DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Object Size and Its Correlated Properties: Insights Into the Representation of Shape in the Ventral Visual Stream

Citation

Magri, Caterina. 2019. Object Size and Its Correlated Properties: Insights Into the Representation of Shape in the Ventral Visual Stream. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

Permanent link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:42013139

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Object Size and its Correlated Properties: insights into the representation of shape in the ventral visual stream

A DISSERTATION PRESENTED
BY
CATERINA MAGRI
TO
THE DEPARTMENT OF PSYCHOLOGY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
PSYCHOLOGY

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
SEPTEMBER 2019

©2019 – Caterina Magri
all rights reserved.

Thesis advisor: Professor Talia Konkle Caterina Magri

Object Size and its Correlated Properties: insights into the representation of shape in the ventral visual stream

Abstract

High-level visual cortex follows a large-scale organization related to the real-world size of objects, where some regions respond more to pictures of small inanimate objects while other regions respond more to large inanimate objects.

In this thesis, I test the hypothesis that size, a visually inferred property, might be an organizational driver because it is meaningfully correlated with properties that are both more abstract and more visual than size itself in nature. Specifically, I investigate the relationship between the real-world size of an object and its motor-relevance (Chapter 1) or its perceived curvature (Chapter 2). In both studies, I find that the real-world size organization is stronger and more reliable when responding to objects that follow the natural covariance with size (small and manipulable, or small and curvy), compared to objects that do not (big and manipulable, or big and curvy). These observations lend support to the hypothesis that the large-scale organization by real-world size is systematically related to a number of covarying properties. These covarying properties could either be behaviorally relevant (e.g. manipulability) or important features being encoded from earlier stages of visual processing (e.g. curvature). Different possible theories of why this organization arises are discussed.

The thesis shifts its focus on Chapter 3 to the finer-grained similarity spaces of inanimate objects and how these reflected in responses to objects across the ventral stream. Specifically, I examined how neural responses to objects across the ventral stream hierarchy are related to three different behavioral spaces: object similarity measured implicitly by visual search, shape similarity as estimated by an explicit sorting task, and unguided object similarity estimated by the same explicit sorting task but without any instructions. I found that the object similarity space measured by visual search was well correlated with posterior regions of the ventral stream, and after a relatively sharp transition, the shape-based similarity space was well correlated with anterior regions of the ventral stream. Based on these data, I suggest that this relatively sharp transition along the ventral stream hierarchy reflects a transition from more pictorial to more structural representation of inanimate object information.

Overall, this thesis provides new theoretical contributions in the understanding the nature of the object representation along the ventral visual stream, suggesting that visual brain regions are sensitive to object properties at multiple levels of abstraction, related to both perceptual and action-related properties, with a transition from more pictorial to more structural levels of representation.

Contents

0 Introduction 1
- 0.1 Real-world size . . . 2
- 0.2 The role of behaviorally-relevant properties: size and motor-relevance . . . 4
- 0.3 The role of mid-level visual features: size and curvature . . . 6
- 0.4 Exploring the feature space: relating brain and behavior . . . 10

1 Size, manipulability and stability contribute to neural responses to inanimate objects 13
- 1.1 Abstract . . . 13
- 1.2 Introduction . . . 15
- 1.3 Results . . . 18
- 1.4 Discussion . . . 33
- 1.5 Methods . . . 37

2 Behavioral and Neural Associations between Object Size and Curvature 49
- 2.1 Abstract . . . 49
- 2.2 Introduction . . . 51
- 2.3 Results . . . 55
- 2.4 Discussion . . . 65
- 2.5 Conclusion . . . 68
- 2.6 Methods . . . 69

3 Comparing facets of behavioral object representation with brains and models 79
- 3.1 Abstract . . . 79
- 3.2 Introduction . . . 81
- 3.3 Results . . . 84
- 3.4 Discussion . . . 94
- 3.5 Conclusion . . . 97
- 3.6 Methods . . . 98

4 Conclusion 108
4.1 What drives the organization in high-level visual cortex? 109
4.2 Representational changes along the ventral stream 112
4.3 Conclusions 113

Appendix A Appendix to Chapter 1 115
A.1 Supplementary ROI Analyses 116
A.2 Supplementary Searchlight Analyses 118

Appendix B Appendix to Chapter 2 120

Appendix C Appendix to Chapter 3 124
C.1 Supplementary Methods – Predictive Modeling of Brain and Behavior 125
C.2 Supplementary Results – Predictive Modeling of Brain and Behavior 128

References 141

This thesis is dedicated to my parents, Tito and Marina, and my sister, Chicca.

Acknowledgments

THANK YOU to my advisor Talia Konkle, who has pushed me intellectually and motivated me to always give my best, all the while being sympathetic and supportive. I will always look up to her intelligence, brilliancy and positive energy.

Thank you to George Alvarez, Susan Carey and Marge Livingstone, the committee members for this thesis.

Thank you to Alfonso Caramazza and Angelika Lingnau, who have had important roles in shaping me as a scientist.

Thank you to Bria Long, I'm lucky to know and work with such a smart and resourceful person.

Thank you to Stella Shuchen and Rocco Chiou, whom have been essential collaborators for one of the projects in this thesis.

Thank you to the Konkle Lab: Daniel Janini, Leyla Tarhan, Emilie Josephs, Jeongho Park, John-Mark Taylor, Ruosi Wang, Jacob Prince and Arturo Deza, for all the thoughtful comments and assistance.

Thank you to all members of the Vision Lab, their insights and ideas have had an essential role for all the work presented here.

Thank you to Katie Porter, Anna Leshinskaya, Stefano Anzellotti, Ella Striem-Amit, Costi Rezlescu and Gilles Vannuscorps who have been of support and provided me with useful feedback in my early years as a grad student.

Thank you to James Bursley, for having always provided intellectual and emotional support.

Thank you to fellow grad students Monica Burns, Helen Jing, Katherine Berkovitz, Jason Nemirow. Grad school was made so much better thanks to you.

Thank you to Angela Volorio and Alice Desbuleux, with whom I shared the joys and miseries of the dissertation writing process.

Thank you to the members of the Harvard Mountaineering Club, and to the friends I share my love for climbing and the outdoors with. Outdoor activities have been pivotal for finding intellectual clarity in this journey; and they gave me the chance to connect with some truly great people.

And thank you to my parents and my sister, who have always been there for me through all these years, no matter the ocean between us.

0

Introduction

Consider the processing stages needed to recognize an object, such as a pen sitting on a desk. Our visual system first extracts from the retinal input simple, low-level features such lines and edges. Then, through a sequence of progressively higher-level stages, more complex features get encoded and represented, such as curvature, elongation, texture. Eventually, it is able to parse the figure from its background, to recognize it as a pen from different points of views, when occluded, and at different

scales.

How does our brain accomplishes such a feat, what kind of properties are being processed and encoded in order to achieve object representation?

Among all properties, real-world size is one of the primary identifiable ones in guiding neural object representation. However, since its first identification, it was recognized that it might not be one independent dimension, in that in the world, real-world size tends to covariate with several other object properties. In the first two chapters, this thesis seeks to investigate the relationship between real-world size and two of these properties: one that is more action-based in nature, manipulability, and one that is more perceptual in nature, curvature.

The third chapter moves towards characterizing the representational similarity space for a high number of inanimate objects. Specifically, I explore how the brain similarity space relates to different behavioral object similarity spaces.

In this Introduction, I will first introduce the notion of real-world size, how our visual system automatically processes it and how it relates to the large-scale neural organization. I will review what is known about the relationship between real-world size and manipulability, to introduce and anticipate the topic of Chapter 1. Then, I will review what is known about the relationship between real-world size and curvature, to introduce and anticipate the topic of Chapter 2. Finally, I'll review existing literature on what is known about the correspondence between behavioral and neural similarity spaces, to introduce and anticipate the topic of Chapter 3.

0.1 Real-world size

Several lines of evidence point to the fact that objects' real-world size is a property relevant for object representation (Konkle & Oliva, 2011; Paivio, 1975). When instructed to draw objects, people rely on a canonical visual size representation, drawing small objects smaller than large objects, given a blank

paper of equal size (Konkle & Oliva, 2011). When asked to judge which object is smaller or larger in real life between object pairs presented at different visual sizes, people are slower when the target object is presented at a visual size incongruent with its real-world size (Paivio, 1975). Further, when asked to judge which object is visually smaller or larger between pairs of objects of different real-world sizes, people are slower at responding in incongruent trials, when the visually larger objects if actually smaller in real life (and vice versa), as compared to congruent trials (Konkle & Oliva, 2012a). All this evidence points to the fact that real-world size information, an inferred physical property, is accessed quickly and automatically when thinking about or recognizing a depicted object.

Support for an important role of real-world size in the nature of object representation also comes from neural evidence: real-world size well-explains the large-scale organization of objects in high-level-visual cortex (Konkle & Oliva, 2012b), with stronger responses for big objects over small objects in the ventro-medial occipito-temporal cortex, stronger responses to small objects more laterally, and then a return to a big objects preference more dorsally. Multivariate patterns in category-preferring regions throughout the occipito-temporal cortex have been observed to be more similar for objects close in real-world size, further suggesting that it could have a role in accounting for the underlying neural representations of object categories in higher-order visual areas (Julian et al., 2016). Thus, size seems to have an important role for object representation, both influencing behavioral judgments of object recognition, and accounting for the structure of object-selective cortex.

However, while all these observations show that real-world size is a good descriptor of neural responses in occipito-temporal cortex, they do not answer questions about the *nature* of the underlying representations (Konkle & Oliva, 2012b). The observed behavioral and neural responses might be due to the fact that size, in the real-world, correlates with a number of properties, both visual and more abstract. Support for the hypothesis that real-world size might not be the underlying property that is being encoded comes from the fact that these regions are not driven by whether an object is conceived as big or small in the world through an imagery task (i.e. imagining a giant apple still activates small ob-

ject regions, Konkle & Oliva (2012b)). This suggests that there are underlying features covarying with object size that are being represented and encoded by our visual system. What might these features be?

There are two, non-mutually exclusive hypotheses about the way in which our visual system comes to encode real-world size as a relevant property (Konkle & Oliva, 2012b): the first relates to long-range connections with behaviorally-relevant networks, the second to downstream biases from earlier stages of the visual system driven by physical and gravitational constraints of the real-world. I will address each one separately in the next two sections.

0.2 The role of behaviorally-relevant properties: size and motor-relevance

One hypothesis for the emergence of a real-world size organization posits that it might serve as a proxy for behaviorally-relevant properties that distinguish small and large objects. There is a clear way in which small and big objects usually serve different purposes: smaller objects tend to be manipulated and handheld, while big objects tend to be used as landmarks for navigation and for support (Mahon et al., 2007; Mahon & Caramazza, 2011; Julian et al., 2016; Konkle & Oliva, 2012b). Thus, it is possible that one such behaviorally relevant property might be manipulability.

There is strong neural evidence for a preferred encoding of tools and manipulable objects in the ventral stream. For example, tool selectivity has been observed in the lateral aspect of the occipito-temporal cortex: for example, in this territory, there is a region with preference for tools (Chao et al., 1999; Bracci et al., 2012) which is closely overlapping with effector-specific regions (Bracci et al., 2012), and persists in the congenitally blind (Peelen et al., 2013; Wang et al., 2015). This cortex generally shows connectivity with fronto-parietal networks supporting actions (Simmons & Martin, 2012; Bi et al., 2015; Konkle & Caramazza, 2016), and is causally involved in tool-action discrimination judgments (Perini et al., 2014). Tool selective regions do not seem to store information about object use: apraxic

patients, who are impaired in their knowledge about how to use an object, as still able to name them (Johnson-Frey et al., 2005; Caramazza & Mahon, 2003). However, there is also some evidence that the connectivity between tool-selective regions and the parietal lobule, known to be involved in action planning and execution, modulates performance in object recognition (Mahon et al., 2007). Overall, these studies suggest that manipulability might be an important dimension underlying the representation of objects in small-object regions.

On the other hand, work focused on scene understanding has characterized object responses of the ventromedial occipito-temporal cortex, an area known to also show a preference for large objects over small objects (Konkle & Oliva, 2012b), along a series of other properties. Among the properties that have been identified are the invocation of a local space, the tendency to stay fixed in the world, and the usefulness as a landmark (Troiani et al., 2012; Mullally & Maguire, 2011; Julian et al., 2016; Auger & Maguire, 2013; Epstein, 2014). The common theme of all these properties is that they tend to relate to navigability, rather than action. Being this area also involved in scene processing, it is reasonable to think that its coding of real-world size could be relevant for spatial navigation (Konkle & Oliva, 2012b; Julian et al., 2016). Relatedly, there are complementary inanimate, large-object preferring regions along the dorso-medial aspect of visual object cortex (alternatively labeled "OPA" or "TOS" in different literature; Konkle & Oliva (2012b); Kamps et al. (2016)), whose responses are less well understood.

Given the relevance of action-related and navigation-related properties in regions largely overlapping with small- and big-object regions, it is possible that networks primarily involved in these behaviors might shape the tuning in occipitotemporal cortex, and real-world size may serve as a proxy to guide the sort of actions and interactions that are involved in these behaviors. In this account, the previously observed real-world size organization may be more directly related to an underlying organization by manipulability, if most small objects tend to be manipulable and most big object tend to be non-manipulable.

To test this, in Chapter 1 I am going to look at the degree to which motor-relevance, a dimension

related to manipulability but extendable to other effectors beyond the hands, might drive a large-scale organization in the visual cortex independently from real-world size by employing a 2 x 2 stimulus set of small motor-relevant, small non-motor-relevant, big motor-relevant and big non-motor-relevant objects.

0.3 The role of mid-level visual features: size and curvature

An alternative hypothesis for the emergence of a real-world size organization is that it stems from earlier perceptual influences more closely related to the input. That is, differences in the way small and big objects look would lead to systematic biases in the visual input affecting visual processing at a later stage (Konkle & Oliva, 2012b).

Because of the gravitational and physical constraints of the world, and because of the way we interact with them, small and big objects also differ in the way they look: small objects tend to be curvier and more elongated; on the other hand, big objects tend to be boxier in shape, with straight corners (Konkle & Oliva, 2012b; Long et al., 2016). Further, objects' real-world size constrains the way in which we can experience and look at objects: small objects tend to be seen at a closer distance and they tend to occupy the center of the visual system; while big objects tend to be seen from farther away and to occupy more of the periphery of the visual field (Konkle & Oliva, 2011; Cate et al., 2011). Thus, there are multiple visual features correlating with size. Here, we're focusing on one of them, the mid-level visual feature of curvature.

Behaviorally, curvature has long been recognized has one of the primary features being extracted during intermediate stages of visual processing. For example, Asada & Brady (1986) developed the curvature primal sketch, a model of object contour based on the parameterization of curvature which is stable when presented at different scales. An important role to curvature for object recognition is also given Biederman's recognition-by-components theory (Biederman, 1987). This theory hypothesizes

that the brain possesses representations of simple three-dimensional shapes called geons, hypothesized to possess non-accidental shape properties allowing them to be viewpoint-invariant, among which the straightness versus curvature of the contour is one of the most important. These basic shapes, when combined together, allow for the recognition of objects: an ice cream cone is, for example, an upside down cone with a sphere on top. Thus, curvature has long been considered one of the most important features to compute and implement for the goal of object recognition.

Further behavioral evidence for the importance of mid-level visual features in driving real-world size representations comes from work of our group involving *texforms*, artificial stimuli generated by submitting objects to an algorithm which degrades them so as to preserve some mid-level features, but disrupting their identification (Freeman & Simoncelli, 2011). While identification was disrupted, people were still above chance in judging whether an object was big or small (Long et al., 2016).

The effect of texform information in driving size representation has also been successfully tested using the Familiar-Size Stroop paradigm (Konkle & Oliva, 2012a; Long & Konkle, 2017b), in which participants are presented with pairs of texforms, one produced from an object that is small in the real world, the other produced from an object that is big in the real world. The two texforms are presented at two different visual sizes on the screen, and participants are asked to judge which of the two texforms is visually bigger or smaller. For congruent trials, the visual size of the texform in the screen corresponds to the real-world size of the original object (e.g., the visually small texform is produced from a small object, the visually big texform is produced from a big object); for incongruent trials, the visual size of the texform in the screen does not correspond to the real-world size of the originals (e.g., the visually small texform is produced from a big object, the visually big texform is produced from a small object). Participants were faster at reporting the correct visual size when presented with congruent than incongruent trials (Long & Konkle, 2017b).

It thus seems that some of the features preserved by the texform algorithm are sufficient in order to automatically access information about object size. Among the mid-level features of the original object

that might be preserved in the texform version, curvature information seems to be a likely candidate. First, big objects tend to be rated consistently as boxier, and small objects as curvier, both for original images and for texforms (Long et al., 2016). Second, texforms' perceived real-world size correlates with curvy-to-boxy ratings of curvature, with texforms rated as smaller being also rated as curvier, and texforms rated as bigger being also rated as boxier (Long & Konkle, 2017a). Finally, perceived curvature differences between small and big objects were found to be good predictors of the magnitude of the observed Size-Stroop effect (Long & Konkle, 2017a).

These observations strongly support the hypothesis that curvature might be a key property supporting behavioral judgments related to the size of objects, even before the object is recognized.

0.3.1 Neural evidence for importance of curvature

Aside from affecting behavioral judgments, curvature has a prominent role also in shaping neural responses all along the ventral stream.

Animal studies have extensively proven an important role of curvature in shaping neural responses in the visual cortex of monkeys. Area V4 in the macaque monkey brain has been observed to be more responsive to hyperbolic and polar stimuli, characterized by curvy lines, than to cartesian stimuli characterized by straight lines (Gallant et al., 1996). Many neurons in the area are responsive to specific orientations of curves, show a bias for convex features (Pasupathy & Connor, 1999), and show size invariance with respect to curvature preference, suggesting a possible mechanism through which size invariance might be computed along the visual processing stream (El-Shamayleh & Pasupathy, 2016). Coding for curvature information has been observed also more anteriorly in IT, and it seems to be encoded independently from other properties such as aspect ratio (Kayaert et al., 2005). By employing fMRI, a network of three curvature-selective patches has been observed in the macaque's brain along the ventral visual pathway, one in dorsal V4 (posterior curvature-biased patch), a middle one located ventrally on the posterior superior temporal sulcus (middle curvature-biased patch), and one located

more anteriorly (anterior curvature-biased patch; Yue et al. (2014)).

Further credit to the idea that the feature of curvature is a building block for object processing comes from the exploration of retinotopic maps in higher-order visual cortex: curvature, end-stopping and eccentricity maps tend to correlate in the macaque's visual cortex, with centrally-biased patches corresponding to end-stopped sites and preferring curvy contours, and peripherally biased regions preferring straight stimuli (Ponce et al., 2017; Srihasam et al., 2014). These observations have led to recent proposals suggesting that both eccentricity and curvature reflect a proto-organization of the ventral stream that constrains the formation of stimuli-specific patches which emerge only after training in juvenile monkeys (Srihasam et al., 2014, 2012). These studies point to curvature being an important feature being computed in the ventral stream, not only relatively early in retinotopic areas, but also at later stages in areas thought to process higher order visual information.

Human neuroimaging studies from our group employing the texform stimuli have further explored the degree to which mid-level properties, among which curvature, might underlie the organization by animacy and size (Konkle & Oliva, 2012b; Konkle & Caramazza, 2013). When employing texforms, neural organization by animacy and, within inanimate regions, by size is still preserved in the occipito-temporal cortex (Long et al., 2018), suggesting a strong role of natural statistics with mid-level features in building this large-scale organization by domains. Further, Long and colleagues (Long et al., 2018) collected ratings on a curvy-to-boxy scale for the same stimuli, and found that this intuitive curvature rating performed reasonably well at predicting neural responses in occipito-temporal cortex. These results suggest that curvature information might be underlying the neural division by animacy and size in high-level visual cortex.

All these studies support the hypothesis that curvature might have a pivotal role not only in affecting behavioral judgments, but also in the shaping of neural responses along the ventral stream.

To test this hypothesis, in Chapter 2 I will present a 2 x 2 stimulus set of small curvy, small boxy, big curvy and big boxy objects to two independent sets of participants in a behavioral and a fMRI

task. My question is whether behavioral judgments and neural responses are affected by the natural covariance between size and curvature, or whether these two properties are coded independently.

0.4 Exploring the feature space: relating brain and behavior

While we gain understanding about the underlying shape-space by exploring broad, interpretable domains such as size, manipulability and curvature, an alternative approach is the one of characterizing this space by looking at a high number of objects varying in multiple dimensions in condition-rich designs. Studies in this direction have often tried to characterize the neural feature space by comparing it to similarity spaces produced from behavioral judgments or deep neural networks (de Beeck et al., 2008; Mur et al., 2013; King et al., 2019).

Work from our group has recently looked at the correspondence between visual search behavior and neural responses (Cohen et al., 2017). In visual search, object similarity is indirectly tracked through reaction times, in that people take a longer time locating a target among distractors if target and distractors are more similar (Duncan & Humphreys, 1989). The nature of the similarity tracked by the visual search task is hypothesized to relate most to low- and mid-level features, among which shape is a likely candidate (Wolfe and Horowitz, 2017).

Cohen and colleagues observed surprising results in how well category-level visual search behavior relates to neural responses throughout occipito-temporal cortex, with a strong correspondence between neural responses and behavior both in ventral and dorsal divisions of the ventral stream (Cohen et al., 2017). However, neural responses in this study were recorded for object categories, not at the item level, and participants were instructed at the beginning of each block for the category they were looking for. Thus, they activated a template of the object category during the search. While these results suggest that the nature of the representation in object-selective cortex might reflect more mid-level than high-level representations, the fact that similarities were only recorded at this categor-

ical level did not allow for a more fine-grained characterization of the neural similarity space in the occipito-temporal cortex.

Other studies in the past have looked at neural correlates of both physical and perceived similarity of objects. Indeed, physical and perceived similarity do not always correspond: Biederman identified non-accidental properties of objects, whose differences are thought to be at the basis of shape classification, and should not be affected by rotations on depth. On the other hand, metric properties are those properties of objects which are affected by rotations in depth (1987). Humans tend to be more sensitive to changes in non-accidental properties, than to changes in metric properties (Biederman & Bar, 1999), and this is reflected by the neural tuning of IT cortex in monkeys (Kayaert et al., 2003). That is, there is some evidence that neurons in IT are not tuned so much based on how similar two objects are, but on how similar they are perceived to be.

Relatedly, a human neuroimaging study also looked at how physical and perceived similarities are encoded in the ventral stream (Haushofer et al., 2008). Perceptual differences in this case were tracked with an implicit measure, by asking participants to detect whether two abstract objects presented sequentially and very briefly were identical or not after displaying them for a very short interval (Haushofer et al., 2008). The rate to which two objects were considered identical by participants was considered a measure of their perceptual similarity. When relating this behavioral similarity space to neural activation, a gradient was found, with more posterior regions more closely matching information about physical similarity (as computed through aspect ratio), and more anterior regions more closely matching perceptual similarities (Haushofer et al., 2008). These findings suggest that different parts of the ventral stream might potentially reflect different perceptual features, and that implicit behavioral judgments could potentially relate to neural responses in at least some of these areas.

Another paradigm that has been employed to collect object similarity judgments, this time through explicit judgments, has been the multi-arrangement method, in which participants are asked to arrange objects within a circular arena, so that similar objects are nearby each other in the arena (Kriegesko-

rte & Mur, 2012). It has been observed that, when participants are not given specific instructions as to what kind of similarity to follow, the similarity space that emerges from this method tends to be more categorical than the neural space in IT (Mur et al., 2013). However, if participants are specifically required to arrange objects either by shape or by semantic category (Bracci & Op de Beeck, 2016), it emerges that while a shape-based similarity shape reflects neural responses both in posterior and anterior occipito-temporal cortex, a category-based one is only related to more anterior regions. Overall, these results suggest that the multi-arrangement task can be utilized to collect similarity spaces that are more perceptual if people are instructed to specifically pay attention to this measure of similarity, and that a shape-based space might more closely match neural responses throughout the ventral stream as compared to a category-based one.

While studies reviewed so far have looked at perceptual similarity both with implicit and explicit measures, a direct comparison of how well each of these reflects neural responses at the item-level has not yet been performed.

In Chapter 3, I develop an item-level odd-one-out visual search paradigm that participants perform online in order to obtain similarity measures for 72 inanimate objects. I also compare this space with two different kinds of explicit judgments in a multi-arrangement task: in the shape sorting task, participants are instructed to arrange objects by shape; in the unguided sorting task, a different set of participants was not given any specific instructions on what similarity measure to follow. Given what we know about the multi-arrangement task, we expect the unguided sorting to be more categorical, and the shape sorting to rely more on visual features of objects.

The aim of this work is to examine how well each of these object similarity spaces correlates with OTC responses, and where it correlates best along the ventral stream hierarchy.

1

Size, manipulability and stability contribute to neural responses to inanimate objects

1.1 Abstract

In human occipito-temporal cortex, brain responses to depicted inanimate objects have a large-scale organization by real-world object size. Critically, the size of objects in the world is systematically related

to behaviorally-relevant properties: small objects are often grasped and interacted with (e.g. forks), while large objects tend to be less motor-relevant (e.g. tables), though this relationship does not always have to be true (e.g. picture frames and wheelbarrows). To determine how these two factors interact, we measured brain activity with functional magnetic resonance imaging while participants viewed a stimulus set of small and large objects with either low or high motor-relevance. The results revealed that the size organization was evident for objects with both low and high motor-relevance; and a motor-relevance map was also evident across both big and small objects. Targeted contrasts revealed that typical combination—small motor-relevant vs. large non-motor relevant—yielded more robust topographies than the atypical covariance contrast—small non-motor relevant vs. big motor-relevant. In subsequent exploratory analyses, a factor analyses revealed that the construct of motor-relevance was better explained by two underlying factors: one more related to manipulability, and the other to whether an object moves or is stable. The factor related to manipulability better explained responses in lateral small-object preferring regions, while the factor related to object stability (lack of movement) better explained responses in ventromedial large-object preferring regions. Taken together, these results reveal that the structure of neural responses to objects of different sizes also further reflects action-relevant properties of manipulability and stability.

1.2 Introduction

To recognize objects in the real world, a large portion of the brain is utilized—the ventral visual stream (Goodale & Milner, 1992; Ishai et al., 1999) . Within this cortex there is a large-scale organization of responses distinguishing large and small inanimate objects (Konkle & Oliva, 2012b; Konkle & Caramazza, 2013). Some proposals have raised the possibility that the large-scale divisions in this cortex arise to support different behavioral needs (e.g. manipulating objects vs navigating in environments), linked to distinct underlying long-range brain networks (Mahon & Caramazza, 2011; Konkle & Caramazza, 2016). Thus, in the present work we examined how both the size of an object and its motor relevance can explain brain responses to object categories.

Intuitively, the relationship between an object's size and the degree of interaction it affords in the world is evident: small objects tend to be graspable and easy to hold up, while larger objects tend to be used as support or for navigation. However, other relationships are possible, including non-moving small objects like picture frames, and interactive big objects like pianos. Here, we operationalize the dimension dealing with object interaction as motor-relevance. While similar conceptually to the property of manipulability that has been investigated in the past (Mahon et al., 2007), motor-relevance spans a wider definition in that it includes actions performed not only with the hands, but with other parts of the body as well. With this broader conceptualization, size can more easily dissociate from degree of interaction: large objects like swing sets and wheel barrows tend to involve movements beyond hand-performed actions, while for small objects such as picture frames and smoke alarms there is no clear motor interaction associated.

To date, there is substantial evidence that motor-relevance is an important construct for visual object responses, particularly along the lateral aspect of the occipito-temporal cortex, a region known to show a preference for small inanimate over large inanimate object (Konkle & Oliva, 2012b; Konkle & Caramazza, 2013). For example, in this territory, there is a region with preference for tools (Chao

et al., 1999; Bracci et al., 2012) which is closely overlapping with effector-specific regions (Bracci et al., 2012), and persists in the congenitally blind (Peelen et al., 2013; Wang et al., 2015). This cortex generally shows connectivity with fronto-parietal networks supporting actions (Simmons & Martin, 2012; Bi et al., 2015; Konkle & Caramazza, 2016), and is causally involved in tool-action discrimination judgments (Perini et al., 2014). While these studies have typically focused on objects that require hand-performed actions, the present study allows us to test whether the inclusion of non-hand-performed actions through the motor-relevance dimension would further explain neural responses and extend them to large objects.

In the ventromedial temporal cortex, another tool-preferring region has been observed by some studies when contrasting tools to animals in the medial fusiform gyrus (Chao et al., 1999; Whatmough et al., 2002; Mahon et al., 2007; Garcea & Mahon, 2014). However, the property of motor-relevance in this region is less clear: for example, responses to manipulable tools are not higher than to other inanimate objects (Mahon et al., 2007; Chen et al., 2018), though there is still some evidence for sensitivity for tools in neural adaptation signals (Mahon et al., 2007). Instead, research focused on scene understanding have characterized these ventral-medial object responses along other kinds of object properties that are less related to motor-relevance. For example, responses in ventromedial scene-related regions are best predicted by objects invoking a local space that tend to stay fixed in the world, and are useful as a landmarks (Troiani et al., 2012; Mullally & Maguire, 2011; Julian et al., 2016; Auger & Maguire, 2013; Epstein, 2014). These same areas are also known to show a preference for large over small inanimate objects (Konkle & Oliva, 2012b; Konkle & Caramazza, 2013).

