
UNIVERSITY
OF TRENTO - Italy

# Disentangling the representations of object shape and object category in the brain: the animate-inanimate distinction

Daria Proklova

CIMeC - Center for Mind/Brain Sciences

Doctoral School in Cognitive and Brain Sciences

29th cycle

University of Trento

December 2016

Supervisor: Prof. Marius V. Peelen

# Table of Contents

**Table of contents 3**

**Chapter 1 5**

**General Introduction**

The animate-inanimate distinction in the brain……6

Evidence for shape- and category-driven organization of the ventral temporal cortex……..8

Disentangling object shape and object category……10

**Chapter 2 13**

**Creating a shape-matched stimulus set and quantifying visual dissimilarity**

Abstract……13

Introduction……14

The stimulus set……14

Quantifying visual dissimilarity……15

Computational methods……15

Behavioral methods……17

Experiment 1 (Overall visual dissimilarity)……19

Experiments 2 and 3 (Outline and texture dissimilarity)……20

Discussion……23

**Chapter 3 25**

**Disentangling object shape and object category in human ventral temporal cortex: An fMRI study**

Abstract……25

Introduction……26

Materials and methods……28

Results……34

Discussion……42

**Chapter 4** **47**

**Disentangling object shape and object category with MEG**

Abstract……47

Introduction……48

Materials and methods……49

Results……56

Discussion……68

**Chapter 5** **71**

**General Discussion**

Conclusions……76

**References** **78**

**Supplementary Information** **85**

**Acknowledgements** **88**

**Appendix** **89**

**Disentangling Representations of Object Shape and Object Category in Human Visual Cortex: The Animate–Inanimate Distinction**

# Chapter 1

# General Introduction

The classical view of the visual system distinguishes between two visual pathways (Ungerleider & Mishkin, 1982; Goodale & Milner, 1992). According to this view, after visual information reaches primary visual cortex (V1) it is processed along the dorsal visual stream that stretches into parietal cortex, as well as along the ventral visual stream that proceeds towards the temporal cortex. The two streams differ in the visual properties they process: the dorsal stream, also known as "where" or "how" pathway, processes spatial properties of objects to guide reaching and grasping actions, thus subserving "vision-for-action" (Milner & Goodale, 1995). In contrast, the ventral visual pathway, also known as "what" pathway, is traditionally thought to process visual object properties and participate in object recognition, underlying "vision-for-perception". This pathway culminates in ventral temporal cortex (VTC), which contains visual object representations (Grill-Spector & Weiner, 2014). There is substantial evidence in the literature that these representations are not organized randomly but have an underlying structure. However, the principles and the dimensions that drive VTC organization remain unclear (Kourtzi & Connor, 2011).

In the classical neurophysiology literature, object recognition is thought to be achieved by a hierarchy of visual areas in the ventral stream, with each successive area sensitive to more complex stimulus features: starting with V1 (which processes simple visual features such as orientation) and ending in VTC, which represents complex shape configurations (Hubel & Wiesel, 1968; Tanaka, 1996; Riesenhuber et al., 2000). In the past two decades, accumulating evidence from the functional magnetic resonance imaging (fMRI) literature suggested that object representations in VTC are, at least in some cases, organized based on object category. Several cortical areas were discovered that respond preferentially to stimuli from a particular category. The category-selective areas include the fusiform face area (FFA) that responds more strongly to visually presented faces compared to other object categories (Kanwisher et al., 1997), the parahippocampal place area (PPA) that exhibits selectivity for landmarks and scenes (Epstein & Kanwisher, 1998), the extrastriate body area (EBA) that responds preferentially to images of bodies and body parts (Downing et al., 2001; Peelen & Downing,

2005), and the visual word form area (VWFA) that is selective to written words (Cohen & Dehaene, 2004). However, only a handful of all possible object categories are associated with dedicated functional areas, as demonstrated by a study that tested brain responses to 20 object categories but did not reveal any additional category-selective areas apart from FFA, PPA, and EBA (Downing et al., 2006).

The functional regions described above were discovered using a classical univariate approach by contrasting BOLD responses to different object categories. The use of a new data analysis method called multivariate pattern analysis (MVPA) has provided a breakthrough in research of object representations in the brain (for a review on the topic, see Haxby, 2012). Instead of comparing the magnitude of BOLD response, MVPA allows one to measure and compare spatial activity patterns elicited by different categories. In a landmark study, Haxby et al. (2001) demonstrated that different object categories elicit distinct patterns of response in the VTC. Importantly, these patterns could be used to discriminate between the two categories even after the regions that responded maximally to these categories were excluded from analysis, challenging the idea of focal category-selective areas. In general, MVPA has proved to be more sensitive to the fine-grained representational structure of VTC compared to traditional univariate analyses (Haxby et al., 2001) and was used extensively in the past 15 years to study the widespread patterns of activity associated with different object categories. A version of MVPA called representational similarity analysis (RSA; Kriegeskorte, Mur, & Bandettini, 2008) allows to measure brain response patterns to large number of individual objects (instead of averaging responses for stimuli within a category) in order to explore the large-scale organization of object responses. Kriegeskorte et al. (2008b) used this approach and found that VTC response patterns of a large number of objects clustered according to object category: specifically, responses to animate and inanimate stimuli formed two distinct clusters, with responses to faces and bodies forming sub-clusters within a larger animate cluster. These results suggest that object animacy is an important organizing principle for object representations in the ventral stream. In the next section I will focus more closely on the animate-inanimate distinction in the brain and show that this idea is supported by literature in the fields of neuropsychology, monkey neurophysiology, and human fMRI.

**The animate-inanimate distinction in the brain**

Evidence from neuropsychology suggests that the distinction between animate and inanimate objects is an important conceptual distinction in the brain (Warrington, 1981; Caramazza & Shelton, 1998; Forde & Humphreys, 1999). Specifically, knowledge pertaining to each of these categories can be selectively impaired after brain damage, with some patients showing an impairment in naming animals but not the inanimate objects, and other patients having the opposite deficit (Warrington & Shallice, 1984; Caramazza & Shelton, 1998). However, the interpretation of these deficits has been a source of debate. According to the so-called sensory/functional theory (SFT), conceptual knowledge is organized based on modality (e.g., visual, auditory, etc.), and different modalities are differentially important for recognizing animate and inanimate objects. Assuming that sensory (e.g., visual) properties are more important for identifying animals than inanimate objects, the deficit in processing animals can be caused by the damage to a visual semantic subsystem and not "animate" semantic subsystem (Warrington & McCarthy, Warrington & Shallice, 1984). However, the SFT has been challenged by other studies that argue that animate-inanimate distinction is a truly semantic distinction in the brain, with distinct neural mechanisms underlying the perception of these domains (Caramazza & Shelton, 1998).

Moreover, a large amount of neuroimaging evidence suggests that animate and inanimate objects also have different neural representations in the ventral visual stream. Generally, responses to these categories follow lateral-to-medial organization in the posterior ventral temporal cortex, with animate stimuli (e.g., faces, animals) eliciting higher activation in the lateral fusiform gyrus and inanimate stimuli (e.g., houses, tools) more strongly activating the medial fusiform gyrus (Chao, Haxby, & Martin, 1999; Downing et al., 2006; Bell et al., 2009; Wiggett, Pritchard, & Downing, 2009). The category-selective regions described in the previous section fit in this organization, with FFA and EBA (selective to faces and bodies, respectively) located in the lateral portion of the fusiform gyrus, and PPA (selective to houses and scenes) located in the medial fusiform gyrus.

Evidence from monkey neurophysiology further supports the idea of an animate-inanimate distinction in the ventral stream. Kiani et al. (2007) measured responses to large number of images in the inferior temporal (IT) cortex of the monkey and showed that IT population responses cluster according to object category, with animate and inanimate objects forming

distinct clusters. A human fMRI study using the same stimuli found strikingly similar representational structure in human VTC, showing that large-scale patterns of activity in VTC cluster according to object animacy (Kriegeskorte et al., 2008b).

Recently, it has been suggested that the distinction between VTC representations of animate and inanimate stimuli is not a clear-cut binary distinction, but instead animacy is a continuous dimension, with object's "degree of animacy" defined as the distance from the prototype, i.e., human (Connolly et al., 2012; Sha et al., 2014). In one study, researchers compared responses to different animal species and observed a pattern similar to classical lateral-to-medial organization (e.g., Chao et al., 1999), with "more animate" stimuli (e.g., primates) eliciting greater activation in lateral portions of VTC and "less animate" stimuli (bugs) activating medial portions of VTC mores strongly, similar to inanimate objects (Connolly et al., 2012). However, this study had several limitations, including the absence of inanimate stimuli for control and the small number of animal species sampled. A more recent study addressed these limitations and reported similar results, suggesting that graded animacy is a major organizing principle in VTC (Sha et al., 2014).

To summarize, a large amount of evidence suggests that the animate-inanimate distinction is an important organizing principle for object representations in VTC. However, the factors driving this organization remain yet unknown. Specifically, it is unclear whether the animate-inanimate distinction in VTC is driven by object animacy per se or, alternatively, by other object properties that covary with category: for example, visual features such as object shape. Indeed, apart from belonging to different categories, animate and inanimate objects also differ in how they look: most animals share common features such as head and legs, faces have oval shape, bodies are elongated. On the other hand, inanimate objects such as buildings or tools tend to have more right angles (Nasr, Echavarria, & Tootell, 2014). In the next sections, I will review the evidence supporting shape- or category-based accounts of VTC organization and propose a way to disentangle these two factors in the representation of animate and inanimate objects.

**Evidence for shape- and category-driven organization of the ventral temporal cortex**

There is an ongoing debate about the source of the apparent category selectivity in VTC. Objects belonging to the same category often share visual features (e.g., oval shape for faces, angles for landmarks, four limbs for some animals). According to the object form topology account, information about object shape is represented topologically in VTC, with different object categories evoking distributed and overlapping responses in VTC defined by their shape (Haxby et al., 2000; Ishai et al., 1999). Recent studies have suggested that category-selective regions in VTC are responding to the characteristic shape features of their preferred categories, rather than category information per se (Watson et al., 2014; Rice et al., 2014; Andrews et al., 2015). For example, one study has shown that FFA responds preferentially to oval shapes with more dots in the upper part than in the lower part (Caldara et al., 2006). Another series of studies demonstrated that PPA preferentially responds to cardinal orientations (Nasr et al., 2012) and rectilinear angles (Nasr et al., 2014), visual features that are characteristic for buildings. Rice et al. (2014) used MVPA to measure VTC response patterns to several object categories and found that within- and between-category correlations of category-evoked neural patterns can be explained by the corresponding correlations between the visual properties of objects from different categories. Finally, a study using neuronal recordings in monkeys failed to find animate-inanimate distinction in large-scale VTC response patterns and concluded that the observed patterns were better explained by object shape (Baldassi et al., 2013).

Despite the emerging evidence in favor of shape-based organization of VTC, a recent review has pointed out that shape cannot fully account for the observed category-selective patterns of activity showing that visual features are neither sufficient nor necessary to explain category selectivity (Peelen & Downing, submitted). The evidence that visual features are not sufficient to evoke category-selective responses comes from studies using ambiguous stimuli such as Mooney faces (Mooney, 1957). The response in face-selective areas increases only after the subject recognizes the stimulus as a face (Andrews et al., 2004), despite the visual input being the same. In another study, tool-selective activity increased after the subjects learned to use novel objects as tools (Weisberg et al., 2007).

Visual features are also not necessary to evoke category-specific responses. In congenitally blind subjects, the same regions preferentially respond to living and non-living stimuli as in

sighted subjects, despite these subjects not having any visual experience (Mahon et al., 2009). However, this does not necessarily mean that object representations in congenitally blind individuals lack shape information. The shape conveyed by vision-to-sound substitution is known to activate object-selective areas in blind participants (Amedi et al., 2007). A recent study demonstrated that even in congenitally blind participants, patterns of activity in occipito-temporal cortex reflect shape similarity of objects presented with the auditory cue (Peelen et al., 2014). This shows that shape information seems to be present in VTC of both sighted and blind individuals.

**Disentangling object shape and object category**

Until recently, the two accounts of VTC organization - based on shape or category - have never been directly tested against each other. In many studies, the visual similarity between stimuli from different categories was measured post-hoc (e.g., Kriegeskorte et al., 2008b), but there was no attempt to pre-select the stimuli in a way to control for shape. Category and shape information can be disentangled by using a factorial design, in which objects from different categories are controlled for shape (e.g., Bracci & Op de Beeck, 2016). Several recent studies used this approach to study responses to tools and visually matched non-tools with fMRI (Macdonald & Culham, 2015), and body parts and matching clothes with MEG (Kaiser, Azzalini, & Peelen, 2016).

Here we introduce a stimulus set designed to disentangle information about object shape and object animacy. This stimulus set consists of images of animate and inanimate objects that are carefully matched for shape (see Chapter 2). We also describe a behavioral paradigm that allows for measuring visual dissimilarity of the stimuli based on participants' performance in a simple visual search task. Three behavioral visual search experiments (described in Chapter 2) were conducted to measure different aspects of visual dissimilarity of the stimuli. In Chapter 3, we tested the hypothesis that visual dissimilarity can fully explain the category-related patterns of activity in human ventral temporal cortex. Using fMRI, we demonstrated that this was not the case: category information was still present in VTC activity patterns after controlling for visual dissimilarity of the stimuli. Finally, in Chapter 4 we describe two MEG experiments in which we tested whether it was possible to disentangle information about object category and object shape reflected in MEG sensor patterns. This allowed us to look at

the time courses of shape and category information separately. We found that information about object shape, outline, and texture was reflected robustly in the MEG signal. Surprisingly, after regressing out visual dissimilarity, we found no residual information about object category in MEG signals, suggesting that MEG is less sensitive to this information compared to fMRI.

# Chapter 2

# Creating a shape-matched stimulus set and quantifying visual dissimilarity

**Abstract**

In order to disentangle object shape and object category and examine their separate contributions to brain representations, it is important to select an appropriate stimulus set as well as to properly quantify visual dissimilarity of the stimuli. Here, we introduce a stimulus set that includes images of objects from two categories: animate and inanimate objects. Importantly, for each animate stimulus there is an inanimate stimulus carefully matched for overall shape features (i.e., snake and rope). This stimulus set allows one to study category-selective brain responses while controlling for object shape. However, to ensure that our stimuli were indeed closely matched for shape and to obtain the concrete measure of visual dissimilarity for further analyses, it is important to select an appropriate method to quantify visual dissimilarity of the stimuli. Here, we provide an overview of existing methods for quantifying visual dissimilarity, from computational models of vision to behavioral methods. Finally, we describe a behavioral visual search task that is well suited to quantify perceptual similarity of the stimuli. Using this task, we conducted three behavioral visual search experiments to measure three different aspects of visual dissimilarity of our stimulus set: overall visual dissimilarity of the intact stimuli, outline dissimilarity, and texture dissimilarity.

## Introduction

Multiple neuroimaging studies that investigated category-selective brain responses have also controlled for object shape in some manner. Often this is done post hoc by using a computational model of visual similarity to show that the category-related responses cannot be explained by visual features (e.g., Kriegeskorte et al., 2008). However, these models do not capture all of the visual features that could be influencing the category-related responses. Very few studies have tried to explicitly match stimuli from different categories for shape (Bracci & Op de Beeck, 2016; Kaiser, Azzalini, & Peelen, 2016).

In the present study, our goal was to create a stimulus set that would allow us to study independent contributions of shape and category (specifically, animacy) to the brain responses measured with fMRI and MEG. To this end, we created a stimulus set that included images of animate and inanimate objects. Critically, for each animate object there was an inanimate object that was similar to it in shape (e.g., a bird and an airplane).

## The stimulus set

The stimulus set is shown in Figure 2.1. It consists of 16 individual stimuli which are organized into four sets of four objects, with four objects within each set having a roughly similar shape. Two objects of each set were animate, and two objects were inanimate. In addition, there were four exemplars of each object (e.g., four images of a snake), resulting in 16 stimuli per set and a total of 64 stimuli. The complete stimulus set is shown in Supplementary Figure 2.1. All images were gray scaled and matched for luminance and contrast using the SHINE toolbox (Willenbockel et al., 2010).

Since stimulus selection and matching was based on subjective and possibly superficial shape similarity judgements, we needed a more objective measure to quantify the visual dissimilarity of the stimuli and confirm our initial selection. In the next section, we review existing methods of measuring visual dissimilarity and introduce a behavioral visual search task that we used for this purpose.

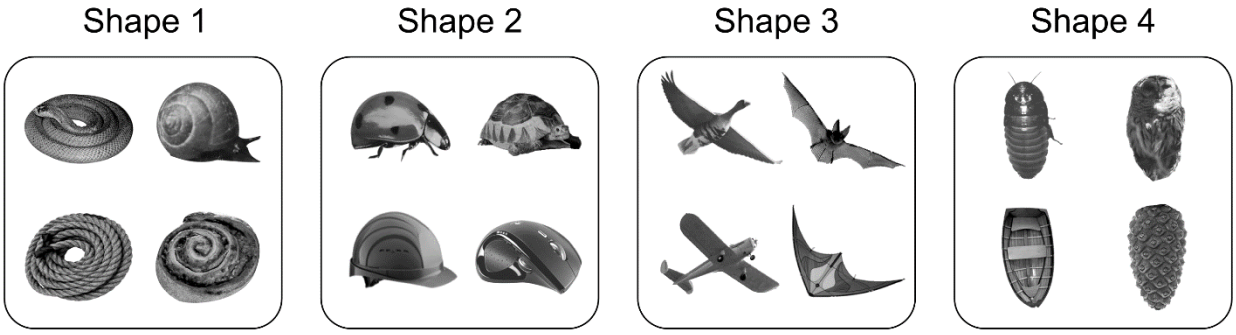



**Fig. 2.1. Stimulus set.** The stimulus set consists of sixteen unique stimuli, divided into four shape-defined subsets. All objects within each set have similar shape. Two objects of each set are animate (upper row), and two objects are inanimate (bottom row). In addition, we used four versions of each stimulus (see Supplementary Figure 1 for the complete stimulus set).

## Quantifying visual dissimilarity

Various methods of quantifying perceptual similarity can be roughly divided into two main classes: computational and behavioral. In computational methods, the visual similarity is measured using some external model, such as pixel-wise similarity or biologically inspired models of visual processing. In behavioral methods, the visual dissimilarity is inferred from various behavioral measures, such as dissimilarity ratings, or reaction times (RTs) in behavioral tasks such as categorization or visual search. In the following sections, we will briefly describe these two classes of methods and provide some examples.

## Computational methods

One approach to quantify visual dissimilarity of the stimuli is to use computational models. Different algorithms can be used to measure physical characteristics of the stimuli. Moreover, such models are often inspired by different stages of processing in the visual system, modeling object representations at different levels with the ultimate goal of relating them to behavior.

The most straightforward way to measure physical dissimilarity between two images is by measuring their pixel-wise dissimilarity (e.g., Cutzu & Edelman, 1998, Grill-Spector et al., 1999). However, sometimes two images can differ a lot in terms of pixels, but appear similar

to a human observer. One way to obtain a more sensitive dissimilarity measure is to estimate the transformation relating one stimulus to another and define dissimilarity as the number of matching errors in the process of transformation (e.g., Belongie, Malik, & Puzicha, 2002). Still, these models are limited when it comes to complex, naturalistic stimuli. Moreover, physical features of the stimuli could be very different from the perceptual features used by humans to judge visual similarity, as shown by studies in which physical and perceptual similarity are orthogonalized (Op de Beeck, Torfs, & Wagemans, 2008).

One way to model perceptual similarity of the stimuli is to create computational models that are based on our knowledge of the visual system, and then compare the performance of the model with that of humans. The main example is HMAX, a biologically-inspired model of object recognition, which is organized by a hierarchy of layers modelling different areas of the visual system, starting with V1 and culminating in inferior temporal cortex, IT (Reisenhuber & Poggio, 1999). This model has been shown to accurately model human behavioral performance in visual categorization task (Serre, Oliva, & Poggio, 2007). However, HMAX failed to fully explain the representational geometry in IT, in particular, the clustering of object representations based on category (Kriegeskorte et al., 2008).

Finally, one of the most recent advancements in the field of modelling object shape perception came with the use of deep neural networks (DNNs). The most common examples of DNNs are convolutional neural networks (CNNs), which are composed of several convolutional layers as well as fully connected layers, with the image label as the final output. CNNs were shown to classify images with surprisingly high accuracy (Krizhevsky, Sutskever, & Hinton, 2012), even reaching performance level close to that of humans (Szegedy et al., 2014). They can also be used to quantify shape dissimilarity. In a recent study, the outputs of deep neural networks were shown to correlate with perceived shape as measured by human ratings (Kubilius, Bracci, & Op de Beeck, 2016), suggesting that DNNs might be using similar object shape features to judge visual similarity as those used by a human observer. This might be the most accurate computational model of human visual processing so far.

**Behavioral methods**

Computational models are powerful tools, but they are still only the approximations of the human ability to judge perceptual dissimilarity of stimuli. Even if model performance matches that of human, the features used by the two could still be different. It is therefore important to have a more direct measure of perceived visual similarity that is derived from actual human behavioral data.

One of the most common ways to measure perceived dissimilarity is to present participants with pairs of stimuli and ask them to rate how visually dissimilar the two stimuli are on a numerical scale (e.g., Cortese & Dyre, 1998). However, with large stimulus sets this will require a large number of trials, which might be difficult for participants and affect their ratings. Moreover, this method only allows for comparison of two stimuli at a time, potentially missing how all of the stimuli within a stimulus set relate to each other.

An interesting alternative is a method known as inverse multidimensional scaling (inverse MDS; De Leeuw & Groenen, 1997). Multidimensional scaling (MDS) refers to the process of visualizing the distances between n-dimensional vectors, for example, on a 2D plane (Torgerson, 1958; Borg & Groenen, 2005). Inverse MDS is the opposite process: by taking the arrangement of points on a 2D plane, one can infer the distances, and, therefore, the dissimilarity between them. This can be done by displaying all stimuli on a screen and asking participants to arrange the stimuli according to their visual similarity, moving similar images more closely together (Kriegeskorte & Mur, 2012). The advantage of this method as compared to pairwise ratings is that it allows participants to arrange all of the stimuli in one setting, capturing the dissimilarity not just of two stimuli at a time, but of all of the stimuli in relation to each other, providing a continuous measure of dissimilarity. This method is also well suited for large stimulus sets.

Another way to infer visual dissimilarity from behavioral measures is to engage participants in a task that requires processing of stimuli and then measure how behavior changes as a function of visual similarity between the stimuli. One such measurement is reaction time (RT) in behavioral tasks such as categorization or visual search.

Mohan & Arun (2012) introduced a visual search task in which participants were asked to search for a target surrounded by an array of identical distractors. When the target is very

visually different from the distractors, it "pops out" and is very easy to spot, which is reflected in short RTs. However, this task gets more difficult as the target and distractor become more visually similar, which leads to longer RTs. By computing pairwise RTs for each target-distractor pair, we can obtain distance matrix that contains pairwise dissimilarity values for all of the stimuli.

This method is similar to pairwise ratings in a way that it requires the number of trials that is equal to the number of all possible pairwise comparisons within a stimulus set. However, the trials are shorter because the participants are not required to actively think about the rating but instead are performing a largely automatic bottom-up task. This way, participants are likely to experience less fatigue, making this task appropriate for our stimulus set size.

Another advantage of this method is that the category of the stimuli is irrelevant to the task. Participants are not given any cue about the identity of the target before the search array appears, and the visual features of the stimuli are enough to accurately detect the oddball without necessarily processing the identity of all the stimuli. Finally, participants have to respond quickly, so there is less time to allow category similarity to affect their judgement of visual similarity, as could be the case with behavioral ratings. Since we are interested in disentangling visual and category dissimilarity, this task is ideal to obtain a visual dissimilarity measure that is not influenced by category.

**Measuring overall visual dissimilarity with the visual search task**

We set out to measure the pairwise visual dissimilarity for our stimuli using a visual search task similar to the one described by Mohan and Arun (2012). A series of three behavioral visual search experiments was conducted, each focused on measuring a separate aspect of visual dissimilarity. In all three experiments, participants searched for an oddball target surrounded by identical distractor objects (Figure 2.2A). The response time (RT) in this task is a measure of visual similarity: the longer the response time for locating the oddball stimulus, the more visually similar are the target and the distractor object. Experiment 1 measured overall visual similarity, Experiment 2 measured outline visual similarity, and Experiment 3 measured texture visual similarity.

*Experiment 1 (Overall visual dissimilarity)*

To quantify overall pairwise visual similarity of the stimulus set, 18 participants were tested in a behavioral experiment (two men; mean age = 22.5 years, *SD* = 2.97 years). Stimuli were presented on a 17-in. CRT monitor, and presentation was controlled using Psychtoolbox (Brainard, 1997). Each search display contained 16 objects placed in a 4 × 4 grid, with one oddball target and 15 identical images of the distractor object. The location of the 16 objects in the grid was randomized. The size of the target and seven of the distractors was 100 × 100 pixels, which corresponded to 2.9° visual angle. The remaining distractors differed in size, with four being 120% of the target size (3.4° visual angle) and four being 80% of the target size (2.3° visual angle). This was done to minimize the effects of stimulus alignment that could facilitate the oddball detection. Participants had to indicate whether the oddball target appeared on the left side or on the right side of the screen. Importantly, no information about the category of the oddball target was provided. The search display remained on the screen until the response, followed by 500-msec fixation, after which the next trial started. The experiment consisted of four blocks. In each block, only one of the four exemplars of each object was used (e.g., always the same snake within one block), resulting in 16 unique objects and 240 trials per block (for all possible target-distractor pairings of the 16 stimuli). Accuracy was high (97.2%) and was not further analyzed.

RTs were averaged across corresponding target-distractor object pairs and across blocks. The data from the visual search experiment served to create a matrix of overall pairwise visual dissimilarity. For this purpose, we took the inverse of these RTs (1/RT) for each stimulus pair as a measure of dissimilarity. The resulting visual dissimilarity matrix consisted of one visual dissimilarity value for every pairwise combination of objects (Figure 2.2B). Multidimensional scaling analysis (using *cmdscale* function in MATLAB, The MathWorks, Natick, MA) revealed that stimuli from the same shape sets clustered together (Fig. 2.2D), whereas there was no apparent categorical organization. Furthermore, within each shape set, we did not find any evidence for categoricality, with the average visual dissimilarity within categories (e.g., snake–snail) being equal to the average visual dissimilarity across categories (e.g., rope–snail), $t(17) = 1.22$, $p = .239$. These results confirm our intuitive shape sets and show that there were no obvious visual properties that covaried with category membership.

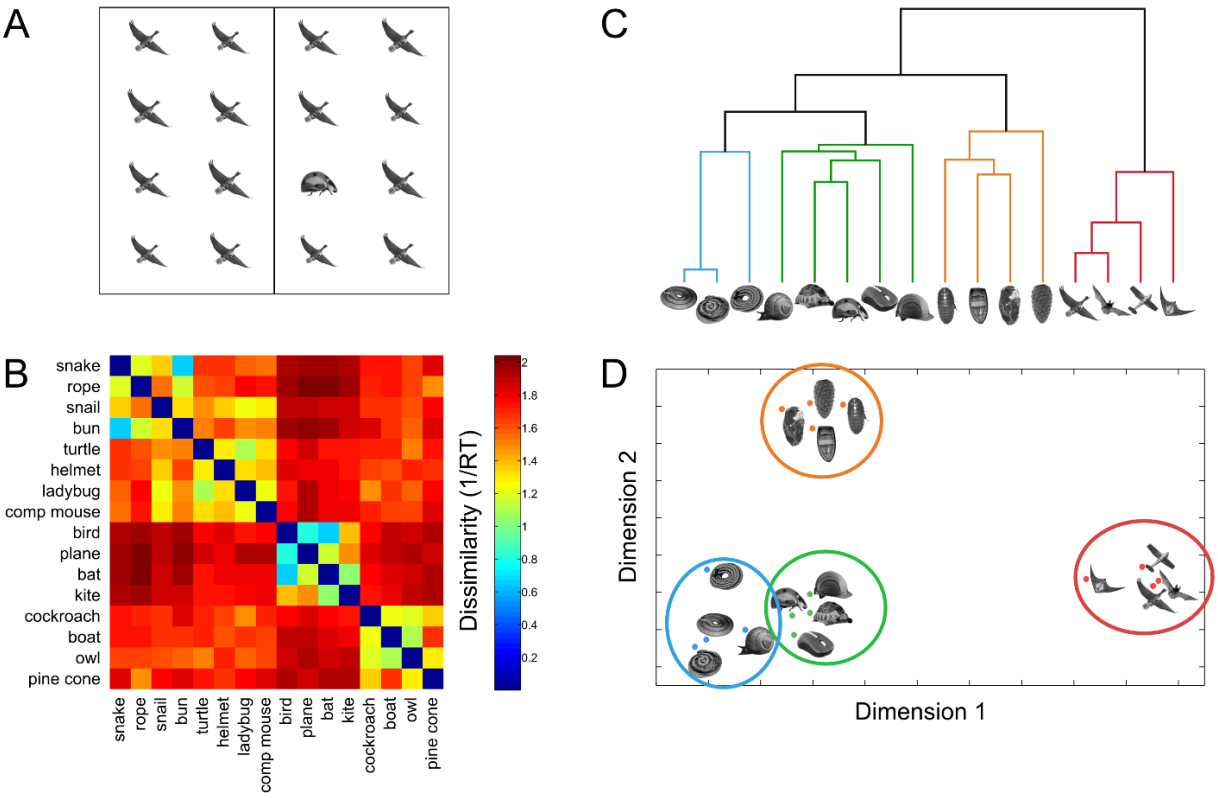



**Fig. 2.2. Measuring overall visual dissimilarity.** (A) In the visual search experiment, participants indicated whether an oddball target in a 4 × 4 search array was located to the right or to the left of the vertical display midline. No prior information about the target was given, so that participants had to rely on bottom-up visual differences to perform the task. (B) From the RTs in this experiment, we created a visual dissimilarity matrix, where each element represents the inverse RT for a pair of stimuli, averaged across the two respective target-distractor pairings. High dissimilarity values correspond to short RTs, reflecting that target and distractor were visually dissimilar. (C) Hierarchical clustering plot based on the visual dissimilarity matrix, showing that with the exception of one stimulus (snail) the stimuli from four original shape sets cluster together. (D) Multidimensional scaling representations of the visual dissimilarity matrix confirmed that stimuli from the same shape sets clustered together.

*Experiments 2 and 3 (Outline and texture dissimilarity)*

To measure pairwise similarity of the outline shape of the stimuli, 18 new participants (three men; mean age = 23.3 years, *SD* = 3.4 years) were tested in Experiment 2. The experiment was identical to Experiment 1 except that the stimulus set consisted of outline drawings of the

stimuli, created by automatically tracing the outline contours of binarized silhouette versions of the original stimuli (Figure 2.3A). Accuracy was high (98.1%) and was not further analyzed.

To measure pairwise texture similarity of the stimulus set, 18 participants (eight men; mean age = 25.7 years, *SD* = 5.5 years) were tested in Experiment 3. Two of the participants had also participated in Experiment 2. The experiment was identical to Experiments 1 and 2 except that the stimulus set consisted of circular texture patches, created by masking the original images with a circular aperture that covered about 20% of the image (Figure 2.3B). The aperture was centered on the mean pixel coordinate of the image. Any blank spaces were filled in using the clone stamp tool in Photoshop. This circular masking abolishes outline shape information, while leaving inner features and texture properties of the stimuli largely intact. It should be noted that these patches still contained some local structure (e.g., the inner contour of the snake) and may thus capture some internal shape features in addition to texture properties. Accuracy was high (98.1%) and was not further analyzed.

