset, so that the units that have similar functional properties will fall close in the simulated physical space. A parameter called α in the spatial loss function indicates how much the neighbouring units must be correlated with each other; following (Margalit et al., 2024, we used a value of α = 0.25, as it has been demonstrated to be the optimal value for the emergence of VTC-like topographic organization. These networks include 8 layers implementing topographic constraints, with different surface areas across layers to simulate the hierarchy of the ventral visual stream, from V1 to high-level VTC. We use five different random initializations of the network weights.

Data analyses

Univariate: For the TDANN only, we performed simulated univariate analysis by testing the topographic organization and selectivity profile of the five different random initializations of the network in response to our six object categories; most analyses were conducted on the last layer that qualitatively showed the clearest clustering by categories, which we called VTC-like layer (as in Margalit et al., 2024). Specifically, we tested 1) the clustering of units selective for the different object categories within the simulated physical cortical space in the VTC-like layer and 2) the selectivity profile of the top-50 most selective neurons for each category in the VTC-like layer.

Overlap: To examine whether object categories in the VTC-like layer of the TDANN exhibit a similar relationship to those found in the OTC, we measured the overlap in selectivity between units across different conditions. We followed the method introduced by Margalit et al. (2024). Specifically, the simulated cortical sheet was partitioned into 1 mm wide square sections. In each section, we assessed the proportion of units that were selective ($t > 3.5$) for two

categories (e.g., hands and tools, hands and faces, etc.) in pairs. The overlap between these categories was determined by analysing the frequency of selectivity co-occurrence of the two categories within each section. Essentially, if the selectivity frequency for one category can predict the selectivity for the other, the unit populations are considered to overlap. This overlap is measured using an index that ranges from 0 to 1: a score of 0 means the presence of units selective for one category (e.g., hands) always predicts the absence of units selective for the other (e.g., tools); a score of 0.5 indicates no predictability between the two categories; and a score of 1 signifies perfect overlap, where the presence of units selective for one category always coincides with the presence of the other category.

Multivariate: For all networks, we presented our stimulus set and extracted the feature activations from the convolutional and fully-connected layers across the network hierarchy for the DANNs, and from the eight topographic layers for the TDANN. We generated RDMs for each layer by correlating pairwise the features extracted by the networks for each stimulus. As for neural data, for each layer in each network, we performed the RSA analysis testing three models (shape, animacy, and action) and computed the action-effector and grasp indices. Moreover, we computed multidimensional scaling on the matrix of the last convolutional layer of the two ResNet and of the VTC-like layer of the TDANN, to explore its multidimensional profile more in detail. Statistical significance for all results was assessed via 10,000 permutation tests ($p = .001$).

Results

To investigate how action-related properties influence object topography in visual cortex, we designed a stimulus set organized along two dimensions: animacy (body parts vs. inanimate objects) and action. Specifically, the three inanimate categories vary along two action-related properties: action effector and graspability (Figure 2.1). Tools are both action effectors and graspable; manipulable objects are graspable but not effectors; and non-manipulable objects are neither effectors nor graspable. The three body parts also differed in action relevance: low for faces, higher for bodies, and highest for hands. Action-related properties for all categories were behaviorally validated (see methods for details).

To investigate the degree to which animacy and the two properties of the action dimension can predict the object topography in visual cortex, we combined univariate (e.g., functional profile, overlap analysis) and multivariate analyses of fMRI data to examine both the large-scale spatial distribution and the underlying representational content in lateral and ventral OTC. In parallel, we evaluated the ability of DANNs to capture this organization to assess where current models align with, or diverge from, biological systems.

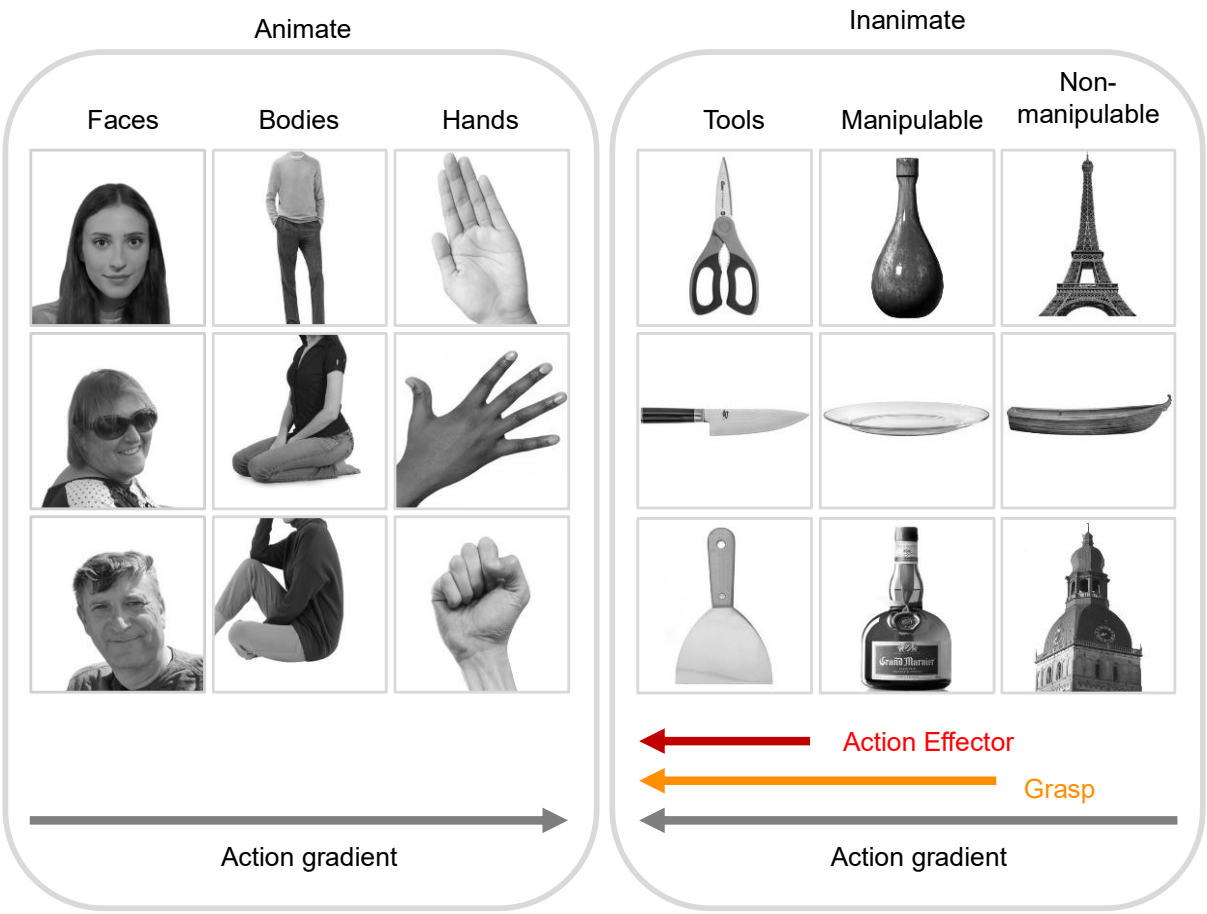


Figure 2.1. Stimulus set. Images were divided into 6 categories varying along two dimensions, animacy and action. For inanimate objects, action was characterized by two properties, action-effector (red) and graspability (orange). The three inanimate objects were matched for visual shape and orientation, to avoid confounds based on the overall shape (e.g., the elongation) of the stimuli. Note that the actual images have been replaced with faces of people that provided consent for publication.

Action properties differentially shape object topography in ventral and lateral OTC

To investigate object space organization in ventral and lateral OTC (VOTC and LOTC, respectively), we first mapped the activation response for each category (versus all others, $t > 3.5$, $p < .05$ FDR corrected at the cluster level) onto the whole-brain surface (Figure 2.2a). Beyond replicating the known parallel organization of category selective responses in lateral and ventral OTC (Pillet et al., 2024), the whole-brain analysis confirmed a dissociation

between the VOTC and LOTC in the left hemisphere (Figure 2.2a) based on the activation patterns for object classes with varying degrees of action-related information. Whereas in VOTC we found the typical medial-to-lateral animacy division with no overlap between animate and inanimate categories (Grill-Spector & Weiner, 2014), in LOTC we observed overlapping responses between animate and inanimate conditions with a different degree of action properties. From dorsal-posterior to ventral-anterior, we observed selective and partly overlapping activations for bodies, hands, tools, and manipulable objects, with a convergence and high degree of overlap for the animate and the inanimate categories characterized by the highest degree of action properties: hands and tools. The action-based organization was particularly evident when comparing activations of inanimate objects. Specifically, we found a consistent action-related gradient in LOTC, with a smooth transition across the cortical surface where the activation to object categories characterized by different action properties changes systematically according to the two action-related properties. This gradient was characterized by a large activation cluster for tools which are both action-effector and graspable, a smaller cluster for manipulable objects which are only graspable, and no significant activation for non-manipulable objects which are neither action effector nor graspable; the opposite pattern was observed in VOTC, with a larger cluster for non-manipulable relative to manipulable objects, which in turn revealed a larger activation relative to tools. The action-related topographic organization in LOTC was also observed at the level of individual participants, without spatial normalization or smoothing (see Figure 2.2c for an example participant). Unlike the left hemisphere, the right hemisphere did not show any action-related organization, as neither tool nor object selectivity were observed (see Supplementary Fig.S2.1, Fig. 2.2, and Fig. 2.3 for right hemisphere results). In the remainder of the paper, all analyses refer to the left hemisphere.

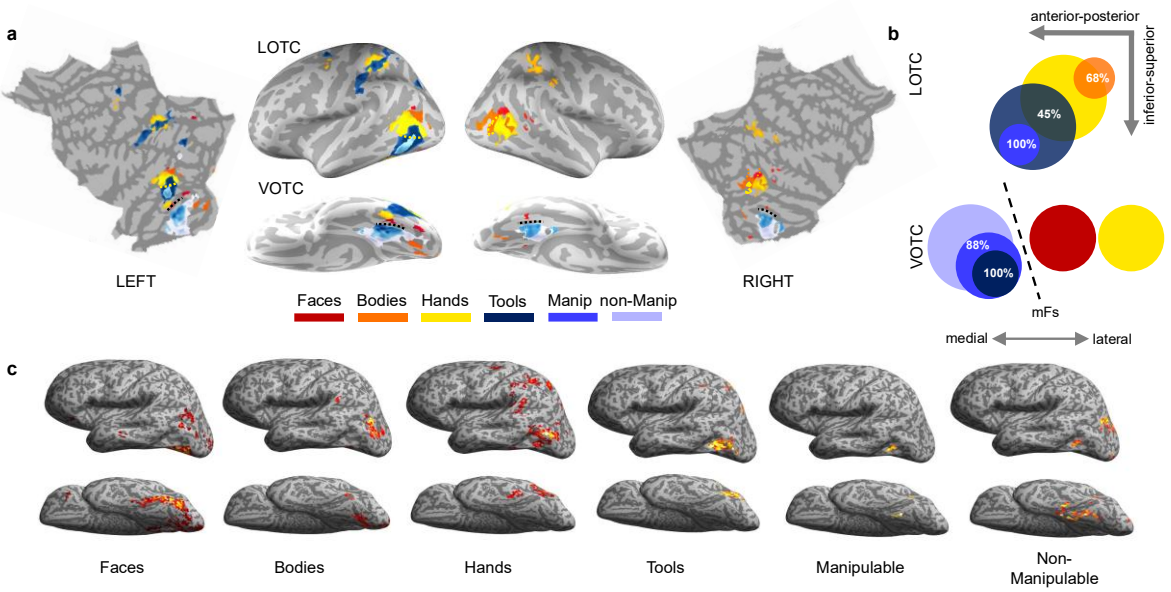


Figure 2.2. Action-related topography of occipitotemporal cortex. a) Whole-brain results. Response for each category (vs. all) was visualized on a freesurfer average brain surface using BrainSurfer (https://www.mathworks.com/matlabcentral/fileexchange/91485-brainsurfer), with a threshold of $t > 3.5$ ($p < .05$ FDR corrected at the cluster level), excluding activations within early visual cortex (approximately V1-V2-V3) to focus on the regions of interest in LOTC and VOTC. Color-coded dashed lines indicate overlap between activations. The black dashed line indicates the mid-fusiform sulcus. b) Category overlap visualization. The size of each circle represents the approximate size of the category-selective cluster in VOTC and LOTC in the left hemisphere. c) Single subject results on the unsmoothed native surface of one representative participant ($t > 3.5$, FDR cluster corrected at $p < .05$). For all panels, VOTC = Ventral Occipitotemporal Cortex. LOTC = Lateral Occipitotemporal Cortex, and red = faces; orange = bodies; yellow = hands; dark blue = tools; blue = manipulable objects; light blue = non-manipulable objects.

These results were further confirmed by the overlap analysis, which allowed us to further assess the spatial relationship between categories, with the underlying rationale that spatial proximity and overlap in the cortex suggest shared features (Ritchie et al., 2024a). We quantified the extent of activation overlap between each category by calculating an overlap index for each pairwise combination of regions, separately for the ventral and lateral OTC (see

methods, Figure 2.2b). The index represents the number of voxels common between the areas, varying from 0 (no voxels in common) to 1 (the smaller area falls completely within the larger). In LOTC, from dorsal-posterior to ventral-anterior a large overlap could be observed between hands and bodies (0.68), between hands and tools (0.45), and between tools and manipulable objects (1.0, where manipulable objects fall completely within the larger tool cluster), but no overlap could be observed for the other combinations. On the contrary, in VOTC, no overlap could be observed between animate and inanimate categories, nor between faces and hands; inanimate objects, instead, presented a strong overlap with each other, with tools falling completely within the manipulable object cluster (1.0), and manipulable showing an extended overlap with non-manipulable objects (0.88), thus further confirming the opposite gradient in LOTC and VOTC for objects characterised by a different degree of action properties. A schematic visualization of category overlap is shown in Figure 2.2b.

To further characterize the spatial and functional profile of the different object topography observed in LOTC and VOTC, we plotted the beta values for each condition extracted from a series of partially overlapping spheres covering a broad region of visual cortex including a wide portion of ventral and lateral OTC from the parahippocampal cortex (PHC) to the transverse occipital sulcus (TOS) (see methods, and Figure 2.3a). The vector of ROIs analysis confirmed that from lateral to ventral OTC, the response profile for all inanimate objects follow a similar activation trend but with an opposite response strength based on action-related properties of objects: tools, which are both action effectors and graspable, show the highest response peak in LOTC and the lowest in VOTC; manipulable objects, which are graspable but do not serve as effectors, show the intermediate response in both LOTC and VOTC; and non-manipulable objects which are neither action effectors nor graspable show the lowest response in LOTC but the highest in VOTC (Figure 2.3a).

Overall, these results indicate that the topography of objects in lateral and ventral OTC is driven by their different degree of action properties, as measured by their action-effector and grasp properties. To verify this, we plot the peak response (1 sphere) for each condition in

ventral and lateral OTC (Figure 2.3b). Results of paired two-tailed t-tests confirm that, within the inanimate object cluster, tools elicit the highest activation across all three LOTC object peaks ($p < .01$ for all contrasts; Bonferroni corrected for n = 5 comparisons). In contrast, non-manipulable objects elicit the highest response across all three VOTC object peaks ($p < .001$ for all contrasts), except in VOTC-tool where non-manipulable and manipulable did not differ from each other ($p = .41$). Hands elicit the highest activation in the LOTC animate peaks (LOTC-hand, LOTC-face) compared to all other object categories ($p < .001$ for all contrasts) except for LOTC-body where bodies elicited the highest response ($p < .003$ for all contrasts). Finally, whereas faces show the typical selectivity in VOTC (VOTC-face and VOTC-body: $p < .001$ for all contrasts), we also observed a small but selective cluster for hands in the occipitotemporal sulcus, located lateral to the face cluster, which shows significant higher activation for hands than for all other categories including faces and bodies (VOTC-hand: $p < .001$ for all contrasts). This region likely corresponds to the left counterpart of the fusiform body area [38], a region that has been also called OTS-limbs (Weiner & Grill-Spector, 2010). Here, we report its selective activation for hands specifically and not bodies in general, thus confirming the possibility of dissociating the activation to hand stimuli from the one to whole bodies not only in lateral (Bracci et al., 2010) but also in ventral OTC (see also Pillet et al., 2024).

Overall, these results support the conclusion that the parallel object representations in LOTC and VOTC encode distinct object properties, and specifically point to the presence of an opposite organization within ventral and lateral OTC, with the latter being sensitive to object categories that contains a different degree of action information, as indexed by the consistent topographic organization for objects and body parts with different action-related properties and convergence between inanimate (tools) and animate (hands) categories that share effector properties.

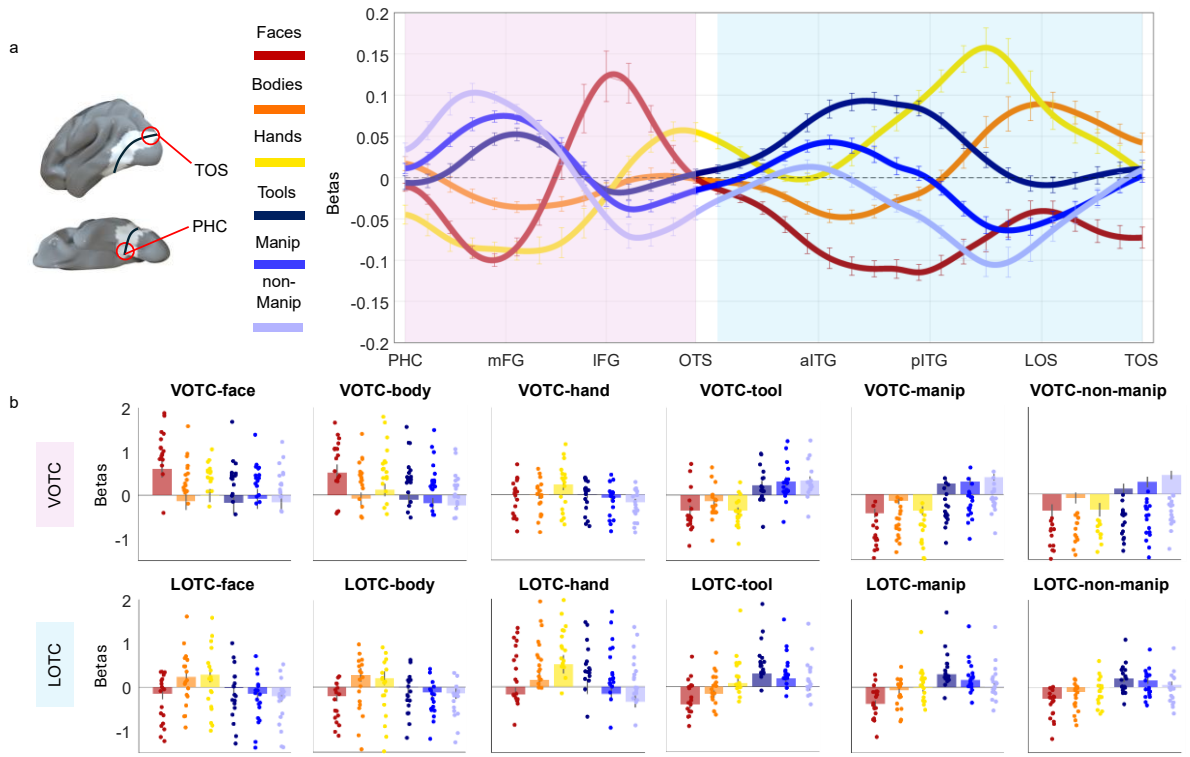


Figure 2.3. Distinct object topographies in lateral and ventral OTC. a) Vector-of-ROIs analysis. The vector was generated by fitting a spline (drawn black line) connecting the PHC and the TOS and passing through a set of anchor points which coordinates were based on classically defined category-selective areas (i.e., face, body, hand, object) from previous studies. Partially overlapping spheres (n = 34) were generated along this spline, and they correspond to the ROIs analysed. Standard univariate analyses were performed on each of the ROIs (see methods for details), which are visualized in white with a surface projection using Surf Ice (https://www.nitrc.org/projects/surfice/). Normalized activation (against the average of all categories) is plotted for each category as a function of the position of the vector along the cortex. The x-axis corresponds to each sphere along the vector, with labels for major anatomical landmarks; the y-axis corresponds to the normalized beta values. The vector was broadly divided into a ventral component (pink shade) and a lateral component (light blue shade). Error bars represent ± 1 SEM across participants (n = 18). b) Beta values are plotted for each category's peak activation (one sphere) separately for the ventral occipitotemporal cortex (VOTC) and lateral occipitotemporal cortex (LOTC). Error bars represent ± 1 SEM across subjects (n = 18 participants). Each data point reflects the beta value extracted from one subject's ROI at the category's peak activation. PHC = Parahippocampal Cortex. mFG = medial Fusiform Gyrus. lFG = lateral Fusiform Gyrus. OTS = Occipitotemporal Sulcus. aITG = anterior Inferior Temporal Gyrus. pITG = posterior

Inferior Temporal Gyrus. LOS = Lateral Occipital Sulcus. TOS = Transverse Occipital Sulcus. For all panels, red = faces; orange = bodies; yellow = hands; dark blue = tools; blue = manipulable objects; light blue = non-manipulable objects. Source data are provided as a Source Data file.

Topographic DANNs successfully mimic animacy division in VOTC but fail to replicate action-based topography in LOTC

The above results show that the lateral and ventral OTC are characterized by a different topographic organization: whereas in VOTC the animacy of objects strongly drives the organization of representations giving rise to the well-documented animacy division, in LOTC the topographic organization is driven by the degree of object action properties with a gradient from posterior-superior to anterior-inferior. Here, we test whether topographic deep artificial neural networks (TDANNs), a type of computational model developed to capture the topographic organization of ventral visual cortex (Margalit et al., 2024), can mimic the action-related organization observed in lateral OTC. TDANNs allow testing whether a model designed to capture general topographic organization as a by-product of minimizing wiring-length (Chklovskii et al., 2002) can account for object topography in visual cortex, thus suggesting that brain-like representations and their spatial organization can co-emerge with biologically inspired spatial constraints.

The network architecture was based on a ResNet-18 backbone, pre-trained with a self-supervised contrastive-learning object recognition task (Chen et al., 2020). We tested five different random initializations of the network's weights. We fed the networks with the images from our experiment and extracted the activation maps for each topographic layer, selecting the last VTC-like layer for further analyses (consistent with Margalit et al., 2024). A unit was defined as selective if its response for a specific category passed a set threshold (defined as $t > 3.5$, with a contrast of category > all). This uncorrected threshold was chosen for visualization purposes only (Figure 2.4a). The subsequent functional selectivity analysis was performed on the first 50 most selective units. To investigate whether TDANNs replicate the

topography and functional profile of category activations in visual cortex, we visualized their respective spatial distribution in the simulated cortical space and plotted the activation profiles for the 50 most selective units per category. Results are shown in Figure 2.4. Despite some variations between the five initializations – especially in the clustering's strength – two main findings could be observed (Figure 2.4a): first, in all networks, units selective for animate and inanimate objects formed separate clusters, such that when a unit responded to a body-part it did not respond to an inanimate object and vice versa; second, no organization based on action properties was observed. Specifically, tools and hands did not activate the same units, and no smooth overlap based on action properties was found among the three object categories.

To quantify these observations and compare TDANNs with brain results, we performed an overlap analysis (as in Margalit et al., 2024). Specifically, we measured the co-occurrence of units selective for each category by using an overlap score ranging from 0 (the presence of one category always predicts the absence of the other) to 0.5 (no relationship) to 1 (perfect co-occurrence). Statistical significance was tested via 10,000 permutation tests. Results (Figure 2.4b) confirmed significant overlap within animate (score: 0.68, $p < .001$) and inanimate (score: 0.74, $p < .001$) categories relative to the between-category overlap (animate-inanimate, score: 0.51). In other words, units that responded to a body part or an inanimate object also responded significantly to other categories within the same superordinate class. Second, the overlap score between action effector categories such as hands and tools (score: 0.59) was not significantly higher than the overlap between hands and other manipulable objects (score: 0.594, $p = .37$), as well as the overlap between tools and manipulable objects (score: 0.79) was not significantly larger relative to the overlap between tools and non-manipulable objects (score: 0.72, $p = .24$), nor relative to the overlap between manipulable and non-manipulable (score: 0.72, $p = 0.33$) thus, showing no action-related organization in TDANNs.

Visual exploration of Figure 2.4a suggests that, in addition to the separation between animate and inanimate categories, there seem to be additional differences in the organization of categories. Specifically, whereas the spatial distribution of units selective for the different body parts seem a bit scattered around, the inanimate objects mostly activated a similar portion of the cortical space. To investigate the functional profile of the TDANN units, we extracted the activation profiles for the 50 most selective units for each category and plotted the results (Figure 2.4c shows results averaged across the five initializations). Here, the focus was not on unit selectivity per se (e.g., do tool units respond to tools more than all other categories) but rather the degree to which a unit that responds to one category also responds to other categories (e.g., do tool units respond to other categories as well?). Overall, the results show that while a certain degree of category-selectivity could be found for the different body-parts, as different units selectively activated for each body part independently from the other body parts, the top-units for each inanimate object category responded to the other inanimate objects to a similar degree. Indeed, the selectivity of units chosen based on their response for faces, bodies, and hands was significantly higher for their preferred category compared to all other categories (for all contrasts, $p < .001$; permutation test $n = 10{,}000$). In contrast, units selected for their response to tools, manipulable and non-manipulable objects did not differ in selectivity across other inanimate object categories (for all contrasts, $p > .05$), while being more selective for their preferred category than for the animate categories (for all contrasts, $p < .001$; permutation test $n = 10{,}000$). Thus, similar to what we observed in visual cortex, TDANNs units that respond to one inanimate object category do also respond to the other inanimate object categories, but differently from human VTC, we did not observe any differential response gradient from high to low (tools > manipulable > non-manipulable) as observed in LOTC or from low to high (non-manipulable > manipulable > tools) as observed in VOTC. Finally, differently from visual cortex, units that respond to tools did not seem to activate hand units, thus confirming results from the TDANNs overlap analysis.

Overall, these results show that TDANNs primarily distinguish between animate from inanimate objects, with additional functional selectivity for individual body-parts, and a weaker, or absent, distinction among inanimate object categories. These results mirror the pattern of overlap found in VOTC, which also showed a separation between animate and inanimate object categories, with further clustering for hands and faces. However, no action gradient organization, as found in LOTC, could be observed in TDANNs.

Altogether, these analyses on networks implementing biologically-inspired topographic constraints reveal their ability to capture visual features important to distinguish animacy and to capture – to a certain extent – the selectivity for body-parts, but cannot replicate the action-related organization observed in visual cortex.

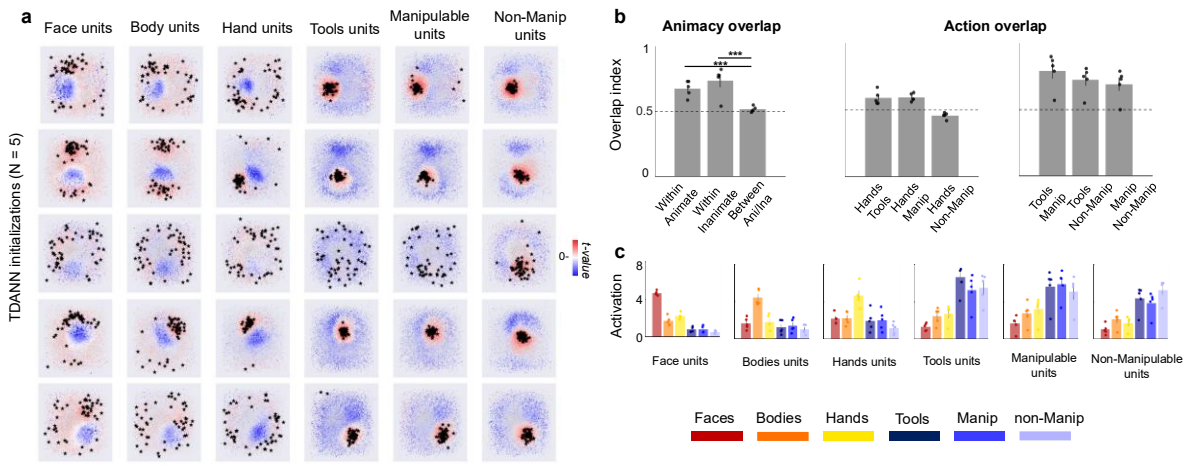


Figure 2.4. TDANNs replicates animacy but not action-related organization of OTC. a) Spatial distribution of each category (as defined by t-values) on the simulated cortical space of the VTC-like layer of five random initializations of the TDANN. Rows correspond to each of the five initializations. Stars represent the location of the top-50 most selective units for that category. Category-selective units (positive t values) are shown in red while units not selective for that category (negative t values) are shown in blue. b) Overlap analysis. Statistical significance was assessed using permutation tests (10,000 randomizations on the mean overlap score across initializations). Stars represent statistical significance at the minimum resolvable p-value (p = .0001), corresponding to the 10,000-permutation limit. Error bars correspond to ± 1 SEM across the random initializations. Black dashed line represents baseline (overlap of 0.5 means no correlation between the presence of two categories). Each data point

represents the value from a single TDANN initialization (n = 5 model initializations). c) Selectivity profile of the top-50 most selective units for each category (red = faces; orange = bodies; yellow = hands; dark blue = tools; blue = manipulable objects; light blue = non-manipulable objects), based on the activation of the VTC-like layer (as in a). Each data point corresponds to one TDANN model initialization (n = 5 model initializations). Error bars indicate ± 1 SEM across model initializations. A baseline overlap of 0.5 denotes chance-level correspondence between category-selective units. Source data are provided as a Source Data file.

VOTC and LOTC support distinct object feature spaces

Our results reveal a different object organization in LOTC and VOTC and that TDANNs are able to capture only part of visual cortex topographic organization. Next, we employ multivariate analyses to further investigate what properties underlie this object space. Specifically, we use representational similarity analysis (RSA, Kriegeskorte et al., 2008) to investigate how the action and animacy dimensions relate in both visual cortex and DANNs. We created three models, each reflecting a distinct dimension: the action model capturing action-related information for each object category; the animacy model capturing the body-parts\inanimate objects divide; and the shape model capturing the average aspect-ratio of each category (see methods), added to account for visual properties relevant in OTC (Bao et al., 2020; Coggan & Tong, 2023). The animacy and action models were generated from participants who judged a random subset of stimuli (n = 36) on each dimension (see methods). The models were orthogonal: animacy vs. action-effector (r = –0.08); animacy vs. shape (r = 0.08); action-effector vs. shape (r = –0.16). Dissimilarity matrices (Figure 2.5a) support our predictions: the animacy model clearly separated body-parts from inanimate objects; the action-effector model showed a graded continuum: as the action-related properties of body parts and objects increased, their correlation strengthened.

We assessed how these dimensions are represented across lateral and ventral OTC, by correlating neural activity patterns in each vector-of-ROIs sphere with the three models (Figure

2.5a). Results showed that while animacy was strongly represented across the entire swath of cortex and reached the noise ceiling in ventro-medial regions of OTC, the action dimension reached its highest peak within LOTC, specifically between posterior ITG and LOS and its lowest peak in VOTC, in correspondence of the highest peak for animacy. Interestingly, throughout both ventral and lateral OTC, the effect for object shape closely followed the trend of the action model, suggesting that regions encoding action-related properties of objects also represent their shape properties. To quantify this trend, we perform pairwise correlations between the effects of each model along the vector. Results confirmed that shape and action did indeed show a small but significant correlation along the vector ($r = 0.18$, $t_{(17)} = 3.2$, $p = .0044$; for all RSA results, Bonferroni correction for n = 3 comparisons; $p < .016$). On the contrary, a significant negative correlation was observed between the action and the animacy models ($r = -0.4$, $t_{(17)} = -8.2$, $p < .001$), whereas no correlation was found between shape and animacy ($r = -0.05$, $t_{(17)} = -1.2$, $p = .24$).

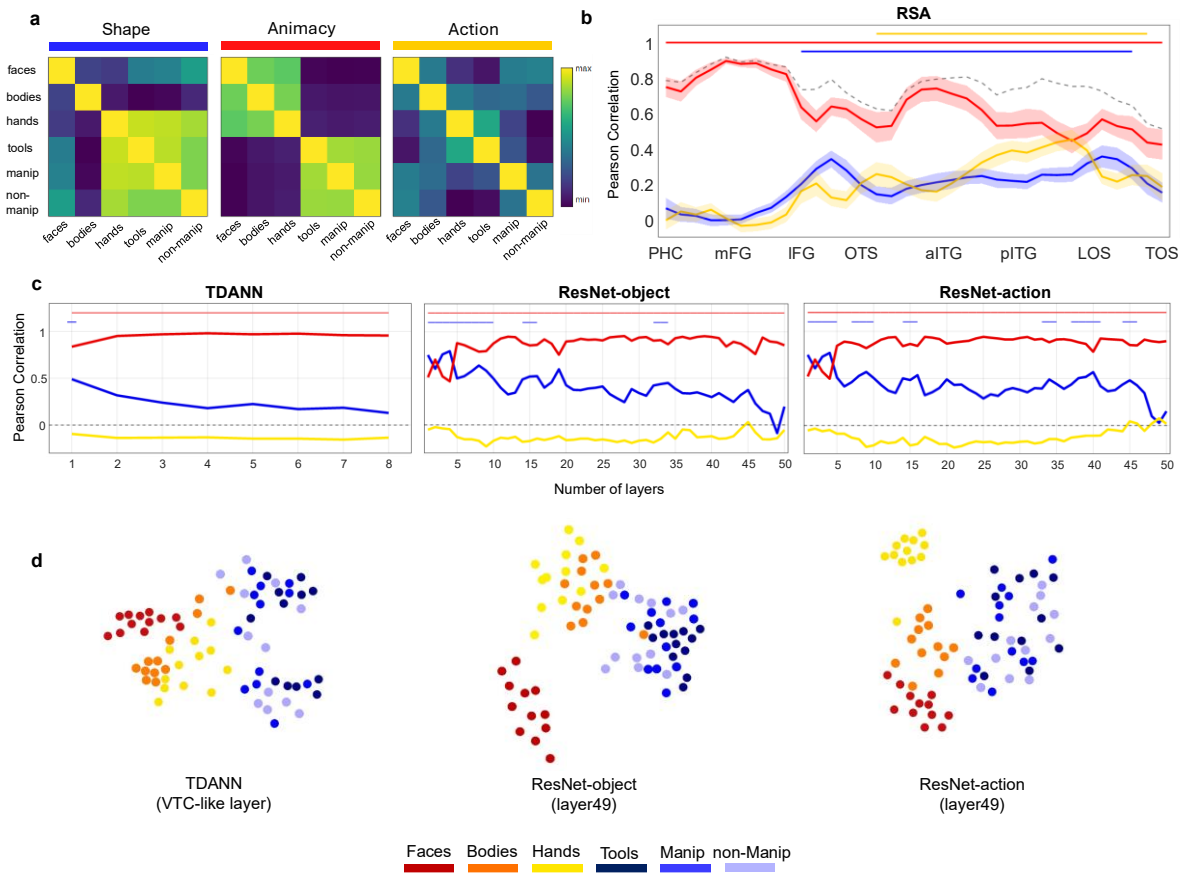


Figure 2.5. The distinct role of action and animacy in visual cortex and DANNs. a) RSA Models: the shape model (blue) captures the aspect-ratio of the stimuli, whereas the animacy (red) and action (yellow) models are based on behavioral ratings (see methods). b) Vector-of-ROIs RSA results. The dashed line represents the noise ceiling boundary which indicates the highest best possible fit to the neural data that a model can achieve given the noise in the data. Statistical significance was assessed using two-sided one-sample t-tests, and horizontal lines indicate statistical significance (vs. baseline) for each model ($p < .0014$ Bonferroni corrected for $n = 34$ comparisons). The shaded area around the line indicates ± 1 SEM across participants ($n = 18$). PHC = Parahippocampal Cortex. mFG = medial Fusiform Gyrus. lFG = lateral Fusiform Gyrus. OTS = Occipitotemporal Sulcus. aITG = anterior Inferior Temporal Gyrus. pITG = posterior Inferior Temporal Gyrus. LOS = Lateral Occipital Sulcus. TOS = Transverse Occipital Sulcus. c) RSA results for the three DANNs. Statistical significance was assessed using permutation tests (10,000 random shuffles of category labels). Color-coded lines on top of each graph indicate the layers where each model reach statistical significance relative to baseline ($p < .001$). d) MDS for the DANNs (ResNet-Object and ResNet-Action) last convolutional layer (layer 49) and the TDANN VTC-like layer. Results for the TDANN refers to one of its initializations. (red = faces; orange = bodies; yellow = hands; dark blue = tools; blue = manipulable objects; light blue = non-manipulable objects). Source data are provided as a Source Data file.

