Tracking the Spatiotemporal Neural Dynamics of Real-world Object Size and Animacy in the Human Brain

Seyed-Mahdi Khaligh-Razavi[1], Radoslaw Martin Cichy[2], Dimitrios Pantazis[1], and Aude Oliva[1]

Abstract

■ Animacy and real-world size are properties that describe any object and thus bring basic order into our perception of the visual world. Here, we investigated how the human brain processes real-world size and animacy. For this, we applied representational similarity to fMRI and MEG data to yield a view of brain activity with high spatial and temporal resolutions, respectively. Analysis of fMRI data revealed that a distributed and partly overlapping set of cortical regions extending from occipital to ventral and medial temporal cortex represented animacy and real-world size. Within this set, parahippocampal cortex stood out as the region representing animacy and size stronger than most other regions. Further analysis of the detailed representational format revealed differences among regions involved in processing animacy. Analysis of MEG data revealed overlapping temporal dynamics of animacy and real-world size processing starting at around 150 msec and provided the first neuromagnetic signature of real-world object size processing. Finally, to investigate the neural dynamics of size and animacy processing simultaneously in space and time, we combined MEG and fMRI with a novel extension of MEG–fMRI fusion by representational similarity. This analysis revealed partly overlapping and distributed spatiotemporal dynamics, with parahippocampal cortex singled out as a region that represented size and animacy persistently when other regions did not. Furthermore, the analysis highlighted the role of early visual cortex in representing real-world size. A control analysis revealed that the neural dynamics of processing animacy and size were distinct from the neural dynamics of processing low-level visual features. Together, our results provide a detailed spatiotemporal view of animacy and size processing in the human brain. ■

INTRODUCTION

Object properties that are universally applicable to any object are of primary importance to our understanding of the visual world by providing basic ordering and categorization to our perceptions. Animacy and real-world size are two such universal properties that are pertinent to perception and of high behavioral relevance (Konkle & Oliva, 2011; Chao, Haxby, & Martin, 1999). Every object has a physical size, from small scales, such as a strawberry we can pick up with our fingers, to large scales, such as a building that we can use as a landmark (Konkle & Oliva, 2011, 2012; Mullally & Maguire, 2011). Similarly, every object is either animate or inanimate (Shutts, Markson, & Spelke, 2009). Animacy and real-world size are independent properties of each other, in that animate objects can range from small to large and so can inanimate objects.

Previous research has shown that animacy and object size processing is carried out in a distributed and partly overlapping network of brain regions in the occipital, temporal, and parietal lobes (animacy: Sha et al., 2015; Bell, Hadj-Bouziane, Frihauf, Tootell, & Ungerleider, 2009; Mahon, Anzellotti, Schwarzbach, Zampini, & Caramazza, 2009; Wiggett, Pritchard, & Downing, 2009; Kriegeskorte et al., 2008; Downing, Chan, Peelen, Dodds, & Kanwisher, 2006; Chao et al., 1999; size: Konkle & Caramazza, 2013, 2017; Fabbri, Stubbs, Cusack, & Culham, 2016; Bainbridge & Oliva, 2015; Konkle & Oliva, 2012; Murray, Boyaci, & Kersten, 2006). Some previous studies have also looked at the temporal dynamics of animacy processing (Clarke, Devereux, Randall, & Tyler, 2015; Carlson, Simmons, Kriegeskorte, & Slevc, 2014; Cichy, Pantazis, & Oliva, 2014; Carlson, Tovar, Alink, & Kriegeskorte, 2013; Liu, Agam, Madsen, & Kreiman, 2009; Thorpe, Fize, & Marlot, 1996), showing that the information related to animacy is processed rapidly within the first 100 msec after stimulus onset.

However, crucial questions about the spatial and temporal neural dynamics underlying animacy and real-world size processing remain unanswered. Concerning the spatial dynamics, it remains unknown how in detail animacy and size information are represented. For example, do all regions that process animacy use the same representational format, or does the representational format differ across regions? Similarly, it remains unknown how animacy and size information interact and are integrated

[1]Massachusetts Institute of Technology, [2]Free University Berlin

across the overlapping set of regions: Do all regions process animacy and size in parallel in the same fashion, or does a single region stand out? Concerning the temporal dynamics, it remains unknown how animacy and real-world size processing is orchestrated in time. Given that the properties real-world size and animacy are independent of each other, one might expect them to be processed independently in the brain, engaging different cortical sites at different moments. Alternatively, given the universality and importance of those two properties, the brain might use a distributed code where real-world size and animacy are processed in parallel and redundantly in the same cortical sites and with similar temporal dynamics.

To reveal the processing of animacy and real-world size in the human brain, we recorded brain data with fMRI and MEG while participants viewed a set of object images that differed in animacy and real-world size independently. To reveal the spatial dynamics, we applied representational similarity analysis (RSA; Nili et al., 2014; Kriegeskorte & Kievit, 2013; Kriegeskorte, Mur, & Bandettini, 2008; Kriegeskorte, Mur, Ruff, et al., 2008) to fMRI data to investigate the loci of animacy and size processing. To reveal the temporal dynamics, we applied multivariate pattern classification and RSA to MEG data in a time-resolved fashion (Cichy et al., 2014; Carlson et al., 2013). Finally, we combined MEG and fMRI data using a novel extension of MEG–fMRI fusion by representational similarity (Cichy et al., 2016; Cichy et al., 2014) to reveal the neural dynamics specific to the processing of animacy and size resolved simultaneously in space and time.

METHODS

Experimental Design

The MEG and fMRI data have been fully described in a previous publication (Cichy et al., 2016) and are available upon request. Here, we repeat the pieces of information that are necessary for reproduction.

Participants

Fifteen healthy volunteers (five women; age: mean = 26.60 years, SD = 5.18 years) participated in this experiment. All participants were right-handed with normal or corrected-to-normal vision and provided written consent. The studies were conducted in accordance with the Declaration of Helsinki and approved by the local ethics committee (institutional review board of the Massachusetts Institute of Technology).

Stimulus Set and Stimulus Presentation Parameters

The stimulus set contained 118 real-world object images on real backgrounds. The full image set is available at brainmodels.csail.mit.edu/images/stimulus_set.png. The image set contained 27 animate objects (nine small, three medium, and 15 large) and 91 inanimate objects (33 small, 32 medium, and 26 large). For examples, see Figure 1A. For both MEG and fMRI recordings, images were presented at the center of a screen at 4.0° visual angle overlaid with a gray fixation cross. The presentation duration was 500 msec per image.

fMRI Data Acquisition and Preprocessing

We used the fMRI data as reported in Cichy et al. (2016), Experiment 2. The experiment consisted of two sessions of 9–11 runs of 486 sec in duration each. In each run, every image was presented once. Image order was randomized with the restriction that the same condition was not presented on consecutive trials. Twenty-five percent of the trials were null trials during which only a gray background was presented and the fixation cross changed luminance for 100 msec. Participants were instructed to respond to the change in luminance with a button press. SOA was 3 sec, or 6 sec with a preceding null trial.

MRI data were acquired on a 3-T Trio scanner (Siemens) with a 32-channel head coil. Structural images were obtained using a standard T1-weighted sequence (192 sagittal slices, field of view [FOV] = 256 mm^2, repetition time = 1900 msec, echo time = 2.52 msec, flip angle = 9°). Functional data were obtained with the following parameters: gradient-echo EPI sequence: repetition time = 750 msec, echo time = 30 msec, flip angle = 61°, FOV read = 192 mm, FOV phase = 100% with a partial fraction of 6/8, through-plane acceleration factor of 3, bandwidth = 1816 Hz/Px, resolution = 3 mm^3, slice gap = 20%, slices = 33, and ascending acquisition.

The fMRI data were preprocessed using SPM8 (www.fil.ion.ucl.ac.uk/spm/). For each participant, fMRI data from both sessions were realigned and coregistered to the T1 structural scan acquired in the first MRI session. Then, MRI data were normalized to the standard Montreal Neurological Institute template. A general linear model (GLM) was used to estimate the fMRI response to the 118 image conditions. Image onsets and durations entered the GLM as regressors and were convolved with a hemodynamic response function. Movement parameters were included as nuisance parameters. Additional regressors modeling the two sessions were included in the GLM. The estimated condition-specific GLM parameters were converted into t values by contrasting each condition estimate against the implicitly modeled baseline. The t values were used in a later analysis for constructing fMRI dissimilarity matrices.

MEG Data Acquisition and Preprocessing

Participants completed one session of 15 runs of 314 sec in duration each. In each run, every image was shown

twice, and the sequence of image presentations was randomized. Trial onset asynchrony varied between 0.9 and 1 sec. Every three to five trials, an image of a paper clip was shown to which participants were asked to respond with an eye blink and a button press. Participants were instructed to not blink at other times.

The MEG signals were acquired from 306 channels (204 planar gradiometers, 102 magnetometers; Elekta Neuromag TRIUX, Elekta) at a sampling rate of 1 kHz, filtered between 0.03 and 330 Hz.

To denoise data, temporal source space separation (maxfilter software, Elekta; Taulu & Simola, 2006; Taulu, Kajola, & Simola, 2004) was applied before further analyzing data with Brainstorm (Tadel, Baillet, Mosher, Pantazis, & Leahy, 2011). In detail, for each trial, we extracted peristimulus data from −100 to +700 msec, removed the prestimulus baseline mean from the signals, and smoothed data with a 20-msec sliding window.

Definition of fMRI ROIs

We parcellated the whole cortex into 60 nonoverlapping ROIs in each hemisphere. ROIs were defined through using two probabilistic atlases of the human brain (Wang, Mruczek, Arcaro, & Kastner, 2015; Tzourio-Mazoyer et al., 2002). First, we used the retinotopically defined ROIs from the Wang et al. (2015) probabilistic atlas. As this atlas does not cover the whole cortex, for the remaining cortex, we used the anatomically defined ROIs of the automated anatomical labeling toolbox (Tzourio-Mazoyer et al., 2002). To create nonoverlapping ROIs from probabilistic maps, we assigned each voxel to the ROI of highest probability; the aggregate probability over all ROIs was ≥33%.

All analyses and statistics were conducted for all ROIs, but for ease of visualization, the results are shown only for those ROIs that had any significant effect after accounting for multiple comparisons across all tests. From the ROIs discussed in the article, the following ROIs were defined using the Wang et al. (2015) atlas: early visual cortex (EVC), lateral occipital cortex (LO), temporal occipital cortex (TO), ventral occipital cortex (VO), parahippocampal cortex (PHC), intraparietal sulcus (IPS), and superior parietal lobe (SPL). EVC was defined as the combination of V1, V2, and V3 masks. We combined these three regions because visual stimuli were presented at 4° visual angle, which is an eccentricity at which the distinction between V1, V2, and V3 is difficult as they are located close together in the foveal confluence. IPS and SPL corresponded to IPS0 and SPL1 in the nomenclature of Wang et al.'s (2015) atlas. LO, TO, VO, and PHC were combinations of masks LO1 and LO2, TO1 and TO2, VO1 and VO2, and PHC1 and PHC2, respectively. Fusiform gyrus (Fusi), inferior temporal gyrus (ITG), and middle temporal (MT) were defined using the automated anatomical labeling atlas (Tzourio-Mazoyer et al., 2002).