In the present study, we sought to explore the degree to which the large-scale organization by real world size is affected by the real-world natural covariance between size and motor-relevance. To do so, we built a 2 x 2 Dataset of small motor-relevant, small non-motor-relevant, big motor-relevant and big non-motor-relevant objects (432 total images, 18 categories per condition, 6 exemplars per category; Figure 1.1a). We presented these objects to participants undergoing functional neuroimaging,

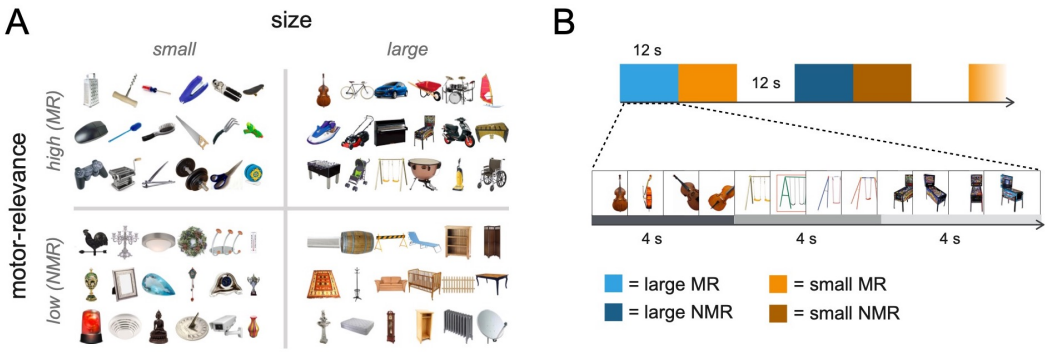


Figure 1.1: Stimuli and fMRI Design. a): Stimulus set: 72 objects were divided into four conditions (small motor-relevant, small non-motor-relevant, large motor-relevant, large non-motor-relevant). Each object category included six different pictures of the same object, for a total of 432 individual images. b) fMRI task: subjects were presented with four consecutive pictures of the same object category in a 4s object category mini-block. Three category mini-blocks were included in a 12s main block, corresponding to one of the 4 main conditions (e.g. large motor relevant objects). The task for participants was to detect when a red frame surrounded one of the objects, which happened once per 12s block.

and examined the degree to which the occipitotemporal cortex jointly or independently coded these two dimension. The fMRI experiment consisted of a mini-block design (Figure 1.1b, see Methods) which enabled us to analyze the data at two different levels of granularity. First, we compared the broad relationship among object size and motor-relevance properties along a 2x2 design (big vs small, motor-relevant vs non-motor relevant; Figure 1.1a). Second, we examined brain responses at the level of object categories (72 categories total). This allowed us to more deeply explore the role of motor-relevance and other related factors in describing brain responses.

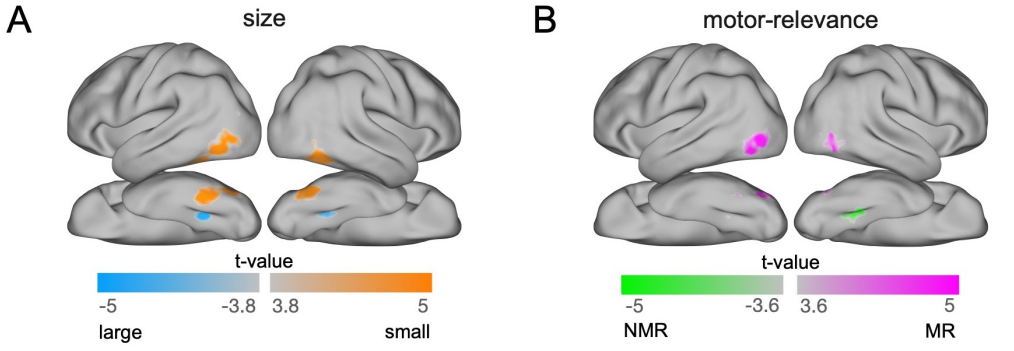
1.3 Results

1.3.1 Whole-brain conjunction analysis

We first examined whether the large-scale neural organization by real-world size of inanimate objects is preserved when disrupting the natural covariance between size and motor-relevance. To do so, we conducted a conjunction analysis to isolate regions which show a large vs small object difference that holds when comparing motor-relevant objects as well as non-motor-relevant objects (see Methods).

When looking at the real-world size organization, we found regions that showed a preference for small over large objects, and for large over small objects, regardless of the motor-relevance preference. The whole brain maps can be visualized in Figure 1.2a, with the group MNI coordinates at the bottom of the figure. With respect to small objects, we observed activation bilaterally in ITG and in the left hemisphere and along the lateral aspect of the fusiform gyrus. In the left hemisphere, the region with a preference for small objects expanded posteriorly to the lateral occipital cortex. With respect to large objects, along the ventral surface of the occipito-temporal cortex, we observed bilateral regions corresponding to the parahippocampal cortex. A preference for large objects was also present more dorsally in the lateral surface in the transverse occipital sulcus (TOS) in individual participants, but this region did not survive the group-level conjunction test at the FDR threshold depicted in the Figure 1.2a. Overall, this group-level conjunction analysis reveals that, even when taking into account motor-relevance, the real-world size of objects drives large-scale differential responses across the ventral stream.

Next, we examined whether also motor-relevance, when disrupting its natural covariance with size, would elicit a large-scale organization across this cortex. Another whole-brain conjunction analysis was conducted, comparing motor-relevant objects vs non-motor-relevant objects, requiring this relationship to hold for both large object and small objects independently. The results are mapped in Figure 1.2b, with the group MNI coordinates at the bottom of the figure. Along the lateral surface,



Conjunction Contrasts	x	y	z	Peak t-value
small: *small MR > large MR & small NMR > large NMR*				
L ITG	-40	-66	-8	8.28
L Fus	-42	-44	-18	7.46
L LO	-40	-88	-6	4.26
R ITG	46	-56	-10	7.41
large: *large MR > small MR & large NMR > small NMR*				
L PHC	-28	-50	-2	4.64
R PHC	32	-46	-6	-5.28

Conjunction Contrasts	x	y	z	Peak t-value
MR: *small MR > small NMR & large MR > large NMR*				
L LO	-46	-82	0	7.64
R LO	42	-70	2	5.65
NMR: *small NMR > small MR & large NMR > large MR*				
L PHC	-26	-44	-14	-5.80
R PHC	28	-36	-18	-5.71

Figure 1.2: Whole Brain Random-Effects Conjunction Analysis. a) Size conjunction contrasts are shown in which both small motor-relevant > large motor-relevant and small non-motor-relevant > large non-motor-relevant relationships hold. a) Motor-relevance conjunction contrasts are shown in which both small motor-relevant > small non-motor-relevant and large motor-relevant > large non-motor-relevant relationships hold. Contrasts were thresholded at q(FDR)<0.05.

we observed regions with a stronger response to motor-relevant objects, with peaks in LO bilaterally. Along the medial aspect of the ventral surface, we found stronger responses to non-motor-relevant objects primarily in the right hemisphere, with the peak activation located slightly anterior to the large-object preferring region. This analysis reveals that the dimension of motor-relevance also elicits stable response differences in parts of object-responsive cortex.

Notably, these conjunction maps yielded relatively similar organizations. Specifically, lateral regions prefer both small and motor-relevant objects; while ventromedial regions prefer both large and non-motor-relevant objects. These findings indicate that neither size alone nor motor-relevance alone are sufficient to explain the large-scale organization of responses to objects along occipito-temporal cortex. Further, these results highlight that there is a clear association between the two dimensions and how they map across the cortex.

1.3.2 Typical-atypical Contrasts

To further explore the relationship between size and motor-relevance, we compared maps between two contrasts: a typical contrast (small motor-relevant > large non-motor-relevant), which follows the real-world covariation of small objects being more motor-relevant than large objects; and an atypical contrast (small non-motor-relevant > large motor-relevant; Figure 1.3a). These maps are shown in Figure 1.3b and c.

When looking at surface maps produced by the typical contrast (Figure 1.3b), we find extensive activations along the lateral and ventral surfaces. On the other hand, the surface maps produced by the atypical contrast (Figure 1.3c) are less extensive and also less reliable in a split-half analysis (typical split-half map correlations: M = 0.61, SD = 0.18; atypical split-half map correlations: M = 0.43, SD = 0.24; t_{17} = -2.70, $p < 0.05$).

These topographic observations mirror the ecological covariation of these factors (Figure 1.3a): while it is theoretically possible to artificially dissociate size and motor-relevance as we have done in

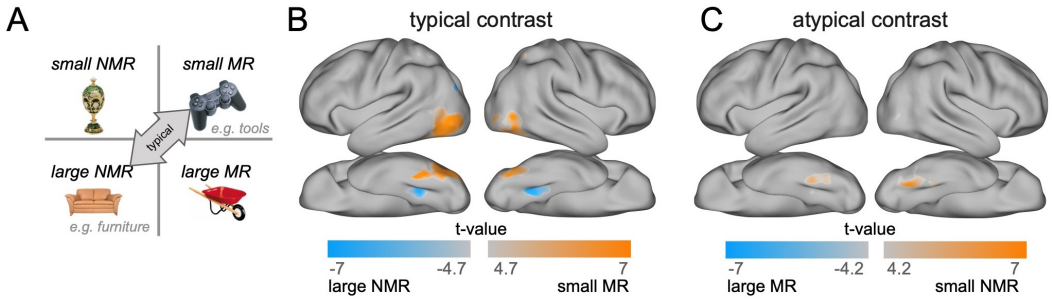


Figure 1.3: Typical/Atypical contrast Analysis. a) simplified representation of the size x motor-relevance design. Arrow points to the typical contrast conditions, representing the covariation commonly observed in the environment. b) Typical contrast small motor-relevant > large non-motor-relevant. c) Atypical contrast small non-motor-relevant > large motor-relevant. Contrasts were thresholded at q(FDR)<0.05.

this stimulus set, in every day experience small objects tend to be more motor-relevant and big objects less so. And we found this covariation to also be evident in the neural data (see Figure 1.3b,c): taken together, these data show the importance of the interaction between size and motor-relevance in driving more reliable and robust neural responses, thus revealing the importance of the natural covariation between these two properties in the visual system.

1.3.3 Exploratory Analyses

Factor Analysis

We next explored the possibility that the relationship between size and motor-relevance may be more multifaceted than what can be gathered from a simple 2 x 2 design. Our goal for this analysis was to understand and explore how other, related properties that have been proposed in the past relate to our findings and predict neural responses at the category level. To do so, we collected ratings along five additional dimensions of our 72 object categories and submitted them to a factor analysis together with size and motor-relevance (see Figure 1.4a, left of the table).

First, manipulability ratings were obtained, which focused on hand movements specifically, in or-

A

Dimension	Question	Pairwise Correlations							Factor Analysis			
		size	MR	man	IntEnv	SD	PV	MBI	Var	f_{manip} 34%	f_{size} 25%	f_{stab} 24%
size	How big is this object in every day life?	0.99	0.03	-0.09	0.35	0.72	-0.16	0.03	0.12	**0.95**	-0.15	
motor-relevance	Interactions require moving hands/body?	0.03	0.97	**0.92**	0.76	0.06	0.66	0.73	**0.81**	0.02	**0.53**	
manipulability	Particular hand movements involved?	-0.09	**0.92**	0.97	0.67	-0.18	0.67	0.66	**0.91**	-0.15	0.37	
interaction envelope	How many hands are used?	0.35	0.76	0.67	0.97	0.24	0.47	0.55	**0.77**	0.31	0.29	
spatial definition	Is a background evoked?	0.72	0.06	-0.18	0.24	0.82	-0.06	0.15	-0.04	**0.83**	0.11	
position variability	Change position in the environment?	-0.16	0.67	0.67	0.47	-0.06	0.97	0.69	0.38	-0.11	**0.70**	
motion-based identity	Motion important in identifying object?	0.03	0.73	0.66	0.55	0.15	0.69	0.92	0.39	0.09	**0.82**	

B

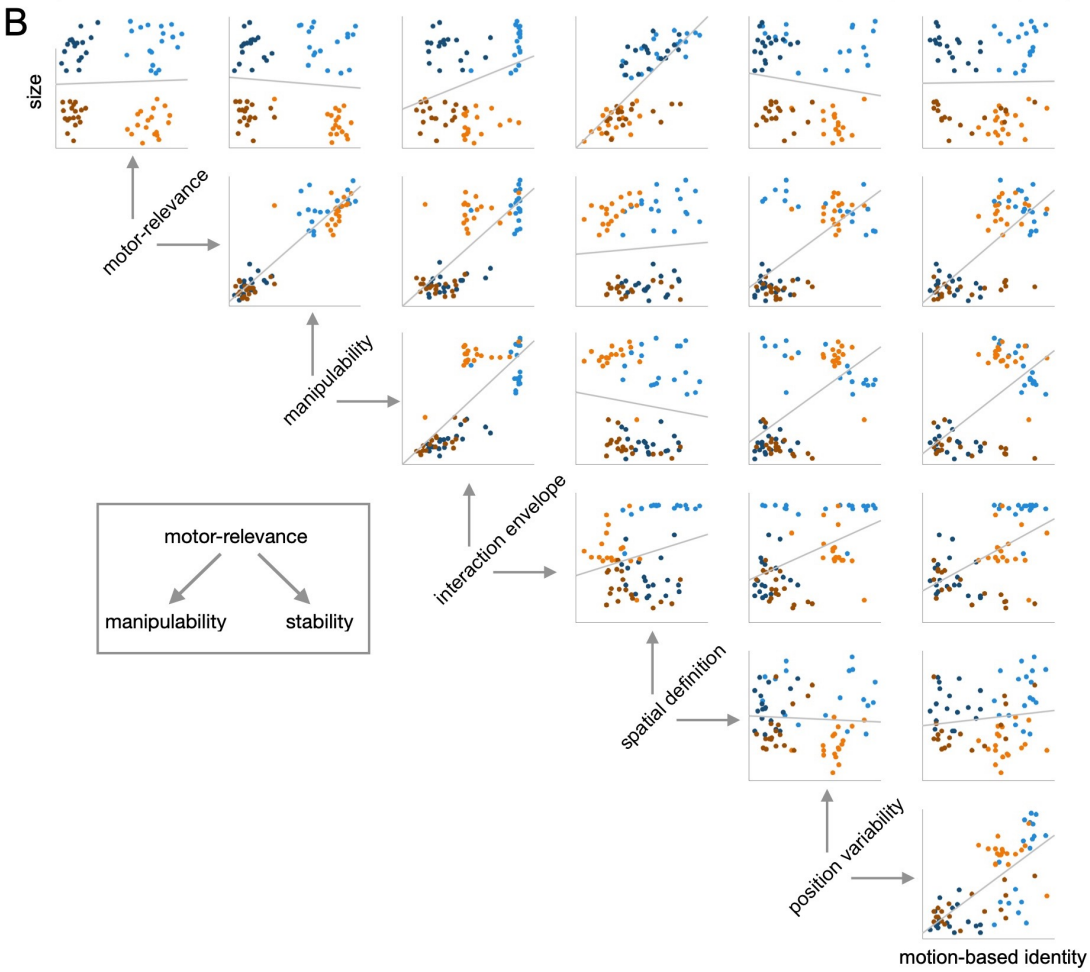


Figure 1.4: Object dimension similarity and Factor Analysis. a) For each rated object property, the table shows the dimension label, the question asked to elicit ratings on the dimensions, the correlation to the other dimensions, where the ratings reliability is shaded, and the dimension's loading in a factor analysis. b) Scatter plots are shown for all pairs of 7 dimensions, where each dot reflects an object category and is color-coded based on its size and motor-relevance (light orange – small motor relevant; dark orange: small non-motor relevant; light blue: large motor relevant; dark blue: large non-motor relevant).

der to understand the extent to which our motor-relevance dimension correlated with this previously studied dimension (Mahon et al., 2007). We also examined a number of dimensions that have been previously proposed as significant modulators of tool regions or scene regions, namely interaction envelope (Bainbridge & Oliva, 2015), spatial definition (Mullally & Maguire, 2011), and motion-based identity (Beauchamp, 2005). Finally, based on our observation regarding low responses to vehicles in Large-PHC, we obtained "position variability" ratings, capturing how often an object changes its position in the environment. Note that these properties were defined after having seen the data, and thus any role they play in explaining neural data requires independent follow-up studies to confirm.

The relationship among these ratings and our original size and motor-relevant dimensions are detailed in Figure 1.4. All ratings had a relatively high reliability (Figure 1.4a, shaded diagonal numbers), with a range of similarity relationships between any pair of dimensions (Figure 1.4b).

There are a couple observations to note from inspecting these pairwise correlations scatter plots. First, by design, size and motor-relevance were completely uncorrelated. Second, the correlation between motor-relevance and manipulability was at ceiling, virtually identifying the same property. As mentioned in the Introduction, most motor-relevant objects involved hands for their interaction, thus including whole-body actions in the motor-relevance score had a minimal effect (the only exception being the skateboard, which was high in motor-relevance but low on manipulability).

Next, we extracted the latent dimensions within these ratings using factor analysis (see Methods). A parallel analysis procedure determined that three factors should be retained, which were sufficient to explain more than 80% of all variance among the ratings. The property loadings for each factor are shown to the right in Figure 1.4a.

The first factor was loaded mostly by manipulability, motor-relevance and interaction envelope—it seemed thus to be most related to actions and hand-object interactions. The second factor was loaded by size and spatial definition, and was thus most related to the physical presence of the object. Finally, motion-based identity, position variability and motor-relevance loaded strongly on the third factor,

which seemed thus to capture the mobility/movability of the object.

Considering these factors within the context of our main design, our motor-relevance dimension was effectively split into two factors, one related to hand-object interactions (f_{manip}) and one related to object's mobility/movability ($f_{stability}$), and both of these were separate from the more physical properties of size and spatial definition (f_{size}). Thus, while our stimulus set was collected as varying along only two primary factors (size and motor-relevance), the factor analysis indicated that this stimulus set is better characterized along three factors (size, manipulability and stability).

We next explored how well the three factors extracted from our ratings can explain the neural response profiles, both in a targeted ROI analysis and in a broader searchlight analysis. Note that these analyses were unplanned and should be considered post hoc and exploratory in nature.

ROI Analysis

To explore how well the three factors could explain fMRI responses, we employed a linear modeling analysis explaining the activation profile in individually defined small-object preferring inferior temporal gyrus (Small-ITG, from the contrast [Small > Large], see Figure 1.5a top) and large-object preferring parahippocampal cortex (Large-PHC, from the contrast [Large > Small], see Figure 1.5a bottom) with the three factors obtained from the factor analysis and their interactions as regressors (see Methods). Note that the ROIs were produced by utilizing two runs of the main experiment, and the average beta activation from the remaining runs was used for this analysis.

Our goal was to see which weighted combination of factors would best associated with each region's response variation to the 72 objects. This finer-grained analysis looking at responses to all 72 object categories was possible due to our nested fMRI protocol design (see Methods), in which the neural responses could be modeled at the category-level for each of the 72 categories independently (see Figure 1.1b). Note that these three factors were determined only from behavioral ratings of our stimulus set. Further, to perform this and the following searchlight analysis, we flip $f_{stability}$ to correlate it with

degree of fixedness, rather than with degree of mobility, for easier interpretation of the results. This transformation does not affect the significance of the statistical results.

First, we determined the appropriate model complexity by evaluating the adjusted R^2 ($adjR^2$) of three nested models: a model with the three factors as main effects, a second model with all two-way interactions included, and a third model with also the three-way interaction included (see Methods). The model with the highest $adjR^2$ was considered to be the one best fitting the region, and the significant coefficients were explored.

From this analysis, we find that both Small-ITG and Large-PHC were best explained by a model with all three factors and their two-way interactions (Small-ITG: $adjR^2 = 0.33$; $F_{6,65} = 6.90$, $p < 0.001$; Large-PHC: $adjR^2 = 0.27$; $F_{6,65} = 5.36$, $p < 0.001$), as compared to the model with only main effects (Small-ITG: $adjR^2 = 0.23$; Large-PHC: $adjR^2 = 0.21$) and the model with main effects, two- and three-way interactions (Small-ITG: $adjR^2 = 0.32$; Large-PHC: adj. $R^2 = 0.26$).

Next, we examined which factors had significant coefficients, and whether they were positive or negative, to understand the models' relationship to the data (Figure 1.5b).

In Small-ITG, the results indicate that size alone, and size in combination with the manipulability factor best predicted neural response magnitude: there was a significant effect of f_{size}, as well as the interaction of f_{size} and f_{manip} (see Figure 1.5b top). The interaction between these factors further indicates that it is specifically small objects involved in hand-object actions which drive the strongest responses in this region. Thus, the distinction of hand-object action / manipulability is a better descriptor of the responses over the more general concept of motor-relevance.

In Large-PHC, we observed a different pattern: the coefficients that significantly explained the model were f_{size} and the interaction of f_{size} and $f_{stability}$ (see Figure 1.5b bottom). Thus, responses in this region were driven most by big objects that are stable in the environment. This result in Large-PHC clarifies what observed in earlier analyses employing the 2 x 2 design: rather than the low-action component of motor-relevance, it might be the fixedness component of this construct the one most

driving representation in this area. That is, the stability component of non-motor-relevance is the one to best predict response magnitude in Large-PHC.

To better understand these relationships, we visualize them in Figure 1.5c: for each ROI, the activation to each category is plotted on the y-axis as a function of the manipulability factor for Small-ITG, and the stability factor for Large-PHC. Each dot is color-coded based on the f_{size} value for the corresponding object. This plot reveals that the manipulability factor drives activation in Small-ITG but mostly for small objects (increasing activity for more manipulable small objects, see Figure 1.5c top). In contrast, inspection of the scatterplot for Large-PHC suggests that activation is driven by stability but mostly for big objects (increasing activity for more stable large objects, see Figure 1.5c bottom). These relationships are also evident in the activation profiles of the two ROIs in Figure 1.6: For Small-ITG, tools are the small objects that drive the most activation (e.g. nail clipper, hand rack, yo-yo; see Figure 1.6a), while for Large-PHC, means of transport tend to be the big objects that drive the least activation (e.g. bike, scooter, jetski; see Figure 1.6b).

Taken together, these analyses begin to characterize the nature of the ecological covariation between size and motor-relevance, clarifying how these object properties fit together. For example, no one factor alone (size, manipulability, stability) best accounted for the responses in any ROI. Rather, the manipulability (or hand-relevance) of an object matters more for small objects in the lateral cortex, and the stability of an object matters more for big objects in the ventral cortex. Finally, these findings are consistent with prior literature that has focused on either the ventral or lateral surface separately, unifying them in one common data set and analysis procedure.

Additional ROIs

We also conducted these same analysis in the classic scene regions parahippocampal place area and occipital place area (PPA and OPA; Epstein (2014)), and object-responsive lateral occipito-temporal cortex (LOC; Grill-Spector et al. (2001)), to more directly relate our findings to previous literature.

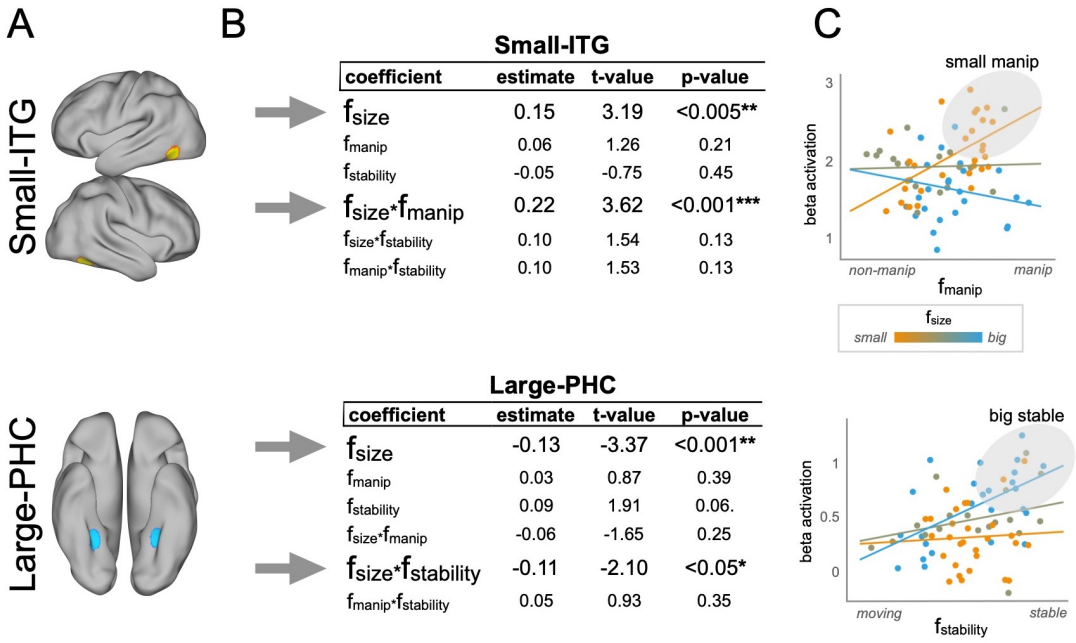


Figure 1.5: ROI analysis. a) Visualization of size ROIs Small-ITG and Large-PHC in one example subject. b) GLMs with ROIs' activity as dependent variable and factor scores as independent variables. Shown are estimates for all predictors included in the model, t-value and p-value. The t-value was measured by dividing the coefficients by their standard errors. The p-value tested the hypothesis of achieving that t-value if the coefficient were to be zero. c) Top: scatterplot of Small-ITG's activation by f_{manip} (x-axis) and f_{size} (y-axis). Dots are color-coded based on the f_{size} weight from blue (low-weight, large size) to orange (high weight, small size). Bottom: scatterplot of Large-PHC's activation by $f_{stability}$ (x-axis) and f_{size} (y-axis). Same color-coding scheme as described for the plot above.

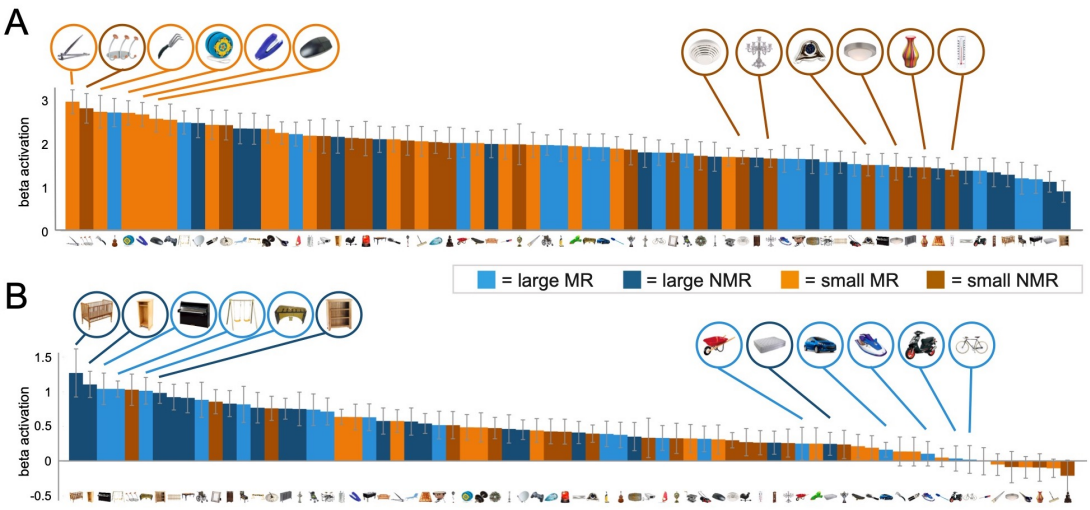


Figure 1.6: Barplots of ROIs activation. Mean activation to each of the 72 object categories are plotted for each region (a: Small-ITG; b: Large-PHC), where the categories are rank-ordered from most to least active, The object categories are color-coded based on their size and motor-relevance conditions (light orange – small motor relevant; dark orange: small non-motor relevant; light blue: large motor relevant; dark blue: large non-motor relevant). Error bars reflect ±1 SEM across subjects.

These regions were functionally defined at the individual level with an independent Category Localizer (see Methods), and are adjacent and partially overlapping with the size ROIs (Konkle & Oliva, 2012b).

The findings in these well-studied category-preferring regions generally converge with what was observed for the size ROIs, and are reported in Appendix A: in LOC, there was a significant contribution of f_{size}, and a significant contribution of the interaction between f_{size} and f_{manip} (see Figure A.1 left). In PPA, there was a significant contribution of the size factor, and marginally of stability and of the interaction between size and stability (see Figure A.1 right). Further, we conducted these analyses in an additional occipito-dorsal region identified with the [Large > Small] size contrast from two runs of the main experiment (Large-TOS) and in adjacent scene-selective occipital place area (OPA). In Large-TOS, we found a result parallel to our observation in PPA: a significant contribution of f_{size} and of the interaction between f_{size} and $f_{stability}$ (see Figure A.2 and A.3). However, this result did not replicate in OPA where only f_{size} had a significant contribution (see Figure A.4).

Searchlight Analysis

By focusing on the ROIs produced by the size contrast, we were able to explore whether other factors beyond size and motor-relevance were important in these ROIs—to which the answer is clearly yes: motor-relevance was indeed split into two sub-dimensions, one more related to manipulability, and one more related to movement, which dissociated in explaining small- and large-object ROIs. However, by performing a targeted ROI analysis, we might have missed additional regions with different associations between these factors.

Given the complexity of the number of possible interactions among these factors, we approached the searchlight analysis with a more targeted comparison. Specifically, we mapped which regions showed a better model fit among two candidate models:

- model 1, that associates size and manipulability [$f_{size} + f_{manip} + f_{size} * f_{manip}$]
- model 2, that associates size and stability [$f_{size} + f_{stability} + f_{size} * f_{stability}$]

All analyses were conducted within the most reliable object-responsive voxels (Tarhan & Konkle (2019), see Methods). At each searchlight, neural responses from a voxel and its neighbors were extracted and used as predicted variable for both candidate models. That voxel was then assigned to the model that best fit the neural responses based on $adjR^2$. The results are shown in Figure 1.7.