We performed the same RSA analyses in the TDANNs and in two non-topographic models, both based on the ResNet-50 architecture but trained with different task objectives: object recognition with ImageNet (Deng et al., 2008; ResNet-object) and action recognition with Moments-in-Time (Monfort et al., 2019; ResNet-action). This allowed us to test whether training objectives influence the networks' representational space and whether action recognition training improves the representational correspondence with LOTC.

The RSA analysis performed in DANNs revealed different results compared to visual cortex. Across all networks, regardless of architecture (topographic or non-topographic) or training task (object or action recognition), animacy was the dominant dimension, highly significant throughout the network hierarchies and outperforming other models in most layers (Figure 2.5c). Shape was the second-best model, with high correlations along the networks' hierarchy

dropping in the final layers, in line with previous reports (Zeman et al., 2020). The action model never reached significance in any layer or model. Furthermore, differently from what we observed in visual cortex, action and shape were not significantly correlated across DANNs' layers (Pearson r = –0.14; p = 0.34). Together, these results show that neither training task is sufficient to produce a brain-like action-related organization in the networks.

To further inspect the DANNs feature space, for each model we projected the dissimilarity matrix of the last convolutional layer (layer 49) of the two ResNet and the VTC-like layer of the TDANN into a two-dimensional plot by using multidimensional scaling (MDS; Figure 2.5d). Confirming the RSA results, the animacy division appears to be the main dimension emerging in the representational space of all DANNs with no evidence for any action gradient. In addition, an effect of shape was observed in the arrangement of inanimate objects. That is, differently from body parts, which show some clustering based on category, objects that by design were matched for shape, show an arrangement based on visual properties such as aspect-ratio and orientation.

Lateral OTC represents action-effector and (to a lesser extent) grasping properties of objects

Up to now, we have shown that distinct object dimensions are represented in ventral and lateral OTC. Here, we further characterise the specific action-related properties underlying this object space. To this aim, we calculated two indices derived from the correlational matrices obtained with multivariate analysis (see methods): the action-effector index and the grasp index. The indices measure distinct properties of the object categories, specifically the possibility of an object to be an end-effector (the action-effector index), which differentiates tools (e.g., a pair of scissors or a knife) from other graspable objects (e.g., a bottle or a glass) and is shared between hands and tools, and the possibility of an object to be grasped (the grasp index), which differentiates manipulable objects from large non-manipulable objects that cannot be grasped (e.g., a building or a vehicle). The action-effector index was calculated by

taking the correlation between each body-part with tools and from that subtracting the correlation between each body-part and manipulable objects; the grasp index was calculated by taking the correlation between each body-part with manipulable objects and from that subtracting the correlation between each body-part and non-manipulable objects (see methods). Results are shown in Figure 2.6.

This analysis revealed that the driving factor underlying the object space in LOTC is the action-effector property of objects, followed by a smaller but significant effect for object grasp. More specifically, the action-effector index shows that across the whole LOTC, hands are strongly associated with objects that are characterized by effector properties, such as tools, compared to other manipulable objects which share graspable properties with tools but do not serve as action effectors (Figure 2.6a, left). This effect is specific for hands, as whole bodies do not show the same pattern and faces even show a negative index (which indicates higher correlation with objects that are not action-effectors). These results show that while the action-effector effect is present throughout most LOTC, its strength follows closely the response profile of hands, suggesting that univariate hand-selectivity supports an object space with one of the main dimensions being action-related. To directly test this relationship, we computed the correlation between the effector index and the activation for the different object categories along the vector-of-ROIs. Throughout the vector, the effector index was significantly correlated with the hand's response profile ($r_{(17)} = 0.38$; $t_{(17)} = 4.46$, $p < .001$; Bonferroni correction for n = 6 comparisons; $p < .0083$) but not with the response profile for either faces, bodies, or tools (faces: $r_{(17)} = -0.09$; bodies: $r_{(17)} = 0.08$; tools: $r_{(17)} = 0.032$;) and it was negatively correlated with the response profile for manipulable and non-manipulable objects (manipulable: $r_{(17)} = -0.2$; $t_{(17)} = -3.61$, $p = .0022$; non-manipulable: $r_{(17)} = -0.28$; $t_{(17)} = -4.7$, $p < .001$).

The grasp index (Figure 2.6a, right) reveals a smaller but significant effect in some regions of LOTC, showing that hands are also associated with manipulable objects more than to non-manipulable objects. This effect was not observed for bodies and faces. Confirming the other analyses, no significant grasp index was found in VOTC. Finally, in line with the weaker grasp-

related effect, only a modest relationship was found between univariate selectivity for hands and the grasp index (hands: r = 0.22; p = .031) which however did not survive Bonferroni correction for multiple comparisons (n = 6; p > 0.0083).

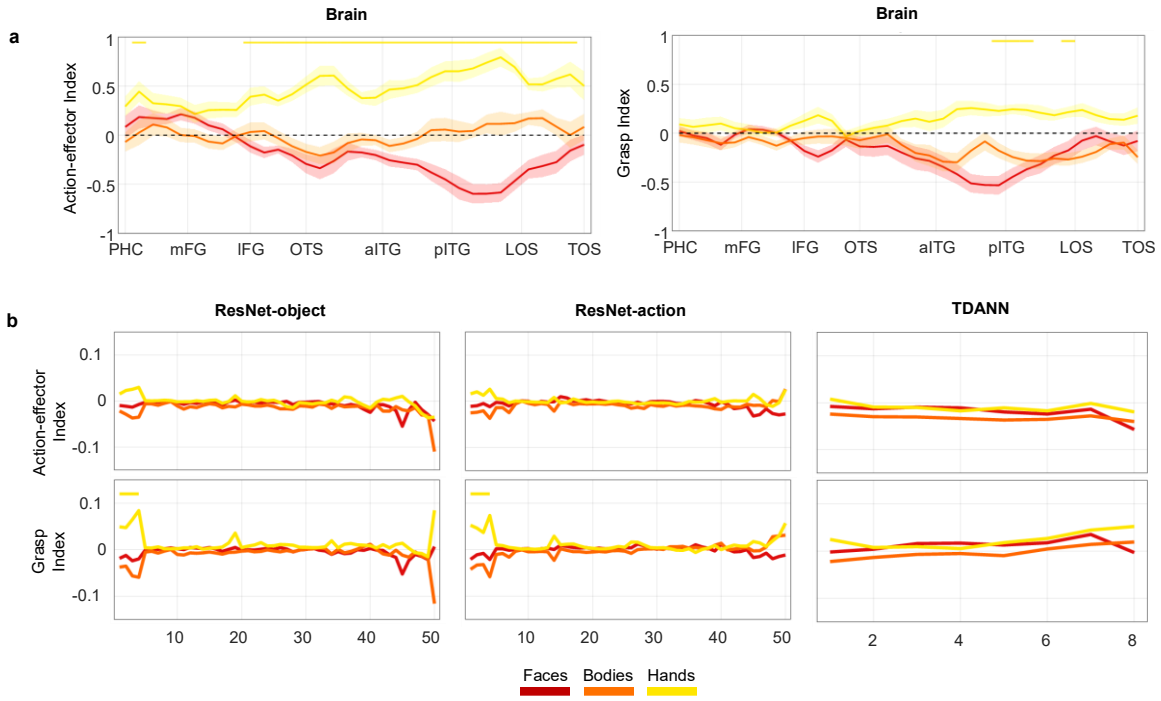


Figure 2.6. Index analysis. a) Vector-of-ROIs action-effector index (right*)* and grasp index (left). The shaded area around the line indicates ± 1 SEM across participants (n = 18). Statistical significance was assessed using two-sided one-sample t-tests, and color-coded lines at the top of each plot indicate spheres along the vectors where each index reached significance, Bonferroni corrected for the number of spheres (n = 34; *p* = .0014). PHC = Parahippocampal Cortex. mFG = medial Fusiform Gyrus. lFG = lateral Fusiform Gyrus. OTS = Occipitotemporal Sulcus. aITG = anterior Inferior Temporal Gyrus. pITG = posterior Inferior Temporal Gyrus. LOS = Lateral Occipital Sulcus. TOS = Transverse Occipital Sulcus. b) Action-effector index (top) and grasp index (bottom) for the three artificial networks tested. Statistical significance was assessed using permutation tests (10,000 random shuffles of category labels), and color-coded lines at the top of each plot indicate layers where each index reached significance (*p* < 0.001). In all panels, red = face indices; orange = body indices; yellow = hand indices. Source data are provided as a Source Data file.

Although DANNs do not show any action properties, for completeness and to test possible similarities or differences with visual cortex we calculated the action-effector and grasp indices for all layers and DANNs (Figure 2.6b). In agreement with the above results, no network shows either action-effector or grasp effects; the two indices did not reach significance ($p > .05$) at any stage of the hierarchy of any of the networks, except for a small effect in the first four layers of both non-topographic networks for the grasp index.

Discussion

Our study identifies action as a fundamental dimension shaping the topographic organization of the visual cortex. We demonstrate that the left lateral occipitotemporal cortex (LOTC) exhibits a dorsal-posterior to ventral-anterior gradient where body parts and inanimate objects are topographically organized based on their action-related properties. The combination of action effector and graspability contributes to explain the spatial organization of voxels that show a preferential response to bodies (Downing et al., 2001), hands (Bracci et al., 2010; Pillet et al., 2024; Orlov et al., 2010), tools (Chao et al., 1999), and manipulable objects (Almeida et al., 2023). While DANNs replicate aspects of ventral stream organization (e.g., animacy), they entirely lack the action-related topography observed in lateral OTC. Together, our results show that the action dimension is an important organizing principle of lateral OTC and highlight remaining gaps between biological and artificial systems.

Previous work emphasised how the combination of multiple object dimensions and principles may result in the topography-by-selectivity that is observed in high-level visual cortex (Grill-Spector & Weiner, 2014; Arcaro & Livingstone, 2024; Bracci & Op de Beeck, 2023; Contier et al., 2024; Huth et al., 2012; Mahon & Caramazza, 2011; Op de Beeck et al., 2019; Peelen & Downing, 2017; Prince et al., 2024; Ritchie et al., 2025; Magri et al., 2021), with proposals stressing the role of shape, animacy, and real-world size (Konkle & Oliva, 2012; Konkle & Caramazza, 2013; Op de Beeck et al., 2008), among others. Previous studies have already shown the relevance of action in explaining aspects of LOTC object space (Lignau & Downing, 2015; Kabulska et al., 2024; Tarhan et al., 2020; Tucciarelli et al., 2019). For example, overlapping responses in left LOTC between tools and hands, or tools and graspable food might reflect shared end-effector properties (Bracci et al., 2012) and action-related affordances (Ritchie et al., 2024a). Our results are in line with these previous findings and lift them up to a whole new level by revealing that a large-scale topographic organization is responsible for these earlier findings. More specifically, this approach enables us to move beyond post-hoc interpretations of visual cortex category organization (e.g., faces in lateral

FG, tools in medial FG), allowing us to generate novel predictions about the spatial organization of new object categories – to be tested in future experiments – that share similar action-related features. Based on where these categories fall within a multidimensional feature space, we can predict their alignment within the topographic layout of OTC. For instance, as food items share grasping properties with manipulable objects and are not action effectors, we expect them to map along the same action-based dimension and to partially overlap with manipulable objects, but not with hands.

Furthermore, we demonstrate that lateral and ventral OTC represent different object features, with their topographic organization exhibiting opposing response patterns that depend on the degree of action properties associated with objects. In left LOTC, the action-based topography culminated at the intersection between animate (hands) and inanimate (tools) as both being end-effectors. Dorsally and posteriorly, hands overlap with bodies and inferiorly and anteriorly, tools overlap with manipulable objects which share with tools grasping properties but not end-effector properties. This organization is consistent across participants (even in unsmoothed, native surface) and cannot be explained by differences in object size or shape as tools and manipulable objects are matched for real-word size and all object categories are controlled for their overall shape. The opposite object pattern can be observed in VOTC, with higher and more extended activation for non-manipulable than manipulable objects, and tools being embedded within the manipulable object cluster in medial VOTC. These findings challenge views that tool representations in VOTC reflect action-related properties (Mahon et al., 2007), suggesting instead that they encode general object features – such as surface properties (Cant & Goodale, 2007) or weight (Gallivan et al. 2014) – shared across manipulable and non-manipulable objects to support recognition of inanimate objects in general rather than tools specifically (Cortinovis et al., 2025b; Mahon & Almeida, 2024).

The opposite activation patterns observed in ventral and lateral OTC aligns with the proposal of a third lateral pathway dedicated to (inter)action recognition (Lingnau & Downing, 2015; Wurm & Caramazza, 2022; Pitcher & Ungerleider, 2021; Weiner & Grill-Spector, 2013; see

Ritchie et al., 2024b for a critical discussion). The studies characterizing this pathway have proposed a posterior-to-anterior organization, from perceptual to conceptual action-related and a medial-to-dorsal organization, from inanimate to animate processing and from transitive to social actions (Lingnau & Downing, 2015; Wurm & Caramazza, 2022; Papeo et al., 2019; Wurm et al., 2017). In this framework, the anatomical location of the LOTC action-based topography falls within a posterior and inferior region of the lateral visual pathway, suggesting their contribution to perceptually based action-related representations of objects.

But what is the origin of this action-based dimension? Although our experiment does not directly address this question, two alternatives might be considered. First, the action dimension might be perceptual in nature: for instance, hands and commonly used tools often visually appear together, which may explain why they are closely mapped in LOTC. According to the principle of minimizing wiring cost, which shapes known organizational patterns in both visual (Durbin & Mitchinson, 1990; Chklovskii et al., 2002) and motor cortices (Graziano & Aflalo, 2007), such visual co-activation may promote the proximity of hand and tool populations in LOTC. Alternatively, this dimension might be tied to motor experience with tools (e.g., learned associations between hands and tools during object interaction) reflecting how we engage with objects through action (but see Striem-Amit et al., 2017). Supporting this view, evidence shows that LOTC is active not only when viewing body parts or tools, but also during actual movements (Orlov et al., 2010; Astafiev et al., 2004). It is also plausible that multiple constraints might play a joint role in the emergence of this action-based topography, originating both from bottom-up visual factors (e.g., visual statistics) and top-down factors (e.g., behavioural goals) to ultimately represent object properties useful to support behaviour (Bracci & Op de Beeck, 2023; Op de Beeck et al., 2019).

Interestingly, studies have found that areas within the lateral visual pathway show higher sensitivity to dynamic than static stimuli (Beauchamp et al., 2002; Küçük et al., 2024). While the choice of static stimuli in the current study allowed us to have higher control on possible confounding variables (i.e., shape), future studies may employ dynamic stimuli such as short

video clips of people performing actions that may not only replicate but even extend the relevance of behaviourally-relevant properties in explaining the object space in LOTC (Haxby et al., 2020).

Univariate and multivariate results revealed interesting couplings between object dimensions in visual cortex. Notably, object action and object shape representations were closely intertwined in lateral OTC, offering key insights into the functional organization of high-level visual cortex. The coupling of shape and action in lateral OTC highlights how object shape directly informs interaction potential. For instance, elongation – a mid-level shape property which characterizes most tools – is known to drive responses in tool-selective cortex (Chen et al., 2018). Critically, however, our results go beyond these intrinsic associations between object category and shape (Bao et al., 2020; Coggan & Tong, 2023): even after controlling for shape, we observed robust action-shape coupling in lateral OTC, demonstrating that shape and action are distinct yet interacting dimensions.

DANNs results revealed both convergence and divergence with the functional and spatial organization of the visual cortex. Prior studies using topographic artificial neural networks (Blauch et al., 2022; Lu et al., 2025; Margalit et al., 2024) or self-organizing maps (Cowell & Cottrell, 2013; Doshi & Konkle, 2023; Zhang et al., 2024) have shown that principles like minimization of wiring length yield emergent macro- and mesoscale structures resembling those in visual cortex, including clusters for faces, bodies, scenes, and objects, and large-scale gradients of animacy and real-world size. Here, we confirm that while these networks capture the large-scale clusters based on animacy, and to a certain extent the category clusters for faces, bodies and hands, they could not capture the action-based object topography and the category clusters for the three inanimate object categories.

This failure may stem from DANNs' reliance on mid-level visual features—such as shape and texture—that often correlate with object category in natural datasets. While this works well for animate categories (possibly because of curvature features; Long et al., 2018)), it breaks down

for inanimate categories when visual features are controlled, as in our study. In these cases, DANNs default to encoding lower-level properties like orientation or aspect ratio, leading to weak category-specific clustering for inanimate objects (Figure 2.5b-d). Thus, a tight control of visual features is especially important when comparing visual cortex and DANNs, as the two systems may represent objects in an apparent similar way but actually use different visual features that are confounded in the natural environment or uncontrolled stimulus sets (Bracci et al., 2023; Mahner et al., 2025).

Neither differences in training regimes (supervised vs. self-supervised) nor in computational objectives (e.g., object vs action recognition) improved alignment with LOTC. While networks trained on action recognition did show some differences, such as a separated hand cluster compared to object-trained models (Figure 2.5d), they still failed to capture the action-related organization observed in LOTC. Why do models trained on action recognition do not show any better alignment with LOTC relative to standard object recognition models? One possibility is that the action categories used during training are too abstract. For instance, the label opening could refer to actions as different as opening a box or opening one's eyes (Monfort et al., 2021), thereby failing to isolate action-effector relationships that drive LOTC responses. More generally, although these models are trained on short video clips, rather than static images, they process actions as static patterns across frames, lacking sensitivity to temporal dynamics, predictive processing, and temporal integration that humans naturally rely on (Lake et al., 2017). Finally, human action perception is shaped not only by motion but also by social context and affordances (Chartouny et al., 2024), factors that are entirely absent from current DANN models (Lake et al., 2017). For instance, the comparison between DANNs and visual cortex is especially revealing when considering the case of shape: while both systems are sensitive to aspects of shape, such as elongation and aspect-ratio, shape information might be used for different purposes: exclusively for categorization in DANNs, where shape is indicative of category membership, and for more varied behaviorally-relevant goals in the brain, such as grasping, manipulation, and functional use of objects. This divergence may arise because

DANNs are trained on passive visual tasks (e.g., classification), whereas biological vision is inherently linked to action planning and sensorimotor experience. A promising direction may involve training models through reinforcement learning in embodied agents, where tasks are grounded in action. For example, agents could learn to evaluate an object's graspability or identify the specific parts relevant for grasping and functional use (Yang et al., 2023) or learning actions in social contexts while interacting with humans (Chartouny et al., 2024). Overall, while TDANNs represent a step forward in modelling visual cortex organization, we point to the necessity of using more ecological, varied tasks – beyond object or action classification – and the inclusion of biological constraints (Qian et al., 2024) to fully model OTC object space (but see Finzi et al., 2023).

In summary, this study demonstrates the critical role of the action dimension as an organizing principle of object representations in lateral occipitotemporal cortex. While artificial neural networks successfully replicated animacy-based organization, they failed to capture the action-based topography observed in the brain, despite their prominence in human functional organization. These findings underscore the importance of behaviorally relevant object properties in shaping the visual cortex's topography and advance our understanding of how multidimensional representations support object vision in the human brain.

Chapter 3 - Object dimensions underlying food selectivity in visual cortex

Abstract

Recent work has identified two food-selective areas in lateral and ventro-medial occipitotemporal cortex (OTC) that respond to food images independently of mid-level features like shape, texture or colour, and partially overlapping with tool responses. However, previous studies did not distinguish between ventral and lateral functional profile of food areas. Here, across two fMRI experiments, we characterize the dimensions that give rise to food selectivity in lateral and ventral OTC. Our results reveal a dissociation between lateral and ventral food-selective areas, with food selectivity emerging from distinct representational constraints in each region. Specifically, in the first study, we identified two food clusters: a ventral OTC cluster responding selectivity to food and much less to all other categories in both hemispheres, and a lateral OTC cluster exhibiting more complex responses, specifically responding to food, tools, manipulable objects, and shapes in the left hemisphere and to food, bodies, and shapes in the right hemisphere. In the second study, we orthogonally varied surface properties for object colour and ensemble statistics, and found again a dissociation between ventral and lateral areas, with colour shaping only the ventral OTC food cluster, and ensemble statistics acting in opposite direction in ventral (higher sensitivity to groups of objects) and lateral (higher sensitivity to single objects) food clusters. Finally, topographic artificial neural networks implementing architectural constraints meant to capture OTC spatial organization similarly exhibited two dissociable clusters of food-selective units based on sensitivity to ensemble statistics (but not colour). Together, these findings suggest that lateral food selectivity reflects action-relevant properties, whereas ventral food selectivity arises from sensitivity to surface-based visual features critical for food identification.

Introduction

The occipitotemporal cortex (OTC) is organized in areas that respond selectively to ecologically relevant object categories, including faces, bodies, hands, tools, and scenes (Downing et al., 2006; Kanwisher et al., 2010). These category-selective areas are embedded within broader cortical gradients that track dimensions such as eccentricity, curvature, animacy, real-world size, and action (Hasson et al., 2003; Cortinovis et al., 2025a; Konkle & Caramazza, 2013; Konkle & Oliva, 2012; Kriegeskorte et al., 2008a; Yue et al., 2020). For example, face-selective areas respond to objects that typically appear in the fovea, are curvilinear, animate, and small (Arcaro & Livingstone, 2024; Grill-Spector & Weiner, 2014). Likewise, hand- and tool-selective areas are spatially adjacent and partially overlapping, reflecting shared sensitivity to action-related properties (Bracci & Peelen, 2013). A similar nested spatial organization is observed in Topographic Artificial Neural Networks (TDANNs), a model that adds biologically-inspired constraints within its architecture by imposing correlations among neighbouring units (Margalit et al., 2024; see also Deb et al., 2025 and Lu et al., 2025).

Recent research has turned its attention to another category of objects: food. Food is among the most salient object categories in daily life. Accordingly, research has examined how the brain responds to food images, identifying activations not only in visual cortex (around the fusiform gyrus) but also in taste- and reward-related areas such as the insula and orbitofrontal cortex (Adamson & Troiani, 2018; Avery et al., 2021; Huerta et al., 2014; Simmons et al., 2005; Van der Laan et al., 2011). Recent reports revealed robust food-selective responses in ventral visual cortex (Jain et al., 2023; Khosla et al., 2022; Pennock et al., 2023), comparable in magnitude to classic category-selective responses. These studies identified two food-selective "stripes" in stereotypical locations: one medial (between face- and scene-selective areas) and one lateral to face-selective cortex. Critically, these responses could not be fully explained by low- or mid-level visual features such as colour, curvature, or texture, but instead reflected genuine food category selectivity.

Building on these discoveries, researchers have proposed several properties that may drive food selectivity (reviewed in Henderson et al., 2025). Here we focus on three candidates: two visual (colour and ensemble statistics) and one action-related (graspability/manipulability). Colour is an obvious candidate, as it conveys food-specific information such as ripeness or calorie content (Foroni et al., 2016), and it is a behaviorally-relevant property that can be exploited to rapidly recognize food objects (Sato, 2021). Selective responses to chromatic information have been documented in both macaque and human ventral visual cortex (Lafer-Sousa et al., 2013; 2016), including regions near food-selective areas. Indeed, one study argued that "colour-biased" patches are actually food-selective (Pennock et al., 2023). A less explored property is ensemble configuration: food often appears as collections of similar elements (e.g., peas, spaghetti, a bunch of grapes). While ensemble perception is well studied behaviourally (Alvarez, 2011; Whitney & Yamanshi Leib, 2018), its neural bases in ventral OTC are less clear. Some evidence points to regions near the parahippocampal place area that respond to textures and ensembles (Cant & Xu, 2012), potentially overlapping with medial food-selective cortex. Finally, food is inherently action-related: it must be grasped and manipulated before consumption. A recent study reported overlapping food and tool activations in ventral and lateral OTC, suggesting graspability/manipulability as a potential driving factor for its selectivity (Ritchie et al., 2024a).

However, while food-selective responses have been identified in both ventral and lateral OTC, there is currently no clear evidence for a distinct computational role among the two clusters. One study reported a partial dissociation between the two based on image content (close-up views of food vs. food-related scenes) using PCA (Jain et al., 2023), but most analyses have either not tested for, or failed to find, differences in their feature sensitivity. In light of recent accounts proposing a division of labour between ventral and lateral OTC in supporting object recognition versus action processing (Lingnau & Downing, 2015; Wurm & Caramazza, 2022), food represents an ideal category to test whether ventral and lateral regions contribute differently to food-related processing in relation to their respective computational goals.

Based on this evidence, we hypothesize that food selectivity in visual cortex reflects two dimensions: 1) a visual dimension capturing surface-level properties such as colour and ensemble statistics, that might be relevant for object recognition and thus likely drive responses in the ventral OTC pathway; and 2) an action-related dimension, defined by the type of interaction an object affords, that might be relevant for action processing, and thus likely driving responses in the lateral OTC pathway. To test these hypotheses, we conducted an fMRI study in which participants completed (1) an experiment to identify food-selective areas and their topographic relationship to classic category-selective areas (the *category* experiment), and (2) a *surface property* experiment investigating the contributions of visual properties (colour and configuration) to food selectivity. Across the two studies, we observed a robust dissociation between ventral and lateral food-selective areas: ventral clusters were primarily tuned to surface properties, especially colour and, to a lesser extent, ensemble configuration, whereas lateral clusters were more strongly linked to graspability and action-related properties of food stimuli, consistent with the hypothesized role of each pathway. Performing the same analyses on the TDANNs revealed that these networks similarly organized food responses in two clusters that could be dissociated based on their sensitivity to ensemble statistics information, resembling the dissociation in visual cortex; however, no colour organization could instead be observed across these clusters.

Methods

fMRI experiment and analyses

Participants

A total of 20 participants took part in the fMRI study, participating in both experiments. Two participants were excluded due to excessive head motion (exceeding one voxel in translation or rotation), resulting in a final sample of 18 participants (9 females, sex self-reported; mean age = 24 years, SD = 3.3). All participants were right-handed, had normal or corrected-to-normal vision, and no history of neurological disorders. Written informed consent was obtained from all participants, and the experimental procedures were approved by the Ethics Committee of the University of Trento.

Stimuli

The *category* experiment included 8 categories. The set comprises faces, headless bodies, hands, food, tools, manipulable objects, scenes, and meaningless spiky shapes. Food category included both graspable food (e.g., hot-dog, sandwich) and non-directly graspable food (e.g., pasta, fruit salad). Tools were defined as hand-held objects that are typically used to physically and directly act on another object or surface (e.g., pliers, knife), and generally manipulable objects were defined as objects that can be grasped but are not usually used as action-effectors (e.g., bedside lamp, book); tools and manipulable objects were matched in terms of their overall shape and orientation (Bracci & Peelen, 2013). Each category included 48 greyscale squared (400x400) images with a white background. Part of the images were used in Matić et al. (2020) and Cortinovis et al. (2025a); part of the images of food were selected among the localizer developed by Jain et al. (2022) and Ritchie et al. (2024a); meaningless spiky shapes were taken from Op de Beeck et al. (2006); all other images were obtained through internet searches.

The surface properties experiment included 8 conditions, varying along three orthogonal dimensions: 1) category (food vs inanimate objects), 2) colour (coloured vs. greyscale), and

3) ensemble/configuration (images were either presented as groups of homogenous ensembles of objects, such as a group of apples, vs. the corresponding single object, such as a single apple on a naturalistic background). The stimuli were selected to minimize potential confounds between categories. For instance, inanimate objects included images that were as colourful as food images, and food stimuli were chosen to span a wide range of colours including both warm and cool tones. All images were sourced through internet searches. Examples of stimuli for both sets are shown in Figure 3.1.

Scanning procedure

The fMRI study took place in two sessions in two separate days within a week. In the first session we collected the data for the *category* experiment (6 runs per participant) and in the second session the surface properties experiment (6 runs per participant). The anatomical scan was collected during the first session. Additional data were collected for a separate experiment (not reported here). The design was the same for both experiments. Each run lasted 336 seconds (168 volumes). Images were presented for 400 ms each, with an inter-stimulus interval (ISI) of 266 ms, organized into 8-second blocks (12 images per block). For each participant and run, a fully randomized sequence of all conditions was repeated 4 times, separated with 16-second fixation blocks at the beginning and end of the run. The entire experiment therefore consisted of 24 repetitions of each category block. Stimuli were displayed using the Psychophysics Toolbox (Brainard, 1997) with MATLAB (version 2021b; The MathWorks) and projected onto a screen subtending 8 × 8 degrees of visual angle, viewed via a mirror on the head coil. Participants were instructed to maintain fixation on a central cross and press a button whenever the same image appeared twice consecutively within a block (one repeat per block). Behavioral performance was quantified by calculating response accuracy (mean across experiments = 93.7%, SD = 3.1%) and reaction time (RT; mean across experiments = 588.6 ms, SD = 25.8 ms). Accuracy was defined as the proportion of correctly identified target stimuli, with responses considered correct if made within two trials following the target, accounting for the rapid stimulus presentation (400 ms).

Imaging parameters

The fMRI data was collected using a 3T Siemens scanner with a 64-channel head coil in the Center for Mind/Brain Sciences at the University of Trento. MRI volumes were collected using echo planar (EPI) T2*-weighted sequence, with repetition time (TR) of 2 s, echo time (TE) of 28 ms, flip angle (FA) of 75°, and field of view of 220 mm. Each volume contained 69 axial slices, covering the whole brain, with matrix size 200x200 mm and 2x2x2 mm voxel size. Slices were acquired with a multiband (multi-slice) sequence, with slice acceleration factor = 3. Anatomical images were acquired using the T1-weighted acquisition and MP-RAGE sequence, with a resolution of 1x1x1 mm.

Preprocessing

Preprocessing was performed using the Statistical Parametric Mapping software package (SPM12, Wellcome Trust Centre for Neuroimaging, London) and MATLAB (R2021b, The MathWorks). Functional images underwent the following preprocessing steps: spatial realignment to the first image to correct for head motion, slice-timing correction, coregistration with anatomical images, and spatial smoothing using a Gaussian kernel (4 mm FWHM). When performing group-level analysis, the data was normalized to the Montreal Neurological Institute's (MNI) ICBM152 template. Prior to preprocessing, we defined participants' exclusion criteria as follows: runs where head movement exceeded one voxel (in translation or rotation, 2 mm) were excluded; if more than half of the runs (6 total per experiment) were excluded for a given participant, that participant was fully excluded from further analysis. Based on this criterion, two participants were excluded entirely, along with 3 runs from 3 participants (one run each) for the *category* experiment and 4 runs from 3 participants (one run in two participants and two runs in one participant) for the surface properties experiment. The preprocessed signal was modelled for each voxel, for each participant, and for each condition using a general linear model (GLM). For both experiments, the GLM included 8 regressors of interest (one per experimental condition) and 6 nuisance regressors (motion correction

parameters: x, y, z for translations and rotations). Predictors' time courses were modelled by convolving the haemodynamic response function (HRF) with a boxcar function.

Data analysis

Whole-brain univariate analysis

First, we performed a group random-effects analysis to visualize the spatial clustering and the reciprocal relationship of category-selective areas for both experiments. Each condition in the action experiment was contrasted against the average of all the others, whereas each condition in the surface properties experiment was contrasted against the opposite condition (i.e., colour vs. grayscale, ensemble vs. single, food vs. objects). Results were thresholded at $p < .001$ uncorrected at the voxel level and $p < .05$ FDR corrected at the cluster level and visualized on a Freesurfer average surface using freeview (Fisch, 2012).

ROI definition

We identified ROIs at both the individual and at the group level for both studies. First, for both experiments, we identified areas selectively responding to food images on the native volume of each individual. These areas were defined with a contrast of category vs. average of all others, with a threshold of $p < .001$ uncorrected or – if no ROI could be selected at this threshold – $p < .01$ uncorrected. If necessary (e.g., when the activation in ventral or lateral formed a contiguous cluster with each other or with activations in early visual cortex), ROIs were restricted to a cube of 6 mm width centred on the activation peak. The areas defined in one experiment were used to test functional selectivity in the other experiment and viceversa, thus ensuring independence of the data used for statistical analysis from the ROI definition. Food areas were defined bilaterally and independently for ventral and lateral OTC.

To test overlap between categories, we identified additional ROIs using the same criteria as above. For the category experiment, we identified face, body, hand, tool, and scene-selective areas with a contrast of $p < .001$ uncorrected. Manipulable objects activations could not be

defined in most participants at either $p < .001$ or $p < .01$ uncorrected). All ROIs were defined bilaterally, except tool-selective areas that exhibited a strong left-lateralization.

