

## ORIGINAL ARTICLE

# Similarity-Based Fusion of MEG and fMRI Reveals Spatio-Temporal Dynamics in Human Cortex During Visual Object Recognition

Radoslaw Martin Cichy<sup>1,3</sup>, Dimitrios Pantazis<sup>2</sup>, and Aude Oliva<sup>1</sup><sup>1</sup>Computer Science and Artificial Intelligence Laboratory and <sup>2</sup>McGovern Institute for Brain Research, MIT, Cambridge, MA, USA and <sup>3</sup>Department of Education and Psychology, Free University Berlin, Berlin, GermanyAddress correspondence to Radoslaw Martin Cichy, Computer Science and Artificial Intelligence Laboratory, MIT, 32-D430, Cambridge, MA, USA.  
Email: rmcichy@mit.edu

## Abstract

Every human cognitive function, such as visual object recognition, is realized in a complex spatio-temporal activity pattern in the brain. Current brain imaging techniques in isolation cannot resolve the brain's spatio-temporal dynamics, because they provide either high spatial or temporal resolution but not both. To overcome this limitation, we developed an integration approach that uses representational similarities to combine measurements of magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI) to yield a spatially and temporally integrated characterization of neuronal activation. Applying this approach to 2 independent MEG–fMRI data sets, we observed that neural activity first emerged in the occipital pole at 50–80 ms, before spreading rapidly and progressively in the anterior direction along the ventral and dorsal visual streams. Further region-of-interest analyses established that dorsal and ventral regions showed MEG–fMRI correspondence in representations later than early visual cortex. Together, these results provide a novel and comprehensive, spatio-temporally resolved view of the rapid neural dynamics during the first few hundred milliseconds of object vision. They further demonstrate the feasibility of spatially unbiased representational similarity-based fusion of MEG and fMRI, promising new insights into how the brain computes complex cognitive functions.

**Key words:** fMRI, MEG, multimodal integration, representational similarity analysis, visual object recognition

## Introduction

A major challenge of cognitive neuroscience is to map the diverse spatio-temporal neural dynamics underlying cognitive functions under the methodological limitations posed by current neuroimaging technologies. Noninvasive neuroimaging technologies such as functional magnetic resonance imaging (fMRI) and magneto- and electroencephalography (M/EEG) offer either high spatial or high temporal resolution, but not both simultaneously. A comprehensive large-scale view of brain function with both high spatial and temporal resolution thus necessitates

integration of available information from multiple brain imaging modalities, most commonly fMRI and M/EEG (for review, see Dale and Halgren 2001; Debener et al. 2006; Rosa et al. 2010; Huster et al. 2012; Jorge et al. 2014).

Here, we provide a novel approach to combine MEG and fMRI based on 2 basic principles. First, we assume “representational similarity”: if neural representations of 2 conditions are similarly represented in fMRI, they should also be similarly represented in MEG (Kriegeskorte 2008; Kriegeskorte and Kievit 2013; Cichy et al. 2014). Second, we assume “locality” of neural representation:

information is represented in neuronal populations in locally restricted cortical regions, rather than being distributed across the whole brain. Based on these assumptions, linking the similarity relations in MEG for each millisecond with the similarity relations in a searchlight-based fMRI analysis (Haynes and Rees 2005; Kriegeskorte et al. 2006) promises a spatio-temporally resolved account of neural activation.

We used the proposed methodology to yield a novel characterization of the spatio-temporal neural dynamics underlying a key cognitive function: visual object recognition. Visual object recognition recruits a temporally ordered cascade of neuronal processes (Robinson and Rugg 1988; Schmolesky et al. 1998; Bullier 2001; Cichy et al. 2014) in the ventral and dorsal visual streams (Ungerleider and Mishkin 1982; Milner and Goodale 2008; Kravitz et al. 2011). Both streams consist of multiple brain regions (Felleman and Van Essen 1991; Grill-Spector and Malach 2004; Wandell et al. 2007; Op de Beeck et al. 2008; Grill-Spector and Weiner 2014) that encode different aspects of the visual input in neuronal population codes (Haxby et al. 2001; Pasupathy and Connor 2002; Haynes and Rees 2006; Kiani et al. 2007; Meyers et al. 2008; Kreiman 2011; Konkle and Oliva 2012; Tong and Pratte 2012; Cichy et al. 2013).