From this analysis, we can gather two main conclusions. First, we found that the voxels with best fitting models (i.e., the ones with highest F-value, see Figure 1.7 center) focused mostly in the lateral occipitotemporal and in the ventromedial occipitotemporal regions, in areas mostly overlapping with small- and big-object preferring Small-ITG and Large-PHC. Thus, our targeted ROIs seemed to cover most OTC areas where voxels' information could be explained by a combination of the three factors.

Second, we found that the best-fitting voxels resulting from the searchlight largely confirmed our targeted ROI analyses:

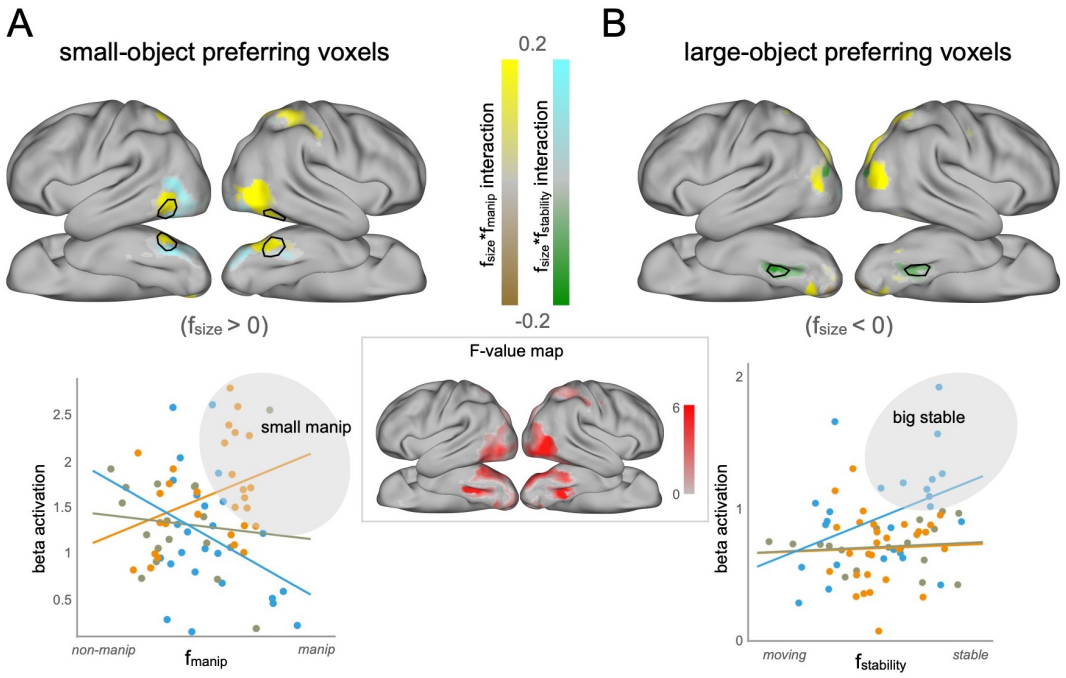


Figure 1.7: Searchlight Analysis. a) Top: surface map of interaction coefficients for small-object preferring voxels (i.e., voxels with a positive f_{size} coefficient for the winning model). In voxels in which the winning model was model 1, that voxel is colored according to its f_{size} * f_{manip} interaction coefficient estimate in a brown to yellow scale. In voxels in which the winning model was model 2, that voxel is colored according to its f_{size} * $f_{stability}$ interaction coefficient estimate in a green-to-cyan scale. Bottom: scatterplot of activation by f_{manip} (x-axis) and f_{size} (y-axis) for all voxels showing a positive f_{size} * f_{manip} interaction in lateral OTC. Dots are color-coded based on the f_{size} weight from blue (low-weight, large size) to orange (high weight, small size). b) Top: surface map of interaction coefficients for large-object preferring voxels (i.e., voxels with a negative f_{size} coefficient for the winning model). In voxels in which the winning model was model 1, that voxel is colored according to its f_{size} * f_{manip} interaction coefficient term in a brown-to-yellow scale. In voxels in which the winning model was model 2, that voxel is colored according to its f_{size} * $f_{stability}$ interaction coefficient estimate in a green-to-cyan scale. Bottom: scatterplot of activation by $f_{stability}$ (x-axis) and f_{size} (y-axis) for all voxels showing a negative f_{size} * f_{manip} interaction in ventromedial OTC. Dots are color-coded based on the f_{size} weight from blue (low-weight, large size) to orange (high weight, small size).

- Small-object preferring voxels in the lateral occipitotemporal region (i.e., the ones with positive f_{size} coefficients for the winning model) presented a positive interaction of the size and manipulability factors (Figure 1.7a top, colored in yellow), similarly to what observed in Small-ITG: that is, the higher f_{size} (the smaller the object), and the higher f_{manip} (the more manipulable the object), the stronger the activation.

- Similarly, we found that large-object preferring voxels in ventromedial occipitotemporal region (i.e., the ones with negative f_{size} coefficients for the winning model) presented a negative interaction of the size and stability factors (Figure 1.7b top, colored in green), similarly to what observed for Large-PHC: that is, the lower f_{size} (the bigger the object), and the higher $f_{stability}$ (the more stable the object), the stronger the activation.

For ease of comparison, we have drawn a black outline on the searchlight coefficient maps of the size ROIs Small-ITG (Figure 1.6a) and Large-PHC (Figure 1.6b) for one example subject, from which it is clear that the targeted ROI analyses do in fact highlight the major structures in neural responses that are well-explained by our three factors.

At lower F-value levels, we observed two more nuanced relationships. However, given the lower fit of model 1 or model 2 within these voxels, one can only draw tentative inferences from these results. First, a significant portion of small-object preferring voxels in the lateral occipitotemporal surface (Figure 1.7a top, colored in cyan) are fit best by models with size and stability, and show a negative interaction between the size and stability factors. Inspection of activation to all 72 object categories in this area shows a preference for moving over non-moving things (see Appendix A, Figure A.5a) which is in line with past observations of a movement-preferring region MT+ in the lateral surface.

Second, we also observed large-object preferring voxels in the dorsal occipito-parietal region, where most voxels (Figure 1.7b top, colored in yellow) indicate a positive interaction of the size and manipulability factors while a few voxels (colored in green) indicate a negative interaction of the size and

stability factors. However, inspection of the scatter plots for these regions showed that these interactions were not as simple and easily interpretable as the ones observed in areas neighboring our size ROIs (see Appendix A, Figure A.5b and c). From these results, it is clear that there are systematically varying object responses in this dorsal occipitotemporal region, but none of the three factors or their interactions cleanly provides a description of what explains the object activations.

1.4 Discussion

Here, we examined whether the large-scale organization of neural responses to object size could be better explained by a related, but importantly different factor: motor-relevance. This was not the case—we found that the contrasts of small vs large objects elicited a clear size organization for both motor-relevant and non-motor relevant objects. We also found that motor-relevance, independent of size, accounted for some of the large-scale structure of neural responses. This observation confirms the importance of motor interaction as an explanatory dimension in the ventral visual stream, and suggests the intriguing possibility that its representation might decouple from object size. Further analyses showed the two dimensions to not be completely independent, in that more typical combinations (small and motor-relevant, large and non-motor-relevant) elicited more reliable representations than less typical ones.

Our subsequent exploratory analyses helped to refine and clarify these results. Specifically, a factor analysis revealed that the construct of motor-relevance could be divided into two underlying dimensions, one better correlated with manipulability (i.e., related to hand performed actions), and one better related to an object's physical presence and stability. Further ROI and searchlight exploratory analyses revealed that the manipulability factor was a better descriptor of activation in small-object preferring lateral OTC, while the stability factor was a better descriptor of activation in large-object preferring ventromedial OTC.

Broadly, these results begin to characterize the ecological covariation between size and many behaviorally relevant object properties, highlighting clear dissociations between the ventral stream regions with a preference for the inanimate domain. In the next two sections, we situate these findings in the literature, and discuss how these results inform the deeper theoretical question of what drives the neural organization of inanimate objects.

1.4.1 Dissociations within the inanimate-domain

Objects are characterized by numerous visual properties, size being one of them. However, size tends to correlate with other more or less visual properties. Rather than considering this issue in terms of competing, mutually-exclusive alternatives, we believe that more nuanced possibilities have yielded deeper answers: specifically, there are suites of partially related factors that predict neural responses to inanimate objects, with major division between small, manipulable objects on one side and large, stable, navigationally relevant objects on the other (Julian et al., 2016; Mullally & Maguire, 2011; Troiani et al., 2012; Auger & Maguire, 2013; Bracci et al., 2012; Macdonald & Culham, 2015; Bainbridge & Oliva, 2015). Overall, these findings point to a richer organization than simply object size alone, where different brain regions may be sensitive to size and other factors for different reasons.

For example, along lateral OTC and middle temporal gyrus, there is a mosaic of partially overlapping regions whose response magnitudes are predicted by a range of related factors, from more primitive object shape (Grill-Spector & Weiner, 2014), to real-world size (Konkle & Oliva, 2012b; Julian et al., 2016), to tools and other objects that are manipulated by the hands (Bracci et al., 2012; Macdonald & Culham, 2015; Mruczek et al., 2013; Lewis, 2006). These regions are also adjacent to motion-selective cortex and body-preferring regions (Grill-Spector & Weiner, 2014), including a region that jointly responds to both tools and hands, spanning the animate/inanimate divide (Bracci & Peelen, 2013). While the number of regions and exact functional divisions along this cortex are still being clarified, from our results and these others, it is clear that these regions are most highly related to small, movable, actionable items.

Along ventromedial OTC, where scene-preferring PPA is located, there are now a number of converging results characterizing what dimensions are best associated with responses to isolated objects. These collective factors include the degree to which the object invokes a local tridimensional space, its size, its tendency to stay fixed in the world, and can be generally summarized by an overarching con-

struct of stable spatial properties and landmark suitability (Troiani et al., 2012; Mullally & Maguire, 2011; Julian et al., 2016; Auger & Maguire, 2013; Epstein, 2014). While these studies examining PPA responses differ in their stimulus sets, designs, and even mode of presentation (e.g. pictures, mental imagery), they all leverage condition-rich designs and factor analyses to characterize the configuration of related object properties. While our observation of a role of spatial stability in PHC cannot in isolation lead to strong inferences, given that it is the product of a post hoc analysis, it joins this collection of studies, adding to the diversity of stimulus sets and neuroimaging designs that show converging evidence that size and position fixedness best explain the object responses in this region.

Our results, jointly with these collections of studies, suggest that a division based on manipulability for small objects and stability for large objects is fundamental in the organization of inanimate object responses. Further, they introduce a methodological implication for high-level vision research, in which the aim is not to find the single best high-level descriptive property to explain neural responses, but instead to understand the nature of the collective factors that capture the complexity of a region's responses.

1.4.2 What drives this object organization?

Why is it important to discover the suite of high-level dimensions that describes a regions response? To discuss this, we want to clarify the difference between understanding which properties of a stimulus lead to a particular neural tuning—that is, what is being computed by a neural population—and understanding the deeper question of why we have this particular tuning to begin with—that is, why is this neural population tuned to these features and not others.

With regard to the nature of feature tuning in occipitotemporal cortex, there is good evidence that low- and mid-level visual features such as curvature and texture comprise a relatively strong component of the tuning, with support from neuroimaging (Andrews et al., 2010; Rajimehr et al., 2011; Bracci & Op de Beeck, 2016; Long et al., 2018). These findings point to a basic role of visual input in shaping

the representational space in occipitotemporal cortex. Consistent with this account, recent evidence indicates that neural response structure is very well predicted and modeled with deep convolutional neural networks (Güçlü & van Gerven, 2015; Khaligh-Razavi & Kriegeskorte, 2014; Yamins et al., 2014; Kubilius et al., 2016), where feature tuning reflects increasingly intricate combinations of images statistics. However, the answer we gain at this level is the product of weighting thousands of model parameters, which may not be as satisfying or informative at a higher level and for broader theories.

In contrast, we suggest that characterizing the feature tuning at a more cognitive level, as we do here in terms of manipulability and spatial stability, serves a broader theoretical role by providing suggestive evidence towards the deeper question of why we have this specific tuning. In this approach we ask: what are the high-level descriptors that give an intuitive sense of the (mid-level shape) properties encoded by different neural populations? With our results, we are not trying to claim that manipulability, stability and size are coded in small- and large-object regions at a high-level, that is, independent of shape or visual information. Rather, we propose that these well-fitting high-level descriptors provide explanatory value by suggesting that particular regions of cortex are read-out for those high-level judgments. And, more provocatively, these properties may play a causal role in guiding the kinds of shape features that come to be tuned in service of high-level information about objects' real-world size, their stability, and their manipulability.

1.5 Methods

1.5.1 Participants

For the fMRI experiment, 21 healthy adults with normal or corrected-to-normal vision participated in a 2-hour fMRI session (age, 18-30 years; 11 females). Three were later removed for excessive head movement, and the remaining 18 subjects were analyzed. Each subject gave informed consent according to procedures approved by the Institutional Review Board at the University of Trento.

For the online behavioral experiments, ratings were collected from 644 total participants (464 subjects for the stimulus set development and validation over the dimensions of motor-relevance [n=161], size [n=168] and recognition [n=135]; and 36 subjects for each of the dimensions of manipulability, spatial definition, position variability, motion-based identity and interaction envelope).

1.5.2 Stimuli

The stimulus set consisted of images from 72 object categories, where each object category belonged to one of four conditions (small motor-relevant, small non-motor-relevant, large motor-relevant and large non-motor-relevant) with 18 object categories per condition. Each object category included images of six exemplars of that category, for a total of 432 pictures in the stimulus set (Figure 1a). Each image depicted an isolated object on a white background, and the image was sized to 512 x 512 pixels with the object's longest axis reaching the border of the image.

Stimuli validation. To develop and validate the stimulus set, we performed three separate behavioral studies using Amazon Mechanical Turk on an initial larger set of 20 object categories per condition with 7 unique exemplars for each object category (560 total images). Our aim was to construct a stimulus set in which items were recognizable, and where the size and motor-relevance dimensions were balanced.

1. Recognition ratings. Observers were presented with an item and reported whether they were familiar with this object (yes/no), and how easily they recognized the object on a scale from 1 (very hard to recognize) to 5 (very easy to recognize).

2. Motor-relevance ratings. Participants were presented with an item and rated the degree to which the object made them think of moving their hands or other parts of their body, using a scale from 1 (not at all) to 7 (very strongly).

3. Size Ratings. Participants were presented with an item and judged the size of an object as they would encounter it in everyday life, on a scale from 1 to 7. To provide a reference for the scale, pictures of a wristwatch (small size), a hiking backpack (medium size) and a bed (large size) were provided beside the values of 1, 4 and 7 on the scale.

For all three experiments, each participant made judgments for 20 items (5 items from each of the four conditions). Only one exemplar for an object category was presented to each participant. Ratings were obtained for all 560 objects combining data from all participants, so that each image was rated by 6 observers along the measures of size and motor-relevance and by 5 observers along the measure of recognition.

In order to eliminate those items that were least recognizable, we dropped one item from each object category based on the familiarity score in the recognition experiment. We next eliminated two object categories that scored low in familiarity in order to balance the ratings from the other two experiments (Motor-relevance and Size experiments). This resulted in a new stimulus set with 18 objects per condition (72 total) and 6 items per object. All results and analysis that follow refer to this new stimulus set.

To ensure these factors were balanced, we conducted several statistical tests. As expected, motor-relevant objects had higher motor-relevance scores than non-motor-relevant objects ($t_{35} = 32.93$, $p < 0.001$), and large objects had higher size ratings than small objects ($t_{35} = 19.82$, $p < 0.001$). Further,

large and small objects did not differ in their average motor-relevance scores (t_{35} = 1.63, p=0.11). The same was true when looking separately at large motor-relevant and small motor-relevant objects (t_{17} = 1.05, p = 0.31), and large non-motor-relevant and large non-motor-relevant object categories (t_{35} = 1.35, p = 0.2). Similarly, motor-relevant and non-motor-relevant objects did not differ in their average size score (t_{35} = -0.72, p = 0.5) and neither did large motor-relevant and large non-motor-relevant relevant objects (t_{17} = 0.96, p = 0.35), nor small motor-relevant and small non-motor-relevant objects (t_{17} = -1.76, p = 0.1).

1.5.3 Image Acquisition

Imaging data were acquired using a BioSpin MedSpec 4T scanner (Bruker) with an eight-channel coil. Functional data were obtained with an echo-planar 2D imaging sequence (repetition time TR: 2000ms; echo time TE: 33ms; flip angle: 73º; slice thickness: 3 mm; gap: 0.99 mm, with 3x3 in-plane resolution and 34 slices). Volumes were acquired in ascending interleaved order of slice acquisition.

1.5.4 Experimental Design and Statistical Analysis

Design and procedure

During the fMRI experiment, subjects viewed images of the small motor-relevant, small non-motor-relevant, large motor-relevant, large non-motor-relevant objects in a classic blocked design. In each run, stimuli from each of the 4 conditions were presented 6 times in 12s stimulus blocks, with 6 rest blocks of 12s interspersed. Images were each presented for 800ms with a 200ms blank between each stimulus. Fixation blocks also appeared at the beginning and end of each run, for 4 and 12 seconds respectively. Participants were instructed to maintain fixation and to press a button when a red frame surrounded one of the objects, which happened once per block. The full experiment consisted of six runs each lasting 6 minutes and 12 seconds.

Additionally, we added sub-structure in each stimulus block, to support more exploratory analyses about the relationship between all 72 object categories (see Figure 1b). Specifically, each stimulus block was comprised of 4s mini blocks. In each mini block, four different exemplars from the same category were shown. For example, in a 12s block of large motor-relevant objects, a participant would see 4 different cellos, followed by 4 different swings, followed by 4 different pinball machines. Within each run all 72 object categories were presented exactly one time. The specific 4 exemplars that were displayed in each block were randomly chosen without repetition from the set of 6 possible exemplars from that object category.

Each image was presented in isolation at a 8 x 8º visual angle. Stimulus presentation was performed using the Psychophysics Toolbox package (Brainard, 1997) in MATLAB.

Functional Localizer

Three independent functional localizer runs were performed to identify category-selective regions. These runs consisted of a blocked design, in which each block included one of the following stimuli: 1) objects 2) scrambled objects 3) scenes 4) bodies 5) hands. Each run lasted 6 minutes and 36 seconds, with 50 6-second stimulus blocks (10 per condition) and ten 8-second rest blocks interspersed. In each stimulus-block, six different stimuli from the same condition were presented, where each image was shown for 800ms followed by 200ms of fixation. Additionally, there was an orthogonal motion manipulation: for each of the 5 stimulus categories, in half of the blocks the object was presented at the center of the screen, and in the other half the object moved from the center out in one of 8 randomly directions randomly. Observers were instructed to maintain fixation throughout the experiment and press the button when an exact image repeated back to back.

Two of the 21 subjects performed 8 experimental runs (instead of 6) and only 2 Localizer runs (instead of 3). Data from the all experimental runs for these subjects were used for data analysis. Due to a technical error, localizer runs for one of these subjects did not include scrambled objects.

For the purpose of the current project, we were only interested in identifying category-selective regions Lateral Occipital Complex (LOC; Objects > Scramble) (Grill-Spector, 2003), the Parahippocampal Place Area (PPA; Scenes > Objects) (Epstein & Kanwisher, 1998) and the Occipital Place Area (OPA; Scenes > Objects) (Dilks et al., 2013). We were not able to localize area LOC in the subject for whom the scrambled objects condition was lacking. To compare the coordinates of our contrasts of interest with the location of well-known category-selective regions, we also localized extrastriate body area (EBA; Downing et al., 2001) and the hand-selective area (Bracci et al., 2012).

fMRI Data Analysis

Functional data were analyzed using SPM12 and bspmview running on MATLAB (2014b and 2016a). Results were displayed with Caret (Van Essen et al., 2001) for brain maps and ggplot2 (Wickham & Chang, 2009) in RStudio for bar plots and scatterplots.

The raw functional images were submitted to preprocessing, where the first 4 volumes were discarded from each run, slice scan-time correction was performed, followed by 3D motion correction, normalization and spatial smoothing (8 mm FWHM kernel). Data were modeled using standard general linear models (GLM). The first GLM included regressors for each of the four main conditions, with run regressors and motion correction parameters included as nuisance factors. The second GLM modeled each object category as a separate condition for a total of 72 regressors, with run regressors and motion correction parameters included as nuisance factors.

Regions-of-interest selection

We created our ROIs utilizing a separate set of runs from the ones we used for data-analysis. Spherical ROIs were defined around the individual peaks of activation for the whole-brain size contrast Small > Large (collapsing over motor-relevance) from two experimental runs. Univariate responses were

extracted from the remaining experimental runs for size and motor-relevance conditions. These ROIs were defined with an 12mm sphere (257 voxels) centered around the peak positive and negative voxels in individual subjects, and then intersected with individual whole-brain gray matter masks.

The selected peaks are referred to following the conventions of Konkle & Oliva (2012b) and included small-object preferring bilateral inferior temporal gyrus (Small-ITG, left hemisphere [LH] n=15, right hemisphere [RH] n=17), and large-object preferring bilateral PHC (Large-PHC, LH n=16, RH n=15). After intersecting the 12mm ROI sphere with each subjects' gray-matter mask, the ROIs slightly varied in size across subjects. After intersection, the average size was 244 voxels for Small-ITG and 257 voxels for Large-PHC.

We further identify large-object preferring TOS (Large-TOS) dorsally based on the size contrast, and a series of category-selective regions using the Functional Localizer runs. The regions identified with the Localizer were object-selective lateral occipitotemporal cortex (LOC), scene-selective parahippocampal place area (PPA), and scene-selective occipital-place area (OPA). Analysis and results for Large-TOS and the category-selective ROIs are reported in the Appendix A.

Voxel mask

We produced a mask of the most reliable object-selective voxels in occipital, temporal and parietal regions using the reliability-based voxel selection method for condition-rich designs (Tarhan & Konkle, 2019). To compute the reliability maps, average beta values for odd and even runs were separated in two groups and correlated for each voxel that was included within the gray matter. To establish a reliability threshold, we then correlated the patterns for the same object category at different thresholding levels and picked the threshold at which the average correlation across conditions plateaued, which in this case was r = 0.25. We then produced a mask from the reliability map at this value.

Whole-brain conjunction analysis

To explore the size organization across motor-relevance, we performed a random-effects conjunction analysis at the individual level between small motor-relevant > large motor-relevant and small non-motor-relevant > large non-motor-relevant. The conjunction analysis between two contrast maps, a new map is produced in which the weighted β-value, corresponding to the smallest t-value between the two contrast maps, is assigned to the corresponding voxel. This resulted in eighteen conjunction maps (one for each participant) which were then submitted to a one-sample t-test, and the resulting t-values were assigned to the corresponding voxels on a new map. The voxels with highest value in these new t-maps are those that are strongly activated or deactivated by both contrasts in most subjects. The resulting maps were thresholded with FDR $q < 0.05$. Similarly, to explore the motor-relevance organization across size, we performed the same analysis but considering small motor-relevant > small non-motor-relevant and large motor-relevant > large non-motor-relevant contrasts.

Typical-Atypical contrast analysis

To further explore the relationship between size and motor-relevance, we compared maps between two contrasts: a typical contrast [small motor-relevant > large non-motor-relevant], which follows the real-world covariation of small objects being more motor-relevant than large objects; and an atypical contrast [small non-motor-relevant > large motor-relevant]. Individual maps for each contrast were submitted to a second-level analysis to obtain a group map, which was FDR-thresholded at $q<0.05$.

To assess the reliability of the two contrasts, we further produced individual maps for each contrast separately for odd and even runs and correlated them. We then submitted individual correlations for the two separate contrasts to a t-test to assess whether there was a significant different in reliability between the maps produced by the two contrasts.

1.5.5 Exploratory Analysis

Object Property Ratings

Ratings for the 72 objects in our dataset were collected on a series of other dimensions to identify the ones that were best associated with the response profiles of our ROIs. All studies were conducted on Amazon Mechanical Turk with 36 judgments per object category. Each of the 36 subjects rated one exemplar of each of the 72 object categories, with six subjects rating the same exemplars within each category. Subjects were not presented with two exemplars from the same category. The following object properties were rated:

- Manipulability. Manipulability tracks the degree to which hand-interactions with objects are prominent. Participants were presented with an item and rated the degree to which the object made them think of moving their hands, using a scale from 1 (not at all) to 7 (very strongly). The question asked to participants was very similar to the one posed to collect the motor-relevance dimension, but here we focus on actions performed with the hands rather than with other parts of the body.

- Motion-based Identity. The degree to which an object's parts are expected to move and change position might be a relevant property for representing objects. Subjects were asked to rate on a scale from 1 to 7 the degree to which motion is important in determining the identity of the object.

- Spatial Definition. This property is defined as "the degree to which objects evoke a sense of surrounding space" (Mullally & Maguire, 2011). Subjects were asked to rate on a scale from 1 to 7 the degree to which a background is evoked when looking at a specific object.

- Position Variability. Objects can be more or less likely to change location in the environment

(i.e., a car is more likely to change location than a swing). Subjects were asked to rate on a scale from 1 to 7 the degree to which an object is likely to change position in everyday life.

- Interaction Envelope. "Interaction envelope" was operationalized as a measure of the amount of space needed to interact with the object (e.g., a hamburger needs to be manipulated with two hands that cover its whole surface, thus it has a larger interaction envelope than a coffee mug which just needs one hand on its handle to be manipulated), following Bainbridge et al. (Bainbridge & Oliva, 2015). Subjects had to judge on a scale from 0 to 2 how many hands are often used when interacting with an object.

Pairwise correlations

We compared the similarity across our seven dimensions (size, motor-relevance and the additional ones collected post hoc), and the reliability of each of the rated dimensions, in the following way. First, we divided subjects into two groups and computed the mean rating for each group and each object category separately for each dimension, resulting in two group vectors of 72 mean ratings for each dimension. To compute the reliability of a dimension, we correlated the two group vectors with each other. To compute the correlation between two different dimensions, we correlated each dimension's group vectors with the other dimension's group vectors, and averaged across the four combinations.

Factor Analysis

To understand how our critical dimensions of size and motor-relevance related to the variety of stimulus properties from the subsequent stimulus norming studies, we conducted a maximum-likelihood factor analysis on all seven dimensions (size, motor-relevance, manipulability, motion-based identity, spatial definition, position variability and interaction envelope) with a "varimax" rotation of the coefficients. This factor analysis was implemented with the factanal function in R. To determine the num-

ber of meaningful factors to extract from our dimensions, we performed a "parallel" analysis (Horn, 1965). This procedure involves generating random data and submitting them to the same factor analysis, iterated 5000 times. The average of the eigenvalues resulting from the parallel random-data factor analyses is then compared to the factors from the observed data: if the average eigenvalues from the parallel factor is smaller than the one from the data factor, then the data factor is kept. This procedure was implemented with the fa.parallel function in R. The three resulting factors can be summarized as relating to real-world size and context (f_{size}), manipulability (f_{manip}) and degree of movement ($f_{stability}$).

For the following ROI and searchlight analyses, for ease of interpretation, we flipped the direction of $f_{stability}$ so as to positively correlate with degree of stability.

ROI analysis with Factors as predictors

Following the factor analysis on the behavioral ratings of the objects, we performed a linear model analysis to predict the size ROIs responses from the resulting factor scores and their interactions. To estimate the response profile of each ROI, beta weights for each of the 72 object categories were extracted from the GLMs of the main experiment runs (which were not used to produce the ROIs). These betas were averaged across voxels, resulting in a vector of 72 values reflecting the ROIs overall response activation to each category. These ROI activation profiles were computed for each subject using subject-specific ROIs and GLMs.

We next sought to explain these ROI activation profiles using the factors from the factor analysis above and a linear modeling approach. All linear models to predict an ROI activation were implemented with the function lm in R. We first looked for hemispheric interactions by submitting our size ROIs to a model with the three factors, hemisphere and their interaction as coefficients. No hemispheric interaction significantly explained either Small-ITG or Large-PHC's activation; hemisphere was thus removed from the models.

Next, we examined which of three nested models best explained activation in our size ROIs by com-

paring their resulting adjusted R square (adjR2). The three models were: one with the three factors as predictors, with no interaction terms [f_{size} + f_{manip} + $f_{stability}$]; a second model that also included the two-way interaction terms [f_{size}*f_{manip} + f_{size}*$f_{stability}$ + f_{manip} *$f_{stability}$]; and a third model that also included the three-way interaction term [f_{size}*f_{manip}*$f_{stability}$]. The model with the highest adjR2 was considered the one most fitting the region and the significant coefficients were explored. If two models had comparable adjR2 (rounded at the second decimal), the simpler model was picked. The same analysis was also performed on Large-TOS and in category-selective regions defined from the Functional Localizer (LOC-sphere, PPA-sphere and OPA-sphere; see Appendix A).

To evaluate the quality of the fit of the different coefficients, we looked at the resulting associated t- and p-values. T-values were computed by dividing coefficients by their standard errors, while p-values tested the significance against the hypothesis of obtaining given t-values if the coefficients were not actually contributing in explaining the dependent variable.

Searchlight Regression Analysis

Following our findings in the spherical ROIs, we next explored which factors would best fit voxels throughout object-responsive cortex.

The searchlight analysis was performed within this mask in the following way: within the voxel mask, a 6mm sphere was drawn around a voxel, and all voxels falling within that sphere were considered part of the neighborhood for that voxel. This procedure was performed iteratively so that each voxel within the object-selective mask was at the center of the sphere. In each searchlight ROI, voxel activation for all 72 object categories was averaged across all voxels within the sphere, and the resulting vector used as the dependent variable for two models.

For each search sphere, we fit two models. The first model related size and manipulability, with f_{size}, f_{manip}, and their interaction [f_{size} + f_{manip} + f_{size} * f_{manip}]. the second model related size and stability, with f_{size}, $f_{stability}$ and their interaction [f_{size} + $f_{stability}$ + f_{size} * $f_{stability}$]. For each given searchlight,

adjusted R square values for these two models were extracted and compared, and the model with the highest positive adjR2 value was selected and assigned to that searchlight's central voxel.