Functional selectivity analyses

We examined the functional selectivity of the ROIs in both the action experiment and the surface properties experiment. In the *category* experiment, we assess food selectivity relative to a range of visual categories, to test whether food-selective areas are truly selective to food stimuli or whether they also respond to other categories based on their shared action related properties. In the surface properties experiment, we evaluate how ventral and lateral food areas might differentially respond to surface properties such as colour and ensemble statistics. Importantly, by defining ROIs using data from the other experiment, we ensured statistical independence between ROI definition and functional selectivity assessment.

Overlap analyses

Anatomical overlap between ROIs was measured for both experiments. We used the same ROIs as defined for the functional selectivity at the native level, and adopted a procedure used in previous studies (e.g., Bracci et al., 2012). Specifically, we calculated the number of voxels in common between two ROIs (e.g., hands and food or ensemble and colour) and divided it by the smaller of the two ROIs. This gives us an index ranging from 0 (no overlap between the two categories) to 1 (full overlap, where the smaller of the two ROIs falls completely within the bigger of the two).

Vector-of-ROIs

To investigate the general topographic organization of category-selective areas and dimensions in visual cortex, we conducted a vector-of-ROIs analysis. This procedure involves the construction of a series of partially overlapping spheres along a vector. The vector was generated by fitting a spline connecting two points: one medial, around the parahippocampal

gyrus, and one dorsal and lateral, around the transverse occipital sulcus. Coordinates for these two points were taken from Konkle & Caramazza (2013). Between these two points, a series of anchor points was defined to constraint the vector to pass through previously known areas responding to specific object categories. The anchor points coordinates were based on published studies on category selectivity and specifically passed through areas responding to faces (Julian et al., 2012), small objects (Konkle & Oliva, 2012), tools (Bracci et al., 2012), and bodies (Julian et al., 2012). In addition to these areas, we also added an anchor point that was based on the coordinates of the food-selective area in the ventral OTC, to make sure that the vector passed through the relevant regions showing food preference. The coordinates were selected independently across experiments: for the category experiment data, we used the coordinates corresponding to the peak of food selectivity identified in the surface properties experiment, and vice versa. Along the vector, a series of spheres of 5 mm, spaced 3 mm between each other, was generated. We extracted the beta values for both experiments and plotted the activation for each condition in each of the sphere and analysed their functional profile.

Representational similarity analysis

To investigate the representational space underlying food-selective areas we adopted a Representational Similarity Analysis (RSA) approach (Kriegeskorte et al., 2008). First, we computed correlation matrices by calculating the pairwise correlations of all conditions (separately for both experiments) based on the food-selective ROIs, leading to 8x8 matrices (for bilateral ventral and lateral food ROIs for both experiments).

Then, for the category experiment we performed further calculations on the similarity patterns. Specifically, to quantify the multivariate relationships between stimulus categories, we performed an index analysis based on the correlation matrices obtained from the RSA. We computed two action indices – food indices and hand indices – to assess the relative similarity of each category to other action related categories. This action relation has two components:

graspability\manipulability (that is shared by food, tools, and manipulable objects), and end-effector specificity (that is shared by hands and tools exclusively). The food index measures the similarity between food and other object categories (tools, manipulable objects, and hands). For each participant, we calculated the pairwise correlations between food and each of the other categories (food–tool, food–manipulable, and food–hand). To isolate the relative similarity of each pairing, we subtracted the average of the two remaining correlations from the correlation of interest: i.e., food&tool – avg(food&hand and food&manipulable); food&manipulable – avg(food&hand and food&tool); food&hand – avg(food&tool and food&manipulable). The hand index were computed analogously, reflecting the similarity between hands and other object categories: i.e., hand&tool – avg(hand&food and hand&manipulable); hand&manipulable – avg(hand&tool and hand&food); hand&food – avg(hand&tool and hand&manipulable). These indices quantify whether a given category (food or hand) shows greater similarity to action related categories compared to unrelated ones, thus providing a concise metric of the representational organization observed in the category experiment data.

For the surface properties experiment we correlated the obtained matrices with two theoretical models that captures the two orthogonal dimensions as by design of our stimulus set: colour (coloured vs greyscale images) and ensemble (groups of objects vs single object). We ran semi-partial correlations between the models and each of the ROIs and tested for the ability of each model to explain independent portions of variance. Once again, ROIs from independent data were used (i.e., food ROIs from the category experiment were used for the colour experiment and viceversa). As by design, the models were orthogonal ($r = 0$). The lower bound of the noise ceiling was calculated by correlating each participant matrix with the group average matrix (iteratively excluding the participant that is being correlated).

Topographic Artificial Neural Networks

We tested a model – the Topographic Artificial Neural Network (TDANN) – of the spatial organization of the ventral visual stream to investigate if the set of principles used by the model leads to a similar clustering and functional responses of food images as in ventral and lateral OTC. We adopted the same model developed by Margalit et al. (2024) and used in our previous work (Cortinovis et al., 2025a). Here, we briefly describe its main characteristics.

The TDANN is based on a ResNet-18 architecture; it includes nine topographic layers (meant to capture the entire hierarchical organization of the ventral visual stream); each topographic layer is a grid of units, and each unit is assigned to a specific coordinate before training based on preset optimizations such as coarse retinotopic mapping. During training, the model concurrently optimizes two loss functions: a task loss, where the model is trained with a self-supervised contrastive learning task (SimCLR; Chen et al., 2020) on ImageNet (Deng et al., 2009), and the spatial loss, where units that are neighbouring in the layer must have correlated patterns; the strength of correlation is determined by a parameter called alpha, that is set at 0.25. After training, we submit the model to the same analysis as done in visual cortex. We focus on the last "VTC-like" topographic layer, meant to capture the organization of high-level ventral visual cortex. To ensure replicability of results, we test five random initializations of the model's weights.

Specifically, we performed simulated univariate analysis by testing the topographic organization and selectivity profile of the five different random initializations of the network in response to our eight object categories. Specifically, we tested (1) the clustering of units selective for the different object categories within the simulated physical cortical space in the VTC-like layer and (2) the selectivity profile of the top-25 most selective units for each category in the VTC-like layer. For more precise functional selectivity analysis, we identify spatially contiguous clusters of food-selective units in the VTC-like layer by adopting a patch-based clustering approach using the DBSCAN algorithm (Ester et al., 1996). Food selectivity was

first quantified in each unit using a t-statistic comparing responses to food images against all other categories. Units exceeding a fixed selectivity threshold ($t > 4.5$) were considered supra-threshold candidates. DBSCAN clustering was then applied to the spatial coordinates of these units to identify contiguous patches, with a fixed neighbourhood radius of 0.3. For each model instance, we analysed the two largest significant food-selective patches. The procedure was performed on both experiments stimulus sets. We ran the same functional selectivity analyses as done in visual cortex: briefly, we analysed the responses of food-selective units separately for the two clusters; we used the clusters identified with the category localizer to test the responses of the conditions in the surface properties experiment and viceversa.

Results

In this study, we investigated the potential differential profile of food responses in the ventral and lateral occipitotemporal cortex (OTC). Specifically, the aim was to test if these regions exhibit differential sensitivity to action-related versus visual surface properties reflecting the functional distinction between the two pathways. Two fMRI experiments in the same group of participants were conducted. In the *category* experiment, participants viewed stimuli from eight object categories to examine the relationship between food selectivity and other categories, specifically those characterized by action-related properties, such as graspability/manipulability (shared with tools and manipulable objects) and end-effector properties (shared by hands and tools but not by food); in the *surface properties* experiment, participants were presented with images of food and objects that varied orthogonally along two dimensions: colour (coloured vs. grayscale images), and configuration (ensemble vs. single objects), to test how surface properties modulate responses in food-selective areas. Example stimuli from both experiments are shown in Figure 3.1.

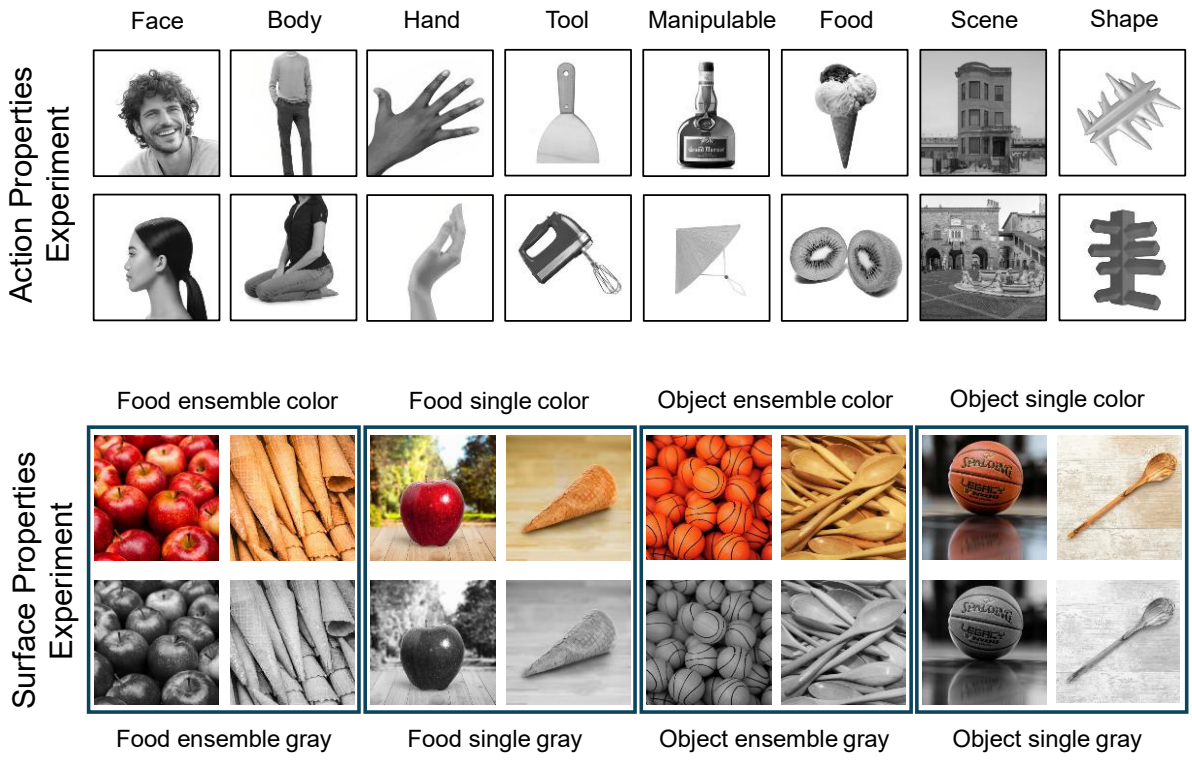


Figure 3.1. Stimulus set. Examples of images from the category experiment (above) and the surface properties experiment (below).

The topographic organization of food-selective areas in ventral and lateral OTC

Using data from the *category* experiment, we mapped food-selective activations and examined their spatial organization relative to other category-selective areas at the group level ($p < .001$ uncorrected at the voxel level and $p < .05$ FDR-corrected at the cluster level) across both ventral and lateral OTC (Figure 3.2). Our results replicated and extended recent reports (Jain et al., 2023; Khosla et al., 2022; Pennock et al., 2023; Ritchie et al., 2024a) highlighting two separate food clusters: one in ventral OTC, located between face- and scene-selective areas, and one in lateral OTC, strongly lateralized in the left hemisphere.

The lateral food-selective cluster co-localize in regions selective to tools and other manipulable objects, anterior to – but nonoverlapping with – hand-selective areas. Tool-selective responses extended more posteriorly, overlapping partially with the hand-selective cluster, while the body-selective activation (smaller in the left than in the right hemisphere) was even more posterior, overlapping with hand- but not tool-related clusters. This result is consistent with our recent proposal (Cortinovis et al., 2025a) that in lateral OTC, object representations are organized according to an action-related principle progressing from effector-specific to graspable/manipulable properties along a posterior-to-anterior axis, therefore predicting that food-selective areas should emerge in locations associated with graspable (e.g., tools and manipulable objects), but not action effector only representations (e.g., hands, but see Ritchie et al., 2024a). In line with this prediction, an overlap analysis (Figure 3.2c) performed on single-subject native space maps (see Methods) confirmed that in lateral OTC, food voxels overlap substantially with tool voxels (score = 0.51, $p < .017$) but not with hand voxels (score = 0.03, $p = .13$). In contrasts, hand and tool areas showed substantial overlap (score = 0.35, $p < .017$).

Next, to visualize the relation between food representations and object surface properties, using data from the surface properties experiment, we mapped each property – colour and ensemble statistics – and assessed their relative spatial arrangement with food areas within ventral and lateral OTC (Figure 3.2b). The food > object contrast confirmed the results observed in the category experiment, revealing two food clusters: one in ventral and one in lateral OTC. As for surface properties, the ensemble configuration contrast (ensembles > single objects) produced robust activation in bilateral early visual cortex extending more anteriorly to ventral OTC, whereas the reverse contrast (single objects > ensembles) yielded more lateral (and anterior) activations. While the contrast (grayscale > colour) did not reveal any activation, the colour contrast (colour > grayscale), revealed multiple colour-selective clusters in ventral OTC in line with prior findings (Lafer-Sousa et al., 2016; Pennock et al., 2023): a posterior cluster likely corresponding to classically defined area V4, and a more anterior cluster in close proximity to the ventral food cluster. The overlap analysis (Figure 3.2d) performed on single-subject native space maps (see Methods) quantified these spatial relationships. The posterior colour cluster showed minimal overlap with ventral food clusters and failed to survive correction for multiple comparisons in either hemisphere (left: score = 0.07, $p = .013$; right: score = 0.04, $p = .04$; Bonferroni $p > .0125$). In contrast, the anterior colour cluster exhibited substantial overlap with the ventral OTC food clusters in both hemispheres (left: score = 0.43; right: score = 0.29; both $p < .0125$). Finally, no overlap was observed between the lateral OTC food clusters and any colour selective region, consistent with the known ventral–lateral dissociation of colour responses (Brouwer & Heeger, 2009).

Together, results from the two experiments suggest a dissociation between ventral and lateral OTC: lateral food clusters reflect graspable properties shared by food, tools, and other manipulable objects, whereas ventral food clusters are related to surface properties, especially colour.

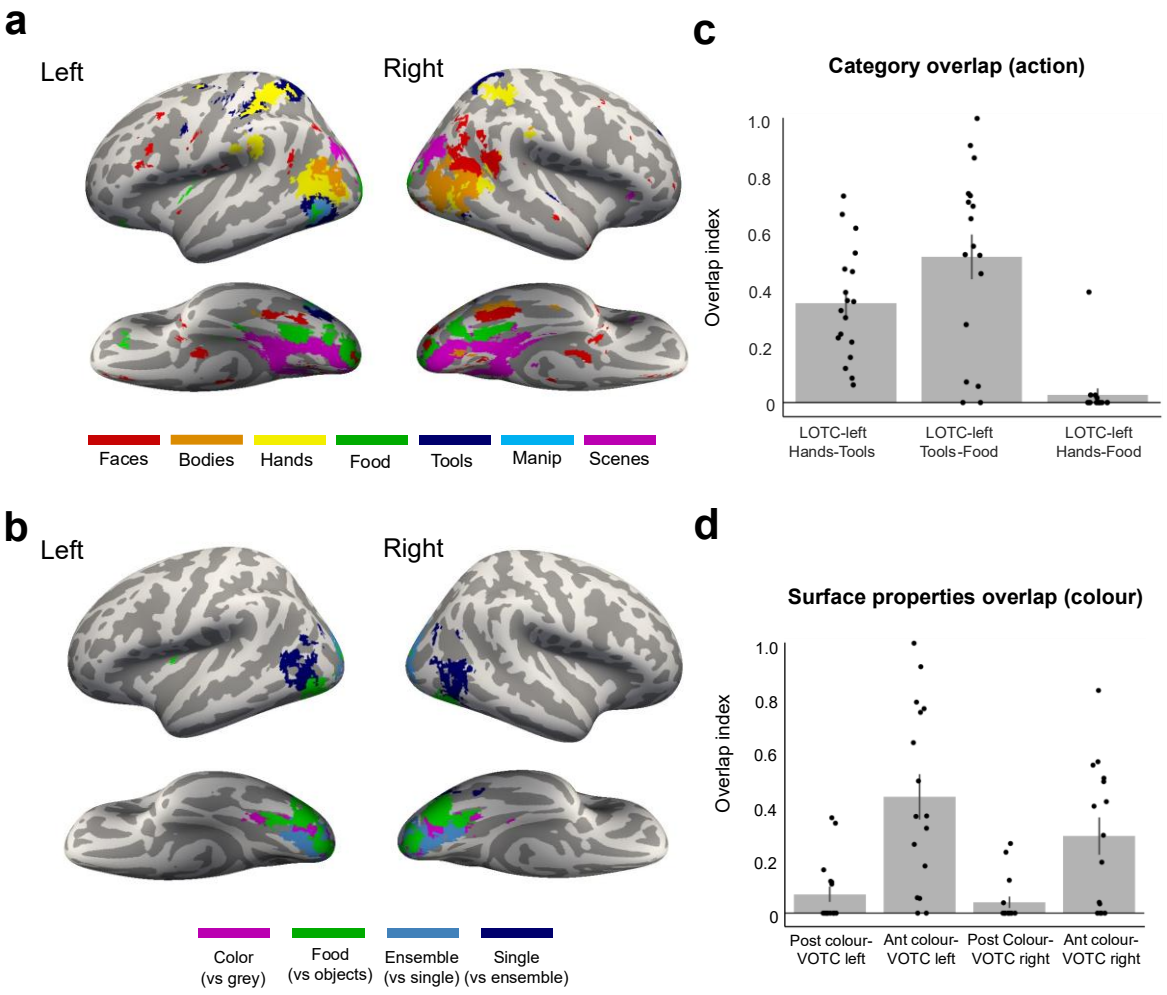


Figure 3.2. The topography of food-selective areas in occipitotemporal cortex. a) Group-level whole-brain results for the category experiment. Response for each category (vs. all) was visualized on a freesurfer average brain surface using freeview, with a threshold of $p < .001$ uncorrected at the voxel level and $p < .05$ FDR corrected at the cluster level). **b)** Group-level whole-brain results for the surface properties experiment. Same visualization as a) **c)** Overlap index is plotted for each pairwise overlap among three categories (hands, food, tools). **d)** Overlap index is plotted for each pairwise overlap among two conditions (colour and food) and four areas (VOTC left and right, and posterior and central colour clusters. Error bars represent ± SEM across subjects. Each data point reflects the overlap index value from one subject.

Distinct functional profile of ventral and lateral food-selective areas

To characterize the functional tuning of ventral and lateral OTC food areas, ROIs were identified independently in each participant's native volume (see Methods) and used to measure responses to the stimulus categories within each study. Across both studies, results confirm a functional dissociation between ventral and lateral OTC food areas.

In the *category* experiment, VOTC-food areas showed exclusive selectivity, responding strongly to food, with minimal activation to other categories (all contrasts $p < .001$). In contrast, LOTC-food areas exhibited a more complex response profile based on lateralization. In the left hemisphere, responses to food were higher than baseline (defined as the mean of all the other categories; $t_{(17)} = 9.2$; $p < .001$) but food responses were indistinguishable from tools and meaningless shapes ($p > .05$), which, together with manipulable objects, all elicited significant higher responses than baseline (all $p < .001$), as predicted by the left lateralized action gradient (Cortinovis et al., 2025a). In the right hemisphere, LOTC-food similarly exhibited higher responses to food than baseline ($t_{(17)} = 6.1$; $p < .001$) but responded equally strongly to food, shapes, and bodies ($p > .05$) but significantly higher than tools ($t_{(17)} = 3.3$; $p = .004$) and manipulable objects ($t_{(17)} = 4.4$; $p < .001$).

This pattern aligns with the anatomical positioning of food-selective areas within OTC. VOTC-food areas, located between face- and scene-selective areas, appear to occupy a distinct territory not strongly engaged by other category domains. LOTC-food areas, by contrast, overlap with regions implicated in action-related object processing (i.e., manipulable objects and tools) in the left hemisphere, and with the body selective area in the right hemisphere.

In the *surface properties* experiment, VOTC exhibits a complex response pattern driven by the interplay of category preference, colour, and ensemble statistics. In VOTC-food, robust activation was observed for coloured food images ($p < .001$, Bonferroni corrected for 8 comparisons) but not for greyscale versions ($p > .05$). This effect held regardless of whether the stimuli were depicted as single objects or ensembles (both $p < .001$). Interestingly,

inanimate objects also elicited a strong response indistinguishable from food ($p > .05$) but specifically when presented as coloured ensembles ($t_{(16)} = 4.9$; $p = .002$ in the left hemisphere; in the right hemisphere the effect did not survive correction for multiple comparisons; $t_{(17)} = 2.5$; $p = .03$). This high response was not observed for single-coloured objects ($p = .4$) and all grayscale objects (regardless of format) failed to elicit comparable activation (all $p > .05$).

In contrast to VOTC, LOTC food exhibited a response profile driven by object configuration rather than colour. Specifically, while LOTC-food retained a strong preference for food over inanimate objects (all $p < .001$), this response was not modulated by colour ($p > .05$). Instead, the region displayed a bias toward single-object processing: responses tended to be higher for single food items compared to ensembles, although this difference did not reach statistical significance ($p = .08$).

Overall, this pattern suggests a functional dissociation between ventral and lateral food-selective areas in OTC, with the former sensitive to surface properties of objects, such as colour and ensemble statistics, whereas the latter is tuned to single object configuration but not to colour.

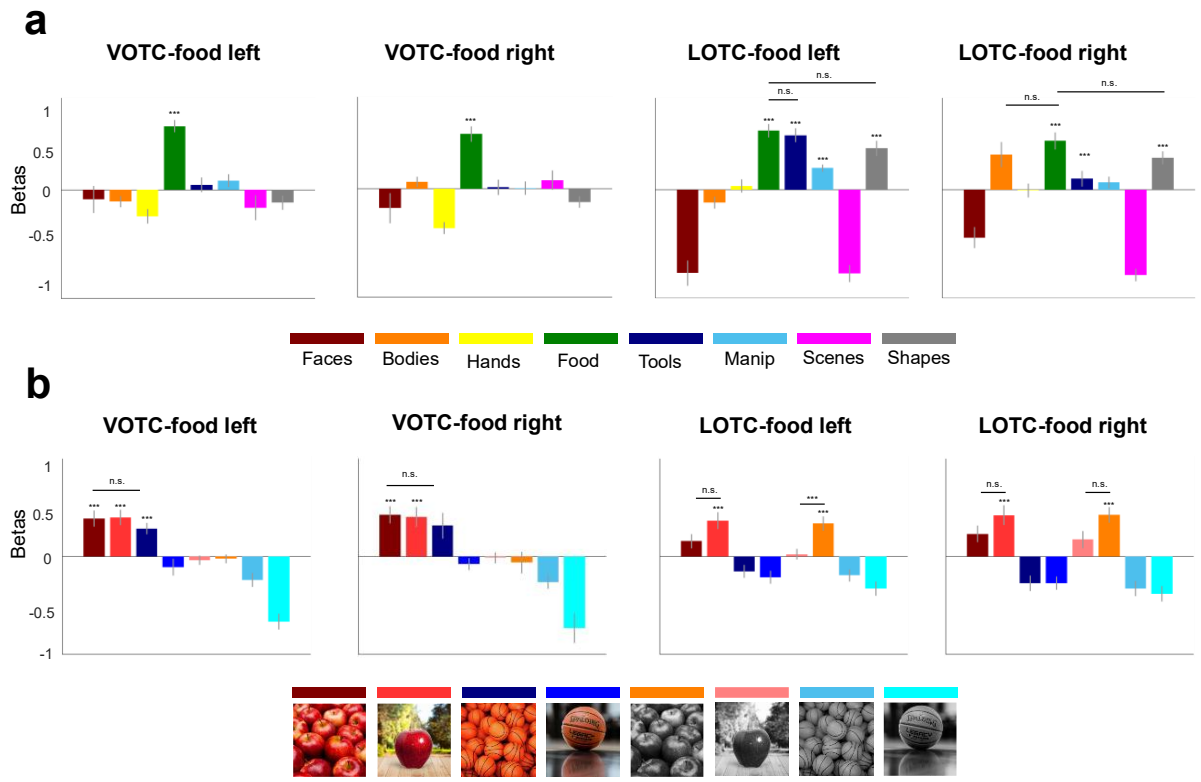


Figure 3.3. Functional selectivity analysis. Normalized beta values (against the average of all categories) are plotted for each category in the four food-selective areas identified in each individual participant in ventral and lateral OTC for **a)** the Category Experiment and **b)** the Surface Properties Experiment. Error bars represent ± 1 SEM across subjects. Stars represent statistical significance (vs. baseline) Bonferroni corrected at $p < .00625$.

These results were confirmed with the vector of ROI analysis, which samples a broad portion of OTC by generating spheres along a vector from the parahippocampal gyrus to the transverse occipital sulcus (see Methods). For the *category* experiment, given the relationship found between food, tools, and manipulable objects, we focus on the left hemisphere (right hemisphere results can be found in the supplementary information). For the surface properties experiment, results were averaged across hemispheres as no hemispheric differences were found. Results are shown in Figure 3.4.

Overall, this analysis replicated the functional and topographic patterns described earlier. For the *category* experiment, in ventral OTC, responses to food peaked around the medial fusiform

gyrus between scene (parahippocampal cortex) and face (fusiform gyrus) selective cortex, with numerically higher activations compared to inanimate objects. In lateral OTC, by contrast, the response profile for food follows the same trajectory of other object categories (tools, manipulable objects, and meaningless shapes). Moving posteriorly, response profiles transitioned from food, then tools, to hands and finally to bodies.

For the *surface properties* experiment, ventral OTC, around the medial fusiform gyrus (showing strong food responses in the *category* experiment) responded robustly to coloured ensemble stimuli, regardless of whether depicting food or objects, and images of coloured food when presented as single objects. In contrast, lateral OTC responded preferentially to food, independent of colour, but only when images were presented as single objects, not ensembles.

Taken together, these findings confirm a dissociation between ventral and lateral visual cortex in relation to food specific processing: both pathways exhibit high activation to food stimuli, but responses to food in the ventral pathway are more selective and dissociated from the responses to other inanimate objects and represent surface properties such as colour and ensemble statistics. On the contrary, food responses in the lateral pathway are associated with other inanimate object categories, closely replicate the organizational action gradient reported in our previous report (Cortinovis et al., 2025a), and food elicits a response only in a single-object configuration, suggesting a dependence on discrete, graspable shape structure rather than surface cues. Together, these results suggest that the two pathways process distinct information, with the ventral OTC representing primarily surface properties and the lateral OTC sensitive to action-related properties. In the following sections we turn to multivariate analysis to further characterize this dissociation.

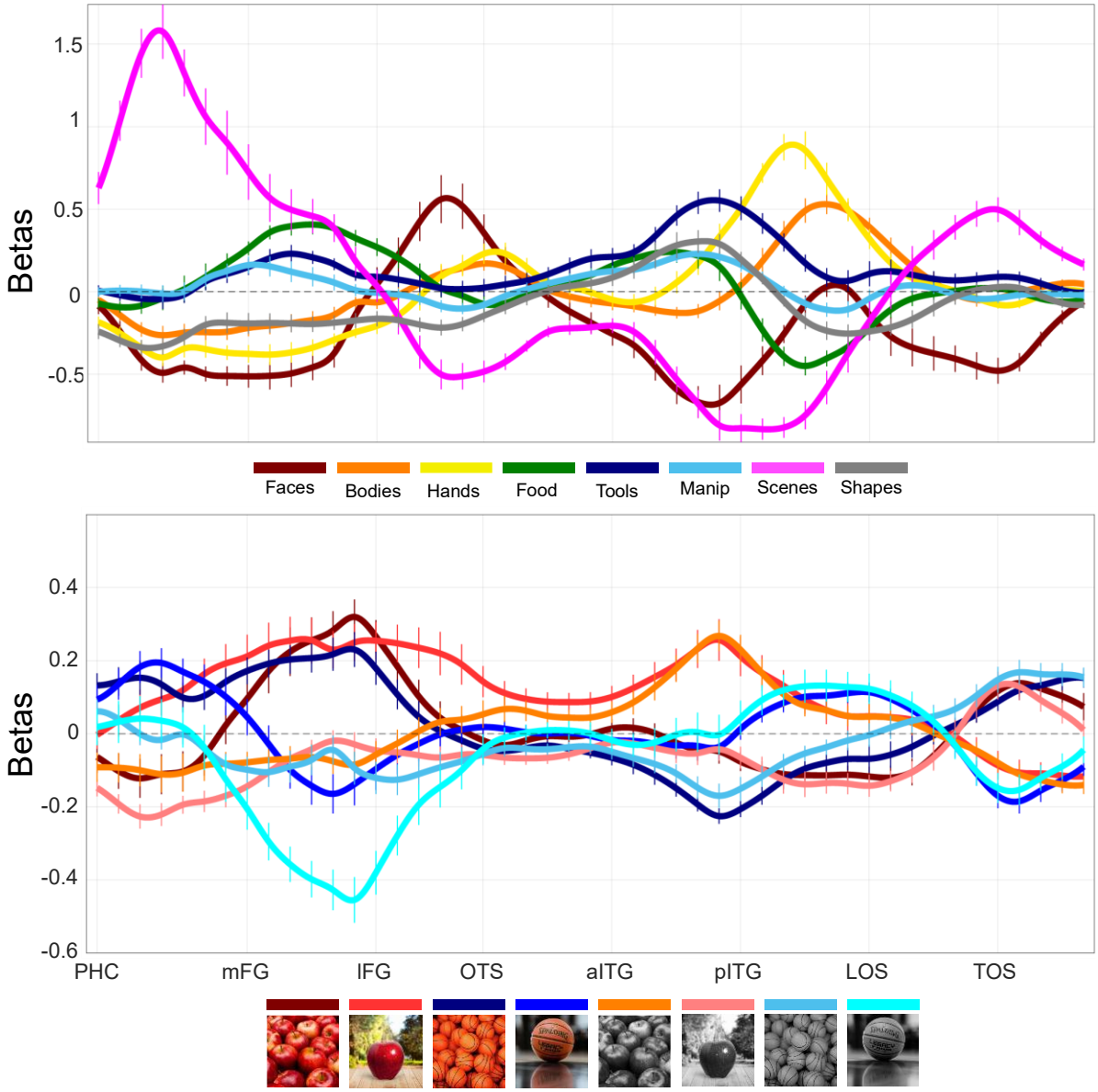


Figure 3.4. Vector-of-ROI analysis. A spline connecting distinct anchor points across ventral and lateral OTC is fitted and partially overlapping spheres (corresponding to the ROIs analysed) along the spline are generated. Functional selectivity is extracted from each sphere along the vector (see methods for details). The activation for each category is plotted as a normalized activation against the average of all other categories. *Top:* vector-of-ROIs for the *category* experiment (only left hemisphere); *Bottom*: vector-of-ROIs for the *surface properties* experiment (averaged across hemispheres). The x-axis corresponds to each sphere along the vector, with labels for major anatomical landmarks; the y-axis corresponds to the normalized beta values. Error bars represent ± 1 SEM across participants (n = 18). PHC = Parahippocampal Cortex. mFG = medial Fusiform Gyrus. lFG = lateral Fusiform Gyrus. OTS =

Occipitotemporal Sulcus. aITG = anterior Inferior Temporal Gyrus. pITG = posterior Inferior Temporal Gyrus. LOS = Lateral Occipital Sulcus. TOS = Transverse Occipital Sulcus.

Multivariate analysis reveals the object space underlying food responses

The above results showed that ventral and lateral pathways both contain food responses but differ in their sensitivity to food-related properties. To examine the multivariate representational structure of food-selective areas within the two pathways and their relation to other categories, we computed representational similarity matrices (RSMs) for each of the four food-selective ROIs identified in the independent *surface properties* experiment (see above). For each ROI, we calculated pairwise correlations among the response patterns elicited by the eight stimulus categories, yielding 8 × 8 RSMs. The resulting RSMs for the four food ROIs are shown in Figure 3.5. To better characterize the representational distances between key categories, and more specifically to test the relationship in object space between categories characterized by distinct action-related properties, we derived two indices from these matrices (see methods for details). The food (action) index quantified the representational distance between food and other action-related categories (hands, tools, and manipulable objects). The hand (action) index quantified the distance between hands and other inanimate categories (food, tools, and manipulable objects). Together, these indices provide a measure of the association in object space of food, inanimate objects, and hands.

The multivariate patterns closely mirrored the univariate findings. As illustrated in Figure 3.5, the food index revealed that food-selective ROIs exhibited high similarity between food and inanimate categories sharing graspability (tools and manipulable objects) whereas food representations were strongly dissimilar from hands. One-sample two-tailed t-tests revealed that across all ROIs, food–hand similarity was significantly lower than both food–tool and food–manipulable similarity (all Bonferroni-corrected $p < .013$), indicating a consistent representational separation between food and hands that persists across ventral and lateral OTC and in both hemispheres.

The hand indices revealed a strong association between hands and tools specifically in the left lateral OTC, with no further robust patterns observed in other regions. One-sample two-tailed t-tests showed that in the left LOTC, hands were represented as significantly more similar to tools than to either food or manipulable objects (both Bonferroni-corrected $p < .017$). Importantly, this effect cannot be explained by an animacy distinction, as hands and tools, which differ in animacy, exhibited the highest similarity, exceeding hand–food and hand–manipulable similarity in line with the pattern observed for the food index. This selective hand–tool similarity was not observed in bilateral VOTC or right LOTC ($p > .05$). Furthermore, hand–tool similarity was significantly greater in left LOTC compared to each of the other ROIs ($p < .001$), highlighting a unique functional profile in this region in which hands and tools occupy the closest positions in representational space, while hands and food occupy the most distant positions.

These multivariate results replicate and extend our previous findings (Cortinovis et al., 2025a), reinforcing the proposed action-related organizational gradient: food, tools, and manipulable objects share representational properties linked to graspability (and inanimacy), whereas hands and tools (but not food) share effector-related dimensions, and hands and food remain largely distinct in representational space. Having characterized the action feature space that differentiate lateral and ventral food selective areas, in the next section we turn to the characterization of the surface properties that also differentiate between the two.