Analysis of 2 MEG–fMRI data sets from independent experiments yielded converging results: Visual representations emerged at the occipital pole of the brain early (~50–80 ms), before spreading rapidly and progressively along the visual hierarchies in the dorsal and ventral visual streams, resulting in prolonged activation in high-level ventral visual areas. This result for the first time unravels the neural dynamics underlying visual object recognition in the human brain with millisecond and millimeter resolution in the dorsal stream and corroborates previous findings regarding the successive engagement of the ventral visual stream in object representations, available until now only from intracranial recordings. Further region-of-interest analyses revealed that dorsal and ventral regions showed MEG–fMRI correspondence in representations significantly later than early visual cortex.

In sum, our results offer a novel description of the complex neuronal processes underlying visual recognition in the human brain and demonstrate the efficacy and power of a similarity-based MEG–fMRI fusion approach to provide a refined spatio-temporally resolved view of brain function.

## Materials and Methods

### Participants

We conducted 2 independent experiments. Sixteen healthy volunteers (10 female, age: mean  $\pm$  SD = 25.87  $\pm$  5.38 years) participated in Experiment 1 and fifteen (5 female, age: mean  $\pm$  SD = 26.60  $\pm$  5.18 years) in Experiment 2. 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).

### Experimental Design and Task

The stimulus set for Experiment 1 was  $C = 92$  real-world object images on a gray background, comprising human and animal faces, bodies, as well as natural and artificial objects (Fig. 2A). This stimulus set was chosen as it had yielded strong and consistent neural activity in previous studies, including monkey

electrophysiology (Kiani et al. 2007), fMRI (Kriegeskorte et al. 2008), and ROI-based MEG–fMRI fusion (Cichy et al. 2014). The stimulus set for Experiment 2 was  $C = 118$  real-world object images on real backgrounds (Fig. 4A). This stimulus set had no strong categorical structure, in particular involving human bodies and faces, with each image being from a different entry-level category. The rationale of this choice was 2-fold: first, to present objects in a more ecologically valid fashion than in Experiment 1 through the use of natural backgrounds and second, to show that MEG–fMRI fusion does not necessarily depend on particular stimulus characteristics and object categories, such as silhouette images that drive V1 strongly, and categories, such as human bodies and faces that drive high-level visual stream strongly.

In both experiments, images were presented at the center of the screen at 2.9° (Experiment 1) and 4.0° (Experiment 2) visual angle with 500 ms duration, overlaid with a gray fixation cross. We adapted presentation parameters to the specific requirements of each acquisition technique.

For MEG, participants completed 2 sessions of 10–15 runs of 420 s each for Experiment 1, and 1 session of 15 runs of 314 s each for Experiment 2. Image presentation was randomized in each run. Trial onset asynchrony was 1.5 or 2 s (Experiment 1; see Supplementary Fig. 1) and 0.9–1 s (Experiment 2). SOA was shortened for Experiment 2, as no significant effects in MEG–fMRI fusion were observed after 700 ms post stimulus onset, resulting in faster data collection and increased participant comfort and compliance. Participants were instructed to respond to the image of a paper clip shown randomly every 3–5 trials (average 4) with an eye blink and a button press.

For fMRI, each participant completed 2 sessions. For Experiment 1, each session consisted of 10–14 runs of 384 s each, and for Experiment 2, each session consisted of 9–11 runs of 486 s each. Every image was presented once per run, and image order was randomized with the restriction that the same condition was not presented on consecutive trials. In both experiments, 25% of all trials were null trials, that is, 30 in Experiment 1 and 39 in Experiment 2. During null trials, only a gray background was presented, and the fixation cross turned darker for 100 ms. Participants were instructed to respond to the change in luminance with a button press. TOS was 3 s, or 6 s with a preceding null trial.