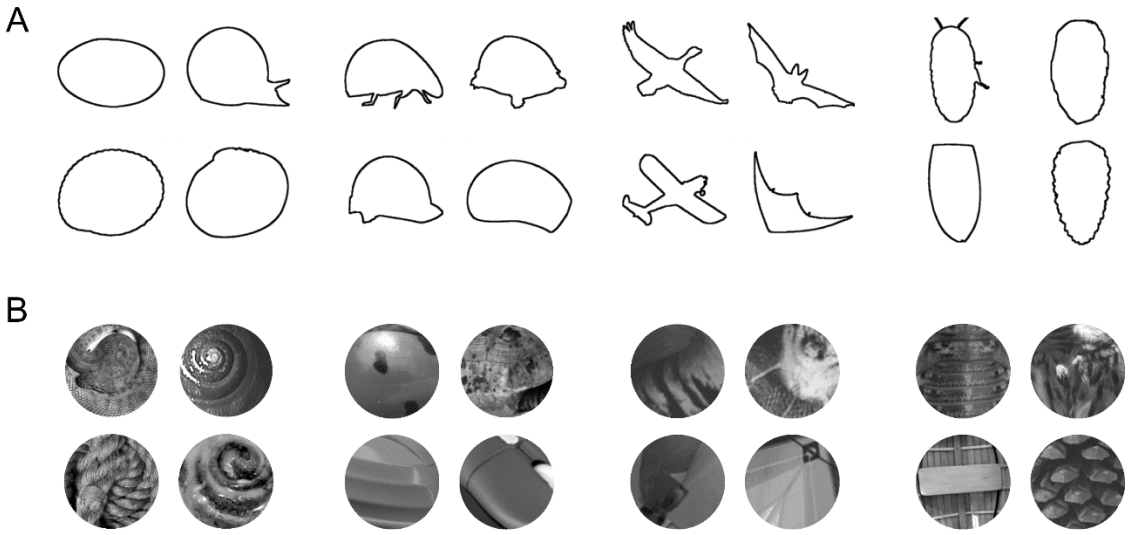



**Figure 2.3. Outline and texture of the original stimuli.** (A) To measure outline dissimilarity of our stimuli, we used the outline drawings created by automatically tracing the outline contours of binarized silhouette versions of the original stimuli. (B) We also created a stimulus set consisting of texture patches taken from the center of each of the original stimuli to quantify texture dissimilarity. In this stimulus set, the information about object shape is lost, while the inner stimulus features are largely preserved.

Data for Experiments 2 and 3 were analyzed as in Experiment 1, resulting in dissimilarity matrices representing outline and texture dissimilarity of the stimulus set (Figure 2.4A and B, center). Further analyses showed that overall visual dissimilarity could be nearly perfectly predicted by a linear combination of outline and texture dissimilarity, with the optimal weights being 0.75 and 0.25, respectively. Linear combinations of dissimilarity matrices were computed by using data from different single participants (e.g., 0.75 × outline of one participant + 0.25 × texture of another participant), each of which was then correlated with the average visual dissimilarity matrix (with one participant left out). The resulting average correlation ($r$ = .77) approached the noise ceiling of visual dissimilarity ($r$ = .82, computed by correlating each participant's visual dissimilarity matrix with the group-averaged visual dissimilarity matrix, leaving out this participant). Finally, the combined model was significantly more strongly correlated with visual dissimilarity than was either outline or texture alone ($p$ < .001, for both comparisons). These analyses show that overall visual dissimilarity is influenced both by outline and texture properties, which together almost fully explain overall visual dissimilarity.

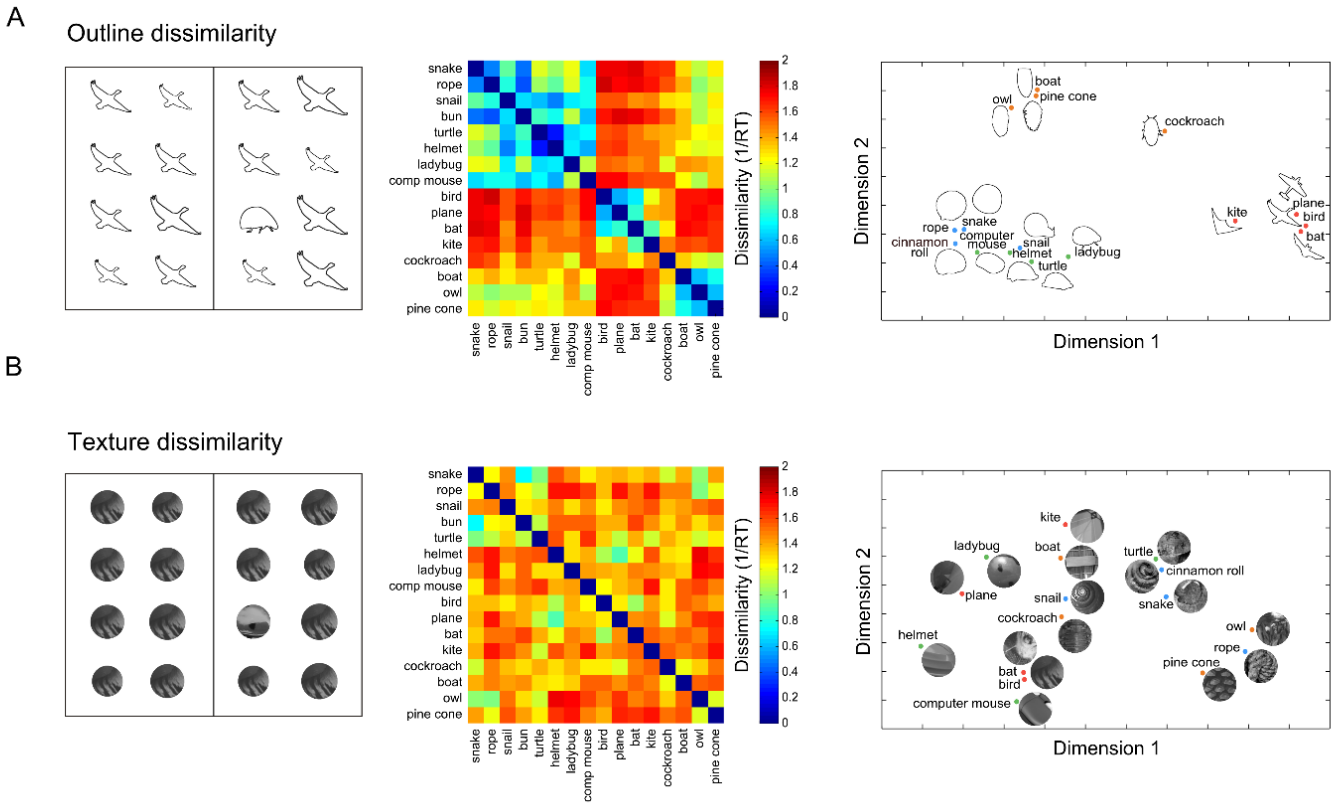



**Figure 2.4. Measuring outline and texture dissimilarity.** Two additional visual search experiments were conducted to measure (A) outline dissimilarity, and (B) texture dissimilarity of the stimuli. As

in Experiment 1, participants searched for an oddball target in a 4 × 4 search array (left) and responded whether it was located to the right or to the left of the vertical display midline. From the RTs in these experiments, we created visual dissimilarity matrices (center), where each element represents the inverse RT for a pair of stimuli, averaged across the two respective target–distractor pairings. Multidimensional scaling representations of the two visual dissimilarity matrices are shown in the right panels. The two experiments differed only in the stimuli used: (A) outline drawings of the original stimuli and (B) internal textures of these images.

## Discussion

In this chapter, we introduced a stimulus set designed to disentangle information about object category (specifically, animacy) and object shape. This stimulus set consisted of images of animate and inanimate objects, with each animate object being closely matched for shape with an inanimate object.

Moreover, we described a behavioral visual search task that can be used to quantify different aspects of visual dissimilarity of the stimuli. We conducted three visual search experiments to measure overall dissimilarity, outline dissimilarity, and texture dissimilarity. In the first experiment using the original stimuli we confirmed that the stimuli within each of the four shapes are indeed perceptually similar to each other (Figure 2.2). Interestingly, we showed that both texture and outline contribute to overall visual dissimilarity of the stimuli, and that the overall visual dissimilarity can be almost perfectly described as a linear combination of outline and texture dissimilarity.

Apart from quantifying the visual properties of our stimulus set and confirming the intuitive stimulus selection, these experiments provided us with visual dissimilarity matrices that could be used to relate neuroimaging data with the perceptual similarity of the stimuli. In particular, we aimed to use these matrices as predictors in representational similarity analysis (RSA) to disentangle the independent contributions of visual and category dissimilarity to neural dissimilarity (fMRI Experiment, Chapter 3). We then used them to investigate the time course of visual and category information reflected in MEG sensor patterns (MEG Experiment, Chapter 4).

# Chapter 3

## Disentangling object shape and object category in human ventral temporal cortex: An fMRI study.

**Abstract**

Objects belonging to different categories evoke reliably different fMRI activity patterns in human occipitotemporal cortex, with the most prominent distinction being that between animate and inanimate objects. An unresolved question is whether these categorical distinctions reflect category-associated visual properties of objects or whether they genuinely reflect object category. Here, we addressed this question by measuring fMRI responses to animate and inanimate objects that were closely matched for shape and low-level visual features. Univariate contrasts revealed animate- and inanimate-preferring regions in ventral and lateral temporal cortex even for individually matched object pairs (e.g., snake–rope). Using representational similarity analysis, we mapped out brain regions in which the pairwise dissimilarity of multivoxel activity patterns (neural dissimilarity) was predicted by the objects' pairwise visual dissimilarity and/or their categorical dissimilarity. Visual dissimilarity was measured as the time it took participants to find a unique target among identical distractors in three visual search experiments, where we separately quantified overall dissimilarity, outline dissimilarity, and texture dissimilarity. All three visual dissimilarity structures predicted neural dissimilarity in regions of visual cortex. Interestingly, these analyses revealed several clusters in which categorical dissimilarity predicted neural dissimilarity after regressing out visual dissimilarity. Together, these results suggest that the animate–inanimate organization of human visual cortex is not fully explained by differences in the characteristic shape or texture properties of animals and inanimate objects. Instead, representations of visual object properties and object category may coexist in more anterior parts of the visual system.

## Introduction

Large-scale patterns of fMRI activity spanning the ventral temporal cortex (VTC) distinguish animate from inanimate object categories (e.g., Kriegeskorte, Mur, Ruff, et al., 2008), with animate objects evoking higher BOLD responses in lateral VTC and inanimate objects evoking higher BOLD responses in medial VTC (e.g., Mahon et al., 2007; Downing, Chan, Peelen, Dodds, & Kanwisher, 2006; Chao, Haxby, & Martin, 1999). Within these broader regions, focal regions exhibit selective responses to more specific categories, including regions selective for buildings and scenes, faces, tools, body parts, and words (Peelen & Downing, 2005; Cohen & Dehaene, 2004; Downing, Jiang, Shuman, & Kanwisher, 2001; Chao et al., 1999; Epstein & Kanwisher, 1998; Kanwisher, McDermott, & Chun, 1997). Although the selectivity for object categories in VTC has been widely replicated, particularly the animate–inanimate distinction, the factors driving this selectivity are still under debate (Andrews, Watson, Rice, & Hartley, 2015; Grill-Spector & Weiner, 2014; Mahon & Caramazza, 2011; Op de Beeck, Haushofer, & Kanwisher, 2008; Martin, 2007).

One of the key questions is whether category-specific patterns of brain activity reflect genuine categorical distinctions (Caramazza & Shelton, 1998) or whether these can be alternatively explained by factors that covary with category membership, such as shape properties. Because of the close association between certain visual properties and category membership, it is to be expected that category-selective regions are optimized for processing these visual properties and/or that these regions are located in parts of the visual system that have visual and retinotopic biases that are optimal for processing the visual features that are characteristic of the category. However, although specific visual properties often characterize object categories, these two dimensions (visual, categorical) are not identical and can indeed be experimentally dissociated. For example, although most tools are elongated, this shape property can be dissociated from the conceptual properties associated with tools (e.g., that tools are manipulable and used as effectors; Bracci & Peelen, 2013). For a visually more homogenous category such as animals, this distinction is more challenging but may still be addressed by testing responses to visually less typical examples (e.g., snakes) and, conversely, testing responses to inanimate objects that share visual features with animals (e.g., mannequins, dolls, statues). These considerations raise the intriguing question of whether

category selectivity in VTC reflects selectivity for conceptual category or selectivity for visual properties that characterize a category.

According to the object form topology account, category-selective fMRI responses in VTC reflect the activation of object form representations that are mapped onto VTC in a continuous fashion (Haxby, Ishai, Chao, Ungerleider, & Martin, 2000; Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999). The selective response to animals in VTC may thus arise from selectivity for characteristic animal shape(s) rather than selectivity for animacy per se. A recent monkey study provided support for this hypothesis, showing that the organization of animate and inanimate object representations in monkey inferotemporal cortex primarily reflects visual similarity rather than semantic similarity (Baldassi et al., 2013; but see Kiani, Esteky, Mirpour, & Tanaka, 2007). Further support for the visual similarity account comes from fMRI studies showing that category-selective regions in VTC respond selectively to visual properties that are characteristic of the regions' preferred categories, even for otherwise meaningless stimuli (i.e., in the absence of category recognition). For example, the fusiform face area was shown to respond more strongly to oval shapes with a greater number of black elements in the top half than to oval shapes with a greater number of elements in the bottom half, although none of these stimuli were recognized as faces (Caldara et al., 2006). Similarly, the parahippocampal place area, located within the medial inanimate-preferring VTC, was shown to respond preferentially to objects made up of cardinal orientations and right angles, features typical of manmade objects, buildings, and scenes (Nasr, Echavarria, & Tootell, 2014).

Recent evidence against a “visual properties” account of category selectivity in VTC comes from studies in congenitally blind individuals. These individuals, with no visual experience, show a categorical organization of VTC that is remarkably similar to that observed in sighted individuals (Ricciardi, Bonino, Pellegrini, & Pietrini, 2014). For example, aurally presented words describing large inanimate objects, versus animals, activate medial VTC in both blind and sighted groups (He et al., 2013; Mahon, Anzellotti, Schwarzbach, Zampini, & Caramazza, 2009). Using a variety of presentation methods, most of the category-selective VTC regions found in sighted individuals have now also been reported in blind individuals, often at nearly identical anatomical locations in the two groups (Striem-Amit & Amedi, 2014; Peelen et al., 2013; Reich, Szwed, Cohen, & Amedi, 2011; Wolbers, Klatzky, Loomis, Wutte, & Giudice, 2011; Buchel, Price, & Friston, 1998). These studies show that the processing of visual features is

not necessary for some category-selective responses to develop. However, they do not exclude the possibility that category selectivity in VTC nevertheless reflects shape properties of objects. This is because VTC has been shown to extract object shape from nonvisual input modalities (Amedi et al., 2007; Amedi, von Kriegstein, van Atteveldt, Beauchamp, & Naumer, 2005), with VTC activity patterns reflecting the shape similarity of objects in both blind and sighted groups (Peelen, He, Han, Caramazza, & Bi, 2014).

This study was designed to investigate the contribution of shape similarity in the representation of animate and inanimate object categories in VTC. Participants viewed pictures of a variety of animals that systematically differed in their shape, grouping into four shape clusters (Chapter 2, Figure 2.1). Importantly, inanimate control objects were selected to closely match the animals in terms of their shape, following the same four shape clusters. This design allowed us to test whether animate- and inanimate-preferring regions (localized with a standard functional localizer) maintain their selectivity for carefully matched animate–inanimate pairs (e.g., snake vs. rope) and whether this is true for a variety of animals (e.g., birds, insects, reptiles) and inanimate objects (e.g., plane, rope, pine cone). In addition to analyses measuring activation differences, we used representational similarity analysis (RSA) to map out regions in which neural similarity reflected the objects' visual and/or categorical similarity (animate/inanimate). For this purpose, we quantified pairwise visual similarity using visual search tasks designed to measure different aspects of visual similarity (overall visual similarity, outline similarity, and texture similarity; see Chapter 2 and Figure 3.1).

## Materials and methods

### Participants

Eighteen participants (seven men; mean age = 25 years, *SD* = 2.4 years) were scanned at the Center for Mind/Brain Sciences of the University of Trento. All participants gave informed consent. All procedures were carried out in accordance with the Declaration of Helsinki and were approved by the ethics committee of the University of Trento. One participant was excluded from all analyses because of excessive head movement.

## Stimuli

The stimuli of the main experiment were organized into four sets of four objects. The four objects within each set all had a roughly similar shape. Two objects of each set were animate, and two objects were inanimate (see Chapter 2, Fig 2.1). In addition, there were four exemplars of each object (e.g., four images of a snake), resulting in 16 stimuli per set and a total of 64 stimuli. All images were gray scaled, placed on gray background and matched for luminance and contrast using the SHINE toolbox (Willenbockel et al., 2010). Stimulus presentation was controlled using the Psychtoolbox (Brainard, 1997). Images were back-projected on a translucent screen placed at the end of the scanner bore. Participants viewed the screen through a tilted mirror mounted on the head coil. Stimuli were presented foveally and subtended a visual angle of approximately 4.5°.

## Visual Search Experiments

To provide a measure of pairwise visual similarity of the stimulus set, a series of three behavioral visual search experiments was conducted. The experiments are described in detail in Chapter 2. In all three experiments, participants searched for an oddball target surrounded by identical distractor objects (Figure 3.1A). The response time in this task is a measure of visual similarity (Mohan & Arun, 2012): The longer the response time for locating the oddball stimulus, the more visually similar are the target and the distractor object. Experiment 1 measured overall visual similarity, Experiment 2 measured outline visual similarity, and Experiment 3 measured texture visual similarity.

## Main fMRI Experiment Procedure

The main fMRI experiment consisted of eight runs. Each run consisted of 80 trials that were composed of 64 object trials and 16 fixation-only trials. In object trials, a single stimulus was presented for 300 ms, followed by a 3700-ms fixation period (Figure 3.2A). In each run, each of the 64 images appeared exactly once. In fixation-only trials, the fixation cross was shown for 4000 ms. Trial order was randomized, with the constraints that there were exactly eight 1-back repetitions of the same category (e.g., two snakes in direct succession) within the object trials and that there were no two fixation trials appearing in direct succession. Each run

started and ended with a 16-s fixation period, leading to a total run duration of 5.9 min. Participants were instructed to press a button whenever they detected a 1-back repetition.

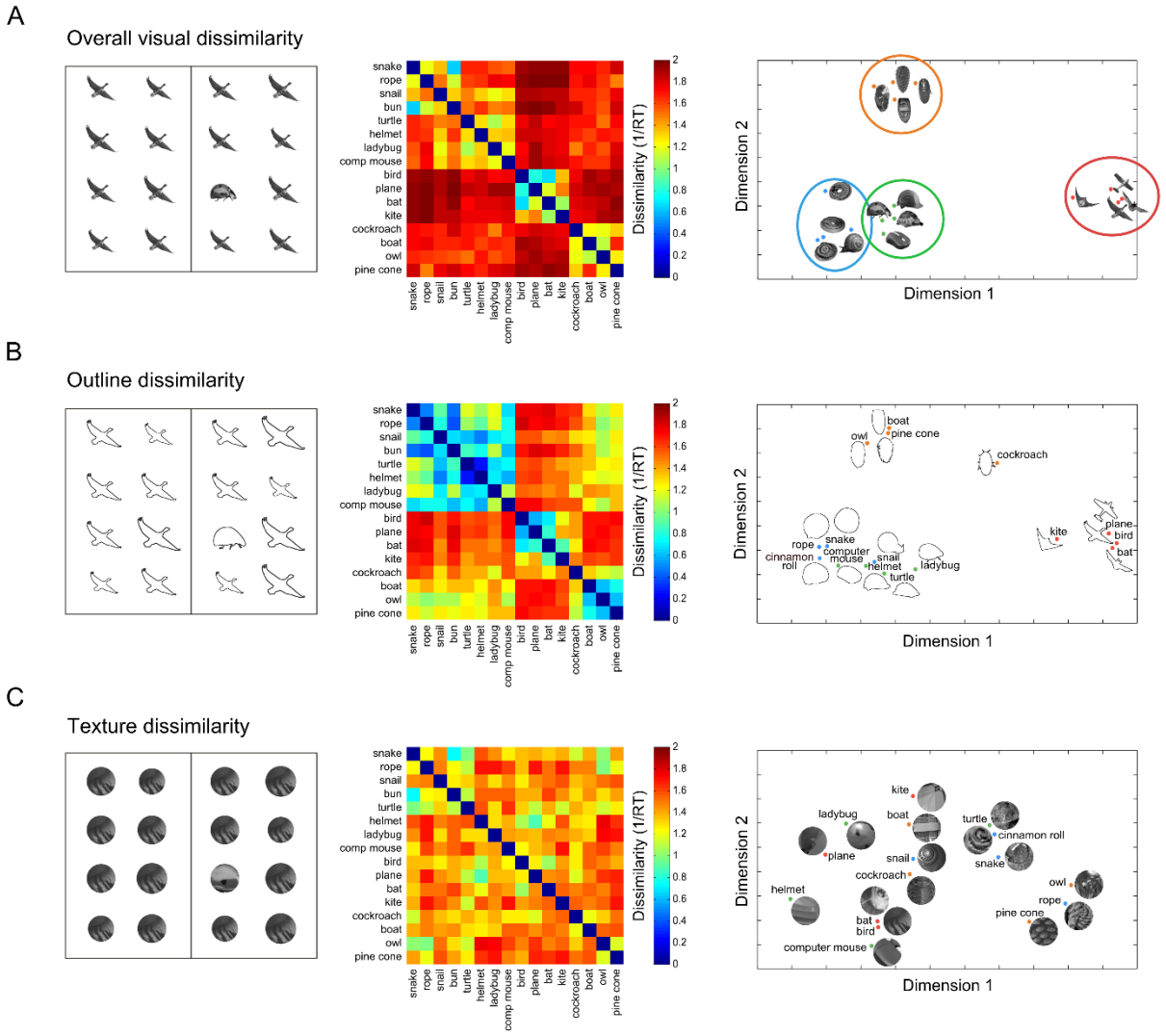



**Figure 3.1. Visual search experiments.** In three visual search experiments, participants indicated whether an oddball target in a 4 × 4 search array (left) was located to the right or to the left of the vertical display midline. No prior information about the target was given, so that participants had to rely on bottom–up visual differences to perform the task. From the RTs in these experiments, we created visual dissimilarity matrices (center), where each element represents the inverse RT for a pair of stimuli, averaged across the two respective target–distractor pairings: High dissimilarity values thus correspond to short RTs, reflecting that target and distractor were visually dissimilar. Multidimensional scaling representations of the three visual dissimilarity matrices are

shown in the right panels. The three experiments differed only in the stimuli used: (A) the original images used in the fMRI experiment, (B) outline drawings of these images, and (C) internal textures of these images.

Functional Localizer Experiment Procedure

In addition to the main experiment, participants completed one run of a functional localizer experiment. During the localizer, participants viewed grayscale pictures of 36 animate and 36 inanimate stimuli in a block design (Figure 3.2B). Animate stimuli included five different types of animals (mammals, birds, fish, reptiles, and insects). Inanimate stimuli included five types of inanimate objects (cars, chairs, musical instruments, tools, and weapons). These stimuli were not matched for their shape (thus, this design resembled the standard animate–inanimate contrast used in previous studies). Each block lasted 16 s, containing 20 stimuli that were each presented for 400 ms, followed by 400-ms blank interval. There were eight blocks of each stimulus category and four fixation-only blocks per run. The order of the first 10 blocks was randomized and then mirror reversed for the other 10 blocks. Participants were asked to detect 1-back image repetitions, which happened twice during every nonfixation block.

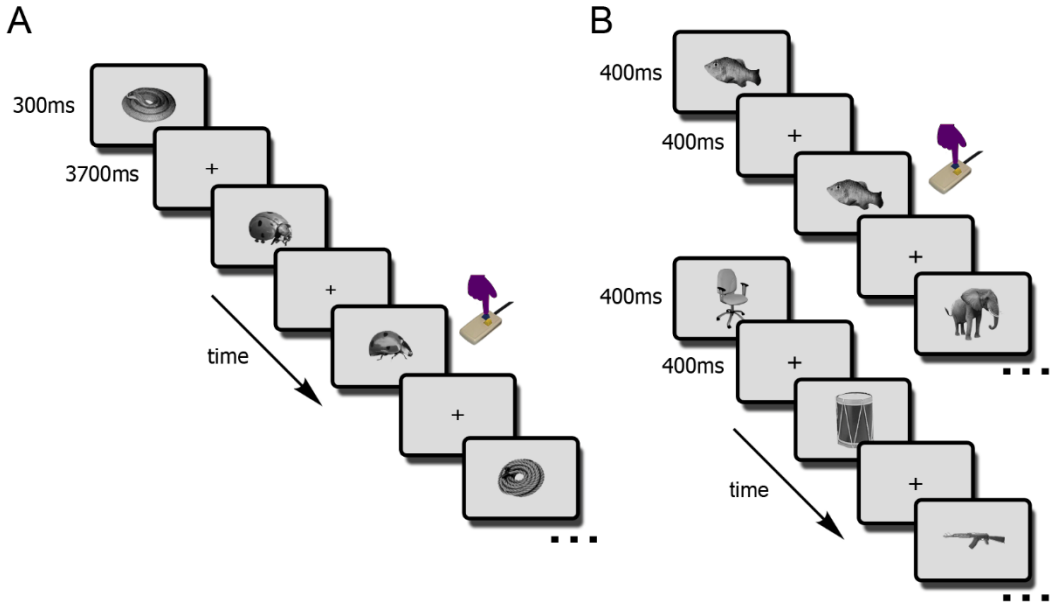



**Figure 3.2. fMRI paradigm.** (A) In the main fMRI experiment, participants viewed images of animate objects and shape-matched inanimate objects (see Chapter 2, Figure 2.1 for further

stimulus examples). Trial order was randomized, and participants detected, by button press, 1-back object-level repetitions (here, two ladybugs). (B) In the functional localizer experiment, participants viewed blocks of animate (top sequence) and inanimate (bottom sequence) stimuli. All stimuli were different from the ones used in the main experiment. Each block lasted 16 s, and participants detected, by button press, 1-back image-level repetitions (here, the fish image).

### fMRI Acquisition

Imaging data were acquired using a MedSpec 4-T head scanner (Bruker Biospin GmbH, Rheinstetten, Germany), equipped with an eight-channel head coil. For functional imaging, T2*-weighted EPIs were collected (repetition time = 2.0 sec, echo time = 33 msec, 73° flip angle, 3 × 3 × 3 mm voxel size, 1-mm gap, 34 slices, 192-mm field of view, 64 × 64 matrix size). A high-resolution T1-weighted image (magnetization prepared rapid gradient echo; 1 × 1 × 1 mm voxel size) was obtained as an anatomical reference.

### fMRI Preprocessing and Modeling

The neuroimaging data were analyzed using MATLAB and SPM8. During the preprocessing, the functional volumes were realigned, coregistered to the structural image, resampled to a 2 × 2 × 2 mm grid, and spatially normalized to the Montreal Neurological Institute 305 template included in SPM8. For the univariate analysis, the functional images were smoothed with a 6-mm FWHM kernel, whereas for the multivariate analysis, the images were left unsmoothed. For the main experiment, the BOLD signal of each voxel in each participant was modeled using 22 regressors in a general linear model, with 16 regressors for each of the objects (e.g., one regressor for all snakes) and six regressors for the movement parameters obtained from the realignment procedure. For the functional localizer data, the signal was modeled using two regressors (animate and inanimate objects) and six movement regressors. All models included an intrinsic temporal high-pass filter of 1/128 Hz to correct for slow scanner drifts.

### Univariate Analysis

Univariate random effects whole-brain analyses were performed separately for the localizer and the main experiment, contrasting animate with inanimate objects. Statistical maps were thresholded using a voxel-level threshold of $p < .001$ (uncorrected) and a cluster-level

threshold of $p < .05$ (family-wise error [FWE] corrected). In addition, regions activated in the localizer were defined as ROIs. Within these ROIs, beta estimates for the conditions of the main experiment were extracted and averaged across the voxels of each ROI. These beta values were statistically compared using ANOVAs and *t* tests.

RSA

Representational similarity analysis (RSA; Kriegeskorte, Mur, & Bandettini, 2008) was used to relate the visual and categorical dissimilarity of the objects to neural dissimilarity. RSA was performed throughout the whole brain using searchlight analysis (Kriegeskorte, Goebel, & Bandettini, 2006), implemented in the CoSMoMVPA software package (Oosterhof, Connolly, & Haxby, 2016). Each spherical searchlight neighborhood consisted of 100 voxels, centered on every voxel in the brain. For each of these spheres, we correlated the activity (beta values) between each pair of conditions from the main experiment across the voxels of the sphere, leading to a 16 × 16 symmetrical correlation matrix with an undefined diagonal. This matrix was transformed into a neural dissimilarity matrix by subtracting the correlation values from 1.

In a first analysis, neural dissimilarity matrices were related to the visual dissimilarity matrix and the categorical dissimilarity matrix using multiple regression analysis (see Figure 4.5A). The visual dissimilarity matrix was derived from RTs in a visual search experiment (Chapter 2, Experiment 1; Figure 3.1A), whereas the categorical dissimilarity matrix reflected whether two objects were from the same category (0) or from different categories (1). All dissimilarity matrices were *z* normalized. The multiple regression analysis yielded beta estimates for the two predictors of neural dissimilarity (visual and categorical dissimilarity), reflecting the independent contributions of these predictors in explaining neural dissimilarity. These two beta estimates were obtained for all spheres, resulting in two whole-brain maps for each participant. These maps were then tested against zero using random effects analyses (*t* tests), thresholded using a voxel-level threshold of $p < .001$ (uncorrected) and a cluster-level threshold of $p < .05$ (FWE corrected). In the second analysis, neural dissimilarity matrices were related to outline dissimilarity (Figure 3.1B), texture dissimilarity (Figure 3.1C), and categorical dissimilarity (Figure 3.6A). In all other respects, the analysis was the same as the first analysis described above.

## Results

Univariate Results

Whole-brain Analysis

The contrast between animate and inanimate objects in the functional localizer experiment revealed a characteristic medial-to-lateral organization in VTC (Figure 3.3A). In line with previous findings, animate stimuli more strongly activated regions around the lateral fusiform gyrus (left hemisphere [LH]: 1936 $mm^3$, peak Montreal Neurological Institute coordinates: $x$ = −40, $y$ = −48, $z$ = −22; right hemisphere [RH]: 3424 $mm^3$, peak coordinates: $x$ = 42, $y$ = −52, $z$ = −20), and inanimate stimuli preferentially activated more medial regions around parahippocampal gyrus (LH: 3760 $mm^3$, peak coordinates: $x$ = −28, $y$ = −46, $z$ = −12; RH: 3184 $mm^3$, peak coordinates: $x$ = 34, $y$ = −42, $z$ = −12). In addition to these ventral regions, animate stimuli preferentially activated a more posterior and lateral region, around middle temporal gyrus (LH: 3528 $mm^3$, peak coordinates: $x$ = −48, $y$ = −80, $z$ = 0; RH: 7800 $mm^3$, peak coordinates: $x$ = 52, $y$ = −74, $z$ = −2).

The same animate–inanimate comparison was performed for the main experiment, in which the animate and inanimate stimuli were closely matched for shape and low-level visual features (see Methods). This contrast revealed a significant animacy organization: Animate stimuli more strongly activated two clusters in the RH, again around fusiform gyrus (808 $mm^3$, peak coordinates: $x$ = 42, $y$ = −52, $z$ = −20) and middle temporal gyrus (1120 $mm^3$, peak coordinates: $x$ = 50, $y$ = −74, $z$ = 0). Similar to the functional localizer results, inanimate-preferring regions were found around bilateral parahippocampal gyrus (LH: 4736 $mm^3$, peak coordinates: $x$ = −26, $y$ = −52, $z$ = −16; RH: 3456 $mm^3$, peak coordinates: $x$ = 24, $y$ = −40, $z$ = −16). Figure 3.3, A, shows the animate- and inanimate-preferring clusters from the localizer and the main experiment as well as their overlap.

These results indicate that the medial-to-lateral animacy organization is also found when controlling for shape differences of animate and inanimate objects. As can be seen in Figure 3.3A, activity was stronger in the localizer than in the main experiment. This effect is hard to interpret, however, given the many differences between the localizer and the main experiment (e.g., block design vs. event-related design, stimulus duration, the specific animals and objects included, etc.). Indeed, the purpose of this analysis was not to compare the

strength of activity between localizer and main experiment directly but to show that the medial-to-lateral organization is remarkably similar in both experiments. Finally, although some of the left hemisphere clusters did not survive multiple comparisons correction in the main experiment, the functionally localized ROIs maintained their selectivity in the main experiment in both hemispheres, as reported in the next section.