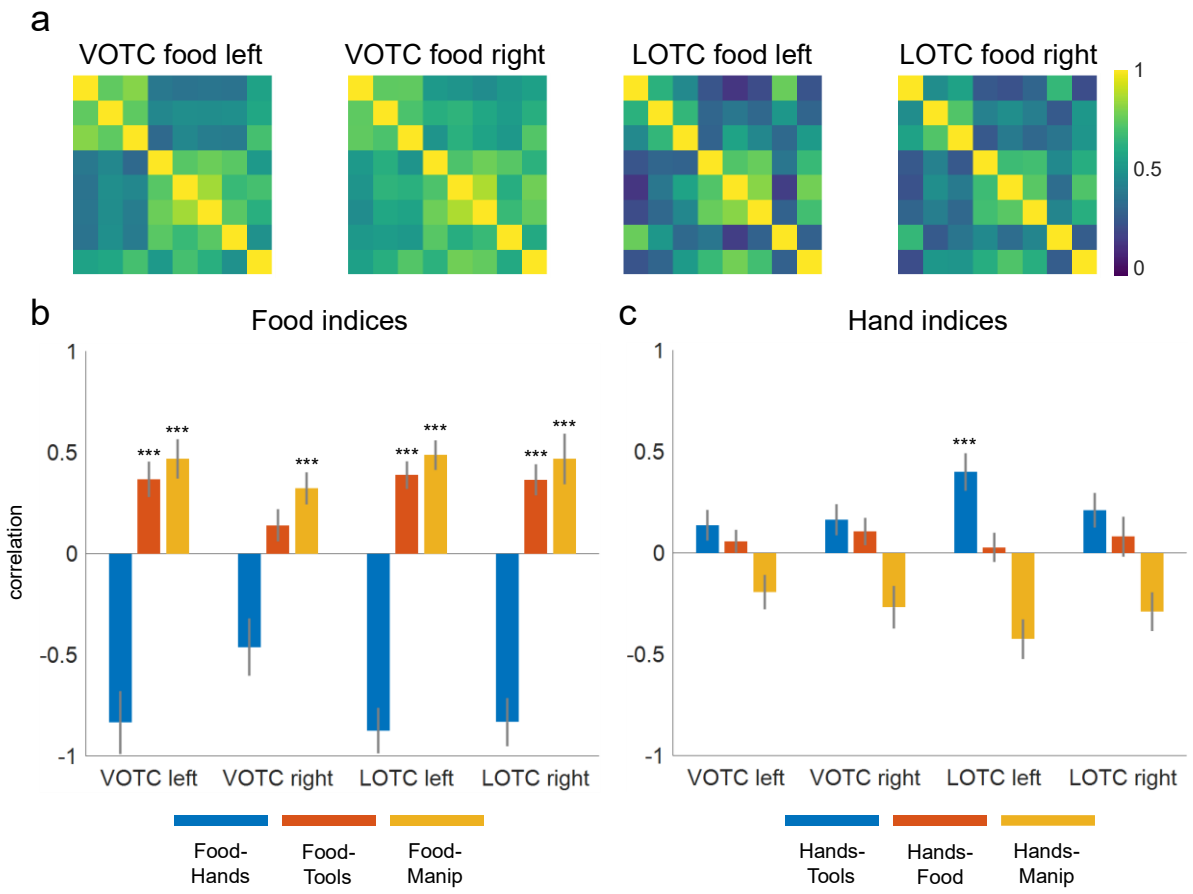


Figure 3.5. Multivariate analysis. a) Representational Similarity Matrices for the four food selective areas. The patterns of responses for the eight conditions were correlated with each other to obtain an 8x8 similarity matrix, one for each food-selective area in ventral and lateral OTC and in the left and right hemisphere. **b)** Food index analyses. The indices indicate the relationship in multivariate space between food and other categories (tools, manipulable objects, and hands). **c)** Hand index analysis. The indices indicate the relationship in multivariate space between hands and other categories (tools, manipulable objects, and food). Statistical significance was tested using two-sided one-sample t-tests (Bonferroni corrected for 12 comparisons). Stars represent significance at $p < .004$.

The role of surface properties in OTC food representations

The analysis reported in previous sections revealed that, while action-related properties mostly explain representations in the left lateral food-selective area, surface properties modulate both areas in a complex way. Here, to better isolate the contribution of each orthogonal dimension in eliciting these responses, we first ran representational similarity analysis (RSA; see

methods) to directly test the contribution of each surface properties in the representations underlying food selectivity, and second, we adopted more specific vector-of-ROIs analyses compared to the one performed above to clarify what property is driving more activity in an area.

First, to clarify how the two surface properties (colour, configuration) contribute to the representational structure observed in food-selective cortex, we conducted an RSA. We constructed 8×8 RDMs for each of the four food-selective ROIs identified with the *category* experiment by computing pairwise dissimilarities across the eight conditions of the *surface properties* experiment (Figure 3.6b). We then correlated each empirical RDM with two model RDMs reflecting the two orthogonal surface properties (colour and ensemble statistics) using semi-partial correlations to estimate the unique contribution of each dimension while controlling for the other (Figure 3.6c).

In line with the above results, this analysis revealed a clear dissociation between ventral and lateral food clusters. In ventral OTC, both left and right hemispheres showed the strongest unique association with the colour model (VOTC-left: $r = .5$, $p < .001$; VOTC-right: $r = .44$, $p = .002$), which explained the largest proportion of unique variance in these ROIs (34% and 32%, respectively). Ensemble configuration also contributed in ventral OTC, albeit to a lesser degree (VOTC-left: $r = .27$, $p = .03$, 23% variance; VOTC-right: $r = .27$, $p = .04$, 25% variance) though the results did not survive multiple comparison correction (both $p > .025$).

By contrast, lateral OTC showed a markedly different profile. Here, object configuration emerged as the dominant dimension, with both LOTC-left and LOTC-right demonstrating strong semi-partial correlations with ensemble (LOTC-left: $r = .3$, $p = .001$, 20% variance; LOTC-right: $r = .34$, $p < .001$, 19% variance). Colour accounted for only minimal variance in lateral OTC (LOTC-left: $r = 0$, $p = .9$, 4% variance; LOTC-right: $r = .04$, $p = .5$, 7% variance).

Second, we visualized responses to each dimension independently along the vector (e.g., contrasting all coloured vs. grayscale conditions across spheres). This analysis allows us to clarify the direction of the information represented: for example, while LOTC exhibits a high correlation with the ensemble model, indicating a representational difference between single and ensemble conditions in this area, this analysis cannot tell us the direction of the effect, as LOTC might respond higher to either single or group of objects. Spheres showing reliably greater activation for one level of a dimension were interpreted as biased toward that feature. Results can be visualized in Figure 3.6b. They reveal that food-responsive ventral clusters are sensitive to both colour and ensemble statistics properties of objects; viceversa, lateral food clusters respond exclusively to single object configuration and are not sensitive to colour information. This pattern aligns with prior demonstrations that anterior medioventral OTC is tuned to surface-based properties such as colour and texture/ensemble statistics (Cant & Xu, 2012; Cavina-Pratesi et al., 2010), while lateral OTC is more sensitive to shape-defined, single-object structure (e.g., Grill-Spector et al., 2001).

Together, these analyses demonstrate that ventral food-selective areas are primarily driven by surface-based dimensions such as colour and secondarily ensemble statistics, whereas lateral food areas prefer food in a single object configuration, consistent with the functional dissociation between ventral and lateral OTC reported in the previous functional analyses.

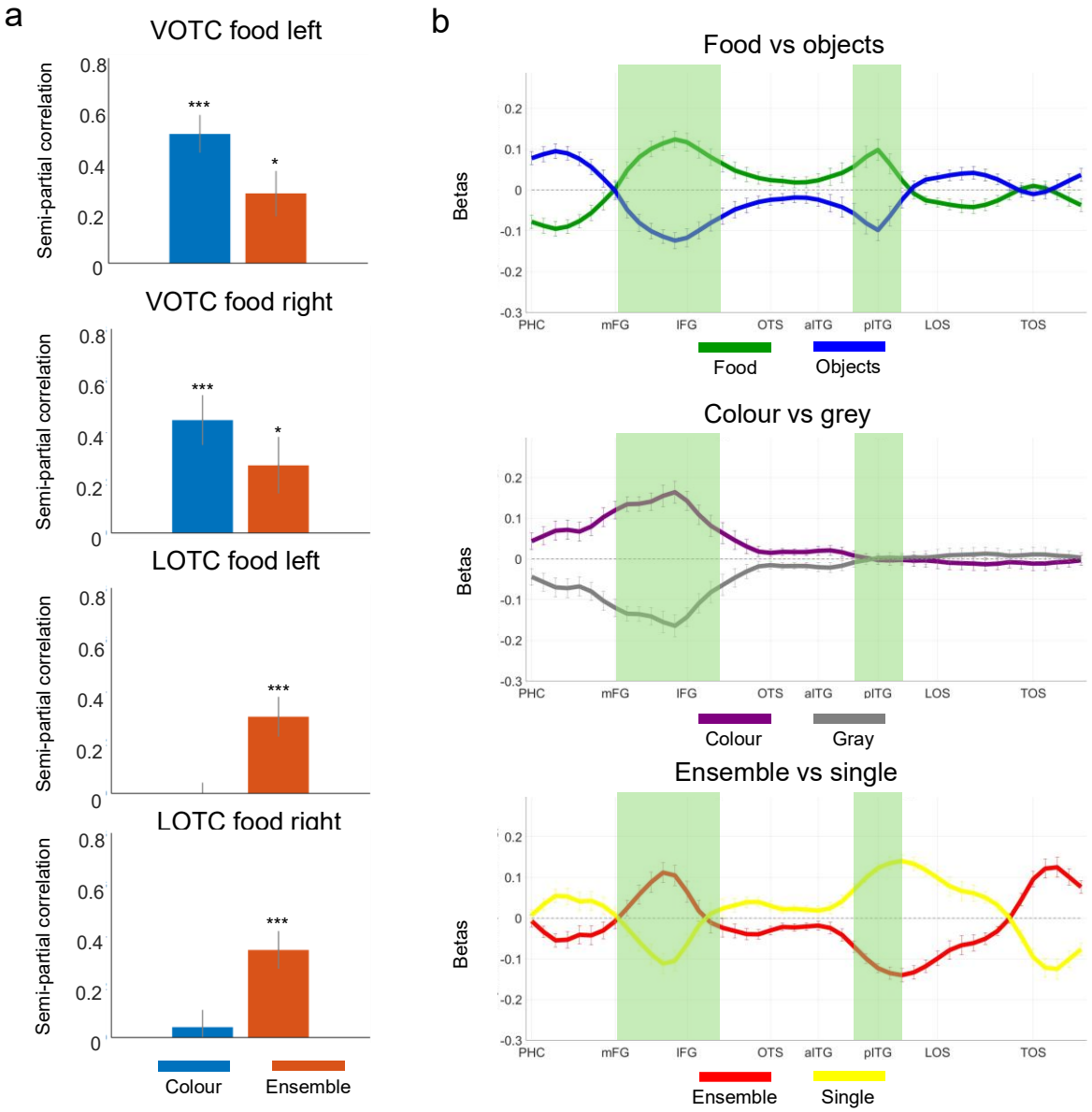


Figure 3.6. Surface Properties analyses of ventral and lateral OTC. a) *Vector-of-ROIs for the three orthogonal dimensions.* For each sphere along the vector, we contrasted the activation for one condition vs. the activation for the other (e.g., food vs objects). Error bars represent ± SEM across subjects (n = 18 participants). Green boxes underlie the spheres that are biased to food vs. inanimate objects. **b)** *RSA results.* Semi-partial correlation was run between the patterns of responses of the four food ROIs and two models representing the two surface properties, colour and ensemble statistics. Stars represent significance with one-sample two-sided t-tests, with *** $p < .001$, and * $p < .05$.

Food selectivity in TDANNs is organized into two clusters with distinct functional properties

Finally, we evaluated a model of the spatial organization of ventral visual cortex – the TDANN (Margalit et al., 2024) – to test whether biologically inspired pressures, namely wiring-length minimization combined with a self-supervised training objective, can account for the topographic organization of food selectivity and its associated functional properties. To this end, we performed univariate analyses analogous to those conducted in human visual cortex. Specifically, we presented both stimulus sets (from the *category* and the *surface properties* experiments) to five independent initializations of the TDANN and examined the spatial organization, functional selectivity, and patterns of overlap in the final "VTC-like" layer, meant to capture the organization of high-level ventral visual cortex.

Results for one initialisation can be visualized in Figure 3.7a-b. When we consider the spatial organization for the eight categories, we observe a tight clustering of the different categories at different locations of the simulated cortical space: units selective to both inanimate objects (tools, manipulable objects), shape, and food occupy very similar portion of the grid, whereas scenes form a tight cluster that partially overlap with the inanimate object cluster. Units selective for body-parts seem to be more scattered around, and fully separate from the scene cluster. When looking instead at the activations of the three orthogonal dimensions from the *surface properties* experiment, we see very few units exhibiting preference for either coloured or greyscale stimuli, whereas units preferring either ensemble vs. single objects or food vs. inanimate objects cluster together.

To better characterize the organization of food-selectivity and to quantify these observations, we identified contiguous clusters of food-responsive units using a DBSCAN algorithm (see methods). When these units were defined based on responses to the *category* experiment, their functional selectivity was assessed across all conditions of the *surface properties* experiment, and viceversa.

Results are shown in Figure 3.7c–f. Across all five model initializations, two spatially contiguous clusters of food-responsive units were consistently identified. When we identify units with the surface properties experiment (Figure 3.7c) and test the functional selectivity using the *category* experiment stimulus set (Figure 3.7d), both clusters revealed high selectivity to food, which was higher compared to all other categories ($p < .0001$, tested with 10000 permutations). Importantly, these two clusters exhibited distinct functional responses when localized with the *category* experiment (Figure 3.7e) and analysed using the *surface properties* experiment stimulus set (Figure 3.7f). Although both clusters responded in virtually the same way to coloured and greyscale stimuli, they differed markedly in their responses to food-related conditions. Specifically, one cluster responded strongly to food stimuli presented in a single-object configuration, but not to food ensembles, and showed a stronger response to single inanimate objects compared to ensembles of those objects. Conversely, the other cluster responded robustly to food stimuli in both single and ensemble configurations, and additionally showed a stronger response to ensemble inanimate objects compared to single inanimate objects (all $p < .0001$, tested with 10000 permutations). This pattern mirrors the dissociation observed between ventral and lateral food-selective clusters in human visual cortex with respect to their differential sensitivity to ensemble statistics, and suggests that TDANN may be able to capture mid-level visual features that contribute to shape the organization of OTC.

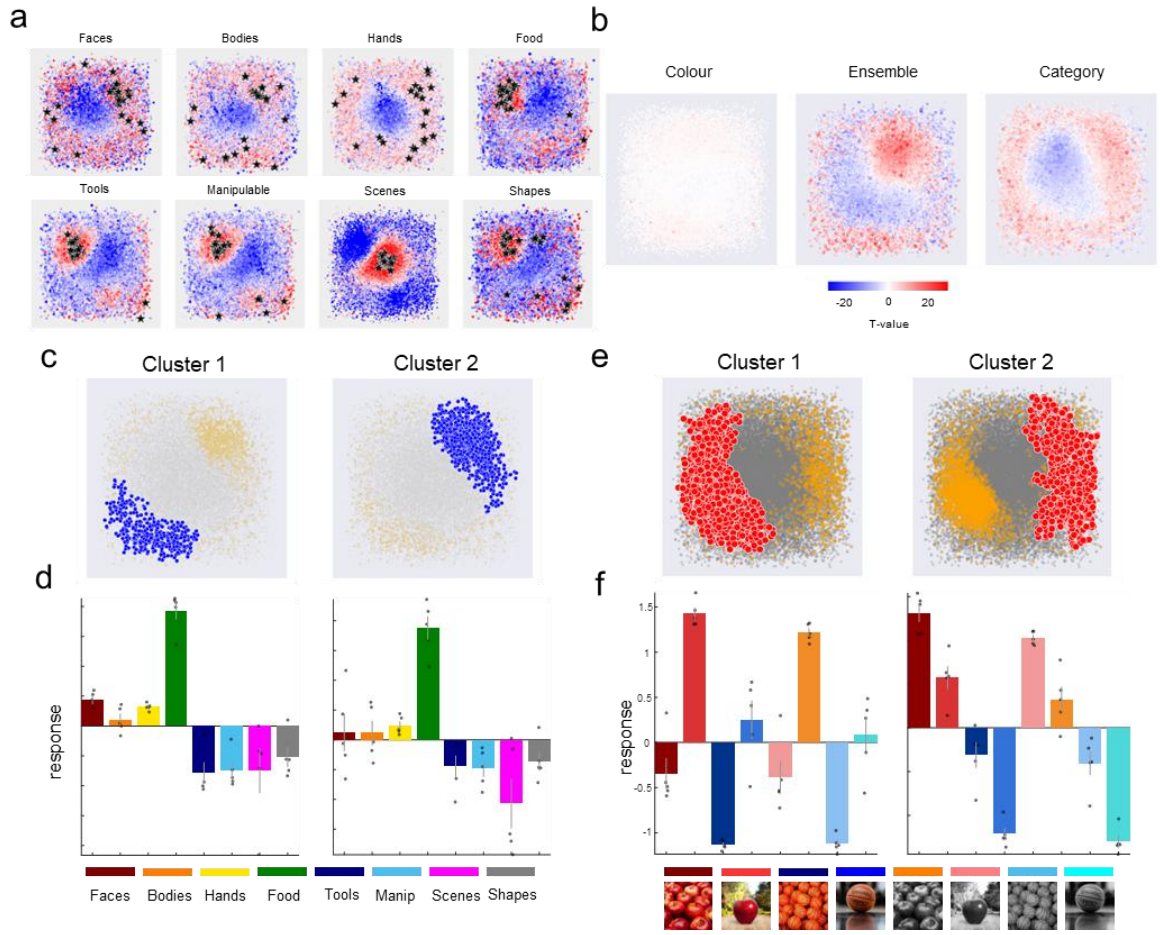


Figure 3.7. Spatial organization of food selectivity in TDANNs. a) Spatial distribution of category selectivity (defined by t-values from the Category Experiment) across the simulated cortical surface of the VTC-like layer for one representative TDANN initialization. Stars indicate the locations of the top 25 most selective units for each category. **b)** Spatial distribution of selectivity for the three orthogonal dimensions (defined by t-values from the Surface Properties Experiment) in the same TDANN initialization shown in (a). For both (a) and (b), category-selective units (positive t-values) are shown in red, whereas non-selective units (negative t-values) are shown in blue. **c)** Spatial distribution of food-selective units in the VTC-like layer for the same TDANN initialization. Orange dots indicate all food-selective units; red dots indicate units assigned to a cluster by the DBSCAN algorithm. **d)** Selectivity profiles of food-selective units in each identified cluster, based on activations in the VTC-like layer. Responses for each condition are normalized relative to the average response to all other conditions, as in the analyses of human visual cortex. Each data point corresponds to one TDANN model initialization (n = 5). Error bars indicate ±1 SEM across model initializations.

Discussion

In the present work, we investigated the object features and representational dimensions that underlie food selectivity in ventral and lateral occipitotemporal cortex (OTC) and found dissociations among the representations supported by the ventral and lateral food areas. We found that food selectivity in left lateral OTC aligns with an action-related gradient: food responses partially overlap and are correlated with categories sharing graspable features (e.g., tools, manipulable objects) but not with hands, which differ in effector-related properties; in contrast, ventral food areas were positioned between face- and scene-selective areas, responded selectively to food, and did not contain action-related representations. Moreover, when we examined the visual features that may be important for driving the emergence of food-selective areas, we found that surface-level properties – especially colour and, to a lesser extent, ensemble statistics – play an important role in shaping the location and multivariate representations of ventral food selective clusters, while little-to-no colour information was present in the lateral food areas. Together, these findings advance our understanding of the feature dimensions that support food representations in OTC and highlight properties that distinguish between the ventral and lateral pathways.

The topography of ventral and lateral food-selective areas

Initial evidence for food-selective areas in human visual cortex came from analyses of the Natural Scenes Dataset (NSD; Allen et al., 2022), revealing two food-selective clusters: one in ventral OTC near the medial fusiform gyrus and one in lateral OTC around the inferior temporal gyrus. These studies argued against explaining food selectivity purely through low- or mid-level visual features such as colour, curvature, or texture (Jain et al., 2023; Khosla et al., 2022; Pennock et al., 2023; Henderson et al., 2025). However, the spatial proximity of these food-selective areas to known colour-, texture-, and ensemble-biased areas in ventromedial OTC (Cant & Xu, 2012; Lafer-Sousa et al., 2016), together with reports of overlap between food and tool responses (Ritchie et al., 2024a), suggests that food selectivity may be interpretable within broader spatial organizational maps of visual cortex.

By using controlled stimulus sets, our study directly examined how the locations of food-related responses relate to both action-related and visual feature properties. We replicated the presence of two food-selective clusters in OTC, one medial and one lateral compared to face-selective areas, and demonstrated that their positions reflect different organizational principles: lateral food selectivity is best understood as part of an action-related gradient (Cortinovis et al., 2025a), whereas ventral food selectivity relates more closely to regions sensitive to surface-level properties, particularly colour and ensemble statistics. Below, we consider these two properties in turn.

Action properties explain the position of food-selective areas in left lateral OTC

Previous work reported overlap between food and tool representations in both ventral and lateral visual cortex and suggested that shared action-related properties such as graspability contribute to this organization (Ritchie et al., 2024a). Our results confirm that in left lateral OTC, food responses overlap with tools and manipulable objects due to shared graspable features; however, we did not find evidence for overlap between food and hands. This pattern aligns with our proposed action-related gradient (Cortinovis et al., 2025a): food and tools overlap because both are graspable, while hands and tools overlap because both serve as effectors. In contrast, hands (which are not graspable) and food (that are not effectors), lack shared action properties and therefore do not overlap.

Whether similar action-related features contribute to food selectivity in ventral OTC remains unclear. Ventral cortex contains little-to-no tool or manipulable-object selectivity, seemingly at odds with earlier claims that the medial fusiform gyrus reflects action-related properties of tools (Mahon et al., 2007). Rather, ventral areas may preferentially process surface-level object features that relate only indirectly to action and that are common among many inanimate object categories (Cortinovis et al., 2025b; Mahon & Almeida, 2024), including, possibly, food. More generally, this framework is consistent with proposals of a division of labour between ventral (involved in object processing) and lateral (supporting action processing) OTC (see

Lingnau & Downing, 2015; Weiner & Grill-Spector, 2013, and Wurm & Caramazza, 2022 for reviews). Future work could examine more precisely which graspability- or manipulability-related dimensions organize food, tools, and other manipulable objects within lateral OTC, extending approaches used to characterize action properties of manipulable objects specifically (Almeida et al., 2023; 2025).

Surface properties (colour and ensemble statistics) explain the position of food-selective areas in ventral OTC

The surface properties experiment demonstrated a strong role for colour in eliciting responses within ventral food-selective areas. Colour is especially diagnostic of food: behavioural and neural studies indicate that it is a behaviourally-relevant feature supporting rapid food detection and recognition more than it does for many other object categories (Conway, 2018; Sato, 2021). Neuroimaging research across humans and macaques consistently identifies colour-biased regions in medial OTC, whereas lateral OTC shows much weaker colour sensitivity (Brouwer & Heeger, 2009; Lafer-Sousa et al., 2013, 2016; Rosenthal et al., 2018; see Mullen, 2019 for a review). Analyses of the NSD similarly identified correspondence between colour-biased patches and food selectivity (Pennock et al., 2023).

However, colour is not the only mid-level feature represented in the medial fusiform gyrus\collateral sulcus; rather, this region likely plays a role in supporting texture and ensemble statistics processing. Much less is known about high-level visual representations of textures, even if some studies found that similar colour-responsive regions in medio-ventral OTC also respond to textures (Cant & Goodale, 2007; 2011; Cavina-Pratesi et al., 2010). Even fewer studies investigated ensemble processing. One study found that regions neighbouring and partially overlapping with the parahippocampal place area support ensemble processing (Cant & Xu, 2012); importantly, this and subsequent studies showed that the responses of this region could not be reduced to lower-level texture statistics (nor colour), but rather possibly rely on higher-level texture representations (Cant & Xu, 2015; 2017; see also Henderson et

al., 2023 for a hierarchical view of texture processing). Again, a ventral-lateral division of OTC is in place: ventral visual cortex supports texture-like representation of objects whereas more lateral regions seem to respond more to the shape of the objects. Our findings support this view by showing that colour, ensemble configuration, and category each explains unique variance in ventral food-selective responses, with colour contributing the most.

Other features

Of course, we do not claim that the (left) lateral food-selective cluster can be fully explained by its action-related properties, nor that colour and ensemble statistics are the only features processed in medio-ventral OTC. Important candidate features not directly addressed here include shape properties such as curvature and aspect ratio, which have been proposed as important organizing principles for both ventral and lateral OTC (Bao et al., 2020; Yue et al., 2020), even if previous studies already showed that curvature alone cannot account for food selectivity (Khosla et al., 2022).

Similarly, our results cannot distinguish ensemble-specific from high-level texture processing, which likely share overlapping mechanisms (Cant & Xu, 2017). While ensemble responses in ventral visual cortex cannot be reduced to low-level statistics (e.g., spatial frequency), high-level texture representations likely contribute substantially (Cant & Xu, 2012; 2017; Henderson et al., 2023). Furthermore, medial food-selective areas lie within regions broadly implicated in material perception, extending beyond colour, texture, and ensemble structure (Paulun et al., 2025; Xiao & Liao, 2025). Thus, food-selective responses likely reflect a constellation of mid- and high-level properties which combination is characteristic of food objects.

Topographic Artificial Neural Networks partially replicate food dissociations

Analyses on the TDANN, a model that is meant to capture visual cortex spatial organization aside from its function, reveals both convergence and divergence with visual cortex. The models exhibited two clusters of food-responsive units that could be dissociated based on

their spatial and functional properties. Although these clusters could not be distinguished based on colour, they diverged in their sensitivity to ensemble statistics. Specifically, mirroring patterns observed in visual cortex, in TDANNs one cluster overlapped more strongly with inanimate-object–responsive units and showed a stronger response to food images presented in isolation, whereas the other cluster responded robustly to food images across both configurations.

These results are consistent with our previous work (Cortinovis et al., 2025a) which showed that TDANNs were able to replicate the organization by shape and animacy of ventral OTC but not the action-related organization of lateral OTC. Taken together, the present and prior results suggest that TDANNs (and ANNs more generally) might be a better model of mid-level (visual) features rather than higher-level behaviorally-relevant features (Mahner et al., 2025), including action-related properties. Even colour, typically considered a mid-level visual feature, supports behaviorally-relevant goals in humans (as it indicates potentiality for consumption; Sato, 2021) but likely not in TDANNs, explaining why no difference was found for colour vs. greyscale stimuli. More broadly, they indicate that the core assumption of the TDANN framework – namely, that jointly optimizing a self-supervised contrastive objective alongside a spatial loss is sufficient to reproduce any type of spatial organization (Finzi et al., 2023; Margalit et al., 2024) – is likely incomplete. Capturing the full structure of these pathways may require more biologically plausible training objectives or datasets, such as those approximating the visual experience of children actively exploring their environment (Long et al., 2024).

Conclusion

In summary, we demonstrate that distinct features drive food selectivity across OTC: lateral selectivity reflects action-related properties such as graspability, whereas ventral selectivity is strongly influenced by surface properties including colour and ensemble statistics. Our findings support a view of visual cortex organization in which category-selective areas reflect

underlying behaviorally-relevant feature dimensions rather than purely categorical representations (Bracci & Op de Beeck, 2023; Peelen & Downing, 2017; Ritchie et al., 2025). In this framework, food-selective areas are not “food detectors” per se but areas tuned to visual and semantic properties reliably associated with food and relevant for interacting with it (Jain et al., 2023; Ritchie et al., 2024a; 2025), properties that depend on the position of each area along a distinct pathway and that are associated with the computational goal each pathway supports.

Chapter 4 - Encoding models reveal fine-grained feature selectivity for bodies, hands, and tools in occipitotemporal cortex

Abstract

Category-selective areas in occipitotemporal cortex (OTC) are typically described in broad terms, such as selectivity for faces, bodies, or scenes. Yet mounting evidence suggests a finer-grained functional organization, with separable responses to specific body parts or objects, supporting distinct feature spaces in line with their differential computational roles. Here, we combined image-level fMRI analyses with artificial neural network (ANN)–based encoding models to test the selectivity and underlying feature sensitivity of category selective areas in ventral and lateral OTC. Using densely sampled fMRI data from three participants across six sessions, we first demonstrated robust dissociations between body, hand, and tool responses at the level of individual images. We then trained area-specific encoding models based on deep neural network features and used them to predict responses to millions of novel images. Results demonstrated reliable dissociation of body-, hand-, and tool-selective areas in both ventral and lateral OTC, and the encoding models trained in each of these areas exhibited robust prediction accuracy and a clear preference for the expected category. Importantly, comparisons between models revealed differential feature sensitivity for areas selective for the same category, consistent with each area's position along ventral vs. lateral OTC and hemispheres. Together, these findings provide evidence for fine-grained specialization within OTC and illustrate how ANN-based encoding models can serve as computational tools to uncover the feature-level basis of category selectivity.

Introduction

A well-established property of ventral temporal cortex is the presence of category-selective areas that respond preferentially to specific object categories, such as faces, body parts, or scenes (Kanwisher, 2010). Recently, artificial neural networks (ANNs) have been adopted to model category selectivity in human visual cortex (Kanwisher et al., 2023). In particular, ANNs have been combined with encoding approaches (Naselaris et al., 2011; Khosla et al., 2022) to generate “virtual” models of category-selective areas and to predict their responses to novel images (Agrawal et al., 2014; Eickenberg et al., 2016; Gu et al., 2022, 2023; Wen et al., 2018). For example, Murty et al. (2021) trained encoding models using ANN-derived features and fMRI data. Participants first underwent standard localizer scans to identify face-, body-, and scene-selective areas, and were then shown a diverse set of images across multiple sessions. Neural responses were used to train separate encoding models for each area. When tested on millions of novel images, these models showed robust category selectivity, validating their functional tuning.

Can encoding models based on ANN features also capture finer-grained functional distinctions observed in the brain? And, most importantly, can they be exploited to test feature differences among areas, even those selective to the same category?

Human neuroimaging provides strong evidence for finer-grained selectivity in both ventral and lateral occipitotemporal cortex (VOTC and LOTC, respectively; Taylor & Downing, 2011). VOTC contains areas selective not only for faces and scenes, but also for whole bodies, hands (Peelen & Downing, 2005; Schwarzlose et al., 2005; Pillet et al., 2024a), and inanimate objects (Mahon et al., 2007). LOTC, by contrast, exhibits differential responses to whole bodies, hands, and tools (Bracci et al., 2010, 2012; Bracci & Peelen, 2013; Peelen et al., 2013), distinctions also observed with high-field fMRI (Pillet et al., 2024b) and intracranial recordings (Ramirez et al., 2024). These ventral and lateral distinctions have been proposed to reflect partially dissociable computational roles, with ventral regions supporting object recognition and lateral regions supporting action-related properties (Lingnau & Downing, 2015; Wurm &

Caramazza, 2022). Category-selective areas also exhibit hemispheric asymmetries, with stronger right-hemisphere responses for bodies and left-hemisphere responses for hands (Pillet et al., 2024a; Bracci et al., 2010). With higher spatial resolution or minimal smoothing, even classic category-selective areas can be further parcellated into multiple clusters, each responsive to the same broad category (Weiner & Grill-Spector, 2010, 2012, 2013; Çukur et al., 2013). For example, multiple limb-selective areas have been identified bilaterally (Weiner & Grill-Spector, 2011), arranged along a body-part map (Orlov et al., 2010) or according to action-related principles (Cortinovis et al., 2025a; Bracci et al. 2015). Converging evidence further shows that the representational content of category-selective areas is itself finer-grained, encoding features beyond canonical category boundaries (Contier et al., 2024; Çukur et al., 2013; van Dyck et al., 2025; Vincken et al., 2023).

Why does the brain instantiate multiple representations of the same category? One possibility is that these areas differ in the specific features they encode, reflecting distinct computational role. For instance, we recently proposed that ventral and lateral tool-sensitive areas respond to different tool-related properties, with the lateral area more sensitive to shape and action-related properties and the ventral area to surface properties (Cortinovis et al., 2025b).

Considering this evidence, our objectives in the present study were twofold. First, we leverage image-level functional analysis and ANN-based encoding models to provide a stringent test of selectivity for closely overlapping areas. Second, we use the encoding models to explore differential feature representations for areas exhibiting the same category selectivity.

Using densely sampled fMRI data from three participants scanned across six sessions, we localized areas selective for whole bodies, hands, and tools using a functional localizer. Participants then viewed 200 images depicting a range of body parts and inanimate objects, allowing assessment of functional selectivity at the image level (Mur et al., 2012). Then, we trained encoding models on these neural responses following the approach of Murty et al. (2021) and tested them on a large set of novel images. To interpret model predictions, we

applied occlusion-based saliency mapping to identify image regions that most strongly contributed to predicted neural responses. Finally, beyond validating fine-grained category selectivity for closely overlapping voxels, we used the encoding models to probe whether areas selective for the same category differ systematically in their underlying feature sensitivity. We predict that such differences would depend on the position of each area along ventral vs. lateral OTC or a specific hemisphere, depending on the computational goal each of these support.

Our results demonstrate that image-level functional selectivity and encoding models can dissociate body-, hand-, and tool-selective areas in ventral and lateral OTC and reveal that areas labelled as selective for the same category can substantially differ in the features they encode, consistent with the computational role of the region (ventral vs. lateral) or hemisphere in which they are embedded.

Methods

Participants

Three right-handed participants (2 females, age range 24-30 years) with normal or corrected-to-normal vision and no history of neurological disorders took part in the study. Participants provided informed consent, and the experimental procedures were approved by the Ethics Committee of the University of Trento.

Stimuli

Two distinct sets of stimuli were used for the main event-related experiment and the localizer. The main experimental stimulus set consisted of 200 coloured images with natural backgrounds, primarily sourced from the THINGS dataset (Hebart et al., 2019) and internet searches. The images belonged to 5 categories: 25 whole bodies, 25 hands, 50 tools, 50 manipulable objects, and 50 non-manipulable objects. Tools were defined as objects that are used with hands to physically interact with other objects or surfaces (e.g., scissors, knives); manipulable objects were graspable but are usually the passive receiver of the action (e.g., books, cups); non-manipulable objects are large non-graspable objects (e.g., vehicles, buildings).

The localizer stimulus set consisted of 72 greyscale images divided into 6 categories: whole bodies, hands, tools, manipulable objects, non-manipulable objects, and chairs. Each category included 12 images presented on a white background, adapted from a set used by Matić et al. (2020). Inanimate object categories were controlled for visual differences like shape and orientation. All images were resized to 400x400 pixels and subtended 8° of visual angle.

Experimental Design and Procedure

Each participant completed six scanning sessions over two weeks, totalling approximately nine hours of fMRI data per participant. Data for both experiments were acquired in each

session. Visual stimuli were presented using the Psychophysics Toolbox (Brainard, 1999) in MATLAB (2021b, The Mathworks), projected onto a screen and viewed through a mirror mounted on the head coil.