RSA

RSA enables relating representations obtained from different modalities such as computational models, MEG, and fMRI data (Kietzmann, Gert, Tong, & König, 2017; Cichy, Khosla, Pantazis, Torralba, & Oliva, 2016; Cichy, Pantazis, & Oliva, 2016; Cichy et al., 2014; Khaligh-Razavi & Kriegeskorte, 2014; Nili et al., 2014; Kriegeskorte & Kievit, 2013; Kriegeskorte, Mur, & Bandettini, 2008; Kriegeskorte, Mur, Ruff, et al., 2008). The basic analytical tool of RSA is the representational dissimilarity matrix (RDM). An RDM is a square symmetric matrix, defined in rows and columns by experimental conditions, in which off-diagonal elements indicate the dissimilarity between the activation patterns associated with two different conditions. The diagonal of the RDM reflects comparisons between identical conditions and is thus set to 0 and excluded from the analysis. An RDM thus summarizes the representational geometry in terms of dissimilarity between condition-specific activation patterns.

Definition of RDMs

MEG RDMs. The analysis was conducted separately for each participant. For each time point t of the peristimulus epoch time, we compared pairwise the dissimilarity of MEG data for all 118 conditions (images) to construct a 118×118 RDM. The measure of dissimilarity was cross-validated decoding accuracy (as in Cichy et al., 2016; Cichy et al., 2014) between the 306-channel MEG response patterns.

fMRI RDMs. For each participant and each brain area, the fMRI RDMs were constructed by computing pairwise dissimilarities (1 minus Pearson's correlation) between condition-specific voxel t value patterns.

Animacy model RDM. The animacy model RDM is binary with entry 1 if two conditions differed in animacy and 0 if both conditions were both animate or inanimate.

Size model RDM. The size model RDM was constructed based on behavioral ratings of size judgment. In detail, we asked 10 participants to arrange the 118 objects into three real-world size categories: (a) small: finger size or one-hand size; (b) medium: size of two hands; and (c) large: body size or larger. Each image was then assigned to the size category voted by most participants. We used these size categories to construct the size model RDM. RDM elements corresponding to comparisons of images of the same size were assigned the entry 0; RDM elements corresponding to comparisons of images of different sizes were assigned the entry 2 if they were large versus small and 1 if otherwise.

Gist model RDM. To evaluate the low visual feature content of the images, we computed an RDM based on

the Gist descriptor (Oliva & Torralba, 2001, 2006). The Gist code is available at people.csail.mit.edu/torralba/code/spatialenvelope/. In detail, we computed image-specific Gist features by subdividing the image into 16 bins, and then for each bin, we determined the filter energy of Gabor filters in eight orientations over four different spatial scales. We then compared (Spearman's r) image-specific Gist features pairwise to construct a Gist model RDM.

Standard RSAs

We conducted three standard RSAs differing in the modalities compared: (i) MEG-to-category model, (ii) fMRI-to-category model, and (iii) MEG-to-fMRI.

MEG-to-category model RSA. We correlated participant-specific MEG RDMs at each time point to the reference category model RDM (size or animacy model RDM). This yielded an MEG-to-category model similarity time course for each participant and category model.

fMRI-to-category model RSA. We correlated participant-specific fMRI RDMs of each ROI to the reference category model RDMs. This yielded an fMRI-to-category model similarity value for each participant, category model, and ROI.

MEG-to-fMRI RSA. Also termed MEG–fMRI fusion (Cichy, Pantazis, et al., 2016), for this analysis, we correlated participant-specific fMRI RDMs for each ROI with the participant-averaged MEG RDM at each time point. This yielded an MEG-to-fMRI similarity map for each participant and time point.

Noise ceiling. To account for the fact that different brain areas have different levels of noise, we estimated the noise ceiling for each brain area (Nili et al., 2014) and the fMRI-to-category model RSA results for each ROI were normalized with the corresponding noise ceiling. The noise ceiling is defined as the mean correlation of each participant's RDM with the participant-averaged RDM, where the average RDM can be thought of as an estimate of the true model's RDM. The noise ceiling indicates the expected correlation of the true model given the noise in the data.

Onset and peak latency. We calculated the onset and peak latencies for participant-averaged results of the MEG-to-category model and MEG-to-fMRI RSA. The peak latencies and standard error are obtained by 10,000 bootstrap resampling of participants.

Statistical testing. For each of these three standard RSAs, namely, MEG-to-model, fMRI-to-model, and MEG-to-fMRI, statistical significance was obtained by a random effects analysis over the participant-specific results. In detail, following the statistical procedure in Nili et al. (2014), we used the Wilcoxon signed-rank test for assessing statistical significance of RDM correlations. This test is nonparametric (Gibbons & Chakraborti, 2011; Hollander & Wolfe, 1999) and thus does not make assumptions about the distribution of the data. When comparing peaks and onset latencies, we used two-sided bootstrap resampling test. In other cases that involved testing if a correlation is significantly above zero, we conducted one-sided signed-rank test with $p < .05$. We used false discovery rate (FDR; Simmons, Nelson, & Simonsohn, 2011; Benjamini & Hochberg, 1995) to correct for multiple comparisons: fMRI-to-category model RSA results were corrected across brain regions, MEG-to-category model RSA was corrected across time points, and MEG-to-fMRI RSA was corrected across both brain regions and time points.

Content-dependent MEG–fMRI Fusion Algorithm

Standard MEG–fMRI fusion reveals the whole spatiotemporal neural dynamics after visual stimulation. To further characterize spatiotemporal neural dynamics specific to the processing of a particular type of information, such as animacy or real-world size (referred to here as "content"), additional algorithmic constraints are required. Toward this aim, we used a statistical procedure called "conjunction inference" (Nichols, Brett, Andersson, Wager, & Poline, 2005) to combine the results of the three standard RSA procedures explained above, namely, (i) MEG-to-fMRI, (ii) MEG-to-category model RSA, and (iii) fMRI-to-category model RSA. In particular, statistical inference as described above yielded significant effects for each procedure alone. To reveal the spatiotemporal dynamics specific to a particular content, we then combined the results of the three procedures with an 'AND' operation. The resulting combined statistical map indicates MEG time points linked with fMRI locations that have a similar representational geometry to each other and to the reference model RDM. Visualizing the obtained 3-D time–location–content statistical map at each millisecond results in a movie of MEG–fMRI correspondence with respect to a prespecified content (e.g., size or animacy), thus revealing the spatiotemporal neural dynamics underlying the processing of the particular content.

RESULTS

General Experimental Rationale

We present the overall experimental rationale to investigate the spatiotemporal neural dynamics of processing the object properties animacy and real-world size in Figure 1. The analyses were based on fMRI and MEG data collected by Cichy, Pantazis, et al. (2016) on 118 different natural images of everyday objects. The object set comprised both animate and inanimate objects in either of
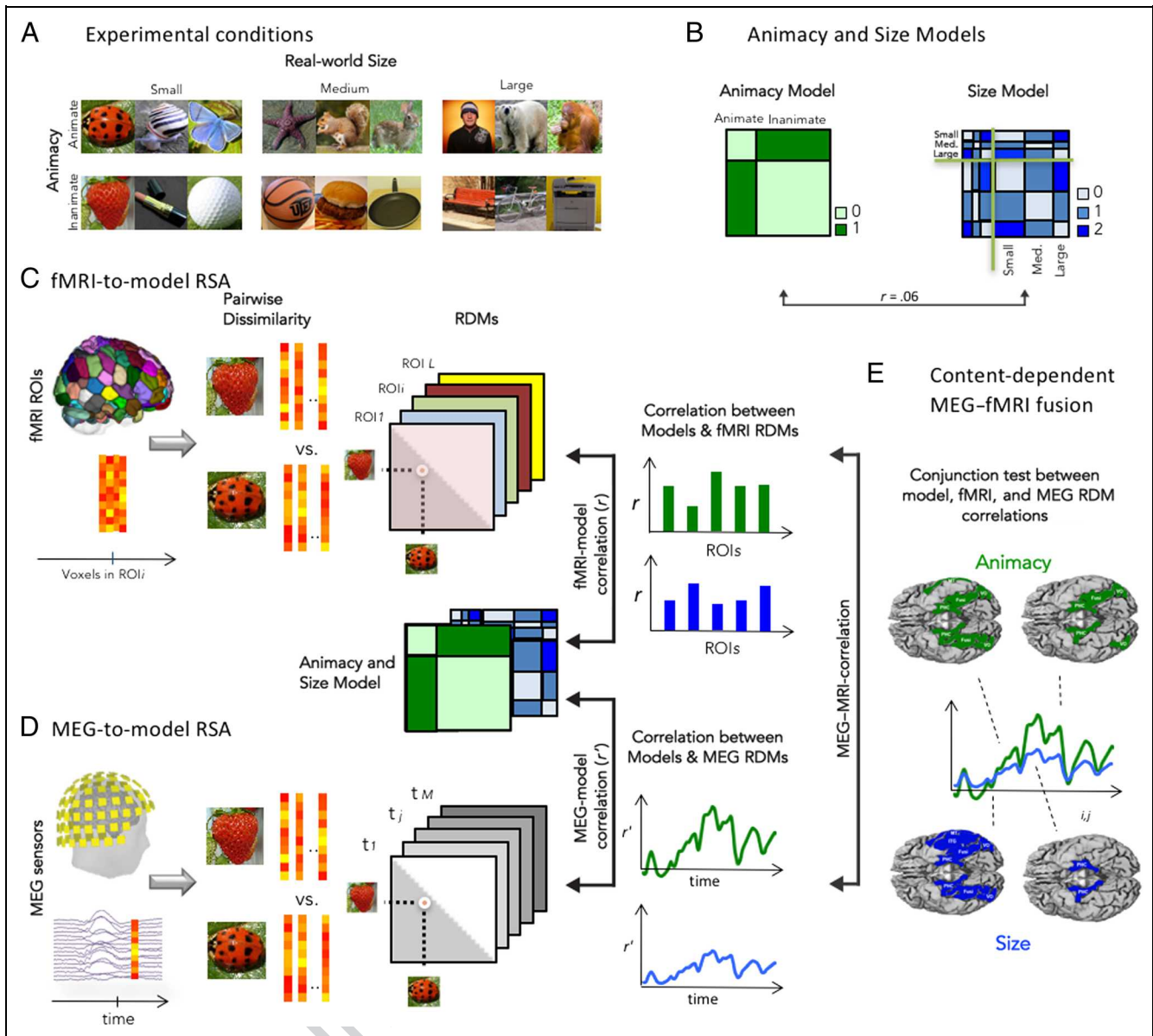
Figure 1. (A) Experimental conditions. MEG and fMRI data were presented while participants viewed 118 object images of real-world objects. The objects were either animate or inanimate and of three sizes: small, medium, and large. Representative examples are shown here (see Supplementary Figure 2 for all images). (B) On the basis of objects' animacy and physical size, we constructed two 118×118 model RDMs. The animacy model RDM identifies a representational space in which objects are clustered based on their animacy. The size model RDM identifies a representational space in which objects are clustered based on their physical size. For illustrative purposes, here, we show cartoony versions of the model RDMs. Here and in further panels, the animacy model is color-coded green and the size model is color-coded blue. There is a small correlation of $r = .06$ between these two model RDMs. (C) fMRI-to-model RSA. We parcellated each hemisphere of cortex into $L = 60$ nonoverlapping ROIs. For each region i (from 1 to L), we correlated condition-specific fMRI activation patterns for all condition pairs, yielding a 118×118 fMRI RDM for each ROI. We then correlated region-specific RDMs with the animacy and size model RDMs. This yielded one value of representational similarity for each ROI and model. (D) MEG-to-model RSA. For each time point t (from -100 to 700 msec in 1-msec steps), we correlated condition-specific MEG sensor activation patterns for all conditions, yielding a 118×118 MEG RDM for each time point. We then correlated time-specific RDMs with the animacy and size model RDMs. This yielded a time course of representational similarity for each model. (E) Content-dependent MEG–fMRI fusion. To reveal the spatiotemporal dynamics during object vision, we conducted MEG-to-fMRI RSA by correlating all region-specific RDMs with all time-point-specific RDMs. To identify the aspects of neural activity specific to the processing of either animacy or size, we conducted a conjunction test on significant results from MEG-to-model RSA, fMRI-to-model RSA, and MEG-to-fMRI RSA for the animacy and size model separately. This yielded a spatiotemporally resolved picture of brain activity for animacy and object size processing. Med. = medium.