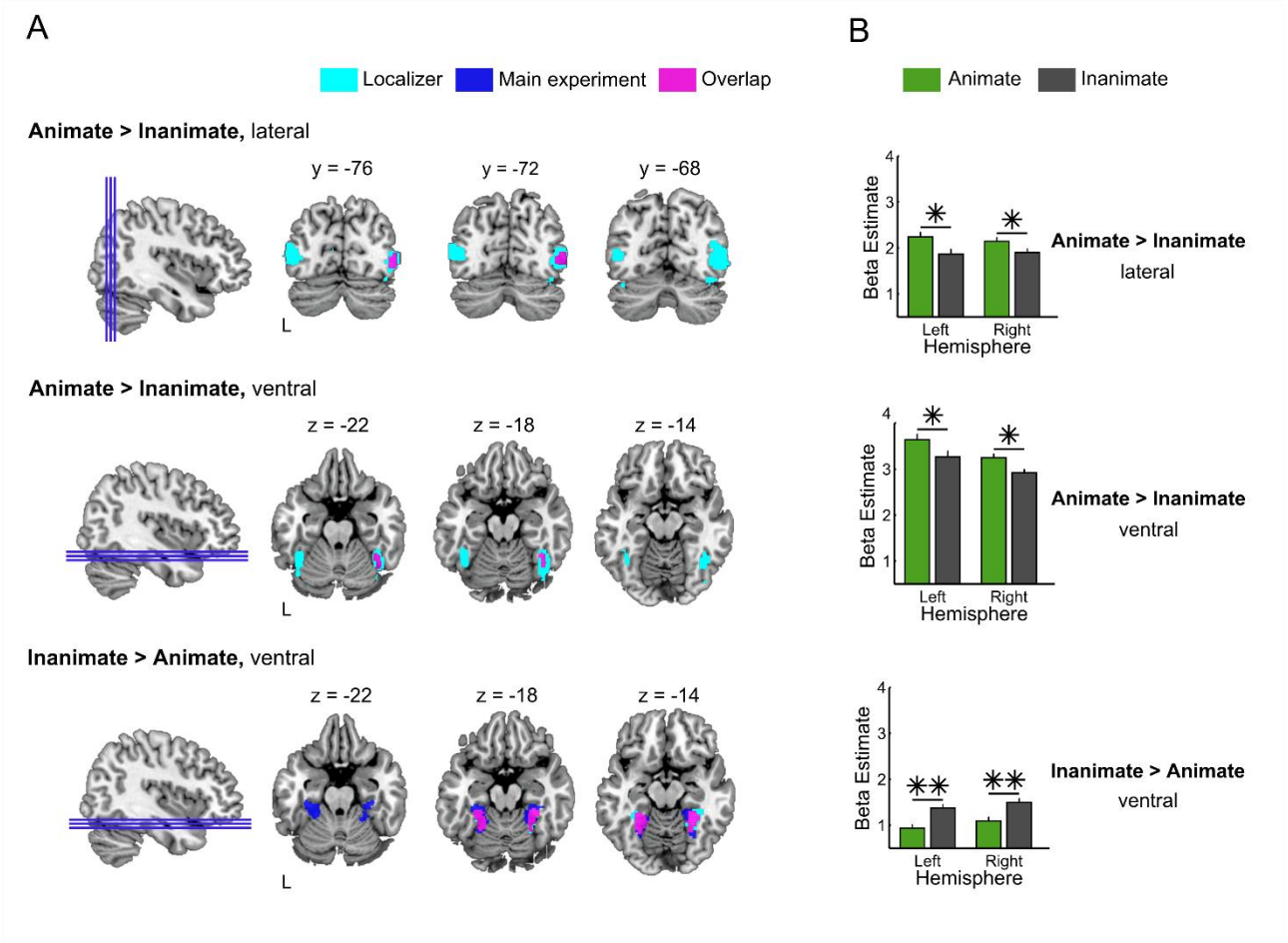



**Figure 3.3. Univariate results.** (A) Results of univariate whole-brain group analysis comparing animate and inanimate conditions in the functional localizer experiment (cyan) and main experiment (blue). Overlapping regions are displayed in purple. Top row shows lateral animate-preferring clusters, middle row shows ventral animate-preferring clusters, and bottom row shows ventral inanimate-preferring clusters. Statistical maps were thresholded at $p < .05$ (FWE corrected) and overlaid on a structural brain template (MRIcron). (B) Results of ROI analyses. The bar graphs show the responses to the conditions in the main experiment within the animate- and inanimate-preferring clusters defined based on the functional localizer experiment (corresponding to the

cyan regions in A), separately for each hemisphere. Green bars indicate average response to animate objects; gray bars indicate average response to inanimate objects. Error bars reflect *SEM* difference. *$p$ < .01, **$p$ < .001.

ROI Analysis

ROI analyses were used to test for selectivity for the conditions in the main experiment within each of the six clusters of the functional localizer (Figure 3.3, A; for coordinates and cluster sizes, see the whole-brain analysis section above). A three-way ANOVA with the factors Animacy (animate, inanimate), Region (animate lateral, animate ventral, inanimate ventral), and Hemisphere (right, left) revealed a critical Region × Animacy interaction ($F$(2, 32) = 42.2, $p$ < .001). Because there were no interactions with hemisphere ($F$ < 2.49, $p$ > .10, for all tests), data were collapsed across hemispheres for all follow-up analyses. Separate Region × Animacy ANOVAs for every pair of regions revealed that the animacy preferences of the two animate-preferring regions were each significantly different from that of the inanimate-preferring region (ventral animate region: $F$(1, 16) = 89.24, $p$ < .001; lateral animate region: $F$(1, 16) = 45.65, $p$ < .001). There was no significant difference in animacy preference between the two animate regions ($F$(1, 16) = 0.59, $p$ = .45). As expected, both animate-preferring regions showed an increased response to the animate stimuli in the main experiment ($t$(16) > 3.53, $p$ < .003, for both tests), whereas the inanimate-preferring region showed a significant preference for the inanimate stimuli ($t$(16) = 5.25, $p$ < .001). These results show that all regions defined in the functional localizer maintained their selectivity in the main experiment (see Figure 3.3, B for results in separate hemispheres).

To explore whether the animate–inanimate organization in the main experiment was driven by some of the stimuli preferentially, we next compared the responses to each of the eight individual animate stimuli and their shape-matched inanimate counterparts. For simplicity and for optimal statistical power, the conditions were recoded into “preferred” (e.g., a snake for an animate-preferring region) and “nonpreferred” conditions (e.g., a snake for an inanimate-preferring region), and responses were then averaged across ROIs, so that each ROI contributed equally. A two-way ANOVA with the factors Preference (preferred, nonpreferred) and Object pair (the eight different pairs) revealed a significant interaction ($F$(7, 112) = 3.68, $p$ = .001), indicating that different object pairs differentially contributed to

the observed animacy organization (Figure 3.4). A significant category preference was observed for six of the eight pairs ($t(16) > 2.16$, $p < .047$, for all tests), with one pair (ladybug–computer mouse) showing a trend ($t(16) = 2.10$, $p = .052$) and one pair (snail–bun) not reaching significance ($t(16) = 0.22$, $p = .83$).

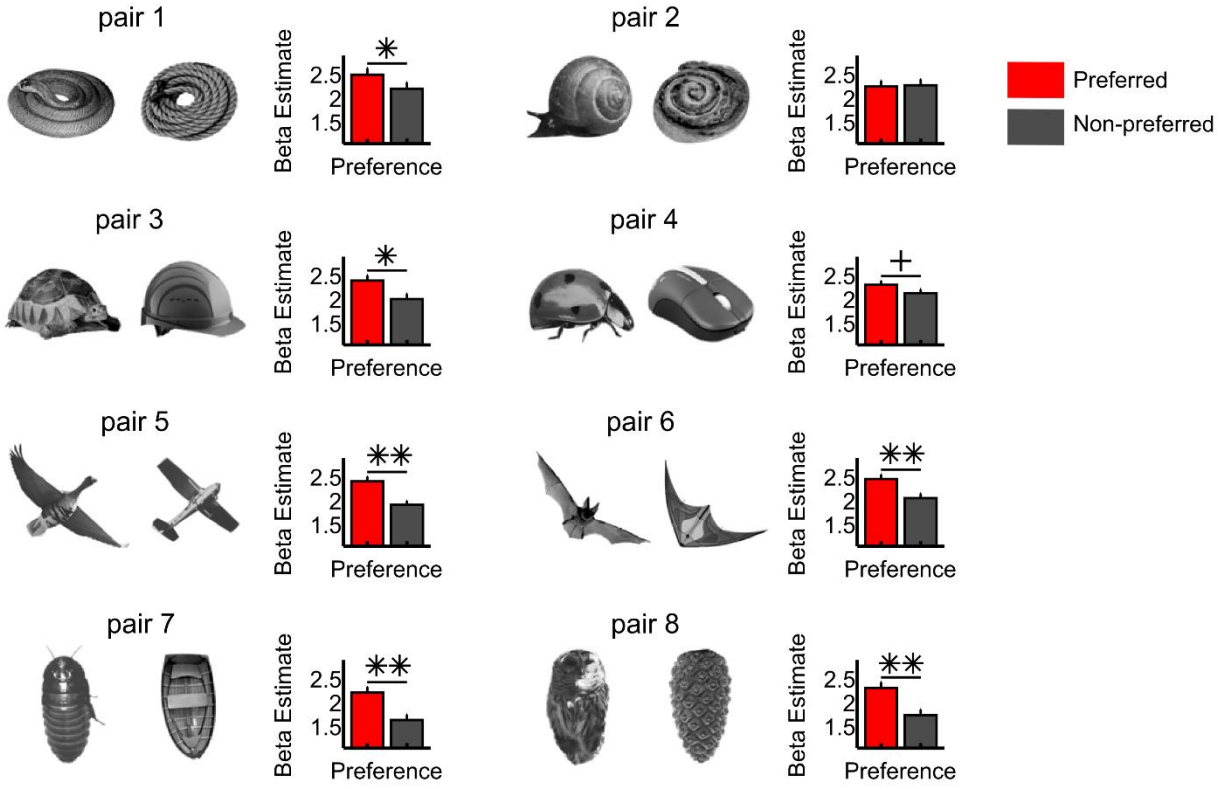



**Figure 3.4. Individual pair analysis.** Category preference analyzed for each shape-matched pair separately. Responses were combined across all regions defined in the localizer experiment (cyan regions in Figure 3.3, A) by recoding responses according to the preference of the region (e.g., snake response in animate-preferring regions and rope response in inanimate-preferring region both contribute to the "preferred" condition, shown in red). Error bars reflect *SEM* difference. +$p$ = .052, *$p$ < .05, **$p$ < .001.

RSA

In addition to showing overall differences in focal regions, objects of different categories evoke distinct multivoxel activity patterns in visual cortex: Within VTC, activity patterns to objects from the same category are more similar than activity patterns to objects from different categories (Haxby et al., 2001). An open question is whether these effects reflect visual differences and/or categorical differences between objects. To address this question, we used RSA to relate neural dissimilarity (based on correlations between multivoxel activity patterns) to visual and categorical dissimilarity. In this analysis, for every spherical neighborhood (100 voxels) of the brain, the pairwise neural dissimilarity structure between the 16 objects was modeled as a linear combination of their pairwise categorical dissimilarity and overall visual dissimilarity (Figure 3.5A; see Methods). Overall visual dissimilarity was quantified using response times in a visual search task (Figure 3.1A), capturing all contributing factors to visual discriminability (potentially beyond the ones we explicitly matched in the univariate comparisons). This analysis yielded two beta maps reflecting the independent contributions of visual and categorical variables in accounting for neural dissimilarity.

Random effects group analysis revealed widespread clusters of voxels in which neural dissimilarity was significantly related to overall visual dissimilarity, including primary visual cortex and large parts of extrastriate visual cortex, extending into VTC (51,864 $mm^3$ in total; peak in the lingual gyrus, $x = -18$, $y = -92$, $z = -8$). These clusters partly overlapped with the animate- and inanimate-preferring regions of the functional localizer experiment (cyan regions in Figure 3A), with visual dissimilarity being significant in 64% of the ventral animate-preferring voxels, 24% of the lateral animate-preferring voxels, and 3% of the ventral inanimate-preferring voxels. Interestingly, categorical dissimilarity was independently reflected in two clusters: one in the right ventral visual cortex (3512 $mm^3$; peak in the fusiform gyrus, $x = 42$, $y = -60$, $z = -18$) and one in the LH (4400 $mm^3$ in total, including a lateral visual cortex part with a local peak in middle occipital gyrus, $x = -42$, $y = -80$, $z = 6$, and a ventral visual cortex part with a local peak in fusiform gyrus, $x = -44$, $y = -52$, $z = -16$). As can be seen in Figure 5B, these clusters partly overlapped with the anterior end of the overall visual dissimilarity clusters. These clusters also partly overlapped with the animate- and inanimate-preferring regions of the functional localizer experiment, with

categorical dissimilarity being significant in 33% of the ventral animate-preferring voxels, 10% of the lateral animate-preferring voxels, and 2% of the ventral inanimate-preferring voxels.

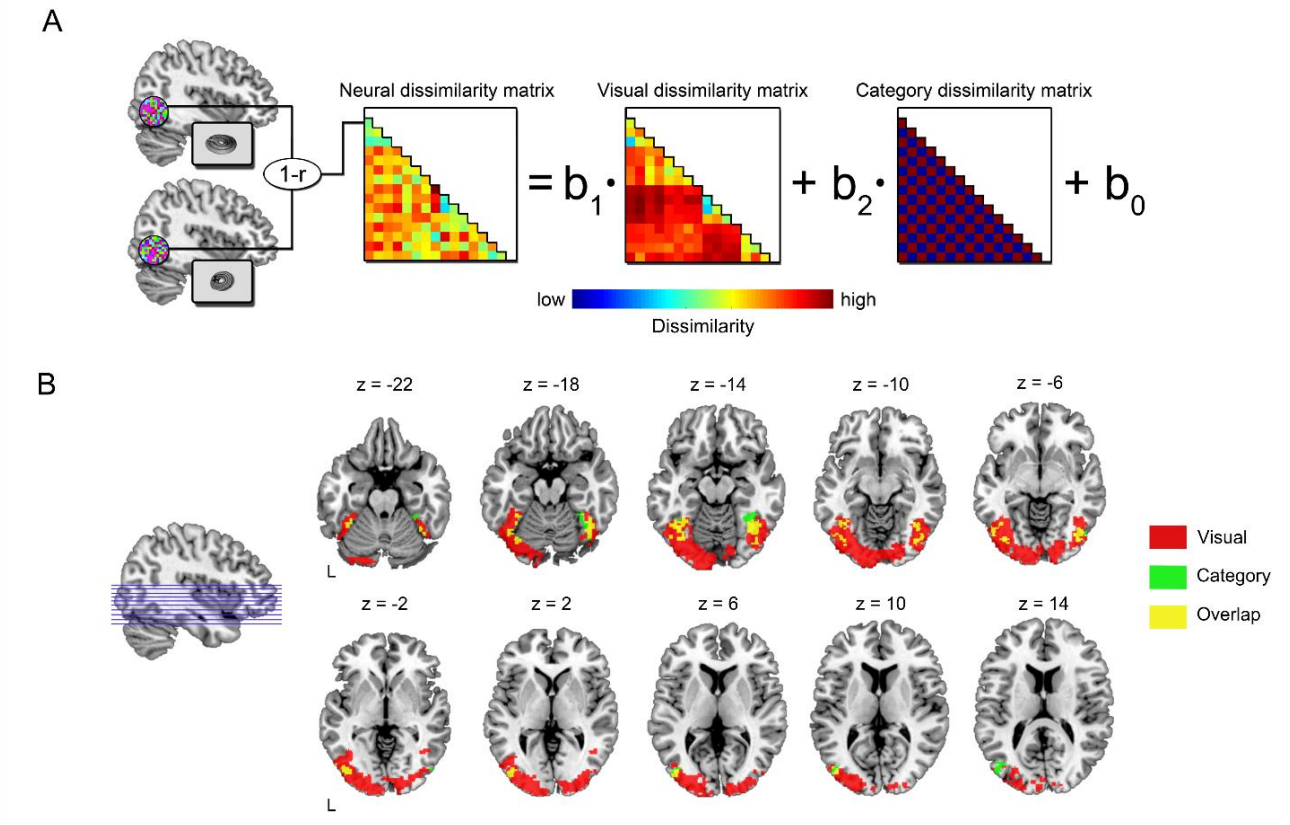



**Figure 3.5. Multivariate searchlight analysis: overall visual dissimilarity and categorical dissimilarity.** (A) Schematic of the linear modeling approach. For every spherical searchlight neighborhood, a 16 × 16 neural dissimilarity matrix was constructed using pairwise correlations of multivoxel activity patterns. These pairwise dissimilarity values (1 − *r*) were modeled by a linear combination of two predictors: One predictor was derived from the RTs in the visual search experiment that served as a proxy for overall visual dissimilarity of the stimuli (Figure 3.2A), and the other predictor reflected pairwise categorical (animate vs. inanimate) dissimilarity. This procedure tested the extent to which the neural dissimilarity structure in a given sphere reflected overall visual dissimilarity (while regressing out categorical dissimilarity) and/or categorical dissimilarity (while regressing out overall visual dissimilarity). (B) Results of whole-brain group analyses testing the value of each predictor versus zero. The analysis identified a large cluster of voxels where neural dissimilarity reflected overall visual dissimilarity (red and yellow voxels),

spanning early visual cortex and extrastriate regions up to VTC. In addition, clusters were identified where neural dissimilarity was independently predicted by the category of the stimuli (yellow and green voxels): These clusters partly overlapped with the overall visual dissimilarity clusters (yellow voxels) and showed local peaks in right and left fusiform gyrus and left middle occipital gyrus. Statistical maps were thresholded at $p < .05$ (FWE corrected) and shown as binary color maps overlaid on a structural brain template (MRIcron).

It is possible that the measure of overall visual dissimilarity captured some visual features better than others. For example, the overall visual dissimilarity structure could be driven more by the outline shape than by texture properties of the stimuli. In this case, if animate and inanimate stimuli consistently differed in their texture, the category-selective regions revealed in the previous analysis could in principle reflect texture information rather than category information. To address this possibility, we quantified the pairwise dissimilarity structure for outline shape and texture independently from each other in two further visual search experiments (see Methods). Using these data, we repeated the RSA, this time modeling pairwise neural dissimilarity using the combination of three predictors: outline shape dissimilarity (Figure 2B), texture dissimilarity (Figure 2B), and categorical dissimilarity (see Figure 6A). This analysis resulted in three beta maps reflecting the independent contributions of outline shape, texture, and category membership to the neural dissimilarity structure.

The results of random effects group analyses on these three maps are shown in Figure 6B. Four widespread clusters of voxels in which neural dissimilarity reflected outline dissimilarity were found in primary visual cortex extending to extrastriate visual cortex (22,584 $mm^3$, occipital and temporal lobe; peak around the left lingual gyrus, $x = -22$, $y = -92$, $z = -8$), right superior parietal cortex (peak $x = 16$, $y = -50$, $z = 60$), right ventral visual cortex (1552 $mm^3$; peak $x = 44$, $y = -64$, $z = -12$), and left temporal lobe around fusiform gyrus (1240 $mm^3$; peak in BA 37, $x = -28$, $y = -46$, $z = -18$). Texture dissimilarity was related to neural dissimilarity in one cluster of voxels in left early visual cortex (2456 $mm^3$; peak in the lingual gyrus, $x = -20$, $y = -84$, $z = -14$). Crucially, categorical dissimilarity was independently reflected in two clusters: one in right ventral visual cortex (2928 $mm^3$; peak in the fusiform gyrus, $x = 42$, $y = -60$, $z = -18$) and one in left lateral visual cortex (3960 $mm^3$; peak in the middle occipital gyrus, $x = -42$, $y = -80$, $z = 6$). Adding overall visual dissimilarity to this analysis as fourth predictor revealed nearly identical category clusters, as would be expected based on the

finding that a linear combination of outline and texture dissimilarity nearly perfectly captured overall visual dissimilarity (see Methods), thus making this variable redundant as additional predictor.

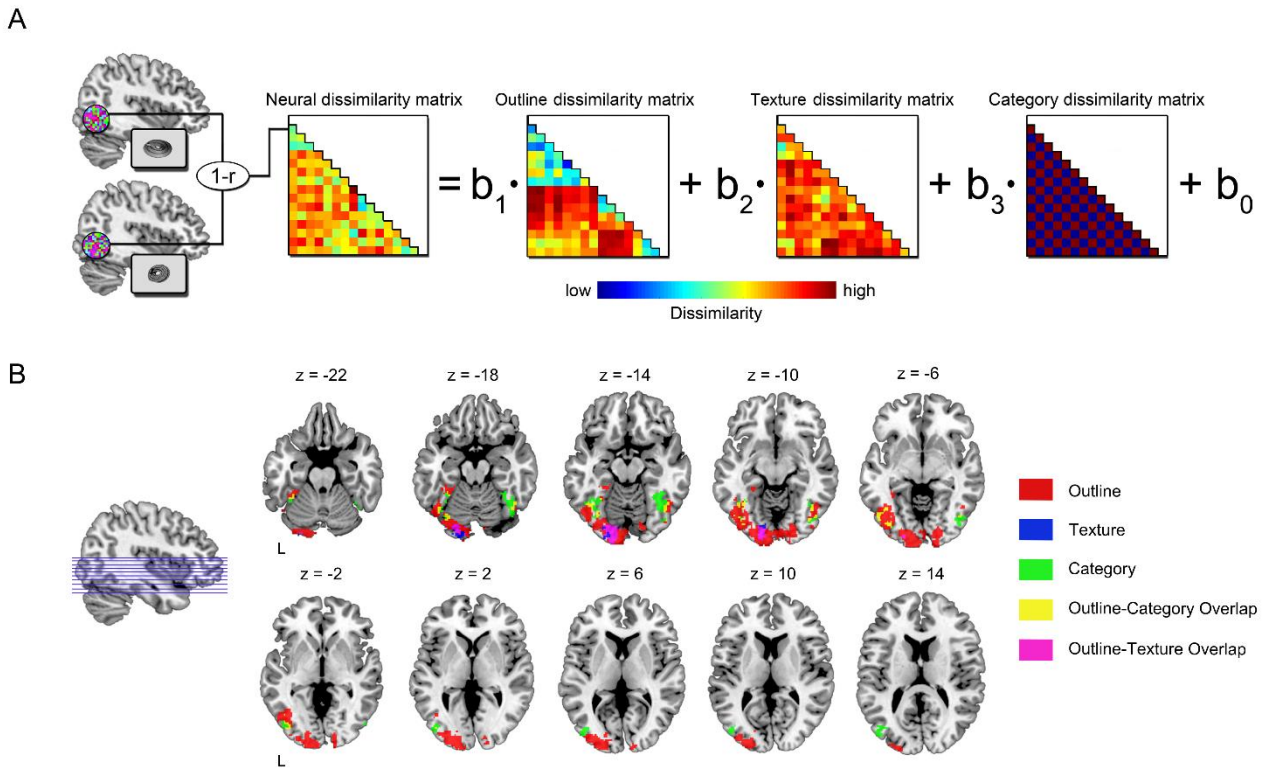



**Figure 3.6. Multivariate searchlight analysis: outline, texture, and categorical dissimilarity.** (A) Schematic of the linear modeling approach similar to the one shown in Figure 3.5 (see the caption for details). In this analysis, neural dissimilarity was modeled using the combination of three predictors: outline dissimilarity, texture dissimilarity, and categorical dissimilarity. (B) Results of whole-brain group analyses testing the value of each predictor versus zero. Clusters of voxels in which neural dissimilarity was predicted by outline dissimilarity (red, yellow, and pink voxels) were found in primary visual cortex, extending into extrastriate visual cortex. Texture dissimilarity predicted neural dissimilarity in a cluster in left early visual cortex (blue and pink voxels), partly overlapping with an outline dissimilarity cluster (overlap indicated in pink). Finally, this analysis revealed clusters of voxels in which neural dissimilarity was predicted by categorical dissimilarity (green and yellow voxels), with local peaks in right fusiform gyrus and left middle occipital gyrus. These clusters partly overlapped with outline dissimilarity clusters (overlap indicated in yellow). There was no overlap between texture dissimilarity clusters and categorical dissimilarity clusters.

Statistical maps were thresholded at $p < .05$ (FWE corrected) and shown as binary color maps overlaid on a structural brain template (MRIcron).

In summary, these results reveal distinct but overlapping representations for outline shape and texture in early visual cortex. Importantly, categorical representations in higher level visual cortex were still present even after regressing out both outline and texture dissimilarity.

## Discussion

In this study, we asked whether the animate–inanimate organization of object responses in human VTC reflects characteristic visual properties of animate and inanimate objects (e.g., characteristic animal shapes or textures) or whether it (partly) reflects a true categorical organization. We approached this question by testing whether the animate–inanimate organization can still be observed when controlling for visual similarity of objects from animate and inanimate domains. A standard functional localizer experiment contrasting activity to a variety of animals with activity to a variety of inanimate objects replicated previous studies, showing animate- and inanimate-preferring regions in VTC and animate-preferring regions in lateral occipitotemporal cortex. Importantly, all of these regions, in both hemispheres, remained selective for their preferred category in the main experiment in which animate and inanimate objects were carefully matched for shape as well as for low-level features such as luminance and contrast. Results were consistent across all but one of the eight animate–inanimate pairs. (We speculate that the lack of preference for the snail condition may relate to snail shells frequently being experienced as inanimate objects, because they are often viewed without an animal inside.) These pairs varied widely in terms of their shape (e.g., snake vs. bird), further supporting the claim that specific shape properties (e.g., presence of limbs) do not fully account for the animate–inanimate organization in VTC. Finally, the inanimate objects also varied widely on various conceptual dimensions that have been linked to inanimate-preferring regions, such as real-world size (Konkle & Oliva, 2012) and manipulability (Mahon et al., 2007). The consistency of results across the pairs suggests that the animate–inanimate organization revealed here is not fully explained by such alternative conceptual properties.

The objects that were contrasted in the univariate analyses were matched for visual similarity by the experimenters. This approach is subjective and assumes that visual similarity can be accurately judged through visual inspection. An important additional aspect of our study was therefore the use of behavioral visual search tasks to quantify different aspects of visual similarity in a naive group of participants. On each trial, participants simply indicated the location of the unique stimulus in an array of identical distractors; that is, there was no predefined target category. Visual differences are the only source of information to locate the target in this task, such that performance (RT) closely reflects the visual similarity of the target and distractor stimuli (Mohan & Arun, 2012). In the first experiment, we measured the visual similarity of the same images used in the fMRI experiment. These data potentially capture not only differences in outline shape but also any other visual property (e.g., texture, extent, spatial frequency) that helps to visually distinguish the target from the distractor, rendering it a measure of overall visual similarity. Moreover, in two additional experiments, we specifically measured outline similarity (using outline drawings) and texture similarity (using texture patches).

RSA with overall visual similarity and category similarity as predictors revealed that activity patterns throughout visual cortex reflected visual similarity, confirming that visual similarity is a dominant organizing principle of both low- and high-level visual cortex (e.g., Andrews et al., 2015). The overall visual similarity matrix derived from the visual search task was additionally used to regress out variance in the neural similarity matrices, testing whether any remaining variance can be attributed to categorical similarity. This analysis revealed clusters in VTC in which this was the case. Similar results were obtained when modeling outline and texture similarity separately as predictors of the neural similarity structure. Although both texture and outline shape where represented independently in visual cortex, category information was still present in some regions. In both analyses, these category clusters were found in the vicinity of animate- and inanimate-preferring regions.

Although our results provide evidence for an animate–inanimate organization of VTC that is not explained by outline shape or texture, we should consider the possibility that there may be remaining visual features that distinguished animals from objects. Clearly, in the absence of other cues, there must be visual properties that allow the observer to recognize the objects and to distinguish, for example, between a snake and a rope—we do not claim that there are

no visual differences between the two objects of each pair. However, it seems unlikely that there were visual features that consistently covaried with category membership across pairs. Furthermore, such consistent features would likely be reflected in the visual similarity measures (Mohan & Arun, 2012) and thus regressed out in the representational similarity analyses. Nevertheless, we acknowledge that we cannot fully exclude that there may be residual visual differences between animals and inanimate objects that do not affect visual similarity as measured in the visual search experiments. For example, it is possible that certain category-specific shape features are not visually salient (and may not even be visible in the image) but become represented once an object is recognized as an animal (e.g., head and eyes).

Interestingly, clusters representing categorical similarity partly overlapped with clusters representing shape similarity at higher levels of the visual system (yellow clusters in Figures 5.5B and 5.6B). This suggests that a shape-based organization coexists with a category-based organization, with neither of these two reducible to the other. This coexistence suggests close mutual interactions between shape and category representations. In one direction, shape properties strongly inform category membership in most real-world situations. For example, the set of midlevel visual features that characterize animals allows for efficiently detecting the presence of an animal in a natural scene (Ullman, Vidal-Naquet, & Sali, 2002; Thorpe, Fize, & Marlot, 1996). In the other direction, category membership provides information about likely visual properties of an object, such as the structure of its parts, and allows for making perceptual predictions, for example, related to characteristic motion patterns of animals. The close proximity and partial overlap of shape and category representations may thus be optimal for real-world behavior in which these levels of representation need to closely interact.

Previous findings have shown that the degree to which a stimulus evokes an "animate" response in VTC depends on the degree to which it shares characteristics with the animate prototype—humans (Sha et al., 2014). This is consistent with earlier findings of strong selectivity for human faces and bodies at the approximate locations of the ventral and lateral animacy clusters in our study (Peelen & Downing, 2005; Downing et al., 2001; Kanwisher et al., 1997). These regions show a graded response profile, responding most strongly to human faces and bodies, followed by mammals, birds, reptiles, and insects (Downing et al., 2006).

These findings are consistent, however, both with a visual similarity interpretation (i.e., differences in visual typicality; Mohan & Arun, 2012) and a conceptual similarity interpretation (e.g., differences in agency; Sha et al., 2014). Therefore, future work is needed to independently manipulate the degree to which animals share visual and conceptual properties with humans to test whether graded animacy effects are primarily reflecting one or both of these properties. Interestingly, our current results show that a reliable animal preference exists even for animals that are visually and conceptually distinct from humans (e.g., snakes, insects).

Together, the present results suggest that the animate–inanimate organization of VTC is not fully explained by local biases for visual features. Instead, we interpret this organization as reflecting the recognition of an object as belonging to a particular domain. In daily life, visual properties are an important cue for categorizing objects, but many other cues also contribute. These cues include information from other modalities (e.g., audition, touch) and, more generally, our expectations, knowledge, goals, and beliefs. Rather than following the visual features falling on the retina, category-selective activity in VTC appears to partly reflect the interpretation, based on all available cues, that the object we look at is animate or inanimate. On this account, category-specific activity that is independent of visual features would reflect a relatively late stage in the object recognition process. Future work could use multivariate analysis of magnetoencephalography data (e.g., Cichy, Pantazis, & Oliva, 2014; Carlson, Tovar, Alink, & Kriegeskorte, 2013) to reveal the temporal dynamics of object categorization using carefully designed stimuli that allow for disentangling visual and categorical similarity. One prediction consistent with our results would be that the initial response in VTC primarily reflects visual similarity, with later stages additionally reflecting category membership. In Chapter 4, we describe an MEG study we conducted to test this prediction.

If not visual features, then what property might drive the animate–inanimate distinction? One proposal is that this distinction reflects agency: the potential of an object to perform self-initiated, complex, goal-directed actions (Sha et al., 2014; Caramazza & Shelton, 1998; Premack, 1990). For example, studies have shown that activity in the right fusiform gyrus—at the approximate location of the ventral animate-preferring region—can be evoked by simple geometric shapes that, through their movements, are interpreted as social agents (Gobbini, Koralek, Bryan, Montgomery, & Haxby, 2007; Martin & Weisberg, 2003; Schultz et al., 2003;

Castelli, Happe, Frith, & Frith, 2000). Other work consistent with this account has shown that animal selectivity in VTC is strongest for animals, such as mammals, that are perceived as having relatively more agentic properties (Sha et al., 2014).

In summary, the present results suggest that the animate–inanimate organization of VTC may not fully reflect visual properties that characterize animals and objects. Results from RSA indicate that visual and categorical representations coexist in more anterior parts of the visual system. Clearly, future work is needed to further exclude the possibility of confounding visual features, to define exactly what dimensions drive the animate–inanimate distinction, to reveal the time course of visual and categorical representations, and to test how these interact to allow for efficient object categorization in our daily life environments.

**Acknowledgements**

We thank Nick Oosterhof for help with data analysis. The research was funded by the Autonomous Province of Trento, Call "Grandi Progetti 2012," project "Characterizing and improving brain mechanisms of attention—ATTEND."

# Chapter 4

# Disentangling object shape and object category with MEG


## Abstract

Recent studies have shown that object category can be decoded from MEG sensor patterns, revealing the time course of object categorization with a millisecond resolution. However, objects belonging to the same category often share characteristic shape features (e.g., most mammals have four legs), so that category decoding in these studies could have partially reflected shape similarity rather than categorical similarity. In the present study, we aimed to disentangle the contributions of object shape and object category to the MEG signal by using animate and inanimate objects that were closely matched for shape and low-level visual features (e.g., snake-rope). In a series of behavioral visual search experiments, different aspects of visual similarity of these objects were quantified (overall similarity, outline similarity, texture similarity). In a previous fMRI study using these stimuli, we found that both shape and category are represented in ventral temporal cortex. Following the analysis approach used in the fMRI study, neural dissimilarity of MEG sensor patterns was modeled using regression analysis, where visual dissimilarity (taken from the search experiments) and categorical dissimilarity served as predictors of neural dissimilarity. The results show that visual object properties are reflected in MEG patterns starting from 80 ms after stimulus onset, with a peak at 110-130 ms. Surprisingly, when regressing out the contribution of visual properties, no residual category information was present in MEG response patterns. These results suggest that MEG sensor patterns evoked by visually presented objects predominantly reflect visual object properties.