Main Experiment: The main experiment used a rapid event-related design. In each run, lasting 5 min and 24 sec, 100 images drawn from the full stimulus set were presented for 500 ms with a 3500 ms inter-stimulus interval (ISI). Blank trials were included approximately 20% of the time. Participants performed a catch trial detection task, pressing a button whenever an image of a bug or plant appeared (~10% of the time). Around 60 runs per participant were collected, totalling on average 15 repetitions per image.

Localizer: The localizer followed a block design. In each of 6 runs, lasting 5 min 36 sec, blocks of images from a single category were presented. Within a block, images were shown for 400 ms with a 266 ms ISI. 5 block repetitions per category were included. Participants performed a one-back task, pressing a button when an image was repeated twice in a row (one repetition per block).

MRI Acquisition and preprocessing

All imaging data were collected at the Center for Mind/Brain Sciences (CIMeC), University of Trento, using a 3T Siemens Prisma scanner with a 64-channel head coil. Functional volumes were acquired for both the localizer and main experiments using a T2*-weighted echo-planar imaging (EPI) sequence with a multiband acceleration factor of 3 (parameters: TR = 2,000 ms; TE = 28 ms; flip angle = 75°; 69 axial slices covering the whole brain; FoV = 220 mm, voxel size of 2 x 2 x 2 mm). A high-resolution T1-weighted anatomical image was also acquired in the first scanning session for each participant using an MPRAGE sequence with a voxel size of 1 x 1 x 1 mm.

Preprocessing was performed using the Statistical Parametric Mapping software (SPM12; Wellcome Trust Centre for Neuroimaging) in MATLAB (R2021b). The pipeline included slice-timing correction, head motion correction via spatial realignment to the first volume of each run, and coregistration of the functional data to the participant's anatomical scan. To maximize spatial precision and avoid possible overlap between neighbouring category-selective voxels, the functional images were analyzed in the native participant space. A FWHM Gaussian kernel of 3 mm was applied to the localizer data to improve signal-to-noise ratio and increase localization's robustness. Runs with head motion exceeding predefined thresholds (2 mm in translation or 1 mm in rotation) were excluded from the analysis.

A general linear model (GLM) was then fitted to the preprocessed data. For the main experiment, each stimulus presentation was modelled as a unique event. For the localizer, the six object categories were modelled as conditions. In both cases, the six motion-correction parameters were included as nuisance regressors. Predictors were generated by convolving a boxcar function with SPM's canonical hemodynamic response function.

Region of Interest (ROI) Selection

Category-selective regions of interest (ROIs) were defined on the native cortical surface of each participant using data from the localizer experiment. Analyses were focused on the ventral and lateral occipitotemporal cortex (VOTC and LOTC respectively). Functional ROIs were identified using individualized contrasts for category, corrected for multiple comparisons at a cluster-level threshold (*FDR* < .05). Areas selective for bodies were identified with a contrast of whole-bodies vs. hands and tools; areas selective for hands were identified with a contrast of hands vs. whole-bodies and tools; areas selective for tools were identified with a contrast of tools vs. all other categories. The contrasts were specifically chosen to functionally dissociate these highly related and often overlapping representations, which – in the case of bodies and hands – are often grouped within a single "extrastriate body area" (EBA) or "fusiform body area" (FBA). Indeed, voxels responding to hands can also respond to tools or

to bodies (and viceversa); here, we selected non-overlapping voxels exclusively responding to one of the three categories.

Functional selectivity analysis

As a first step, we evaluated the selectivity of the identified ROIs. Our design allowed us to test the functional responses of category-selective areas at the image level with a relatively high number of images. We quantified these category preferences using a *d'* index (Mur et al., 2012), defined as:

$$d' = \frac{\mu_{category} - \mu_{others}}{\sqrt{\frac{1}{2}\sigma^2_{category}\sigma^2_{others}}}$$

Where $\mu_{category}$ and μ_{others} are the mean responses to two different categories, and $\sigma^2_{category}$ and σ^2_{others} are their variances. The significance of each category *d'* (vs. the *d'* of each of the other category) was assessed using a permutation test with 10,000 permutations, and all reported results were significant at $p < .0125$ (Bonferroni corrected with $N = 4$ comparisons, each category *d'* vs the other categories *d'*), unless we report no effect. In addition, to provide a complementary assessment of category discriminability, we conducted a Receiver Operating Characteristic (ROC) analysis (Mur et al., 2012). For each category, we performed a one-vs-all comparison, treating the stimuli for one category as the "positive" class and all stimuli from the remaining four categories as the "negative" class. The activation values for all 200 stimuli were used as the ranking variable to plot the true positive rate against the false positive rate across all possible thresholds. The Area Under the Curve (AUC) was then calculated as the primary metric of classification performance. An AUC value of 1.0 indicates perfect classification, where all stimuli from the target category elicit a higher response than any stimulus from the other categories, while an AUC of 0.5 represents performance at chance level. Finally, functional selectivity was further tested with a Top-N rank analysis, which indicates the proportion of stimuli from each category among the top 25 responses (e.g., how many hands there are in the 25 stimuli eliciting the highest activation).

Encoding models

General description

We adopted an analysis framework similar to that of Murty et al. (2021). The encoding model operates by extracting visual features from images using a deep convolutional neural network and learning a mapping from these features to brain activity. This allowed us to predict neural responses to novel images based on their visual content. Then, screening the trained encoding models on large-scale image datasets allowed us to perform a stringent test of category selectivity, mitigating potential biases arising from limited stimulus sampling in fMRI experiments.

Specifically, we first extracted features for the 200 images using a ResNet-50 architecture pre-trained on ImageNet. The network was chosen based on results from previous studies employing similar encoding modelling procedures (Gu et al., 2022, Murty et al., 2021). We evaluated each ResNet-50 layer and selected the one yielding the highest correspondence between predicted and actual fMRI responses. The best-performing layer was the final global average pooling layer, which outputs a 2048-dimensional feature vector per image. These vectors served as predictors, while the target variable was the average fMRI response to each image across the three participants. The relative low number of images (compared to large scale image datasets, for example the Natural Scenes Dataset, Allen et al., 2022) - and hence the possible limits in generalization that can be obtained with this dataset - are due to a trade-off between the number of stimuli and the number of repetitions that can feasibly be presented and the quality that can be achieved at an image-level with fMRI.

We used ridge regression to learn a linear mapping from features to fMRI responses and employed 10-fold cross-validation for model evaluation. The dataset was split into 10 disjoint subsets; in each fold, 9 subsets were used for training and 1 for testing. This procedure was repeated so each subset served once as the test set. Within each fold, the optimal ridge regularization parameter (λ) was selected through a nested 10-fold cross-validation on the

training data to avoid overfitting. Model performance was quantified as the Pearson correlation coefficient (*r*) between predicted and actual responses in each test fold. Final prediction accuracy was calculated as the average correlation across all folds. We trained a separate encoding model for each ROI.

Reliability of fMRI Response Patterns

To evaluate the internal consistency of the stimulus-evoked fMRI response patterns within each ROI, we conducted a split-half reliability analysis (Lage-Castellanos et al., 2019; Murty et al., 2021). For this procedure, the full set of trials for each of the 200 stimuli was randomly partitioned into two equal halves. The responses for each stimulus were then averaged across the trials within their respective half and across all voxels in the ROI. This process was performed for each participant, and the resulting response patterns were then averaged at the group level, yielding two independent group-average response vectors (one for each split). The Spearman rank correlation was calculated between these two vectors. To estimate the reliability of the full dataset, this correlation coefficient was corrected using the Spearman-Brown (SB) prophecy formula. Because this score represents the maximum explainable variance rather than the maximum achievable correlation, we report the noise ceiling in correlation units as $r = \sqrt{r_{SB}}$ (where r_{SB} is the Spearman-Brown score) when comparing against model–data correlations (van Bree et al., 2025). To ensure a stable estimate, this entire process (from random splitting to the corrected correlation) was iterated 100 times, and the final reliability score was taken as the mean of these iterations. To calculate the noise ceiling for the averaged group data, we repeated the same procedure (split-half analysis and Spearman-Brown correction) on the betas averaged across participants. With the entire number of repetitions per stimulus, we obtained reliable fMRI responses (see results). This score functions as a noise ceiling against which we can compare the model performance (Lage-Castellanos et al., 2019).

Image Screening and Datasets

To assess the feature preferences and category selectivity of the encoding models, we screened them against ImageNet (Deng et al., 2009; *n* = 1,281,149), a large-scale image dataset containing diverse object categories traditionally used in machine learning to train models, and ecoset (Mehrer et al., 2021; *n* = 1,444,892), an ecologically motivated dataset designed to more closely mirror human visual diet. This combination ensured a broad coverage of categories, including images of people, objects (and specifically, tools and manipulable objects), and their interactions, which are often underrepresented in well-known vision neuroscience datasets such as the Natural Scenes Dataset (Allen et al., 2022; see Shirakawa et al., 2025).

Using the trained models, we predicted BOLD responses for all images in each dataset. Images were then ranked by predicted activation for each ROI, allowing us to identify the categories and features most strongly associated with neural responses. For a first qualitative assessment of category tuning, we first visualize the top activating images ranked by predicted activation for each model; these visualizations enable direct inspection of the visual features and categories that maximally activated each region. However, since the model tends to respond similarly to images that differ only slightly (e.g., multiple near-identical screwdrivers), the top-ranked images can be dominated by a single object or feature. To obtain a more comprehensive picture of the object categories and features underlying the encoding models responses, we therefore applied an unsupervised clustering analysis to the top 0.1% of images (≈2,500) ranked by predicted activation. Specifically, visual features were extracted from these images using the penultimate layer of a ResNet-50 model pretrained on ImageNet. K-means clustering was performed with a cosine distance metric. We visualized the most activating images contained in each of four most selective clusters and generated inferences based on their common visual features. The aim of this analysis was to test if any object category or specific features not related to the hypothesised preferred category elicit strong and consistent activations in the encoding models.

Interpreting Encoding Model Predictions with Occlusion-Based Saliency Mapping

To understand which parts of the images were driving the model's predictions, we used an occlusion-based saliency mapping technique inspired by Randomized Input Sampling for Explanation (RISE; Petsiuk et al., 2018). This method identifies pixel regions within the image that most contribute to the predicted activation. For this analysis, we selected 25 images from the COCO dataset (Lin et al., 2015). All images depicted scenes that included body-parts (such as visible trunks, arms, legs, etc.), hands, and tools (i.e., hairdryers, brushes, scissors, cutlery). The images were randomly sampled from the COCO dataset by selecting relevant labels indicating categories of interest (e.g., "person", "scissors"), resized and squared to 400x400 pixels. We selected only images in which body-parts, hands, and tools were simultaneously present and sufficiently large to be reliably captured by the masking procedure.

For each image, 2,000 random binary masks (8×8 resolution) were generated and multiplied with the input image to create partially occluded versions. These were passed through the full encoding pipeline (ResNet-50 + ridge regression), and each mask was weighted by the predicted BOLD signal. The final importance map was computed as the weighted sum of these masks, highlighting image regions that most strongly influenced the model's prediction. This process was applied separately to each encoding model. This analysis was used to gain insight into the image features driving model predictions and to compare feature sensitivity across areas selective for the same category.

Comparing Category-Selective Encoding Models

Beyond evaluating models in isolation, we also examined pairwise differences between encoding models to uncover feature-level distinctions among brain areas selective for the same category, focusing on contrasts between ventral and lateral pathways and between hemispheres. For each model pair, we identified images that elicited the largest differences in predicted activation, i.e., stimuli predicted to strongly activate one model but not the other. Such differential predictions provide insights into the unique feature sensitivities of each area

and offer evidence for potential functional subdivisions underlying feature specificity (and, thus, their computational goals) within category-selective areas. The analysis was conducted on the same large-scale stimulus set used for image screening (ImageNet and ecoset combined, N = ≈ 2,500,000 after removing images common between the two datasets). We focused in particular on contrasts between areas selective for the same category but located either in opposite hemispheres or along different processing pathways. For example, we compared predictions between LOTC-hand left and LOTC-hand right (hemispheric comparison), as well as between LOTC-hand left and VOTC-hand left (pathway comparison).

All predictions were z-score normalized prior to comparison. We adopted the same visualisations and clustering approach as described above: we first plotted the top-activating images with the largest positive differences for each pair of areas, highlighting stimuli that preferentially activated one model over the other; then, we applied the unsupervised clustering analysis to the top 0.1% of images, and plot the images contained in each of four cluster, ranked by their average activations. All analyses were conducted using custom MATLAB scripts.

Results

The current work had two main objectives. First, we characterize functional selectivity at the level of individual images in body-, hand-, and tool-selective areas of ventral and lateral OTC. Second, we use ANN-based encoding models to provide stringent tests of category selectivity and to uncover feature-level differences between areas selective for the same category, with a focus on dissociations related to ventral and lateral OTC and to the two hemispheres.

We localized areas selective for whole-bodies, hands, and tools in three participants and recorded event-related responses in those areas to a stimulus set containing 200 images of body-parts (whole-bodies and hands) and inanimate objects (tools, manipulable, and non-manipulable; see Figure 4.1a). First, we tested the functional selectivity of category-selective areas; then, we trained encoding models based on the activations of those areas and evaluated the functional responses of those models.

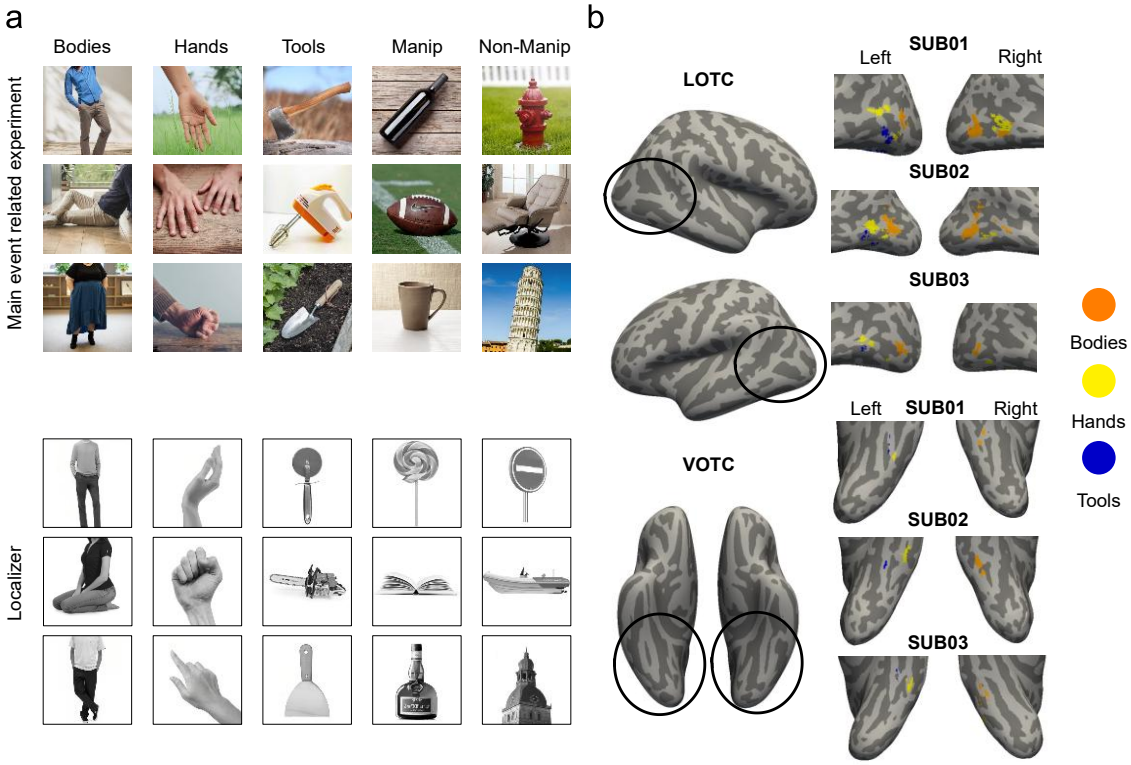


Figure 4.1. Stimulus sets and ROIs. a) Stimuli used in the main experiment (above) and in the localizer (below). Both sets included 5 categories: whole-bodies, hands, tools, manipulable, and non-manipulable objects. **b)** Single subject category-selective activation maps. Activations for body, hand, and tool selective areas in bilateral ventral and lateral occipitotemporal cortex ($p < .001$ uncorrected at the voxel level, FDR < .05 at the cluster level). ROIs were selected in volume space and projected for visualisation on the Freesurfer-reconstructed native surface of each participant (Fischl, 2012).

Ventral and lateral OTC areas selective for whole-bodies, hands, and tools

In each participant, we identified areas selective to images of whole-bodies, hands, and tools in ventral and lateral OTC (*FDR* cluster-corrected at $p < .05$). We only selected those areas that could be reliably localized in all three participants (Figure 4.1b). Whole-bodies activated bilateral regions in LOTC surrounding the Lateral Occipital Sulcus, further extending more anteriorly and superiorly in the right hemisphere; in VOTC, a body-selective area could be reliably identified only in the right hemisphere, around the occipitotemporal sulcus. In LOTC, hands exhibited stronger activations in the left hemisphere, spanning the posterior inferior temporal gyrus and extending more anteriorly and superiorly towards the middle temporal gyrus; albeit smaller, a hand-selective area could also be identified in the right hemisphere; in both cases, hand selectivity was anterior to body selectivity; in VOTC, hand selectivity was identified exclusively in the left hemisphere, around the occipitotemporal sulcus, in a similar position as the right-lateralized ventral body activation. In LOTC, tools elicited a strongly left-lateralized activation in the inferior temporal gyrus, neighbouring but anterior and inferior to the hand cluster; a smaller cluster of responses was observed in VOTC, around the medial fusiform gyrus. Notice that the ROIs were defined with contrasts that were specifically chosen to exclude voxels that respond to more than one category (hands and tools or hands and bodies).

As a first step, we assessed the internal consistency of the fMRI data in the identified ROIs using a split-half procedure using the main event-related experiment data. We found that with

the full set of stimulus repetitions, the response patterns demonstrated consistent reliability (LOTC-body right: Spearman-Brown corrected $r = 0.54$; LOTC-body left: $r = 0.53$; VOTC-body right: $r = 0.48$; LOTC-hand right: $r = 0.58$; LOTC-hand left: $r = 0.58$; VOTC-hand left: $r = 0.3$; LOTC-tool left: $r = 0.48$; VOTC-tool left: $r = 0.38$).

Next, we characterized functional selectivity at the level of individual images, allowing us to assess fine-grained distinctions between closely related categories (e.g., bodies vs. hands), moving beyond averages over all members of a category (e.g., moving beyond comparing the average activations of all hands vs. all bodies). The functional selectivity of each ROI is visualized in Figure 4.2. We quantified selectivity using three primary measures: a *d'* index, a Receiver Operating Characteristic (ROC) analysis, and a top-N analysis. The *d'* index provides a preference score for a category: values around 0 indicates no preference, values above 1 indicates moderate selectivity, and values above 2 indicates strong selectivity. The significance of the *d'* for each ROI's preferred category was tested against the *d'* of the other non-preferred categories using 10,000 permutations. The ROC analysis assesses how well the neural activation can be used to distinguish one category from all others, giving a classification measure where 1 represents perfect separation (i.e., all stimuli from the preferred category elicits more activation than all the other stimuli) and 0.5 represents chance level. Finally, the top-N analysis determines the proportion of stimuli from an area's preferred category that were present within the top 25 most activating images.

The ROIs showed significant and robust category selectivity. For body-selective areas, LOTC-body right: $d' = 2.44$; *AUC* = 0.97 (indicating a 97% probability that a randomly chosen body stimulus would elicit a higher response than a non-body stimulus), and 72% of *top-25* responses were bodies; LOTC-body left: $d' = 2.0$; *AUC* = 0.93; *top-25* = 64%; VOTC-body right: $d' = 2.48$; *AUC* = 0.96; *top-25* = 76%. For hand-selective areas, LOTC-hand right: $d' = 2.1$; *AUC* = 0.94; *top-25* = 68%; LOTC-hand left: $d' = 2.3$; *AUC* = 0.95; *top-25* = 64%; VOTC-hand left: $d' = 1.2$; *AUC* = 0.81; *top-25* = 44%. Finally, for tool-selective areas: LOTC-tool left: $d' = 1.38$; *AUC* = 0.84; *top-25* = 72%; VOTC-tool left: $d' = 0.72$; *AUC* = 0.69; *top-25* = 52%.

The *d'* for the preferred category was statistically significant ($p < .0125$, Bonferroni corrected with 4 comparisons) against the *d'* for the other categories in all ROIs except in VOTC-tool, which generally showed weaker selectivity and more similar responses between all inanimate object categories (*d'* tool vs manipulable: $p = 0.013$; *d'* tool vs non-manipulable: $p = 0.044$); this weaker selectivity suggests that ventral tool-selective responses may reflect broader inanimate object properties rather than tool identity per se, an interpretation we return to later below.

Together, these findings confirm that OTC contains spatially distinct clusters tuned to bodies, hands, and tools. In VOTC, we found right-lateralized body selectivity and left-lateralized hand selectivity but found weak and non-significant tool selectivity relative to other inanimate objects. These results replicate and extend previous parcellations of OTC (Bracci et al., 2010, 2012; Pillet et al., 2024a, 2024b; Rosenke et al., 2021), reinforce hemispheric lateralization for body, hand, and tool responses (Pillet et al., 2024a), and confirm that hand and tool selectivity is stronger in lateral than ventral OTC (Cortinovis et al., 2025b).

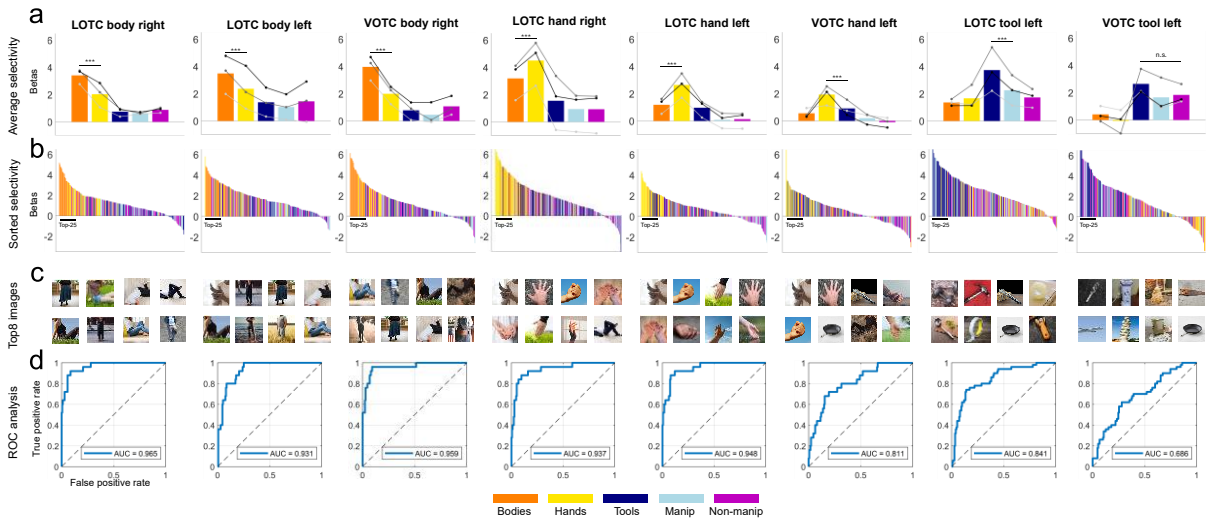


Figure 4.2. Functional selectivity profiles across neural category-selective areas. a) Average selectivity by category. Lines indicate the response for each single subject. Stars indicate significance ($p < .001$) between the highest activation and the second highest activation. **b)** Sorted selectivity (from highest to lowest) for each stimulus, averaged across subjects. Black line under each plot indicates the

top-25 most activating stimuli. **c)** Top-8 images activating each area. **d)** Area under the Receiver Operating Characteristic Curve (AUC). The AUC is a ratio of the true positive rate vs. false positive rate, measuring how well neural responses discriminate between the preferred stimulus class from all other stimulus categories.

Brain encoding models confirm fine-grained category selectivity

While functional selectivity analysis confirmed the possibility of dissociating body-, hand-, and tool-selective responses in OTC, we sought to investigate the feature spaces underlying category-selective areas, while providing a robust and unbiased test of their selectivity. Following the approach of Murty et al. (2021), we trained a series of encoding models to predict the mean fMRI BOLD response within each ROI from image features extracted from a deep neural network (ResNet-50, see methods for details). After training, we evaluated the performance of the models and used them to perform an in-silico screening of large, independent sets of images (ImageNet, Deng et al., 2009; ecoset, Mehrer et al. 2021), to identify the visual features that drove the highest predicted activations, thereby revealing the learned tuning properties of each brain area.

Results show that the encoding models for all category-selective areas were highly successful in predicting their activation patterns. For LOTC-body right the model achieved a prediction accuracy of r = 0.67 against a group noise ceiling (reported as $\sqrt{r_{SB}}$) of 0.8; LOTC-body left: r = 0.59 (noise ceiling [0.74]); VOTC-body right: r = 0.66 (noise ceiling [0.72]); LOTC-hand right: r = 0.67 (noise ceiling [0.78]); LOTC-hand left: r = 0.66 (noise ceiling [0.78]); VOTC-hand left: r = 0.41 (noise ceiling [0.55]); LOTC-tool left: r = 0.51 (noise ceiling [0.64]); VOTC-tool left: r = 0.45 (noise ceiling [0.56])). Notice that the noise ceiling was computed as a (Spearman-Brown corrected) split-half on the averaged group data (contrary to the reliability analysis, which was performed at the subject level), thus representing an overestimation (upper bound) of the noise ceiling, and indicating high reliability of the data.

Since we were able to successfully train area-specific encoding models with good performance, we then proceeded to evaluate their responses to a separate set of images. Specifically, we screened large image datasets and evaluated the response prediction of each encoding model for all images. For a first qualitative assessment, we visualized the images that were predicted to most strongly activate each model (see Supplementary Material); then, to move beyond idiosyncratic repetitions of nearly identical images, we also applied an unsupervised clustering analysis to the top 0.1% of images (≈2,500). Visual features were extracted from the penultimate layer of a ResNet-50 pretrained on ImageNet, and k-means clustering with cosine distance was used to group images by shared visual characteristics. This approach allowed us to identify the most consistent features across top-ranked stimuli, to provide a more comprehensive picture of each model's tuning preferences, and to test for the presence of recurrent features that are not related to the hypothesised preferred category. In this section, we describe the general response of body-, hand-, and tool-trained encoding models; in the following sections we target more specifically the distinct feature properties and representations that distinguish the different areas.

A qualitative inspection of the top-ranked images revealed striking and highly specific tuning preferences for each model (see Figure 4.3 for examples after clustering procedure and Supplementary Materials for the "raw" unclustered top-100 most activating images). For models trained on body-selective areas, the top images consistently depicted whole human bodies, often capturing the full body form rather than isolated body parts. Clustering confirmed this preference: the majority of images within each cluster belonged to human bodies. Similarly, for the models trained on data from hand-selective areas, the top-ranked images (including in each cluster) almost exclusively featured human hands in various postures, both in isolation and interacting with objects. Importantly, the images depict hands and not the general body form, further confirming dissociable responses to hands from other body-parts; additionally, VOTC-hand left seemed to be sensitive to specific types of inanimate objects, such as tools like brooms or brushes. Finally, the model trained on data from LOTC-tool left

identified a preference for tools and highly manipulable objects: the top images in each cluster consisted of a diverse array of graspable, functional objects, including nails, weights, silverware, paperclips, and Swiss-army knives, and hands interacting with objects (such as hammers); visualising the overall most activating images also show strong activations for more stereotypical tools such as hammers, pliers, scissors. Despite their varied appearances, the common characteristic was that they were all highly manipulable man-made objects designed for a specific function; all of them were also handheld objects. Many (but not all) objects had a metallic surface and presented an elongated handle. VOTC-tool left, instead, presented a different pattern, responding to some tool objects (such as brushes) but also to other inanimate – and often squared – objects in general (stoplights, cassette tapes).

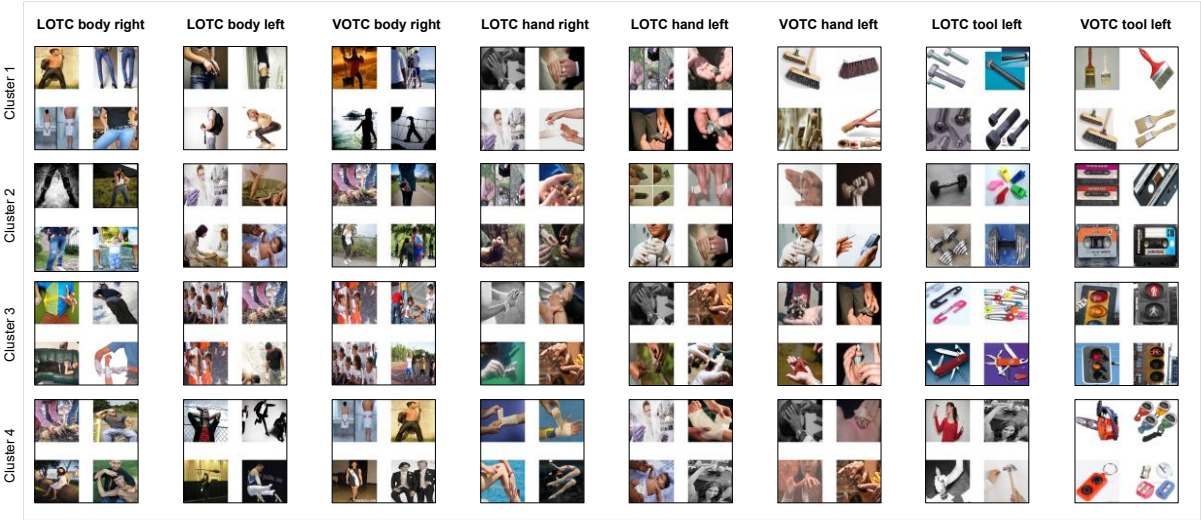


Figure 4.3. Encoding modelling analysis. Predictions for each image in each encoding model were sorted from highest to lowest, and images sharing similar visual content among the top 2500 images (based on the penultimate layer of a ResNet-50) were clustered. Each cluster was ranked based on the average selectivity of the images contained therein. The top 4 images are shown for each cluster.

These findings provide evidence for the expected category selectivity and against the influence of potential stimulus selection bias. Even when challenged with a massive and

diverse stimulus set, the encoding models consistently identified images belonging to each area's preferred category as the most effective stimuli. This suggests that the category selectivity observed in these areas is a genuine and robust property of the visual system.

The tuning to the preferred category in each model was further confirmed by the salience maps analysis (Figure 4.4). We presented encoding models with images containing the exact same scenes depicting body-parts interacting with objects and tools, sampled from the COCO dataset (Lin et al., 2014). Each encoding model generally revealed preference within the image for their respective category: body models responded to parts of the images depicting body-parts such as arms and shoulders (sometimes including – but not specific to – hands), and at the same time showing no preference for inanimate objects; hand models were specifically tuned to hands (and generally not other body-parts); and tool models preferred highly-manipulable objects such as scissors, brushes, or cutlery, with the encoding model trained on VOTC-tool data exhibiting a more general preference for inanimate objects.

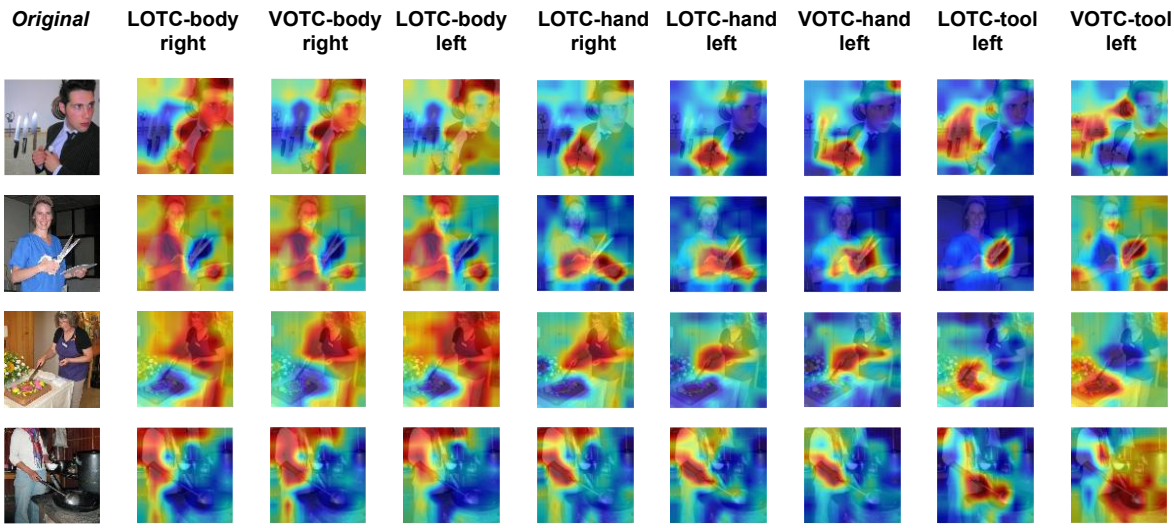


Figure 4.4. Saliency ("heatmaps") analysis. 2000 masks (8x8) were generated that randomly occluded parts of the image, and the effect on the encoding models' predictions were calculated. Areas of the images that elicit a decrease in response are color-coded in red.

Distinct feature sensitivity underlies areas selective for the same category of objects

The analyses conducted so far provide evidence for category-specific processing of objects within each selective area. However, even areas that respond to the same category may process distinct underlying features. For example, tools are often both elongated and metallic; two regions might respond similarly to a tool, while representing different aspects of the stimulus (e.g., shape versus surface properties). To investigate such potential feature sensitivities, we compared the prediction scores of encoding models across areas. Specifically, we computed pairwise differences between the models' predictions for each image and identified the images that maximally activated one model relative to another. We performed these contrasts both across hemispheres (e.g., LOTC-hand right vs. LOTC-hand left) and across pathways (e.g., LOTC-hand left vs. VOTC-hand left). This analysis revealed systematic and interpretable feature-level dissociations between areas selective for the same category, aligning with known distinctions between cortical pathways and hemispheres. Results can be visualized in Figure 4.5, where we show representative images for each contrast.