three sizes, namely, small, medium, and large. Figure 1A displays examples of stimuli in each subgroup. During the study, observers watched the object images presented for 0.5 sec each while their brain activity was recorded with fMRI and MEG in separate sessions, to resolve neuronal activity in both high spatial and temporal resolutions, respectively. Participants performed a task orthogonal to the experimental conditions presented and

were neither told to attend to specific properties of the objects nor informed about the hypotheses of the study. This ensures that the brain responses measured are not specific to a particular task but reflect task-independent and automatic processing of these properties.

To model the effect of size and animacy on brain representations independently of each other, we constructed two RDMs (Figure 1B). These RDMs are models of the assumption that dissimilarity is higher for objects across category boundary than within category boundary.

To reveal the spatiotemporal neural dynamics of real-world size and animacy processing, we used RSA in three different ways. First, to identify the brain regions involved in processing animacy and object size, we correlated the model RDMs with 60 fMRI RDMs, one for each of the 60 nonoverlapping regions in each hemisphere (Figure 1C). This analysis yielded one value of representational similarity between each model and each region. Second, to identify when in time during perception animacy and object size are processed, we correlated model RDMs with time-resolved MEG RDMs built from sensor activation patterns in 306 MEG sensors (Figure 1D). This analysis yielded one value of representational similarity between each model and each time point. Third, for a view on brain activity resolved simultaneously in space and time, we conducted MEG-to-fMRI RSA (termed "MEG–fMRI fusion"; Cichy et al., 2015) by correlating all region-specific RDMs with all time-point-specific RDMs (Figure 1E). To identify the aspects of neural activity specific to the processing of animacy and size, we conducted a conjunction test on significant results from MEG-to-model RSA, fMRI-to-model RSA, and MEG-to-fMRI RSA for the animacy and size model separately. This content-dependent fusion approach yielded a spatiotemporally resolved picture of brain activity for animacy and object size processing.

A Distributed and Partly Overlapping Network of Brain Regions Represents Animacy and Real-world Object Size

To identify the network of brain regions representing animacy and object size, we conducted fMRI-to-model RSA (Figure 1C). Figure 2 shows the brain regions that were significantly correlated with at least one of the model RDMs (animacy: Figure 2A; physical size: Figure 2B; $N = 15$, one-sided signed-rank test, with results FDR-corrected across all brain parcels at $p < .05$).

We found a network of brain areas primarily located in the ventral temporal cortex, with the set of brain regions representing animacy being a subset of the regions representing object size. In detail, animacy was encoded in LO, TO, VO, PHC, Fusi, ITG, MT, and the left IPS. Size was encoded in all those regions as well and, in addition, in the left EVC and right SPL.

Within the identified network, PHC stood out through a combination of characteristics that were unique. First,
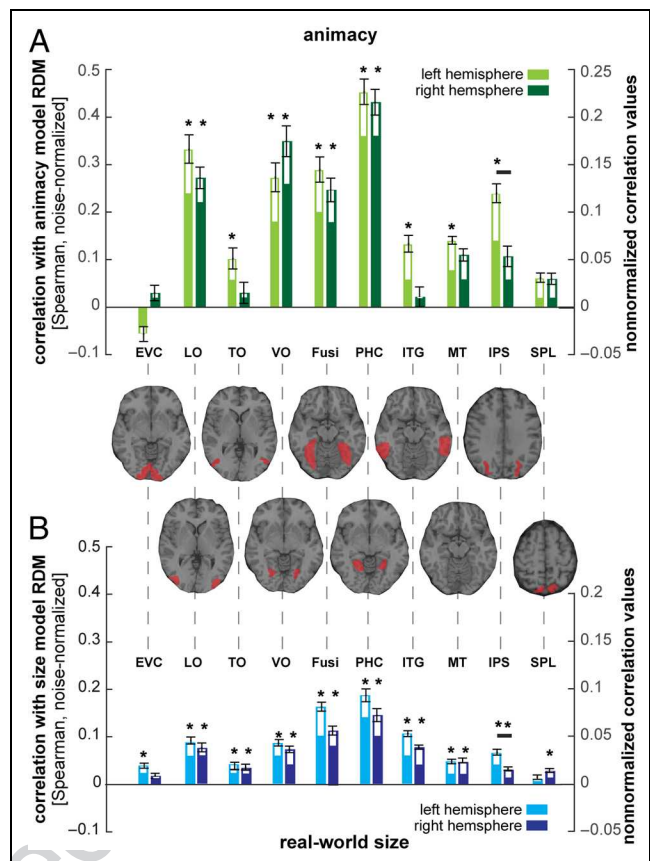

Figure 2. Large-scale network of brain regions involved in processing (A) animacy and (B) real-world object size. Bars show correlations between model RDMs with fMRI RDMs for 10 brain ROIs (out of 60) that had a significant correlation with either the animacy or real-world size model RDM. Error bars indicate *SEM*s. Asterisks indicate significant RDM correlations ($N = 15$, one-sided signed-rank test, $p < .05$ FDR-corrected across the number of ROIs in both hemispheres, i.e., 60 ROIs in two hemispheres = 120). The black horizontal line on top of IPS shows a significant difference between left- and right-hemisphere correlations with the model RDMs—no other region had a significant difference ($N = 15$, two-sided signed-rank test, FDR-corrected at $p < .05$). Given that different ROIs have different noise ceilings, we normalized RDM correlations for each ROI with respect to its noise ceiling (defined as in Nili et al., 2014) so that the correlations between ROIs become more readily comparable. The hollow part of the bars shows noise-normalized correlation values (axis on the left). The nonnormalized correlation values are indicated with the solid bars (axis on the right). Acronyms used in the figure for brain ROIs are as follows: EVC, LO, TO, VO, Fusi, PHC, IPS, and SPL.

PHC in both hemispheres encoded both animacy and object size. Second, the correlation with the size model was higher that with most other regions (for all regions in both hemispheres except left LO and Fusi where the difference was not statistically significant, $N = 15$, two-sided signed-rank test, FDR-corrected at $p < .05$ across all pairs of ROIs). Third, the correlation also with the animacy model was higher than with most other regions (except left and right VO).

To evaluate the robustness of the results, we investigated the processing of size and animacy independent

of each other by repeating the fMRI-to-model RSA within each level of the factors size and animacy separately. In detail, we conducted fMRI-to-size model RSA twice separately for animate and inanimate objects (Figure 3C–E) and fMRI-to-animacy model RSA thrice separately for each size level (Figure 3A and B). In all cases, we found ROIs with a significant correlation with the reference model RDM. However, for small objects, the effect of animacy was particularly small and only observed in the right PHC.

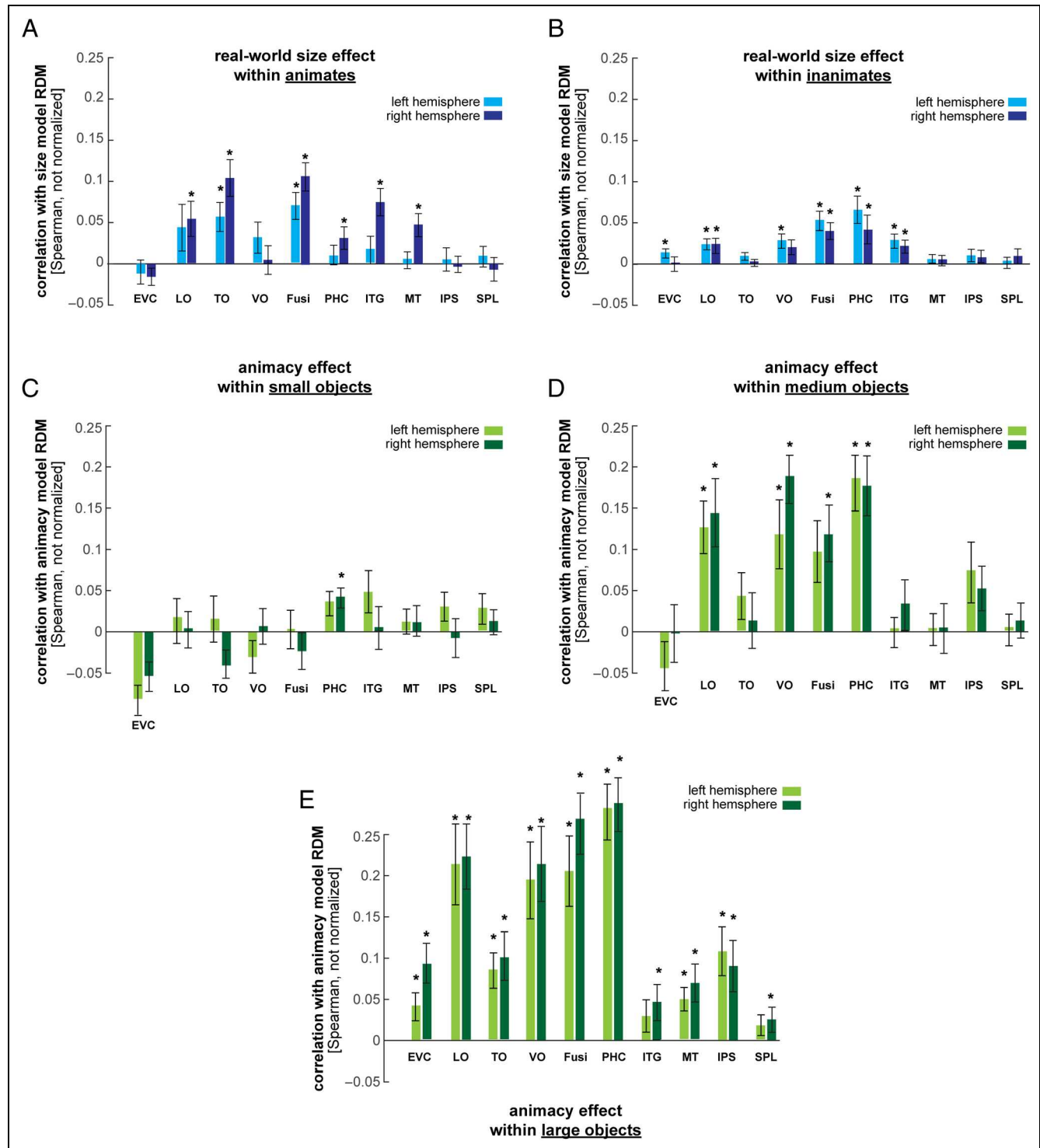In addition, to ameliorate the concern that the scene background might bias the results related to the object

Figure 3. (A, B) Brain regions involved in size processing assessed independently for animate and inanimate objects. (C–E) Brain regions involved in animacy processing assessed independently for small, medium, and large objects. As in Figure 2, bars show correlations between model RDMs with fMRI RDMs for the 10 brain ROIs. Stars show significant correlations. The correlation values are not noise normalized.

size, we determined whether there was a significant correlation between object size and the retinal size of the object. For each image, we calculated the number of pixels covered by the objects and evaluated whether this number differed across the three real-world size categories (i.e., small, medium, and large). Results are reported in Supplementary Figure 3. There was no difference between the three real-world size categories, thus demonstrating that object size effect is not confounded by the pixel size of the objects.