To visualize which models are fitting best where, we looked separately at small-preferring and big-object preferring voxels. Small- and big-object preferring voxels were identified based on whether the coefficient for the main effect of f_{size} was positive (preference for small) or negative (preference for big). We then visualized the coefficient maps for the interaction terms f_{size}*f_{manip} and f_{size}*$f_{stability}$ separately for big- and small-preferring voxels, comparing them with a map of the F-values for the winning model associated with each voxels.

2

Behavioral and Neural Associations between Object Size and Curvature

2.1 Abstract

In the physical world, the size of an object is systematically related to its shape—big objects tend to be boxier, small objects curvier. Here, we examined whether internal object representations are sensitive

to this size/curvature covariance, leveraging a 2 x 2 stimulus set of big and small, boxy and curvy objects. In a behavioral experiment, participants judged either the real-world size or the curvature of these objects. Overall, people were faster and more accurate at judging objects with typical size-curvature combinations relative to those with atypical combinations. These images were also presented to new participants in a neuroimaging experiment, where we found that size and curvature were jointly represented, with more neural representational resources devoted to processing big-boxy and small-curvy combinations than the atypical combinations. This behavioral and neural evidence demonstrates that the visual system is sensitive to the covariance between real-world size and curvature, reflected in more efficient behavioral judgments and more robust neural responses.

2.2 Introduction

We can easily and seamlessly detect and recognize objects around us, and one of the properties that seems to be deeply related to this capacity is the physical size of the object in the world—whether it is small and manipulable or big enough to provide support or shelter. Indeed, as soon as we recognize an object, we automatically access information about its real-world size (Konkle & Oliva, 2012a). The real-world size of objects is a large-scale factor in the structure of object responses in the ventral visual stream (Konkle & Oliva, 2012b; Konkle & Caramazza, 2013; Julian et al., 2016).

Moreover, there is evidence that the visual system processes real-world size information through systematic differences in the way big and small objects look. For example, in behavioral studies we found that big and small objects have systematically different perceptual features, related to differences in texture and coarse form information (Long et al., 2016), and these mid-level features of an object are sufficient to trigger size-related processing without even recognizing the object first (Long & Konkle, 2017a). Further, the large-scale organization of object responses related to real-world size is evident even when viewing textured blobs that only preserve the mid-level features of objects (Long et al., 2018). Indeed, there is clear converging evidence using many different approaches that neural responses in this cortex predominantly reflect mid-level visual features related to shape and texture, rather than very abstract high-level tuning (Bracci & Op de Beeck, 2016; Hong et al., 2016; Kaiser et al., 2016). While these findings highlight the relevance of mid-level features for higher-level processing, they raise an important new challenge to characterize what exactly these mid-level feature differences are and how they relate to the overarching higher-level categorical distinctions.

Here, we focus on probing the links between the real-world size of inanimate objects and the perceptual property of curvature. We have speculated that the physical world gives rise to systematic links between size and shape: big objects are boxier with extended surfaces, to support people and withstand gravity; small objects may take on more possible shapes but may tend to be curvier and hand

held (Konkle & Oliva, 2012b). Behaviorally, we have found that perceptual judgments of how curvy-to-boxy a depicted object is correlates with its perceived real-world size (Long et al., 2016), and that this measure can be used to predict the magnitude of familiar size Stroop effects (Long & Konkle, 2017a). With neuroimaging, we have found that the perceived curvature of objects was a good predictor of the neural similarity of objects across the ventral visual stream (Long et al., 2018). Our aim here is to directly examine the degree to which this mid-level feature of curvature is integrally related to the property of real-world object size, in both behavioral judgments and neural responses.

Curvature has long been recognized has one of the primary features being extracted during intermediate stages of visual processing (Attneave, 1954; Asada & Brady, 1986; Biederman, 1987; Drucker & Aguirre, 2009; Vernon et al., 2016). Neurally, there is extensive evidence for the importance of curvature for the organization of monkey visual cortex. For example, recordings in area V4 in macaque monkeys indicate that this region is more responsive to curvy lines than to straight lines (Gallant et al., 1996), it shows a bias for convex features (Pasupathy & Connor, 1999), and it shows size invariance with respect to curvature preference (El-Shamayleh & Pasupathy, 2016). Indeed, a network of curvature-selective patches has been observed using functional neuroimaging of the macaque's brain along the ventral visual pathway (Yue et al., 2014). Finally, curvature has even been hypothesized to be a critical factor in the emergence of the large-scale topography of object responses in the ventral stream: curvature, end-stopping and eccentricity maps tend to correlate in the macaque's visual cortex, with centrally-biased patches corresponding to end-stopped sites and preferring curvy contours, and peripherally biased regions preferring straight stimuli (Ponce et al., 2017; Srihasam et al., 2014). Thus, there is extensive evidence that curvature is an important factor in the tuning properties of neural responses.

To explore the degree to which real-world size and curvature are coded independently versus dependently, we selected stimuli to play against the natural covariance we hypothesize between size and curvature. Thus, we collected a stimulus set of big boxy objects, big curvy objects, small boxy ob-

Figure 2.1: Stimuli. Stimulus set with 180 images of inanimate objects, with 45 images for each condition of big boxy, big curvy, small boxy, and small curvy objects. All objects were balanced for familiarity, pixel area and aspect ratio across all conditions.

jects, and small curvy objects (180 total images, 45 per condition; see Figure 2.1). Behaviorally, we tested whether people's judgments of size were automatically affected by curvature (and vice versa), even when the opposite dimension was not relevant to the task. With functional neuroimaging, we examined the degree to which the occipitotemporal cortex jointly or independently coded these two dimension.

2.3 Results

2.3.1 Behavioral judgments of size and curvature

To examine the associations and dissociations between the two properties of object size and curvature, we created a 2 x 2 stimulus set consisting of small curvy, small boxy, big curvy and big boxy objects (180 images, 45 per condition; see Figure 2.1). Behavioral norming studies were conducted to balance the image set for size variation across curvature, curvature variation across size, and for overall familiarity (see Methods). We also balanced overall pixel area across all four conditions—even though this factor is also likely meaningfully correlated with the size of the objects, here our goal was to probe the relationship between curvature and other underlying factors that relate to real-world object size.

In the behavioral task, we aimed to assess whether people's judgements of one dimension were automatically influenced by the other dimension, even when it was not relevant to the task. Participants (N = 25) were presented with single objects and had to judge their size—whether the objects were big or small (Size task), in the first block of trials. In the next block of trials, for the same objects, they had to judge curvature – whether the objects were curvy or boxy (Curvature task). Participants completed 4 blocks of trials, alternating tasks, with a full counterbalance of which task was conducted first and what keys were mapped to which options across participants (see Methods; see Figure 2.2a). Participants were instructed to be as fast as possible while maintaining high accuracy.

The results of the behavioral task are shown in Figure 2.2b. In the size task, participants were both faster and more accurate to judge objects as big if the objects were boxy (rather than curvy), and to judge objects as small if the objects were curvy (rather than boxy). That is, we observed a cross-over interaction with faster and more accurate responses for the typical vs atypical size-curvature combinations (2 x 2 ANOVA: size task: interaction for RT: $F_{1,23} = 7.94$, $p = 0.010$, interaction for accuracy: $F_{1,23} = 10.99$, $p = 0.003$; curvature task: interaction for RT: $F_{1,23} = 9.87$, $p = 0.005$, interaction for accuracy: $F_{1,23}=8.090$, $p=0.009$). A similar effect was found for the curvature task: that is, people were

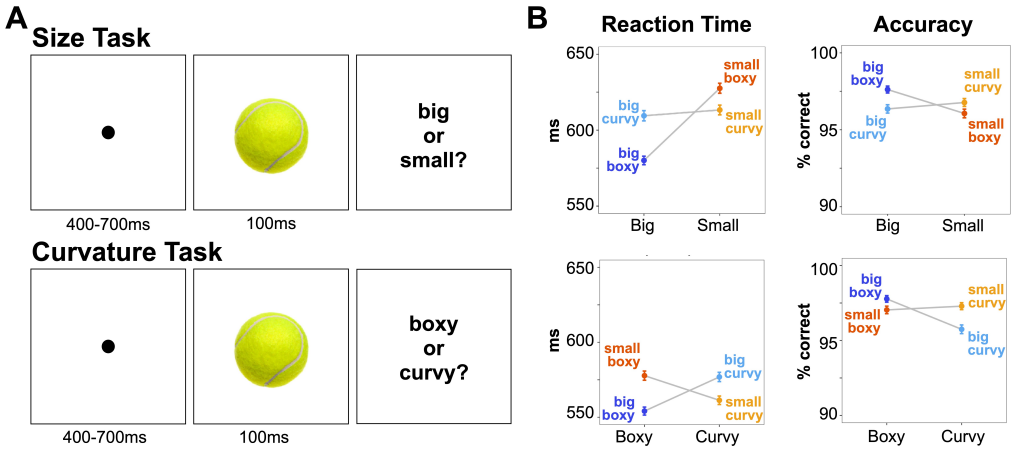


Figure 2.2: Behavioral Task and Results. a): behavioral paradigm for size and curvature tasks. A fixation dot appeared for 400-700ms, followed by the image of one of the objects for 100ms. After the object presentation, the screen was blanked and participants responded within a 2 seconds to respond with a key-press. In some blocks the task was to judge whether the depicted object was a big or small object in the world; in alternating blocks the task was to judge whether the depicted object was boxy or curvy. b): Interaction plots for reaction times (in milliseconds) and accuracy (in percent correct), for the Size and the Curvature Tasks. Error bars reflect the within-subject standard error (n = 24).

faster to judge an object as boxy if it was big (rather than small), and to judge it as curvy if it was small (rather than big; 2 x 2 ANOVA: interaction for RT: $F_{1,23} = 9.87$, $p = 0.005$; interaction for accuracy: $F_{1,23} = 8.09$, $p = 0.009$). These interactions were also confirmed running a mixed-effect model for each task, with reaction times as the dependent variable, with Size, Curvature, and their interaction as fixed effects, and with participants and items as random effects (Size Task: Fixed effect coefficient on Size-Curvature interaction: $F_{1,4216} = 7.31$, $p = 0.007$, Adjusted $R^2 = 0.46$; Curvature task: Fixed effect coefficient on Size-Curvature interaction: $F_{1,4233} = 9.38$, $p = 0.002$, Adjusted $R^2 = 0.49$; see Methods).

These behavioral data demonstrate that behavioral judgements of both size and curvature automatically influence each other, specifically with a big boxy and small curvy alignment.

One important question is the degree to which this result is induced by the experimental context of having to do both size and curvature tasks on the same objects in the same session. Notably, this argument still relies on there being a systematic alignment between size and curvature properties across

participants. However, an interesting question is whether the interaction effects hold even in the first block of trials, having only done one of the tasks. This would indicate an even stronger underlying association between these two dimensions. To examine this possibility, we conducted a post hoc within-subjects analysis using only the first block of trials for all participants. The pattern of data partly held, with a significant interaction for reaction time and accuracy in the size task (RT: $F_{1,11} = 5.45$, $p = 0.039$; accuracy: $F_{1,11} = 9.77$, $p = 0.010$), but no significant interactions in the curvature task (RT: $F_{1,11} = 2.25$, $p = 0.162$; Accuracy: $F_{1,11} = 0.94$, $p = 0.352$). We did find, however, robust interactions for both tasks when employing mixed-effects modeling (Size Task: Fixed effect coefficient on Size-Curvature interaction with RT in the first block: $F_{1,1889} = 5.76$, $p = 0.016$; Adjusted $R^2 = 0.37$; Curvature Task: Fixed effect coefficient on Size-Curvature interaction with RT in the first block: $F_{1,1947} = 9.73$, $p = 0.002$; Adjusted $R^2 = 0.27$; see Figure B.1). This analysis indicates that the behavioral interaction effects relating size and curvature are evident even in the first block of trials.

In addition, the behavioral data showed several other overall reliable patterns. Being big or being boxy led to faster reaction times when judging the size of the depicted object (big main effect: $F_{1,23} = 32.59$, $p < 0.001$; boxy main effect: $F_{1,23} = 4.59$, $p = 0.043$). This is partly consistent with classic literature finding faster RTs for visually larger objects (Osaka, 1976; Sperandio et al., 2009), and with the observations of faster RTs for objects that are bigger in the real world (Konkle & Oliva, 2012a). Additionally, overall reaction times were faster in the curvature task compared to the size task ($F_{1,23} = 4.59$, $p = 0.043$). This pattern of data might indicate that the mid-level feature of curvature is more quickly accessible than the high-level feature of size, but these results should be interpreted with caution, as these results could change with more subtle or more obvious differences between big/small or boxy/curvy conditions.

Taken together, the behavioral data show clear evidence that the real-world size of an object and its curvature properties are jointly processed, even when either dimension is task-irrelevant. Specifically, they show a stronger association for big boxy objects and small curvy objects than the alternative com-

bination. This pattern of results is consistent with the intuitive arguments that these dimensions are intrinsically correlated in visual object image statistics based on the physical constraints of the world.

2.3.2 Neural responses to size and curvature

Size-Curvature Preference Maps

We have shown evidence for the association between size and curvature properties of objects in behavior—are these same associations evident in neural responses to these two properties? To answer this question, we conducted a functional magnetic resonance imaging experiment in new participants (N = 16) who were shown the same 2 x 2 image set of small curvy, small boxy, large curvy, and large boxy objects in a standard blocked design (see Methods). Our goal was to design an analysis that clarifies how neural preferences for big vs small objects are related to neural preferences for boxy vs curvy objects, and to see how these associations are mapped across object-responsive cortex, without presupposing any relationships or regional boundaries in advance.

To do so, we first defined occipitotemporal cortex based on overall activation levels relative to fixation baseline and anatomical landmarks. Then, for each voxel, we computed its size preference [Small – Big] and its curvature preference [Curvy – Boxy]. If size and curvature properties are coded completely independently across this cortex, we would expect to see an equal number of voxels with a big boxy preference, a big curvy preference, a small boxy preference, and a small curvy preference.

Figure 2.3a shows the scatter plot between size preference and curvature preference. Each dot is a voxel, plotted based on its size preference (x-axis) and curvature preference (y-axis). Overall, voxels have stronger size preferences than curvature preferences, and voxels with a strong big object preference tend to have boxy preferences. Figure 2.3b shows the corresponding frequency histogram of the number of voxels falling within each polar angle of the scatterplot. From this graph, it is clear that the distribution of voxels is not uniform (Rayleigh test of uniformity: $z = 588.08$, $p < 0.001$)—there

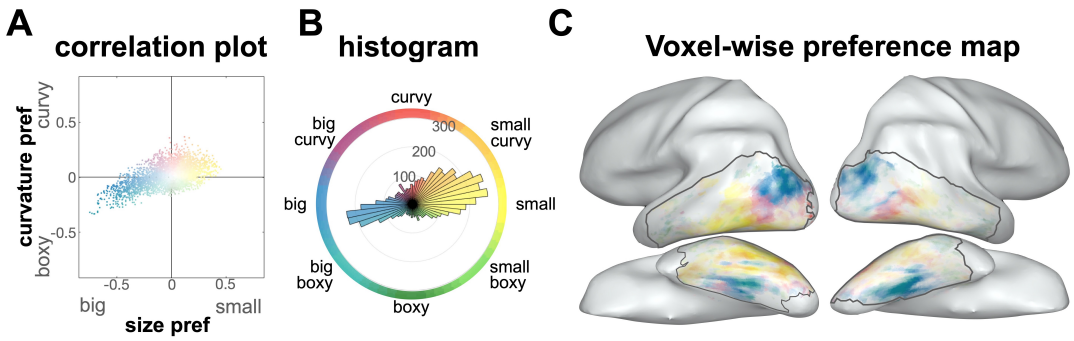


Figure 2.3: Polar scatterplot, radial frequency histogram and brain maps of voxel-wise preference. a): Scatter plot with size preference on the x-axis (small - big) and curvature (curvy - boxy) preference on the y-axis. Each dot corresponds to a voxel in OTC. b): radial frequency histogram of the number of voxels falling on each direction of the scatter plot in A. c): surface map plotting the location of each single voxel in OTC color-coded based on its size and curvature preference following the scheme on the far left. Outlined in gray is the group-level OTC mask within which all voxels with positive values for all four conditions were considered.

is relatively more cortex sensitive to the big-boxy and the small-curvy combinations than other combinations.

To visualize how these voxel preferences are organized across the cortex, we colored each voxel using a two-dimensional color space, where the hue indicates the size-curvature preference, and the saturation indicates the overall strength of preferences, as shown in the correlation plot in Figure 2.3a. The size-curvature preference map is shown in Figure 2.3c (for individual maps, see Figure B.2). Voxels with a small/small curvy preference are localized to the lateral surface, and voxels with a big/big-boxy preferences are localized along both the medial-ventral and medial-dorsal edges of the occipitotemporal cortex. These topographies are consistent with previously mapped response preferences for objects of big and small real-world sizes (Konkle & Oliva, 2012b; Long et al., 2018). Notably, these results clearly highlight the joint sensitivity of big boxy response preferences and small curvy response preferences, using an analysis that could have found any number of possible relationships between these two dimensions across the cortex.

Overall, these size-curvature maps imply that there are more neural representational resources de-

voted to processing big-boxy and small-curvy combinations than the less typical combination of these factors.

Relationship to animacy organization

The size-curvature topographies in Figure 2.3c also show some regions of cortex with a red/pink hue, mostly along the lateral surface, indicating a general curvature preference. This observation suggests that at least some portions of object-responsive cortex are less sensitive to size, but still have a preference for curvier things rather than boxier things. Indeed, we have shown in previous work that, between big and small inanimate object preferring zones, there are zones of cortex with a strong preference for animals (Konkle & Caramazza, 2013; Long et al., 2018), and that on a curvy-to-boxy scale, animals are judged the curviest (see Supplementary Figure S4b in Long et al. (2018)).

To more quantitatively assess this correspondence, we used an independent Localizer to map the tripartite organization related to animacy and object size (Figure 2.4, see Methods). Zones of cortex were defined based on their preference for big objects (blue), animals (purple), or small objects (orange) at the group-level. Figure 2.4 shows the distribution of size-curvature preferences mapped separately for each zone. This analysis confirms the prior results that regions with a big object preference were not uniformly distributed but show a slight boxy preference overall, while regions with a small object preference have a slightly broader distribution of curvature preferences on the curvy side (Rayleigh test of uniformity: small object preferring voxels: $z = 47.76$, $p < 0.001$; big object preferring voxels: $z = 397.77$, $p < 0.001$). Further, this analysis reveals that cortex that responds more to pictures of animals than inanimate objects shows an overall bias towards small curvy objects (Rayleigh test of uniformity for animal-preferring voxels: $z = 372.04$, $p < 0.001$). Overall, there is a preference towards small/small curvy objects in the animal-preferring regions, which is expected given that these zones are known to have systematic structure in the non-preferred conditions, responding to an intermediate degree to small objects and relatively less to big objects (Konkle & Caramazza, 2013). However, the ranging hues

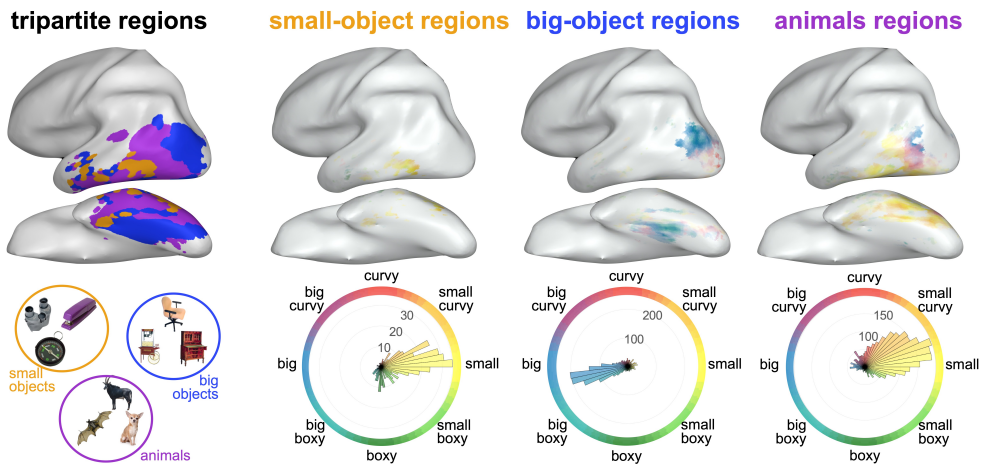


Figure 2.4: Voxel-wise analysis by animal-, big-objects and small-object preferring regions. Far left: visualization of group-level animal-preferring, small-object preferring and big-object preferring regions as produced from the Tripartite Localizer, with example stimuli from the Localizer runs at the bottom. These regions were produced at the group-level by assigning voxels to the condition which had the highest activation among all three, respectively animals-preferring (animals> small and big objects), big-object preferring (big objects > animals and small objects) and small-object preferring (small objects > animals and big objects). On the right, surface maps and histograms of curvature and size preference for these Tripartite regions. We plot the radial frequency histograms for size and curvature preference as described in the Methods and in Figure 2.3, this time restricting the selection of voxels to the corresponding Tripartite region.

evident on the lateral aspects of animate preferring cortex highlight the potential for other organizing distinctions within this cortex.

Taken together, this analysis serves to characterize how neural tuning related to inanimate object size and curvature falls with respect to the major large-scale divisions of the occipitotemporal cortex.

Alternative size-curvature tuning combinations

The size-curvature analyses we have conducted so far focus on how these two dimensions align. However, it does not give us a clear indication of the degree to which one dimension (e.g. curvature) would generalize across the other (e.g. size). To probe for the presence of these voxels, we look at the degree to which curvature preference generalized across size: that is, the proportion of curvy-preferring voxels that were preferring both small curvy over small boxy, and big curvy over big boxy objects.

This analysis revealed that 62% of voxels which showed a curvy over boxy preference, tended to do so both for small and big objects (2164/3487), while only 46% of voxels which showed a boxy over curvy preference did so for both small and big objects (1229/2663 of voxels).

On the other hand, when factoring by size, 73% of voxels which show a small over big preference tend to do so for both boxy and curvy objects, and 72% of voxels with big object preference do so for both boxy and curvy objects (1803/2487 of voxels).

Taken together, these results suggest that size preference tends to generalize across curvature, while curvature preference shows a relatively weaker degree of generalization across size. Analyses are still ongoing to test the significance of these results, as well as to further characterize the tuning profile of populations that do not generalize across size or curvature.

Typical and Atypical Preference Maps

The final analysis of the neuroimaging data focused on targeted contrasts in the 2 x 2 space: the typical contrast, comparing big boxy vs small curvy activations, and the atypical contrast, comparing big curvy and small boxy activations. Doing so enabled us to ask two distinct questions, related to the relative strength and the robustness of the response topographies.

The first analysis was aimed at understanding the relative strengths of curvature and object size properties for the organization of brain responses. We have previously hypothesized that the large-scale organization of object responses by real-world size actually reflects different tuning along mid-level features, like curvature (Konkle & Oliva, 2011; Long et al., 2018). We reasoned that if the typical real-world size organization was relying primarily on curvature differences, then when pitting size and curvature against each other, the response topographies to size should reverse. That is——a region that responds preferentially to big boxy object should respond more to small boxy objects (same curvature), rather than big curvy objects (same size). However, if curvature is not the sole mid-level feature underlying the real-world size organization, then we expect the atypical contrast to follow the original object size topography, perhaps with slightly weaker activations.

The typical contrast topographies and the atypical contrast topographies are shown in Figure 2.5a. As expected, we observed stronger responses to big boxy images in the medial-dorsal and medial-ventral regions, with preferences to small curvy objects on the lateral occipital cortex. Critically, in the atypical contrast, the size-factor predicts the topography—that is, big-boxy regions prefer also big-curvy objects, and the small-curvy regions also prefer small-boxy objects. We quantified this relationship at the single subject level by correlating the two maps (see Methods), and found positive correlations which indicate a size-biased organization, rather than negative correlations which would indicate a curvature-biased organization ($M = 0.39$, $SD = 0.13$). This result is also consistent with our previous analyses that showed voxels in occipitotemporal cortex tended to have stronger size prefer-

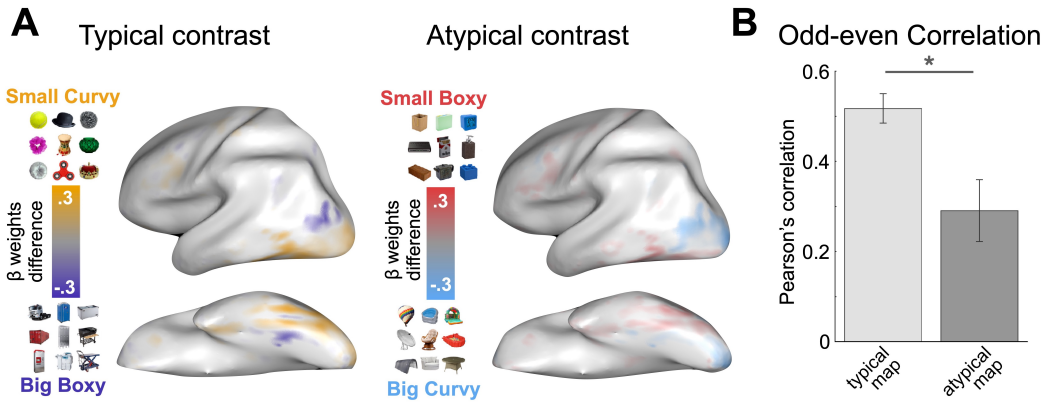


Figure 2.5: Typical/Atypical Maps Correlations. a): Typical [small curvy > big boxy] and Atypical [small boxy > big curvy] Preference maps at the group level. Maps were thresholded at the same level for easier comparison. b): Bar-plots of individual and average odd-even runs correlations for the typical and atypical maps, averaged across subjects. Error-bars represent standard error (n=16).

ences than curvature preferences (Figure 2.3a). More generally, this analysis indicates that curvature (as we have operationalized it here), is only one of the visual properties related to visual differences between big and small objects, and that there are must be others as well (Konkle & Oliva, 2011).

The second question we asked of the typical vs atypical topographies was about their robustness, as measured by split-half reliability. Which contrast yields more reliable organizations? We found that the typical preference map was significantly more reliable than the atypical one (typical split-half map correlations: M = 0.52, SD = 0.13; atypical: M = 0.29, SD = 0.27; $t_{15} = 3.87$, p = 0.002; Figure 2.5b; for individual reliabilities, see Appendix B, Figure B.3). This result highlights that even though the organization of object responses tends to follow size more than curvature (as operationalized here), the responses are clearly affected by their covariation—the organization by object size differences is more robust when aligned with curvature differences than when misaligned.

2.4 Discussion

Here, we investigated how our visual system is affected by the natural covariance between size and curvature by employing a 2 x 2 stimulus set in which these two properties are controlled for.

Behaviorally, we found that people are faster and more accurate when judging objects that follow the natural covariance between size and curvature, suggesting that these two visual features are not independent from each other.

Neurally, we found that the distribution of small-to-big, and curvy-to-boxy preferences in occipitotemporal cortex was not uniform, but also followed this natural covariance.

Overall, these results show that behavioral judgments and neural representations are affected by the natural covariance between size and curvature, strengthening the hypothesis that curvature is a mid-level visual property essential for the internal representations of small and big objects.

Despite having found the two properties of size and curvature to not be independent form each other, they were not equally important in driving neural representations. Indeed, flipping curvature information as to go against the natural world covariance did not flip the normal real-world size organization. What could explain these results?

One hypothesis is that size cannot be reduced completely to curvature, but that there may be additional features other than curvature driving real-world size organization. What could these additional features be? It is possible that texture information, and its natural covariance with main categorical domains such as animacy and size, might have affected our results: together with shape, texture is a mid-level visual-feature that has been observed to drive neural representations in the ventral stream (Grill-Spector & Weiner, 2014) as well as DNN layers representations (Geirhos et al., 2018). While there are clear textural differences between animate and inanimate things (e.g. animals tend to have furs, objects tend to have smoother surfaces), there might also be distinct statistical regularities for small and big inanimate objects that could drive real-world size organization.

Another possibility is that objects' elongation (i.e. width to length ratio), which has been observed to drive representations in the ventral stream (Bracci & Peelen, 2013; Cohen et al., 2017), might be a main driver of the real-world size organization, especially in small object preferring regions that tend to prefer manipulable small objects such as tools, which are typically elongated in order to be hand-held, over other small objects (Magri et al., 2016). This would explain our observation following the voxel-wise analysis that voxels in small-object preferring regions did not show a clear curvature signature: in this area, elongation and width-to-height ratio might be more important.

Another hypothesis for why we failed to see a flipping of the map for the atypical contrast is that there could be limitations in our curvy-to-boxy scale, and thus in the way we picked our stimuli. A curvy-to-boxy scale is a quick and intuitive way to track objects' curvature, and it has been shown to predict reasonably well responses in OTC (Long & Konkle, 2018). However, curvature (i.e., the amount to which a geometric object such as a line deviates from being straight) is a complex construct which requires precise mathematical definitions, and the way we discursively describe it is unlikely to completely encompass its complexities. For example, as an object changes in scale and gets smaller, the absolute value of its curves are going to increase. But curvature can be also defined with size-dependent specifications, as a function of contour length (El-Shamayleh & Pasupathy, 2016). While it has been observed that ventral stream regions which possess scale-invariant representations do also represent normalized over absolute curvature (El-Shamayleh & Pasupathy, 2016), it is not yet clear the degree to which "normalized" versus "absolute" curvature may affect explicit judgments, although it is intuitively more likely that humans may judge curvature on a scale-invariant level.