When contrasting hand-selective areas in the two hemispheres, we found that LOTC-hand left responded strongly (relative to LOTC-hand right) to handheld manipulable objects, typically tools such as hammers, markers, remote controls, and spatulas. By contrast, LOTC-hand right showed a clear preference (relative to LOTC-hand left) for images of bodies, often embedded in richly textured backgrounds. These results make sense in light of the position of each hand-selective area in their respective hemisphere: as shown by previous studies, the left hand-selective area forms part of an action-related gradient that relate hands with tools (Cortinovis et al., 2025a), whereas the hand-selective area, in the right hemisphere, neighbours body-selectivity (Downing et al., 2001), which usually forms a more extended cluster in the right hemisphere (at least in right-handers, Willems et al., 2010). Interestingly, LOTC-hand left did not simply show stronger tool responses than VOTC-hand left. Instead, LOTC-hand left was driven by more complex scenes containing people interacting with animals or multiple objects

arranged in rectilinear (e.g., buildings) or curvilinear (e.g., clustered umbrellas) layouts, whereas VOTC-hand left was more responsive to single elongated objects against simple backgrounds (e.g., columns, brooms on a uniform-colored background). The fact that sensitivity to tools cannot distinguish between ventral and lateral hand-selective areas is again consistent with the action-related gradient: tool selectivity is "sandwiched" between lateral and ventral hand-selective areas, and VOTC-hand has also been implicated in representing action-related object properties (Cortinovis et al., 2025a) or small highly motor relevant objects in general (Magri et al., 2021). The observed differences may therefore reflect distinct sensitivities to low- or mid-level visual features such as elongation and curvilinearity, properties that have been found to explain large portions of OTC in general and hand and tool selectivity in particular (e.g., Chen et al., 2018; Yue et al., 2020), or may underlie sensitivity to complex scenes involving interactions (consistent with the recently proposed interaction-sensitive lateral pathway (Puce, 2024; Wurm & Caramazza, 2022).

Comparisons between ventral and lateral tool-selective areas revealed a similar division of labor. LOTC-tool left was preferentially driven by classically defined tools and manipulable objects, often depicted with hands (e.g., hammers, pens, syringes). In contrast, VOTC-tool left responded more broadly to inanimate objects, particularly large, non-manipulable items such as streetlights, buildings, vehicles, and bell towers. These results support a division between lateral tool-selective regions tuned to action-related objects and ventral regions tuned more generally to large non-graspable inanimate objects (Magri et al., 2021; Zhao et al., 2025).

Finally, differences between body-selective areas across the two hemispheres were less straightforward to interpret. Contrasting models trained on body-selective areas across hemispheres revealed that – perhaps surprisingly – differences were driven primarily by mid-level visual features and material properties. Images driving LOTC-body left more strongly than LOTC-body right included inanimate objects, such as tennis rackets, fire extinguishers, phones, and small metallic objects, such as nails; conversely, images that strongly activated LOTC-body right, but not LOTC-body left were characterized by saturated colors, rectilinear

and repetitive textures, and materials such as textiles, fabrics, and wood. A similar distinction emerged when comparing lateral and ventral regions in the right hemisphere: LOTC-body right responded more strongly to images containing rectilinear features, including some form of textiles, fabrics, wooden pavements, and, additionally, sofas; VOTC-body right was preferentially driven by small circular or stubby objects, such as flyswatters, fire extinguishers, or small phone booths. The relative difficulty of interpreting these contrasts likely reflects greater similarity across body-selective areas compared to hand- and tool-selective areas. Indeed, the average activation difference between LOTC-tool left and VOTC-tool left (≈ 3.6) and between LOTC-hand left and LOTC-hand right (≈ 2.3) was notably larger than that between body areas (≈ 1.7). In other words, the more similar two models' prediction scores are, the less distinctive, and thus less interpretable, the feature differences that separate them become.

Figure 4.5. Distinct feature preferences across areas responding to the same category. Encoding models trained on different areas selective for the same category were compared by pairwise subtracting their prediction scores. The top images were then ranked and clustered. Each cluster was ranked based on average selectivity. Example images are shown for each cluster.

Together, these results demonstrate that body-, hand-, and tool-selective areas exhibit robust image-level functional selectivity, while ANN-based encoding models provide a stringent test

of such selectivity. Areas selective for the same category differ systematically in the features they encode, with these differences aligning with known distinctions between ventral and lateral pathways and between hemispheres (Cortinovis et al., 2025a; Wurm & Caramazza, 2022).

Discussion

In this study, we first characterized functional selectivity for category-selective areas in OTC for bodies, hands, and tools at the level of individual images, and then trained ANN-based encoding models on those responses. These models captured fine-grained category preferences observed in the brain and revealed distinct feature representations underlying selectivity for areas responding to the same category. We show that body-, hand-, and tool-selective areas are reliably dissociable across ventral and lateral OTC and exhibit both strong selectivity for their preferred categories and graded responses to non-preferred categories based on their hypothesised computational goal. Encoding models trained on each area's responses predicted neural activity with high accuracy, and large-scale image screening confirmed that the strongest predicted activations consistently belonged to the expected category. Crucially, comparisons between encoding models revealed systematic differences in feature sensitivity among areas selective for the same category, such as heightened sensitivity to action-related properties in the lateral tool-selective area and more general processing of the features belonging to many inanimate objects in the ventral "tool" area. Together, these findings indicate that category-selective areas in OTC reflect multiple, partially dissociable computational roles that depend on their position along ventral vs. lateral OTC or in distinct hemispheres.

Research has consistently shown that discrete regions in OTC respond selectively to specific object categories (Kanwisher, 2010), including faces, bodies, scenes, words (Downing et al., 2001; Epstein & Kanwisher, 1998; Kanwisher et al., 1997; McCandliss et al., 2003), and more recently food (Khosla et al., 2022b; Jain et al., 2023; Pennock et al., 2023). However, converging evidence suggests that this organization is more fine-grained than this canonical map implies. Areas traditionally treated as unitary, such as FFA or EBA, can be parcellated into multiple clusters with distinct functional profiles (Weiner & Grill-Spector, 2011; Rosenke et al., 2021; Grill-Spector et al., 2017; Pillet et al., 2024a). For body-parts specifically, previous work has identified limb-selective patches arranged in a ring-like fashion around motion-

sensitive cortex (Weiner & Grill-Spector, 2011, 2013), as well as dissociable responses to hands versus whole bodies (Orlov et al., 2010; Bracci et al., 2010, 2015).

Despite this evidence, much of the literature still treats large regions such as EBA as homogeneous units, averaging across heterogeneous subareas. This practice risks obscuring the complexity of cortical organization and underestimates the degree to which OTC responses are distributed across distinct, specialized clusters. Moreover, other forms of category-selectivity, such as for tools or manipulable objects (Chao et al., 1999; Cortinovis et al., 2025b), remains more controversial and are often considered less robust than selectivity for "canonical" categories (Downing et al., 2006). Importantly, such distinctions are not a trivial, secondary phenomena: hand-selective areas, for example, support specific functions distinct from body-selective areas, including sensitivity to hand pose and connections to parieto-frontal circuits for tool use (Bracci et al., 2012; 2018; Knights et al., 2021; Striem-Amit et al., 2017), and even have distinct developmental trajectories (Nordt et al., 2023). Recognizing this finer-grained organization is essential for understanding how OTC supports distinct feature spaces and behavioral goals (Peelen et al., 2017; Bracci & Op de Beeck, 2023).

Our study advances this debate by providing a robust test of functional selectivity for body-, hand-, and tool-selective areas. At the image level, these areas showed distinct and graded response profiles: hand-selective areas responded preferentially to hands over whole bodies and exhibited strong secondary responses to tools; tool-selective areas responded most strongly to tools and highly manipulable objects rather than to all small inanimate objects. This interpretation is consistent with a graded organization extending from effector-specific representations within hand-selective areas and the hand–tool overlap (Bracci et al., 2012, 2013) to more general processing of manipulability in more ventral and anterior regions (Cortinovis et al., 2025a, 2025b), and, together, these findings argue against accounts based on coarse category labels alone.

Beyond investigating OTC functional responses, our work builds on and extends ANN-based encoding approaches used to model responses in ventral visual cortex in general (Güçlü & van Gerven, 2015, 2017; Wen et al., 2018; Agrawal et al., 2014; Gu et al., 2021; Murty et al., 2021; Özçelik & VanRullen, 2023) and body-selective areas in particular (Marrazzo et al., 2023; Yashiro et al., 2025). Following the rationale of Murty et al. (2021), we trained encoding models on the mean response of category-selective areas and tested their functional profiles on large-scale image datasets. Encoding models are particularly important given the inherent constraints of fMRI and the limited sampling of image space achievable even in large-scale neural datasets (e.g., Allen et al., 2022; Hebart et al., 2023; Lahner et al., 2024; St-Laurent et al., 2025). By serving as virtual instantiations of category-selective areas, these models provide a complementary approach to these datasets: once trained, they can be evaluated on millions of images, enabling fine-grained tests of selectivity and representational structure that would be impractical with direct human experiments alone (Murty et al., 2021).

Here, we successfully trained encoding models based on the responses of body-, hand-, and tool-selective areas, enabling interpretation and visualization of the visual features that differentiate both across category-selective areas and within areas responding to the same category. Whereas body-selective areas showed relatively homogeneous sensitivity to body images, hand- and tool-selective areas could be dissociated along an action-related dimension. Left hemisphere hand-selective areas were more sensitive to end-effector objects (e.g., tools such as hammers or scissors), while right hemisphere hand-selective areas showed greater sensitivity to the body form. A similar dissociation emerged within tool-selective cortex: ventral regions showed reduced sensitivity to action-related features and were more broadly tuned to inanimate object properties, whereas lateral regions were more selectively tuned to canonical tools and end-effectors objects. Importantly, these findings demonstrate that encoding models are not limited to testing predefined hypotheses about category selectivity: by probing the functional space of trained models, it is possible to identify the visual features that drive regional selectivity, thereby moving beyond “word models” of

category organization (Goodhill, 2007; Murty et al., 2021). This approach enables principled differentiation between areas that respond to the same category but encode distinct feature dimensions. In line with this, hemispheric differences within hand-selective cortex were captured by differential sensitivity to action-related objects, with left-lateralized tuning for inanimate objects carrying high action information and right hemisphere regions preferentially encoding body-related features. This pattern is consistent with the left-lateralization of the hand–tool overlap (Bracci et al., 2012) and previously reported action-related topographic organization (Cortinovis et al., 2025a).

Moreover, the encoding models allowed us to differentiate areas in the ventral and lateral pathways that are sensitive to tools. While the lateral tool area is more tuned to properties related to high manipulability, the ventral tool area is not selective to tools per se but is biased towards processing inanimate objects in general, including large objects (see Mahon et al., 2007 for a different perspective on tool "selectivity" within the medial fusiform gyrus). Additional visual properties such as elongation and metallic appearance also contributed to model predictions, in line with recent fine-grained characterizations of responses to manipulable objects (Almeida et al., 2023); however, we think that these features, while important, are not the primary features represented within lateral tool areas, that might be instead involved in representing action-related properties of objects. Notice that the features that are mapped via the encoding model are extracted from a CNN, that might be biased towards visual properties and be unable to access the action-related properties characterising tools (see Cortinovis et al., 2025a); thus, shape and material properties may be just a proxy for the action-related features represented by the tool (and hand) areas. Overall, these results are consistent with the distinct processing between lateral and ventral OTC, one involved in representing action-related properties (and shape), and one possibly dedicated to the elaboration of surface properties of objects, that are typical not of tools specifically but of many object categories (Cortinovis et al., 2025b; Mahon & Almeida, 2024).

While our results support the existence of distinct category-selective clusters, they also intersect with a complementary view in which OTC is organized along shared representational dimensions such as animacy (Bao et al., 2020; Konkle & Caramazza, 2013; Kriegeskorte et al., 2008), real-world size (Konkle & Oliva, 2012), aspect-ratio (Bao et al., 2020), and action (Cortinovis et al., 2025a). Recent work suggests that neighbouring and partly overlapping activations for whole-bodies, hands, tools, manipulable objects, and even food in lateral OTC (Almeida et al., 2023; Bracci et al., 2010, 2012; Ritchie et al., 2024a) can be explained by action-related dimensions describing how body parts interact with objects (Cortinovis et al., 2025a). Our findings do not contradict this perspective: while our aim was to dissociate as much as possible functional responses using targeted contrasts to maximise selectivity, we also observe that each selective area presented a graded response to the non-preferred categories. For instance, while hand-selective areas are most strongly driven by hands, the images that next-best activated the area were body-parts (specifically when legs were prominent) and, above all, tools, consistent with an action-related representational gradient, where hands and tools share similar effector properties. Thus, a category-based account coexists with a dimensional framework (Contier et al., 2024; van Dyck et al., 2025), and ANN-based encoding models offer a powerful way to bridge the two. Therefore, our results are consistent with a more graded view of object preference (Mur et al., 2012; Ritchie et al., 2025). Finally, these results open avenues for directly comparing biological and artificial representations in models that capture both functional and spatial organization of visual cortex (Deb et al., 2024; Lu et al., 2025; Margalit et al., 2024; Qian et al., 2024), providing insights into the computational pressures shaping cortical topography (Finzi et al., 2023; Truong & Hasson, 2025).

In summary, our study both confirms and extends classic accounts of category selectivity in OTC. It confirms them by demonstrating robust selectivity for bodies, hands, and tools, and extends them by revealing systematic feature-level differences within and across category-selective areas, offering a computational framework for understanding how functional

specialization emerges in OTC. When provided with this feature information, encoding models can reliably distinguish between whole bodies, hands, and tools, categories that are often grouped together or not considered. Even if not ascribing to a strict modular view of category selectivity, our analyses confirm the presence of information related to hands and tools in visual cortex, suggesting that it is important to consider these categories or the domain they support to have a complete picture of the ventral visual cortex organization and topography.

Chapter 5 - General Discussion

5.1 Summary

In this thesis, we proposed and empirically tested a novel dimension shaping the spatial organization of object categories in occipitotemporal cortex (OTC), and we investigated how this dimension relates to the functional dissociation between lateral and ventral visual pathways. In Chapter 2, building on prior findings of spatial proximity and representational similarity between hands and tools, we identified a broader topographic organization that extends beyond these categories and is structured by action-related properties of visual stimuli, most notably graspability and end-effector features. This action dimension accounts for the systematic arrangement of multiple object categories along the left lateral OTC, including bodies, hands, tools, and manipulable objects. In Chapter 3, we extended this framework to recently-discovered category-selective areas responding to images of food. We showed that food-selective cortex in left lateral OTC overlaps with areas responsive to tools and manipulable objects and is tuned to action-related properties, consistent with the proposed action dimension. In contrast, ventral food-selective areas exhibited sensitivity to surface-based visual properties, including colour and ensemble statistics, revealing a dissociation between lateral and ventral pathways in their representational content. In Chapter 4, we further examined the relationship between category selectivity and representational content by directly comparing body-, hand- and tool-selective areas across hemispheres and pathways. Image-level functional selectivity analyses confirmed robust and dissociable category-selective responses to bodies, hands, and tools, while encoding models clarified how these responses are shaped by distinct visual, semantic, and action-related features, consistent with their anatomical positioning in different hemispheres and pathways. Finally, across chapters, we compared visual cortex with both topographic and non-topographic artificial neural networks (ANNs), highlighting both the promise and current limitations of such

models in reproducing the spatial and functional organization observed in human ventral visual cortex. Together, these findings support a multidimensional account of OTC organization in which action-related properties play a central role in structuring representations within the lateral visual pathway.

In this final chapter, we first discuss the implications of these results for our understanding of the spatial and functional organization of visual cortex, focusing on the proposed division of labor between ventral and lateral pathways. We then consider what inductive biases are necessary for topographic ANNs to better capture the action-based spatial organization of visual cortex. We conclude by outlining two broader theoretical questions raised by this work: what is the origin of the action dimension, and more generally, what functional advantages does cortical clustering confer?

5.2 Multiple constraints shape the spatial and functional organization of ventral and lateral OTC

Understanding how object representations are organized in high-level visual cortex remains a central challenge in cognitive neuroscience. One influential way to conceptualize cortical organization is in topographic terms: a highly multidimensional feature space is projected onto the two-dimensional cortical sheet such that neighbouring neurons (or voxels) respond to neighbouring points in that space (Durbin & Mitchinson, 1990; Kourtzi & Connor, 2011). Previous work, reviewed in Chapter 1, has identified several candidate dimensions that strongly guide this mapping, including animacy, real-world size, and aspect ratio (Bao et al., 2020; Konkle & Caramazza, 2013; Konkle & Oliva, 2012), whereas other work proposed modular organization around category-selective areas (Kanwisher, 2010) that intersect these maps (Op de Beeck et al., 2008; Weiner & Grill-Spector).

However, the previously proposed dimensions cannot fully account for several findings reported here: the spatial and representational proximity of hands and tools and their arrangement along a gradient (Chapter 2), the emergence of food-selective responses in specific regions of left lateral OTC (Chapter 3), or the dissociations observed between bodies and hands and between ventral and lateral representations of the same category (Chapter 4). In the present work, we propose that, in order to fully account for the organization of object categories in OTC, we need to include another relevant dimension that is based on the action-related properties of objects.

If we look at the topographic organization of object categories as investigated in Chapter 2, the location of each category activation in left lateral OTC can be explained by the mapping of multiple dimensions. For instance, dorsal\posterior regions are sensitive to animate categories (bodies and hands) and more ventral\anterior regions prefer inanimate objects (tools and manipulable objects); however, animacy is not sufficient - and neither is size - as not only hands and tools partially overlap, tools also elicit a higher activation and more extended cluster than manipulable objects. Instead, the mapping requires a further dimension: action, and specifically the way our body-part effectors (i.e., hands) interact with the inanimate object effectors (i.e., tools). The concurrent mapping of these two dimensions can explain the posterior-to-anterior gradient of activations found (see vector-of-ROIs, Chapter 2); moreover, it could also successfully predict the location of food-selective cluster in left lateral OTC that strongly overlapped with tools and manipulable objects (common graspable features) but not with hands (hands are not graspable and food objects are not effectors); and, finally, it could explain the functional selectivity at the image level for hands and tools in visual cortex, with sensitivity to action-related effector objects in left lateral hand- and tool-selective cortex, whereas the ventral tool area and the right-lateralized hand area did not show the same selectivity, as they are not part of this action-related topographic organization.

Overall, in this work we demonstrate the role of action-related properties of objects in explaining the complex organization of object categories in visual cortex. However, the distinct

organization exhibited by ventral and lateral OTC warrants deeper discussion: the mapping principles that guide this organization, and therefore the underlying computational goals that require the different mapping, must be different to have such a distinct organization (Aflalo & Graziano, 2011; Graziano & Aflalo 2007). In the next section we focus on the differences between ventral and lateral OTC and discuss how our work informs this distinction.

Ventral vs. Lateral OTC

The results we obtained across all studies are broadly consistent with recent proposals that divide the ventral visual stream into two partially dissociable pathways, one spanning its ventral surface and one its lateral surface (Lingnau & Downing, 2015; Pitcher & Ungerleider, 2021; Puce, 2024; Ritchie et al., 2024b; Weiner & Grill-Spector, 2013; Wurm & Caramazza, 2022). Following Ritchie et al. (2024b), we refer to these as ventral and lateral *pathways*, rather than fully separate *streams*.

The most important account for the current work proposes that the ventral pathway primarily supports object recognition, whereas the lateral pathway supports action recognition (Wurm & Caramazza, 2022). This accounts for the dissociable functional responses found in the two pathways: for instance, colour responses are present in ventral but not lateral OTC (Brouwer & Heeger, 2009), consistent with their role in object recognition but their limited utility for action recognition; motion information and sensitivity to dynamic stimuli, instead, are mostly represented within the lateral surface, as motion is usually more important for action than object recognition (Beauchamp et al., 2002; Pitcher et al., 2019; Küçük et al., 2023).

Our results further inform this division by demonstrating that action-related object properties, (such as graspable and end-effector properties) are supported by the lateral pathway. Specifically, objects that contain a high-level of action-related information drive the responses within a posterior-inferior region in lateral OTC, a region that has been previously associated with the perceptual recognition of actions involving objects (Wurm et al., 2017). Moreover, the effects we found were strongly left-lateralized, consistent with the known role of the left

hemisphere in representing transitive actions and tool use (Brandi et al., 2014; Lewis, 2006). Instead, the right hemisphere is more sensitive to social information and actions that involve other biological agents, a pattern that is supported by recent studies investigating face and body sensitivity (Isik et al., 2017; Saxe, 2006; Gandolfo et al., 2024) and (indirectly) by our own data showing greater sensitivity to faces and whole bodies and increased overlap between hand, face, and body responses in the right hemisphere.

Our second study also highlights some features that are important for the ventral pathway. Consistent with previous work, we found that ventral OTC showed sensitivity to colour (Lafer-Sousa et al., 2016) and ensemble statistics (Cant & Xu, 2012). Importantly, lateral OTC exhibited greater sensitivity to single objects than to groups of objects, a pattern that may reflect its role in encoding graspable and action-relevant properties: a single object (e.g., an apple) affords more precise action information than a collection of objects.

Many studies have reported that category-selective areas are often mirrored on ventral and lateral surfaces (e.g., FFA and OFA, FBA and EBA). While earlier proposals argued that ventral areas are tuned to the whole stimulus and lateral areas process their constituent parts (Taylor & Downing, 2011; Taylor et al., 2007), the pathway framework may better account for this duplication of selectivity. Indeed, under this view, category-selective responses come in pairs because areas positioned in a distinct pathway process features necessary to solve the specific computational goal of each pathway: object recognition in ventral regions and action-related processing in lateral regions. Supporting this interpretation, previous work found that lateral “limb”-selective areas surround motion responses (Weiner & Grill-Spector, 2011), and our own results show that selectivity for tools and hands is stronger in lateral than ventral regions, and lateral tool-selective areas are more strongly tuned to action-related properties than their ventral counterparts (Chapter 4), consistent with their role in supporting action-related properties.

Taken together, these findings suggest that the lateral pathway may be best understood as an (inter)action pathway, supporting both object-oriented actions and social interactions, a pathway that includes further functional subdivisions and hemispheric biases (Puce, 2024). Within this framework, the occipitotemporal sulcus (OTS) emerges as a particularly interesting structure. Although traditionally considered part of the ventral pathway, the results reported here blur the ventral–lateral boundary: in fact, the OTS hand area exhibits action sensitivity (as tested with univariate, multivariate, and encoding modelling analyses), making it similar to the lateral hand-selective area, and tool responses are “sandwiched” between these two areas. OTS may thus represent a transitional anatomical structure along which representational dimensions shift (see for instance its role in representing a shift in eccentricity band, Daniel-Hertz et al., 2025), analogous to the role of the mid-fusiform sulcus for eccentricity, animacy, and real-world size (Weiner et al., 2014). Further work will be necessary to clarify the precise contribution of areas along OTS to object versus action processing.

5.3 Divergences between topographic computational models and the brain and future directions to close this gap

An important goal in computational neuroscience is narrowing the gap between ANNs and the functional and spatial organization of visual cortex. Recent work on topographic deep artificial neural networks (TDANNs) represents a step forward in this direction. Indeed, work with TDANNs demonstrate that implementing biologically-inspired constraints within ANNs allows them to capture not only the functional but also the spatial organization of visual cortex (Blauch et al., 2022; Deb et al., 2025; Lu et al., 2025; Margalit et al., 2024).

However, our results highlight a critical limitation of current topographic models: their failure to capture the action-related organizing dimension observed in lateral OTC. In particular, TDANNs do not reproduce the graded sensitivity to manipulability and end-effector properties

that differentiates hand- and tool-selective areas, nor the hemispheric asymmetries observed in our data. In this section, we outline two of the possible approaches through which we can narrow this gap. One approach requires a fundamental change in paradigm, relying on embodied agents that physically act in the world, whereas the second remains within the TDANN framework (or, more generally, the "neuroconnectionist" framework, Doerig et al., 2023).

Embodiment

One possibility is that the emergence of action-based topography requires embodied agents that actively interact with their environment. In other words, a system must be physically able to perform actions in the environment to learn the association between specific objects, and the consequences of the actions that those objects afford. Visual experience alone would therefore be insufficient, implying the need for a fundamentally different modelling paradigm that integrates perception and action (Bartnik et al., 2025). Recent work in AI-robotics integration is starting to develop systems that are capable of performing flexible actions by imitating the typical ways in which humans grasp and use objects and generally act in the environment, often within a reinforcement learning framework (Arunachalam et al., 2022; Bahl et al., 2023; Qin et al., 2022). However, as most of these robotic systems still learn via visual input (e.g., videos of humans performing actions), it is unclear how necessary it is to physically perform those actions to develop an action-based organization.

Relatedly, converging evidence from human neuroscience challenges the necessity of embodiment – at least in the narrow sense of motor experience (Rietveld & Kiverstein, 2014) – for the computations supported by lateral OTC. For example, individuals born without hands develop a typical hand action observation network (Vannuscorps et al., 2019) and, most importantly, exhibit the overlap between hand- and tool-selective areas despite having no motor experience with hand–tool interactions (Striem-Amit et al., 2017). These results suggest

that action-related organization in lateral OTC may arise from evolutionary or developmental constraints other than direct sensorimotor experience.

The TDANN framework

Considering that embodiment might not be strictly necessary to account for the results presented in the current work, in this section we turn to possible modifications of core ingredients of ANNs that are directly implementable within the TDANN framework. Specifically, we briefly discuss possible changes to the task loss, to the visual diet, to the architecture, or to the spatial loss, changes that might improve their ability to capture the action-related spatial organization of OTC.

Task loss within TDANNs

One possible source of divergence between TDANNs and lateral OTC involves the training objective. The TDANN model used here was trained with self-supervised contrastive learning, a training regime which was shown to be effective in capturing several aspects of ventral visual cortex organization, including the emergence of category-selective clusters (Konkle & Alvarez, 2022; Margalit et al., 2024; Prince et al., 2024) and the division between ventral, lateral, and dorsal visual pathways (Finzi et al., 2023; Tang et al., 2025). Our results, however, suggest that the same training objective is insufficient to elicit the action-related organization observed for object categories such as hands and tools.

However, recent studies suggest that additional pressures related to goal-directed behavior are important to shape higher-level visual cortex organization (Mineault et al., 2021; Reza et al., 2025), including for specific category-selective areas (Dobs et al., 2022; Dwivedi et al. 2021a; 2021b). Importantly, the failure of TDANNs trained with action recognition objectives to match neural data in lateral OTC (including the present results) suggests that simply adding explicit action labels may not be sufficient. Instead, capturing action-related organization may

require training objectives that go beyond object or action classification and more closely reflect how visual representations are used to guide interactions with the environment.

Visual diet

The visual diet itself may be an important inductive bias, as evidenced by recent studies (Conwell et al., 2024). A promising way forward is to train the network with developmental data, exploiting head-mounted cameras that allow collecting hours of egocentric videos from the daily experience of infants, such as the SAYCam dataset (Sullivan et al., 2021), or the recent BabyView Dataset that has been used to train from scratch a self-supervised network (Long et al., 2024). Interestingly, work in this area has shown that the "visual diet" of infants and their fixation patterns progressively switch between the first few months of life to the first couple of years to a visual diet mostly consisting of faces to one mostly consisting of hands and hand-object interactions (Fausey et al., 2016; Frank et al., 2012; Long et al., 2022), highlighting the potential benefit of this approach to capture action-related processing.

Architectural priors and hemispheric biases

A second limitation of current TDANNs concerns the absence of anatomical and hemispheric priors. In the human brain, category-selective areas and larger-scale topographic dimensions exhibit systematic lateralization, with faces and bodies biased toward the right hemisphere and hands, tools, and words toward the left (Rossion & Lochy, 2022; see Blauch et al., 2025 for a review). The present work extends this evidence by demonstrating a hemispheric dissociation in the representation of action-related properties, with left hemisphere regions showing greater sensitivity to end-effector objects.

By contrast, TDANNs are typically initialized without any hemispheric asymmetries or large-scale anatomical constraints, treating cortex as a homogeneous two-dimensional sheet. Incorporating these priors – for example by simulating (recurrent) connectivity or training the two hemispheres with distinct inductive biases – is a promising direction for improving

alignment with biological data. More generally, introducing architectural biases before training may help bridge the gap between TDANNs and visual cortex, in a similar way as how genetic and developmental constraints shape neural organization prior to experience (Versace et al., 2018; Zador, 2019).

The role of the spatial loss

Finally, it is worth considering whether limitations of TDANNs arise from the spatial loss itself. The spatial loss constrains nearby units to develop similar representations, thereby encouraging topographic organization, but it is agnostic to the specific content of that organization. As a result, the form of emergent topography critically depends on the features learned through the task objective and architectural priors. While more biologically-plausible spatial losses have recently been proposed (Lu et al., 2025; Qian et al., 2024; Truong et al., 2025), we argue that the failure to capture action-related organization in lateral OTC is unlikely to be resolved by modifying the spatial loss alone. Instead, future progress will probably depend on refining multiple inductive biases at once.

5.4 Open questions and future directions

The findings reported in this work raise several open questions and point to possible directions for future research. A first set of questions is specific to the action-related dimension proposed here and concerns its origin: how does this dimension emerge over development and evolution, and what constraints shape its organization in OTC? Addressing this issue is critical for understanding whether action-related topography reflects hard-wired innate biases, learning-driven plasticity, or their interaction.

A second, more general set of questions is related to the broader rationale for topographic and category-selective organization in high-level visual cortex. Why does the visual system exhibit clustered representations at all, and what computational advantages does such

organization confer? In this section, I consider existing theoretical and empirical accounts of these issues and outline how developmental, connectivity-based, time-resolved, and computational approaches could help clarify both the emergence and the functional significance of action-related and category-selective organization.

How does the action dimension emerge?

A question left unanswered by our results is the origin of the action dimension, across either evolution or development. To answer this question, other analyses are needed. In particular, developmental studies may help disentangle the role of experience vs. innate constraints on this organization, connectivity analyses could reveal the large-scale and possibly innate spatial constraints shaping object representations, and time-resolved methods could help establish the directionality of information flow underlying the emergence of action-related topography. Here, we briefly consider each in turn.

Developmental work suggests that high-level visual cortex in neonates exhibits a similar organization as that found in adults (before extensive experience in seeing or using objects), an organization that however undergoes refinement during development (Deen et al., 2017). Specifically, while previous studies found that face-, body-, and scene-related responses are already present in the first few days of life (Buiatti et al., 2019; Kosakowski et al., 2022; Saxe et al., 2022), to the best of our knowledge, no study to date has tested if tool or hand responses (or their overlap) are present at birth. Related to tools, the only existing evidence shows that tool responses in dorsal, lateral, and ventral visual cortex develop in parallel and are present at least by age 4, but also undergo further maturation during development, suggesting a role of both innate scaffold and experience (Kersey et al., 2015). Extending these investigations to neonates would elucidate the role of innate priors vs. experience in driving the action-related topography.

In Chapter 1, we reviewed a proposal that argued that the organization of visual cortex is due to long-range connectivity with regions outside of visual cortex itself (Mahon & Caramazza,

2011). One work investigated how connectivity structures the organization-by-dimension exhibited by high-level visual cortex (Konkle & Caramazza, 2017). In their study, the authors found that the tripartite distinction at the univariate level is reflected at the level of resting-state networks, dividing the brain into three large-scale networks encompassing regions responding to either small inanimate objects, big inanimate objects, and animate objects; interestingly, their work found that the network of small objects is connected with parietal regions in intraparietal sulcus involved in object-directed actions, suggesting that small objects are fundamentally tied to their degree of manipulability. Extending this approach, future work could investigate the connectivity between left lateral visual cortex and the rest of the brain in a similar way to determine whether regions organized along the proposed action-effector dimension participate in distinct, action-related functional networks.

A related appealing idea is that the origin of the action-related representations in lateral OTC arises from top-down input from the dorsal stream. Recently, aside from its established role in action guidance, the dorsal stream has been proposed to be involved in global shape perception and other aspects of object recognition (Ayzenberg & Behrmann, 2022; Freud et al., 2016). High-density electroencephalography found that these features are processed *first* in the dorsal stream and only *then* in the ventral stream, and suggested an influence of the former onto the latter (Ayzenberg et al., 2023; Collins et al., 2019; Gurariy et al., 2022). However, this just displaces the question onto the dorsal stream and leaves unresolved the origin of action-based representations within dorsal visual regions themselves, a question that our work currently cannot answer.

A possible unifying account for these results might consider all of these factors (innate connectivity patterns and the role of experience) together. For instance, we can speculate that connectivity (including from the dorsal stream) and other biases (e.g., retinotopy) present at birth may provide the scaffold (shaped during evolution) over which visual and motor experience with objects build a more robust structure that leads to the emergence of the spatial topography observed in OTC. Hebbian learning and wiring length constraints play a role in

this context by explaining why those structures form clustering in the first place and end up close to each other (Chklovskii & Koulakov, 2004; Hebb, 1949; Koulakov & Chklovskii, 2001; see also the following section).

What is the computational benefit of a category-selective organization?

Why does high-level visual cortex exhibit spatially organized and clustered category-selective responses? As reviewed in Chapter 1, several studies propose that topographic maps form for a simple reason: minimizing wiring length, and, therefore, reducing the metabolic cost of an already energetically expensive organ (Chklovskii & Koulakov, 2004). Indeed, in some cases clustering might not even provide any functional benefit, as suggested for example for ocular dominance columns (Horton & Adams, 2005).

With respect to high-level visual cortex, few studies have directly addressed the question of *why* neurons exhibiting similar sparse responses are spatially clustered in the first place (Reddy & Kanwisher, 2006), including for example the mentioned work on the dissociation between faces and objects in ANNs (Dobs et al., 2022; Kanwisher et al., 2023). In other words, does spatial clustering confer a computational advantage compared to a more distributed or scattered type of organization, or is the clustering merely an effect of wiring length constraints? One computational account argues that modularity arises from the need of the ventral visual stream to develop invariant object representations (Leibo et al., 2015). According to this view, invariance can be achieved for many object categories using general representations that transfer across objects, whereas for other categories (e.g., faces and bodies) the relevant transformations do not transfer broadly. In these cases, the system benefits from learning invariance within a specific transformation class, leading to the separation of these categories from others.

While these studies suggest that functional specialization can emerge as a solution reached by different systems to solve specific tasks, they do not explicitly test the benefits of clustering

similarly responsive neurons in a physical space. In particular, it remains unclear whether spatial clustering confers a computational advantage over a more distributed organization, or whether it is instead a byproduct of wiring length constraints. Evidence directly addressing this question has only recently begun to accumulate, suggesting that clustering may indeed provide specific computational advantages.