In summary, the fMRI-to-model RSA revealed a distributed and partly overlapping set of brain regions encoding animacy and size, with PHC representing both object properties stronger than most other regions. Our results corroborate and extend previous results of fMRI studies using univariate methods (Konkle & Caramazza, 2013, 2017; Bainbridge & Oliva, 2015; Konkle & Oliva, 2012) that reported similar brain networks of animacy and real-world size processing.

Fine-grained Differences in the Representation of Animacy across Brain Regions

The finding of a distributed and partly overlapping set of regions for animacy and size poses the question about the nature of animacy and size representations in those regions. One possibility is that size and animacy are represented the same way in all regions with a highly redundant code. Another possibility is that different regions represent animacy or size in different ways.

To investigate this issue qualitatively, we visualized the representational geometry of each of the 10 regions involved in either size or object processing. For this, we conducted multidimensional scaling (MDS; Borg & Groenen, 2005; Kruskal & Wish, 1978) on the region-specific RDMs and plotted the results in the first two dimensions (Figure 4A). Visual inspection suggested that ROIs differed in the representational geometry for animacy. For example, TO showed a tight cluster for animate objects but not for inanimate objects. PHC revealed the opposite pattern, that is, a tight cluster for inanimate but not animate objects. LO had two equally tight clusters for animate and inanimate objects.

To quantify these qualitative observations, we defined three model RDMs (Figure 4B) exploring different assumptions about animacy representations. Model 1 assumes equally tight clustering for animate and inanimate objects. Model 2 assumes tighter clustering for animate objects; and Model 3, for inanimate objects. We then correlated the region-specific fMRI RDMs for each of the 10 regions with the Model 1–3 RDMs. This analysis (Figure 4C) confirmed the qualitative observations quantitatively for LO, TO, and PHC and did not reveal any significant effects that were consistent across hemispheres in any other region. Thus, different ROIs have different representational geometries of animacy. We found that some regions (e.g., LO) did not show a

specific preference for one of the representational geometries, whereas both TO and PHC preferred a different representational geometry of animacy (Models 2 vs. 3). This shows that LO versus TO and PHC have different representational geometries of animacy. This suggests that, although all those regions represent animacy, they do so in different ways, potentially because they weigh features that define animacy differently.

In equivalent fashion, we further explored different representational geometries for object size. However, none of the brain areas showed a consistent and significant preference for one representational geometry over the others.

Temporal Dynamics of Animacy and Object Size Processing during Object Vision

Using MEG-to-model RSA, we investigated when in time during visual perception are animacy and real-world object size processed (Figure 1D). The results are shown in Figure 5A. We determined peak and onset latencies to describe the resulting time courses. Onset latencies were similar for size and animacy but slightly shorter for size (size: 144 ± 1 msec [*SEM*]; animacy: 156 ± 1 msec; $p < .01$, bootstrap resampling of participants). There was no significant difference between the peak latencies for size and animacy (size: 193 ± 29 msec; animacy: 184 ± 23 msec; bootstrap resampling, $p > .05$).

To ascertain the robustness of the results, we investigated the processing of size and animacy independent of each other by repeating the MEG-to-model RSA within each level of the factors size and animacy separately. In detail, we conducted MEG-to-size model RSA twice separately for animate and inanimate objects (Figure 5B) and MEG-to-animacy model RSA thrice separately for each size level (Figure 5C). The analyses revealed significant effects in all cases, demonstrating the robustness of our findings.

Together, we report for the first time a time-resolved neuromagnetic signature of physical object size processing and corroborate previous studies investigating the temporal dynamics of animacy processing by finding comparable peak latencies at ~150 msec (Cichy et al., 2014; Carlson et al., 2013).

Content-dependent MEG–fMRI Fusion Resolves Processing of Animacy and Size Simultaneously in Space and Time

To resolve the neural dynamics of animacy and size processing simultaneously in space and time, we used an extension of MEG–fMRI fusion that made it content dependent (Figure 1E). The general rationale of fMRI–MEG fusion by RSA is this: If an RDM based on MEG patterns at time point t is similar to an RDM based on fMRI patterns at location l, location l might be the source of the signal observed at t. Conducted for all time points
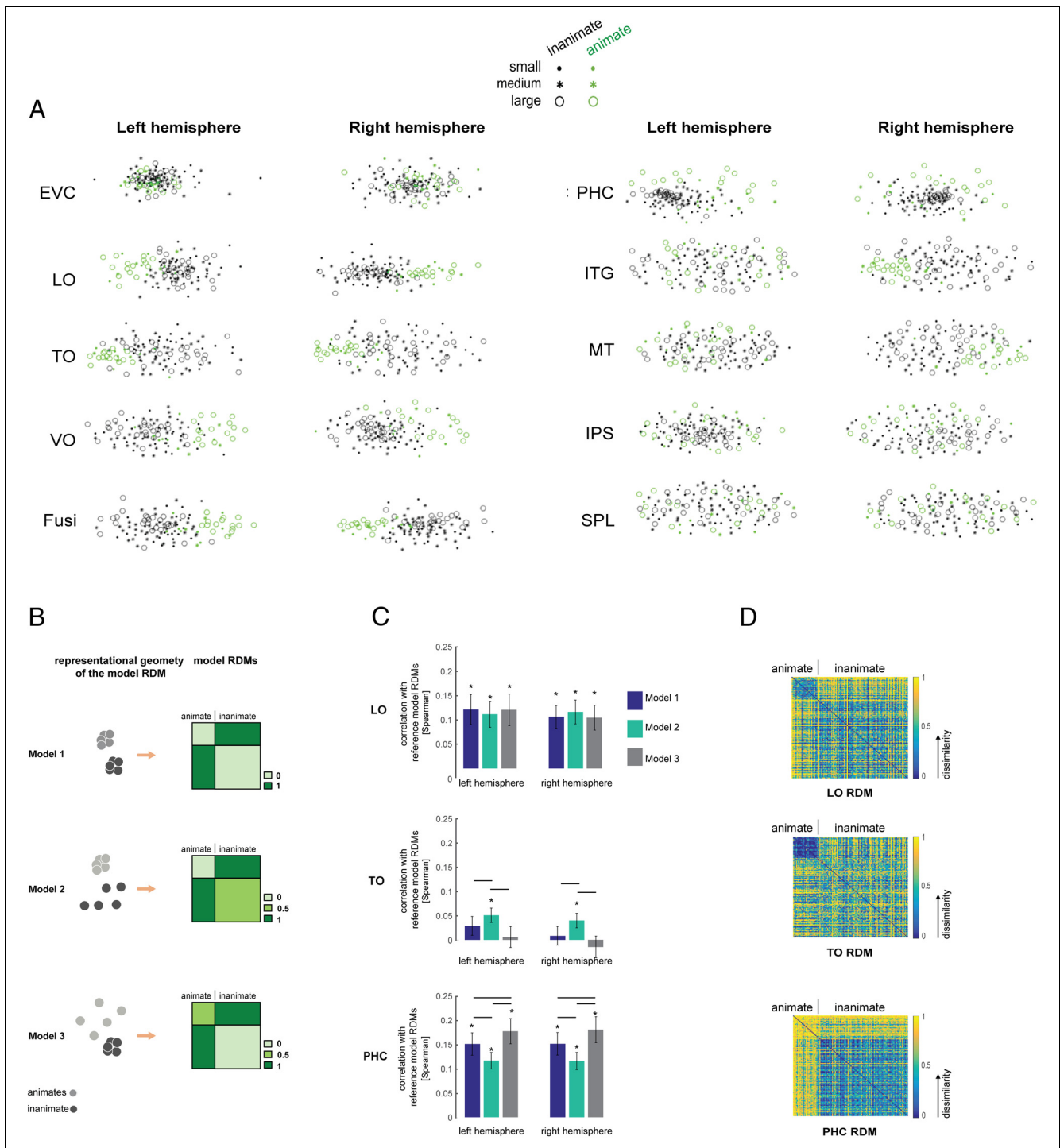
Figure 4. Representational geometries of regions representing animacy and object size. (A) We used nonmetric MDS minimizing the loss function stress (Borg & Groenen, 2005; Kruskal & Wish, 1978) on region-specific fMRI RDMs to visualize how brain regions represent object size and animacy. The MDS results are plotted for the first two dimensions. Each condition (i.e., image) is represented by a colored dot, with the color indicating category (animacy) and the dot shape indicating real-world size. (B) Three model RDMs quantifying three different representational geometries for animacy. Model 1 assumes that both animate and inanimate objects form tight clusters. Model 2 assumes a tighter cluster for animate compared with inanimate objects. Model 3 assumes a tighter cluster for inanimate compared with animate objects. (C) Results of fMRI-to-model RSA for the three different models of animacy. Bars show nonnormalized RDM correlations for regions that showed significant effects that were consistent across hemispheres. Stars above each bar show statistical significance ($N = 15$, one-sided signed-rank test, $p < .05$). A horizontal line over two bars indicates that the two model RDMs perform significantly differently ($N = 15$, two-sided signed-rank test, $p < .05$). Results are corrected for multiple comparisons by FDR correction across all pairwise comparisons. (D) Sample fMRI RDMs for LO, TO, and PHC, each showing a different representational geometry of animacy.