Further, while we pit curviness and boxiness in two opposite sides of a spectrum, and thus in polar opposites in the curvy-to-boxy space, the mathematical opposite of a curve is not a right angle, but a straight line. Thus, rather than quantifying the amount of right angles in a shape, in order to get to a measure that is the opposite of curvy one should compute the amount of straight lines in a shape. While the presence of right angles and straight lines in an object's contour necessarily correlate, they are

not perfectly matched (i.e. straight lines might be longer or shorter). It is not yet clear whether human judgments over an object's boxiness are solely based on the presence of right angles, or whether the amount and length of straight (or almost straight) lines also affects them.

Given these uncertainties, it is possible that our curvy-to-boxy scale missed some features of curvature that are important for the brain in order to process contour information, but that are not explicitly picked up during judgments. Thus, while we have tested and found some clear associations between size and curvature, future work accounting for additional mid-level features (e.g. texture, elongation), and employing a more quantitative way to measure curvature might bring us closer to fully explain the real-world size organization.

One question raised by these results is, why is the visual system sensitive to this natural covariance? One hypothesis supported by a growing body of work is that the processing of mid-level features statistically covarying with categorical information could lead to more efficient coding (Grill-Spector & Weiner, 2014; Bracci & Op de Beeck, 2016; Hong et al., 2016; Long et al., 2018). Several studies have shown that more anterior parts of the visual system have more abstract representations, but that they still retain sensitivity to low-level features. For example the PPA, which shows a preference for navigationally-relevant objects (Janzen & Van Turennout, 2004), is also preferentially selective for vertical-horizontal lines (Nasr & Tootell, 2012). Specifically, distributed coarse coding might allow for the acquisition of invariant object representations, and it might also preserve information about category-independent features such as position, perimeter or curvature (Pasupathy & Connor, 2002; Connor et al., 2007; DiCarlo & Cox, 2007; Hong et al., 2016).

Aside from the natural covariance between an object's shape and its category, a further force for domain organization formation might be due to ecological and gravitational constraints in the way we see and interact with objects. Indeed, while small objects tend to be mostly foveated to and looked at from a closer distance, big objects tend to be located farther away, and at the periphery of the visual field. Recent work has observed a largely overlapping map between eccentricity bias and curvature maps in

the ventral stream, and this maps were largely predictive of the location of category selective regions, thus suggesting they might function as proto-organizing forces for object-selective cortex (Srihasam et al., 2014).

However, we do not exclude a role of long-range connections to behaviorally relevant networks, which has also been advanced as a possible driver of this organization (Mahon et al., 2007; Konkle & Oliva, 2012b). There is a clear way in which small and big objects tend to be behaviorally relevant: small objects tend to be manipulated, while big objects tend to be used for support and navigation. It has been proposed that long-range connections between small-object regions and the action-related network on one side, and large object regions and the navigation-related network on the other might help in the shaping of real-world size organization (Mahon et al., 2007; Konkle & Caramazza, 2013, 2016; Magri et al., 2016).

Thus, the natural covariance between size and curvature might affect object recognition and subsequently object-oriented action. Overall, our work joins a converging set of findings that "high-level" regions still have sensitivity to mid-level and even low-level features, driving the formation of a broad organization by domain which facilitates the processing of information related to objects and what to do with them.

2.5 Conclusion

Rather than assuming that the ventral stream encodes curvature and size as separate and independent entities, here we show it processes the associations between them, which is evidenced by more efficient behavioral processing and more robust neural responses, as well as extended neural territory. These findings broadly fit into the larger hypothesis that the visual system processing system works by associating levels of representation, with mid-level features in service of high-level properties that are useful for behavior.

2.6 Methods

2.6.1 Stimuli

Final Dataset

Both for the behavioral and the fMRI experiment, we employed a stimulus set of 180 small curvy, small boxy, big curvy and big boxy inanimate objects (45 objects per condition). Stimuli were matched in aspect ratio (F_3 = 1.58, p = 0.19) and pixel area (F_3 = 0.94, p = 0.42) across conditions. They were also matched in real-world size across curvature (t_{178} = -0.65, p = 0.52), in curvature across size (t_{178} = 1.26, p = 0.21) and familiarity across all conditions (F_3 = 0.56, p = 0.64) with ratings collected in Amazon Mechanical Turk.

Stimuli and Ratings Collection

An initial Dataset of 336 images of objects (84 objects per condition) was collected manually through Google Images, the background was removed and the images were squared and resized to 550 x 550 pixels per inch. Size rank (n = 9), Curvy-to-boxy (n = 8) and familiarity ratings (n = 18) were collected on Mechanical Turk for all 336 images. Size rank. Participants were presented with individual objects, and asked to rate their size on a 7-point scale based on anchor objects which were not part of the dataset. Curvy-to-boxy ratings. Participants were presented with individual objects, and asked to rate them on a curvy-to-boxy 5-point scale. Familiarity. Participants were presented with individual objects, and asked to rate their familiarity on a 5-point scale. Further, pixel area and aspect ratio of all images were measured to balance them across all conditions.

Stimuli selection Procedure

We manually excluded 25 individual items for several reasons (e.g. their similarity with the animate category or to human body parts, or for the presence of text). We excluded any of our initial items for which there was a mismatch between actual and perceived size, and for which our assumptions about their curvature did not match raters' judgment. Further, we excluded items that were rated as overly familiar (> 4.7) or overly unfamiliar (< 3.5), to equalize familiarity across conditions, for a total of 131 items removed. This resulted on a total of 180 items, 45 per category. We performed t-test comparisons among all ratings and across all 4 conditions and found that the only failure of perfect balance was a significant aspect ratio difference between small curvy and big curvy objects ($t_{88} = 2.77$, $p = 0.007$).

2.6.2 Behavioral Experiment

Participants

Twenty-five naive subjects participated. All participants were affiliated with Harvard University. A power analysis conducted on previously collected pilot data (N = 12), which showed that a minimum of 24 participants was required to achieve a power of 0.75 in effects of interest. One participant showed exceptionally low overall accuracy (accuracy level 53%), we thus excluded that participants as an outlier. The remaining sample consisted of 24 participants (11 males, 13 females, mean age: 21 years, age range: 18-33 years). All had normal or corrected-to-normal vision.

Experimental Setup

Participants were positioned in front of an Apple iMac (1024 × 768 pixels, 60 Hz) and were instructed to sit straight and keep their body position still during the experiment. The experiment was programmed and run in MATLAB 2017a with Psychophysics Toolbox (Brainard, 1997).

Behavioral Experiment Design

Participants were asked to judge the size (Size Judgments) or the curvature (Curvature Judgments) of presented objects for a total of four blocks, two for each kind of judgment. Instruction on the kind of judgments to give were provided at the start of each block. Each trial started with a fixation dot. Duration was randomly set within a 400-700ms range to prevent participants from anticipating the onset of stimuli. Then, an object was presented at the screen center for 100ms. Each object was presented once per block for a total of 180 trials per block and items were all presented at the same size on the screen (350 × 350 pixels). In the Size Judgment blocks, participants were asked to judge if the real-world size of the object was bigger or smaller compared to an anchor object, a trash that was present in the testing room. In the Curvature Judgment blocks, participants were asked to judge if the shape of the object was perceived as boxy or curvy. Responses were given by response keys following which immediate feedback was provided. Participants were instructed to respond as fast as possible by using a keyboard and pressing either ".>" key with the right index finger or "/?" key with the right middle finger. Following responses, feedback was provided, which lasted 500ms for correct responses and 1200ms for wrong responses. If no keypress was detected within 2000ms from stimulus disappearance, the trial would timeout and a 1200ms feedback display would appear, communicating response time took too long. Thus, high accuracy was encouraged. Prior to the actual experiment, participants practiced 6 trials in size task and 6 trials in curvature task. The formal experiment consisted of four blocks, where participants performed either size task or curvature task in each block. Task type switched after 1st block and 3rd block. Each block consist of 180 trials, with every stimulus from the whole set (4 × 45 = 180 stimuli in total) presented once in a randomized sequence. Therefore, each unique stimulus was judged twice for real-world size and twice for perceived curvature. Overall, there were 720 trials in the experiment. Reaction time and accuracy were recorded for every trial.

Task order were counterbalanced across participants so that half of the participants started with a

size task, and the other half started with a curvature task. Task order during training was consistent with the order of the first two blocks of the formal experiment. The mapping of responses (big/small for Size Judgments, boxy/curvy for Curvature Judgments) to response keys (".>" and "/?") was also counterbalanced across participants.

Outliers removal and preprocessing

One participant with a near-chance overall accuracy (53.2%) was excluded from all further analysis. Overall, accuracy of the remaining 24 participants was above 85% (M = 92.12%, SD = 3.70%). For the remaining participants, we excluded inaccurate trials (8% of trials removed on average across subjects), and iteratively eliminated any trials for which RTs were 3 standard deviation above the overall mean for that subject, condition, and judgment (4% of trials removed on average across subjects).

Analysis

We tested whether both reaction times (RTs) and accuracy were affected by the covariance between size and curvature. To do so, we performed a within-subject repeated measure ANOVA, separately for the Size and the Judgment tasks, and separately for RTs and accuracy, with Size and Curvature as within-subject independent variables. We employed the function ezANOVA in R to fit the within-subject repeated measure ANOVA tests. We confirmed results of our ANOVA by employing mixed-effect modeling, with RTs as dependent variable, size, curvature and their interaction as fixed effects, and subjects and items as random effects. Subjects random effects were modeled in terms of both slope and intercept, while the random effects of Items were modeled in terms of intercept only. This was the most complex mixed-effect model that could reach convergence, which is recommended when fitting mixed-effect models (Bates, 2005). We tested how significantly each predictor and random effect contributed to the model by running a likelihood ratio test comparing nested mixed-effect models, in

which the simpler model has some parameters fixed (compare function for table structures in MATLAB). Two separate mixed-effect models were run, a separate one for each task. The mixed-effect models were computed in MATLAB using the fitlm function.

Exploring S-R mapping: first block. To test whether our results were affected by stimulus-response mapping focusing only on the first block of each subject, before any mapping can take place. Similarly to previous analyses, we perform two separate ANOVAs to predict accuracy and RTs, separately for Size and Curvature Judgments, with Size and Curvature as within-subjects independent variables. We performed the same mixed-effect model analysis, but this time restricting data to the first block. We employed the same procedure as before to establish the number of random effects.

2.6.3 fMRI Experiment

Participants

Sixteen healthy observes with normal or corrected-to-normal vision participated in a 2 hour fMRI session (age range: 22 - 37 years; 6 females). All participants were consented using the procedures approved by the Institutional Review Board at Harvard University.

MRI Acquisition

Imaging data were collected using a 32-channel phased-array head coil with a 3T Siemens Prisma fMRI Scanner at the Harvard Center for Brain Sciences. High-resolution T1-weighted anatomical scans were acquired using a 3D MPRAGE protocol (176 sagittal slices; FoV = 256mm, 1 x 1 x 1 mm voxel resolution; gap thickness = 0mm; TR = 2530 ms; TE = 1.69ms; flip angle = 7 degrees). For functional runs, blood oxygenations level-dependent (BOLD) contrast was obtained using a gradient echo-planar T2* sequence (84 oblique axial slices acquired at 25° angle off of the anterior commissure-posterior commissure line; FoV = 204 mm; 1.5 x 1.5 x 1.5 mm voxel resolution; gap thickness = 0 mm, TR = 2000 ms;

TE = 30 ms; flip angle = 80 degrees; multi-band acceleration factor = 3).

fMRI experiment design

Observers viewed images of big curvy, small curvy, big boxy and small boxy objects in a standard blocked design while undergoing functional neuroimaging. Observers were asked to pay attention to each image, and to press a button when the same image repeated back to back, which occurred once per block. Each run had five 10-second blocks per condition, with ten 10-seconds rest blocks interleaved. Each block consisted of ten images, nine unique and one repeated. Each image was presented for 800ms followed by a 200ms blank. All images were presented in isolation on a uniform white background. The total duration of a run was of 5.33 minutes, and participants performed 6 runs.

Localizers

In addition to the main experiment, we ran three separate Localizers: a Tripartite Localizer, a Retinotopy Localizer, and a Category Localizer.

The Tripartite Localizer was run to identify the animate region and the small and big inanimate object regions. In each run, animals, small objects and big objects were presented in a standard blocked design. Participants had to pay attention to images and respond when an image was repeated, which occurred once per block. Within each run, there were six 12-second blocks per condition, and twelve 12-second rest blocks interleaved. The order of the block presentation was randomized. Each block contained 12 total images, 11 unique ones and one repeated. An image was presented for 800ms followed by a 200ms blank. Each run lasted 5.47 minutes and each participant performed a total of four runs.

The Retinotopy Localizer was performed to identify areas V1-V3, as well as curvature preferring areas. Participants payed attention while being presented with vertical meridians, horizontal meridi-

ans, a small circle (fovea), a large circle (periphery), abstract wavy/curvy patterns and abstract straight patterns. Images for these last two conditions were sourced from Shriasam and colleagues (Srihasam et al., 2014), to compare our results with their observations in the monkey brain. Images were presented in a standard blocked design. Vertical meridians, horizontal meridians, small circle and large circle were presented in seven six-second blocks per run, while the straight and curvy patterns were presented in fourteen six-second blocks. The order of the blocks was randomized. In each run there were 14 8-second rest blocks interleaved throughout. A run lasted a total of 7.73 minutes and participants performed a total of 2 runs.

The Category Localizer was performed to identify body-selective, face-selective, object-selective, scene-selective areas, as well as MT+. Participants watched pictures of headless bodies, faces, scenes, inanimate objects, scrambled objects in a standard blocked design. Participants were asked to detect when a picture was repeated back-to-back, which occurred once per block. There were eight six-seconds blocks per condition in one run. In half of the blocks, the image was fixed in the middle of the screen, in the other half, the image moved across the screen at a random direction. This was done in order to identify movement-sensitive area MT+. Within a 6-second block, there were five unique images and one repeated image. Images were presented for 800ms followed by a 200ms blank. In addition, there were eight eight-second rest blocks interleaved throughout the run. Each run lasted a total of 5.33 minutes, and participants performed a variable number of runs, depending on the amount of time available.

Data Preprocessing

We analyzed functional data using BrainVoyager QX software for MALTAB. Preprocessing included slice scan-time correction, 3D motion correction, linear trend removal, temporal high-pass filtering (0.01 Hz cutoff), spatial smoothing (4 mm FWHM kernel), and normalization into Talairach (TAL) coordinates. Data were down-sampled to 3 x 3 mm voxel resolution. We computed General Linear

Models (GLMs) at the single subject level for texforms and original runs separately. GLMs included square-wave regressors for each condition's presentation times (small curvy, small boxy, big curvy, big boxy), which were convolved with a gamma function to approximate the hemodynamic response.

Retinotopy and Occipitotemporal Mask definition

We created and early visual cortex mask by bilaterally defining retinotopic maps V1, V2 and V3 based on anatomical landmarks on BrainVoyager. This was done as to exclude voxels in early visual cortex from further analysis. We also created an occipitotemporal (OTC) mask as the union of manually defined uber-regions ventromedial temporal cortex (vOTC), lateral occipitotemporal cortex (lOTC), and dorsal occipitotemporal cortex (dOTC), both at the individual and the group-level. The OTC mask did not include voxels from early visual cortex.

Tripartite Regions Definition

Group-level animal, small-object and big-object preferring regions were defined in the occipitotemporal cortex based on the average beta weights across subjects for the Tripartite Localizer. Within the group-level occipitotemporal cortex mask, a voxel was assigned to the condition which had the highest positive beta value.

Voxel-wise Analysis

We perform a voxel-wise analysis where, for each voxel, we look separately at size and curvature preference. We retained all voxels that had positive values for all conditions within the group-level OTC mask, which excluded 1937 out of the total 8087 voxels within the mask.

For each of the resulting voxels, we computed the Size [Small > Large, across curvature] and Curvature [Curvy > Boxy, across size] preference, and plot resulting activation with Size and Curvature

preference as main axes in a Cartesian space. These Cartesian coordinates were then transformed to polar coordinates with given angle and radius, and voxels were color-coded based on such a scale. The number of voxels which showed the same preference profile were plotted in a frequency histogram, and color-coded voxels were displayed in a scatterplot and as a brain surface map.

We tested for uniformity of the circular distribution of the data by performing the Rayleigh test (Zar, 1999) using the circstat toolbox on MATLAB (Berens et al., 2009). Given the bimodal distribution of our data, which presented two diametrically opposite modes, we performed a doubling of the angles procedure to transform it in a unimodal distribution. In this procedure, each angle is doubled, and if its result is equal or larger than 360°, then 360 is subtracted from it. In this way, data were transformed into a unimodal distribution fit for submission to a Rayleigh test. We visualize the voxel locations on the brain surface with voxels color-coded based on their angle and radius in the polar space.

We further look at the voxel-wise preference within the animal, small- and big-object preferring regions of the Tripartite Localizer. Tripartite regions were defined at the group-level, voxels with positive values for all four conditions of the main experiment were selected, and polar frequency histograms and surface maps were visualized using the same procedures are before. Voxels with certain types of interactions would not be visible with the previous analysis. To test the degree to which curvature preference generalized across size, and size preference generalized across curvature, we computed the number of voxels in OTC (masked to only select voxels with positive activation for all four conditions) that showed a curvy>boxy preference, and the number of voxels that showed a boxy>curvy preference. For these subsets of voxels, we computed the proportion of voxels whose curvature preference generalized across size. The same analysis was also performed when factoring by size, that is looking at the proportion of small-object and big-object preferring voxels in OTC which showed the size bias across curvatures.

Typical/Atypical Preference Maps

We were interested in testing whether the typical combinations [small curvy, big boxy] would lead to more reliable neural responses than the atypical combinations [small boxy, big curvy]. To do so, we restricted our investigation to all active voxels for the contrast [all conditions > rest] thresholded at t>2 for the Tripartite Localizer data. We then computed a typical [small curvy > big boxy] and an atypical contrast [small boxy > big curvy] from the main experimental run, masked them, and displayed them on the surface. Maps were thresholded at the same level for easier comparison. We first assessed the reliability of typical and atypical contrasts by performing an individual level odd-even correlation. For each subject, we divided the masked beta maps in two groups, depending on whether they were collected in even-numbered runs or odd-numbered runs, and computed the average activation. Beta maps for the average of odd and even runs were then correlated with each other. We then performed a t-test to assess whether the reliability between the typical and atypical contrast maps was significantly different.

3

Comparing facets of behavioral object representation with brains and models

3.1 Abstract

Here, we explored how well different kinds of behavioral similarity judgments of inanimate objects related to neural response to those objects along the ventral stream hierarchy.

Specifically, we first measured similarity implicitly using a visual search task, measuring the time it took participants to find one target amongst others items. We also measured similarity with an explicit shape-guided sorting task, in which participants dragged and dropped similar items nearby while focusing on shape similarity, and not on other object properties. Finally, we measured similarity with a second sorting task, but this time without giving participants any guidance on what measure of similarity to follow (unguided). Our results show that (i) there are clear differences in the measured similarity space of objects across tasks, (ii) the implicit similarity measured by visual search was most strongly correlated with the posterior half of the ventral visual stream, (iii) the explicit shape-guided sorting similarity was most strongly correlated to anterior half along the ventral stream, and (iv) the similarity measured with unguided sorting task was not well-correlated with any visual system responses, despite it being a highly reliable behavioral similarity space. Deep-net hierarchical layers showed a similar relationship with these three behavioral measures, but with notable divergences: the deep-net lacked a strong shape-based similarity space, and showed moderate correlation with the unguided similarity space in the final output layer.

Broadly, these results highlight that different kinds of similarity can be manifest in different behavioral tasks, and reveal a clear joint in the way the visual system represents the similarity of objects. Based on these data, we hypothesize that this relatively sharp transition along the ventral stream hierarchy may reflect a transition from more pictorial to more structural representation of inanimate object information.

3.2 Introduction

The similarity space of objects has been extensively used as a tool to relate representations among minds, brains, and models. However, the psychological construct of "similarity" is not well defined – objects can be similar in different ways. For example, an orange and a tennis ball are more similar if shape is considered as a measure of similarity, while an orange will be more similar to a banana if semantic category is considered as the measure of similarity. Further, the way in which we define shape similarity is also not obvious, in that shape information is constituted of a multitude of features at different levels (DiCarlo & Cox, 2007). This raises the important question of whether and how these different object similarity spaces relate to the structure of object responses in the brain.

Various studies have collected behavioral measures of object similarity and correlated these measures with neural responses along the ventral stream, utilizing different kinds of paradigms (Kiani et al., 2007; Haushofer et al., 2008; Bracci et al., 2015; de Beeck et al., 2008; Peelen & Caramazza, 2012; Carlson et al., 2014; Bracci & Op de Beeck, 2016; Edelman et al., 1998; Jozwik et al., 2016). One of the most popular paradigms in recent years has been the multiple object arrangement method (sorting task for short; Kriegeskorte & Mur (2012)). In this paradigm, participants are required to arrange an initial set of objects for several trials. Over trials, an algorithm zooms into specific subsets and then computes the ideal similarity space given information collected throughout the trials. Interestingly, when instructing participants to arrange items into a space based on their similarity, without further instructions on what this similarity should be (i.e., it was unguided), similarity seemed to follow features that were more categorical than the ones guiding representational space in monkey IT (Mur et al., 2013). This suggests that the sorting task, when no instructions are provided, leads to people relying more strongly on semantic categorical information (e.g. grouping together food, clothing, furniture, etc.) than on more primitive perceptual features (e.g. grouping together based on aspect ratio, curvature, shape).

What happens if participants are instructed to arrange objects based on a specific measure of similarity? Bracci & Op de Beeck (2016) tested two different variations of the sorting task: while one version asked participants to arrange objects specifically based on semantic similarity, the second one asked participants to arrange objects based on their perceived shape similarity. They also carefully selected a stimulus set to separate as much as possible the shape of depicted objects from their broad semantic category (e.g. animals, fruits and vegetables, sporting equipment, etc.). They found that posterior aspects of the visual system showed a neural similarity structure related to the shape-similarity of the objects, and that more anterior parts of the visual system showed a correlation with both shape and category similarity structures (Bracci & Op de Beeck, 2016; Peelen & Caramazza, 2012; Proklova et al., 2016).

Finally, more recently, Cohen et al. (2017) examined brain/behavior similarity where the similarity was measured with a different paradigm: visual search. This method measures similarity implicitly, by measuring the time it takes to find one object among distractors of another kind (Duncan & Humphreys, 1989; Wolfe & Horowitz, 2017). For example, people are quick to find a picture of a face among pictures of cars and slow to fine that same picture of a face among pictures of bodies—this indicates faces are more similar to bodies than to cars. Using this measure of similarity, they found extensive and strong correlations across entire ventral visual stream. They proposed that there may be one ubiquitous representational space across the entire visual stream. However, they only measured brain and behavioral similarity at the level of object categories, effectively presuming category-level relationships, leaving open whether the search and brain relationships actually hold when considering similarity among individual items.

Taken together, these studies indicate that there are different kinds of similarities among objects. Implicit measurements using visual search speeds and explicit measurements focusing on shape-similarity seem to both account better for much of the ventral visual stream (Bracci & Op de Beeck, 2016; Cohen et al., 2017) than unguided similarity measurements (Mur et al., 2013), which tend to be more seman-

tic rather than perceptual. To date, these measures have not been considered at the same time, in the same participants, over the same stimulus set. Further, the relationship between visual search implicit similarity and shape-guided explicit similarity is completely unknown—do they both correlate with brain similarity measures in the same way? The aim of this study is to measure and compare these three similarity spaces and relate them to the neural responses along the visual hierarchy.

One key challenge of our study is the stimulus set of choice: we showed participants objects that cover a range of categories that was narrower than the one employed in previous studies exploring psychological similarity (Mur et al., 2013; Jozwik et al., 2016). Specifically, we did not cover the animate/inanimate distinction, but focused on inanimate objects. Since representations in occipito-temporal cortex are strongly driven by distinctions at this broader categorical level (Mur et al., 2013; Kriegeskorte et al., 2008), this choice provides a stricter and more nuanced test, as compared to previous literature, of the relationship among these behavioral similarity measures and neural responses.

3.3 Results

3.3.1 Visual Search Task Design

In the visual search task, participants had to detect the odd-one-out target among distractors as quickly as they could in a Mechanical Turk study (Figure 3.1a, see Methods). For this experiment, images (n = 72, all inanimate) were gray-scaled and luminance-matched. The measure of dissimilarity is, in this case, given by reaction time (RT): the faster the target is individuated, the more dissimilar the target is from the distractors (Duncan & Humphreys, 1989; Wolfe & Horowitz, 2017). Constructing a 72 x 72 similarity matrix based on RTs requires the collection of 2556 reliable RT measures, too many for it to be feasible to collect them from one individual participant in a single experiment (or Mechanical Turk "hit"). We thus devised a novel way to collect RTs, in which participants performing a hit were presented with only a sub-sample of the whole possible 2556 target-distractor pairs, which consisted of 12 pairs (objects were presented both as target and as distractor within each experiment; Figure 3.1b). From pilot data, we established that, to obtain reliable data, we needed at least 27 individual RT measures per object pair (averaged across 12 presentations per hit), which amounted to 828,144 trials. Thus, we produced a total of 5,751 separate Mechanical Turk hits, with 12 object pairs each, and we ensured that there was no repetition of the same object among these 12 object pairs within an experiment. Further, we ensured that object pairs overlap was not more than 50% for any two of the 5751 Turk hits. A total of 1,272 participants performed the experiment (individuals could perform more than one hit). We eliminated individual subject variability through mixed-effect modeling, by modeling subjects as random intercept. This proved to greatly increase the reliability of the data, that went from 0.15 to 0.54 (average of 50 split-half iterations).

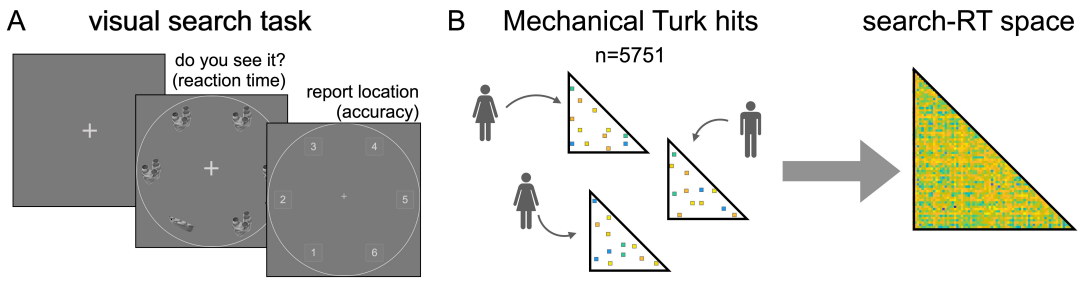


Figure 3.1: Visual Search Task Design. a): visual search task display: after a fixation dot (500ms), six objects appeared in a circular array. Only one of these objects was different (target) while all the others were similar (distractors). Participants had to press the space-bar as soon as they identified the target location. After their response, the images were masked and participants had to report on the keyboard where the target was located. b): Visual Search Experiment Design. The 72 x 72 matrix of 2,556 comparisons was divided into 5,751 separate Mechanical Turk hits in which 12 object pairs were presented. Each object in a pair was presented both as target and as distractor within one hit. Since participants could perform more than one individual hit, the final number of participants (1,272) is smaller than the number of individual hits (5,751). Once all 5,751 hits were collected, we obtained 27 average RTs for each of the 2,556 combinations, which pilot data established was sufficient to obtain reliable data.

3.3.2 MDS plots

To understand which were the main dimensions that drove RT differences among object pairs, we mapped these differences into a Cartesian space through multidimensional scaling (MDS; mapped into 2 dimensions; figure 3.2a left). When visualizing the MDS plot, we can easily identify subdivisions based on objects' mid-level visual features such as curvature or elongation. Further, even though images were gray-scaled and luminance-matched, from the plot it appears that there could also be an influence of lower-level visual features (e.g. pixel area). Thus, this space seems to contain visual feature information of low-to-intermediate complexity. How did this implicit search space compare to the spaces produced with explicit similarity measurements?

To explore this question, we collected similarity ratings during a sorting task (Kriegeskorte & Mur, 2012; Bracci & Op de Beeck, 2016), in which participants were asked to arrange objects based on similarity inside a circular arena. In one experiment, participants were asked to arrange objects by shape (shape-guided sorting space). In the other, they were not given specific instructions about which sim-

ilarity measure to use (unguided sorting space). Given past literature, we expected the shape-guided space to lead to more perceptual dimensions, and the unguided sorting space to lead to more categorical dimensions (Mruczek et al., 2013; Bracci & Op de Beeck, 2016).

When mapping the dissimilarities into a 2-dimensional space with multidimensional scaling, we find that for the shape-guided sorting space (split-half reliability: 0.82; Figure 3.2a center), clustering was primarily based on shape similarity: a squared-to-boxy object gradient emerged, as well as a gradient from objects with rounder shapes to objects with elongated parts (replication of Bracci & Op de Beeck (2016)). From this qualitative analysis, there did not seem to be a strong difference between this space and the search-RT one. However, the correlation between the shape-guided sorting space and the search-RT one was relatively low ($r = 0.34$), suggesting that these two spaces were more dissimilar than what could be inferred from a qualitative observation of the MDS plots.

On the other hand, a qualitative evaluation of the MDS plot produced by the unguided sorting space revealed this space to be more categorical. Indeed, there were higher-order divisions as compared to the search-RT and the shape-guided sorting space (split-half reliability: 0.73; Figure 3.2a right), with clustering based on semantic categories such a foods, musical instruments, clothing items, cleaning supplies. The dissimilarity between the unguided sorting and the search-RT was also reflected by their low correlations ($r = 0.12$). Correlation between the unguided sorting and the shape-guided sorting spaces was 0.29, suggesting that they were also largely different.