One possible advantage is increased robustness to clutter and noise. While biological vision is generally robust to clutter (i.e., the presence of multiple objects in the visual field) and noise, ANNs remain more sensitive to these perturbations (Jang et al., 2021). In visual cortex, evidence shows that a category-selective organization supports robustness to clutter. Specifically, when multiple objects are presented concurrently (e.g., a face and a house), robust discriminative responses are primarily maintained within category-selective voxels (Reddy & Kanwisher, 2007). The most direct evidence comes from macaque studies showing that neurons in category-selective areas preserve their selectivity when preferred objects (e.g., faces) are presented together with other object categories, whereas responses in non-selective regions are more strongly affected by clutter (Bao & Tsao, 2018); critically, this study argues that it is the *clustering* of selective neurons that provide such an advantage. These results were replicated with fMRI in humans (Kliger & Yovel 2020, 2024). Interestingly, Kliger & Yovel (2024) found that clustering and spatial proximity of face- and body-selective activations in ventral visual cortex not only guarantee robustness for the processing of faces and bodies, but it supports robust whole-person processing at the border between the two areas. This result is especially relevant to the current work, as spatial proximity between hands and tools can be interpreted in a similar light: hands and tools may cluster closely in visual cortex because the visual system needs to form robust representations of hand-tool interactions without needing to form a specific area that is exclusively dedicated to this interaction.

From a computational perspective, previous work showed that training networks on datasets that contain noisier stimuli (e.g., blurred images) improves their robustness to perturbations

(Jang et al., 2021; Jang & Tong, 2024). Moreover, ANNs have been shown to respond differently to clutter compared to human visual cortex (Mocz et al., 2023), although this work did not explicitly test how clutter affects category-selective units or the role of clustering. Recent evidence shows that inducing topographic constraints in ANNs similarly enhance robustness to noise (Qian et al., 2025; Truong & Hasson, 2025). Taken together, this evidence suggests that category selectivity and clustering might provide computational benefits and enhance robustness to noise and to the presence of clutter.

However, more direct evidence is needed in both human visual cortex and ANNs. For instance, future studies may exploit emerging frequency-tagging fMRI paradigms (Gao et al., 2018; Laurent et al., 2023) that allow dissociating signals from distinct stimuli presented simultaneously (Ngo et al., 2024; Rafeh et al., 2025). On the computational side, topographic and non-topographic networks could be systematically compared using stimulus sets that include both isolated and multi-object displays, testing how clutter affects both general object-responsive and category-selective units, and whether clustering modulates these effects. If ANNs exhibit patterns similar to those observed in visual cortex, they may provide mechanistic models that can be probed with more complex stimuli, such as natural scenes, enabling the generation of testable hypotheses about how category selectivity and topographic organization support perception in rich and cluttered environments.

Overall, the combination of neuroimaging - including time-resolved methods that combine temporal and spatial resolution such as magnetoencephalography - developmental studies, and computational modelling may hold the answers to these questions.

5.5 Conclusions

This work demonstrates that the organization of object representations in human occipitotemporal cortex cannot be fully explained by visual or semantic dimensions alone.

Across multiple neuroimaging and modelling approaches, I show that action-related object properties constitute a fundamental and independent organizing principle, particularly within left lateral OTC. By integrating evidence from spatial topography, functional selectivity, and encoding and topographic models, the findings reveal that ventral and lateral pathways support distinct but complementary computational goals.

Importantly, the results raise open questions about how action-related dimensions emerge during development and learning and what computational advantages category-selective and action-based organizations confer. The limitations of current topographic neural networks to capture action-related organization suggests that future models will need to incorporate additional constraints, including more ecological visual diets, task demands and richer learning objectives. Addressing these questions will be essential for developing a complete account of how high-level visual representations support behaviorally-relevant goals and for narrowing the gap between biological and artificial vision systems.

References

Adamson, K., & Troiani, V. (2018). Distinct and overlapping fusiform activation to faces and food. *Neuroimage*, *174*, 393-406.

Aflalo, T. N., & Graziano, M. S. (2011). Organization of the macaque extrastriate visual cortex re-examined using the principle of spatial continuity of function. *Journal of neurophysiology*, *105*(1), 305-320.

Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., ... & Kay, K. (2022). A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature neuroscience*, *25*(1), 116-126.

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in cognitive sciences*, *15*(3), 122-131.

Arcaro, M. J., & Livingstone, M. S. (2021). On the relationship between maps and domains in inferotemporal cortex. *Nature Reviews Neuroscience*, *22*(9), 573-583.

Arcaro, M., & Livingstone, M. (2024). A whole-brain topographic ontology. Annual Review of Neuroscience, 47.

Arcaro, M. J., Schade, P. F., Vincent, J. L., Ponce, C. R., & Livingstone, M. S. (2017). Seeing faces is necessary for face-domain formation. *Nature neuroscience*, *20*(10), 1404-1412.

Arunachalam, S. P., Silwal, S., Evans, B., & Pinto, L. (2022). Dexterous imitation made easy: A learning-based framework for efficient dexterous manipulation. *arXiv preprint arXiv:2203.13251*.

Astafiev, S. V., Stanley, C. M., Shulman, G. L., & Corbetta, M. (2004). Extrastriate body area in human occipital cortex responds to the performance of motor actions. *Nature neuroscience*, *7*(5), 542-548.

Avery, J. A., Liu, A. G., Ingeholm, J. E., Gotts, S. J., & Martin, A. (2021). Viewing images of foods evokes taste quality-specific activity in gustatory insular cortex. *Proceedings of the National Academy of Sciences*, *118*(2), e2010932118.

Ayzenberg, V., & Behrmann, M. (2022). Does the brain's ventral visual pathway compute object shape?. *Trends in Cognitive Sciences*, *26*(12), 1119-1132.

Ayzenberg, V., Simmons, C., & Behrmann, M. (2023). Temporal asymmetries and interactions between dorsal and ventral visual pathways during object recognition. *Cerebral Cortex Communications*, *4*(1), tgad003.

Bahl, S., Mendonca, R., Chen, L., Jain, U., & Pathak, D. (2023). Affordances from human videos as a versatile representation for robotics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13778-13790).

Baker, C., & Kravitz, D. (2024). Insights from the evolving model of two cortical visual pathways. *Journal of cognitive neuroscience*, *36*(12), 2568-2579.

Bao, P., She, L., McGill, M., & Tsao, D. Y. (2020). A map of object space in primate inferotemporal cortex. *Nature*, *583*(7814), 103-108.

Bao, P., & Tsao, D. Y. (2018). Representation of multiple objects in macaque category-selective areas. *Nature communications*, *9*(1), 1774.

Barrow, H. G., Bray, A. J., & Budd, J. M. (1996). A self-organizing model of “color blob” formation. *Neural Computation*, *8*(7), 1427-1448.

Bartnik, C. G., Sartzetaki, C., Sanchez, A. P., Molenkamp, E., Bommer, S., Vukšić, N., & Groen, I. I. (2025). Representation of locomotive action affordances in human behavior, brains, and deep neural networks. *Proceedings of the National Academy of Sciences*, *122*(24), e2414005122.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron*, *34*(1), 149-159.

Bednar, J. A., & Wilson, S. P. (2016). Cortical maps. *The Neuroscientist*, *22*(6), 604-617.

Blasdel, G. G. (1992). Orientation selectivity, preference, and continuity in monkey striate cortex. *Journal of Neuroscience*, *12*(8), 3139-3161.

Blauch, N. M., Behrmann, M., & Plaut, D. C. (2022). A connectivity-constrained computational account of topographic organization in primate high-level visual cortex.

Proceedings of the National Academy of Sciences, *119*(3), e2112566119.

Blauch, N. M., Plaut, D. C., Vin, R., & Behrmann, M. (2025). Individual variation in the functional lateralization of human ventral temporal cortex: Local competition and long-range coupling. *Imaging Neuroscience*, *3*, imag_a_00488.

Bonhoeffer, T., & Grinvald, A. (1991). Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature*, *353*(6343), 429-431.

Bougou, V., Vanhoyland, M., Bertrand, A., Van Paesschen, W., Op de Beeck, H., Janssen, P., & Theys, T. (2024). Neuronal tuning and population representations of shape and category in human visual cortex. *Nature communications*, *15*(1), 4608.

Bowers, J. S., Malhotra, G., Dujmović, M., Montero, M. L., Tsvetkov, C., Biscione, V., ... & Blything, R. (2023). Deep problems with neural network models of human vision. *Behavioral and Brain Sciences*, *46*, e385.

Bracci, S., Caramazza, A., & Peelen, M. V. (2015). Representational similarity of body parts in human occipitotemporal cortex. *Journal of Neuroscience*, *35*(38), 12977-12985.

Bracci, S., Caramazza, A., & Peelen, M. V. (2018). View-invariant representation of hand postures in the human lateral occipitotemporal cortex. *NeuroImage*, *181*, 446-452.

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), 1443-1456.

Bracci, S., Cavina-Pratesi, C., Connolly, J. D., & Ietswaart, M. (2016). Representational content of occipitotemporal and parietal tool areas. *Neuropsychologia*, *84*, 81-88.

Bracci, S., Ietswaart, M., Peelen, M. V., & Cavina-Pratesi, C. (2010). Dissociable neural responses to hands and non-hand body parts in human left extrastriate visual cortex. *Journal of neurophysiology*, *103*(6), 3389-3397.

Bracci, S., & Op de Beeck, H. P. (2016). Dissociations and associations between shape and category representations in the two visual pathways. *Journal of Neuroscience*, *36*(2), 432-444.

Bracci, S., & Op de Beeck, H. P. (2023). Understanding human object vision: a picture is

worth a thousand representations. *Annual review of psychology*, *74*(1), 113-135.

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.

Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial vision*, *10*(4), 433-436.

Brandi, M. L., Wohlschläger, A., Sorg, C., & Hermsdörfer, J. (2014). The neural correlates of planning and executing actual tool use. *Journal of Neuroscience*, *34*(39), 13183-13194.

Brants, M., Baeck, A., Wagemans, J., & de Beeck, H. P. O. (2011). Multiple scales of organization for object selectivity in ventral visual cortex. *Neuroimage*, *56*(3), 1372-1381.

Brewer, A. A., & Barton, B. (2016). Maps of the auditory cortex. *Annual review of neuroscience*, *39*(1), 385-407.

Brouwer, G. J., & Heeger, D. J. (2009). Decoding and reconstructing color from responses in human visual cortex. *Journal of Neuroscience*, *29*(44), 13992-14003.

Buiatti, M., Di Giorgio, E., Piazza, M., Polloni, C., Menna, G., Taddei, F., ... & Vallortigara, G. (2019). Cortical route for facelike pattern processing in human newborns. *Proceedings of the National Academy of Sciences*, *116*(10), 4625-4630.

Cant, J. S., & Goodale, M. A. (2007). Attention to form or surface properties modulates different regions of human occipitotemporal cortex. *Cerebral cortex*, *17*(3), 713-731.

Cant, J. S., & Xu, Y. (2012). Object ensemble processing in human anterior-medial ventral visual cortex. *Journal of Neuroscience*, *32*(22), 7685-7700.

Cant, J. S., & Xu, Y. (2015). The impact of density and ratio on object-ensemble representation in human anterior-medial ventral visual cortex. *Cerebral Cortex*, *25*(11), 4226-4239.

Cant, J. S., & Xu, Y. (2017). The contribution of object shape and surface properties to object ensemble representation in anterior-medial ventral visual cortex. *Journal of*

cognitive neuroscience, *29*(2), 398-412.

Cao, R., & Yamins, D. (2024a). Explanatory models in neuroscience, Part 1: Taking mechanistic abstraction seriously. *Cognitive Systems Research*, *87*, 101244.

Cao, R., & Yamins, D. (2024b). Explanatory models in neuroscience, Part 2: Functional intelligibility and the contravariance principle. *Cognitive Systems Research*, *85*, 101200.

Cortinovis, D., Peelen, M. V., & Bracci, S. (2025b). Tool representations in human visual cortex. *Journal of Cognitive Neuroscience*, *37*(3), 515-531.

Cortinovis, D., Truong, N., Op de Beeck, H., & Bracci, S. (2025a). Investigating action topography in visual cortex and deep artificial neural networks. *Nature Communications*.

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature neuroscience*, *2*(10), 913-919.

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.

Chiou, R., Humphreys, G. F., Jung, J., & Ralph, M. A. L. (2018). Controlled semantic cognition relies upon dynamic and flexible interactions between the executive 'semantic control' and hub-and-spoke 'semantic representation'systems. *cortex*, *103*, 100-116.

Chklovskii, D. B., Schikorski, T., & Stevens, C. F. (2002). Wiring optimization in cortical circuits. *Neuron*, *34*(3), 341-347.

Chklovskii, D. B., & Koulakov, A. A. (2004). Maps in the brain: what can we learn from them?. *Annu. Rev. Neurosci.*, *27*(1), 369-392.

Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific reports*, *6*(1), 27755.

Coggan, D. D., & Tong, F. (2023). Spikiness and animacy as potential organizing principles

of human ventral visual cortex. *Cerebral Cortex*, *33*(13), 8194-8217.

Contier, O., Baker, C. I., & Hebart, M. N. (2024). Distributed representations of behaviour-derived object dimensions in the human visual system. *Nature Human Behaviour*, *8*(11), 2179-2193.

Collins, E., Freud, E., Kainerstorfer, J. M., Cao, J., & Behrmann, M. (2019). Temporal dynamics of shape processing differentiate contributions of dorsal and ventral visual pathways. *Journal of Cognitive Neuroscience*, *31*(6), 821-836.

Conwell, C., Prince, J. S., Kay, K. N., Alvarez, G. A., & Konkle, T. (2024). A large-scale examination of inductive biases shaping high-level visual representation in brains and machines. *Nature communications*, *15*(1), 9383.

Cowell, R. A., & Cottrell, G. W. (2013). What evidence supports special processing for faces? A cautionary tale for fMRI interpretation. *Journal of Cognitive Neuroscience*, *25*(11), 1777-1793.

Daniel-Hertz, E., Yao, J. K., Gregorek, S., Hoyos, P. M., & Gomez, J. (2025). An eccentricity gradient reversal across high-level visual cortex. *Journal of Neuroscience*, *45*(2).

de Haas, B., Sereno, M. I., & Schwarzkopf, D. S. (2021). Inferior occipital gyrus is organized along common gradients of spatial and face-part selectivity. *Journal of Neuroscience*, *41*(25), 5511-5521.

Deb, M., Deb, M., & Murty, N. (2025). TopoNets: High performing vision and language models with brain-like topography. *arXiv preprint arXiv:2501.16396*.

Deen, B., Richardson, H., Dilks, D. D., Takahashi, A., Keil, B., Wald, L. L., ... & Saxe, R. (2017). Organization of high-level visual cortex in human infants. *Nature communications*, *8*(1), 13995.

Dehghani, A., Qian, X., Farahani, A., & Bashivan, P. (2025). Credit-based self organizing maps: training deep topographic networks with minimal performance degradation. In *The Thirteenth International Conference on Learning Representations*.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition?. *Neuron*, *73*(3), 415-434.

Dobs, K., Martinez, J., Kell, A. J., & Kanwisher, N. (2022). Brain-like functional specialization emerges spontaneously in deep neural networks. *Science advances*, *8*(11), eabl8913.

Dobs, K., Yuan, J., Martinez, J., & Kanwisher, N. (2023). Behavioral signatures of face perception emerge in deep neural networks optimized for face recognition. *Proceedings of the National Academy of Sciences*, *120*(32), e2220642120.

Doerig, A., Sommers, R. P., Seeliger, K., Richards, B., Ismael, J., Lindsay, G. W., ... & Kietzmann, T. C. (2023). The neuroconnectionist research programme. *Nature Reviews Neuroscience*, *24*(7), 431-450.

Doshi, F. R., & Konkle, T. (2023). Cortical topographic motifs emerge in a self-organized map of object space. *Science Advances*, *9*(25), eade8187.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*(5539), 2470-2473.

Downing, P. E., Chan, A. Y., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain specificity in visual cortex. *Cerebral cortex*, *16*(10), 1453-1461.

Durbin, R., & Mitchison, G. (1990). A dimension reduction framework for understanding cortical maps. *Nature*, *343*(6259), 644-647.

Dwivedi, K., Bonner, M. F., Cichy, R. M., & Roig, G. (2021). Unveiling functions of the visual cortex using task-specific deep neural networks. *PLoS computational biology*, *17*(8), e1009267.

Dwivedi, K., Cichy, R. M., & Roig, G. (2021). Unraveling representations in scene-selective brain regions using scene-parsing deep neural networks. *Journal of cognitive neuroscience*, *33*(10), 2032-2043.

Eickenberg, M., Gramfort, A., Varoquaux, G., & Thirion, B. (2017). Seeing it all: Convolutional network layers map the function of the human visual system. *NeuroImage*, *152*, 184-194.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual

environment. *Nature*, *392*(6676), 598-601.

Erwin, E., Obermayer, K., & Schulten, K. (1995). Models of orientation and ocular dominance columns in the visual cortex: A critical comparison. *Neural computation*, *7*(3), 425-468.

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, *152*, 101-107.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, NY: 1991)*, *1*(1), 1-47.

Finzi, D., Margalit, E., Kay, K., Yamins, D. L., & Grill-Spector, K. (2023). A single computational objective drives specialization of streams in visual cortex. *bioRxiv*, 2023-12.

Fischl, B. (2012). FreeSurfer. *Neuroimage*, *62*(2), 774-781.

Foroni, F., Pergola, G., & Rumiati, R. I. (2016). Food color is in the eye of the beholder: The role of human trichromatic vision in food evaluation. *Scientific reports*, *6*(1), 37034.

Frank, M. C., Vul, E., & Saxe, R. (2012). Measuring the development of social attention using free-viewing. *Infancy*, *17*(4), 355-375.

Freud, E., Plaut, D. C., & Behrmann, M. (2016). 'What'is happening in the dorsal visual pathway. *Trends in cognitive sciences*, *20*(10), 773-784.

Gandolfo, M., Abassi, E., Balgova, E., Downing, P. E., Papeo, L., & Koldewyn, K. (2024). Converging evidence that left extrastriate body area supports visual sensitivity to social interactions. *Current Biology*, *34*(2), 343-351.

Gao, X., Gentile, F., & Rossion, B. (2018). Fast periodic stimulation (FPS): a highly effective approach in fMRI brain mapping. *Brain Structure and Function*, *223*(5), 2433-2454.

Gatys, L., Ecker, A. S., & Bethge, M. (2015). Texture synthesis using convolutional neural networks. *Advances in neural information processing systems*, *28*.

Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., & Brendel, W. (2018, November). ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. In *International conference on learning*

Representations.

Gomez, J., Barnett, M., & Grill-Spector, K. (2019). Extensive childhood experience with Pokémon suggests eccentricity drives organization of visual cortex. *Nature human behaviour*, *3*(6), 611-624.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in neurosciences*, *15*(1), 20-25.

Graziano, M. S., & Aflalo, T. N. (2007). Mapping behavioral repertoire onto the cortex. *Neuron*, *56*(2), 239-251.

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annu. Rev. Neurosci.*, *27*(1), 649-677.

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.

Grill-Spector, K., Weiner, K. S., Kay, K., & Gomez, J. (2017). The functional neuroanatomy of human face perception. *Annual review of vision science*, *3*(1), 167-196.

Grill-Spector, K., Weiner, K. S., Gomez, J., Stigliani, A., & Natu, V. S. (2018). The functional neuroanatomy of face perception: from brain measurements to deep neural networks. *Interface Focus*, *8*(4), 20180013.

Groen, I. I., Dekker, T. M., Knapen, T., & Silson, E. H. (2022). Visuospatial coding as ubiquitous scaffolding for human cognition. *Trends in Cognitive Sciences*, *26*(1), 81-96.

Gross, C. G., Bender, D. B., & Rocha-Miranda, C. D. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science*, *166*(3910), 1303-1306.

Gross, C. G. (2008). Single neuron studies of inferior temporal cortex. *Neuropsychologia*, *46*(3), 841-852.

Güçlü, U., & Van Gerven, M. A. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience*, *35*(27), 10005-10014.

Gupta, P., & Dobs, K. (2025). Human-like face pareidolia emerges in deep neural networks

optimized for face and object recognition. *PLOS Computational Biology*, *21*(1), e1012751.

Gurariy, G., Mruczek, R. E., Snow, J. C., & Caplovitz, G. P. (2022). Using high-density electroencephalography to explore spatiotemporal representations of object categories in visual cortex. *Journal of Cognitive Neuroscience*, *34*(6), 967-987.

Hiramatsu, C., Goda, N., & Komatsu, H. (2011). Transformation from image-based to perceptual representation of materials along the human ventral visual pathway. *Neuroimage*, *57*(2), 482-494.

Haak, K. V., & Beckmann, C. F. (2018). Objective analysis of the topological organization of the human cortical visual connectome suggests three visual pathways. *Cortex*, *98*, 73-83.

Haxby, J. V., Gobbini, M. I., & Nastase, S. A. (2020). Naturalistic stimuli reveal a dominant role for agentic action in visual representation. *Neuroimage*, *216*, 116561.

He, C., Peelen, M. V., Han, Z., Lin, N., Caramazza, A., & Bi, Y. (2013). Selectivity for large nonmanipulable objects in scene-selective visual cortex does not require visual experience. *Neuroimage*, *79*, 1-9.

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* (pp. 770-778).

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. Psychology press.

Henderson, M. M., Tarr, M. J., & Wehbe, L. (2023). A texture statistics encoding model reveals hierarchical feature selectivity across human visual cortex. *Journal of Neuroscience*, *43*(22), 4144-4161.

Henderson, M. M., Tarr, M. J., & Wehbe, L. (2025). Origins of food selectivity in human visual cortex. *Trends in Neurosciences*, *48*(2), 113-123.

Henriksson, L., Mur, M., & Kriegeskorte, N. (2015). Faciotopy—a face-feature map with face-like topology in the human occipital face area. *cortex*, *72*, 156-167.

Hong, H., Yamins, D. L., Majaj, N. J., & DiCarlo, J. J. (2016). Explicit information for category-orthogonal object properties increases along the ventral stream. *Nature neuroscience*, *19*(4), 613-622.

Horton, J. C., & Adams, D. L. (2005). The cortical column: a structure without a function. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 837-862.

Huang, T., Song, Y., & Liu, J. (2022). Real-world size of objects serves as an axis of object space. *Communications biology*, *5*(1), 749.

Huerta, C. I., Sarkar, P. R., Duong, T. Q., Laird, A. R., & Fox, P. T. (2014). Neural bases of food perception: Coordinate-based meta-analyses of neuroimaging studies in multiple modalities. *Obesity*, *22*(6), 1439-1446.

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*, *76*(6), 1210-1224.

Isik, L., Koldewyn, K., Beeler, D., & Kanwisher, N. (2017). Perceiving social interactions in the posterior superior temporal sulcus. *Proceedings of the National Academy of Sciences*, *114*(43), E9145-E9152.

Jacobs, R. A., Jordan, M. I., & Barto, A. G. (1991). Task decomposition through competition in a modular connectionist architecture: The what and where vision tasks. *Cognitive science*, *15*(2), 219-250.

Jacobs, R. A., & Jordan, M. I. (1992). Computational consequences of a bias toward short connections. *Journal of cognitive neuroscience*, *4*(4), 323-336.

Jain, N., Wang, A., Henderson, M. M., Lin, R., Prince, J. S., Tarr, M. J., & Wehbe, L. (2023). Selectivity for food in human ventral visual cortex. *Communications Biology*, *6*(1), 175.

Jagadeesh, A. V., & Gardner, J. L. (2022). Texture-like representation of objects in human visual cortex. *Proceedings of the National Academy of Sciences*, *119*(17), e2115302119.

Jagadeesh, A. V., & Livingstone, M. (2024). Texture bias in primate ventral visual cortex. In *ICLR 2024 Workshop on Representational Alignment*.

Jang, H., McCormack, D., & Tong, F. (2021). Noise-trained deep neural networks effectively predict human vision and its neural responses to challenging images. *PLoS biology*, *19*(12), e3001418.

Jang, H., & Tong, F. (2024). Improved modeling of human vision by incorporating robustness to blur in convolutional neural networks. *Nature Communications*, *15*(1), 1989.

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.

Jozwik, K. M., Kriegeskorte, N., Storrs, K. R., & Mur, M. (2017). Deep convolutional neural networks outperform feature-based but not categorical models in explaining object similarity judgments. *Frontiers in psychology*, *8*, 1726.

Kanwisher, N. (2010). Functional specificity in the human brain: a window into the functional architecture of the mind. *Proceedings of the national academy of sciences*, *107*(25), 11163-11170.

Kanwisher, N. (2025). Animal models of the human brain: Successes, limitations, and alternatives. *Current Opinion in Neurobiology*, *90*, 102969.

Kanwisher, N., Khosla, M., & Dobs, K. (2023). Using artificial neural networks to ask ‘why’questions of minds and brains. *Trends in Neurosciences*, *46*(3), 240-254.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of neuroscience*, *17*(11), 4302-4311.

Kay, K. N. (2018). Principles for models of neural information processing. *NeuroImage*, *180*, 101-109.

Kersey, A. J., Clark, T. S., Lussier, C. A., Mahon, B. Z., & Cantlon, J. F. (2015). Development of tool representations in the dorsal and ventral visual object processing pathways.

Cerebral Cortex, *26*(7), 3135-3145.

Khaligh-Razavi, S. M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS computational biology*, *10*(11), e1003915.

Khosla, M., Murty, N. A. R., & Kanwisher, N. (2022). A highly selective response to food in human visual cortex revealed by hypothesis-free voxel decomposition. *Current Biology*, *32*(19), 4159-4171.

King, M. L., Groen, I. I., Steel, A., Kravitz, D. J., & Baker, C. I. (2019). Similarity judgments and cortical visual responses reflect different properties of object and scene categories in naturalistic images. *NeuroImage*, *197*, 368-382.

Kitada, R., Yoshihara, K., Sasaki, A. T., Hashiguchi, M., Kochiyama, T., & Sadato, N. (2014). The brain network underlying the recognition of hand gestures in the blind: the supramodal role of the extrastriate body area. *Journal of Neuroscience*, *34*(30), 10096-10108.

Kietzmann, T., McClure, P., & Kriegeskorte, N. (2019, January 25). Deep Neural Networks in Computational Neuroscience. *Oxford Research Encyclopedia of Neuroscience.*

Kliger, L., & Yovel, G. (2020). The functional organization of high-level visual cortex determines the representation of complex visual stimuli. *Journal of Neuroscience*, *40*(39), 7545-7558.

Kliger, L., & Yovel, G. (2024). Distinct Yet Proximal Face-and Body-Selective Brain Regions Enable Clutter-Tolerant Representations of the Face, Body, and Whole Person. *Journal of Neuroscience*, *44*(24).

Klyachko, V. A., & Stevens, C. F. (2003). Connectivity optimization and the positioning of cortical areas. *Proceedings of the national academy of sciences*, *100*(13), 7937-7941.

Knights, E., Mansfield, C., Tonin, D., Saada, J., Smith, F. W., & Rossit, S. (2021). Hand-selective visual regions represent how to grasp 3D tools: Brain decoding during real actions. *Journal of Neuroscience*, *41*(24), 5263-5273.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological cybernetics*, *43*(1), 59-69.

Komatsu, H., & Goda, N. (2018). Neural mechanisms of material perception: Quest on Shitsukan. *Neuroscience*, *392*, 329-347.

Konkle, T., & Alvarez, G. A. (2022). A self-supervised domain-general learning framework for human ventral stream representation. *Nature communications*, *13*(1), 491.

Konkle, T., & Oliva, A. (2012). A real-world size organization of object responses in occipitotemporal cortex. *Neuron*, *74*(6), 1114-1124.

Konkle, T., & Caramazza, A. (2013). Tripartite organization of the ventral stream by animacy and object size. *Journal of Neuroscience*, *33*(25), 10235-10242.

Konkle, T., & Caramazza, A. (2017). The large-scale organization of object-responsive cortex is reflected in resting-state network architecture. *Cerebral cortex*, *27*(10), 4933-4945.

Kourtzi, Z., & Connor, C. E. (2011). Neural representations for object perception: structure, category, and adaptive coding. *Annual review of neuroscience*, *34*(1), 45-67.

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.

Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature neuroscience*, *21*(9), 1148-1160.

Kubilius, J., Bracci, S., & Op de Beeck, H. P. (2016). Deep neural networks as a computational model for human shape sensitivity. *PLoS computational biology*, *12*(4), e1004896.

Kubota, E., Yan, X., Tung, S., Fascendini, B., Tyagi, C., Duhameau, S., ... & Grill-Spector, K. (2025). White matter connections of human ventral temporal cortex are organized by cytoarchitecture, eccentricity and category-selectivity from birth. *Nature Human Behaviour*, 1-16.

Küçük, E., Foxwell, M., Kaiser, D., & Pitcher, D. (2024). Moving and static faces, bodies,

objects, and scenes are differentially represented across the three visual pathways. *Journal of Cognitive Neuroscience*, *36*(12), 2639-2651.

Lafer-Sousa, R., & Conway, B. R. (2013). Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. *Nature neuroscience*, *16*(12), 1870-1878.

Lafer-Sousa, R., Conway, B. R., & Kanwisher, N. G. (2016). Color-biased regions of the ventral visual pathway lie between face-and place-selective regions in humans, as in macaques. *Journal of Neuroscience*, *36*(5), 1682-1697.

Laurent, M. A., Audurier, P., De Castro, V., Gao, X., Durand, J. B., Jonas, J., ... & Cottereau, B. R. (2023). Towards an optimization of functional localizers in non-human primate neuroimaging with (fMRI) frequency-tagging. *NeuroImage*, *270*, 119959.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436-444.

Lee, H., Margalit, E., Jozwik, K. M., Cohen, M. A., Kanwisher, N., Yamins, D. L., & DiCarlo, J. J. (2020). Topographic deep artificial neural networks reproduce the hallmarks of the primate inferior temporal cortex face processing network. *BioRxiv*, 2020-07.

Leibo, J. Z., Liao, Q., Anselmi, F., & Poggio, T. (2015). The invariance hypothesis implies domain-specific regions in visual cortex. *PLoS computational biology*, *11*(10), e1004390.

Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center–periphery organization of human object areas. *Nature neuroscience*, *4*(5), 533-539.

Lewis, J. W. (2006). Cortical networks related to human use of tools. *The neuroscientist*, *12*(3), 211-231.

Lingnau, A., & Downing, P. E. (2015). The lateral occipitotemporal cortex in action. *Trends in cognitive sciences*, *19*(5), 268-277.

Livingstone, M. S., & Hubel, D. H. (1984). Anatomy and physiology of a color system in the primate visual cortex. *Journal of Neuroscience*, *4*(1), 309-356.

Lindsay, G. W. (2021). Convolutional neural networks as a model of the visual system: Past, present, and future. *Journal of cognitive neuroscience*, *33*(10), 2017-2031.

Long, B., Konkle, T., Cohen, M. A., & Alvarez, G. A. (2016). Mid-level perceptual features

distinguish objects of different real-world sizes. *Journal of Experimental Psychology: General*, *145*(1), 95.

Long, B., Sparks, R. Z., Xiang, V., Stojanov, S., Yin, Z., Keene, G. E., ... & Frank, M. C. (2024). The BabyView dataset: High-resolution egocentric videos of infants' and young children's everyday experiences. *arXiv preprint arXiv:2406.10447*.

Long, B., Störmer, V. S., & Alvarez, G. A. (2017). Mid-level perceptual features contain early cues to animacy. *Journal of vision*, *17*(6), 20-20.

Long, B., Yu, C. P., & Konkle, T. (2018). Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proceedings of the National Academy of Sciences*, *115*(38), E9015-E9024.

Lu, Z., Doerig, A., Bosch, V., Krahmer, B., Kaiser, D., Cichy, R. M., & Kietzmann, T. C. (2025). End-to-end topographic networks as models of cortical map formation and human visual behaviour. *Nature Human Behaviour*, 1-17.

Macdonald, S. N., & Culham, J. C. (2015). Do human brain areas involved in visuomotor actions show a preference for real tools over visually similar non-tools?. *Neuropsychologia*, *77*, 35-41.

Magri, C., Konkle, T., & Caramazza, A. (2021). The contribution of object size, manipulability, and stability on neural responses to inanimate objects. *NeuroImage*, *237*, 118098.

Mahon, B. Z., & Almeida, J. (2024). Reciprocal interactions among parietal and occipito-temporal representations support everyday object-directed actions. *Neuropsychologia*, *198*, 108841.

Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-specific organization in the human brain does not require visual experience. *Neuron*, *63*(3), 397-405.

Mahon, B. Z., & Caramazza, A. (2011). What drives the organization of object knowledge in the brain?. *Trends in cognitive sciences*, *15*(3), 97-103.

Mahon, B. Z., Milleville, S. C., Negri, G. A., Rumiati, R. I., Caramazza, A., & Martin, A. (2007). Action-related properties shape object representations in the ventral stream.

Neuron, *55*(3), 507-520.

Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-specific organization in the human brain does not require visual experience. *Neuron*, *63*(3), 397-405.

Margalit, E., Lee, H., Finzi, D., DiCarlo, J. J., Grill-Spector, K., & Yamins, D. L. (2024). A unifying framework for functional organization in early and higher ventral visual cortex. *Neuron*, *112*(14), 2435-2451.

Marvi, A., Kanwisher, N., & Khosla, M. (2024). Sparse components distinguish visual pathways and their alignment to neural networks. *Journal of Vision*, *24*(10), 759-759.

Mazurchuk, S., Fernandino, L., Tong, J. Q., Conant, L. L., & Binder, J. R. (2024). The neural representation of body part concepts. *Cerebral Cortex*, *34*(6), bhae213.

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Trends in cognitive sciences*, *7*(7), 293-299.

Merzenich, M. M., Knight, P. L., & Roth, G. L. (1975). Representation of cochlea within primary auditory cortex in the cat. *Journal of neurophysiology*, *38*(2), 231-249.

Miikkulainen, R., Bednar, J. A., Choe, Y., & Sirosh, J. (2005). *Computational maps in the visual cortex*. New York, NY: Springer New York.

Mineault, P., Bakhtiari, S., Richards, B., & Pack, C. (2021). Your head is there to move you around: Goal-driven models of the primate dorsal pathway. *Advances in neural information processing systems*, *34*, 28757-28771.

Mocz, V., Jeong, S. K., Chun, M., & Xu, Y. (2023). Multiple visual objects are represented differently in the human brain and convolutional neural networks. *Scientific Reports*, *13*(1), 9088.

Moreau, Q., Parrotta, E., Pesci, U. G., Era, V., & Candidi, M. (2023). Early categorization of social affordances during the visual encoding of bodily stimuli. *Neuroimage*, *274*, 120151.

Mountcastle, V. B. (1997). The columnar organization of the neocortex. *Brain: a journal of neurology*, *120*(4), 701-722.

Mruczek, R. E., von Loga, I. S., & Kastner, S. (2013). The representation of tool and non-tool object information in the human intraparietal sulcus. *Journal of neurophysiology*, *109*(12), 2883-2896.