## Introduction

Category-specific responses in the brain have been studied extensively using fMRI, resulting in the discovery of cortical areas that respond preferentially to stimuli belonging to a particular category, such as faces or scenes (see Kanwisher, 2010, for review). Moreover, studies using multivariate pattern analysis techniques revealed distributed representations of various object categories in the ventral visual stream (Haxby et al., 2001), with a prominent distinction between representations of animate and inanimate objects (Kriegeskorte et al., 2008b).

While fMRI allows one to examine the fine-grained structure of object representations with high spatial resolution, other neuroimaging techniques such as magneto- and electroencephalography (MEG/EEG) have the potential to reveal the temporal dynamics of category processing in the brain with a millisecond resolution. One example of category-related effect in MEG/EEG data is M170/N170, an ERP component that is greater for faces compared with other object categories (Bentin et al., 1996; Xu et al., 2005). In recent years, multivariate decoding methods have been applied successfully to MEG data analysis, with several studies reporting reliable decoding of different object categories from MEG sensor patterns (e.g., Cichy et al., 2014). In particular, the distinction between animate and inanimate objects in the MEG signal reportedly emerges as early as 80 ms, peaking between 150 and 250 ms (Carlson et al., 2013; Cichy et al., 2014). However, it remains unclear whether the above-chance decoding of object category information reflects the semantic differences between categories or the lower-level visual differences between them.

Several recent studies have addressed this issue by trying to disentangle object category and object shape information reflected in MEG/EEG sensor patterns. One study did this by comparing how well a model that includes both visual and semantic information predicts MEG data compared with the purely visual model, HMAX (Clarke et al., 2015). However, it was shown that the HMAX model is not a good predictor of MEG data, compared to the behavioral perceptual similarity model, making it possible that HMAX did not capture some of the relevant object shape features (Wardle et al., 2016). In another study, researchers compared the EEG response patterns elicited by intact and scrambled images (Coggan et al., 2016), but it is possible that some residual low-level features that were eliminated by scrambling were still driving the response to the intact objects. To our knowledge, only one study has specifically

matched stimuli from different categories for shape and low-level visual features, using images of body parts and shape-matched clothes (Kaiser, Azzalini, & Peelen, 2016), finding that object category can still be decoded even when the stimuli are matched for shape.

In this study, we used similar approach while focusing on the distinction between animate and inanimate objects. Specifically, we used the set of animate and inanimate stimuli closely matched for overall shape features to disentangle object shape and object category information reflected in the MEG sensor patterns. This stimulus set was previously used in an fMRI study, revealing category-specific responses in the ventral temporal cortex even after controlling for visual similarity (See Chapter 2; Proklova, Kaiser, & Peelen, 2016). We conducted two MEG experiments and used the representational similarity analysis (RSA) approach to examine the independent contributions of object shape and object category to the MEG signal, in order to see if MEG sensor patterns distinguish between animate and inanimate stimuli based on their category or if they primarily reflect shape similarity. Moreover, we were interested in the relative time course of visual and category information: if the category information independent of shape similarity is present in MEG signal, it should emerge in the later stages of processing.

## Materials and methods

### Participants

Twenty-nine participants in total with normal or corrected-to-normal vision took part in two experiments, including fourteen volunteers who participated in Experiment 1 (5 females, mean age = 25.6 years, SD = 4 years), and fifteen who participated in Experiment 2 (8 females, mean age = 25.2 years, SD = 3.1 years). All participants provided informed consent and received monetary compensation for their participation. The experimental protocols were approved by the Ethical Committee of the University of Trento, Italy.

### Stimuli

The stimulus set in both experiments was identical to the one used in an earlier fMRI study (Proklova, Kaiser, & Peelen, 2016, see Chapters 2 and 3) and consisted of 16 objects (8 animate and 8 inanimate) divided into 4 shape sets. Each shape set consisted of 2 animals and 2 inanimate objects that were matched for overall shape features (e.g. snake-rope). In

addition, four versions of each stimulus were used, resulting in total of 64 stimuli. In Experiment 1, one additional visual stimulus (hammer, see Figure 4.1, A) was used to serve as an oddball target. All stimuli were matched for luminance and contrast using the SHINE toolbox (Willenbockel et al., 2010).

Experiment 1: Procedure

Participants viewed the visual stimuli while sitting in the dimly lit magnetically shielded room. The stimuli were projected on a translucent screen located 150 cm from the participant. The stimuli were presented centrally on the uniformly gray background and spanned approximately 8-degree visual angle in size. The stimulus presentation was controlled using the Psychophysics Toolbox (Brainard, 1997). Participants completed 12 experimental blocks, with each of the 64 stimuli appearing exactly twice in each block, in random order. In addition, 32 target trials were randomly distributed throughout the block, resulting in 160 trials in total per block. The stimuli were presented for 500 ms, followed by a variable inter-stimulus interval ranging from 1.5 s to 2 s in discrete 20 ms steps. Participants were instructed to maintain fixation and to press the response button and blink each time they see a picture of a hammer. This oddball image appeared on average every 4 trials.

Experiment 2: Procedure

The procedure was largely identical to Experiment 1, differing only in participants' task. Instead of an oddball detection, participants were asked to perform a one-back task, pressing the response button whenever an image of the same object appeared on two consecutive trials (e.g. two snakes). The one-back repetitions happened 12 times in each block.

MEG acquisition and preprocessing

Electrophysiological recordings were obtained using an Elekta Neuromag 306 MEG system (Elekta Neuromag systems, Helsinki, Finland) equipped with 204 gradiometers and 102 magnetometers. Signals were sampled continuously at 1000 Hz rate and band-pass filtered online between 0.1 and 330 Hz. The data were then preprocessed offline using MATLAB and Fieldrip Toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011). Data from all runs were concatenated and high-pass filtered at 1Hz. Trials were epoched from 100 ms before to 600 ms after the stimulus onset. Trials containing eye-blinks and other movement-related artifacts were discarded from further analysis based on the visual inspection. The signal was baseline-

corrected with respect to pre-stimulus window and downsampled to 100 Hz to reduce the processing time and increase the signal-to-noise ratio (see Carlson, Tovar, Alink, & Kriegeskorte, 2013).

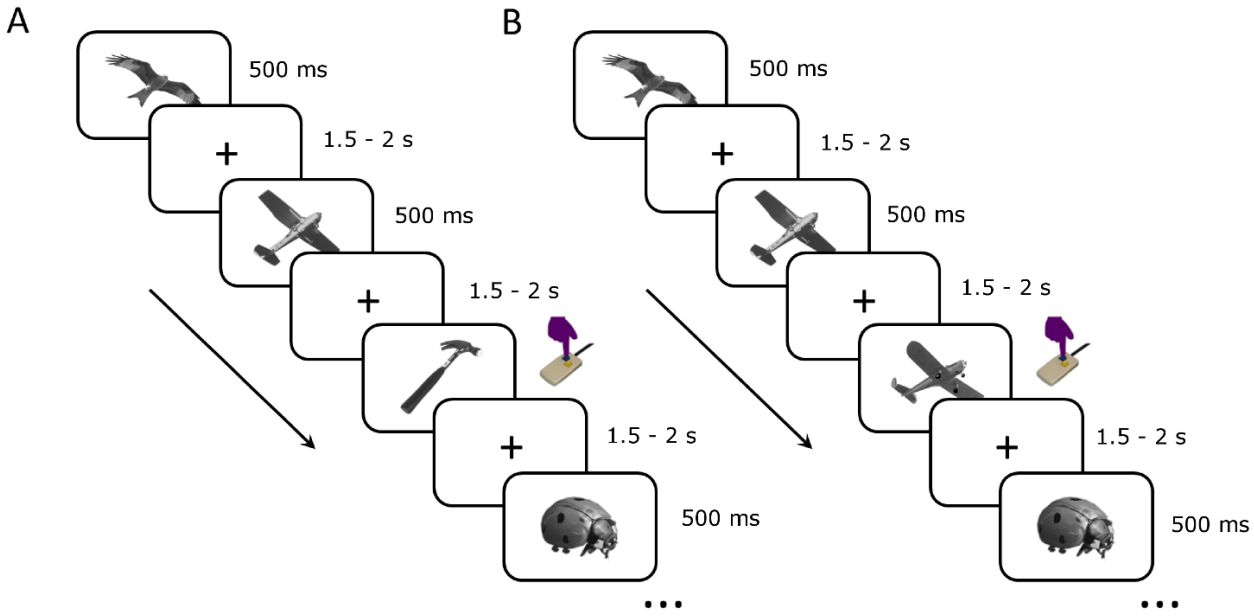



**Figure 4.1. MEG paradigms.** (A) In Experiment 1, individual stimuli were presented centrally for 500 ms, followed by variable ISI of 1.5 – 2 s. Participants were asked to press the response button and blink whenever they saw an oddball image (a hammer). (B) In Experiment 2, the procedure was largely identical to Experiment 1, except that participants performed a one-back task, pressing the button when two images of the same object (e.g., two planes) appeared on two consecutive trials.

Representational similarity analysis

In order to investigate the independent contributions of object shape and object category to the MEG signal, we used representational similarity analysis (RSA, Kriegeskorte, Mur, & Bandettini, 2008). Following the analysis approach used in a previous fMRI study (see Chapter 2; Proklova, Kaiser, & Peelen, 2016), the neural dissimilarity (measured with MEG) at each time point was modeled as a linear combination of category and visual dissimilarity. The procedures used to obtain the neural dissimilarity matrices as well as category and shape predictor matrices are described below.

*Neural dissimilarity*

Following the approach used in previous MEG decoding studies, we chose pairwise decoding accuracy as a measure of neural dissimilarity (Cichy, Pantazis, & Oliva, 2014; Wardle, Kriegeskorte, Grootswagers, Khaligh-Razavi, & Carlson, 2016). For each time point, we trained the linear discriminant analysis (LDA, Duda et al., 2011) classifier to discriminate between all possible pairs of stimuli (Figure 4.2A). For the classification analysis, trials were randomly assigned to four independent chunks, with an equal number of trials for each condition per chunk. Classifiers were trained on data from three chunks and tested on the fourth chunk; this analysis was repeated four times, with each chunk serving as the test set once. To reduce trial-to-trial noise and increase the reliability of data supplied to the classifier, new, "synthetic" trials were created: for every condition and chunk we randomly picked 25% of the original trials and averaged the data across them. This procedure was repeated 500 times (with the constraint that no trial was used more than one time more often than any other trial), producing 500 new trials for each condition and chunk that were then supplied to the classifier. The resulting decoding accuracy time course was smoothed with an averaging filter spanning 3 time points (30 ms). The classifier accuracy at each time point was assessed as the percentage of correctly classified trials (with chance performance being 50%) and was used as a measure of neural dissimilarity between the two conditions. This classification procedure was performed for all pairs of conditions, resulting in a 16 x 16 neural dissimilarity matrix for each time point (Figure 4.2B).

In addition, to provide an alternative neural dissimilarity measure we used Euclidean distance as a measure of pairwise dissimilarity between the stimuli. For each time point, the neural dissimilarity for a pair of conditions was defined as the Euclidean distance between the corresponding channel activity pattern vectors, using *cosmo_pdist* function implemented in CoSMoMVPA toolbox (Oosterhof et al., 2016). This resulted in a 16 x 16 neural dissimilarity matrix for each time point.

*Visual dissimilarity*

Three visual dissimilarity matrices quantifying overall visual dissimilarity, outline dissimilarity, and texture dissimilarity were obtained from separate behavioral visual search experiments (see Chapter 2). In each of the three experiments, participants had to detect an oddball target among the array of identical distractors. This was done for all possible target-distractor pairs among the stimuli. For each pair of stimuli, the corresponding entry of this matrix is the inversed reaction time in a visual search task, using one of the stimuli as a target, and another as a distractor.

*Category dissimilarity*

A category dissimilarity matrix was constructed by assigning zeroes (minimum dissimilarity) to elements of the matrix that corresponded to the pair of objects belonging to the same category, and ones (maximum dissimilarity) for the pairs of objects from different categories.

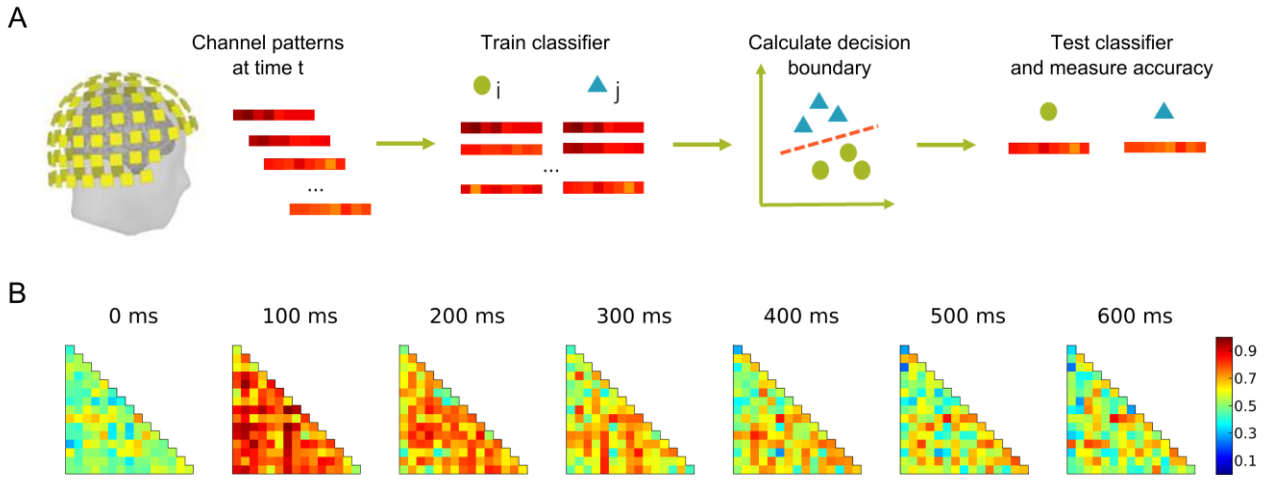



**Figure 4.2. MEG decoding approach.** (A) A schematic of the MEG decoding approach. At each time point, we extracted channel patterns and trained an LDA classifier to discriminate between each pair of conditions using a subset of trials (training set). After the decision boundary was calculated, we tested the classifier on the left-out trials (testing set) and calculated the decoding accuracy, with chance performance being 0.5. This pairwise decoding accuracy was used to create 16 x 16 decoding accuracy matrices for every time point. (B) Decoding accuracy matrices for different time points in one representative subject. At 0 ms, the pairwise decoding accuracies are close to chance

(0.5), but at later time points it rises above chance for some object pairs. Because these matrices are symmetric, only the lower triangles of the matrices are shown.

After constructing neural dissimilarity matrices for each time point, as well as category and visual dissimilarity predictor matrices, we performed two representational similarity analyses (RSA) to determine the contribution of category dissimilarity and different aspects of visual dissimilarity to the MEG signal. In the first analysis, the neural dissimilarity at each time point was modeled as a linear combination of category and overall visual dissimilarity (see Figure 4.3A). This analysis was performed separately for each participant, resulting in two regression weights for each time point per participant. In the second analysis, the neural dissimilarity at each time point was modeled as a linear combination of category, outline, and texture dissimilarity (see Figure 4.3B). This produced three beta weights per participant for each time point. The time courses of the resulting beta estimates for each predictor were then tested against zero. The significance was assessed using the threshold-free cluster enhancement procedure (TFCE; Smith & Nichols, 2009) with default parameters, using multiple-comparison correction based on a sign-permutation test (with null distributions created from 10,000 bootstrapping iterations) as implemented in CoSMoMVPA toolbox (Oosterhof, Connolly, & Haxby, 2016). The resulting z-score time courses were the thresholded at $Z > 1.64$ (i.e., $p < 0.05$, one-tailed) to reveal the time points for which the beta estimate of a given predictor was significantly above zero.

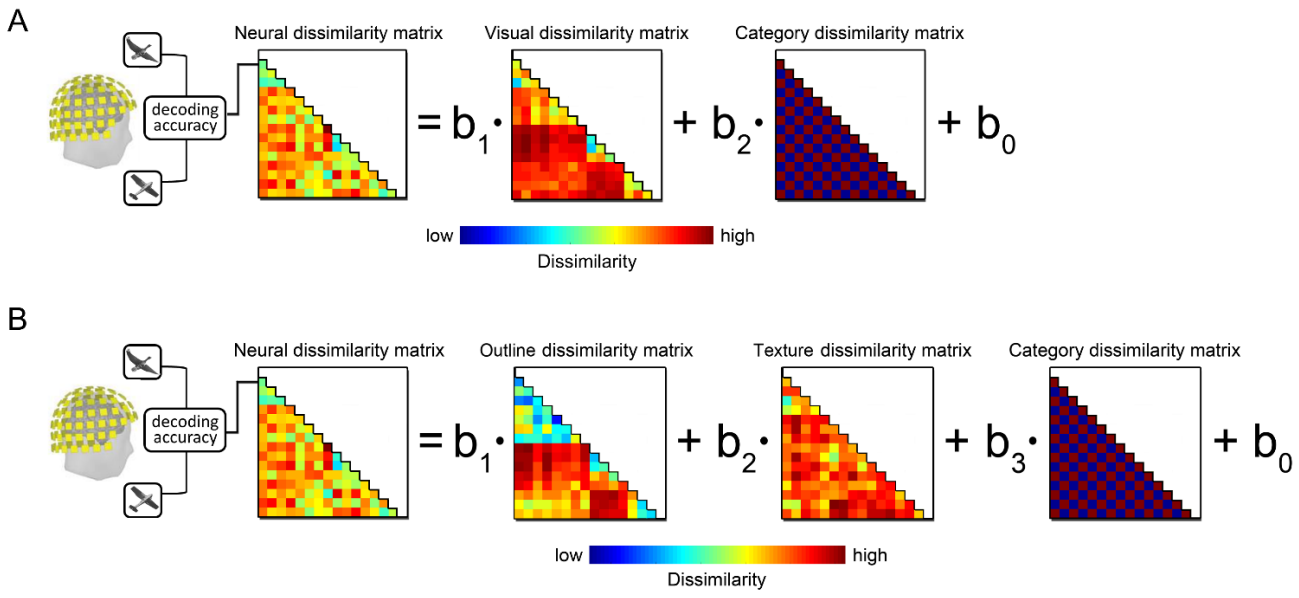



**Figure 4.3. Representational similarity analysis.** (A) At each time point, the neural dissimilarity

matrix was constructed by calculating the pairwise decoding accuracy for all pairs of stimuli. This neural dissimilarity matrix was then modeled as a linear combination of overall visual dissimilarity and category dissimilarity. (B) In the second RSA, the neural dissimilarity at each time point was modeled using outline, texture and category dissimilarity as predictors.

Representational similarity searchlight

To identify time-periods and sensors in which the contribution of category or shape dissimilarity to the MEG pattern dissimilarity was significant, we performed a sensor-space searchlight analysis. Using the same approach as in the previous analyses, we ran two representational similarity searchlights. In the first analysis, for each combination of time and channel the neural dissimilarity was modeled as a combination of visual and category dissimilarity (Figure 4.9), and in the second one neural dissimilarity was modeled using category, outline, and texture dissimilarities as predictors (Figure 4.10).

Prior to running the searchlights, the individual neural dissimilarity matrices were averaged for each 50 ms time window. This was done to improve signal-to-noise ratio and reduce the number of comparisons that would arise from comparing multiple time points and sensors. The individual subject neural RDMs were averaged for each 50 ms time window starting from 0 ms and until 600 ms after stimulus onset, resulting in twelve 50 ms time bins (see Figure 4.9).

These neural dissimilarity matrices were then modeled at each time window as (1) a linear combination of category dissimilarity and overall visual dissimilarity, resulting in two beta estimate maps (see Figure 4.9), and (2) a linear combination of category dissimilarity, outline dissimilarity, and texture dissimilarity, producing three beta estimate maps (see Figure 4.10). The individual subject maps for each predictor were then averaged across participants and tested against zero. Statistical analysis was performed for each time window, using the TFCE procedure described above to reveal the sensors and time windows in which the contribution of a particular predictor to the neural dissimilarity was significantly above chance. The significance was assessed using TFCE procedure described above.

## Results

Decodability of individual stimuli

To assess the quality of our data and evaluate whether stimulus identity could be decoded from the MEG sensor patterns, we averaged the off-diagonal elements of the MEG dissimilarity matrix for each time point. This was done separately for Experiments 1 and 2, as well as for each of the two types of sensors (magnetometers and gradiometers), producing four time courses of average pairwise stimulus decodability (Figure 4.4). Significance was assessed at each time bin, comparing decoding accuracy against chance (50%) and correcting for multiple comparisons using TFCE procedure. In both experiments and for both sensor types, decoding is at chance until about 50 ms, after which the decoding curve rises sharply, becoming significant at 60 ms and peaking at 120 ms. This pattern is consistent with the pattern observed in an earlier study (Cichy, Pantazis & Oliva, 2014). In Experiment 1, the main decoding accuracy peak at 120 ms is followed by a smaller peak at 180-190 ms (Figure 4.4A). In both experiments, the peak pairwise decoding is higher for gradiometers compared to magnetometers: in Experiment 1, peak decoding accuracy is 0.7 for magnetometers (Figure 4.4A, left panel) and 0.75 for gradiometers (Figure 4.4A, right panel). In Experiment 2, the peak decoding accuracy is 0.77 for magnetometers (Figure 4.4B, left panel) and 0.81 for gradiometers (Figure 4.4B, right panel). For both sensor types, average pairwise decoding accuracy is higher in Experiment 2, in which participants were performing the one-back task, suggesting that this task elicited stronger responses for individual stimuli, which led to higher pairwise decodability. Overall, these results replicate earlier MEG decoding studies and show that the individual stimuli can be decoded successfully from the MEG signals.

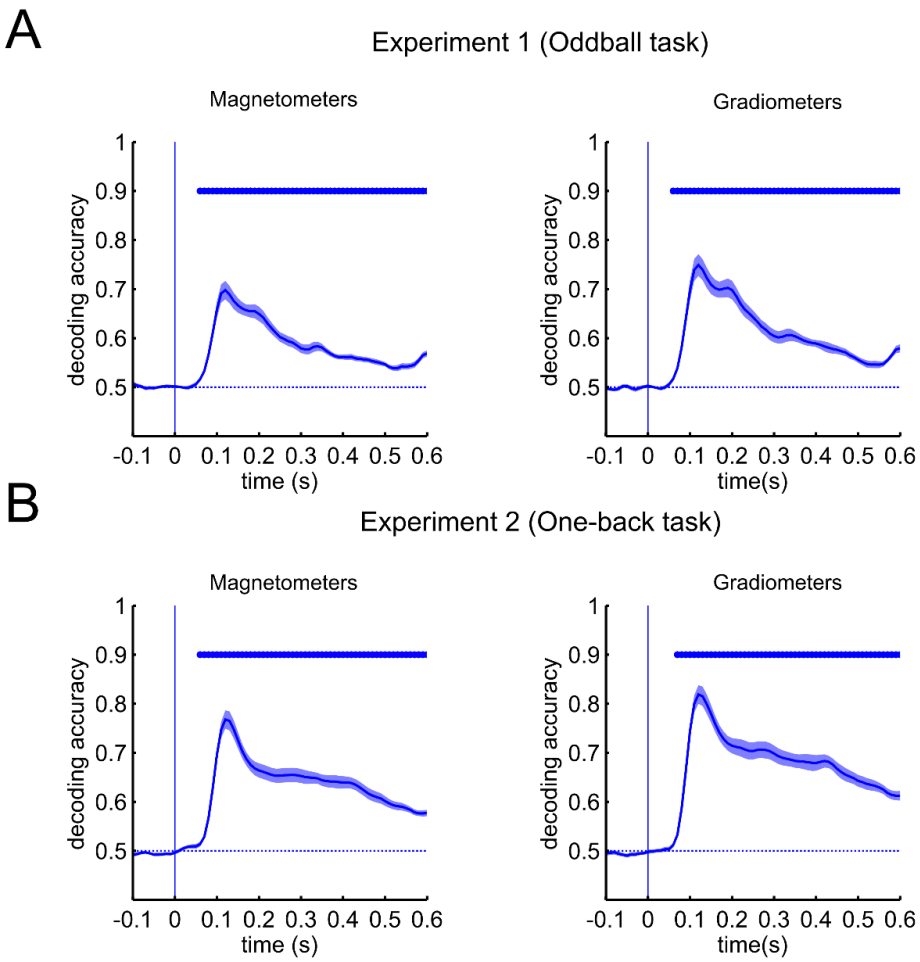



**Figure 4.4. Time courses of average decodability of all stimulus pairs.** (A) Average pairwise stimulus decodability in Experiment 1, shown separately for two sensor types. For both magnetometers and gradiometers, the average pairwise decoding accuracy reaches significance at 60 ms and peaks at 120 ms. The peak decoding accuracy is 0.7 for magnetometers (left panel) and 0.75 for gradiometers (right panel). (B) Individual stimulus decodability in Experiment 2. For both sensor types, average decoding accuracy peaks at 120 ms, with higher peak decoding accuracy for gradiometers (0.81, right panel) compared to magnetometers (0.77, left panel). Shaded areas reflect standard errors of the mean, and asterisks indicate time bins where decoding accuracy was significantly greater than chance (0.5).

Representational Similarity Analysis

To examine the contribution of overall visual dissimilarity and category dissimilarity to the MEG signals, we performed the representational similarity analysis (RSA) modelling the neural dissimilarity at each time point as the linear combination of these two predictors (Figure 4.3A). This analysis produced one beta estimate time course for each predictor. The results of this analysis for Experiment 1 are shown in Figure 4.5A, separately for two types of sensors. For both magnetometers and gradiometers, the beta estimate for overall visual dissimilarity (shown in red) reaches significance at 80 ms, peaking at 120 ms, with a smaller peak at 190 ms. Surprisingly, the beta estimate of category dissimilarity (shown in blue) does not reach significance at any time point. These results suggest that in contrast to robust information about visual stimulus attributes, category information cannot be retrieved from MEG signals when controlling for shape properties. In the next analysis, the neural dissimilarity was modeled as a combination of category, outline, and texture dissimilarity (Figure 4.3B). This analysis produces three time courses that are shown in Figure 4.5B. The outline predictor contributes significantly to MEG signal starting at 80 ms, peaking at 130 ms, followed by a smaller peak at 190 ms. The time course of texture predictor is significant starting from 90 ms, with a peak at 120 ms.

The same analyses were performed for Experiment 2 (Figure 4.6). The first RSA revealed the overall visual dissimilarity peaking at 130 ms for both magnetometers and gradiometers (Figure 4.6A). Again, the category predictor did not reach significance at any time point, showing that even the use of one-back task did not reveal category representations that were independent of object shape. In the second RSA, the outline dissimilarity was significantly contributing to MEG signal with a peak at 140 ms for magnetometers (Figure 4.6B, right panel) and 150 ms for gradiometers (Figure 4.6B, left panel). Texture (shown in yellow) peaked at 110 ms for both types of sensors.

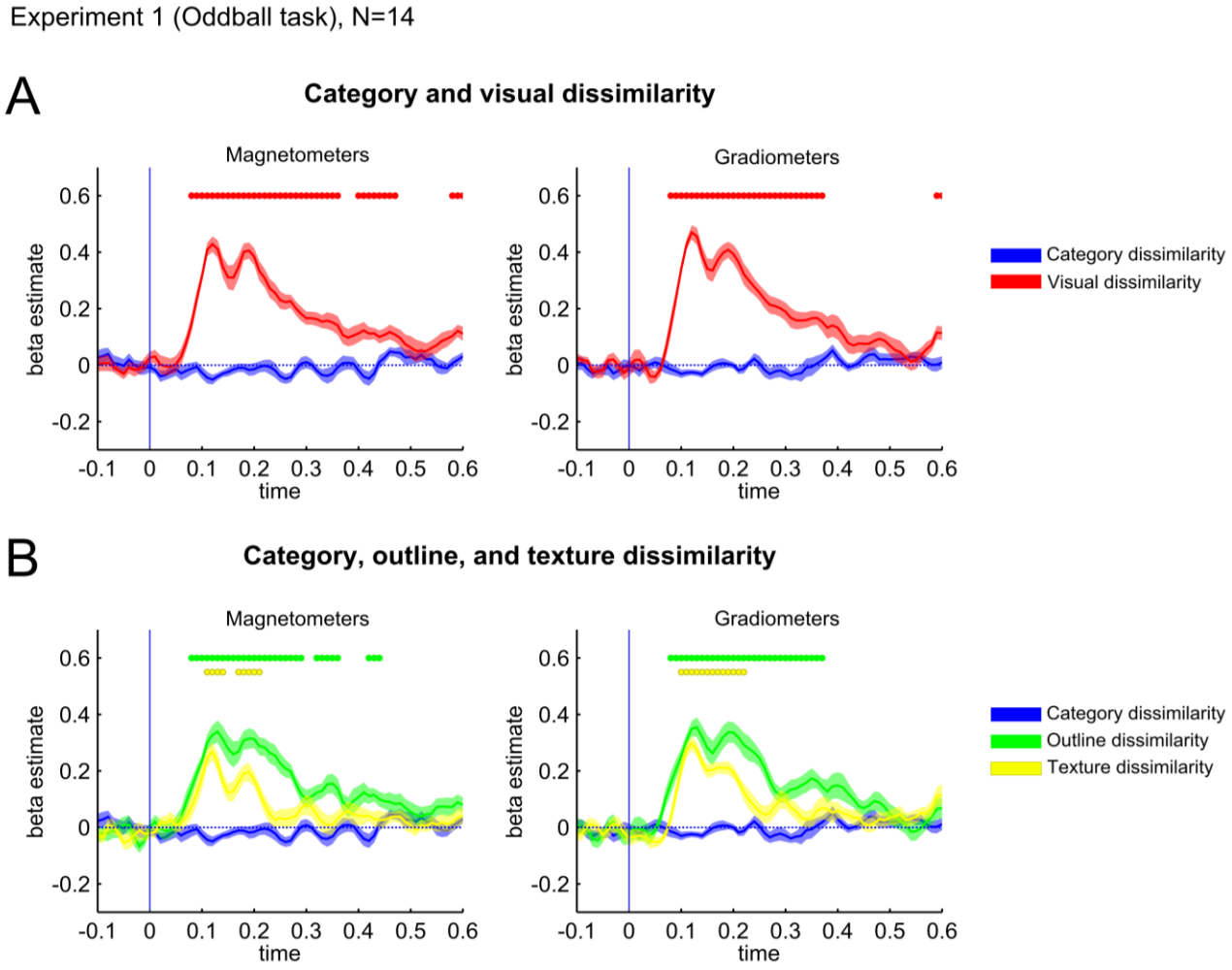



**Figure 4.5**. **Representational similarity analysis results for Experiment 1.** (A) The time courses of regression weights reflecting the contributions of category dissimilarity (in blue) and overall visual dissimilarity (in red) to the neural dissimilarity. For both magnetometers and gradiometers, the beta estimate for overall visual dissimilarity reaches significance at 80 ms, peaking at 120 ms, with a smaller peak at 190 ms. However, regression weights for category dissimilarity do not reach significance at any time point. (B) The time courses of regression weights reflecting the contributions of category dissimilarity (in blue), outline dissimilarity (in green) and texture dissimilarity (in yellow). The outline predictor contributes significantly to MEG signal starting at 80 ms, peaking at 130 ms, followed by the second peak at 190 ms. The time course of texture predictor is significant starting from 90 ms, with a peak at 120 ms. Again, category predictor contribution is not significant. Shaded areas reflect standard errors, and asterisks indicate time bins where beta estimates were significantly greater than zero.