Figure 5. (A) Temporal dynamics of animacy and size processing. Time courses are MEG-to-model similarities averaged across participants. Shading indicates standard errors. (B) Temporal dynamics of size processing assessed independently for animate and inanimate objects. (C) Temporal dynamics of animacy processing assessed independently for small, medium, and large objects. Significant time points ($N = 15$, one-sided signed-rank test, $p < .05$, FDR-corrected for number of time points) are indicated by color-coded horizontal lines above time courses. In B and C, dashed lines indicate results within experimental factors and solid curves indicate the average across them.
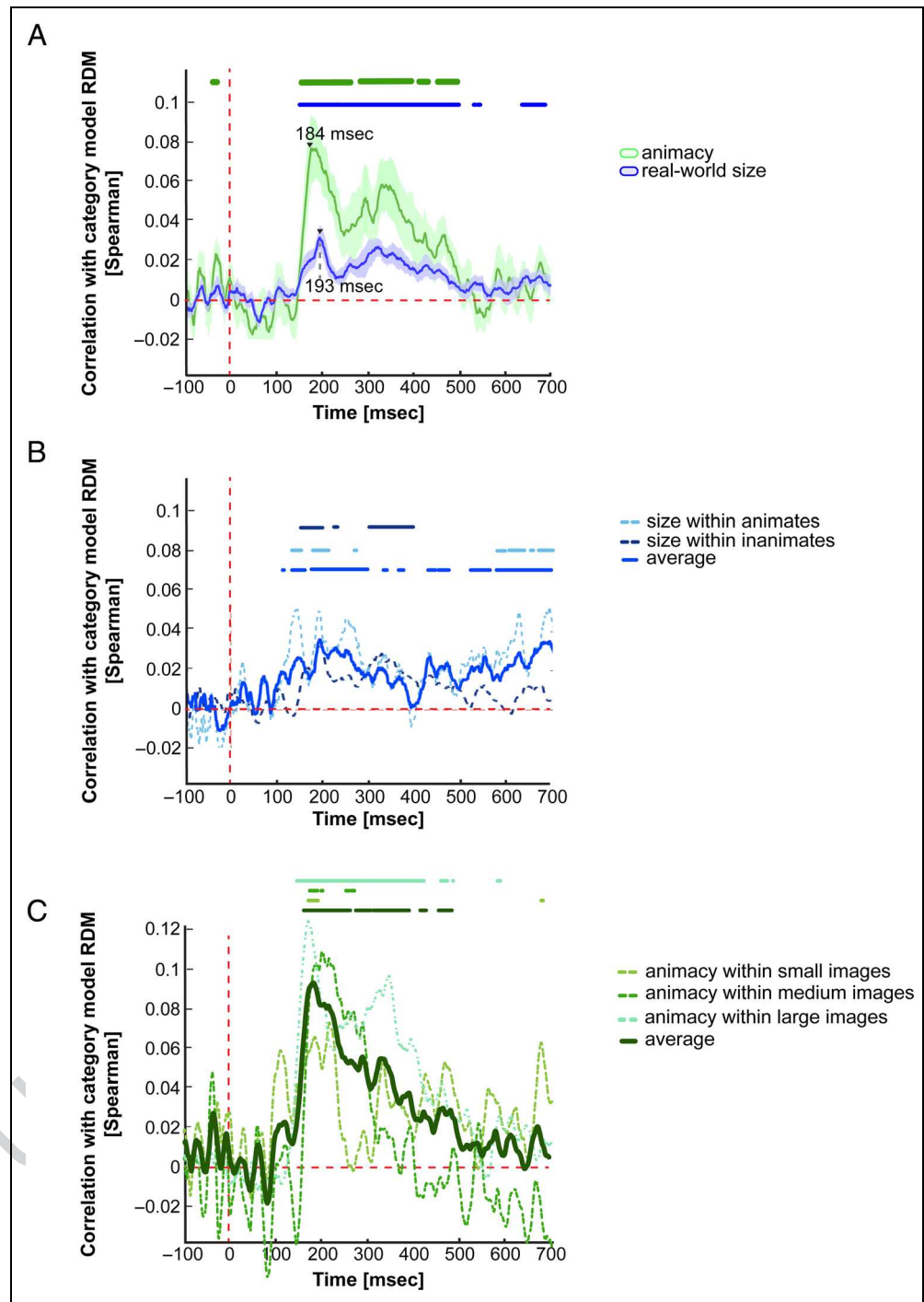
and location combinations, this yields a spatiotemporally resolved map of brain activity. Here, we extend this formulation by the constraint that this spatiotemporal map must have a representational geometry significantly correlated with the theoretical model of animacy or the theoretical model of size (Figure 1E), thus turning it into content-dependent MEG–fMRI fusion. This analysis has the potential to reveal the dynamics of size and animacy processing beyond what MEG or fMRI alone can do by disambiguating the many-to-many mapping between locations and time points.

The full content-dependent MEG–fMRI fusion results for size and animacy processing are shown in Movie 1. Figure 6 shows the results at key time points. The main findings were threefold. First, we found a spatiotemporally distributed and partly overlapping network of brain areas involved in processing both size and animacy from ~156 msec onward. Second, PHC showed distinct behavior from other regions within this network, highlighting it as a key region in animacy and size processing. Third, EVC proved to be a differentiator between animacy and size processing in coding only real-world object size

Movie 1. The top row shows the spatiotemporally resolved movie for animacy processing in green. The bottom row shows the spatiotemporally resolved movie for object size processing in blue. For reference, size and animacy time courses are shown in the center. The image exemplarily shows a snapshot at 170 msec. (www.dropbox.com/s/w92de00rv2h4jny/movie1_AnimSize.mp4?dl=0).
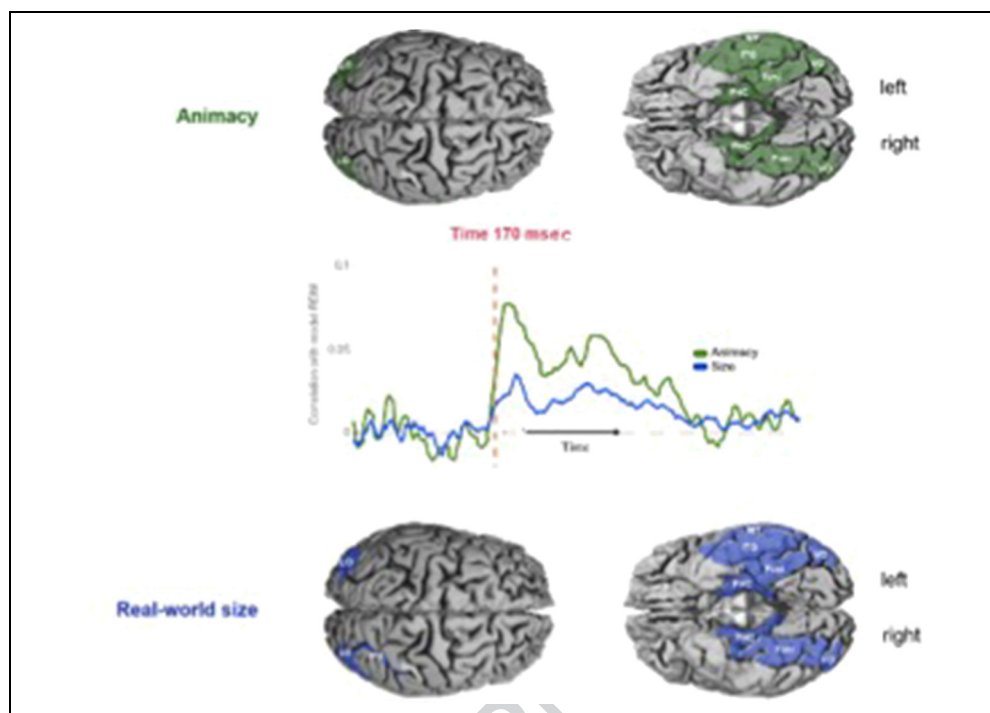
(starting at 144 msec), but not animacy. We detail each main finding below.

Distributed and Partly Overlapping Spatiotemporal Dynamics Underlie Animacy and Real-world Size Processing

We found that the spatiotemporal dynamics underlying animacy and real-world size processing were distributed and partly overlapping (Figure 6). In detail, starting at ~144 msec, several brain regions represented real-world object size, including the left EVC, bilateral PHC, LO, VO, and Fusi. From ~156 to ~350 msec, we found a distributed and partially overlapping set of brain areas including PHC, ITG, VO, LO, Fusi, and TO. In particular, we inspected the regions active at the peak times of animacy and size processing as indicated by MEG-to-model analysis above (Figure 5A). At the peak time for animacy processing, that is, 178 msec, bilateral LO, VO, Fusi, PHC, and left MT and ITG were active.

At the peak time for object size processing, that is, 193 msec, bilateral PHC, right VO, and left MT were active. Moving ahead to the period of 220–260 msec, we found that PHC alone represented both size and animacy. At 350 msec, activity related to both size and animacy processing in several regions reemerged. Finally, at 655 msec, we found that EVC represented real-world size.

In summary, these results delineate the neural dynamics in the overlapping set of brain regions that represent size and animacy, showing that they represent both object properties with similar temporal dynamics.

PHC is a key region for representing animacy and real-world size. The fMRI-to-model RSA (Figure 2) had singled out PHC as a key region in processing both size and animacy. The results of content-dependent MEG–fMRI fusion extend this finding by showing that PHC simultaneously codes for both size and animacy over several hundreds of milliseconds from ~156 to ~430 msec.

A supplementary analysis of standard ROI-based MEG fusion further emphasized the role of PHC in size and animacy processing by showing that peak and onset latencies (Supplementary Table 1) in PHC were later than for many ventral stream regions (bootstrap test, FDR-corrected at $p < .05$). Specifically, in the left hemisphere, PHC's onset latency was significantly later than those of VO, LO, and Fusi, and PHC's peak latency was significantly later than that of LO. In the right hemisphere, PHC's peak latency was significantly later than those of LO and Fusi, and PHC's onset latency was significantly later than that of LO.

In summary, we found that PHC was more strongly engaged in both animacy and size processing than other visual regions and that peak activity in PHC occurred later than in other ventral visual stream regions. Together, this suggests that PHC may be a cortical hub where object-related information is aggregated after being processed in regions lower in the visual processing hierarchy.