Mur, M., Ruff, D. A., Bodurka, J., De Weerd, P., Bandettini, P. A., & Kriegeskorte, N. (2012). Categorical, yet graded–single-image activation profiles of human category-selective cortical regions. *Journal of Neuroscience*, *32*(25), 8649-8662.

Nasr, S., Echavarria, C. E., & Tootell, R. B. (2014). Thinking outside the box: rectilinear shapes selectively activate scene-selective cortex. *Journal of Neuroscience*, *34*(20), 6721-6735.

Nauhaus, I., Nielsen, K. J., Disney, A. A., & Callaway, E. M. (2012). Orthogonal micro-organization of orientation and spatial frequency in primate primary visual cortex. *Nature neuroscience*, *15*(12), 1683-1690.

Nordt, M., Gomez, J., Natu, V. S., Rezai, A. A., Finzi, D., Kular, H., & Grill-Spector, K. (2021). Cortical recycling in high-level visual cortex during childhood development. *Nature human behaviour*, *5*(12), 1686-1697.

Ngo, G. N., Rafeh, R. W., Muller, L. E., Khan, A. R., Menon, R. S., Schmitz, T. W., & Mur, M. (2024). Frequency-tagged fMRI: A platform for fine-grained spatiotemporal analysis of cortical function. *bioRxiv*, 2024-12.

Obermayer, K., Ritter, H., & Schulten, K. (1990). A principle for the formation of the spatial structure of cortical feature maps. *Proceedings of the National Academy of Sciences*, *87*(21), 8345-8349.

Op de Beeck, H. P. (2010). Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses?. *Neuroimage*, *49*(3), 1943-1948.

Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J., & Kanwisher, N. G. (2006). Discrimination training alters object representations in human extrastriate cortex. *Journal of Neuroscience*, *26*(50), 13025-13036.

Op de Beeck, H. P., Haushofer, J., & Kanwisher, N. G. (2008). Interpreting fMRI data: maps, modules and dimensions. *Nature Reviews Neuroscience*, *9*(2), 123-135.

Op de Beeck, H. P., Pillet, I., & Ritchie, J. B. (2019). Factors determining where category-selective areas emerge in visual cortex. *Trends in cognitive sciences*, *23*(9), 784-797.

Orlov, T., Makin, T. R., & Zohary, E. (2010). Topographic representation of the human body De bin the occipitotemporal cortex. *Neuron*, *68*(3), 586-600.

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

Patel, G. H., Kaplan, D. M., & Snyder, L. H. (2014). Topographic organization in the brain: searching for general principles. *Trends in cognitive sciences*, *18*(7), 351-363.

Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *Journal of neurophysiology*, *93*(1), 603-608.

Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature reviews neuroscience*, *8*(8), 636-648.

Peelen, M. V., & Downing, P. E. (2017). Category selectivity in human visual cortex: Beyond visual object recognition. *Neuropsychologia*, *105*, 177-183.

Peelen, M. V., Bracci, S., Lu, X., He, C., Caramazza, A., & Bi, Y. (2013). Tool selectivity in left occipitotemporal cortex develops without vision. *Journal of cognitive neuroscience*, *25*(8), 1225-1234.

Penfield, W., & Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain: A journal of neurology*.

Pennock, I. M., Racey, C., Allen, E. J., Wu, Y., Naselaris, T., Kay, K. N., ... & Bosten, J. M. (2023). Color-biased regions in the ventral visual pathway are food selective. *Current Biology*, *33*(1), 134-146.

Perini, F., Caramazza, A., & Peelen, M. V. (2014). Left occipitotemporal cortex contributes to the discrimination of tool-associated hand actions: fMRI and TMS evidence. *Frontiers in human neuroscience*, *8*, 591.

Pillet, I., Cerrahoğlu, B., Philips, R. V., Dumoulin, S., & Op de Beeck, H. (2024). The position of visual word forms in the anatomical and representational space of visual

categories in occipitotemporal cortex. *Imaging neuroscience*, *2*, 1-28.

Pillet, I., Cerrahoğlu, B., Philips, R. V., Dumoulin, S., & Op de Beeck, H. (2024). A 7T fMRI investigation of hand and tool areas in the lateral and ventral occipitotemporal cortex. *PLoS One*, *19*(11), e0308565.

Pitcher, D., Ianni, G., & Ungerleider, L. G. (2019). A functional dissociation of face-, body-and scene-selective brain areas based on their response to moving and static stimuli. *Scientific reports*, *9*(1), 8242.

Pitcher, D., & Ungerleider, L. G. (2021). Evidence for a third visual pathway specialized for social perception. *Trends in cognitive sciences*, *25*(2), 100-110.

Powell, L. J., Kosakowski, H. L., & Saxe, R. (2018). Social origins of cortical face areas. *Trends in cognitive sciences*, *22*(9), 752-763.

Prince, J. S., Alvarez, G. A., & Konkle, T. (2024). Contrastive learning explains the emergence and function of visual category-selective regions. *Science Advances*, *10*(39), eadl1776.

Proklova, D., Kaiser, D., & Peelen, M. V. (2016). Disentangling representations of object shape and object category in human visual cortex: The animate–inanimate distinction. *Journal of cognitive neuroscience*, *28*(5), 680-692.

Puce, A. (2024). From motion to emotion: Visual pathways and potential interconnections. *Journal of Cognitive Neuroscience*, *36*(12), 2594-2617.

Qian, X., Dehghani, A. O., Farahani, A. B., & Bashivan, P. (2024). Local lateral connectivity is sufficient for replicating cortex-like topographical organization in deep neural networks. *bioRxiv*, 2024-08.

Qin, Y., Wu, Y. H., Liu, S., Jiang, H., Yang, R., Fu, Y., & Wang, X. (2022, October). Dexmv: Imitation learning for dexterous manipulation from human videos. In *European Conference on Computer Vision* (pp. 570-587). Cham: Springer Nature Switzerland.

Rafeh, R. W., Ngo, G. N., Muller, L. E., Khan, A. R., Menon, R. S., Mur, M., & Schmitz, T. W. (2025). Attentional enhancement and suppression of stimulus-synchronized BOLD oscillations. *bioRxiv*, 2025-01.

Rajimehr, R., Firoozi, A., Rafipoor, H., Abbasi, N., & Duncan, J. (2022). Complementary hemispheric lateralization of language and social processing in the human brain. *Cell reports*, *41*(6).

Ramirez, J. G., Vanhoyland, M., Ratan Murty, N. A., Decramer, T., Van Paesschen, W., Bracci, S., ... & Theys, T. (2024). Intracortical recordings reveal the neuronal selectivity for bodies and body parts in the human visual cortex. *Proceedings of the National Academy of Sciences*, *121*(51), e2408871121.

Ratan Murty, N. A., Teng, S., Beeler, D., Mynick, A., Oliva, A., & Kanwisher, N. (2020). Visual experience is not necessary for the development of face-selectivity in the lateral fusiform gyrus. *Proceedings of the National Academy of Sciences*, *117*(37), 23011-23020.

Ratan Murty, N. A., Bashivan, P., Abate, A., DiCarlo, J. J., & Kanwisher, N. (2021). Computational models of category-selective brain regions enable high-throughput tests of selectivity. *Nature communications*, *12*(1), 5540.

Reddy, L., & Kanwisher, N. (2006). Coding of visual objects in the ventral stream. *Current opinion in neurobiology*, *16*(4), 408-414.

Reddy, L., & Kanwisher, N. (2007). Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. *Current Biology*, *17*(23), 2067-2072.

Reich, L., Szwed, M., Cohen, L., & Amedi, A. (2011). A ventral visual stream reading center independent of visual experience. *Current Biology*, *21*(5), 363-368.

Reza, T., Jordan, E., Luo, S. T., Patel, K., Tang, J., & Niemeier, M. (2025). From Tasks to Topology: Dorsal and Ventral Streams Emerge in Optimized Neural Networks. *bioRxiv*, 2025-11.

Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., ... & Kording, K. P. (2019). A deep learning framework for neuroscience. *Nature neuroscience*, *22*(11), 1761-1770.

Ritchie, J. B., Andrews, S. T., Vaziri-Pashkam, M., & Baker, C. I. (2024a). Graspable foods and tools elicit similar responses in visual cortex. *Cerebral Cortex*, *34*(9), bhae383.

Ritchie, J. B., Montesinos, S., & Carter, M. J. (2024b). What is a visual stream?. *Journal of Cognitive Neuroscience*, *36*(12), 2627-2638.

Ritchie, J. B., Wardle, S. G., Vaziri-Pashkam, M., Kravitz, D. J., & Baker, C. I. (2025). Rethinking category-selectivity in human visual cortex. *Cognitive neuroscience*, 1-28.

Ritchie, J. B., Zeman, A. A., Bosmans, J., Sun, S., Verhaegen, K., & de Beeck, H. P. O. (2021). Untangling the animacy organization of occipitotemporal cortex. *Journal of Neuroscience*, *41*(33), 7103-7119.

Rietveld, E., & Kiverstein, J. (2014). A rich landscape of affordances. *Ecological psychology*, *26*(4), 325-352.

Rosenke, M., Van Hoof, R., Van Den Hurk, J., Grill-Spector, K., & Goebel, R. (2021). A probabilistic functional atlas of human occipito-temporal visual cortex. *Cerebral Cortex*, *31*(1), 603-619.

Rossion, B., & Lochy, A. (2022). Is human face recognition lateralized to the right hemisphere due to neural competition with left-lateralized visual word recognition? A critical review. *Brain Structure and Function*, *227*(2), 599-629.

Saadon-Grosman, N., Loewenstein, Y., & Arzy, S. (2020). The 'creatures' of the human cortical somatosensory system. *Brain communications*, *2*(1), fcaa003.

Saenz, M., & Langers, D. R. (2014). Tonotopic mapping of human auditory cortex. *Hearing research*, *307*, 42-52.

Santo, M. G. E., Maxim, O. S., & Schürmann, M. (2017). N1 responses to images of hands in occipito-temporal event-related potentials. *Neuropsychologia*, *106*, 83-89.

Sato, W. (2021). Color's indispensable role in the rapid detection of food. *Frontiers in Psychology*, *12*, 753654.

Sato, T., Uchida, G., Lescroart, M. D., Kitazono, J., Okada, M., & Tanifuji, M. (2013). Object representation in inferior temporal cortex is organized hierarchically in a mosaic-like structure. *Journal of Neuroscience*, *33*(42), 16642-16656.

Saxe, R. (2006). Uniquely human social cognition. *Current opinion in neurobiology*, *16*(2), 235-239.

Saxe, A., Nelli, S., & Summerfield, C. (2021). If deep learning is the answer, what is the question?. *Nature Reviews Neuroscience*, *22*(1), 55-67.

Saygin, Z. M., Osher, D. E., Koldewyn, K., Reynolds, G., Gabrieli, J. D., & Saxe, R. R. (2012). Anatomical connectivity patterns predict face selectivity in the fusiform gyrus. *Nature neuroscience*, *15*(2), 321-327.

Saygin, Z. M., Osher, D. E., Norton, E. S., Youssoufian, D. A., Beach, S. D., Feather, J., ... & Kanwisher, N. (2016). Connectivity precedes function in the development of the visual word form area. *Nature neuroscience*, *19*(9), 1250-1255.

Schwarzlose, R. F., Baker, C. I., & Kanwisher, N. (2005). Separate face and body selectivity on the fusiform gyrus. *Journal of Neuroscience*, *25*(47), 11055-11059.

Scott, L. S., & Arcaro, M. J. (2023). A domain-relevant framework for the development of face processing. *Nature Reviews Psychology*, *2*(3), 183-195.

Seeliger, K., Fritsche, M., Güçlü, U., Schoenmakers, S., Schoffelen, J. M., Bosch, S. E., & Van Gerven, M. A. J. (2018). Convolutional neural network-based encoding and decoding of visual object recognition in space and time. *NeuroImage*, *180*, 253-266.

Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O., & Connolly, A. C. (2015). The animacy continuum in the human ventral vision pathway. *Journal of cognitive neuroscience*, *27*(4), 665-678.

Silson, E. H., Chan, A. W. Y., Reynolds, R. C., Kravitz, D. J., & Baker, C. I. (2015). A retinotopic basis for the division of high-level scene processing between lateral and ventral human occipitotemporal cortex. *Journal of Neuroscience*, *35*(34), 11921-11935.

Silver, M. A., & Kastner, S. (2009). Topographic maps in human frontal and parietal cortex. *Trends in cognitive sciences*, *13*(11), 488-495.

Sirosh, J., & Miikkulainen, R. (1997). Topographic receptive fields and patterned lateral interaction in a self-organizing model of the primary visual cortex. *Neural Computation*, *9*(3), 577-594.

Srihasam, K., Vincent, J. L., & Livingstone, M. S. (2014). Novel domain formation reveals

proto-architecture in inferotemporal cortex. *Nature neuroscience*, *17*(12), 1776-1783.

Striem-Amit, E., Cohen, L., Dehaene, S., & Amedi, A. (2012). Reading with sounds: sensory substitution selectively activates the visual word form area in the blind. *Neuron*, *76*(3), 640-652.

Striem-Amit, E., Vannuscorps, G., & Caramazza, A. (2017). Sensorimotor-independent development of hands and tools selectivity in the visual cortex. *Proceedings of the National Academy of Sciences*, *114*(18), 4787-4792.

St-Yves, G., Allen, E. J., Wu, Y., Kay, K., & Naselaris, T. (2023). Brain-optimized deep neural network models of human visual areas learn non-hierarchical representations. *Nature communications*, *14*(1), 3329.

Sullivan, J., Mei, M., Perfors, A., Wojcik, E., & Frank, M. C. (2021). SAYCam: A large, longitudinal audiovisual dataset recorded from the infant's perspective. *Open mind*, *5*, 20-29.

Swindale, N. V. (1996). The development of topography in the visual cortex: a review of models. *Network: Computation in neural systems*, *7*(2), 161-247.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual review of neuroscience*, *19*(1), 109-139.

Tang, Y., Gokce, A., Al-Karkari, K. J., Yamins, D., & Schrimpf, M. (2025). Diverse perceptual representations across visual pathways emerge from a single objective. *bioRxiv*, 2025-07.

Tarhan, L., & Konkle, T. (2020). Sociality and interaction envelope organize visual action representations. *Nature Communications*, *11*(1), 3002.

Tarigopula, P., Fairhall, S. L., Bavaresco, A., Truong, N., & Hasson, U. (2023). Improved prediction of behavioral and neural similarity spaces using pruned DNNs. *Neural Networks*, *168*, 89-104.

Taylor, J. C., & Downing, P. E. (2011). Division of labor between lateral and ventral extrastriate representations of faces, bodies, and objects. *Journal of Cognitive Neuroscience*, *23*(12), 4122-4137.

Taylor, J. C., Wiggett, A. J., & Downing, P. E. (2007). Functional MRI analysis of body and body part representations in the extrastriate and fusiform body areas. *Journal of neurophysiology*, *98*(3), 1626-1633.

Thorat, S., Proklova, D., & Peelen, M. V. (2019). The nature of the animacy organization in human ventral temporal cortex. *elife*, *8*, e47142.

Truong, N., & Hasson, U. (2025). Improved Robustness and Functional Localization in Topographic CNNs Through Weight Similarity. *arXiv preprint arXiv:2508.00043*.

Truong, N., Pesenti, D., & Hasson, U. (2025). Explaining human comparisons using alignment-importance heatmaps. *Computational Brain & Behavior*, 1-21.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, *311*(5761), 670-674.

Tucciarelli, R., Wurm, M., Baccolo, E., & Lingnau, A. (2019). The representational space of observed actions. *elife*, *8*, e47686.

Ungerleider, L. G. (1982). Two cortical visual systems. *Analysis of visual behavior*, *549*, chapter-18.

van Bree, S., Styrnal, M., & Hebart, M. N. (2025). How Much Variance Does Your Model Explain? A Clarifying Note on the Use of Split-Half Reliability for Computing Noise Ceilings.

van den Hurk, J., Van Baelen, M., & Op de Beeck, H. P. (2017). Development of visual category selectivity in ventral visual cortex does not require visual experience. *Proceedings of the National Academy of Sciences*, *114*(22), E4501-E4510.

van der Laan, L. N., De Ridder, D. T., Viergever, M. A., & Smeets, P. A. (2011). The first taste is always with the eyes: a meta-analysis on the neural correlates of processing visual food cues. Neuroimage, 55(1), 296-303.

van Dyck, L. E., & Gruber, W. R. (2023). Modeling biological face recognition with deep convolutional neural networks. *Journal of cognitive neuroscience*, *35*(10), 1521-1537.

van Dyck, L. E., & Dobs, K. (2025). Category selectivity as a window into behavioral relevance. *Cognitive Neuroscience*, 1-3.

van Dyck, L. E., Hebart, M. N., & Dobs, K. (2025). Multidimensional feature tuning in category-selective areas of human visual cortex. *bioRxiv*, 2025-06.

Vannuscorps, G., F Wurm, M., Striem-Amit, E., & Caramazza, A. (2019). Large-scale organization of the hand action observation network in individuals born without hands. *Cerebral Cortex*, *29*(8), 3434-3444.

Versace, E., Martinho-Truswell, A., Kacelnik, A., & Vallortigara, G. (2018). Priors in animal and artificial intelligence: where does learning begin?. *Trends in cognitive sciences*, *22*(11), 963-965.

Vogels, R. (2022). More than the face: representations of bodies in the inferior temporal cortex. *Annual review of vision science*, *8*(1), 383-405.

Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron*, *56*(2), 366-383.

Weiner, K. S., Golarai, G., Caspers, J., Chuapoco, M. R., Mohlberg, H., Zilles, K., ... & Grill-Spector, K. (2014). The mid-fusiform sulcus: a landmark identifying both cytoarchitectonic and functional divisions of human ventral temporal cortex. *Neuroimage*, *84*, 453-465.

Weiner, K. S., & Grill-Spector, K. (2010). Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *Neuroimage*, *52*(4), 1559-1573.

Weiner, K. S., & Grill-Spector, K. (2011). Not one extrastriate body area: using anatomical landmarks, hMT+, and visual field maps to parcellate limb-selective activations in human lateral occipitotemporal cortex. *Neuroimage*, *56*(4), 2183-2199.

Weiner, K. S., & Grill-Spector, K. (2013). Neural representations of faces and limbs neighbor in human high-level visual cortex: evidence for a new organization principle. *Psychological research*, *77*(1), 74-97.

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual review of psychology*, *69*(1), 105-129.

Wu, W., Wang, X., Wei, T., He, C., & Bi, Y. (2020). Object parsing in the left lateral occipitotemporal cortex: Whole shape, part shape, and graspability.

Neuropsychologia, *138*, 107340.

Wurm, M. F., Caramazza, A., & Lingnau, A. (2017). Action categories in lateral occipitotemporal cortex are organized along sociality and transitivity. *Journal of Neuroscience*, *37*(3), 562-575.

Wurm, M. F., & Caramazza, A. (2022). Two 'what' pathways for action and object recognition. *Trends in cognitive sciences*, *26*(2), 103-116.

Wurm, M. F., & Lingnau, A. (2015). Decoding actions at different levels of abstraction. *Journal of Neuroscience*, *35*(20), 7727-7735.

Yamins, D. L., 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*, *111*(23), 8619-8624.

Yargholi, E., Op de Beeck, H. (2023). Category trumps shape as an organizational principle of object space in the human occipitotemporal cortex. *Journal of Neuroscience*, *43*(16), 2960-2972.

Xu, Y., & Vaziri-Pashkam, M. (2021). Limits to visual representational correspondence between convolutional neural networks and the human brain. *Nature communications*, *12*(1), 2065.

Yue, X., Robert, S., & Ungerleider, L. G. (2020). Curvature processing in human visual cortical areas. *NeuroImage*, *222*, 117295.

Zador, A. M. (2019). A critique of pure learning and what artificial neural networks can learn from animal brains. *Nature communications*, *10*(1), 3770.

Zeman, A. A., Ritchie, J. B., Bracci, S., & Op de Beeck, H. (2020). Orthogonal representations of object shape and category in deep convolutional neural networks and human visual cortex. *Scientific reports*, *10*(1), 2453.

Zhang, Y., Zhou, K., Bao, P., & Liu, J. (2024). A biologically inspired computational model of human ventral temporal cortex. *Neural Networks*, *178*, 106437.

Zhou, D., Fang, Y., Wang, Z., & Xu, R. (2025). TDSNNs: Competitive Topographic Deep Spiking Neural Networks for Visual Cortex Modeling. *arXiv preprint*

arXiv:2508.04270.

Zhuang, C., Yan, S., Nayebi, A., Schrimpf, M., Frank, M. C., DiCarlo, J. J., & Yamins, D. L. (2021). Unsupervised neural network models of the ventral visual stream. *Proceedings of the National Academy of Sciences*, *118*(3), e2014196118.

Additional Material

Chapter 2

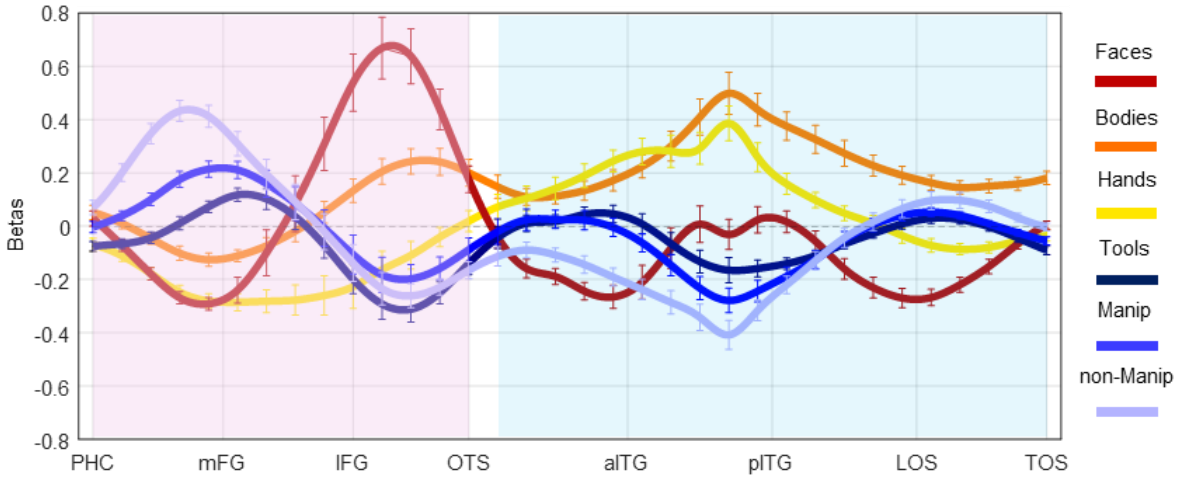


Figure S2.1. Right hemisphere vector-of-ROIs. The same procedure to generate the vector was followed as for the left hemisphere (see methods for details). The spheres along the vector cover an analogous portion of OTC as in the left hemisphere. Normalized activation (against the average of all categories) is plotted for each category as a function of the position of the vector along the cortex. The x-axis corresponds to each sphere along the vector, with labels for major anatomical landmarks; the y-axis corresponds to the normalized beta values. The vector was broadly divided into a ventral component (pink shade) and a lateral component (light blue shade). Contrary to the left hemisphere, no action-related organization can be observed in right lateral OTC. Error bars represent ± 1 SEM across participants (n = 18). Red = faces; orange = bodies; yellow = hands; dark blue = tools; blue = manipulable objects; light blue = non-manipulable objects. Source data are provided as a Source Data file.

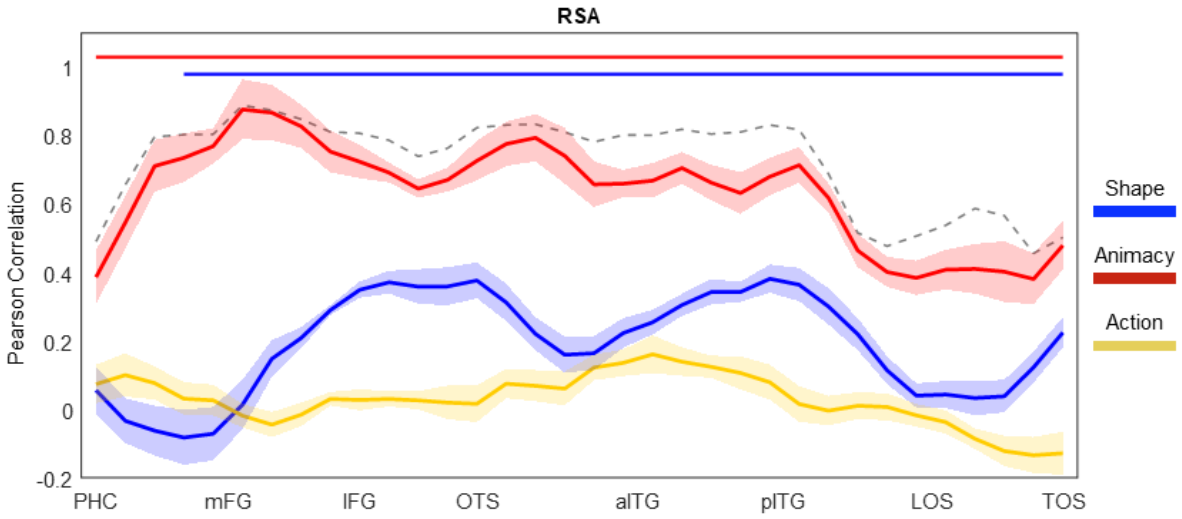


Figure S2.2. Object dimensions in the right hemisphere. Vector-of-ROIs RSA results. The dashed line represents lower bound of the noise ceiling. Two-sided one-sample t-tests were conducted, and horizontal lines indicate statistical significance (vs. baseline) for each model ($p < .0014$ Bonferroni corrected for n = 34 comparisons; blue = shape; red = animacy; yellow = action). The shaded area around the line indicates ± 1 SEM across participants (n = 18). In no sphere of the vector there is a significant effect for the action model, indicating that animacy and – secondarily – shape dominates the object space in the right hemisphere. Source data are provided as a Source Data file.

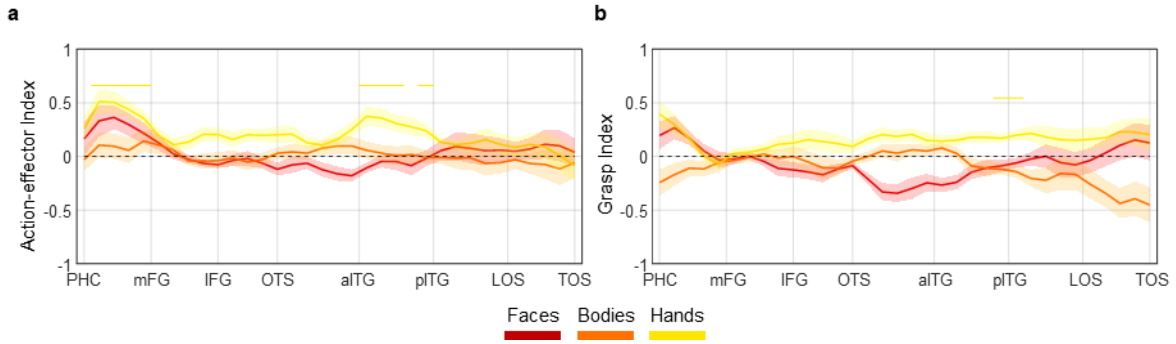


Figure S2.3. Index analysis. a) Vector-of-ROIs action-effector index. b) Vector-of-ROIs grasp index. Color-coded lines at the top of each plot indicate spheres along the vector where each index reached significance, corrected for the number of spheres (n = 34; p = .0015). Red = face indices; orange = body indices; yellow = hand indices. Some effects can be found in right LOTC and VOTC, indicating that, despite the lack of the general action information, hands and tools are moderately correlated with each other also in the right hemisphere. Source data are provided as a Source Data file.

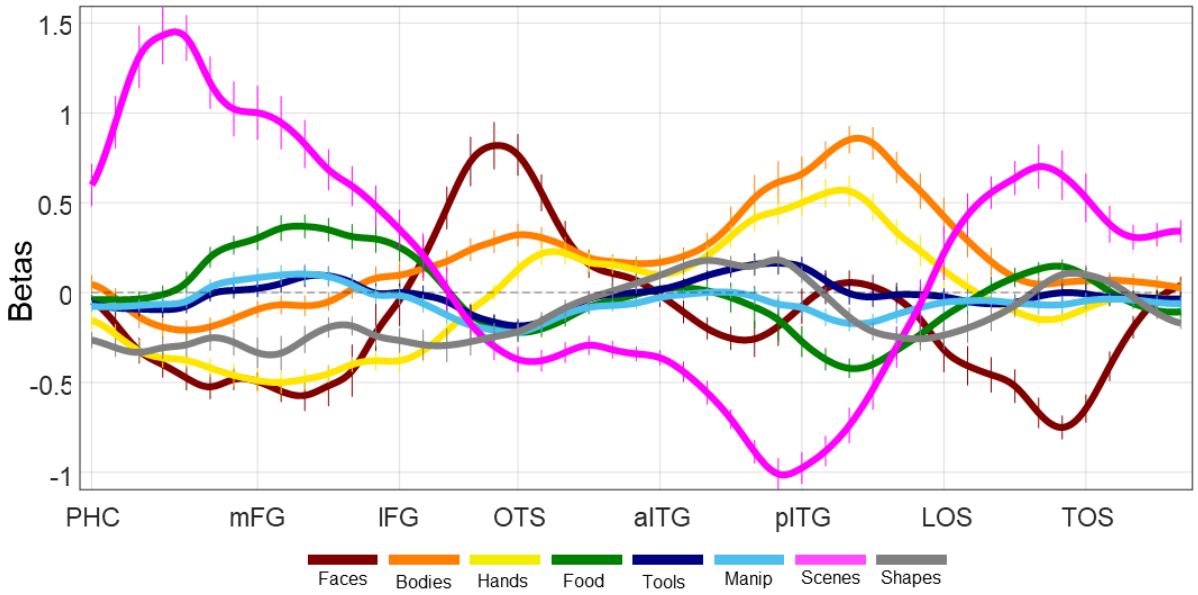
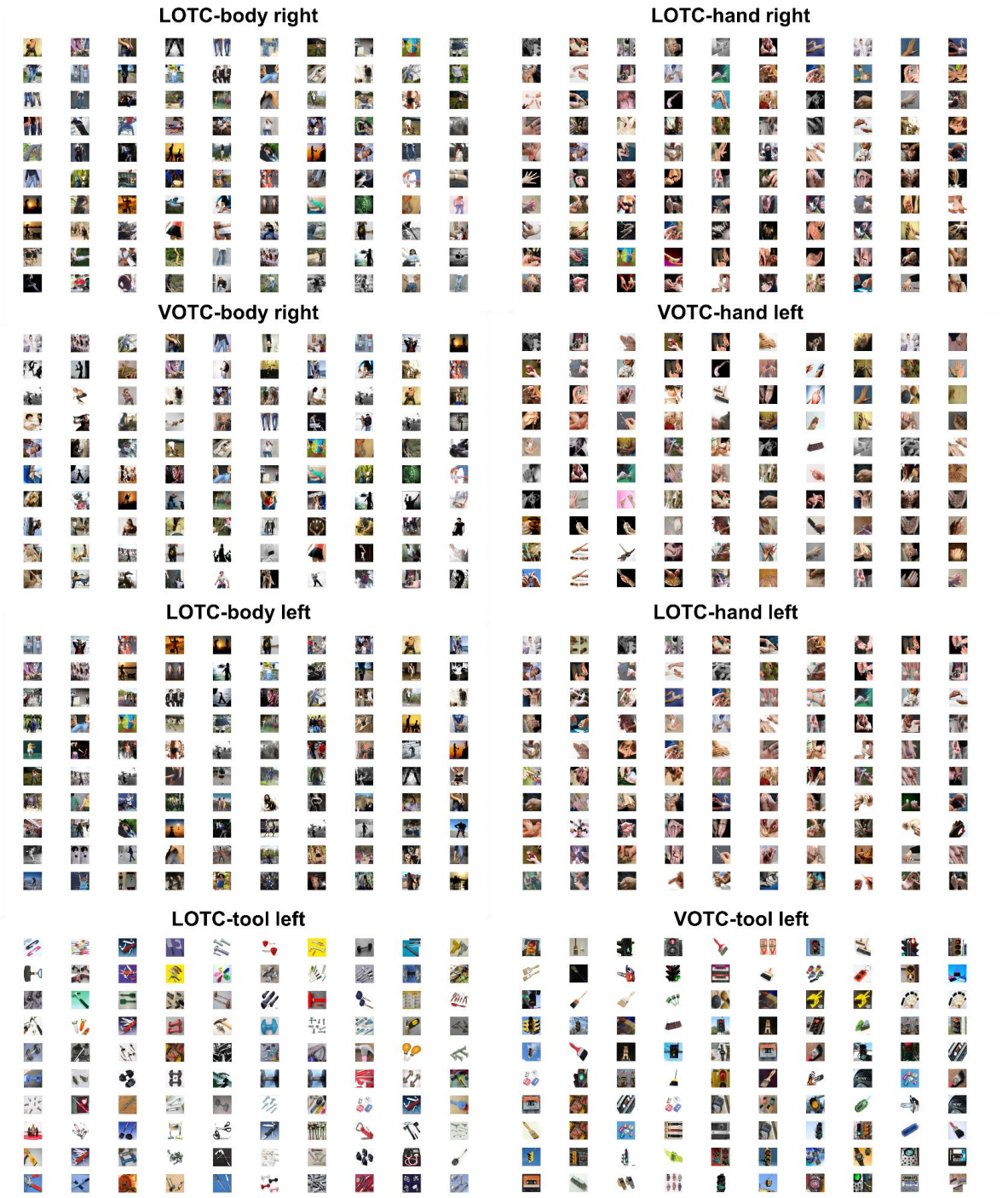


Figure S3.2. Vector-of-ROI analysis for the right hemisphere (Category Experiment). A spline connecting distinct anchor points across ventral and lateral OTC is fitted and partially overlapping spheres (corresponding to the ROIs analysed) along the spline are generated. Functional selectivity is tested for each sphere along the vector (see methods for details). The activation for each category is plotted as a normalized activation against the average of all other categories. The x-axis corresponds to each sphere along the vector, with labels for major anatomical landmarks; the y-axis corresponds to the normalized beta values. Error bars represent ± 1 SEM across participants (n = 18). PHC = Parahippocampal Cortex. mFG = medial Fusiform Gyrus. lFG = lateral Fusiform Gyrus. OTS = Occipitotemporal Sulcus. aITG = anterior Inferior Temporal Gyrus. pITG = posterior Inferior Temporal Gyrus. LOS = Lateral Occipital Sulcus. TOS = Transverse Occipital Sulcus.



Supplementary Figure S4.1. Top-100 most activating images for the body-, hand-, and tool-trained encoding models.