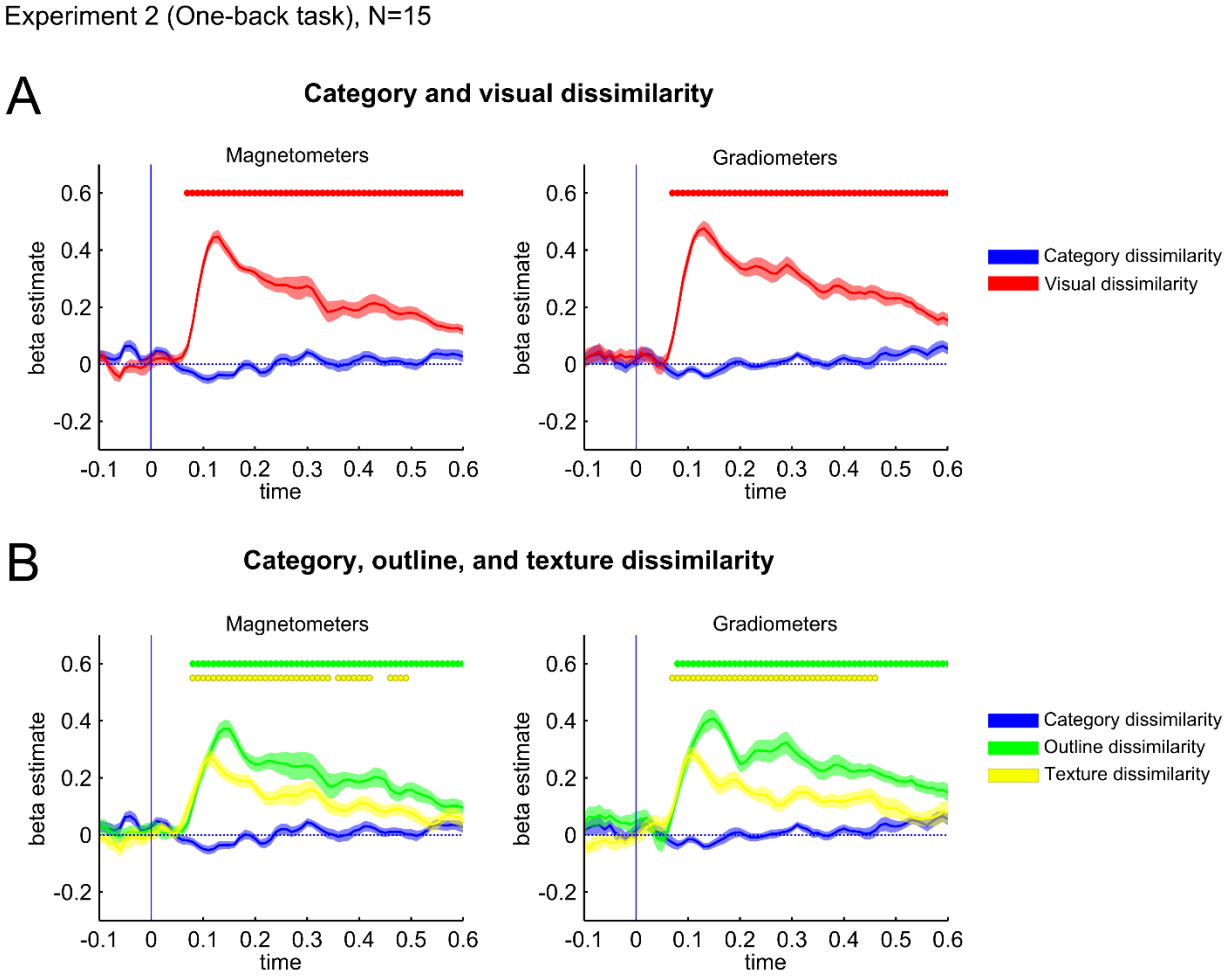



**Figure 4.6**. **RSA results for Experiment 2.** (A) The time courses of regression weights showing the contributions of category dissimilarity (in blue) and overall visual dissimilarity (in red) to the neural dissimilarity. Visual dissimilarity regression weights are significant starting at 80 ms, peaking at 130 ms for both magnetometers and gradiometers, while category dissimilarity is not contributing significantly to neural dissimilarity. (B) The time courses of regression weights reflecting the contributions of category dissimilarity (in blue), outline dissimilarity (in green) and texture dissimilarity (in yellow). The outline dissimilarity contributes significantly to MEG signal starting at 90 ms, with a peak at 140 ms for magnetometers (right panel) and at 150 ms for gradiometers (left panel). Texture predictor regression weights peaked at 110 ms for both types of sensors. Again, no evidence for significant contribution of category predictor is present. Shaded areas reflect standard errors, and asterisks indicate time bins where beta estimates were significantly greater than zero.

Given that the predictor time courses were similar across the two experiments, we combined the data from two experiments (resulting in total of 29 subjects) to increase statistical power and ran the RSA again (Figure 4.7). The overall visual dissimilarity beta estimate peaked at 120 ms for both types of sensors (Figure 4.7A). Interestingly for the first time we observed category dissimilarity predictor reaching significance at two late time points at 520-530 ms (Figure 4.7A, left panel). In the second RSA, the regression weights for outline and texture dissimilarity peaked at 130 ms and 120 ms, respectively, for both types of sensors (Figure 4.7B). Again, the category predictor was significant for the same two time points at 520-530 ms (Figure 4.7B, left panel). Notably, this late category effect was observed at time points when the stimulus was no longer on the screen (after 500 ms), suggesting that it might not be reflecting perceptual processing but, possibly, some other process such as working memory (e.g., holding the stimulus in memory to perform a one-back task).

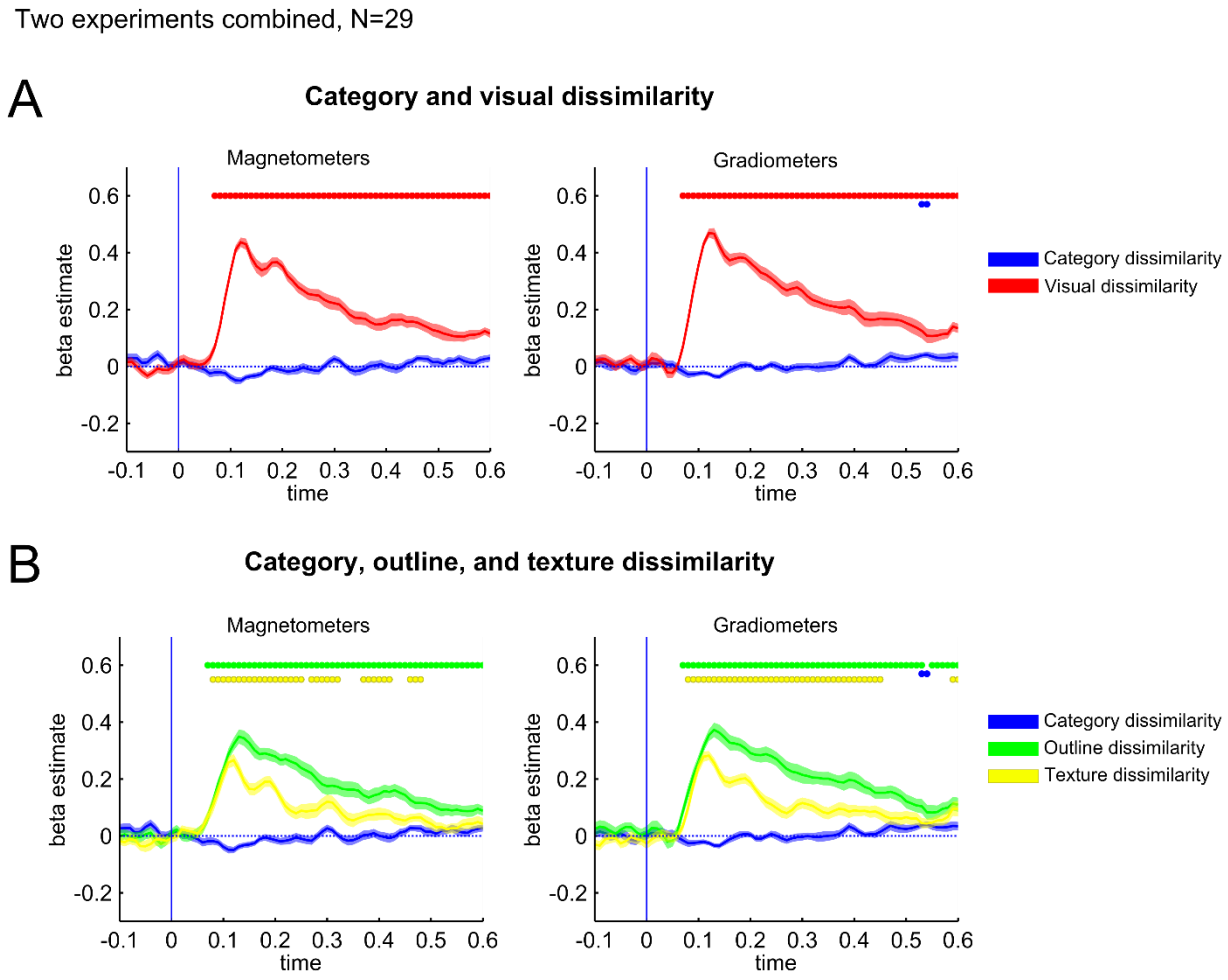



**Figure 4.7. RSA results after combining the data from two experiments.** (A) Contributions of category dissimilarity (in blue) and overall visual dissimilarity (in red) to MEG signal. The overall

visual dissimilarity beta estimate peaked at 120 ms for both types of sensors, and category beta estimate briefly reached significance at two late time points (520-530 ms, right panel). (B) Contributions of category dissimilarity (in blue), outline dissimilarity (in green) and texture dissimilarity (in yellow) to MEG signal. The regression weights for outline and texture dissimilarity peaked at 130 ms and 120 ms, respectively, for both types of sensors. Category regression weight reached significance at the same late time points (520-530 ms, right panel). Shaded areas reflect standard errors, and asterisks indicate time bins where beta estimates where significantly greater than zero.

Euclidean distance as a measure of neural dissimilarity

It is possible that the measure of neural dissimilarity used in the analyses described above (i.e., pairwise decoding accuracy) was not sensitive enough to reveal potential category effects. It is therefore useful to compare it to other commonly used dissimilarity measures, such as the Euclidean distance. We repeated the last analysis (with the data from two experiments combined) using the neural dissimilarity matrices constructed using this new measure: i.e., with pairwise neural dissimilarity between the stimuli defined as the Euclidean distance between the two corresponding activity patterns (se Matherials and Methods).

The comparison between the RSA results using two different neural dissimilarity measures (decoding accuracy and Euclidean distance) are shown in Figure 4.8. Despite the two measures being different, the pattern of results using Euclidean neural dissimilarity predictors (Figure 4.8B) was very similar to the one obtained using decoding accuracy (Figure 4.8A), confirming our previous results.

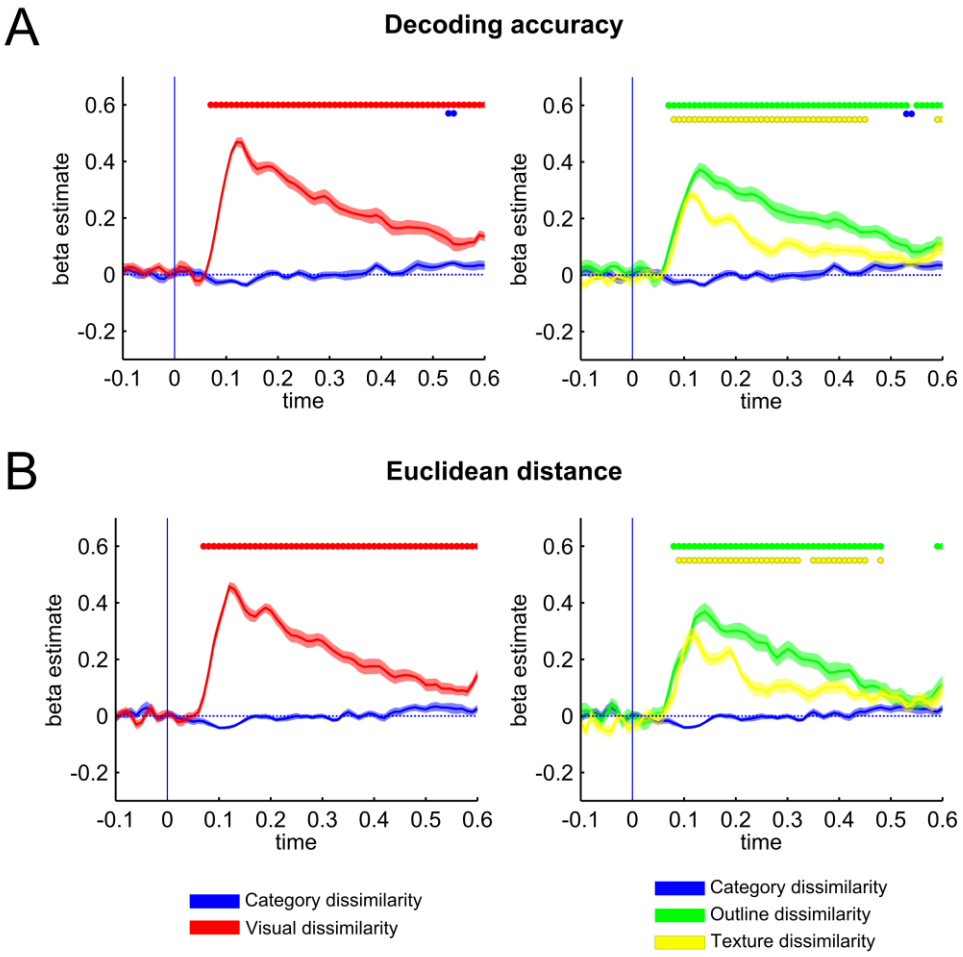



**Figure 4.8. RSA results using two different neural dissimilarity measures.** The analyses were done after combining the data from the two experiments (N=29) and using the data from gradiometers. (A) Pairwise decoding accuracy as a neural dissimilarity measure. These graphs are the same as in Figure 4.7, right panel. (B) Same analyses using Euclidean distance as a neural dissimilarity measure. The overall pattern of results is highly similar between the two measures.

### Representational Similarity Searchlight

To approximately assess the spatial distribution of category- and shape-related signals in the brain, and to determine if the contribution of category dissimilarity was significant in any of the individual sensors, we performed a whole-brain sensor-space representational similarity searchlight. This analysis was done using the data from gradiometers only because only these sensors showed significance for category dissimilarity beta estimates in the previous analyses (see Figure 4.7, right panel). Given that the pattern of results in Experiments 1 and 2 was

highly similar and category significance was only observed after combining the data from both experiments, we pooled data from Experiments 1 and 2 together for the searchlight analysis to maximize power and to examine this late category effect more closely.

For each channel we took the ten neighboring channels and computed the neural dissimilarity matrix for this neighborhood (see Methods). We then used the RSA approach described earlier, modelling the neural dissimilarity for each channel as a combination of category and visual dissimilarity. This procedure was done for each time window, resulting in 12 searchlight maps showing the contribution of visual and category dissimilarity to neural dissimilarity at different spatial locations. Strikingly, this analysis did not reveal any sensors or time windows in which the contribution of category dissimilarity to MEG dissimilarity (as measured by regression weights) was significantly above zero (Figure 4.9A). In contrast, the contribution of visual dissimilarity reached significance in 50-100 ms time window and stayed significant throughout all time windows, starting in a group of posterior sensors and progressing to more anterior sensors at 100-250 ms, peaking at 150-200 ms and gradually receding towards the posterior sensors again (Figure 4.9B). These results suggest that the MEG neural dissimilarity across multiple channel locations predominantly reflects visual, but not category dissimilarity.

We then ran a second RSA searchlight, using category, outline and texture dissimilarity as predictors of neural dissimilarity. The results of this analysis are shown in Figure 4.10. As in the previous analysis, no sensors in any time window exhibited significant contribution of category dissimilarity to MEG signals (Figure 4.10A). On the other hand, the outline dissimilarity beta estimates reached significance at 50-100 ms after stimulus onset in a group of left posterior sensors, spreading to more anterior sensors at 100-150 ms and peaking at 150-200 ms, staying significant in posterior channels throughout all time windows (Figure 4.10B). Similarly, texture dissimilarity beta estimates reached significance in a cluster of central posterior sensors at 50-100 ms and stayed significant at multiple channel locations in all the next time windows, peaking at 150-200 ms (Figure 4.10C). This is further evidence that MEG sensor patterns are strongly reflecting visual features (in this case, outline and texture dissimilarity) and not category dissimilarity.

To summarize, the two searchlight analyses revealed robust contributions of overall visual dissimilarity as well as outline and texture dissimilarity to the MEG sensor patterns. However,

we did not find evidence for significant contribution of category dissimilarity to neural dissimilarity at any time window or sensor, suggesting that MEG sensor patterns primarily reflect object shape information.

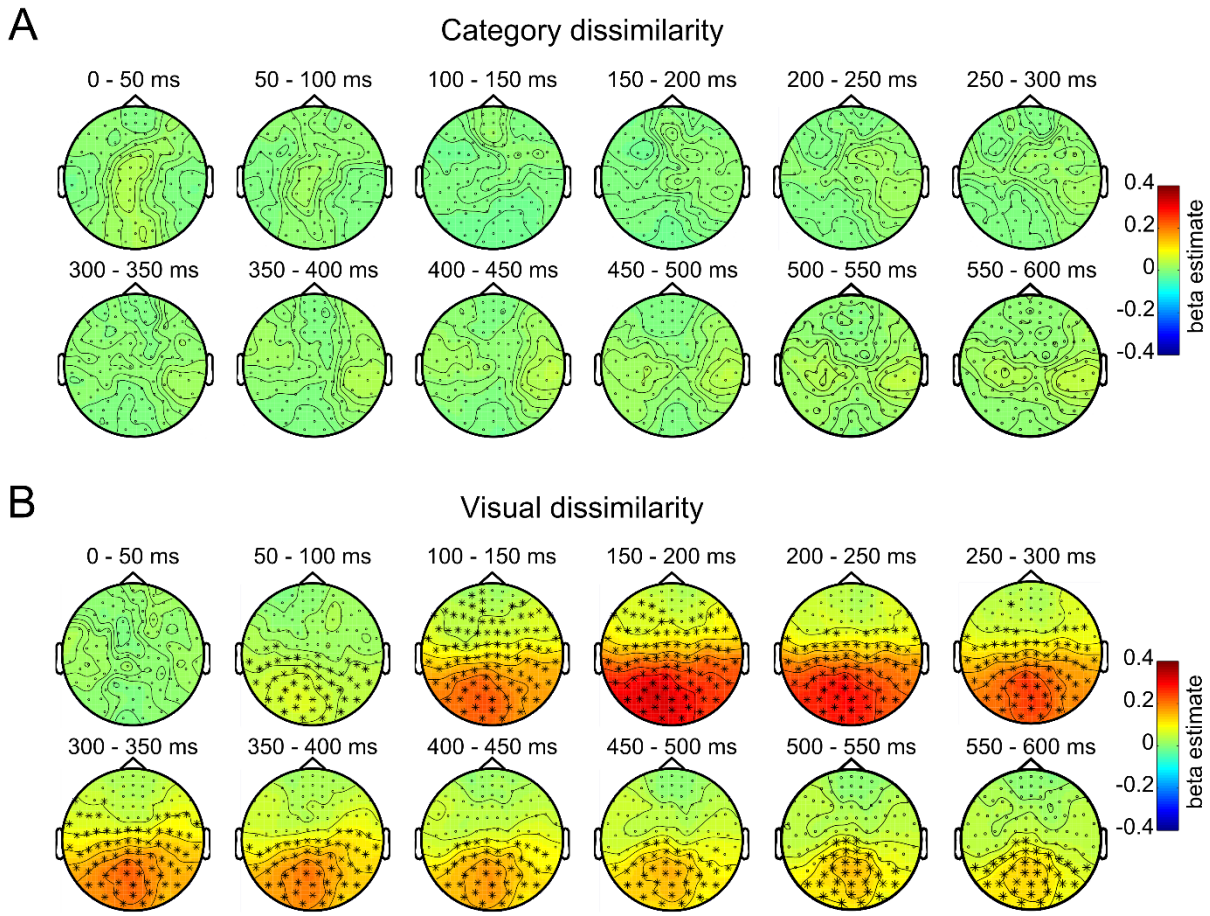



**Figure 4.9**. **RSA searchlight results: category and visual dissimilarity.** Asterisks indicate sensor locations where beta estimates were significantly greater than zero. (A) Searchlight maps showing the regression weights reflecting the contribution of category dissimilarity to MEG patterns across channels (gradiometers) and twelve time windows. The beta estimates for category were not significantly above zero in any of the sensors for all time windows. (B) Searchlight maps showing the visual dissimilarity regression weights across gradiometers for twelve time windows. The beta estimates are significantly above zero for a group of posterior sensors starting from 50-100 ms time window and for all the time windows after that, with activity spreading to more anterior sensors from 100 to 200 ms, and then fading back to posterior sensors. The contribution of visual dissimilarity to MEG signal is peaking in the 150-200 ms time window.

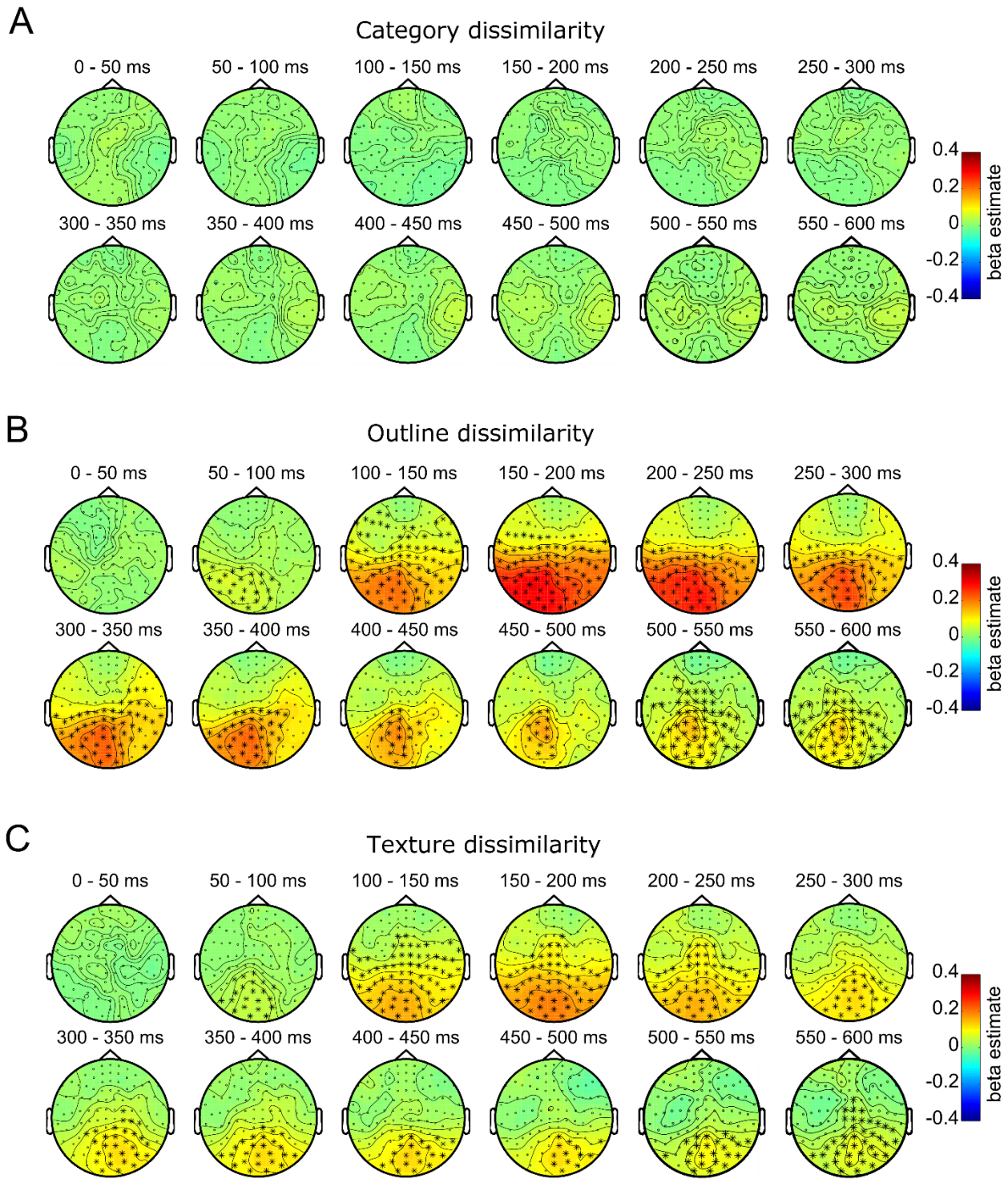



**Figure 4.10**. **RSA searchlight results: category, outline and texture dissimilarity.** Asterisks indicate sensor locations where beta estimates were significantly greater than zero. (A) Beta estimate maps showing the contribution of category dissimilarity to MEG patterns across channels (gradiometers) and twelve 50-ms time windows. The beta estimates for category did not reach significance in any of the sensors or time windows. (B) Beta estimate maps showing the

contribution of outline dissimilarity to MEG signal across gradiometers for twelve time windows. The beta estimates are significantly above zero for a group of posterior sensors starting from 50-100 ms time window and for all the time windows after that, with activity spreading to more anterior sensors from 100 to 200 ms, and then fading back to posterior sensors. The contribution of visual dissimilarity to MEG signal is peaking in the 150-200 ms time window. (C) Beta estimates reflecting contribution of texture dissimilarity to neural dissimilarity. As for the outline maps, the texture regression we reach significance in a group of posterior sensors at 50-100 ms, spreading throughout more anterior sensors from 100 to 250 ms with a peak at 150-200 ms and gradually receding while staying significant in posterior sensors in all the later time windows.

## Discussion

In this study we aimed at disentangling the contribution of shape and category information to the MEG signal. In two MEG experiments, we presented participants with stimuli from two categories (animate and inanimate objects). Critically, the objects from two categories were closely matched for overall shape features, making it possible to disentangle category and shape information. Importantly, in a previous fMRI study these stimuli were shown to elicit category-specific responses in the ventral temporal cortex even after controlling for visual similarity (Chapter 2; Proklova, Kaiser, & Peelen, 2016).

By constructing a decoding accuracy matrix for each time point and averaging all elements, we found that individual stimuli were decodable from the MEG signal with above chance accuracy starting at 60 ms, with the peak at 120 ms (Figure 4.4). This result corroborated earlier MEG studies (e.g., Cichy, Pantazis, & Oliva, 2014) and provided a quality check of the data, confirming that the identity of the stimuli we used could be successfully decoded from MEG signals.

We then ran two representational similarity analyses to examine the independent contributions of category dissimilarity and different aspects of visual dissimilarity to the MEG signal. In the first analysis, neural dissimilarity at each time point was modeled as a linear combination of category and overall visual dissimilarity. In the second analysis, we used category, outline, and texture dissimilarity as predictors of MEG dissimilarity. All the shape-related predictors contributed strongly to MEG dissimilarity, reaching significance at 80 ms

and peaking between 110-130 ms. Strikingly, in both experiments the regression weights reflecting the contribution of category dissimilarity did not reach significance at any time point. Only after pooling the data from the two experiments, we observed a small category effect at 520-530 ms (see Figure 4.7, right panel). However, the category beta estimate at these time points was only slightly above chance and did not show significance in a follow-up whole-brain searchlight analysis (see Figure 4.9A), possibly due to increased number of comparisons. In case this significant finding was reflecting a true effect, the category information became decodable from MEG signals after 500 ms, much later than previously reported (Cichy, Pantazis, & Oliva, 2014; Grootswagers et al., 2016). Importantly, after 500 ms the stimulus was no longer on the screen, but participants still needed to hold it in memory to successfully perform the one-back task, so this signal could be reflecting memory effects and not the perceptual effect. However, we did not observe this effect while looking at the data from Experiment 2 only (Figure 4.6).

Importantly, we used a behavioral measure of visual similarity that was designed to quantify perceptual similarity of the stimuli while not being influenced by the stimulus category. We measured the perceptual similarity of the intact stimuli as well as the dissimilarity of their outline and texture. We found that all of these shape predictors were strongly contributing to MEG responses, showing that MEG is very sensitive to multiple visual object properties.

Several other studies have addressed the issue of separating category and shape information, finding evidence for category information reflected in the MEG or EEG signal after the initial shape-related information before 200 ms. However, these studies used different ways to control for visual dissimilarity. For example, a recent study compared EEG response patterns for images of intact and scrambled objects (Coggan, Baker, and Andrews, 2016). The results showed that both intact and scrambled images elicit similar category-specific patterns of response at the early stages of processing (before 200 ms), but the response to intact images was sustained for a longer time, taken as evidence that this later effect is reflecting semantic category information. However, the use of the scrambled images is not the optimal control for visual object features. It is possible that some residual visual features that were removed during scrambling were driving the response to the intact stimuli. In our study, we specifically measured the visual similarity of intact stimuli, and, after regressing out this visual

information, we did not find any evidence of category information in the MEG sensory patterns.

Another recent study by Clarke et al. (2015) compared how well a visual model and a model that combines visual and semantic information predict the MEG data and found that the model that includes semantic information describes the MEG data better after 200 ms. It is possible that the model that includes semantic features and not just binary category information is a better predictor for the MEG data. Moreover, in this study participants were performing object naming task that could potentially elicit stronger category responses compared to oddball detection and one-back tasks. Changing the task and using a semantic dissimilarity predictor would be useful next steps for our experiment.

An earlier study from our lab found that category information in the MEG signals can be disentangled from shape information by using pictures of hands and torsos and shape-matched images of gloves and t-shirts (Kaiser, Azzalini, & Peelen, 2016). Since images of body parts were used in this study, the extrastriate body area (EBA) was likely involved in their perception. It is possible that the activity in EBA, which is located in the lateral occipital-temporal cortex, is picked up better by MEG compared to more inferior-temporal areas. Indeed, the stimuli we used in a present study have been shown to elicit category-selective responses in ventral temporal cortex, around the fusiform gyrus, which might be more difficult to access with MEG.

Our results, therefore, should not be taken to mean that it is impossible to decode object category information from the MEG response patterns. However, these results demonstrate that different aspects of object shape, including overall shape, outline, and texture, contribute greatly to the MEG response patterns. This is in line with previous studies that showed that perceived shape similarity is a strong predictor of neural dissimilarity measured with MEG (Wardle, Kriegeskorte, Grootswagers, Khaligh-Razavi, & Carlson, 2016). Our results illustrate the importance of properly controlling for object shape features while measuring object category information with MEG.

# Chapter 5

# General Discussion

The main goal of this thesis is to disentangle the effects of object shape and object category on the representations of animate and inanimate objects in the brain. To this end, we created a stimulus set in which shape and category information were dissociated, and conducted fMRI and MEG studies to examine independent contributions of shape and category to brain response patterns. In this chapter, we summarize our findings, describe them in context of existing literature and outline future directions.

In Chapter 2, we introduced a novel stimulus set in which stimuli from two categories (animate and inanimate objects) were carefully matched for overall shape features (e.g. snake – rope), allowing to disentangle the contribution of shape and category to brain activity patterns. This approach has not been used before to address this issue, apart from a few very recent studies (Bracci & Op de Beeck, 2016; Kaiser, Azzalini, & Peelen, 2016). Earlier studies mainly controlled for visual similarity post hoc, by using a computational model to rule out the possibility of visual similarity explaining the observed category-specific responses (e.g., Kriegeskorte et al., 2008). Our approach allows to control for visual dissimilarity more rigorously by purposefully disentangling shape and category information during stimulus selection.

We also described a behavioral paradigm aimed at measuring visual dissimilarity of the stimuli based on participants' performance in a simple visual search task. We conducted three behavioral visual search experiments to measure different aspects of visual dissimilarity of the stimuli. This task was previously used to relate visual dissimilarity to categorization performance (Mohan & Arun, 2012). We chose this task because it allows to measure the perceptual dissimilarity of objects while minimizing the top-down effects. Importantly, using the RTs in these visual search experiments we constructed three visual dissimilarity matrices that were then used as predictors of neural dissimilarity in fMRI and MEG studies.

In Chapter 3, we described an fMRI experiment designed to test the hypothesis that visual dissimilarity of the stimuli can fully explain the category-related patterns of activity in human ventral temporal cortex. Our results showed that this was not the case. First, we found that

lateral-to-medial organization of animate- and inanimate-preferring responses (e.g., Mahon et al., 2007) was preserved even when stimuli from these categories were matched for shape. Second, the representational similarity analysis revealed that category information was still present in VTC activity patterns after accounting for visual similarity of the stimuli. This is in line with earlier studies showing that object animacy is an important factor explaining VTC response patterns (e.g., Chao et al., 1999; Kriegeskorte et al., 2008), and against the interpretation that these patterns reflect visual properties of objects belonging to different categories (e.g., Rice et al., 2014). Importantly, in our study we used a challenging stimulus set, specifically avoiding stimuli such as human faces or bodies that strongly influenced representational structure of VTC in earlier studies (Kriegeskorte et al., 2008), and still observed distinct category-selective patterns of response in VTC.

Our findings are also in line with the results of a recent study that used a very similar approach to match objects from different categories for shape (Bracci & Op de Beeck, 2016). In this study, the researchers found that both shape and category information was reflected in patterns of activity in VTC, as well as in lateral occipito-temporal cortex. Moreover, patterns of activity in VTC were reflecting the animate-inanimate distinction. We observed similar results in our fMRI study, using a slightly different analysis approach (the multiple regression RSA), and after controlling for additional visual features (outline and texture).