3.3.3 Searchlight

After evaluating the three behavioral spaces qualitatively, we sought to relate them to the representational space of the occipito-temporal cortex. To do so, we employed a searchlight analysis, in which we correlate behavior to the similarity space for small clusters of neighboring voxels throughout OTC. The result of this correlation is then assigned to the voxel at the center of a given searchlight. In this way, a map of correlation values is created, representing the degree of similarity between behavioral

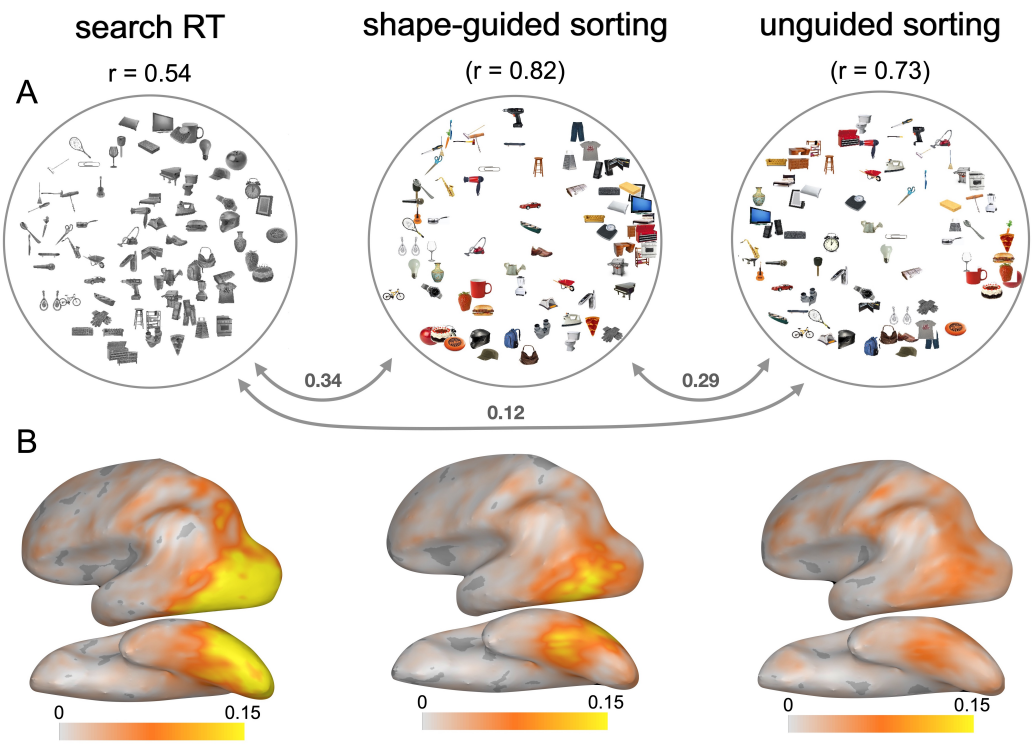


Figure 3.2: MDS plots and Searchlight Analysis. a): MDS plots for the search-RT, the shape-guided sorting and the unguided sorting spaces. Above the plots is displayed the reliability for each task (iterative split-half Pearson correlation). Values above the arrows correspond to Pearson's correlations between spaces. b): Searchlight Analysis results: surface maps of correlations between searchlight neural RDMs and behavioral RDMs. For easier comparisons, all three maps of behavior-brain correlations were thresholded leniently to $0 < r < 0.15$.

and neural RDM at each given point. Results of these analyses can be observed in Figure 3.2b, in which we adopt a lenient threshold which allows the evaluation and comparison among the different searchlight maps.

Literature on the neural representations underlying the occipitotemporal cortex suggests that more perceptual similarity spaces, such as the search-RT and the shape-guided sorting space, are a closer match to neural representations throughout OTC, while the more categorical unguided sorting space would best correlate with more anterior parts of the ventral stream (Mur et al., 2013; Bracci & Op de Beeck, 2016; Peelen & Caramazza, 2012).

When evaluating the fit between the neural responses and search-RT space (Figure 3.2b, left), we find that the search-RT space was extensively correlated with neural patterns in early visual cortex, and extended to more posterior regions of the ventral, lateral and dorsal occipitotemporal cortex. This is partly in line with what was observed by our group with a search task at the category level (Cohen et al., 2017), although in our case we found stronger correlations in early visual cortex, and correlations did not extend quite as anteriorly. This could be explained by differences in the experimental paradigm: while the current paradigm is an odd-one-out search task, Cohen et al. (2017) employed a template-matching task. Further, the similarity space in the current paradigm is at the item-level, rather than the category-level, so it is reasonable to assume that the features underlying the current similarity space tend to relate more to low-level information than the ones of Cohen et al. (2017).

When relating the shape-guided sorting space to the neural space (Figure 3.2b, center), the shape-guided sorting task shows strong correlations with ventral and lateral regions of OTC that seemed to fall more anteriorly than the strongest correlations from the search-RT searchlight analysis. Compared to correlations emerging from the search-RT searchlight map, no strong correlations were found either more dorsally on OTC, or in earlier visual processing regions. This result is in line with our evaluation of the MDS plot and suggests that the visual features underlying the shape-guided sorting space tend to be more complex than the ones defining the search-RT space, possibly representing a holistic

representation of the complex shape of an object, rather than specific features of it such as curves or edges.

When comparing the unguided sorting space to neural responses in a searchlight analysis (Figure 3.2b, right), we find that the unguided sorting space correlated overall less strongly and extensively as compared to the search-RT and the shape-guided sorting space, with strongest correlations in anterior patches of lOTC and vOTC.

Overall, these observations are in line with literature showing encoding of intermediate level features throughout OTC (Hong et al., 2016; Srihasam et al., 2014). Further, these results replicate previous observations that the representational space produced by an unguided version of the multi-arrangement task produces more semantic and categorical distinctions (Mur et al., 2013; Jozwik et al., 2016). Finally, they lend support to previous literature pointing to categorical and semantic information being stored more anteriorly in the ventral stream (Peelen & Caramazza, 2012; Proklova et al., 2016; Bracci & Op de Beeck, 2016).

After a qualitative evaluation of the differences among the three behavioral spaces in how well they related to each other and to OTC, we assessed that the three behaviors seemed to represent similarities among objects at different levels of complexity. Next, we sought to explore this by again relating behavior to neural responses, but this time focusing on the ventral stream, and by selecting a series of neighboring "slices" that went from more posterior to more anterior coordinates. Further, we also related our behavioral spaces to layers of a DNN, and compared the fit of behavior with both brain and models. In this way, we could observe how well different stages of visual processing related to behavior.

3.3.4 Relating behavior to hierarchical stages in brain and models

Posterior-to-Anterior Neural Gradient: Correlations

We sought to explore how well the behavioral spaces related to neural populations in the ventral stream along a posterior-to-anterior gradient. To do so, for each individual, we produce partly overlapping "neural slices" of object-responsive voxels at different y coordinates (6mm thickness) along the ventral stream, going from more posterior to more anterior regions. We then produce the neural RDM for the voxels contained within each given slice, and correlate it with the three behavioral RDMs. Results of the correlations are plotted to the left of Figure 3.3a, with the split-half reliability of each neural slice in a dashed gray line.

The plot shows a clear dissociation in the way the behavioral spaces relate to neural slices from more posterior to more anterior coordinates: while more posterior slices show higher correlations with the search-RT space, more anterior regions show higher correlations with the shape-guided sorting one. On the other hand, the unguided sorting space did not seem to show strong correlations at any point in the gradient.

These results point to what was already suggested by the searchlight analysis: while more posterior ventral regions encode visual features at a lower- or intermediate-level as reflected by the implicit task, more anterior ventral regions are more strongly correlated with complex shape features like the ones represented in the shape-guided sorting task.

Posterior-to-Anterior Neural Gradient: Unique Variance

What this analysis did not tell us was the degree to which the correlation between a behavior and a neural slice was unique to that behavioral space, or whether it was also shared with the others. It is possible that whatever is shared in common by the behavioral spaces composes a big part of the correlation between the ventral stream and each of the tasks. Or, it could be that one or more of the

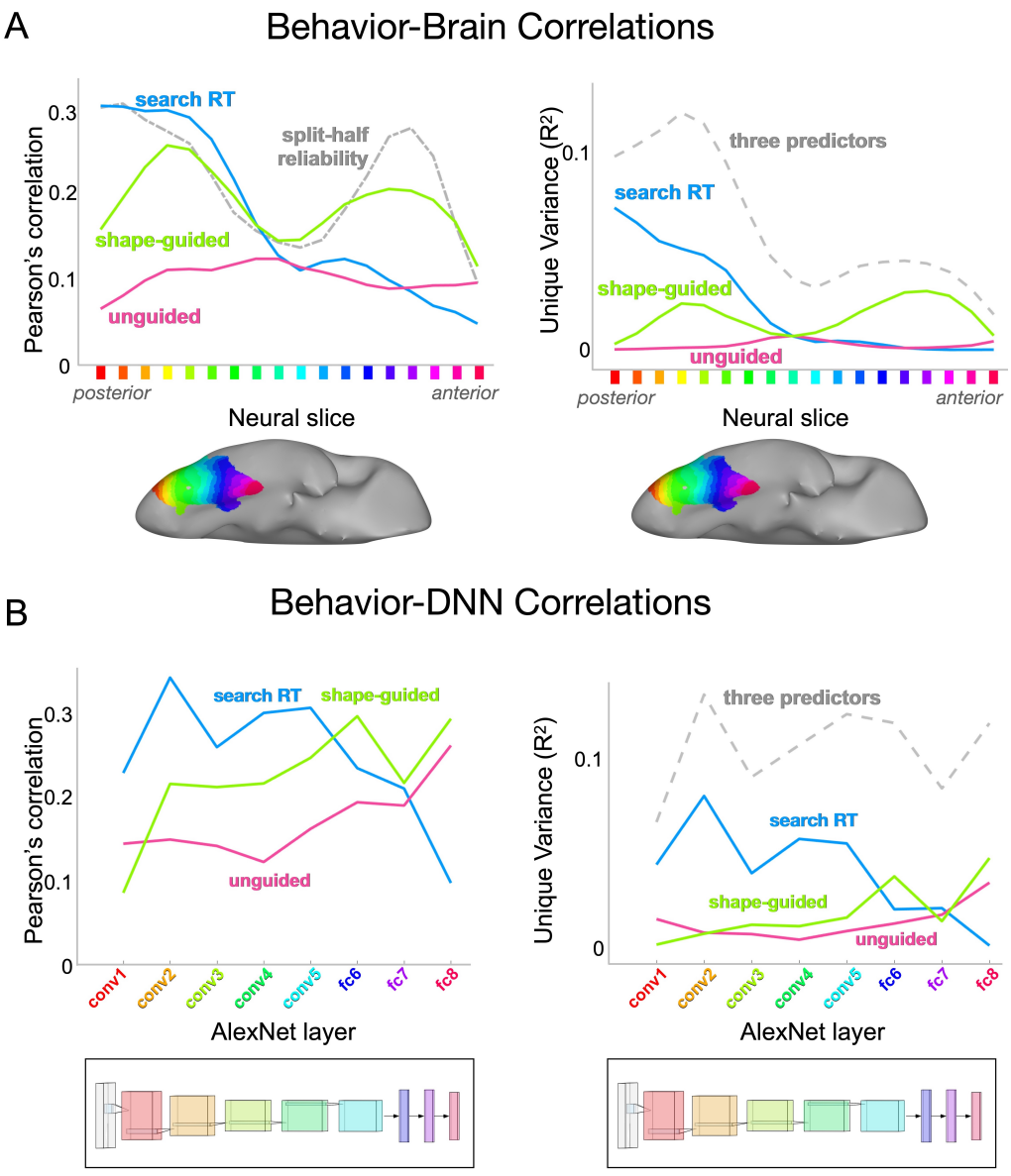


Figure 3.3: Relating behavior to hierarchical stages in brain and models. a): Left: Pearson's correlations between behavioral task RDMs and neural slice RDM, from more posterior to more anterior position (see Methods). Gray dashed line: split-half reliability of neural slice. Right: unique variance explained by one task for a given neural slice, when accounting for the explained variance that is shared with the other two tasks. Gray dashed line: explained variance at each slice for a model with all three tasks as predictors. Bottom of graphs: depiction of neural slices in ventral stream for an example subject. b): Left: Pearson's correlations between behavioral task RDMs and AlexNet layers RDMs. Right: unique variance explained by one task for a given AlexNet layer, when accounting for the explained variance that is shared with the other two tasks. Gray dashed line: explained variance at each layer for a model with all three tasks as predictors. Bottom of graphs: summary description of AlexNet architecture.

behavioral spaces correlates to the brain in a mostly unique way.

To explore these possibilities, we computed the amount of unique variance (R^2) explained by each model separately by performing a commonality analysis (see Methods). The plot on the right of Figure 3.3a shows the unique variance explained at each neural slice by the three separate behavioral models, as well as the variance explained by a model with all three behavioral spaces as predictors (common and shared variance; in a dashed gray line).

From this plot, it is further clear the dissociation between posterior neural slices, in which search-RT contributes in explaining more unique variance, and anterior neural slices, in which shape-guided sorting contributes in explaining more unique variance. To the degree to which the search-RT space and the shape-sorting space contain features of different complexity, this reflects a difference in feature complexity along the ventral stream. To further explore our intuition that this switch is due to a difference in degree of feature complexity between the behavioral spaces, we are next going to perform this same analysis, but correlating the behavioral spaces with RDMs produced from a hierarchical convolutional deep neural network.

DNN layers: Correlations

To assess how well all three behaviors related to the DNN layer hierarchy, we measured the activation of AlexNet (8 layers, 5 convolutional and 3 fully connected, pre-trained on ImageNet) to all 72 objects from our Dataset, and produced 8 RDMs (one per layer, based on Euclidean distance). Then, we computed the correlation between each layer's RDM and the three similarity spaces (see Methods; Figure 3.3b, left). The correlations show that the search-RT space was most correlated with early to mid-layers of the DNN, suggesting that this space contained visual features of low- and intermediate-level complexity, in line with what we previously observed qualitatively with the MDS plots. On the other hand, the shape-guided sorting space showed a slower increase in correlation as the hierarchical stage of the model progressed. However, its performance did not fully surpass the one of the search-

RT until the very end. Finally, the categorical unguided sorting space showed the lowest correlation overall. Like the shape-guided sorting task, its correlation also reached a peak only at the last read-out fully-connected layer 8 of the DNN.The same profile of results was observed when weighting the behavioral spaces by the reliabilities so as to make them comparable across tasks (see Appendix C, Figure C.1).

DNN layers: Unique Variance

After looking at the simple correlations, we explored the degree to which each individual layer was uniquely explained by each of the three behavioral models by employing commonality analysis, with the same approach employed earlier for the neural slices (see Figure 3.3b right). The commonality analysis further replicates our observations: the search-RT space is the one contributing the most unique variance, and it does so in the earlier and intermediate DNN layers. In comparison, neither the shape-guided sorting, nor the unguided sorting spaces seem to contribute a lot of unique variance. Specifically, the shape-guided sorting space does not contribute much in uniquely explaining later layers of the DNN, as it would have been expected given the brain fits (Figure 3.3a).

Overall, these results confirm that our behavioral spaces reflect features of different complexity. Specifically, while the search-RT space best relates to low- and mid-level visual features, the sorting spaces reflect more complex features. This is partly in line with both the searchlight and the posterior-to-anterior gradient analysis: indeed, earlier stages of visual processing and earlier layers of a DNN were better reflected by the lower- and intermediate-level features of the search-RT space. However, while we find that later visual processing stages better reflected the complex shape features of the shape-guided sorting space, this space did not strongly correlate with later DNN layers. This mismatch in representations between later visual processing stages and later DNN layers lends support to past work showing that DNNs classification performance, unlike human object recognition, relies more on information on object texture than on object shape (Geirhos et al., 2018; Long & Konkle, 2018).

3.4 Discussion

Here, we have explored the features underlying explicit and implicit behavioral similarity measures.

In line with past literature, by visualizing the similarity spaces with multidimensional scaling, we found that the search-RT and the shape-guided sorting tasks reflected object features along more visual dimensions as compared to the unguided sorting ask (Mur et al., 2013; Cohen et al., 2017; Bracci & Op de Beeck, 2016). However, the search-RT space seemed to reflect perceptual features at a lower- and intermediate-level, while the shape-guided sorting space contained more complex shape-related features. The unguided sorting space seemed to organize objects along more categorical and semantic dimensions.

After qualitatively assessing the differences among these similarity spaces, we evaluated how well they reflected neural responses in the ventral stream. We found that the two behavioral similarity spaces containing more visually-based features well-reflected different hierarchical stages in the ventral stream. On the other hand, when evaluating our behavioral models fit to a DNN, we found only a partial correspondence with the ventral stream: while earlier stages of processing were well-correlated with the search-RT space in both brain and model, later stages of processing had a good fit with shape-guided sorting space only in brain, while no behavioral space showed a satisfactory fit with later DNN layers.

In results reported in Appendix C, we have tested how well individual AlexNet layers, AlexNet as a whole, and our three behavioral models performed in predicting broad neural sectors of OTC with weighted representational modeling and found predictive performances largely fitting our observations (see Figure C.2), but failing to identify the nuances that could be observed when exploring neural populations at a smaller scale. Future work could employ this predictive modeling approach in order to explore the fit between brain and behavior, and brain and AlexNet, at a smaller scale (i.e. through a searchlight or a posterior-to-anterior gradient analysis).

Next, we're going to address separately the implications of our findings for the representational spaces underlying the ventral stream, compare our results with previous findings, and discuss what they may suggest about how information is computed in DNNs as compared to the brain.

3.4.1 Representations in the ventral stream

Given past literature, we hypothesized that more categorical information was going to more closely reflect neural responses more anteriorly in the ventral stream (Peelen & Caramazza, 2012; Proklova et al., 2016; Bracci & Op de Beeck, 2016), and more perceptually-based information could show fits both in anterior and posterior regions of the ventral stream (Bracci & Op de Beeck, 2016; Cohen et al., 2017; Hong et al., 2016; Long et al., 2018). However, no study had so far tested the explicit and implicit shape similarity judgments one against the other in how well they reflected neural responses in the ventral stream.

By relating the behavioral spaces to the brain through a searchlight analysis and with a posterior-to-anterior gradient along the ventral stream, we observed that the search-RT space better reflected earlier visual processing regions, while the shape-guided sorting space better reflected later visual processing regions. On the other hand, the more categorical space of the unguided sorting task did not show high correlations anywhere along the ventral stream.

While our group in the past theorized a common similarity space shared across all OTC (Cohen et al., 2017), these results conflict with such a possibility, at least when considering item-level similarity spaces. More likely, our analyses suggest that the search-RT and the shape-guided sorting space may highlight two sub-spaces in the ventral stream involved at two different levels of visual processing (for theoretical ideas on how similarity spaces might develop along the ventral stream, see DiCarlo et al. (2012)): an earlier stage, which is best reflected by the implicit similarity measure, is involved in the detection of pictorial information about objects such as curvature and shape. At some point during visual processing, a radical transformation of information would occur following which object infor-

mation could become more complex, and resistant to simple transformations. In this later stage, the explicit similarity judgments of the shape-guided sorting task would best reflected the neural populations space.

3.4.2 Comparison with previous findings

Why did we find different results from Cohen et al. (2017) when relating brain and visual search task? First, Cohen and colleagues employ a template-based version of the visual search task to test category-level responses in high-level visual cortex (Cohen et al., 2017). In the version of the task we adopted, participants did not prepare a mental representation of the target object at the beginning of a trial. Second, we looked at item-level RT differences among 72 objects, for a total of 2556 target-distractor comparisons, rather than at category-level ones. Thus, the search-RT space as produced by the present study paradigm might contain different features distinguishing objects at the item level, and relate differently to representations throughout OTC. It is thus not surprising to observe relatively more posterior fits to the ventral stream in the case of this item-level search-RT space.

The present work is also partly in contradiction with findings from Bracci & Op de Beeck (2016) where, employing similar unguided and shape-guided paradigms, they found that shape and category information coexisted in more anterior regions of the ventral stream. In the present study, we find that information in these areas seems to primarily relate to shape, and that category information is virtually non-existent. Future work could get at this more directly by explicitly instructing participants to sort based on semantic categories.

3.4.3 Relating DNNs and the brain

While the shape-guided space was a good match for the most anterior parts of the ventral stream, it did not perform comparably well in explaining later layers of a DNN. This result is supported by recent

literature suggesting that most DNNs might be biased towards classifying texture rather than shape information, thus accomplishing object recognition in a different way from how the brain performs it (Geirhos et al., 2018; Long & Konkle, 2018).

The tasks developed here might be useful for future work seeking to diagnose the underlying shape-space of DNNs, and could potentially be employed to train DNNs at performing object classification in a way that is closer to how the brain performs it.

3.5 Conclusion

The present work sought to compare different behavioral paradigms in how well they reflected neural responses along the ventral stream, as well as testing how well they related to similarity spaces of increasing complexity produced from layers of a DNN. We demonstrate that a reliable similarity space which well-reflects early and intermediate visual processing stages can be achieved with a new, large-scale online odd-one-out visual search paradigm. On the other hand, we also show evidence that explicit similarity judgments in a sorting task paradigm are strongly affected by the type of instruction given, with similarity judgments based on more visual features leading to better fits for later stages of visual processing.

Overall, these results give further insight into the representational stages underlying the processing of visual information along the ventral stream, and point to differences in the way models and brain might accomplish object recognition.

3.6 Methods

3.6.1 Stimulus Set

For both the fMRI experiment, the sorting tasks and the visual search task, we employed the same 72 pictures of inanimate object categories on white background ranging different sizes. Objects were presented in color on a white background for the fMRI experiment and the sorting tasks. For the visual search task, we removed the background, gray-scaled and luminance-matched the objects to remove some of the effects of low-level visual properties which might affect reaction times.

3.6.2 fMRI experiment

Participants

Thirteen healthy observes with normal or corrected-to-normal vision participated in a 2 hour fMRI session (age range: 19 - 32 years; 8 females). Three participants were excluded for excessive motion, for a total of 10 subjects that underwent analysis after preprocessing.

Functional runs

We presented the 72 objects on a display as participants lied in the scanner during a mini-block design. A total of 10 participants were collected, and 8 six-minute runs per participant. Each run included 72 mini-blocks for all objects, and 4 15-s rest periods interleaved throughout. During each run, participants were asked to think about the objects presented in a different way, specifically to: 1) how to interact with them, 2) where they're found, 3) what they're for or 4) their physical properties. The purpose of this manipulation was unrelated to the goals of this study. In addition to thinking about those properties, participants were also asked to detect when an object was surrounded by a frame.

MRI Acquisition

Imaging data were acquired on a BioSpin MedSpec 4T scanner (Bruker) using an eight-channel head coil. Functional data were collected using an echo-planar 2D imag- ing sequence (TR, 2000ms; TE, 33ms; flip angle, 73°; slice thickness, 3mm; gap, 0.99mm, with 3 x 3 in-plane resolution). Volumes were acquired in the axial plane parallel to the anteroposterior commissure in 34 slices, with ascending interleaved slice acquisition.

Data Analysis

Functional data were analyzed using Brain Voyager QX software and MATLAB. Preprocessing included slice scan-time correction, 3D motion correction, linear trend removal, temporal high-pass filtering (0.01 Hz cutoff), spatial smoothing (6 mm FWHM kernel), and transformation into Talairach (TAL) coordinates. General linear model analyses included square-wave regressors for each condition's presentation times, convolved with a gamma function to approximate the hemodynamic response.

3.6.3 Behavioral experiments

Unguided and Guided sorting task

For both experiments, participants were required to arrange objects displayed on a computer screen inside a circular arena based on their similarity. This experiment aims to compute a measure of psychological similarity as a function of distance, thus assuming a geometric model of mental representation (Kriegeskorte & Mur, 2012).

The experiment followed an adaptive trial design, in which all 72 objects were presented at the first trial, and consequent trials vary based on the information gathered and the dissimilarity evidence available so far (see Kriegeskorte & Mur (2012) for further details and code). At the end of the first

trial, the distance between the objects corresponds to an estimate of the representational dissimilarity matrix (RDM), which is then recomputed at the end of each trial based on the available evidence. The subset presented at the following trial is determined based on the current information available from the estimated RDM. Rather than considering absolute distances, the scaled distances between objects were used to estimate the RDM.

The whole experiment lasted one hour plus time for rest, which occurred every eight trials, and could last as long as participants wanted. Before starting the experiment, participants performed a practice trial on eight geometrical and abstract shapes to familiarize themselves with the interface.

Two different versions of the experiment were performed, and unguided version to record categorical similarities between objects, and a guided version to record shape similarities between objects.

In the unguided version of the experiment, no specific instructions on the kind similarity to consider was provided. This followed the classic paradigm followed in past studies and it's known to lead to a representational space which is more categorical and semantic (Mur et al., 2013; Jozwik et al., 2016).

In the guided version, participants were given more precise instructions, and asked to arrange objects specifically by shape similarity, ignoring alternative categories such as color or function.

Participants and Preprocessing: Sorting Tasks

We collected 26 participants for the unguided, and 25 participants for the guided version of the study. Subjects were excluded if the Fisher z-transformed average correlation with other subjects negatively exceeded a z-score threshold of 2. After preprocessing, 23 participants for the guided and 25 for the unguided were retained. A group RDM was then computed by averaging the single-subject RDMs.

Visual Search Task

We collected reaction times (RTs) during a visual search task as a measure of target-distractor similarity (Duncan & Humphreys, 1989). During the visual search task, participants had to fixate on a fixation cross at the center of a circular arena on a computer screen. After 500s, 6 images of objects were then displayed on a circular array, all the objects were the same (distractors) except on (target), subjects were required to respond as quickly as they saw the distractor by pressing a key (the space-bar). All images were presented on a gray-scale to prevent the predominance of hue over shape and curvature (Alexander et al., 2019).

Reaction time was recorded from object presentation to the space-bar press. After the space-bar was pressed, objects were masked for 250ms. If more than 5 seconds passed without a response, a warning appeared and the experiment was paused until participants were ready again which was signaled by a button press. Once they were ready, the same trial they did not respond to was repeated.

If participants responded within the 5 second interval with a space-bar, the images were masked for 250ms with noise and then six numbered rectangular frames appeared in place of the images, with numbers from one to six. Participants had to report where the object was in the array by pressing the correct number, and if they did not respond within 10 seconds a warning appeared, urging them to repeat the trial again when ready, which they signaled by a button press. If they responded incorrectly, another warning appeared and the next trial was delayed by 3 seconds, which motivated participants to respond accurately. If participants responded correctly, then the array disappeared, and the next trial resumed after 500ms.

In an individual experiment, each object pair was presented both as target and as distractor six times (once per position in the array), for a total of 12 individual trials per object pair, and target-distractor asymmetries were later averaged to compute an individual mean RT for a specific object pair.

Pilot data revealed that reliable data could be obtained with at least 27 individual mean RTs per

object pair. Obtaining 27 individual mean RTs for each of the 2556 object pairs, each of which was presented 12 times, amounts to a total of 828,144 trials.

We decided to divide all conditions in a counterbalanced way among 5751 separate trial sets, employing Amazon Mechanical Turk. Participants could perform more than one individual trial set.

Each trial set consisted of 12 object pairs, and we ensured that there was no repeating object among these 12 pairs within a trial set. We also ensured that there was no more than 50% overlap of conditions (i.e. object pairs) among all 5751 trial sets, to later improve the removal of individual subject variability.

Participants and Preprocessing: Visual Search

- Trials Trimming. For each participant (n = 1,272), we removed trials in which they responded incorrectly, reporting the target at the wrong location. Then, we iteratively removed trials whose z-transformed RT was above 2.5: at each iteration, the highest z-scores above 2.5 was removed, and the z-scores recomputed without that data point. The exclusion process terminated when there were no more z-scored RTs above 2.5. For each subject, we excluded a whole condition/pair if more than 33% of the trials (i.e., if more than 4 out of the 12 trials) were missing.

- Target-Distractor Asymmetries. After testing on pilot data that the average highest between-pair difference between fastest and slowest response was higher in magnitude than the average highest target-distractor difference, we decided to average across targets and distractors RT for the same pair, and, in this way, were able to construct a 72 x 72 symmetric RT matrix.

- RT Matrix Normalization. To account for between-subject differences without losing between-group differences, we submitted the log-transformed 2556 RT pairings to a mixed-effect model, with the target-distractor pair as predicting fixed-effect variable and subject as predictive random-effect variable. Note that we collapsed across target-distractor asymmetries, thus each object in

a pair was target in half of the data points and distractor in the other half. In this way, we were able to remove the individual differences as random-effect from the "true" differences in RT based on target-distractor dissimilarity. To ran the linear mixed-effect model by employing effect coding, in which the mean RT for each level of the categorical variable "target-distractor pair" is compared to the grand mean. This resulted in 2555 fixed-effect coefficients and an intercept, as well as a random-effect of the intercept for each subject. Since in effect coding the constraint is that coefficients must sum up to zero, we were able to compute the last coefficient as the difference between zero and the sum of the 2555 coefficients. We the added the intercept value, which corresponds to the average normalized RT, to the 2,556 fixed-effect coefficients in order to make all values positive.

- Dissimilarity Matrix construction. To build the dissimilarity matrix for the fixed-effect coefficients of the RTs, we subtracted from the overall highest value each single coefficient. No further transformation was needed, as the difference between fixed-effect coefficients corresponds to their perceived dissimilarity, with lower values corresponding to lower RTs and larger target-distractor difference, and vice versa.

3.6.4 Deep Neural Network setup

We submitted our Dataset to AlexNet, an 8-layer deep convolutional neural network pre-trained with ImageNet (1.2 million images), and extracted the features for each layer. AlexNet comprises 8 layers, 5 convolutional and 3 fully connected ones. For the five convolutional layers, features were summed across the image plane. We computed 72 x 72 RDMs for each layer based on Euclidean distance.