The EVC: From Representing Low-level Visual Features to High-level Real-world Size Information

We observed that the left EVC represented real-world size information starting at 144 msec (±1 *SEM*; Figure 6)
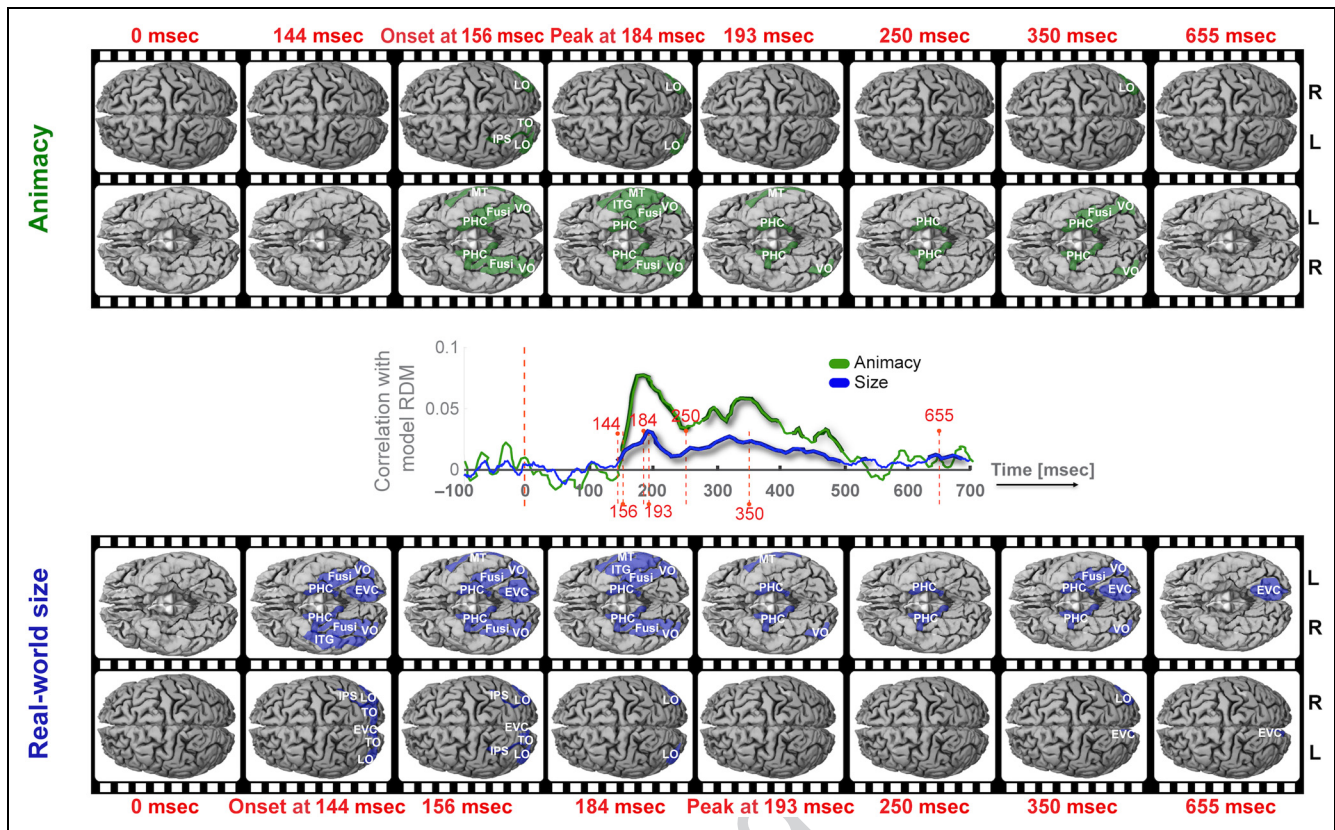
Figure 6. Spatiotemporal dynamics of animacy and real-world size revealed by content-dependent MEG–fMRI fusion. Snapshots of Movie 1 detailing the full spatiotemporally resolved neural representation related to animacy (top) and size (bottom) processing. The middle row reproduces the time courses of size and animacy processing as shown in Figure 5A. Vertical red lines mark particular time points of interest. L = left; R = right.

but did not code for animacy at any time point. This singles out EVC as a differentiator between real-world size and animacy processing.

This result poses the question whether the observed real-world size-related activity reflects feedforward processing of low-level visual information or whether it reflects feedback processing of high-level visual information. To investigate this, we conducted a supplementary standard ROI-based fusion analysis, revealing the temporal dynamics of EVC independent of content (Supplementary Figure 1, Supplementary Table 1). We reasoned that standard content-independent analysis of EVC activity would reveal the dynamics of low-level visual information processing and thus allow discerning potential differences in timing that might differentiate between the two options. We found that EVC had an early onset latency of 78 msec (± 4 *SEM*) and a peak latency of 128 msec (± 2 *SEM*) in the left hemisphere (for the right EVC: onset = 72 msec [± 4 *SEM*], peak = 129 msec [± 12 *SEM*]), likely reflecting feedforward processing of low-level visual information (for further evidence, see also the next section). The temporal precedence and thus dissociation of peak and onset latencies (bootstrap test, $p < .05$) observed in the content-independent analysis compared with the content-dependent analysis argues against the idea of feedforward processing of low-level features as underlying the

real-world object size signal in EVC. It instead favors the view that the observed representation of real-world size in EVC reflects processing high-level visual information feedback from higher visual areas.

In summary, we find evidence of two distinct processing phases in EVC: one early phase of processing low-level visual information and a later phase (starting at ~144 msec) of processing higher-level visual information such as real-world size.

The Spatiotemporal Dynamics of Processing Animacy and Size Are Distinct from Processing Low-level Visual Features

Animacy and real-world size are high-level visual properties that are not trivially encoded in low-level visual features, as demonstrated by the complex computational transformations of low- to high-level visual features necessary for successful image categorization in artificial systems (Kietzmann, McClure, & Kriegeskorte, 2017; Krizhevsky, Sutskever, & Hinton, 2012; Rajaei, Khaligh-Razavi, Ghodrati, Ebrahimpour, & Abadi, 2012; Serre, Oliva, & Poggio, 2007). Under this assumption, the neural activity related to animacy and object-size processing described here reflects processing of high- rather than low-level visual properties.

To control whether this assumption in fact holds in our experimental setting, we investigated whether the spatiotemporal dynamics of low-level visual processing are distinct from the dynamics of processing animacy and size. To capture the low-level visual feature content of the stimulus set, we used the Gist descriptor (Oliva & Torralba, 2001, 2006). Gist is a computer vision image descriptor based on localized responses of Gabor filters at different scales and orientations. It has been shown to predict fMRI responses of early and midlevel visual areas previously (Khaligh-Razavi, Henriksson, Kay, & Kriegeskorte, 2017), making it a reasonable model of low-level visual processing in the cortex.

Importantly, as the GIST model has been found to predict several scene layout-related properties such as openness, expanse, and indoor/outdoor (Oliva & Torralba, 2001, 2006; Oliva, 2005), this also serves as a control for the scene layout. We constructed a model RDM based on Gist features for the 118 object images and compared it with the size and also animacy model RDMs. We found that the correlation was very low (i.e., Gist-size model correlation was .045 and Gist-animacy model correlation was .008), which makes it unlikely that the size or animacy effect observed here was driven by low-level to midlevel visual features or the scene layout. To further investigate the role of low-level visual features, we calculated size/animacy correlations with MEG, after regressing out the Gist effect (Supplementary Figure 4). Comparing time courses where the Gist RDM was partialed out with the time courses where it was not (same as reported in

Figure 5) shows a significant difference only early between 93 and 130 msec for animacy and 79 and 150 msec for size. This further demonstrates that the later components of the size and animacy effects are unlikely to be affected by low-level/midlevel visual features as captured by the Gist features.

We also conducted the fMRI-to-model RSA (Figure 1C), MEG-to-model RSA (Figure 1D), and content-dependent MEG–fMRI fusion (Figure 1E) using a model RDM constructed from the Gist descriptor of the stimulus set. The fMRI-to-Gist-model RSA revealed significant correlations with bilateral EVC and left V4, but no other visual areas (Figure 7A). The MEG-to-Gist-model RSA revealed a time course different from the one revealed for the animacy and size model (as shown in Figure 5A). In particular, for Gist, both the onset latency at 67 msec (±6 msec [*SEM*]) and the peak latency at 120 msec (±7 msec [*SEM*]) were significantly smaller than the onset and peak latencies for either animacy or real-world size (bootstrap resampling test, $p < .001$).

Finally, content-dependent MEG–fMRI fusion for Gist further revealed that the spatiotemporal dynamics of processing Gist were different from processing animacy and size (Figure 7C and Movie 2). The identified dynamics started in EVC at 84 msec, followed by V4 20 msec later, and persisted in EVC for almost the entire percept duration until 500 msec. This is in stark contrast to the onset of processing animacy and size at ~150 msec in a set of high-level visual regions for animacy and size.

Figure 7. Spatial and temporal dynamics of processing low-level visual features as captured by the Gist descriptor. Results of (A) fMRI-to-Gist model RSA. Each bar shows correlation between the RDM of the given brain area and the gist RDM. Correlations are normalized by the noise ceiling of each brain ROI. Nonnormalized RDM correlation values are as follows: left EVC = .07, right EVC = .06, and right V4 = .05. (B) MEG-to-Gist model RSA. Time series has onset at 67 msec (±6 *SEM*) and peak at 120 msec (±7 *SEM*). (C) Gist-dependent MEG–fMRI fusion: We followed the same procedure as explained in Figure 1 for size and animacy, except that here we used the Gist model RDM. For a full spatiotemporally resolved neural activity related to Gist, see Movie 2. The Gist RDM had a correlation of .045 with the size RDM and a correlation of .008 with the animacy model RDM. L = left; R = right.
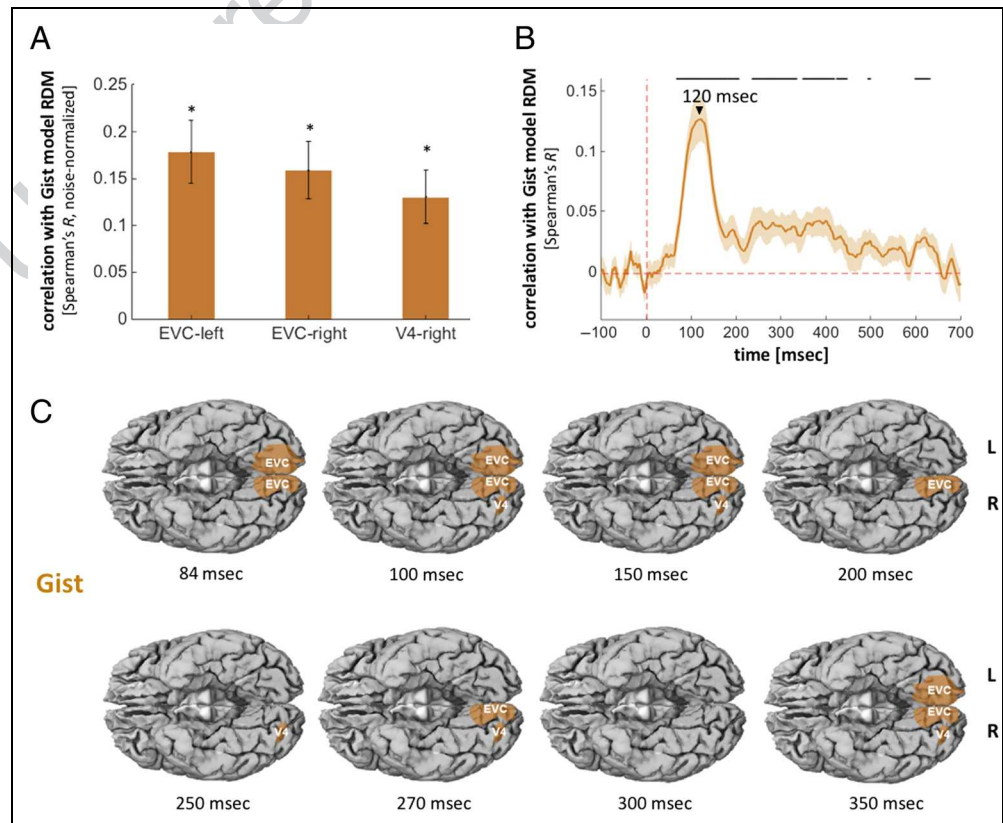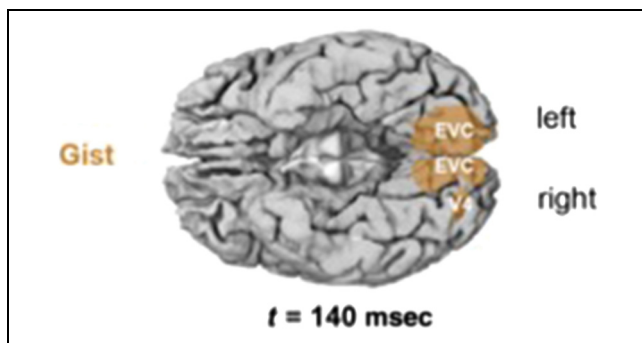
Movie 2. Spatiotemporally resolved neural activity related to Gist descriptor. The image exemplarily shows a snapshot at 140 msec. (www.dropbox.com/s/v2vqzpdx881mqhm/movie2_Gist.mp4?dl=0).