There is still a possibility that in our stimulus selection and visual dissimilarity measure we overlooked some residual visual features that could explain category-selective responses in VTC. Indeed, no existing method allows to control for all of the visual features. The shape-matching approach also has its limits, because the perfectly-matched stimuli would be indistinguishable. It would therefore be important to replicate our findings using different approaches to controlling for shape. For example, one can transform the stimuli in a way that loosely preserves mid-level features (such as texture) as well as coarse object form, but renders the stimuli unrecognizable (Freeman & Simoncelli, 2011; Long, Konkle, Cohen, & Alvarez, 2016) and then measure the responses to such images (“texforms”) in the ventral stream. If category-selective responses are still observed for these stimuli, it would suggest that the animate-inanimate distinction might be driven by mid-level features.

Another interesting next step would be to test deep neural networks (DNNs) on our stimulus set and then compare the DNN dissimilarity to perceptual dissimilarity. This would test

whether DNNs and humans rely on similar object features (Kubilius, Bracci & Op de Beeck, 2016). The DNN dissimilarity could then be used as a predictor of neural dissimilarity measured with fMRI and MEG.

Chapter 4 of this thesis is devoted to two MEG experiments in which we tested whether it was possible to disentangle object shape and object category information reflected in MEG sensor patterns, with the eventual goal of comparing the time courses of these two predictors. We found that information about object shape, outline, and texture was reflected robustly in the MEG signal. However, after regressing out visual dissimilarity, we found no residual information about object category in MEG signals.

The fact that different aspects of visual dissimilarity predicted MEG response patterns so well is consistent with an earlier study that showed that perceptual similarity is reflected robustly in the MEG signal (Wardle, Kriegeskorte, Grootswagers, Khaligh-Razavi, & Carlson, 2016). A novel contribution of our study is that we showed that texture and outline similarity (and not just the overall visual similarity) also predict MEG response patterns. Moreover, we were able to examine the independent time courses of these predictors.

At first glance, it may seem like our fMRI and MEG findings are in conflict with each other: despite using the same stimuli, experimental design, and similar analysis approach, we observed category-specific response patterns that were independent of shape with fMRI, but not with MEG. However, it is important to note that the results of the two studies do not necessarily contradict each other, and can be reconciled in several ways. It could be that RSA approach used in fMRI study was not ideally suited for MEG data analysis; for example, because of the lower spatial resolution of MEG. Although we chose RSA to make fMRI and MEG experiments more comparable, other analysis techniques can be applied to MEG data in order to examine it more closely. Further analyses could include cross-decoding approach similar to the one used by Kaiser, Azzalini, & Peelen (2016), an MEG decoding study that observed shape-independent category responses for body parts. Finally, another possibility is that the category-preferring clusters found in the fMRI study were located in areas that were not well accessible to MEG, namely, in ventral temporal cortex. This could also explain differences with the study by Kaiser et al. (2016), in which stimuli of body parts were used, potentially activating more lateral portions of the temporal cortex (e.g., EBA) that could be more accessible with MEG.

Although shape information was robust in both fMRI and MEG experiments, the shape-independent category information was only revealed in the fMRI experiment. One could argue that even in fMRI experiment shape was the dominant factor in VTC representations, with category-selective clusters less prominent compared to shape-selective clusters (Figure 3.5). However, it is important to note that in the fMRI experiment we did not aim to compare the relative contributions of shape and category to VTC activity patterns. Instead, the aim of the fMRI experiment was to test whether category-selective VTC responses could be explained fully by the shape features that are characteristic for animals and inanimate objects. The existence of shape-independent category clusters in VTC suggests that object category (animacy) is an important organizing principle for VTC representations, even if these clusters are smaller in size compared to shape clusters.

However, it is still possible that these results could be influenced by the way we defined animacy and shape similarity of the stimuli. For example, it could be that in our stimulus set object animacy was defined too narrowly (i.e., a stimulus could be either animate or inanimate, with no gradations of animacy), while shape was defined more broadly using a continuous measure. Could this explain the robust shape responses in both experiments and the absence of category information in the MEG signal? It is possible, and could partially be addressed by using a category predictor that takes into account the "animacy continuum" (Connolly et al., 2012; Sha et al., 2014), with the "degree of animacy" defined as the similarity to the animate prototype (i.e., human). However, in the studies described in this thesis we were specifically interested in the binary definition of object animacy, because this is how animacy was defined in the body of literature on the animate-inanimate distinction in the ventral stream (e.g., Ishai et al., 1999; Kriegeskorte et al., 2008). In other words, our studies tested the "strong" version of the animacy hypothesis, i.e. whether animate-inanimate distinction in the brain responses (with animacy defined in a binary way) is reflecting a true categorical distinction or whether it simply reflects shape properties of these categories. This is also why we specifically avoided "highly animate" stimuli according to the animacy continuum, such as humans and four-legged mammals. Importantly, even using this strict definition of animacy, we were able to find shape-independent category-selective responses in the VTC in the fMRI experiment.

One could also argue that the visual similarity of the stimuli as defined in these experiments (via the visual search tasks) was dominated by low-level shape properties such as silhouette overlap, rather than more high-level shape properties such as perceived shape similarity as rated by human participants (Bracci & Op de Beeck, 2016). It would therefore be informative to collect behavioral ratings of the shape similarity of our stimuli and compare our visual dissimilarity measure with the one obtained from these ratings. However, we speculate that these two measures might produce similar results, given that in visual search task participants had to rely not only on overall silhouette similarity (which was high for the stimuli belonging to the same shape set), but also on the inner texture and other distinctive features of the stimuli. This is also reflected in the fact that overall visual dissimilarity could be closely approximated as the linear combination of outline and texture dissimilarity, and each of these features contributed separately to the fMRI and MEG signal.

An interesting new analysis approach (similarity fusion; Cichy, Pantazis, & Oliva, 2016) can be used to relate the data from the fMRI and MEG experiments described in this thesis. By taking an fMRI dissimilarity matrix from a particular ROI (for example, category-selective areas or early visual cortex) and correlating it with an MEG dissimilarity matrix at every time point, one can obtain a time course of when the representational geometry of this ROI is reflected in MEG sensor patterns. It would be particularly interesting to see if the representational structure of category-selective clusters from out fMRI study is reflected in MEG patterns at some time point.

Given that we were able to find evidence for category selectivity with fMRI, we take our MEG results to mean that MEG is less sensitive to this information compared with fMRI. However, we do not claim that shape-independent category information cannot in principle be found in MEG signals (see Kaiser, Azzalini, & Peelen, 2016). The main finding of our MEG study is that MEG is very sensitive to various visual object properties, and it is therefore important to properly control for visual dissimilarity of stimuli in studies that report the decodability of object category from MEG sensor patterns (e.g., Cichy, Pantazis, & Oliva, 2014).

## Conclusions

To summarize, the studies described in this thesis aimed at disentangling the effects of object shape and object category on the representations of animate and inanimate objects in the brain. To do this, we created a stimulus set in which shape and category information were dissociated, and conducted fMRI and MEG studies to examine independent contributions of shape and category to brain response patterns. Using fMRI, we found that the animate-inanimate distinction in the ventral temporal cortex (VTC) was observed even after controlling for visual similarity of the stimuli. These results suggest that the category-selective responses in VTC do not simply reflect the overall shape features characteristic of these categories. We then conducted an MEG study to examine the time course of shape and category information in the brain. We found that different aspects of visual dissimilarity are contributing robustly to MEG sensor patterns. Surprisingly, when regressing out the contribution of visual properties, no residual category information was present in MEG response patterns. We conclude that MEG sensor patterns evoked by visually presented objects predominantly reflect visual object properties. Together with the results of the fMRI study, these findings suggest that MEG is less sensitive to object category information that is independent of shape information.

## References

Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. Nature Neuroscience, 10, 687–689.

Amedi, A., von Kriegstein, K., van Atteveldt, N. M., Beauchamp, M. S., & Naumer, M. J. (2005). Functional imaging of human crossmodal identification and object recognition. Experimental Brain Research, 166, 559–571.

Andrews, T. J., & Schluppeck, D. (2004). Neural responses to Mooney images reveal a modular representation of faces in human visual cortex. Neuroimage, 21, 91–98.

Andrews, T. J., Watson, D. M., Rice, G. E., & Hartley, T. (2015). Low-level properties of natural images predict topographic patterns of neural response in the ventral visual pathway. Journal of Vision, 15, 3.

Baldassi, C., Alemi-Neissi, A., Pagan, M., DiCarlo, J. J., Zecchina, R., & Zoccolan, D. (2013). Shape similarity, better than semantic membership, accounts for the structure of visual object representations in a population of monkey inferotemporal neurons. PLoS Computational Biology, 9, e1003167.

Bell, A. H., Hadj-Bouziane, F., Frihauf, J. B., Tootell, R. B., Ungerleider, L. G. (2009). Object representations in the temporal cortex of monkeys and humans as revealed by functional magnetic resonance imaging. Journal of Neurophysiology, 101, 688–700.

Belongie, S., Malik, J., & Puzicha, J. (2002). Shape matching and object recognition using shape contexts. IEEE Trans. PAMI, 24 (4), 509–522.

Borg, I., & Groenen, P. J. F. (2005). Modern Multidimensional Scaling – Theory and Applications. New York: Springer.

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, 18247–18258.

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

Brainard, D. H. (1997). The psychophysics toolbox. Spatial Vision, 10, 433–436.

Buchel, C., Price, C., & Friston, K. (1998). A multimodal language region in the ventral visual pathway. Nature, 394, 274–277.

Caldara, R., Seghier, M. L., Rossion, B., Lazeyras, F., Michel, C., & Hauert, C. A. (2006). The fusiform face area is tuned for curvilinear patterns with more high-contrasted elements in the upper part. Neuroimage, 31, 313–319.

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain the

animate-inanimate distinction. Journal of Cognitive Neuroscience, 10, 1–34.

Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. Journal of Vision, 13, 1.

Castelli, F., Happe, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. Neuroimage, 12, 314–325.

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, 913–919.

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. Nature Neuroscience, 17, 455–462.

Cichy, R. M., Pantazis, D., & Oliva, A. (2016). Similarity-based fusion of MEG and fMRI reveals spatio-temporal dynamics in human cortex during visual object recognition. Cerebral Cortex, 26, 3563–3579.

Clarke, A., Devereux, B. J., Randall, B., & Tyler, L. K. (2015). Predicting the Time Course of Individual Objects with MEG. Cerebral Cortex, 25 (10), 3602-3612.

Coggan, D. D., Baker, D. H., & Andrews, T. J. (2016). The Role of Visual and Semantic Properties in the Emergence of Category-Specific Patterns of Neural Response in the Human Brain. eNeuro 3:ENEURO.0158-16.2016.

Cohen, L., & Dehaene, S. (2004). Specialization within the ventral stream: The case for the visual word form area. Neuroimage, 22, 466–476.

Cohen, M.A., Konkle, T., Rhee, J., Nakayama, K., and Alvarez, G.A. (2014) Processing multiple visual objects is limited by overlap in neural channels. Proceedings of the National Academy of Sciences USA. 111, 24, 8955-8960.

Cohen, M.A., Nakayama, K., Konkle, T., Stantic, M., and Alvarez, G.A. (2015) Visual awareness is limited by the representational architecture of the visual system. Journal of Cognitive Neuroscience, 27, 2240-2252.

Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y.-C., Abdi, H., & Haxby, J. V. (2012). The representation of biological classes in the human brain. The Journal of Neuroscience, 32, 2608–2618.

Cortese, J. M., Dyre, B. P. (1996). Perceptual similarity of shapes generated from fourier descriptors. J Exp Psychol Hum Percept Perform., 22, 133–143.

Cutzu, F., & Edelman, S. (1998). Representation of object similarity in human vision: psychophysics and a computational model. Vision Res, 38, 2229–2257.

De Leeuw J., Groenen P. J. (1997). Inverse multidimensional scaling. J. Classif. 14, 3–2110.

Downing, P. E., Chan, A. W., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain

specificity in visual cortex. Cerebral Cortex, 16, 1453–1461.

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

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. Nature, 392, 598–601.

Farah, M. J. (1990). Visual Agnosia: Disorders of Object Recognition and What they tell us about Normal Vision. The MIT Press.

Forde, E. M. E. and Humphreys, G. W. (1999). Category-specific recognition impairments: A review of important case studies and influential theories. Aphasiology, 13: 169–193.

Franconeri, S.L., Alvarez, G.A. & Cavanagh, P. (2013). Flexible cognitive resources: competitive content maps for attention and memory. Trends in Cognitive Sciences, 17, 134–141.

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

Gobbini, M. I., Koralek, A. C., Bryan, R. E., Montgomery, K. J., & Haxby, J. V. (2007). Two takes on the social brain: A comparison of theory of mind tasks. Journal of Cognitive Neuroscience, 19, 1803–1814.

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

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron, 24, 187–203.

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. Nature Reviews Neuroscience, 15, 536–548.

Grootswagers, T., Wardle, S.G., & Carlson, T.A. (In Press). Decoding dynamic brain patterns from evoked responses: A tutorial on multivariate pattern analysis applied to time-series neuroimaging data. Journal of Cognitive Neuroscience.

Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: The early beginnings. Neuroimage, 62, 852–855.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science, 293, 2425–2430.

Haxby, J. V., Ishai, I. I., Chao, L. L., Ungerleider, L. G., & Martin, I. I. (2000). Object-form topology in the ventral temporal lobe. Response to I. Gauthier (2000). Trends in Cognitive Sciences, 4, 3–4.

He, C., Peelen, M. V., Han, Z., Lin, N., Caramazza, A., & Bi, Y. (2013). Selectivity for large

nonmanipulable objects in scene-selective visual cortex does not require visual experience. Neuroimage, 79, 1–9.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol., 160, 106–154.

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, U.S.A., 96, 9379–9384.

Kaiser, D., Azzalini, D. C., & Peelen, M. V. (2016) Shape-independent object category responses revealed by MEG and fMRI decoding. Journal of Neurophysiology, 115, 2246-2250.

Kanwisher, N. (2010). Functional specificity in the human brain: a window into the functional architecture of the mind. Proceedings of the National Academy of Sciences USA, 107, 11163–11170.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. Journal of Neuroscience, 17, 4302–4311.

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, 4296–4309.

Konen, C. S., Behrmann, M., Nishimura, M., Kastner, S. (2011). The functional neuroanatomy of object agnosia: a case study. Neuron, 71, 49–60.

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

Kourtzi, Z. & Connor, C.E. (2011) Neural representations for object perception: structure, category, and adaptive coding. Annual Review of Neuroscience, 34, 45–67.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. Proceedings of the National Academy of Sciences, U.S.A., 103, 3863–3868.

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

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron, 60, 1126–1141.

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

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. NIPS 2012: Neural Information Processing Systems.

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

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: 95-109.

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

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

Mahon, B. Z., & Caramazza, A. (2011). What drives the organization of object knowledge in the brain? Trends in Cognitive Sciences, 15, 97–103.

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

Martin, A. (2007). The representation of object concepts in the brain. Annual Review of Psychology, 58, 25–45.

Martin, A., & Weisberg, J. (2003). Neural foundations for understanding social and mechanical concepts. Cognitive Neuropsychology, 20, 575–587.

Mohan, K., & Arun, S. P. (2012). Similarity relations in visual search predict rapid visual categorization. Journal of Vision, 12, 19.

Mooney, C. M. (1957). Age in the development of closure ability in children. Can J Psychol. 11, 219–26.

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

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

Oostenveld, R., Fries, P., Maris, E., Schoffelen, JM (2011) FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. Computational Intelligence and Neuroscience Volume 2011, Article ID 156869

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. Frontiers in Neuroinformatics, 10, 27.

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

Op de Beeck, H. P., 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), 10111–23.

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, 1225–1234.

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

Peelen, M. V., He, C., Han, Z., Caramazza, A., & Bi, Y. (2014). Nonvisual and visual object shape representations in occipitotemporal cortex: Evidence from congenitally blind and sighted adults. Journal of Neuroscience, 34, 163–170.

Premack, D. (1990). The infant's theory of self-propelled objects. Cognition, 36, 1–16.

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, 680-692.

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

Riesenhuber, M., Poggio, T. (2002). Neural mechanisms of object recognition. Current Opinion in Neurobiology, 12, 162–168.

Ricciardi, E., Bonino, D., Pellegrini, S., & Pietrini, P. (2014). Mind the blind brain to understand the sighted one! Is there a supramodal cortical functional architecture? Neuroscience & Biobehavioral Reviews, 41, 64–77.

Rice G. E., Watson D. M., Hartley T., Andrews T. J. (2014). Low-level image properties of visual objects predict patterns of neural response across category-selective regions of the ventral visual pathway. Journal of Neuroscience, 34, 8837–8844.

Schultz, R. T., Grelotti, D. J., Klin, A., Kleinman, J., Van der Gaag, C., Marois, R., (2003). The role of the fusiform face area in social cognition: Implications for the pathobiology of autism. Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 358, 415–427.

Serre, T., Oliva, A., & Poggio, T. (2007) A feedforward architecture accounts for rapid categorization. PNAS, 104, 6424–6429.

Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O. (2014). The animacy continuum in the human ventral vision pathway. Journal of Cognitive Neuroscience, 27, 665–678.

Striem-Amit, E., & Amedi, A. (2014). Visual cortex extrastriate body-selective area activation in congenitally blind people “seeing” by using sounds. Current Biology, 24, 687–692.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2014). Going deeper with convolutions. CoRR, abs/1409.4842.

Tanaka, K. (1996). Inferotemporal cortex and object vision. Annual Review of Neuroscience, 19, 109–139.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. Nature, 381, 520–522.

Torgerson, W. S. (1958). Theory and Methods of Scaling. New York: Wiley.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. Nature Neuroscience, 5, 682–687.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In "Analysis of Visual Behavior" (D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield, eds.), pp. 549–586. MIT Press: Cambridge, MA.

Wardle, S., Kriegeskorte, N., Grootswagers, T., Khaligh-Razavi, S.-M., & Carlson, T. (2016). Perceptual similarity of visual patterns predicts dynamic neural activation patterns measured with MEG. NeuroImage, 132, 59-70.

Warrington, E. K. (1981). Concrete word dyslexia. British Journal of Psychology, 72, 175–196.

Warrington, E. K., & McCarthy, R. (1983). Category specific access dysphasia. Brain, 106, 859–878.

Warrington, E.K., & Shallice, T. (1984). Category-specific semantic impairments. Brain, 107, 829–853.

Watson D. M., Hartley T., Andrews T. J. (2014). Patterns of response to visual scenes are linked to the low-level properties of the image. NeuroImage, 99, 402–410.

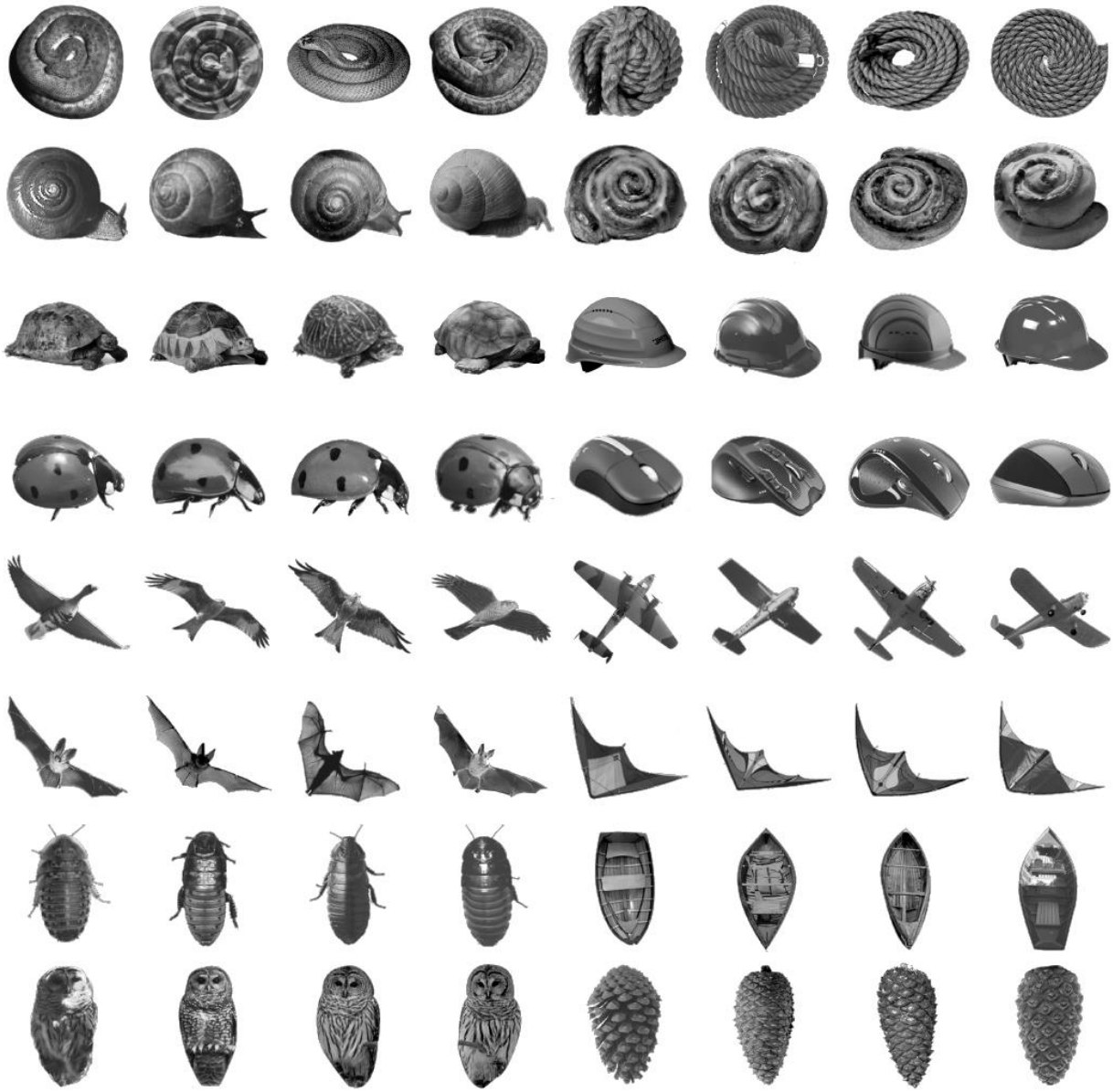
Weisberg J., van Turennout M., & Martin A. (2007). A neural system for learning about object function. Cerebral Cortex 17, 513–521.

Wiggett, A. J., Pritchard, I. C., & Downing, P. E. (2009). Animate and inanimate objects in human visual cortex: Evidence for task-independent category effects. Neuropsychologia, 47, 3111–3117.

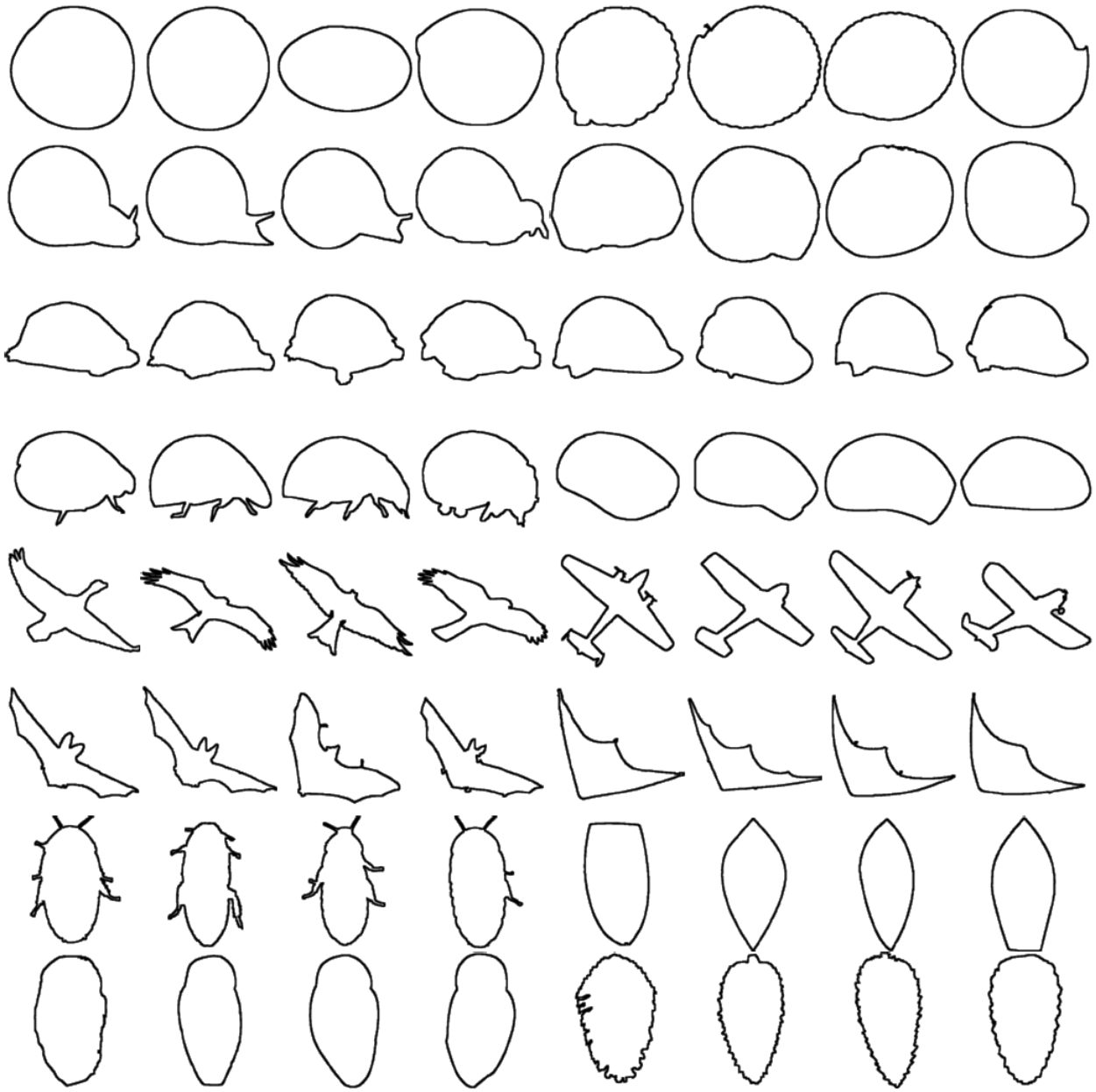
Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. Behavior Research Methods, 42, 671–684.

Wolbers, T., Klatzky, R. L., Loomis, J. M., Wutte, M. G., & Giudice, N. A. (2011). Modality-independent coding of spatial layout in the human brain. Current Biology, 21, 984–989.

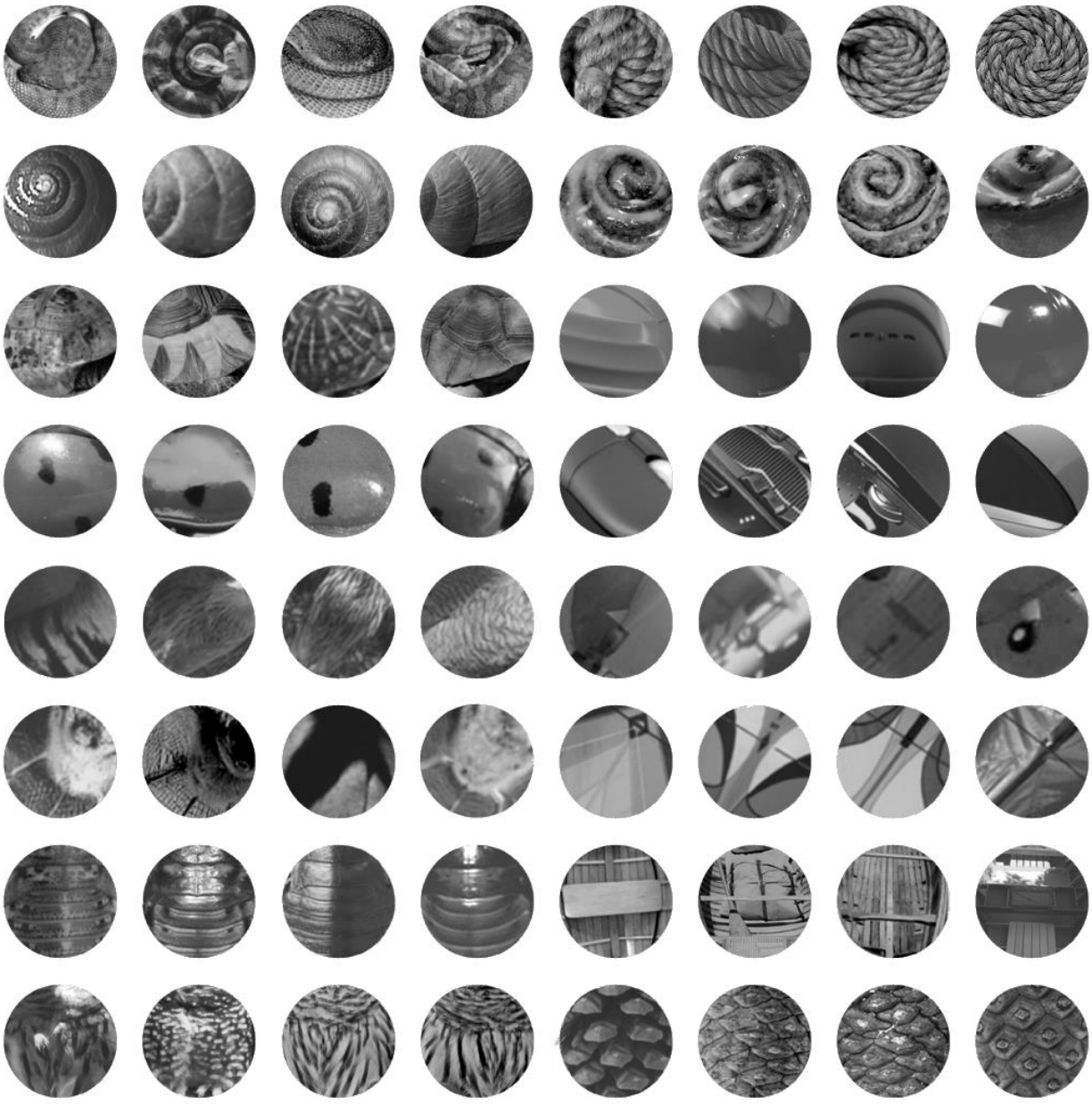
## Supplementary Information



**Supplementary Figure 1. The complete stimulus set.** The complete stimulus set included four versions for each of the original 16 images (Figure 2.1). This resulted in 16 images for each of the four shape subsets and 64 images in total.



**Supplementary Figure 2. The complete stimulus set: Outlines.** The set of 64 object outlines was obtained by tracing the outline of the binarized silhouette versions of the original stimuli.



**Supplementary Figure 3. The complete stimulus set: Textures.** The set of 64 circular texture patches was created by masking the original images with a circular aperture in the center of the image.

## Acknowledgements

First and foremost, I would like to thank my supervisor Marius Peelen for making me a part of his team and providing guidance and support at every stage of the project. I was lucky to work in the lively and dynamic lab which encouraged me to learn fast and prioritize my research. I am grateful for all the freedom that I had.

I would also like to thank my colleague Daniel Kaiser, from whom I learned a lot and who was always available to answer every little question, share the code and provide useful insights. Thanks for all the car rides to and from Mattarello and for psychological support during my first fMRI and MEG testing days.

Thanks to Nick Oosterhof for creating the great CosmoMVPA toolbox and for always being there to answer my questions about it and provide useful advice about data analysis.

Thanks to all the wonderful members of the Peelen Lab whom I met during my time here: Reshanne Ruhnau, Talli Brandman, Timo Stein, Ludwig Barbaro, Elisa Battistoni, Chiara Avancini, Elena Aggius-Vella, Sushrut Thorat. It was really inspiring to work alongside you!

Thanks to all of my Rovereto friends who helped to make it my favorite place in the world. I would particularly like to mention Talli, Daniel S., Nick O., Matt, Hayley, Patrik, Mia, Asya & Denis, Katya D., Yanina, and Vassiki. Special thanks to Anna Blumenthal, my friend since my very first days in Italy, who provided extremely useful feedback on this thesis.

Finally, I would like to thank Giacomo, who was there for me in the worst and the best times, encouraging and supporting me no matter what, making sure I have something other than work in my life, and setting the best possible example.