3.6.5 Reliability estimation

To have an estimate of the reliability of our spaces, we performed an iterative split-half correlation. Reliability was measured by correlating using Pearson's correlation, plotting.

- Search Task. To get an estimate of reliability, we ran an iterative split-half correlation for both log-transformed average RT coefficients and for the fixed-effects resulting from the mixed-effect model, for a total of 50 iterations. At each iteration, the Dataset was split into two random halves with an equal number of subjects for each group and submitted to two separate mixed-effect models. The fixed-effect coefficients resulting from the two models were then correlated, employing both Pearson's r and Kendall's tau. Finally, the correlation values were averaged across iterations to get an estimate of test-retest reliability. The average correlation between fixed effects across iterations was compared to the average correlation between the log-transformed RTs for the two sets. If the mixed-effect model did indeed remove individual subject noise, the fixed coefficients should have been more highly correlated between the fixed effects than the log-RTs. The fact that we did observe the fixed effect correlations to be higher points to the soundness of our normalization method.

- Sorting tasks. Reliability and noise ceiling estimates were computed by iteratively (n=1000). dividing the subjects in two separate groups, averaging the RDMs and correlating them using Pearson and Kendall's tau correlation separately.

3.6.6 Exploring the Behavioral Spaces: MDS plots and correlations

To have an estimate of the similarity of our models' spaces, we computed the pairwise correlations. In addition, we qualitatively estimated the main dimensions of the representational spaces by employing multidimensional scaling (MDS; criterion: metric stress; distance measure: Euclidean distance for guided and unguided sorting task, RT distance for visual search).

3.6.7 Relating behavior and brain: Searchlight Analysis

We explored how our behavioral models related to neural representations in OTC more locally by employing searchlight analysis.

To do so, for each individual subject, for every voxel in the gray matter, we computed the neighboring voxels within a spherical searchlight area of 3 voxel radius (mean number of neighbors $\approx$ 113). For each sphere, we then compute the neural RDM (distance: 1 - correlation). Then, each searchlight is correlated separately with the dissimilarity matrices from the three space (search-RT, unguided sorting and shape-guided sorting).

Average correlations were computed across subjects to produce three separate group maps, one per searchlight analysis (Figure 1b,c and d). All three maps from the three behavioral models were thresholded at $0 > r > 0.15$.

3.6.8 Posterior-to-Anterior Neural Slices Construction

To assess whether there was overall difference in how well our three behavioral models reflected neural responses in the posterior-to-anterior gradient, we wanted to compare our behavioral RDMs to neural RDMs for "neural slices" of voxels going from more posterior to more anterior regions of the ventral stream.

To build our neural slices, we first produced a mask from the intersection between OTC, drawn at the individual level on BrainVoyager based on anatomical landmarks, and the 1000 most active voxels for the contrast [All Objects > Rest] at the group-level.

To produce a "neural slice," for a given Talairach y coordinate, we extracted the voxels which were in the ±3mm range from that coordinate in the mask. We then produced, for the voxels included at each slice, a neural RDM based on Euclidean distance, and the values were re-scaled in the zero-to-one range.

Each neural slice RDM was then correlated with the three separate behavioral RDMs, and the resulting Pearson's correlation values were displayed in a linear plot with neural slice on the x-axis, and Pearson correlation on the y-axis.

To produce group-level neural slices RDMs, we averaged neural slices sharing the same y coordinate across participants. All participants shared common y coordinates with the others for most of their neural slices: there were 18 out of 18 shared neural slices/coordinates for three participants and 18 out of 19 shared neural slices/coordinates for six participants. One participant shared 17 out of 18 neural slices/coordinates with the others.

Of the participants with additional slices/coordinates, two participants had an additional neural slice more posteriorly than the others, and other four participants had an additional neural slice more anteriorly than the others. For these cases, the additional neural slice RDM was averaged with the neighboring slice.

This resulted in 18 separate RDMs per participant, and the ones sharing the same y coordinates were averaged together across participants to produce group-level RDMs.

3.6.9 Relating behavior to hierarchical stages in brain and models: Pearson's Correlations

- Neural Slices. We plotted the Pearson's correlations between group-level neural slice RDMs and behavioral RDMs, and overlaid them with the iterative split-half reliability for the given neural slice (1,000 iterations).

- AlexNet layers. We assess the similarity between our three behavioral spaces and the 8 separate DNN layers by computing the Pearson's correlations between each of the behavioral RDMs and each of the DNN layer RDMs. To allow easier comparison of similarity between DNN and behavior across tasks, we also weighted the correlations by dividing them for the individual

reliabilities.

3.6.10 Relating behavior to hierarchical stages in brain and models: Commonality Analysis

We then computed the amount of unique variance contributed by each separate behavioral model in explaining a given neural slice or DNN layer. To do so, we performed a commonality analysis (Groen et al., 2018), in which we ran four separate regression models with the RDM for a given neural slice or AlexNet layer as predicted variable.

Of the four models, one model had all three behaviors as predictors (search-RT, shape-guided sorting and unguided sorting), and the other three had only two of the behaviors as predictors, to satisfy all pairwise combinations of the three behavioral spaces. To compute the amount of unique variance contributed by one of the behaviors (e.g., search-RT), we subtracted the R squared (R^2) for the model with the two other behavioral spaces as predictors (e.g., shape and unguided sorting) form the R^2 of the model with all three behavioral spaces as predictors. This was done to compute the unique variance of each behavior. The unique variances so computed were then displayed in a linear plot, and the R square for the three-predictor model was overlaid in a dashed gray line.

4

Conclusion

In this dissertation, I have investigated how visual information about objects is organized and represented in the visual cortex. To do so, I have employed two primary approaches: in Chapters 1 and 2, I have explored the way in which the natural covariance of size with lower- and higher-level properties affects the large-scale organization of objects in occipitotemporal cortex; in Chapter 3, I looked at the finer-grained differences among object representations at the item-level, and how their representations change along the ventral stream.

More specifically, the first chapter tested the relationship between size and manipulability, and how the natural covariance between these two properties is represented in higher-level visual cortex. I found that manipulability and size are both encoded in high-level visual cortex, and that they are not independent from each other. Rather, small-object regions are driven by manipulable small objects. Post hoc analyses further showed that a property related to stability might be a strong driver for object representations in large-object preferring regions.

After looking at the relationship between real-world size and the behaviorally-relevant property of manipulability, the second chapter investigated the relationship of real-world size and the mid-level visual property of curvature. I found that, also in this case, the relationship between size and curvature was not orthogonal, but that encoding of object information was affected by the natural covariance between these two properties.

Finally, the third chapter looked more closely at the way in which the representational space is transformed along the ventral stream, by testing different behavioral similarity spaces, and finding that two behavioral measures of perceptual similarity best related to different processing stages along the ventral stream.

These findings inform questions concerning (i) what gives rise to the organizational principles underlying object representations in human visual cortex, and (ii) what is the representational space and what are the features underlying this organization. Here, I am going to discuss these issues in turn.

4.1 What drives the organization in high-level visual cortex?

In this section, I will discuss how these findings leverage on two alternative explanations for what may drive the large-scale organization of OTC.

On the one hand, the finding that the natural covariance between size and curvature affects neural responses (Chapter 2) gives support to the "visual bias" hypothesis. According to this hypothesis, the

origin of the organization of objects representation in OTC stems from regularities in the real-world natural statistics which would then bias and constraint the way information is encoded from earlier to later stages of visual processing. That is, since size tends to relate in a specific way with objects' visual features, these regularities would then affect the visual input, and the way this information is transferred and processed at later stages (Konkle & Oliva, 2012b; Srihasam et al., 2014).

On the other, results unveiling the relationship between size, manipulability and spatial stability (Chapter 1) are in line with the "connectivity hypothesis," according to which the organization of high-level visual cortex may be driven by long-range connections with behaviorally-relevant networks, such as networks involved in action or navigation (Mahon & Caramazza, 2011).

While the two chapters seem to defend diametrically opposing views, it should be noted that the size x manipulability stimulus set employed in Chapter 1 did not correct for low- and mid-level visual features which may covary with size and manipulability. Thus, bias from earlier stages of visual processing cannot be excluded to account at least in part for this organization, and I actually believe it to play an important role.

However, there are multiple ways in which low-level visual biases could constraint the mapping and organization of object representation in the ventral stream. One is through the pre-set innate structure of visual cortex: that is, manipulability and navigability information may covary meaningfully with visual low- and mid-level features such as curvature and eccentricity, and thus lead to this information being mapped and constrained by innate visual processing pathways (Konkle & Oliva, 2012b; Srihasam et al., 2014).

Alternatively, experience we have with our surrounding environment in the early years may lead to the shaping of visual processing information: the behavioral relevance of both manipulability and navigability lends support to the possibility that our attention might be drawn to the features most diagnostic of these properties, which would not only be shaping visual cortex information innately, but also through learning mechanisms (Long et al., 2019).

The extreme consistency across subjects on the location of both large-scale and category-selective regions would suggest that an underlying proto-map of innate visual biases is to the least already in place (Srihasam et al., 2014). On the other hand, experience has also an important role, as is proven by the existence of category-selective regions for words, which could not have feasibly emerged through evolutionary processes (Dehaene et al., 2010). It is thus likely that these forces together are involved in driving the observed organizational structure.

While visual biases do seem to play a role in shaping the organization of the visual cortex, long-range connectivity with behaviorally-relevant areas has also been reported. In support of the connectivity account small-, big- and animal-selective regions have been shown to connect to separate sub-networks (Konkle & Caramazza, 2016), and it is indeed possible that the tuning of OTC for small manipulable and large stable objects might be partly driven by these long-range connections. However, further studies need to be performed in order to understand whether this connectivity is innately present (Mahon & Caramazza, 2011), or emerges with pruning during development in response to differential tuning from the visual cortex (Konkle & Oliva, 2012b; Konkle & Caramazza, 2016; Innocenti & Price, 2005). On this second account, the long-range connectivity would thus follow from the organization of high-level visual cortex driven by visual biases and experience (Konkle & Oliva, 2012b; Srihasam et al., 2014), rather than being its cause. Future work exploring the neural development of connectivity could help clarify whether such connections are built-in or pruned following experience.

Overall, this thesis does not confirm neither of these hypotheses, and does not clarify whether representations in high-level cortex are driven more by proto-maps and neural blue-prints, or by experience and neural pruning. However, it paves the way for further research into these areas, by proving what had already been hypothesized from the first observation of a large-scale organization of OTC by real-world size: that size constitutes an interpretable, high-level descriptor for a multitude of complex and covarying dimensions which, together, lead to the shaping of representations in occipitotemporal cortex (Konkle & Oliva, 2012b).

4.2 Representational changes along the ventral stream

This thesis primarily deals with the featural space(s) underlying the organization of object representations in OTC, and what dimensions they may contain.

Related to the hypothesis of retinotopic mapping extending to higher-order regions in driving visual object organization, recent studies have shown encoding of mid- and intermediate-level features all along the ventral stream: not only at earlier stages of visual processing in more posterior regions, but also more anteriorly, alongside invariant object representations and categorical information (Hong et al., 2016; Bracci & Op de Beeck, 2016). Similarly, our group as previously advanced the hypothesis of a stable and widespread representational space underlying the whole occipitotemporal cortex that was well-predicted by visual search behavior (Cohen et al., 2017). This literature would suggests a certain degree of homogeneity in the representational space underlying the visual cortex.

However, when comparing the neural representational space to different behavioral similarities in Chapter 3, such a homogeneity was not observed. Indeed, I found that two perceptual item-level similarity measures that seem to relate to, respectively, lower-level (i.e. pictorial, related to edges and contour) and higher-level (i.e. related to complex object shape) visual features best reflected neural responses at different points along the ventral stream, respectively more posteriorly and more anteriorly.

This implies that the representational space along the ventral stream might not be as homogeneous as previously observed: there rather seems to be a point along the ventral stream in which the underlying shape-space is radically changed, presumably through pooling and pruning of information, which allows for higher-level visual information to emerge, but that does not completely throw out all lower-level visual features information (Hong et al., 2016). More specifically, the reasons for such a transition could signify a change from a stage in which the encoded information is more pictorial and related to the object's contour, to a stage in which it becomes more structural and invariant to

transformations (DiCarlo et al., 2012).

Interestingly, while I did find good fits between the lower-level implicit similarity space and intermediate DNN layers, the fit was not as good for the higher-level similarity space and later layers of the model. Deep-neural networks constitute the state-of-the-art in the modeling of visual cortex, and have been a powerful instrument in the interpretation and exploration of the representational spaces underlying the ventral stream (Güçlü & van Gerven, 2015; Yamins et al., 2014). However, more recent studies have observed idiosyncrasies between human judgments and DNNs that hint to differences in the underlying computations and representations: a growing body of literature shows how applying small perturbations to images that leave human judgments unaffected result in misclassifications for models (Geirhos et al., 2017). Differently from the visual cortex, they seem to relate more on texture rather than shape in order to accomplish object recognition (Long et al., 2019; Geirhos et al., 2018).

Further study and investigation of the implicit measure of similarity I developed, along with perceptual and categorical explicit similarity measures, could help us further understand the similarity space underlying the ventral stream, and the transformations this space goes through at later stages of visual processing, as well as help us build models that more closely match human judgments.

4.3 Conclusions

This dissertation contributes to the field by exploring the influence of covarying properties in shaping neural responses to object size in high-level visual cortex. I found that, when looking at the large scale organization by size, following the natural covariance between size and manipulability/stability on one side, and size and curvature on the other, led to stronger and more stable neural representations.

This dissertation does not directly test opposing theories on what may have driven this organization to begin with (innate long-range connectivity blueprints? Innate visual biases from earlier visual regions?). However, by showing that neural responses are tuned by the natural covariance of size with

other properties, this thesis paves the way to future research on this area.

Further, by comparing the neural representational space to different measures of similarity, I explored the content of this space and developed a new behavioral paradigm with which it may be possible to test and probe computational models of human vision.

Overall, this work has given new insights into how the brain can accomplish the incredible feat of transforming and representing information in such a way as to allow us to see the world that surrounds us.

A

Appendix to Chapter 1

A.1 Supplementary ROI Analyses

A.1.1 Methods

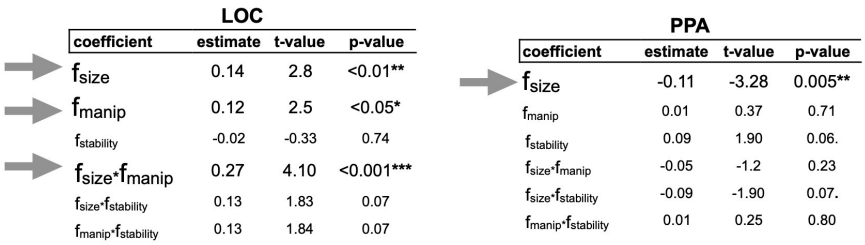
Large-TOS and Categorical ROIs selection

In addition to the small-ITG and large-PHC, we identified left and right large-object selective TOS (Large-TOS , see Figure S1a; n=15 for left hemisphere, n=13 for right hemisphere; average size of 226 voxels).

To connect our main findings with previous literature, we also defined gray-matter masked spherical ROIs (12mm radius) around the peak of activation for object-selective LOC and scene-selective PPA and OPA. PPA and OPA were localized in each participant using the contrast of Scenes > Objects from the functional Localizer runs (n=18), LOC was localized in each participant using a contrast of Objects > Scrambled from the same runs (n=17). The average size was 228 voxels for LOC, 257 voxels for PPA and 248 voxels for OPA. LOC was localized in all but one subject due to a technical issue related to stimuli presentation.

ROIs overlap

When comparing the location of the spherical ROIs, there was systematic overlap between Small-ITG and LOC, between Large-PHC and PPA and between Large-TOS and OPAKonkle & Oliva (2012b). To quantify this, ROI overlap was computed by extracting, for each subject, the number of voxels shared between Small-ITG and LOC, between Large-PHC and PPA and between Large-TOS and OPA. The average of the shared voxels was then computed across subjects and divided by the average extension of LOC, PPA or OPA across subjects. On average, 33 percent of left LOC overlapped with left Small-ITG and 8 percent of right LOC overlapped with right Small-ITG: right Small-ITG was located more anteriorly and dorsally than its left counterpart, while left and right LOC were rela-



LOC

coefficient	estimate	t-value	p-value
f_{size}	0.14	2.8	<0.01**
f_{manip}	0.12	2.5	<0.05*
$f_{stability}$	-0.02	-0.33	0.74
f_{size}*f_{manip}	0.27	4.10	<0.001***
f_{size}*$f_{stability}$	0.13	1.83	0.07
f_{manip}*$f_{stability}$	0.13	1.84	0.07

PPA

coefficient	estimate	t-value	p-value
f_{size}	-0.11	-3.28	0.005**
f_{manip}	0.01	0.37	0.71
$f_{stability}$	0.09	1.90	0.06.
f_{size}*f_{manip}	-0.05	-1.2	0.23
f_{size}*$f_{stability}$	-0.09	-1.90	0.07.
f_{manip}*$f_{stability}$	0.01	0.25	0.80

Figure A.1: LOC and PPA: regression results.

tively symmetrical. Additionally, 47 percent of left PPA overlapped with Large-PHC while 50 percent of right PPA overlapped with right Large-PHC. Finally, 13 percent of left OPA overlapped with left Large-TOS and 22 percent of right OPA overlapped with right Large-TOS.

A.1.2 Results

LOC and PPA.

We observed a pattern of results mostly convergent with our main results when running regression models in LOC and PPA with fsize, f_{manip} and $f_{stability}$ as predictors (see A.1).

Similarly to what observed with Small-ITG, LOC was significantly associated to f_{size}, f_{manip} and their interaction (see Table A.1, left). This result differs only slightly from small-ITG, in which there was no significant contribution of f_{manip} alone.

PPA was significantly associated to f_{size} and marginally to the interaction between f_{size} and $f_{stability}$ (see Table A.1,right). This result is largely in line with the one observed in Large-PHC.

Large-TOS and OPA.

We found a significant contribution of f_{size} and of the interaction between f_{size} and $f_{stability}$ in Large-TOS (Figure A.2b), but inspection of scatterplot and barplot of activation to individual objects didn't

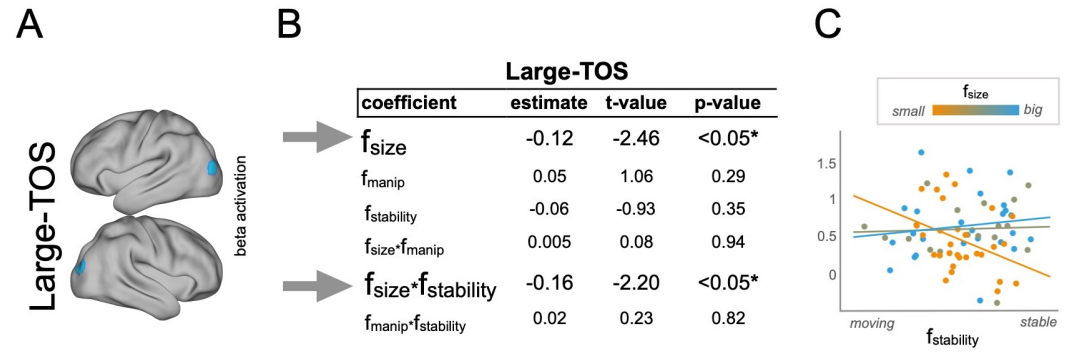


Figure A.2: Large-TOS: ROI analysis and scatterplot.

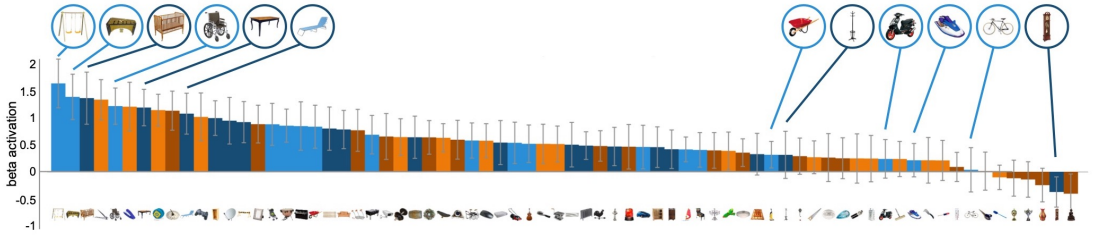


Figure A.3: Large-TOS: Barplot.

reveal a clear direction of the interaction (Figure A.2c and Figure A.3). Finally, OPA was only significantly associated to f_{size}, without other factors or interactions significantly contributing (Table A.4).

A.2 Supplementary Searchlight Analyses

When performing the searchlight analysis, we observed additional sub-regions which we did not cover in the main analyses, as the model fits tended to be lower. Here, we plot the activation within these sub-regions for each of our 72 objects.

Figure A.5a depicts the sub-region composed of small-object preferring voxels which show a positive f_{size}*$f_{stability}$ interaction, showing a preference for moving over stable big objects, and an overall trend in preferring moving over non-moving across object size. This area might be partly overlapping

OPA

coefficient	estimate	t-value	p-value
f_{size}	-0.11	-2.17	<0.05*
f_{manip}	0.04	0.87	0.39
$f_{stability}$	0.03	0.50	0.61
f_{size}*f_{manip}	0.00	0.03	0.98
f_{size}*$f_{stability}$	-0.07	-0.96	0.34
f_{manip}*$f_{stability}$	0.08	1.15	0.25

Figure A.4: OPA: Regression results.

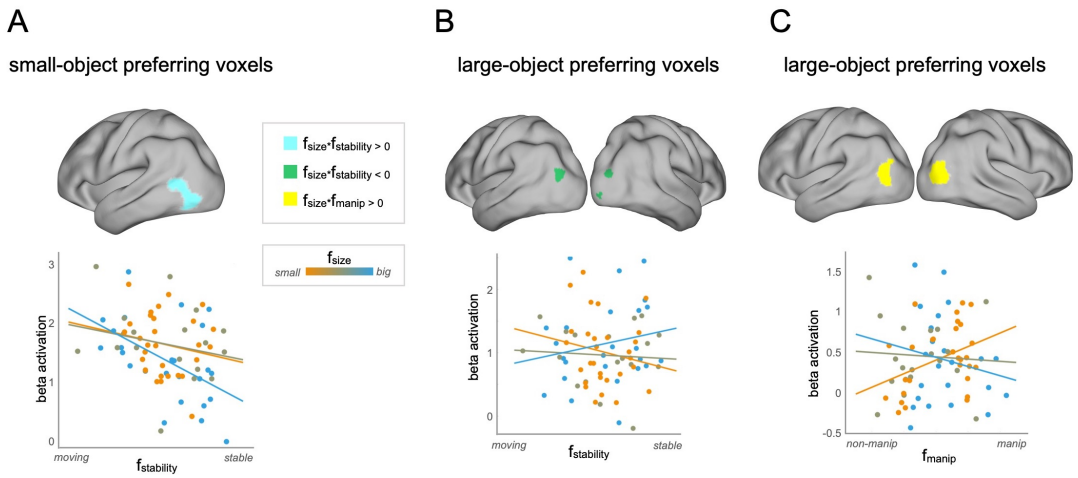


Figure A.5: Supplmenetary Searchlight Results.

with region MT+ which is known to show a preference for moving over non-moving stimuli.

Figure A.5b depicts the sub-region composed of large-object preferring voxels which show a negative f_{size}*$f_{stability}$ interaction, showing a preference for stable over moving big objects. Curiously, this region seems to be partly overlapping with Large-TOS and OPA.

Figure A.5c depicts the sub-region composed of large-object preferring voxels which show a positive f_{size}*f_{manip} interaction, showing a preference for manipulable over non-manipulable small objects, and a trend towards preferring non-manipulable over manipulable small objects.

B

Appendix to Chapter 2

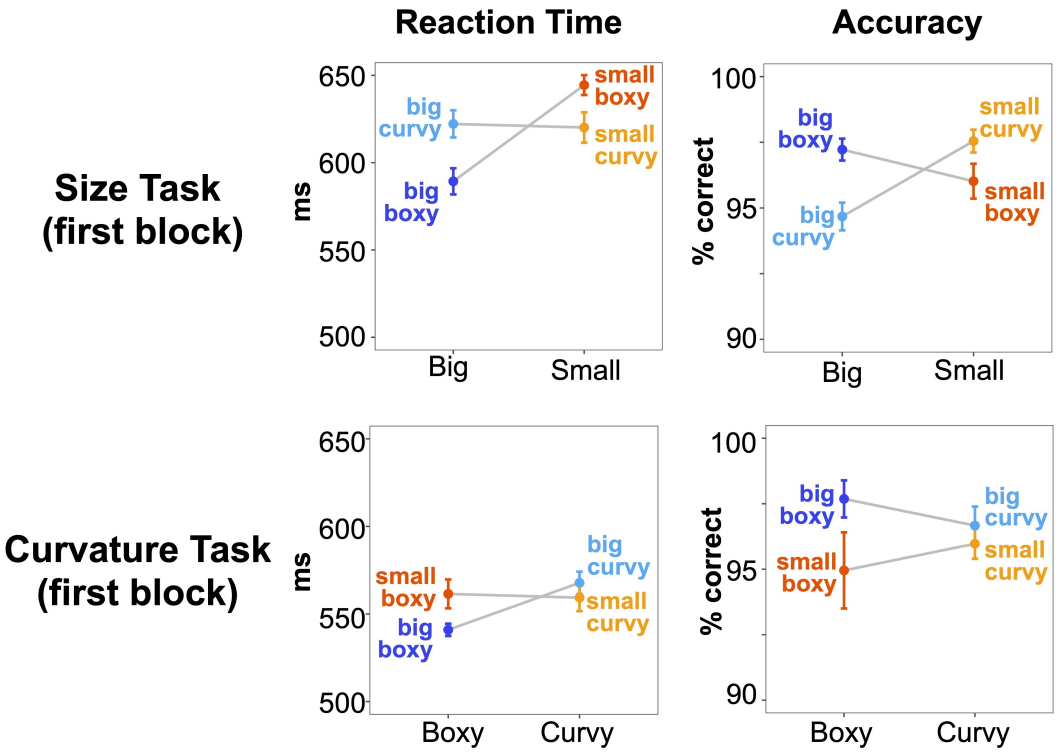


Figure B.1: Behavioral Results form First Block. Interaction plots for reaction times (RT; in milliseconds) and accuracy (in percent correct), for the Size and the Curvature Tasks when only the first block is considered. Error bars reflect the within-subject standard error (n = 12).

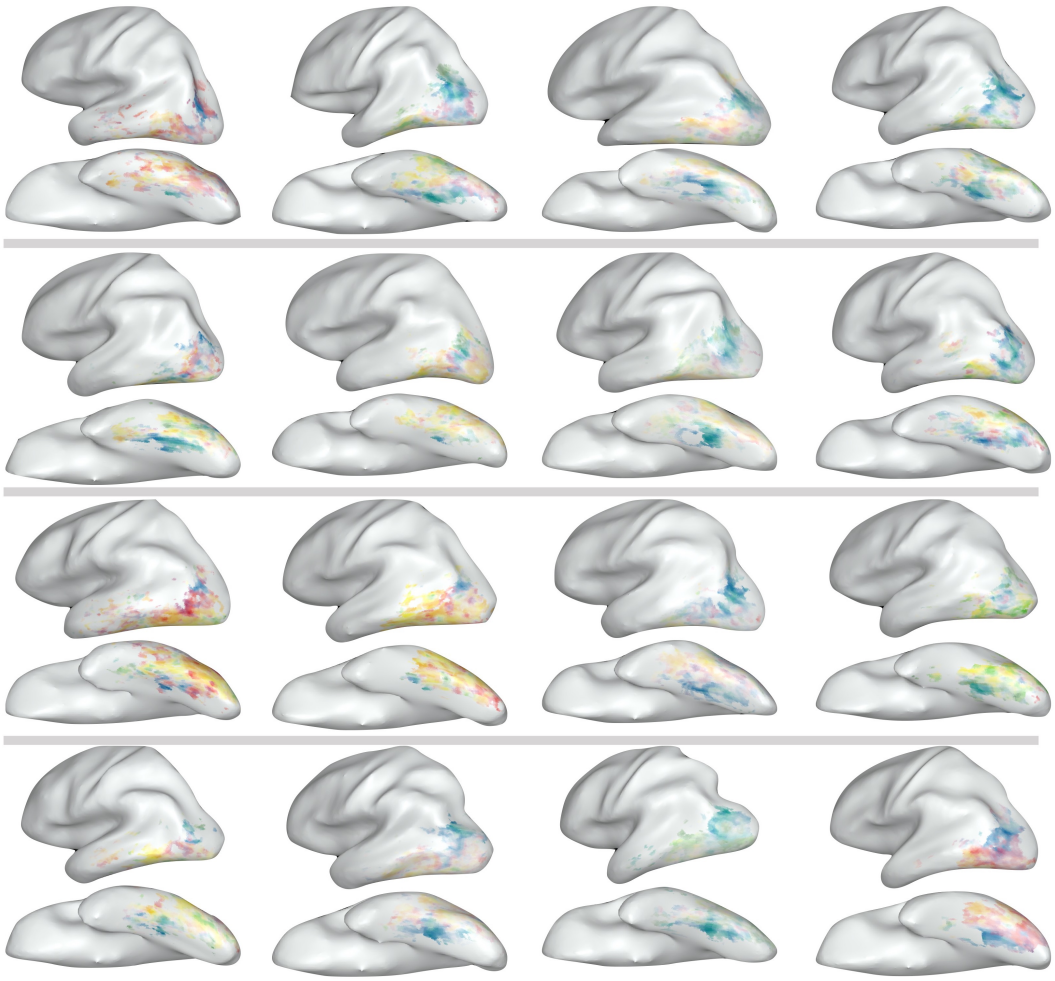


Figure B.2: Individual Voxel-wise Preference Maps for Size and Curvature Preference. Individual surface maps plotting the location of each single voxel in OTC color-coded based on its size and curvature preference following the scheme on Figure 2.3.