A further important difference was the timing and role of EVC activation. In EVC, Gist processing started at 84 msec, which was earlier than size processing that started at around 144 msec (Figures 5 and 6; $p < .05$, based on bootstrap resampling test). This strengthens the idea that the observation of size processing in EVC is driven by feedback from higher visual areas.

DISCUSSION

Summary

We identified the spatiotemporal neural dynamics of animacy and object size by applying RSA to fMRI and MEG data. Investigation of the cortical regions representing animacy and size using fMRI revealed a distributed and partly overlapping set of cortical regions. This set of regions extended from occipital to ventral and medial temporal cortex and parietal cortex. Within the set, PHC was singled out as the region representing animacy and size stronger than most other regions. Fine-grained analysis of representational geometry revealed that different regions processed animacy in distinct ways. Investigation of the time course of animacy and size processing using MEG revealed overlapping temporal extent starting at around 150 msec and provided the first time-resolved neuromagnetic signature of real-world object size processing. Using an extension of MEG–fMRI fusion that shows neural dynamics specific to particular content, we revealed simultaneously in space and time the neural dynamics related to processing animacy and size. We found partly overlapping and distributed dynamics, with PHC singled out as a region that represented size and animacy persistently when other regions did not. A control analysis revealed that the neural dynamics of processing animacy and size were distinct from the neural dynamics of processing low-level visual features.

Limitations of the Content-dependent MEG–fMRI Fusion Method

Although MEG–fMRI fusion in general and in the content-dependent formulation is a promising method to resolve neural activity in space and time, acknowledgment of its limitations is necessary to guide interpretation. For one, the method in its formulation here compares neural activity measured at largely different spatial scales: Whereas sensor-space MEG patterns reflect activity from all of the cortices, fMRI searchlight patterns reflect activity in local neural populations. Thus, depending on how neural signals from different sources impact MEG sensor-level activation patterns, correspondences might go unnoticed. Similarly, for MEG–fMRI fusion to detect spatiotemporal correspondences, relevant representational structure must be present in both MEG and fMRI. However, there is no guarantee that conditions evoking significant and robustly different activation patterns in one imaging modality will do so in the other imaging modality, too. For example, Proklova, Kaiser, and Peelen (2017) found that, whereas the animacy of objects in a particularly controlled stimulus set was clearly detectable from fMRI patterns, it was not detectable from MEG sensor patterns. In summary, as MEG–fMRI fusion depends on two methods with particular sensitivities and limitations that both establish correlations rather than causality, both positive and negative findings have to be interpreted cautiously.

A Distributed and Partly Overlapping Set of Cortical Regions Represents Object Animacy and Real-world Size

Our multivariate analyses based on representational similarity showed that object animacy and size are represented in a distributed and partly overlapping set of regions, rather than a single region. This result is consistent with previous work, demonstrating shared and distributed networks of brain regions involved in processing object properties and stimulus domains (Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016; Behrmann & Plaut, 2013; de Beeck, Brants, Baeck, & Wagemans, 2010; O'Toole, Jiang, Abdi, & Haxby, 2005; Spiridon & Kanwisher, 2002; Haxby et al., 2001; Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999). Thus, our results further strengthen the view that the brain processes information in a distributed rather than localized fashion. The advantage of such a visual architecture might be manifold: For example, distributed representations are more robust to impairments, and importantly, distributed but integrated large-scale networks are essential to ensure rapid and accurate visual object recognition (Behrmann & Plaut, 2013; Catani, 2007; Mesulam, 1990).

Our results go beyond previous work in two ways. First, we described the detailed representational geometry used by the brain to represent animacy. Although TO, LO, and PHC all represented animacy, we showed that they did so in different ways. This qualifies the notion of distributed networks for visual information processing by showing that different nodes of the network conduct related but different processing. Our results thus allow detailed predictions about behavioral

deficits that might result from damage to those regions or disturbance in processing therein. Future neuropsychological or brain stimulation studies using TMS might test these predictions.

Our results also go beyond previous work in revealing the full spatiotemporal dynamics in the network of regions processing animacy and size (Konkle & Caramazza, 2013, 2017; Fabbri et al., 2016; Bainbridge & Oliva, 2015; Konkle & Oliva, 2012). We demonstrated that, starting from 160 msec after stimulus onset, both size and animacy are represented in partially overlapping brain areas with similar temporal dynamics. Within this spatiotemporally shared network, PHC was singled out among other regions in three ways: It simultaneously coded for both size and animacy from ~160 to ~430 msec, it was the only region representing both size and animacy from 220 to 260 msec, and it had the highest RDM correlation with the reference animacy and size model RDMs. These results combined indicate a critical role for PHC in representing size and animacy in the human brain. They further suggest that PHC is a key cortical node that combines the outputs of two parallel processing streams for size and animacy.

Previous studies did not report a special role for PHC in processing animacy and real-world size but rather in processing scenes (Epstein, 2005; Epstein & Kanwisher, 1998). This discrepancy might be explained by differences in the nature of the stimulus sets used to probe the visual system. Previous studies investigating animacy and size processing used silhouette images without background (Konkle & Caramazza, 2013, 2017; Cichy et al., 2014; Carlson et al., 2013; Connolly et al., 2012; Konkle & Oliva, 2011, 2012; Bell et al., 2009; Wiggett et al., 2009; Kriegeskorte et al., 2008). In contrast, here, we presented images on natural backgrounds as they appear in the real world. Furthermore, we found that (after controlling for low-level visual features) PHC indeed differentiated between scenes that contain animates versus scenes that contain inanimates. Thus, size and animacy processing in PHC might depend on the presence of a real-world background. We hypothesize that there is an interaction here between animacy processing in PHC and the object images being on a natural background. In other words, PHC only processes animates in the presence of a background scene. This observation further highlights the importance of using naturalistic stimulus sets to understanding the workings of the visual system under real-world conditions.

Finally, we found that processing of low-level visual features as captured by the Gist descriptor occurred predominantly in low-level and midlevel visual areas rather than midlevel and high-level visual areas. This result is consistent with the notion that animacy and size are high- rather than low-level visual properties (Khaligh-Razavi et al., 2017; Konkle & Oliva, 2012; Kriegeskorte et al., 2008) and thus being processed at later stages in the hierarchical processing cascade of the ventral visual system.

Time Courses of Processing Animacy and Real-world Object Size

Our results concur with previous studies investigating the time course of object processing (Clarke et al., 2015; Carlson et al., 2013, 2014; Cichy et al., 2014; Liu et al., 2009; Thorpe et al., 1996) in revealing a similar time course for the processing of animacy. The peak of animacy processing at 178 msec fell within the peak confidence interval reported in previous MEG studies at 152–302 msec (Cichy et al., 2014) and at 150–240 msec (Carlson et al., 2013).

Our results go beyond previous studies by, to our knowledge for the first time, reporting the time course of real-world object size processing. Importantly, we were able to directly compare the speed of processing real-world size with that of animacy with high precision, as we studied these two object properties using the same stimulus set and the same imaging technique and analysis pipeline. Compared with animacy processing, size processing was slightly faster. In detail, the onset latency for size processing at 144 msec was ~10 msec earlier than the onset latency for animacy processing at 156 msec. However, the peak latency of size processing was not significantly different from that of animacy. Together with the observation that a partly overlapping set of brain regions underlies animacy and size processing, this shows that neural activity underlying animacy and size processing is distributed and overlapping in both space and time.

Compared with the timing of processing low-level visual features as captured by the Gist descriptor, onsets and peak latencies for animacy and size processing were clearly and significantly later in time. This result further strengthened the notion that size and animacy are object properties processed at a late stage of the hierarchical processing cascade of the ventral visual system. Together with the fMRI results, this differentiates the processing of animacy and size from low-level visual processing in both space and time.

The Role of Feedback in the EVC Representation of Real-world Size

Representing real-world size, as opposed to retinal size, is a critical and useful feature of our visual system, referred to as object size constancy (Sperandio, Chouinard, & Goodale, 2012). As objects move with respect to one another, their retinal size varies continuously. If our visual system were to solely represent objects based on their retinal size, without an understanding of their real-world size, the world around us would appear distorted and unstable.

Previous studies have shown that real-world object size is already represented at the first stage of cortical visual processing, that is, primary visual cortex (Sperandio et al., 2012; Schwarzkopf, Song, & Rees, 2011; Murray et al., 2006). Our content-dependent fusion results extend those studies by detailing the temporal dynamics

and the role of feedback in the representation of object size in EVC through three observations.

First, we observed that the left EVC represented real-world object size starting at ~150 msec, which is significantly later than the onset at 78 msec and peak at 128 msec of overall visual activity in EVC (Supplementary Table 2). This temporal delay strongly suggests that size representations in EVC emerged through feedback processing rather than feedforward processing.

Second, we further observed that, at 80 msec, EVC represented Gist features earlier in time than it represented object size. Given that the Gist model assesses low-level visual processing in early visual areas, whereas object size was most strongly represented in high-level visual cortex, this further suggests that size representations in EVC emerged through feedback. The relative temporal succession of processing low- and high-level visual information in EVC observed here relatively matches results in nonhuman primates (Supèr, Spekreijse, & Lamme, 2001; Lee, Mumford, Romero, & Lamme, 1998). Single-cell electrophysiology in monkeys revealed that recurrent dynamics in the primary visual cortex emerge in the afterburst starting at 100 msec and have a role in processing higher-level visual information (Lamme & Roelfsema, 2000; Lee et al., 1998).

Finally, examining brain regions active after stimulus offset revealed that, at 655 msec, only the left EVC was representing real-world size. This might also suggest involvement of feedback given that there is no incoming feedforward visual signal at this time.