# Appendix

**Disentangling representations of object shape and object category in human visual cortex: the animate-inanimate distinction.**

Proklova, D., Kaiser, D., & Peelen, M.V. (2016)

*Journal of Cognitive Neuroscience, 28, 680-692.*

# Disentangling Representations of Object Shape and Object Category in Human Visual Cortex: The Animate–Inanimate Distinction

**Daria Proklova*, Daniel Kaiser*, and Marius V. Peelen**

## Abstract

■ Objects belonging to different categories evoke reliably different fMRI activity patterns in human occipitotemporal cortex, with the most prominent distinction being that between animate and inanimate objects. An unresolved question is whether these categorical distinctions reflect category-associated visual properties of objects or whether they genuinely reflect object category. Here, we addressed this question by measuring fMRI responses to animate and inanimate objects that were closely matched for shape and low-level visual features. Univariate contrasts revealed animate- and inanimate-preferring regions in ventral and lateral temporal cortex even for individually matched object pairs (e.g., snake–rope). Using representational similarity analysis, we mapped out brain regions in which the pairwise dissimilarity of multivoxel activity patterns (neural dissimilarity) was predicted by the objects' pairwise visual dissimilarity and/or their categorical dissimilarity. Visual dissimilarity was measured as the time it took participants to find a unique target among identical distractors in three visual search experiments, where we separately quantified overall dissimilarity, outline dissimilarity, and texture dissimilarity. All three visual dissimilarity structures predicted neural dissimilarity in regions of visual cortex. Interestingly, these analyses revealed several clusters in which categorical dissimilarity predicted neural dissimilarity after regressing out visual dissimilarity. Together, these results suggest that the animate–inanimate organization of human visual cortex is not fully explained by differences in the characteristic shape or texture properties of animals and inanimate objects. Instead, representations of visual object properties and object category may coexist in more anterior parts of the visual system. ■

## INTRODUCTION

Large-scale patterns of fMRI activity spanning the ventral temporal cortex (VTC) distinguish animate from inanimate object categories (e.g., Kriegeskorte, Mur, Ruff, et al., 2008), with animate objects evoking higher BOLD responses in lateral VTC and inanimate objects evoking higher BOLD responses in medial VTC (e.g., Mahon et al., 2007; Downing, Chan, Peelen, Dodds, & Kanwisher, 2006; Chao, Haxby, & Martin, 1999). Within these broader regions, focal regions exhibit selective responses to more specific categories, including regions selective for buildings and scenes, faces, tools, body parts, and words (Peelen & Downing, 2005; Cohen & Dehaene, 2004; Downing, Jiang, Shuman, & Kanwisher, 2001; Chao et al., 1999; Epstein & Kanwisher, 1998; Kanwisher, McDermott, & Chun, 1997). Although the selectivity for object categories in VTC has been widely replicated, particularly the animate–inanimate distinction, the factors driving this selectivity are still under debate (Andrews, Watson, Rice, & Hartley, 2015; Grill-Spector & Weiner, 2014; Mahon & Caramazza, 2011; Op de Beeck, Haushofer, & Kanwisher, 2008; Martin, 2007).

One of the key questions is whether category-specific patterns of brain activity reflect genuine categorical distinctions (Caramazza & Shelton, 1998) or whether these can be alternatively explained by factors that covary with category membership, such as shape properties. Because of the close association between certain visual properties and category membership, it is to be expected that category-selective regions are optimized for processing these visual properties and/or that these regions are located in parts of the visual system that have visual and retinopic biases that are optimal for processing the visual features that are characteristic of the category. However, although specific visual properties often characterize object categories, these two dimensions (visual, categorical) are not identical and can indeed be experimentally dissociated. For example, although most tools are elongated, this shape property can be dissociated from the conceptual properties associated with tools (e.g., that tools are manipulable and used as effectors; Bracci & Peelen, 2013). For a visually more homogenous category such as animals, this distinction is more challenging but may still be addressed by testing responses to visually less typical examples (e.g., snakes) and, conversely, testing responses to inanimate objects that share visual features with animals (e.g., mannequins, dolls, statues). These considerations raise the intriguing question of whether category selectivity in VTC

University of Trento
*These authors contributed equally to this work.



Journal of Cognitive Neuroscience 28:5, pp. 680–692
doi:10.1162/jocn_a_00924

reflects selectivity for conceptual category or selectivity for visual properties that characterize a category.

According to the object form topology account, category-selective fMRI responses in VTC reflect the activation of object form representations that are mapped onto VTC in a continuous fashion (Haxby, Ishai, Chao, Ungerleider, & Martin, 2000; Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999). The selective response to animals in VTC may thus arise from selectivity for characteristic animal shape(s) rather than selectivity for animacy per se. A recent monkey study provided support for this hypothesis, showing that the organization of animate and inanimate object representations in monkey inferotemporal cortex primarily reflects visual similarity rather than semantic similarity (Baldassi et al., 2013; but see Kiani, Esteky, Mirpour, & Tanaka, 2007). Further support for the visual similarity account comes from fMRI studies showing that category-selective regions in VTC respond selectively to visual properties that are characteristic of the regions' preferred categories, even for otherwise meaningless stimuli (i.e., in the absence of category recognition). For example, the fusiform face area was shown to respond more strongly to oval shapes with a greater number of black elements in the top half than to oval shapes with a greater number of elements in the bottom half, although none of these stimuli were recognized as faces (Caldara et al., 2006). Similarly, the parahippocampal place area, located within the medial inanimate-preferring VTC, was shown to respond preferentially to objects made up of cardinal orientations and right angles, features typical of manmade objects, buildings, and scenes (Nasr, Echavarria, & Tootell, 2014).

Recent evidence against a "visual properties" account of category selectivity in VTC comes from studies in congenitally blind individuals. These individuals, with no visual experience, show a categorical organization of VTC that is remarkably similar to that observed in sighted individuals (Ricciardi, Bonino, Pellegrini, & Pietrini, 2014). For example, aurally presented words describing large inanimate objects, versus animals, activate medial VTC in both blind and sighted groups (He et al., 2013; Mahon, Anzellotti, Schwarzbach, Zampini, & Caramazza, 2009). Using a variety of presentation methods, most of the category-selective VTC regions found in sighted individuals have now also been reported in blind individuals, often at nearly identical anatomical locations in the two groups (Striem-Amit & Amedi, 2014; Peelen et al., 2013; Reich, Szwed, Cohen, & Amedi, 2011; Wolbers, Klatzky, Loomis, Wutte, & Giudice, 2011; Buchel, Price, & Friston, 1998). These studies show that the processing of visual features is not necessary for some category-selective responses to develop. However, they do not exclude the possibility that category selectivity in VTC nevertheless reflects shape properties of objects. This is because VTC has been shown to extract object shape from nonvisual input modalities (Amedi et al., 2007; Amedi, von Kriegstein, van Atteveldt, Beauchamp, & Naumer, 2005), with VTC activity patterns reflecting the shape similarity of objects in both blind and sighted groups (Peelen, He, Han, Caramazza, & Bi, 2014).

This study was designed to investigate the contribution of shape similarity in the representation of animate and inanimate object categories in VTC. Participants viewed pictures of a variety of animals that systematically differed in their shape, grouping into four shape clusters (Figure 2A, right). Importantly, inanimate control objects were selected to closely match the animals in terms of their shape, following the same four shape clusters. This design allowed us to test whether animate- and inanimate-preferring regions (localized with a standard functional localizer) maintain their selectivity for carefully matched animate–inanimate pairs (e.g., snake vs. rope) and whether this is true for a variety of animals (e.g., birds, insects, reptiles) and inanimate objects (e.g., plane, rope, pine cone). In addition to analyses measuring activation differences, we used representational similarity analysis (RSA) to map out regions in which neural similarity reflected the objects' visual and/or categorical similarity (animate/inanimate). For this purpose, we quantified pairwise visual similarity using visual search tasks designed to measure different aspects of visual similarity (overall visual similarity, outline similarity, and texture similarity; Figure 2).

## METHODS

### Participants

Eighteen participants (seven men; mean age = 25 years, *SD* = 2.4 years) were scanned at the Center for Mind/Brain Sciences of the University of Trento. All participants gave informed consent. All procedures were carried out in accordance with the Declaration of Helsinki and were approved by the ethics committee of the University of Trento. One participant was excluded from all analyses because of excessive head movement.

### Stimuli

The stimuli of the main experiment were organized into four sets of four objects. The four objects within each set all had a roughly similar shape. Two objects of each set were animate, and two objects were inanimate (see Figures 2A and 4). In addition, there were four exemplars of each object (e.g., four images of a snake), resulting in 16 stimuli per set and a total of 64 stimuli. All images were gray scaled, placed on gray background and matched for luminance and contrast using the SHINE toolbox (Willenbockel et al., 2010). Stimulus presentation was controlled using the Psychtoolbox (Brainard, 1997). Images were back-projected on a translucent screen placed at the end of the scanner bore. Participants viewed the screen through a tilted mirror mounted on the head coil. Stimuli were presented foveally and subtended a visual angle of approximately 4.5°.

### Visual Search Experiments

To provide a measure of pairwise visual similarity of the stimulus set, a series of three behavioral visual search experiments was conducted. In these experiments, participants searched for an oddball target surrounded by identical distractor objects (Figure 2). The response time in this task is a measure of visual similarity (Mohan & Arun, 2012): The longer the response time for locating the oddball stimulus, the more visually similar are the target and the distractor object. Experiment 1 measured overall visual similarity, Experiment 2 measured outline visual similarity, and Experiment 3 measured texture visual similarity.

### Experiment 1

To quantify overall pairwise visual similarity of the stimulus set, 18 new participants were tested in a behavioral experiment (two men; mean age = 22.5 years, $SD$ = 2.97 years). Stimuli were presented on a 17-in. CRT monitor, and presentation was controlled using Psychtoolbox (Brainard, 1997). Each search display contained 16 objects placed in a 4 × 4 grid, with one oddball target and 15 identical images of the distractor object. The location of the 16 objects in the grid was randomized. The size of the target and seven of the distractors was 100 × 100 pixels, which corresponded to 2.9° visual angle. The remaining distractors differed in size, with four being 120% of the target size (3.4° visual angle) and four being 80% of the target size (2.3° visual angle). Participants had to indicate whether the oddball target appeared on the left side or on the right side of the screen. No information about the category of the oddball target was provided. The search display remained on the screen until the response, followed by 500-msec fixation, after which the next trial started. The experiment consisted of four blocks. In each block, only one of the four exemplars of each object was used (e.g., always the same snake within one block), resulting in 16 unique objects and 240 trials per block (for all possible target–distractor pairings of the 16 stimuli). Accuracy was high (97.2%) and was not further analyzed.

RTs were averaged across corresponding target–distractor object pairs and across blocks. The data from the visual search experiment served to create a matrix of overall pairwise visual dissimilarity, to be used as a predictor in the fMRI analysis. For this purpose, we took the inverse of these RTs (1/RT) for each stimulus pair as a measure of dissimilarity. The resulting visual dissimilarity matrix consisted of one visual dissimilarity value for every pairwise combination of objects (Figure 2A, center). Multidimensional scaling analysis (using *cmdscale* function in MATLAB, The MathWorks, Natick, MA) revealed that stimuli from the same shape sets clustered together (Figure 2A, right), whereas there was no apparent categorical organization. Furthermore, within each shape set, there was no evidence for categoricality, with the average visual dissimilarity within categories (e.g., snake–snail) being equal to the average visual dissimilarity across categories (e.g., rope–snail), $t(17) = 1.22$, $p = .239$. These results confirm our intuitive shape sets and show that there were no obvious visual properties that covaried with category membership.

### Experiments 2 and 3

To measure pairwise similarity of the outline shape of the stimuli, 18 new participants (three men; mean age = 23.3 years, $SD$ = 3.4 years) were tested in Experiment 2. The experiment was identical to Experiment 1 except that the stimulus set consisted of outline drawings of the stimuli, created by automatically tracing the outline contours of binarized silhouette versions of the original stimuli (see Figure 2B). Accuracy was high (98.1%) and was not further analyzed.

To measure pairwise texture similarity of the stimulus set, 18 participants (eight men; mean age = 25.7 years, $SD$ = 5.5 years) were tested in Experiment 3. Two of the participants had also participated in Experiment 2. The experiment was identical to Experiments 1 and 2 except that the stimulus set consisted of circular texture patches, created by masking the original images with a circular aperture that covered about 20% of the image (see Figure 2C). The aperture was centered on the mean pixel coordinate of the image. Any blank spaces were filled in using the clone stamp tool in Photoshop. This circular masking abolishes outline shape information, while leaving inner features and texture properties of the stimuli largely intact. It should be noted that these patches still contained some local structure (e.g., the inner contour of the snake) and may thus capture some internal shape features in addition to texture properties. Accuracy was high (98.1%) and was not further analyzed.

Data for Experiments 2 and 3 were analyzed as in Experiment 1, resulting in dissimilarity matrices representing outline and texture dissimilarity of the stimulus set (Figure 2B and C, center). Further analyses showed that overall visual dissimilarity could be nearly perfectly predicted by a linear combination of outline and texture dissimilarity, with the optimal weights being 0.75 and 0.25, respectively. Linear combinations of dissimilarity matrices were computed by using data from different single participants (e.g., 0.75 × outline of one participant + 0.25 × texture of another participant), each of which was then correlated with the average visual dissimilarity matrix (with one participant left out). The resulting average correlation ($r = .77$) approached the noise ceiling of visual dissimilarity ($r = .82$, computed by correlating each participant's visual dissimilarity matrix with the group-averaged visual dissimilarity matrix, leaving out this participant). Finally, the combined model was significantly more strongly correlated with visual dissimilarity than was either outline or texture alone ($p < .001$, for both comparisons). These analyses show that overall visual dissimilarity is influenced both by outline and texture

properties, which together almost fully explain overall visual dissimilarity.

### Main fMRI Experiment Procedure

The main fMRI experiment consisted of eight runs. Each run consisted of 80 trials that were composed of 64 object trials and 16 fixation-only trials. In object trials, a single stimulus was presented for 300 msec, followed by a 3700-msec fixation period (Figure 1A). In each run, each of the 64 images appeared exactly once. In fixation-only trials, the fixation cross was shown for 4000 msec. Trial order was randomized, with the constraints that there were exactly eight 1-back repetitions of the same category (e.g., two snakes in direct succession) within the object trials and that there were no two fixation trials appearing in direct succession. Each run started and ended with a 16-sec fixation period, leading to a total run duration of 5.9 min. Participants were instructed to press a button whenever they detected a 1-back repetition.

### Functional Localizer Experiment Procedure

In addition to the main experiment, participants completed one run of a functional localizer experiment. During the localizer, participants viewed grayscale pictures of 36 animate and 36 inanimate stimuli in a block design (Figure 1B). Animate stimuli included five different types of animals (mammals, birds, fish, reptiles, and insects). Inanimate stimuli included five types of inanimate objects (cars, chairs, musical instruments, tools, and weapons). These stimuli were not matched for their shape (thus, this design resembled the standard animate–inanimate contrast used in previous studies). Each block lasted 16 sec, containing 20 stimuli that were each presented for 400 msec, followed by 400-msec blank interval. There were eight blocks of each stimulus category and four fixation-only blocks per run. The order of the first 10 blocks was randomized and then mirror reversed for the other 10 blocks. Participants were asked to detect 1-back image repetitions, which happened twice during every nonfixation block.

### fMRI Acquisition

Imaging data were acquired using a MedSpec 4-T head scanner (Bruker Biospin GmbH, Rheinstetten, Germany), equipped with an eight-channel head coil. For functional imaging, T2*-weighted EPIs were collected (repetition time = 2.0 sec, echo time = 33 msec, 73° flip angle, 3 × 3 × 3 mm voxel size, 1-mm gap, 34 slices, 192-mm field of view, 64 × 64 matrix size). A high-resolution T1-weighted image (magnetization prepared rapid gradient echo; 1 × 1 × 1 mm voxel size) was obtained as an anatomical reference.

### fMRI Preprocessing and Modeling

The neuroimaging data were analyzed using MATLAB and SPM8. During the preprocessing, the functional volumes were realigned, coregistered to the structural image, resampled to a 2 × 2 × 2 mm grid, and spatially normalized to the Montreal Neurological Institute 305 template included in SPM8. For the univariate analysis, the functional images were smoothed with a 6-mm FWHM kernel, whereas for the multivariate analysis, the images were left unsmoothed. For the main experiment, the BOLD signal of each voxel in each participant was modeled using 22 regressors in a general linear model, with 16 regressors for each of the objects (e.g., one regressor for all snakes) and six regressors for the movement parameters obtained from the realignment procedure. For the functional

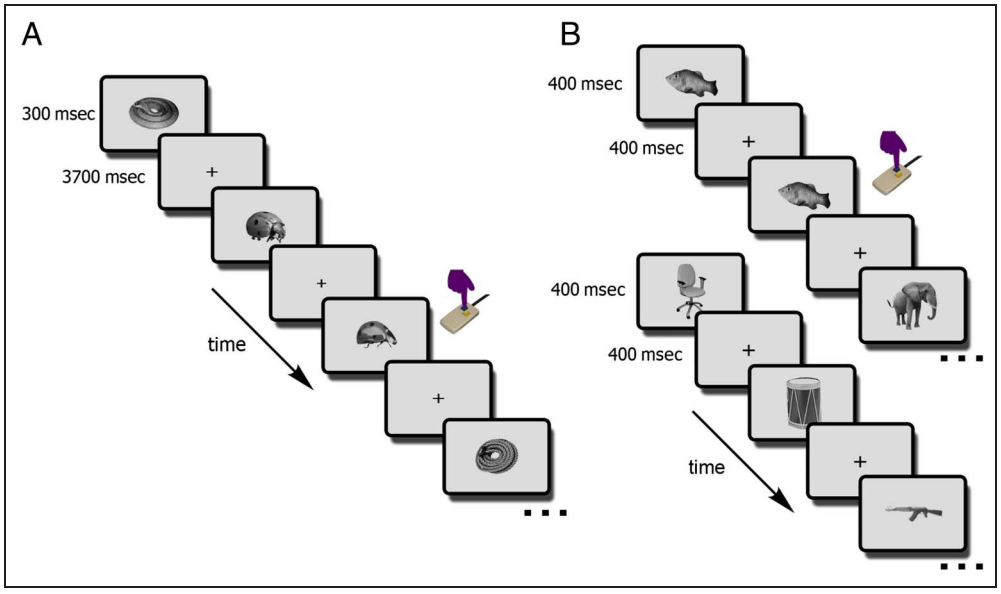



**Figure 1.** fMRI paradigm. (A) In the main fMRI experiment, participants viewed images of animate objects and shape-matched inanimate objects (see Figure 4 for further stimulus examples). Trial order was randomized, and participants detected, by button press, 1-back object-level repetitions (here, two ladybugs). (B) In the functional localizer experiment, participants viewed blocks of animate (top sequence) and inanimate (bottom sequence) stimuli. All stimuli were different from the ones used in the main experiment. Each block lasted 16 sec, and participants detected, by button press, 1-back image-level repetitions (here, the fish image).

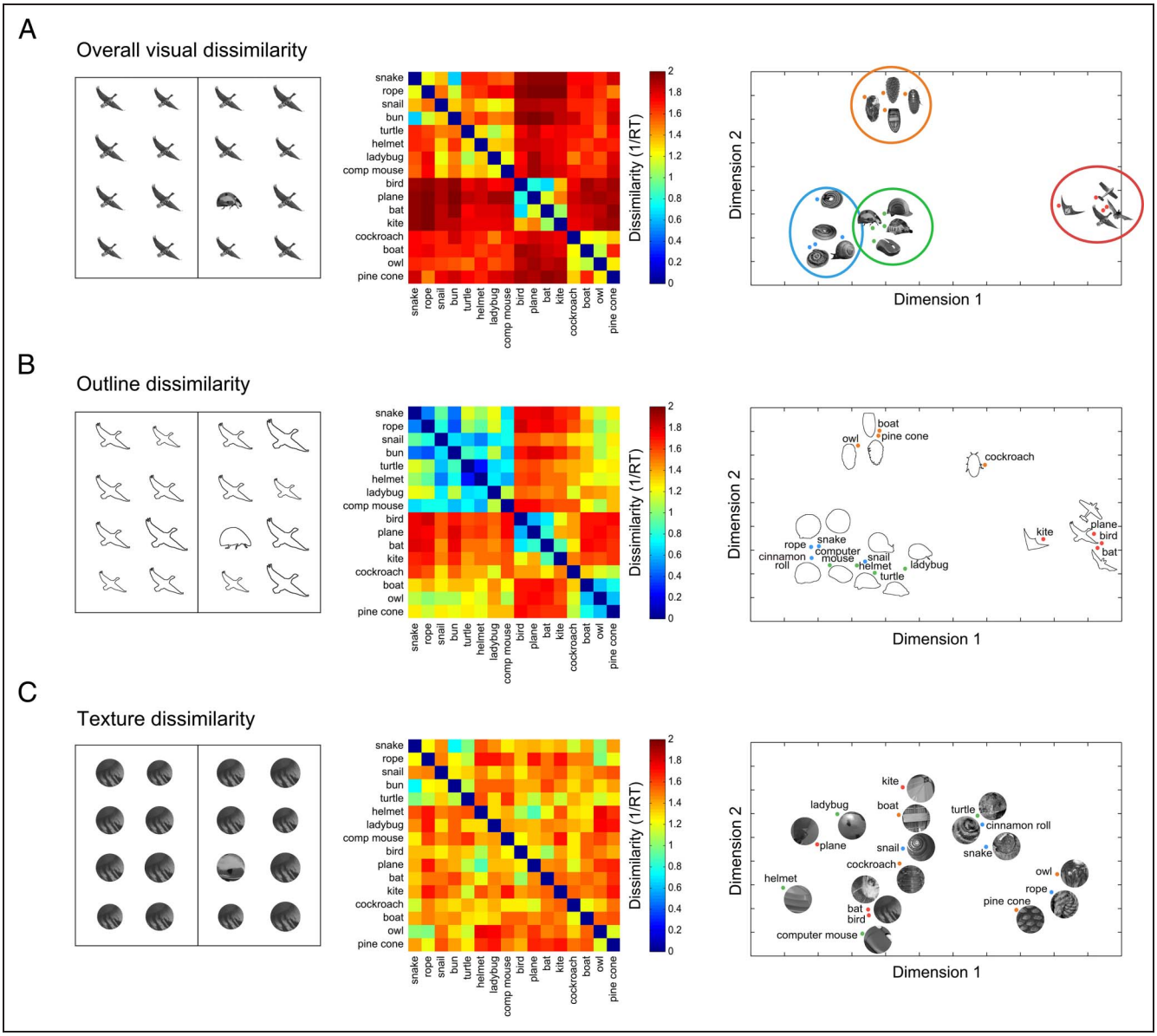



**Figure 2.** Visual search experiments. In three visual search experiments, participants indicated whether an oddball target in a 4 × 4 search array (left) was located to the right or to the left of the vertical display midline. No prior information about the target was given, so that participants had to rely on bottom–up visual differences to perform the task. From the RTs in these experiments, we created visual dissimilarity matrices (center), where each element represents the inverse RT for a pair of stimuli, averaged across the two respective target–distractor pairings: High dissimilarity values thus correspond to short RTs, reflecting that target and distractor were visually dissimilar. Multidimensional scaling representations of the three visual dissimilarity matrices are shown in the right panels. The three experiments differed only in the stimuli used: (A) the original images used in the fMRI experiment, (B) outline drawings of these images, and (C) internal textures of these images.

localizer data, the signal was modeled using two regressors (animate and inanimate objects) and six movement regressors. All models included an intrinsic temporal high-pass filter of 1/128 Hz to correct for slow scanner drifts.

### Univariate Analysis

Univariate random effects whole-brain analyses were performed separately for the localizer and the main experiment, contrasting animate with inanimate objects. Statistical maps were thresholded using a voxel-level threshold of $p < .001$ (uncorrected) and a cluster-level threshold of $p < .05$ (family-wise error [FWE] corrected). In addition, regions activated in the localizer were defined as ROIs. Within these ROIs, beta estimates for the conditions of the main experiment were extracted and averaged across the voxels of each ROI. These beta values were statistically compared using ANOVAs and $t$ tests.

### RSA

RSA (Kriegeskorte, Mur, & Bandettini, 2008) was used to relate the visual and categorical dissimilarity of the objects to neural dissimilarity. RSA was performed throughout the

whole brain using searchlight analysis (Kriegeskorte, Goebel, & Bandettini, 2006), implemented in the CoSMoMVPA software package (www.cosmomvpa.org). Each spherical searchlight neighborhood consisted of 100 voxels, centered on every voxel in the brain. For each of these spheres, we correlated the activity (beta values) between each pair of conditions from the main experiment across the voxels of the sphere, leading to a 16 × 16 symmetrical correlation matrix with an undefined diagonal. This matrix was transformed into a neural dissimilarity matrix by subtracting the correlation values from 1.

In a first analysis, neural dissimilarity matrices were related to the visual dissimilarity matrix and the categorical dissimilarity matrix using multiple regression analysis (see Figure 5A). The visual dissimilarity matrix was derived from RTs in a visual search experiment (Experiment 1; Figure 2A), whereas the categorical dissimilarity matrix reflected whether two objects were from the same category (0) or from different categories (1). All dissimilarity matrices were $z$ normalized. The multiple regression analysis yielded beta estimates for the two predictors of neural dissimilarity (visual and categorical dissimilarity), reflecting the independent contributions of these predictors in explaining neural dissimilarity. These two beta estimates were obtained for all spheres, resulting in two whole-brain maps for each participant. These maps were then tested against zero using random effects analyses ($t$ tests), thresholded using a voxel-level threshold of $p < .001$ (uncorrected) and a cluster-level threshold of $p < .05$ (FWE corrected). In the second analysis, neural dissimilarity matrices were related to outline dissimilarity (Figure 2B), texture dissimilarity (Figure 2C), and categorical dissimilarity (Figure 6A). In all other respects, the analysis was the same as the first analysis described above.

## RESULTS

### Univariate Results

#### *Whole-brain Analysis*

The contrast between animate and inanimate objects in the functional localizer experiment revealed a characteristic medial-to-lateral organization in VTC (Figure 3A). In line with previous findings, animate stimuli more strongly activated regions around the lateral fusiform gyrus (left hemisphere [LH]: 1936 $mm^3$, peak Montreal Neurological Institute coordinates: $x = -40, y = -48, z = -22$; right hemisphere [RH]: 3424 $mm^3$, peak coordinates: $x = 42, y = -52, z = -20$), and inanimate stimuli preferentially activated more medial regions around parahippocampal gyrus (LH: 3760 $mm^3$, peak coordinates: $x = -28, y = -46, z = -12$; RH: 3184 $mm^3$, peak coordinates: $x = 34, y = -42, z = -12$). In addition to these ventral regions, animate stimuli preferentially activated a more posterior and lateral region, around middle temporal gyrus (LH: 3528 $mm^3$, peak coordinates: $x = -48, y = -80, z = 0$; RH: 7800 $mm^3$, peak coordinates: $x = 52, y = -74, z = -2$).

The same animate–inanimate comparison was performed for the main experiment, in which the animate and inanimate stimuli were closely matched for shape and low-level visual features (see Methods). This contrast revealed a significant animacy organization: Animate stimuli more strongly activated two clusters in the RH, again around fusiform gyrus (808 $mm^3$, peak coordinates: $x = 42, y = -52, z = -20$) and middle temporal gyrus (1120 $mm^3$, peak coordinates: $x = 50, y = -74, z = 0$). Similar to the functional localizer results, inanimate-preferring regions were found around bilateral parahippocampal gyrus (LH: 4736 $mm^3$, peak coordinates: $x = -26, y = -52, z = -16$; RH: 3456 $mm^3$, peak coordinates: $x = 24, y = -40, z = -16$). Figure 3A shows the animate- and inanimate-preferring clusters from the localizer and the main experiment as well as their overlap.

These results indicate that the medial-to-lateral animacy organization is also found when controlling for shape differences of animate and inanimate objects. As can be seen in Figure 3A, activity was stronger in the localizer than in the main experiment. This effect is hard to interpret, however, given the many differences between the localizer and the main experiment (e.g., block design vs. event-related design, stimulus duration, the specific animals and objects included, etc.). Indeed, the purpose of this analysis was not to compare the strength of activity between localizer and main experiment directly but to show that the medial-to-lateral organization is remarkably similar in both experiments. Finally, although some of the LH clusters did not survive multiple comparisons correction in the main experiment, the functionally localized ROIs maintained their selectivity in the main experiment in both hemispheres, as reported in the next section.

#### *ROI Analysis*

ROI analyses were used to test for selectivity for the conditions in the main experiment within each of the six clusters of the functional localizer (Figure 3A; for coordinates and cluster sizes, see the whole-brain analysis section above). A three-way ANOVA with the factors Animacy (animate, inanimate), Region (animate lateral, animate ventral, inanimate ventral), and Hemisphere (right, left) revealed a critical Region × Animacy interaction ($F(2, 32) = 42.2, p < .001$). Because there were no interactions with hemisphere ($F < 2.49, p > .10$, for all tests), data were collapsed across hemispheres for all follow-up analyses. Separate Region × Animacy ANOVAs for every pair of regions revealed that the animacy preferences of the two animate-preferring regions were each significantly different from that of the inanimate-preferring region (ventral animate region: $F(1, 16) = 89.24, p < .001$; lateral animate region: $F(1, 16) = 45.65, p < .001$). There was no significant difference in animacy preference between the two animate regions ($F(1, 16) = 0.59, p = .45$). As expected,

both animate-preferring regions showed an increased response to the animate stimuli in the main experiment ($t(16) > 3.53$, $p < .003$, for both tests), whereas the inanimate-preferring region showed a significant preference for the inanimate stimuli ($t(16) = 5.25$, $p < .001$). These results show that all regions defined in the functional localizer maintained their selectivity in the main experiment (see Figure 3B for results in separate hemispheres).

To explore whether the animate–inanimate organization in the main experiment was driven by some of the stimuli preferentially, we next compared the responses to each of the eight individual animate stimuli and their shape-matched inanimate counterparts. For simplicity and for optimal statistical power, the conditions were recoded into "preferred" (e.g., a snake for an animate-peferring region) and "nonpreferred" conditions (e.g., a snake for an inanimate-preferring region), and responses were then averaged across ROIs, so that each ROI contributed equally. A two-way ANOVA with the factors Preference (preferred, nonpreferred) and Object pair (the eight different pairs) revealed a significant interaction ($F(7, 112) = 3.68$, $p = .001$), indicating that different object pairs differentially contributed to the observed animacy organization (Figure 4). A significant category preference was observed for six of the eight pairs ($t(16) > 2.16$, $p < .047$, for all tests), with one pair (ladybug–computer mouse) showing a trend ($t(16) = 2.10$, $p = .052$) and one pair (snail–bun) not reaching significance ($t(16) = 0.22$, $p = .83$).

## RSA

In addition to showing overall differences in focal regions, objects of different categories evoke distinct multivoxel activity patterns in visual cortex: Within VTC, activity patterns to objects from the same category are more similar than activity patterns to objects from different categories (Haxby et al., 2001). An open question is

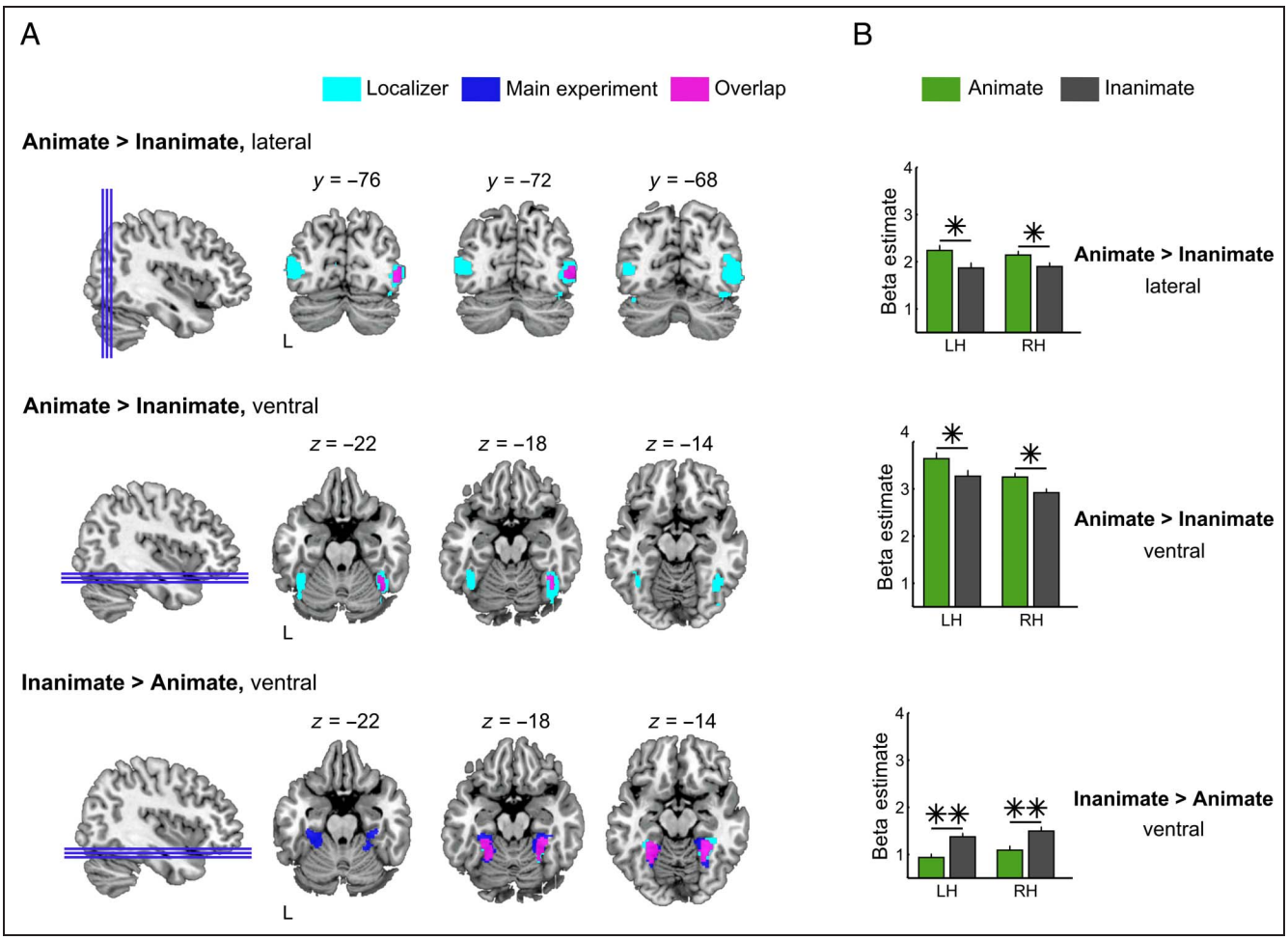



**Figure 3.** Univariate results. (A) Results of univariate whole-brain group analysis comparing animate and inanimate conditions in the functional localizer experiment (cyan) and main experiment (blue). Overlapping regions are displayed in purple. Top row shows lateral animate-preferring clusters, middle row shows ventral animate-preferring clusters, and bottom row shows ventral inanimate-preferring clusters. Statistical maps were thresholded at $p < .05$ (FWE corrected) and overlaid on a structural brain template (MRIcron). (B) Results of ROI analyses. The bar graphs show the responses to the conditions in the main experiment within the animate- and inanimate-preferring clusters defined based on the functional localizer experiment (corresponding to the cyan regions in A), separately for each hemisphere. Green bars indicate average response to animate objects; gray bars indicate average response to inanimate objects. Error bars reflect *SEM* difference. *$p < .01$, **$p < .001$.