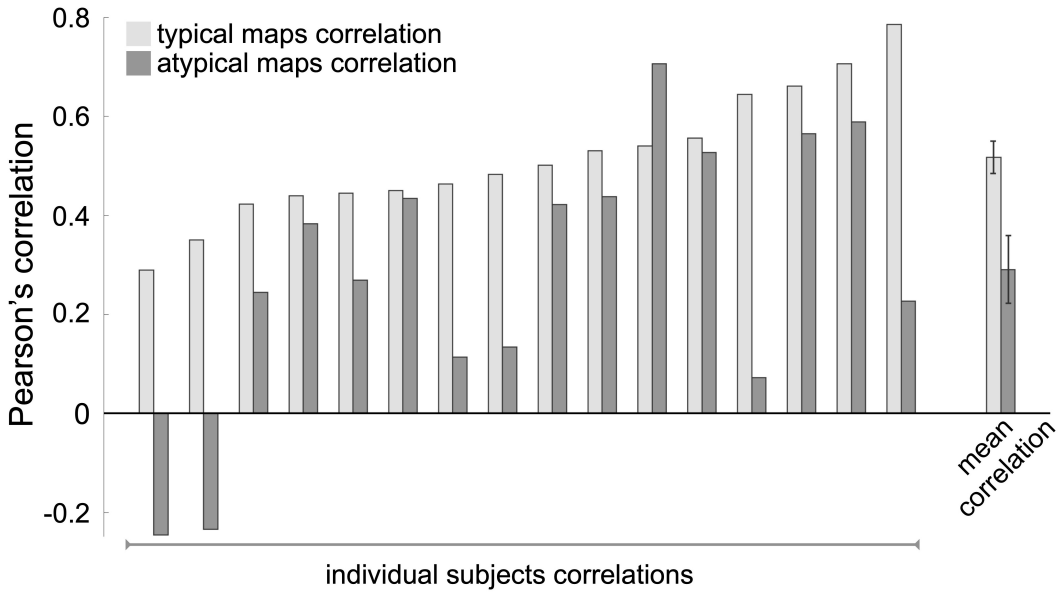


Figure B.3: Individual-level Odd-Even Correlations for Typical and Atypical Maps. Bar-plots of individual and average odd-even runs correlations for the typical and atypical maps. The mean correlation bars (rightmost) were computed by averaging the values across subjects, and are the same plotted in Figure 2.5b of the main manuscript. Error-bars represent standard error (n = 16).

C

Appendix to Chapter 3

C.1 Supplementary Methods – Predictive Modeling of Brain and Behavior

C.1.1 Neural sectors selection

Four broad neural sectors were selected based on anatomical landmarks. Early Visual cortex (EarlyV), which included areas V1, V2 and V3, was based on retinotopy maps for a contrast of horizontal versus vertical lines. The ventral stream was divided among two broad sectors, ventral and dorsal occipital-temporal cortex (vOTC and lOTC). Finally, a neural sector was selected for the dorsal stream, spanning occipital and parietal lobe (occipital-parietal cortex, OPC).

C.1.2 Similarity matrix construction

Brain

For each sector, the mean activation for all objects was computed and intersected with the gray matter mask. Then, the top 1000 most active voxels were selected from each sector, and used a to compute a 72 x 72 RDM for that sector using correlation distance. This resulted in four separate 72 x 72 similarity matrices, one per neural sector.

AlexNet

We computed feature-level RDMs for each layer based on Euclidean distance separately for each feature in each layer.

C.1.3 Predictive modeling

To obtain a predicted estimate of the true model from our predictive models (individual DNN layers, collapsed DNN layers, or behavioral models), we employed linear regression with 9-fold cross-validation, in which data is divided into 9 groups, and 1/9th of the data is removed at each iteration,

the beta weights are computed with the retained data, and a prediction of the removed data is produced with the beta weights values that have just been computed. At the end of the process, beta weights computed at each iteration are averaged, and an average predictive matrix is produced by combining together the predictions computed for the corresponding conditions.

This method allows us to obtain a measure of the degree to which a model is able to generalize and predict new data it was not trained to explain, it is thus a more powerful tool than simple regression. We employed non-linear least square regression (NNLS) to compute the beta weights.

Predicting behavior with DNN layers

We built nine separate model to predict our behavioral similarity spaces: 8 models consisted of the feature RDMs from each of the 8 DNN layers separately, the ninth model consisted of all RDM features from all layers. This resulted in a total of 3 (behavioral spaces) x 9 separate models. The predicted variable was the average RDM across subjects for the sorting tasks, and the RT space for the visual search task. The predictive variables for each model were the RDMs corresponding to the individual features for that layer (dissimilarity measure: Euclidean distance). The model corresponding to layer conv1 of AlexNet thus comprised 96 predictors, conv2 comprised 256 predictors, and so on.

Predicting brain with DNNs and behavior

Each of the four neural sectors was predicted with a total of 12 separate models:

- 8 DNN layers models (similarly as above for predicting behavior),
- one model with all RDM features of all layers used as predictors,
- 3 behavioral models.

The predicted variable was the average neural RDM for that specific neural sector.

Noise Ceiling for Neural sectors. Noise ceiling was computed with an iterative split-half Kendall correlation procedure, in which at each iteration individual subject RDMs were split into two random groups, averaged and correlated (n iterations = 1000).

C.1.4 Inferential analysis of model performance

Inferential analyses was also performed with the RSA toolbox developed by Nili et al. (2014). The predictive power of a certain model is established by computing the Kendall's rank correlation coefficient tau between each candidate RDM (the RDMs constructed based on each candidate model) and the reference RDM. Error bars on the bar plots indicate the standard error, estimated by a condition bootstrapping procedure (1000 iterations), which simulates the variability of the estimates across condition samples.

C.2 Supplementary Results – Predictive Modeling of Brain and Behavior

C.2.1 Predicting Behavior with DNNs as models

The results of weighted representational modeling when predicting behavior with DNNs can be seen in Figure C.1. When predicting the search-RT space, we can observe a similar profile to the one observed with simple Pearson's correlation (Figure 3.3b): also in this case, earlier and intermediate layers performed best at predicting this implicit task compared to the explicit tasks (Figure C.1 left). When collapsing all layers together in one unique model, performance was not greatly improved.

Also in line with our main results, when predicting the two sorting tasks (Figure C.1 center and right) we find that for both spaces later, fully connected layers perform better as compared to earlier layers, with performance reaching noise ceiling. When collapsing all layers in one unique model, performance showed a slight improvement in both cases, suggesting that to some degree different layers predicted different aspects of the spaces.

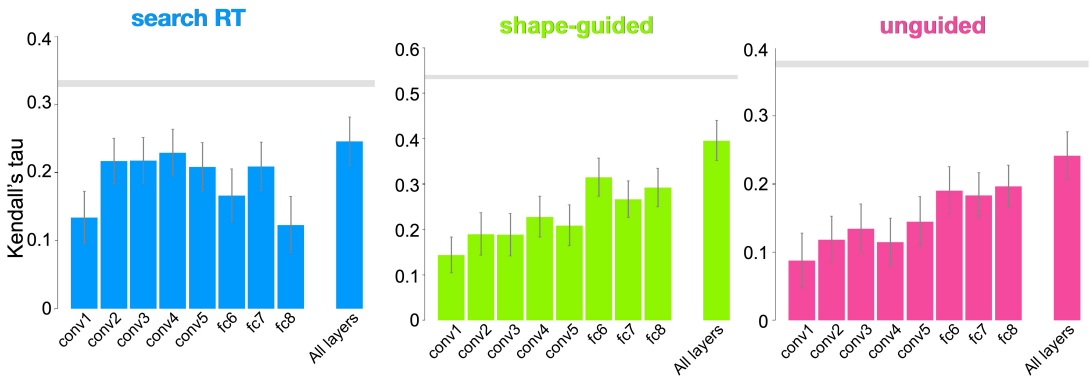


Figure C.1: Predicting Behavior with DNNs as models. Kendall's tau correlations between reference RDMs (left: search-RT, center: shape-guided sorting; right: unguided sorting) and candidate RDMs. RDMs were reconstructed with weighted representational modeling, in which the missing data at each iteration is reconstructed with the computed beta weights. Candidate RDMs were reconstructed based on weights given to the feature RDMs composing each model. Each individual AlexNet layer was either considered as an individual model, or all layers were collapsed together in one unique model (rightmost bar of each barplot). The reference RDM for the unguided sorting and the shape-guided sorting is the average RDM across all participants. Error bars were computed with a condition bootstrapping procedure (n = 1,000). Noise ceiling was computed with an iterative split-half Kendall correlation procedure, in which at each iteration individual subject RDMs were split into two random groups, averaged and correlated (1,000 iterations for sorting tasks, 50 iterations for search task). The final correlation value was the average across all iterations.

C.2.2 Predicting Brain with DNNs and behavior as models

While we investigated in the main analyses how well the three tasks related to neural responses in OTC, we did so by employing simple correlations for local, small clusters of voxels, or for slightly larger sections of cortex in a posterior-to-anterior gradient analysis.

However, we also sought to investigate how the behavioral measures would relate to neural spaces at a larger scale, by employing a stronger inferential approach, and to relate their performance to DNNs, which are currently the state-of-the-art in predicting the neural representational space. Thus, we explored how the search-RT space, the two sorting tasks and DNNs are able to predict broad neural sectors of OTC.

For this analysis, we built neural RDMs for each of four broad neural sectors of OTC: early visual cortex (EarlyV), lOTC, vOTC, and occipital parietal cortex (OPC). We then employed weighted representational modeling to predict the neural RDM with as models:

- individual layers of AlexNet, with the features within each layers modeled as separate predictors;
- all layers of AlexNet collapsed together as a unique model, with all features of all layers modeled as separate predictors;
- a separate model for each of our three behavioral spaces, with only on predictor per model.

Results are displayed in Figure C.2.

EarlyV

In Early visual cortex (far left in Figure C.2), neither the DNN layers separately, nor the behavioral spaces performed particularly well, as the correlation between predicted and real RDM was far from reaching noise ceiling. When looking at the predictive performance between the DNN layers, contrary

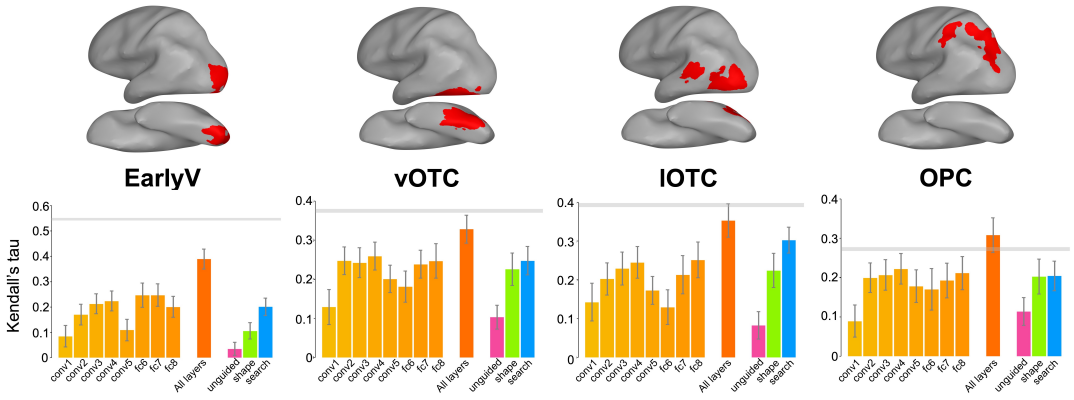


Figure C.2: Predicting Brain with DNNs and behavior as models. Top: Neural sectors for one exemplar subject. For each sector, beta weights from only the top 1000 most active voxels were retained for the contrast [All Conditions > rest]. From left to right: Early Visual Cortex (EarlyV), ventral occipito-temporal cortex (vOTC), lateral occipito-temporal cortex (lOTC), occipito parietal cortex (OPC). Bottom: Correlation between reference RDM (the neural RDM for that neural sector, averaged across subjects) and the candidate RDM coming from different models (light orange: individual AlexNet layers; dark orange: AlexNet layers combined together; pink: unguided sorting space; green: shape-guided sorting space; blue: RT space). Error bars were computed with a condition bootstrapping procedure. Noise ceiling was computed with an iterative split-half Kendall correlation procedure, in which at each iteration individual subject RDMs were split into two random groups, averaged and correlated. The final correlation value was the average across all iterations (n = 1,000).

to what would be expected, earlier layers did not perform better than later ones. Further, all DNN layers together performed significantly better, but still did not reach noise ceiling.

These results suggest that none of our models well reflects the neural shape-space at this early stage, and specifically it does not seem to be the case that early layers of the DNN contained the same low-level visual features that EarlyV seems to care about.

vOTC and lOTC

In ventral and lateral occipitotemporal cortex (vOTC and lOTC, second and third from left in Figure C.2), both individual DNN layers, the shape-sorting space and the search-RT space perform reasonably well, but do not reach noise ceiling. The unguided-sorting space, on the other hand, does not perform particularly well compared to the other spaces. When looking at the predictive performance

between the DNN layers, we found comparable predictive performance for intermediate and later layers. When collapsing all DNN layers together, we find that performance approaches noise ceiling.

These results suggest that the features being represented both by the DNN layers and by two of our behavioral models better resemble the shape-space in this broad neural sectors. Further, given the relatively poor performance of the unguided-sorting space, these features seem to be more related to perceptual properties than to categorical and semantic information.

Finally, the fact that performance improved when collapsing AlexNet layers together suggests that visual features of different complexity, from intermediate- to higher-level might be represented in this areas, in line with our main results.

OPC

Finally, in OPC (far right in Figure C.2), similarly to what observed for vOTC and lOTC, performance of individual DNN layers, of the shape-guided sorting and the search-RT spaces was good. In this case, performance of all models was closer to noise ceiling. Also in this case, DNN layers at multiple hierarchical stages performed similarly well, and collapsing all layers together led to better performance of the DNN.

Overall, these results suggest that the similarity space in OPC, like to the ones in vOTC and lOTC, can be interpreted as being composed of perceptual features from multiple hierarchical stages.

Overall, performance of our model was significantly better in higher-order visual areas as compared to EarlyV. Although most DNN layers performed significantly better than chance, we did not find any layer to perform particularly well when modeled separately, and no layer was able to reach noise ceiling. When modeling all the layers of the DNN together, performance of the DNN significantly

improved. These results suggest that, within broad sections of higher-order visual areas in the ventral and dorsal stream, features at multiple hierarchical stages of complexity are being represented, and that different DNN layers explain different aspects of these representational space.

When comparing performance between DNNs and the more perceptual behavioral spaces (i.e. shape-guided sorting and search-RT spaces), we found that our behavioral models performed similarly to well-performing DNN layers. On the other hand, the model based on a more categorical behavioral measure (i.e., the unguided-sorting space) performed significantly worse than all other models.

These results suggest that, also when considering the neural space at this broader scale, the neural representational spaces in both the ventral and dorsal stream reflect more perceptual, rather than more categorical information.

References

Alexander, R., Nahvi, R., & Zelinsky, G. (2019). Specifying the precision of guiding features for visual search. *Journal of experimental psychology. Human perception and performance*.

Andrews, T. J., Clarke, A., Pell, P., & Hartley, T. (2010). Selectivity for low-level features of objects in the human ventral stream. *Neuroimage*, 49(1), 703–711.

Asada, H. & Brady, M. (1986). The curvature primal sketch. *IEEE transactions on pattern analysis and machine intelligence*, (1), 2–14.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological review*, 61(3), 183.

Auger, S. D. & Maguire, E. A. (2013). Assessing the mechanism of response in the retrosplenial cortex of good and poor navigators. *Cortex*, 49(10), 2904–2913.

Bainbridge, W. A. & Oliva, A. (2015). Interaction envelope: Local spatial representations of objects at all scales in scene-selective regions. *Neuroimage*, (pp. 1–9).

Bates, D. (2005). Fitting linear mixed models in r. *R news*, 5(1), 27–30.

Beauchamp, M. S. (2005). See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Current opinion in neurobiology*, 15(2), 145–153.

Berens, P. et al. (2009). Circstat: a matlab toolbox for circular statistics. *J Stat Softw*, 31(10), 1–21.

Bi, Y., Han, Z., Zhong, S., Ma, Y., Gong, G., Huang, R., Song, L., Fang, Y., He, C., & Caramazza, A. (2015). The White Matter Structural Network Underlying Human Tool Use and Tool Understanding. *Journal of Neuroscience*, 35(17), 6822–6835.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological review*, 94(2), 115.

Biederman, I. & Bar, M. (1999). One-shot viewpoint invariance in matching novel objects. *Vision research*, 39(17), 2885–2899.

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., 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. & Op de Beeck, H. (2016). Dissociations and Associations between Shape and Category Representations in the Two Visual Pathways. *Journal of Neuroscience*, 36(2), 432–444.

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. (1997). The psychophysics toolbox. *Spatial vision*, 10, 433–436.

Caramazza, A. & Mahon, B. Z. (2003). The organization of conceptual knowledge: the evidence from category-specific semantic deficits. *Trends in cognitive sciences*, 7(8), 354–361.

Carlson, T. A., Simmons, R. A., Kriegeskorte, N., & Slevc, L. R. (2014). The emergence of semantic meaning in the ventral temporal pathway. *Journal of cognitive neuroscience*, 26(1), 120–131.

Cate, A. D., Goodale, M. A., & Kohler, S. (2011). The role of apparent size in building- and object-specific regions of ventral visual cortex. *Brain Research*, 1388(C), 109–122.

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.

Cohen, M. A., Alvarez, G. A., Nakayama, K., & Konkle, T. (2017). Visual search for object categories is predicted by the representational architecture of high-level visual cortex. *Journal of Neurophysiology*, 117(1), 388–402.

Connor, C. E., Brincat, S. L., & Pasupathy, A. (2007). Transformation of shape information in the ventral pathway. *Current opinion in neurobiology*, 17(2), 140–147.

de Beeck, H. P. O., Torfs, K., & Wagemans, J. (2008). Perceived shape similarity among unfamiliar objects and the organization of the human object vision pathway. *Journal of Neuroscience*, 28(40), 10111–10123.

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., & Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *science*, 330(6009), 1359–1364.

DiCarlo, J. J. & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in cognitive sciences*, 11(8), 333–341.

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

Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 33(4), 1331–6a.

Drucker, D. M. & Aguirre, G. K. (2009). Different spatial scales of shape similarity representation in lateral and ventral loc. *Cerebral Cortex*, 19(10), 2269–2280.

Duncan, J. & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological review*, 96(3), 433–458.

Edelman, A., Arias, T. A., & Smith, S. T. (1998). The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications*, 20(2), 303–353.

El-Shamayleh, Y. & Pasupathy, A. (2016). Contour Curvature As an Invariant Code for Objects in Visual Area V4. *Journal of Neuroscience*, 36(20), 5532–5543.

Epstein, R. & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392(6676), 598.

Epstein, R. A. (2014). Neural systems for visual scene recognition. *Bar M Kveraga K(Eds) Scene vision*.

Freeman, J. & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature neuroscience*, 14(9), 1195–1201.

Gallant, J. L., Connor, C. E., Rakshit, S., Lewis, J. W., & Van Essen, D. C. (1996). Neural responses to polar, hyperbolic, and cartesian gratings in area v4 of the macaque monkey. *Journal of neurophysiology*, 76(4), 2718–2739.

Garcea, F. E. & Mahon, B. Z. (2014). Parcellation of left parietal tool representations by functional connectivity. *Neuropsychologia*, 60(C), 131–143.

Geirhos, R., Janssen, D. H., Schütt, H. H., Rauber, J., Bethge, M., & Wichmann, F. A. (2017). Comparing deep neural networks against humans: object recognition when the signal gets weaker. *arXiv preprint arXiv:1706.06969*.

Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., & Brendel, W. (2018). Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv preprint arXiv:1811.12231*.

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

Grill-Spector, K. (2003). The neural basis of object perception. *Current opinion in neurobiology*, 13(2), 159–166.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision research*, 41(10-11), 1409–1422.

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.

Groen, I. I., Greene, M. R., Baldassano, C., Fei-Fei, L., Beck, D. M., & Baker, C. I. (2018). Distinct contributions of functional and deep neural network features to representational similarity of scenes in human brain and behavior. *Elife*, 7, e32962.

Güçlü, U. & van Gerven, M. A. J. (2015). Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 35(27), 10005–10014.

Haushofer, J., Livingstone, M. S., & Kanwisher, N. (2008). Multivariate patterns in object-selective cortex dissociate perceptual and physical shape similarity. *PLoS biology*, 6(7), e187.

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.

Innocenti, G. M. & Price, D. J. (2005). Exuberance in the development of cortical networks. *Nature Reviews Neuroscience*, 6(12), 955.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences*, 96(16), 9379–9384.

Janzen, G. & Van Turennout, M. (2004). Selective neural representation of objects relevant for navigation. *Nature neuroscience*, 7(6), 673.

Johnson-Frey, S. H., Newman-Norlund, R., & Grafton, S. T. (2005). A Distributed Left Hemisphere Network Active During Planning of Everyday Tool Use Skills. *Cerebral cortex (New York, N.Y. : 1991)*, 15(6), 681–695.

Jozwik, K. M., Kriegeskorte, N., & Mur, M. (2016). Visual features as stepping stones toward semantics: Explaining object similarity in it and perception with non-negative least squares. *Neuropsychologia*, 83, 201–226.

Julian, J. B., Ryan, J., & Epstein, R. A. (2016). Coding of object size and object category in human visual cortex. *Cerebral Cortex*, 27(6), 3095–3109.

Kaiser, D., Azzalini, D. C., & Peelen, M. V. (2016). Shape-independent object category responses revealed by meg and fmri decoding. *Journal of neurophysiology*, 115(4), 2246–2250.

Kamps, F. S., Julian, J. B., Kubilius, J., Kanwisher, N., & Dilks, D. D. (2016). The occipital place area represents the local elements of scenes. *Neuroimage*, 132, 417–424.

Kayaert, G., Biederman, I., Op de Beeck, H. P., & Vogels, R. (2005). Tuning for shape dimensions in macaque inferior temporal cortex. *European Journal of Neuroscience*, 22(1), 212–224.

Kayaert, G., Biederman, I., & Vogels, R. (2003). Shape tuning in macaque inferior temporal cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 23(7), 3016–3027.

Khaligh-Razavi, S.-M. & Kriegeskorte, N. (2014). Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. *PLoS Computational Biology*, 10(11), e1003915–29.

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of neurophysiology*, 97(6), 4296–4309.

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.

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. (2016). The Large-Scale Organization of Object-Responsive Cortex Is Reflected in Resting-State Network Architecture. *Cerebral Cortex*, (pp. 1–13).

Konkle, T. & Oliva, A. (2011). Canonical visual size for real-world objects. *Journal of Experimental Psychology: human perception and performance*, 37(1), 23.

Konkle, T. & Oliva, A. (2012a). A familiar-size Stroop effect: Real-world size is an automatic property of object representation. *Journal of Experimental Psychology: Human Perception and Performance*, 38(3), 561–569.

Konkle, T. & Oliva, A. (2012b). A Real-World Size Organization of Object Responses in Occipitotemporal Cortex. *Neuron*, 74(6), 1114–1124.

Kriegeskorte, N. & Mur, M. (2012). Inverse MDS: Inferring Dissimilarity Structure from Multiple Item Arrangements. *Frontiers in Psychology*, 3, 245.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 2, 4.

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.

Lewis, J. W. (2006). Cortical Networks Related to Human Use of Tools. *The Neuroscientist*, 12(3), 211–231.

Long, B. & Konkle, T. (2017a). A familiar-size Stroop effect in the absence of basic-level recognition. *Cognition*, 168, 234–242.

Long, B. & Konkle, T. (2017b). A familiar-size stroop effect in the absence of basic-level recognition. *Cognition*, 168, 234–242.

Long, B. & Konkle, T. (2018). The role of textural statistics vs. outer contours in deep cnn and neural responses to objects. In *Conference on Computational Cognitive Neuroscience* (pp.4).

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), 95–109.

Long, B., Moher, M., Carey, S., & Konkle, T. (2019). Real-world size is automatically encoded in preschoolers' object representations. *Journal of Experimental Psychology: Human Perception and Performance*, 45(7), 863.

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 of the United States of America*, 115(38), E9015–E9024.

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(C), 35–41.

Magri, C., Konkle, T., & Caramazza, A. (2016). Visual object responses of the ventral stream reflect both size and motor-relevance. *Journal of Vision*, 16(12), 505–505.

Mahon, B. Z. & Caramazza, A. (2011). What drives the organization of object knowledge in the brain? *Trends in cognitive sciences*, (pp. 1–7).

Mahon, B. Z., Milleville, S. C., Negri, G. A. L., Rumiati, R. I., Caramazza, A., & Martin, A. (2007). Action-related properties shape object representations in the ventral stream. *Neuron*, 55(3), 507–520.

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.

Mullally, S. L. & Maguire, E. A. (2011). A new role for the parahippocampal cortex in representing space. *Journal of Neuroscience*, 31(20), 7441–7449.

Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P. A., & Kriegeskorte, N. (2013). Human Object-Similarity Judgments Reflect and Transcend the Primate-IT Object Representation. *Frontiers in Psychology*, 4, 128.

Nasr, S. & Tootell, R. B. (2012). A cardinal orientation bias in scene-selective visual cortex. *Journal of Neuroscience*, 32(43), 14921–14926.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. *PLoS computational biology*, 10(4), e1003553.

Osaka, N. (1976). Reaction time as a function of peripheral retinal locus around fovea: effect of stimulus size. *Perceptual and motor skills*, 43(2), 603–606.

Paivio, A. (1975). Perceptual comparisons through the mind's eye. *Memory & Cognition*, 3(6), 635–647.

Pasupathy, A. & Connor, C. E. (1999). Responses to contour features in macaque area v4. *Journal of neurophysiology*, 82(5), 2490–2502.

Pasupathy, A. & Connor, C. E. (2002). Population coding of shape in area v4. *Nature neuroscience*, 5(12), 1332.

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.

Peelen, M. V. & Caramazza, A. (2012). Conceptual object representations in human anterior temporal cortex. *Journal of Neuroscience*, 32(45), 15728–15736.

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.

Ponce, C. R., Hartmann, T. S., & Livingstone, M. S. (2017). End-stopping predicts curvature tuning along the ventral stream. *Journal of Neuroscience*, 37(3), 648–659.

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.

Rajimehr, R., Devaney, K. J., Bilenko, N. Y., Young, J. C., & Tootell, R. B. H. (2011). The "Parahippocampal Place Area" Responds Preferentially to High Spatial Frequencies in Humans and Monkeys. *PLoS biology*, 9(4), e1000608–11.

Simmons, W. K. & Martin, A. (2012). Spontaneous resting-state BOLD fluctuations reveal persistent domain-specific neural networks. *Social Cognitive and Affective Neuroscience*, 7(4), 467–475.

Sperandio, I., Savazzi, S., Gregory, R. L., & Marzi, C. A. (2009). Visual reaction time and size constancy. *Perception*, 38(11), 1601–1609.

Srihasam, K., Mandeville, J. B., Morocz, I. A., Sullivan, K. J., & Livingstone, M. S. (2012). Behavioral and Anatomical Consequences of Early versus Late Symbol Training in Macaques. *Neuron*, 73(3), 608–619.

Srihasam, K., Vincent, J. L., & Livingstone, M. S. (2014). Novel domain formation reveals proto-architecture in inferotemporal cortex. *Nature neuroscience*, 17(12), 1776.

Tarhan, L. & Konkle, T. (2019). Reliability-based voxel selection. *bioRxiv*, (pp. 703603).

Troiani, V., Stigliani, A., Smith, M. E., & Epstein, R. A. (2012). Multiple Object Properties Drive Scene-Selective Regions. *Cerebral Cortex*.

Van Essen, D. C., Drury, H. A., Dickson, J., Harwell, J., Hanlon, D., & Anderson, C. H. (2001). An integrated software suite for surface-based analyses of cerebral cortex. *Journal of the American Medical Informatics Association*, 8(5), 443–459.

Vernon, R. J., Gouws, A. D., Lawrence, S. J., Wade, A. R., & Morland, A. B. (2016). Multivariate patterns in the human object-processing pathway reveal a shift from retinotopic to shape curvature representations in lateral occipital areas, lo-1 and lo-2. *Journal of Neuroscience*, 36(21), 5763–5774.

Wang, X., Peelen, M. V., Han, Z., He, C., Caramazza, A., & Bi, Y. (2015). How Visual Is the Visual Cortex? Comparing Connectional and Functional Fingerprints between Congenitally Blind and Sighted Individuals. *Journal of Neuroscience*, 35(36), 12545–12559.

Whatmough, C., Chertkow, H., Murtha, S., & Hanratty, K. (2002). Dissociable brain regions process object meaning and object structure during picture naming. *Neuropsychologia*, 40(2), 174–186.

Wickham, H. & Chang, W. (2009). ggplot2: An implementation of the grammar of graphics. r package version 0.8. 3.

Wolfe, J. M. & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, 1, 1–8.

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

Yue, X., Pourladian, I. S., Tootell, R. B., & Ungerleider, L. G. (2014). Curvature-processing network in macaque visual cortex. *Proceedings of the National Academy of Sciences*, 111(33), E3467–E3475.

Zar, J. H. (1999). *Biostatistical analysis*. Pearson Education India.

This thesis was typeset using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. The body text is set in 11 point Egenolff-Berner Garamond, a revival of Claude Garamont's humanist typeface. The above illustration, "Science Experiment 02", was created by Ben Schlitter and released under CC BY-NC-ND 3.0. A template that can be used to format a PhD thesis with this look and feel has been released under the permissive MIT (XII) license, and can be found online at github.com/suchow/Dissertate or from its author, Jordan Suchow, at suchow@post.harvard.edu.