Future Directions and Limitations

In this study, we used an already existing neuroimaging data set (Cichy et al., 2016) to showcase the potential of the content-dependent fusion method for studying the representation of visual properties, in particular, real-world size and animacy. Although the data and stimulus set were not primarily designed to investigate real-world size or animacy and were thus not fully controlled for those two object properties, it did allow for a set of analyses that revealed novel insights into their neural representations.

The stimulus set here consisted of images of real-world objects on natural backgrounds. Although this ensures some ecological validity of our findings, it precludes a systematic analysis of which features drive object size and animacy processing. For this, fully controlled stimulus sets are required that are designed with the intention to differentiate between candidate features. Univariate fMRI analyses (Julian, Ryan, Hamilton, & Epstein, 2016; Bainbridge & Oliva, 2015; Konkle & Oliva, 2012) suggest that the cortical representation of object size might reflect object properties that correlate with size, such as the interaction envelope, or whether objects are used as landmarks in navigation (Julian et al., 2016). A recent study (Long, Yu, & Konkle, 2017) also suggests a significant contribution for midlevel visual features accounting for the large-scale organization of the ventral visual stream when processing object size and animacy. We see great potential for future studies with fully controlled experimental results that use content-dependent fusion to unravel the features that underlie animacy and size processing.

UNCITED REFERENCES

Bi, Wang, & Caramazza, 2016
Buckingham, Cant, & Goodale, 2009
Charest, Kievit, Schmitz, Deca, & Kriegeskorte, 2014
Cichy & Teng, 2017
Dale & Halgren, 2001
Debener, Ullsperger, Siegel, & Engel, 2006
DiCarlo, Zoccolan, & Rust, 2012
Fabre-Thorpe, 2011
Grill-Spector & Weiner, 2014
Groppe, Urbach, & Kutas, 2011
Huster, Debener, Eichele, & Herrmann, 2012
Jorge, Van Der Zwaag, & Figueiredo, 2014
Kanwisher, McDermott, & Chun, 1997
Kourtzi & Kanwisher, 2000
Kriegeskorte & Kreiman, 2012
Leeds, Seibert, Pyles, & Tarr, 2013
Naselaris, Stansbury, & Gallant, 2012
Rosa, Daunizeau, & Friston, 2010
Rousselet, Thorpe, & Fabre-Thorpe, 2004a
Rousselet, Thorpe, & Fabre-Thorpe, 2004b
Salmela, Salo, Salmi, & Alho, 2016

Acknowledgments

REFERENCES

Bainbridge, W. A., & Oliva, A. (2015). Interaction envelope: Local spatial representations of objects at all scales in scene-selective regions. Neuroimage, 122, 408–416.
Behrmann, M., & Plaut, D. C. (2013). Distributed circuits, not circumscribed centers, mediate visual recognition. Trends in Cognitive Sciences, 17, 210–219.
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.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B: Methodological, 57,* 289–300.

Bi, Y., Wang, X., & Caramazza, A. (2016). Object domain and modality in the ventral visual pathway. *Trends in Cognitive Sciences, 20,* 282–290.

Buckingham, G., Cant, J. S., & Goodale, M. A. (2009). Living in a material world: How visual cues to material properties affect the way that we lift objects and perceive their weight. *Journal of Neurophysiology, 102,* 3111–3118.

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

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

Catani, M. (2007). From hodology to function. *Brain, 130,* 602–605.

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.

Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D., & Kriegeskorte, N. (2014). Unique semantic space in the brain of each beholder predicts perceived similarity. *Proceedings of the National Academy of Sciences, U.S.A., 111,* 14565–14570.

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

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.

Cichy, R. M., & Teng, S. (2017). Resolving the neural dynamics of visual and auditory scene processing in the human brain: A methodological approach. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 372,* 20160108.

Clarke, A., Devereux, B. J., Randall, B., & Tyler, L. K. (2015). Predicting the time course of individual objects with MEG. *Cerebral Cortex, 25,* 3602–3612.

Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y. C., et al. (2012). The representation of biological classes in the human brain. *Journal of Neuroscience, 32,* 2608–2618.

Dale, A. M., & Halgren, E. (2001). Spatiotemporal mapping of brain activity by integration of multiple imaging modalities. *Current Opinion in Neurobiology, 11,* 202–208.

de Beeck, H. P. O., Brants, M., Baeck, A., & Wagemans, J. (2010). Distributed subordinate specificity for bodies, faces, and buildings in human ventral visual cortex. *Neuroimage, 49,* 3414–3425.

Debener, S., Ullsperger, M., Siegel, M., & Engel, A. K. (2006). Single-trial EEG–fMRI reveals the dynamics of cognitive function. *Trends in Cognitive Sciences, 10,* 558–563.

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

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

Epstein, R. (2005). The cortical basis of visual scene processing. *Visual Cognition, 12,* 954–978.

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

Fabbri, S., Stubbs, K. M., Cusack, R., & Culham, J. C. (2016). Disentangling representations of object and grasp properties in the human brain. *Journal of Neuroscience, 36,* 7648–7662.

Fabre-Thorpe, M. (2011). The characteristics and limits of rapid visual categorization. *Frontiers in Psychology, 2,* 243.

Gibbons, J. D., & Chakraborti, S. (2011). *Nonparametric statistical inference* (pp. 977–979). Berlin, Germany: Springer.

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.

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology, 48,* 1711–1725.

Hollander, M., & Wolfe, D. A. (1999). *Nonparametric statistical methods.* Wiley.

Huster, R. J., Debener, S., Eichele, T., & Herrmann, C. S. (2012). Methods for simultaneous EEG–fMRI: An introductory review. *Journal of Neuroscience, 32,* 6053–6060.

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature, 532,* 453–458.

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.

Jorge, J., Van Der Zwaag, W., & Figueiredo, P. (2014). EEG–fMRI integration for the study of human brain function. *Neuroimage, 102,* 24–34.

Julian, J. B., Ryan, J., Hamilton, R. H., & Epstein, R. A. (2016). The occipital place area is causally involved in representing environmental boundaries during navigation. *Current Biology, 26,* 1104–1109.

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.

Khaligh-Razavi, S. M., Henriksson, L., Kay, K., & Kriegeskorte, N. (2017). Fixed versus mixed RSA: Explaining visual representations by fixed and mixed feature sets from shallow and deep computational models. *Journal of Mathematical Psychology, 76,* 184–197.

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

Kietzmann, T. C., Gert, A. L., Tong, F., & König, P. (2017). Representational dynamics of facial viewpoint encoding. *Journal of Cognitive Neuroscience, 29,* 637–651.

Kietzmann, T. C., McClure, P., & Kriegeskorte, N. (2017). Deep neural networks in computational neuroscience. *bioRxiv,* 133504.

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

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

Konkle, T., & Oliva, A. (2011). Canonical visual size for real-world objects. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 23.

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

Kourtzi, Z., & Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. *Journal of Neuroscience, 20,* 3310–3318.

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences, 17,* 401–412.

Kriegeskorte, N., & Kreiman, G. (Eds.). (2012). *Visual population codes: Toward a common multivariate framework for cell recording and functional imaging*. Cambridge, MA: MIT Press.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (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.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23,* 571–579.

Lee, T. S., Mumford, D., Romero, R., & Lamme, V. A. (1998). The role of the primary visual cortex in higher level vision. *Vision Research, 38,* 2429–2454.

Leeds, D. D., Seibert, D. A., Pyles, J. A., & Tarr, M. J. (2013). Comparing visual representations across human fMRI and computational vision. *Journal of Vision, 13,* 25.

Liu, H., Agam, Y., Madsen, J. R., & Kreiman, G. (2009). Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron, 62,* 281–290.

Long, B., Yu, C. P., & Konkle, T. (2017). A mid-level organization of the ventral stream. *bioRxiv,* 213934.

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.

Mesulam, M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Annals of Neurology, 28,* 597–613.

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

Murray, S. O., Boyaci, H., & Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. *Nature Neuroscience, 9,* 429–434.

Naselaris, T., Stansbury, D. E., & Gallant, J. L. (2012). Cortical representation of animate and inanimate objects in complex natural scenes. *Journal of Physiology, Paris, 106,* 239–249.

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage, 25,* 653–660.

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

Oliva, A. (2005). Gist of the scene. In *Neurobiology of attention* (pp. 251–256).

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

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research, 155,* 23–36.

O'Toole, A. J., Jiang, F., Abdi, H., & Haxby, J. V. (2005). Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of Cognitive Neuroscience, 17,* 580–590.

Proklova, D., Kaiser, D., & Peelen, M. (2017). MEG sensor patterns reflect perceptual but not categorical similarity of animate and inanimate objects. *bioRxiv,* 238584.

Rajaei, K., Khaligh-Razavi, S. M., Ghodrati, M., Ebrahimpour, R., & Abadi, M. E. S. A. (2012). A stable biologically motivated learning mechanism for visual feature extraction to handle facial categorization. *PLoS One, 7,* e38478.

Rosa, M. J., Daunizeau, J., & Friston, K. J. (2010). EEG-fMRI integration: A critical review of biophysical modeling and data analysis approaches. *Journal of Integrative Neuroscience, 9,* 453–476.

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004a). How parallel is visual processing in the ventral pathway? *Trends in Cognitive Sciences, 8,* 363–370.

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004b). Processing of one, two or four natural scenes in humans: The limits of parallelism. *Vision Research, 44,* 877–894.

Salmela, V., Salo, E., Salmi, J., & Alho, K. (2016). Spatiotemporal dynamics of attention networks revealed by representational similarity analysis of EEG and fMRI. *Cerebral Cortex, 28,* 549–560.

Schwarzkopf, D. S., Song, C., & Rees, G. (2011). The surface area of human V1 predicts the subjective experience of object size. *Nature Neuroscience, 14,* 28–30.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences, U.S.A., 104,* 6424–6429.

Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O., et al. (2015). The animacy continuum in the human ventral vision pathway. *Journal of Cognitive Neuroscience, 27,* 665–678.

Shutts, K., Markson, L., & Spelke, E. S. (2009). The developmental origins of animal and artifact concepts. In B. Hood & L. Santos (Eds.), *The origins of object knowledge* (pp. 189–210). Oxford, UK: Oxford University Press.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22,* 1359–1366.

Sperandio, I., Chouinard, P. A., & Goodale, M. A. (2012). Retinotopic activity in V1 reflects the perceived and not the retinal size of an afterimage. *Nature Neuroscience, 15,* 540–542.

Spiridon, M., & Kanwisher, N. (2002). How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron, 35,* 1157–1165.

Supèr, H., Spekreijse, H., & Lamme, V. A. (2001). Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). *Nature Neuroscience, 4,* 304–310.

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience, 2011,* 8.

Taulu, S., Kajola, M., & Simola, J. (2004). Suppression of interference and artifacts by the signal space separation method. *Brain Topography, 16,* 269–275.

Taulu, S., & Simola, J. (2006). Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Physics in Medicine and Biology, 51,* 1759.

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

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage, 15,* 273–289.

Wang, L., Mruczek, R. E., Arcaro, M. J., & Kastner, S. (2015). Probabilistic maps of visual topography in human cortex. *Cerebral Cortex, 25,* 3911–3931.

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.