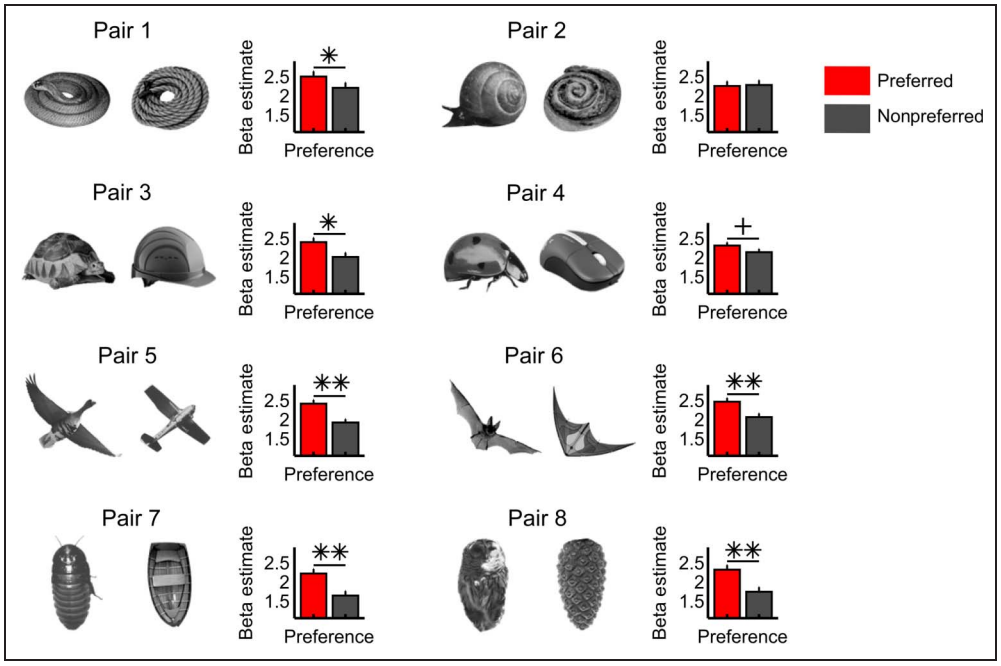



**Figure 4.** Individual pair analysis. Category preference analyzed for each shape-matched pair separately. Responses were combined across all regions defined in the localizer experiment (cyan regions in Figure 3A) by recoding responses according to the preference of the region (e.g., snake response in animate-preferring regions and rope response in inanimate-preferring region both contribute to the "preferred" condition, shown in red). Error bars reflect *SEM* difference. $+p = .052$, $*p < .05$, $**p < .001$.

whether these effects reflect visual differences and/or categorical differences between objects. To address this question, we used RSA to relate neural dissimilarity (based on correlations between multivoxel activity patterns) to visual and categorical dissimilarity. In this analysis, for every spherical neighborhood (100 voxels) of the brain, the pairwise neural dissimilarity structure between the 16 objects was modeled as a linear combination of their pairwise categorical dissimilarity and overall visual dissimilarity (Figure 5A; see Methods). Overall visual dissimilarity was quantified using response times in a visual search task (Figure 2A), capturing all contributing factors to visual discriminability (potentially beyond the ones we explicitly matched in the univariate comparisons). This analysis yielded two beta maps reflecting the independent contributions of visual and categorical variables in accounting for neural dissimilarity.

Random effects group analysis revealed widespread clusters of voxels in which neural dissimilarity was significantly related to overall visual dissimilarity, including primary visual cortex and large parts of extrastriate visual cortex, extending into VTC (51,864 $mm^3$ in total; peak in the lingual gyrus, $x = -18, y = -92, z = -8$). These clusters partly overlapped with the animate- and inanimate-preferring regions of the functional localizer experiment (cyan regions in Figure 3A), with visual dissimilarity being significant in 64% of the ventral animate-preferring voxels, 24% of the lateral animate-preferring voxels, and 3% of the ventral inanimate-preferring voxels. Interestingly, categorical dissimilarity was independently reflected in two clusters: one in the right ventral visual cortex (3512 $mm^3$; peak in the fusiform gyrus, $x = 42, y = -60, z = -18$) and one in the LH (4400 $mm^3$ in total, including a lateral visual cortex part with a local peak in middle occipital gyrus, $x = -42, y = -80, z = 6$, and a ventral visual cortex part with a local peak in fusiform gyrus, $x = -44, y = -52, z = -16$). As can be seen in Figure 5B, these clusters partly overlapped with the anterior end of the overall visual dissimilarity clusters. These clusters also partly overlapped with the animate- and inanimate-preferring regions of the functional localizer experiment, with categorical dissimilarity being significant in 33% of the ventral animate-preferring voxels, 10% of the lateral animate-preferring voxels, and 2% of the ventral inanimate-preferring voxels.

It is possible that the measure of overall visual dissimilarity captured some visual features better than others. For example, the overall visual dissimilarity structure could be driven more by the outline shape than by texture properties of the stimuli. In this case, if animate and inanimate stimuli consistently differed in their texture, the category-selective regions revealed in the previous analysis could in principle reflect texture information rather than category information. To address this possibility, we quantified the pairwise dissimilarity structure for outline shape and texture independently from each other in two further visual search experiments (see Methods). Using these data, we repeated the RSA, this time modeling pairwise neural dissimilarity using the combination of three predictors: outline shape dissimilarity (Figure 2B), texture dissimilarity (Figure 2B), and categorical dissimilarity (see Figure 6A). This analysis resulted in three beta maps reflecting the independent contributions of outline shape, texture, and category membership to the neural dissimilarity structure.

The results of random effects group analyses on these three maps are shown in Figure 6B. Four widespread clusters of voxels in which neural dissimilarity reflected

outline dissimilarity were found in primary visual cortex extending to extrastriate visual cortex (22,584 $mm^3$, occipital and temporal lobe; peak around the left lingual gyrus, $x = -22, y = -92, z = -8$), right superior parietal cortex (peak $x = 16, y = -50, z = 60$), right ventral visual cortex (1552 $mm^3$; peak $x = 44, y = -64, z = -12$), and left temporal lobe around fusiform gyrus (1240 $mm^3$; peak in BA 37, $x = -28, y = -46, z = -18$). Texture dissimilarity was related to neural dissimilarity in one cluster of voxels in left early visual cortex (2456 $mm^3$; peak in the lingual gyrus, $x = -20, y = -84, z = -14$). Crucially, categorical dissimilarity was independently reflected in two clusters: one in right ventral visual cortex (2928 $mm^3$; peak in the fusiform gyrus, $x = 42, y = -60, z = -18$) and one in left lateral visual cortex (3960 $mm^3$; peak in the middle occipital gyrus, $x = -42, y = -80, z = 6$). Adding overall visual dissimilarity to this analysis as fourth predictor revealed nearly identical category clusters, as would be expected based on the finding that a linear combination of outline and texture dissimilarity nearly perfectly captured overall visual dissimilarity (see Methods), thus making this variable redundant as additional predictor.

In summary, these results reveal distinct but overlapping representations for outline shape and texture in early visual cortex. Importantly, categorical representations in higher level visual cortex were still present even after regressing out both outline and texture dissimilarity.

## DISCUSSION

In this study, we asked whether the animate–inanimate organization of object responses in human VTC reflects characteristic visual properties of animate and inanimate

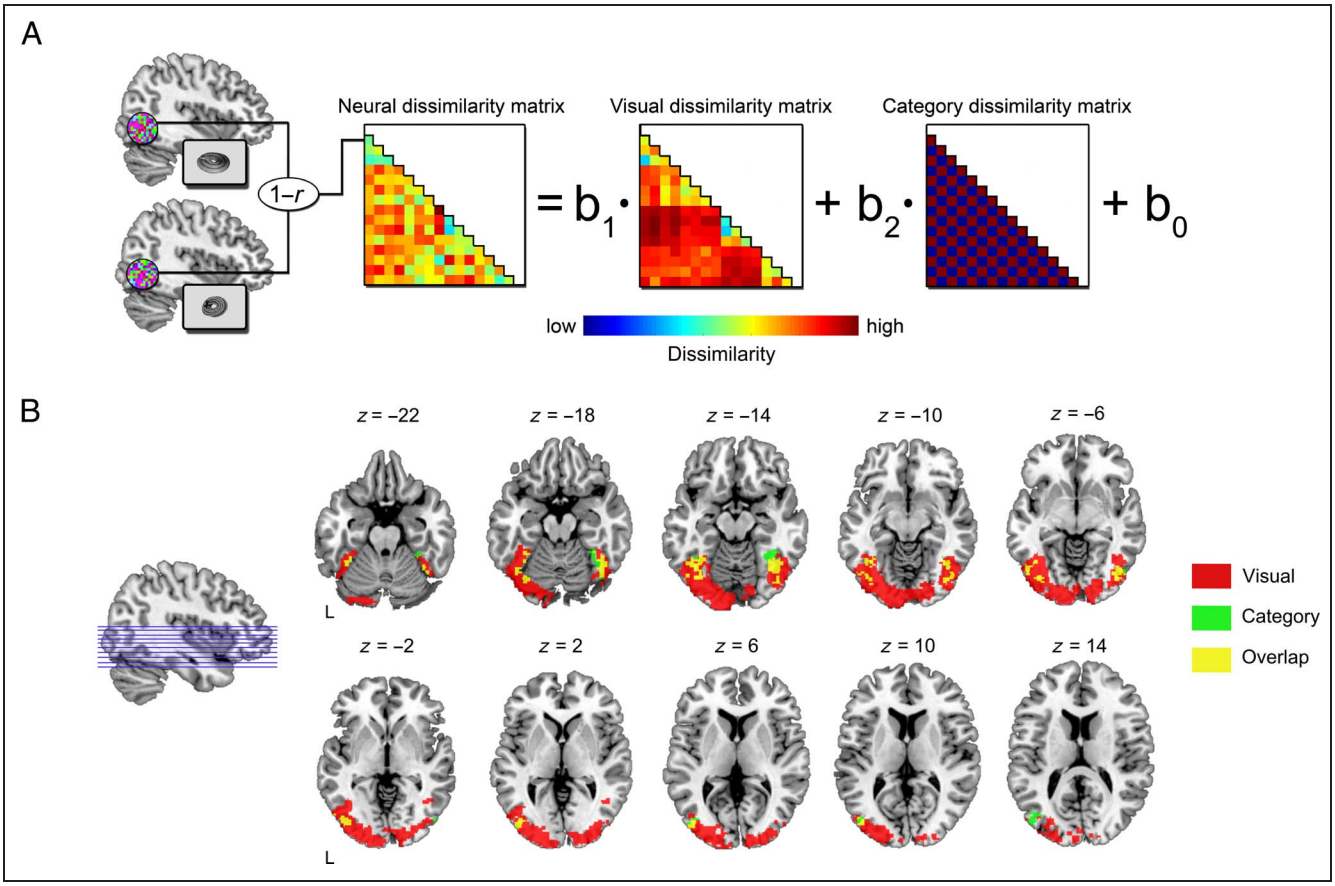



**Figure 5.** Multivariate searchlight analysis: overall visual dissimilarity and categorical dissimilarity. (A) Schematic of the linear modeling approach. For every spherical searchlight neighborhood, a 16 × 16 neural dissimilarity matrix was constructed using pairwise correlations of multivoxel activity patterns. These pairwise dissimilarity values ($1 - r$) were modeled by a linear combination of two predictors: One predictor was derived from the RTs in the visual search experiment that served as a proxy for overall visual dissimilarity of the stimuli (Figure 2A), and the other predictor reflected pairwise categorical (animate vs. inanimate) dissimilarity. This procedure tested the extent to which the neural dissimilarity structure in a given sphere reflected overall visual dissimilarity (while regressing out categorical dissimilarity) and/or categorical dissimilarity (while regressing out overall visual dissimilarity). (B) Results of whole-brain group analyses testing the value of each predictor versus zero. The analysis identified a large cluster of voxels where neural dissimilarity reflected overall visual dissimilarity (red and yellow voxels), spanning early visual cortex and extrastriate regions up to VTC. In addition, clusters were identified where neural dissimilarity was independently predicted by the category of the stimuli (yellow and green voxels): These clusters partly overlapped with the overall visual dissimilarity clusters (yellow voxels) and showed local peaks in right and left fusiform gyrus and left middle occipital gyrus. Statistical maps were thresholded at $p < .05$ (FWE corrected) and shown as binary color maps overlaid on a structural brain template (MRIcron).

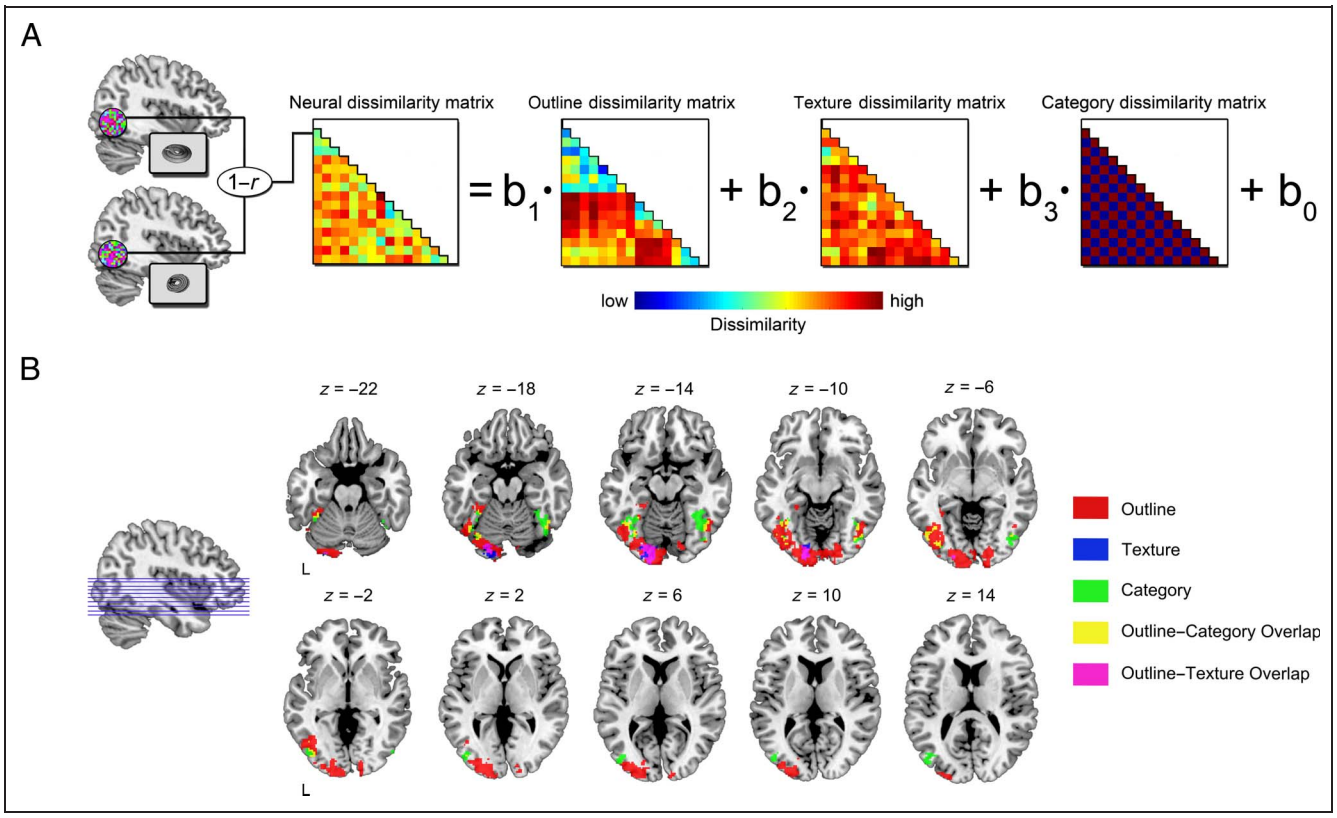



**Figure 6.** Multivariate searchlight analysis: outline, texture, and categorical dissimilarity. (A) Schematic of the linear modeling approach similar to the one shown in Figure 5 (see the caption for details). In this analysis, neural dissimilarity was modeled using the combination of three predictors: outline dissimilarity, texture dissimilarity, and categorical dissimilarity. (B) Results of whole-brain group analyses testing the value of each predictor versus zero. Clusters of voxels in which neural dissimilarity was predicted by outline dissimilarity (red, yellow, and pink voxels) were found in primary visual cortex, extending into extrastriate visual cortex. Texture dissimilarity predicted neural dissimilarity in a cluster in left early visual cortex (blue and pink voxels), partly overlapping with an outline dissimilarity cluster (overlap indicated in pink). Finally, this analysis revealed clusters of voxels in which neural dissimilarity was predicted by categorical dissimilarity (green and yellow voxels), with local peaks in right fusiform gyrus and left middle occipital gyrus. These clusters partly overlapped with outline dissimilarity clusters (overlap indicated in yellow). There was no overlap between texture dissimilarity clusters and categorical dissimilarity clusters. Statistical maps were thresholded at $p < .05$ (FWE corrected) and shown as binary color maps overlaid on a structural brain template (MRIcron).

objects (e.g., characteristic animal shapes or textures) or whether it (partly) reflects a true categorical organization. We approached this question by testing whether the animate–inanimate organization can still be observed when controlling for visual similarity of objects from animate and inanimate domains. A standard functional localizer experiment contrasting activity to a variety of animals with activity to a variety of inanimate objects replicated previous studies, showing animate- and inanimate-preferring regions in VTC and animate-preferring regions in lateral occipitotemporal cortex. Importantly, all of these regions, in both hemispheres, remained selective for their preferred category in the main experiment in which animate and inanimate objects were carefully matched for shape as well as for low-level features such as luminance and contrast. Results were consistent across all but one of the eight animate–inanimate pairs. (We speculate that the lack of preference for the snail condition may relate to snail shells frequently being experienced as inanimate objects, because they are often viewed without an animal inside.) These pairs varied widely in terms of their shape (e.g., snake vs. bird), further supporting the claim that specific shape properties (e.g., presence of limbs) do not fully account for the animate–inanimate organization in VTC. Finally, the inanimate objects also varied widely on various conceptual dimensions that have been linked to inanimate-preferring regions, such as real-world size (Konkle & Oliva, 2012) and manipulability (Mahon et al., 2007). The consistency of results across the pairs suggests that the animate–inanimate organization revealed here is not fully explained by such alternative conceptual properties.

The objects that were contrasted in the univariate analyses were matched for visual similarity by the experimenters. This approach is subjective and assumes that visual similarity can be accurately judged through visual inspection. An important additional aspect of our study was therefore the use of behavioral visual search tasks to quantify different aspects of visual similarity in a naive group of participants. On each trial, participants simply

indicated the location of the unique stimulus in an array of identical distractors; that is, there was no predefined target category. Visual differences are the only source of information to locate the target in this task, such that performance (RT) closely reflects the visual similarity of the target and distractor stimuli (Mohan & Arun, 2012). In the first experiment, we measured the visual similarity of the same images used in the fMRI experiment. These data potentially capture not only differences in outline shape but also any other visual property (e.g., texture, extent, spatial frequency) that helps to visually distinguish the target from the distractor, rendering it a measure of overall visual similarity. Moreover, in two additional experiments, we specifically measured outline similarity (using outline drawings) and texture similarity (using texture patches).

RSA with overall visual similarity and category similarity as predictors revealed that activity patterns throughout visual cortex reflected visual similarity, confirming that visual similarity is a dominant organizing principle of both low- and high-level visual cortex (e.g., Andrews et al., 2015). The overall visual similarity matrix derived from the visual search task was additionally used to regress out variance in the neural similarity matrices, testing whether any remaining variance can be attributed to categorical similarity. This analysis revealed clusters in VTC in which this was the case. Similar results were obtained when modeling outline and texture similarity separately as predictors of the neural similarity structure. Although both texture and outline shape where represented independently in visual cortex, category information was still present in some regions. In both analyses, these category clusters were found in the vicinity of animate- and inanimate-preferring regions.

Although our results provide evidence for an animate–inanimate organization of VTC that is not explained by outline shape or texture, we should consider the possibility that there may be remaining visual features that distinguished animals from objects. Clearly, in the absence of other cues, there must be visual properties that allow the observer to recognize the objects and to distinguish, for example, between a snake and a rope—we do not claim that there are no visual differences between the two objects of each pair. However, it seems unlikely that there were visual features that consistently covaried with category membership across pairs. Furthermore, such consistent features would likely be reflected in the visual similarity measures (Mohan & Arun, 2012) and thus regressed out in the representational similarity analyses. Nevertheless, we acknowledge that we cannot fully exclude that there may be residual visual differences between animals and inanimate objects that do not affect visual similarity as measured in the visual search experiments. For example, it is possible that certain category-specific shape features are not visually salient (and may not even be visible in the image) but become represented once an object is recognized as an animal (e.g., eyes).

Interestingly, clusters representing categorical similarity partly overlapped with clusters representing shape similarity at higher levels of the visual system (yellow clusters in Figures 5B and 6B). This suggests that a shape-based organization coexists with a category-based organization, with neither of these two reducible to the other. This coexistence suggests close mutual interactions between shape and category representations. In one direction, shape properties strongly inform category membership in most real-world situations. For example, the set of midlevel visual features that characterize animals allows for efficiently detecting the presence of an animal in a natural scene (Ullman, Vidal-Naquet, & Sali, 2002; Thorpe, Fize, & Marlot, 1996). In the other direction, category membership provides information about likely visual properties of an object, such as the structure of its parts, and allows for making perceptual predictions, for example, related to characteristic motion patterns of animals. The close proximity and partial overlap of shape and category representations may thus be optimal for real-world behavior in which these levels of representation need to closely interact.

Previous findings have shown that the degree to which a stimulus evokes an "animate" response in VTC depends on the degree to which it shares characteristics with the animate prototype—humans (Sha et al., 2014). This is consistent with earlier findings of strong selectivity for human faces and bodies at the approximate locations of the ventral and lateral animacy clusters in our study (Peelen & Downing, 2005; Downing et al., 2001; Kanwisher et al., 1997). These regions show a graded response profile, responding most strongly to human faces and bodies, followed by mammals, birds, reptiles, and insects (Downing et al., 2006). These findings are consistent, however, both with a visual similarity interpretation (i.e., differences in visual typicality; Mohan & Arun, 2012) and a conceptual similarity interpretation (e.g., differences in agency; Sha et al., 2014). Therefore, future work is needed to independently manipulate the degree to which animals share visual and conceptual properties with humans to test whether graded animacy effects are primarily reflecting one or both of these properties. Interestingly, our current results show that a reliable animal preference exists even for animals that are visually and conceptually distinct from humans (e.g., snakes, insects).

Together, the present results suggest that the animate–inanimate organization of VTC is not fully explained by local biases for visual features. Instead, we interpret this organization as reflecting the recognition of an object as belonging to a particular domain. In daily life, visual properties are an important cue for categorizing objects, but many other cues also contribute. These cues include information from other modalities (e.g., audition, touch) and, more generally, our expectations, knowledge, goals, and beliefs. Rather than following the visual features falling on the retina, category-selective activity in VTC appears to partly reflect the interpretation,

based on all available cues, that the object we look at is animate or inanimate. On this account, category-specific activity that is independent of visual features would reflect a relatively late stage in the object recognition process. Future work could use multivariate analysis of magnetoencephalography data (e.g., Cichy, Pantazis, & Oliva, 2014; Carlson, Tovar, Alink, & Kriegeskorte, 2013) to reveal the temporal dynamics of object categorization using carefully designed stimuli that allow for disentangling visual and categorical similarity. One prediction consistent with our results would be that the initial response in VTC primarily reflects visual similarity, with later stages additionally reflecting category membership.

If not visual features, then what property might drive the animate–inanimate distinction? One proposal is that this distinction reflects agency: the potential of an object to perform self-initiated, complex, goal-directed actions (Sha et al., 2014; Caramazza & Shelton, 1998; Premack, 1990). For example, studies have shown that activity in the right fusiform gyrus—at the approximate location of the ventral animate-preferring region—can be evoked by simple geometric shapes that, through their movements, are interpreted as social agents (Gobbini, Koralek, Bryan, Montgomery, & Haxby, 2007; Martin & Weisberg, 2003; Schultz et al., 2003; Castelli, Happe, Frith, & Frith, 2000). Other work consistent with this account has shown that animal selectivity in VTC is strongest for animals, such as mammals, that are perceived as having relatively more agentic properties (Sha et al., 2014).

In summary, the present results suggest that the animate–inanimate organization of VTC may not fully reflect visual properties that characterize animals and objects. Results from RSA indicate that visual and categorical representations coexist in more anterior parts of the visual system. Clearly, future work is needed to further exclude the possibility of confounding visual features, to define exactly what dimensions drive the animate–inanimate distinction, to reveal the time course of visual and categorical representations, and to test how these interact to allow for efficient object categorization in our daily life environments.

## Acknowledgments

We thank Nick Oosterhof for help with data analysis. The research was funded by the Autonomous Province of Trento, Call "Grandi Progetti 2012," project "Characterizing and improving brain mechanisms of attention—ATTEND."

Reprint requests should be sent to Marius V. Peelen, Center for Mind/Brain Sciences, University of Trento, Corso Bettini 31, 38068 Rovereto (TN), Italy, or via e-mail: marius.peelen@unitn.it.

## REFERENCES

Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., et al. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nature Neuroscience, 10,* 687–689.

Amedi, A., von Kriegstein, K., van Atteveldt, N. M., Beauchamp, M. S., & Naumer, M. J. (2005). Functional imaging of human crossmodal identification and object recognition. *Experimental Brain Research, 166,* 559–571.

Andrews, T. J., Watson, D. M., Rice, G. E., & Hartley, T. (2015). Low-level properties of natural images predict topographic patterns of neural response in the ventral visual pathway. *Journal of Vision, 15,* 3.

Baldassi, C., Alemi-Neissi, A., Pagan, M., Dicarlo, J. J., Zecchina, R., & Zoccolan, D. (2013). Shape similarity, better than semantic membership, accounts for the structure of visual object representations in a population of monkey inferotemporal neurons. *PLoS Computational Biology, 9,* e1003167.

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,* 18247–18258.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436.

Buchel, C., Price, C., & Friston, K. (1998). A multimodal language region in the ventral visual pathway. *Nature, 394,* 274–277.

Caldara, R., Seghier, M. L., Rossion, B., Lazeyras, F., Michel, C., & Hauert, C. A. (2006). The fusiform face area is tuned for curvilinear patterns with more high-contrasted elements in the upper part. *Neuroimage, 31,* 313–319.

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain the animate-inanimate distinction. *Journal of Cognitive Neuroscience, 10,* 1–34.

Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision, 13,* pii: 1.

Castelli, F., Happe, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage, 12,* 314–325.

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,* 913–919.

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience, 17,* 455–462.

Cohen, L., & Dehaene, S. (2004). Specialization within the ventral stream: The case for the visual word form area. *Neuroimage, 22,* 466–476.

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

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

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392,* 598–601.

Gobbini, M. I., Koralek, A. C., Bryan, R. E., Montgomery, K. J., & Haxby, J. V. (2007). Two takes on the social brain: A comparison of theory of mind tasks. *Journal of Cognitive Neuroscience, 19,* 1803–1814.

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

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293,* 2425–2430.

Haxby, J. V., Ishai, I. I., Chao, L. L., Ungerleider, L. G., & Martin, I. I. (2000). Object-form topology in the ventral temporal

lobe. Response to I. Gauthier (2000). *Trends in Cognitive Sciences, 4,* 3–4.
He, C., Peelen, M. V., Han, Z., Lin, N., Caramazza, A., & Bi, Y. (2013). Selectivity for large nonmanipulable objects in scene-selective visual cortex does not require visual experience. *Neuroimage, 79,* 1–9.
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, U.S.A., 96,* 9379–9384.
Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302–4311.
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,* 4296–4309.
Konkle, T., & Oliva, A. (2012). A real-world size organization of object responses in occipitotemporal cortex. *Neuron, 74,* 1114–1124.
Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences, U.S.A., 103,* 3863–3868.
Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience, 2,* 4.
Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., et al. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron, 60,* 1126–1141.
Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-specific organization in the human brain does not require visual experience. *Neuron, 63,* 397–405.
Mahon, B. Z., & Caramazza, A. (2011). What drives the organization of object knowledge in the brain? *Trends in Cognitive Sciences, 15,* 97–103.
Mahon, B. Z., Milleville, S. C., Negri, G. A., Rumiati, R. I., Caramazza, A., & Martin, A. (2007). Action-related properties shape object representations in the ventral stream. *Neuron, 55,* 507–520.
Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology, 58,* 25–45.
Martin, A., & Weisberg, J. (2003). Neural foundations for understanding social and mechanical concepts. *Cognitive Neuropsychology, 20,* 575–587.
Mohan, K., & Arun, S. P. (2012). Similarity relations in visual search predict rapid visual categorization. *Journal of Vision, 12,* 19.
Nasr, S., Echavarria, C. E., & Tootell, R. B. (2014). Thinking outside the box: Rectilinear shapes selectively activate scene-selective cortex. *Journal of Neuroscience, 34,* 6721–6735.
Op de Beeck, H. P., Haushofer, J., & Kanwisher, N. G. (2008). Interpreting fMRI data: Maps, modules and dimensions. *Nature Reviews Neuroscience, 9,* 123–135.
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,* 1225–1234.
Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *Journal of Neurophysiology, 93,* 603–608.
Peelen, M. V., He, C., Han, Z., Caramazza, A., & Bi, Y. (2014). Nonvisual and visual object shape representations in occipitotemporal cortex: Evidence from congenitally blind and sighted adults. *Journal of Neuroscience, 34,* 163–170.
Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition, 36,* 1–16.
Reich, L., Szwed, M., Cohen, L., & Amedi, A. (2011). A ventral visual stream reading center independent of visual experience. *Current Biology, 21,* 363–368.
Ricciardi, E., Bonino, D., Pellegrini, S., & Pietrini, P. (2014). Mind the blind brain to understand the sighted one! Is there a supramodal cortical functional architecture? *Neuroscience & Biobehavioral Reviews, 41,* 64–77.
Schultz, R. T., Grelotti, D. J., Klin, A., Kleinman, J., Van der Gaag, C., Marois, R., et al. (2003). The role of the fusiform face area in social cognition: Implications for the pathobiology of autism. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 358,* 415–427.
Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O., et al. (2014). The animacy continuum in the human ventral vision pathway. *Journal of Cognitive Neuroscience, 27,* 665–678.
Striem-Amit, E., & Amedi, A. (2014). Visual cortex extrastriate body-selective area activation in congenitally blind people "seeing" by using sounds. *Current Biology, 24,* 687–692.
Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381,* 520–522.
Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience, 5,* 682–687.
Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods, 42,* 671–684.
Wolbers, T., Klatzky, R. L., Loomis, J. M., Wutte, M. G., & Giudice, N. A. (2011). Modality-independent coding of spatial layout in the human brain. *Current Biology, 21,* 984–989.