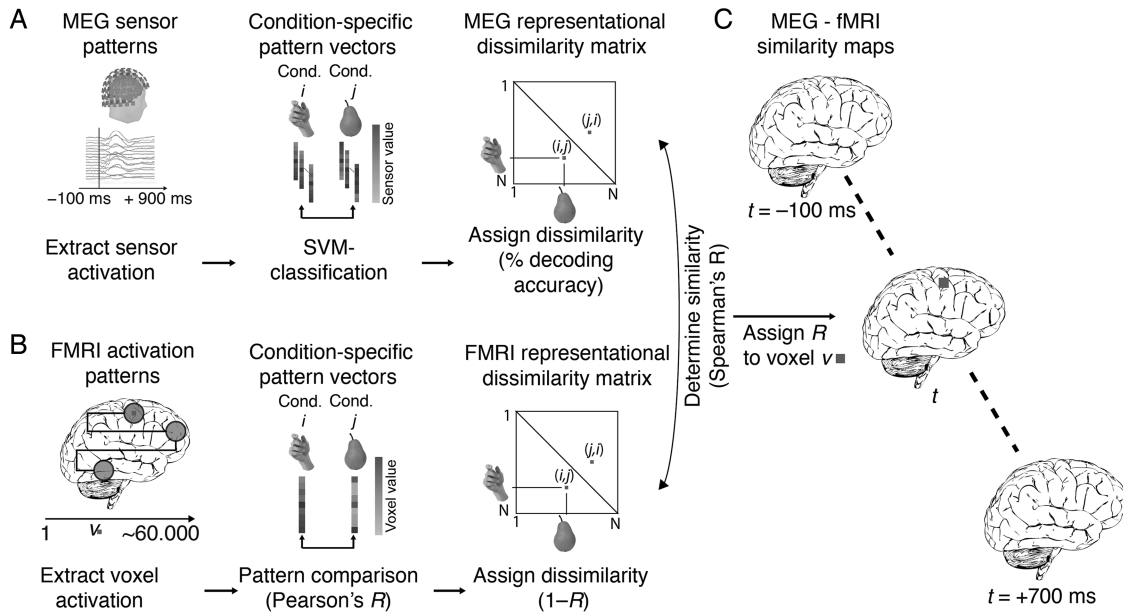
Note that the fMRI/MEG data acquired for Experiment 1 (as described below) has been published previously (Cichy et al. 2014), combining MEG and fMRI data in a region-of-interest-based analysis.

### MEG Acquisition and Analysis

We acquired MEG signals from 306 channels (204 planar gradiometers, 102 magnetometers, Elekta Neuromag TRIUX, Elekta, Stockholm, Sweden) at a sampling rate of 1 kHz, filtered between 0.03 and 330 Hz. We applied temporal source space separation (maxfilter software, Elekta, Stockholm; Taulu et al. 2004; Taulu and Simola 2006) before analyzing data with Brainstorm (Tadel et al. 2011). For each trial, we extracted peri-stimulus data from –100 ms to +700 ms, removed baseline mean, and smoothed data with a 20-ms sliding window. We obtained 20–30 trials for each condition, session, and participant.

Next, we determined the similarity relations between visual representations as measured with MEG by multivariate pattern analysis (Fig. 1). The rationale of using machine classifiers is that the better a classifier predicts condition labels based on patterns of MEG sensor measurements, the more dissimilar the MEG





**Figure 1.** Spatially unbiased MEG–fMRI fusion analysis scheme. (A) MEG analysis. We analyzed MEG data in a time point-specific fashion for each millisecond  $t$  from  $-100$  to  $+700$  ms with respect to stimulus onset. For each pair of conditions, we used support vector machine classification to determine how well conditions were discriminated by multivariate MEG sensor patterns. Repeating this procedure for each condition pair yielded a condition-by-condition ( $C \times C$ ) matrix of decoding accuracies, constituting a summary of representational dissimilarities for a time point (MEG representational dissimilarity matrix, MEG RDM). (B) fMRI analysis. We used searchlight analysis to reveal representational dissimilarity in locally constrained fMRI activity patterns. In detail, for each voxel  $v$  in the brain, we extracted fMRI patterns in its local vicinity (4-voxel radius) and calculated condition-wise dissimilarity ( $1 - \text{Spearman's } R$ ), resulting in a condition-by-condition ( $C \times C$ ) fMRI representational dissimilarity matrix (fMRI RDM), constituting a summary of representational dissimilarities for a voxel's vicinity. (C) fMRI–MEG fusion. In the space of representational dissimilarities, MEG and fMRI results become directly comparable. For each time-specific MEG RDM, we calculated the similarity (Spearman's  $R$ ) to each fMRI searchlight's fMRI RDM, yielding a 3-D map of MEG–fMRI fusion results. Repeating this procedure for each millisecond yielded a movie of MEG–fMRI correspondence in neural representations, revealing spatio-temporal neural dynamics.

patterns are and thus the underlying visual representations. As such, classifier performance can be interpreted as a dissimilarity measure. Here, we used a linear SVM classifier, which can automatically select MEG sensors that contain discriminative information robustly in noisy data. This obviates the need for an a priori human-based selection that might introduce a bias by favoring particular sensors. An analysis approach that would weigh the contribution of all channels equally, such as a correlation-based approach (see fMRI analysis below) might be less suitable, because it would be strongly influenced by noisy MEG channels. For example, in situations when comparatively few MEG channels contain information, noisy channels might strongly change the similarity relations between conditions in the MEG and thus impact the fusion with fMRI. Further, note that the fMRI searchlight used below does select features by constraining voxels to a local sphere for each searchlight, whereas in the MEG analysis all sensors (equivalent to all fMRI voxels) are used.

We conducted the multivariate analysis with linear SVMs independently for each subject and session. For each peri-stimulus time point from  $-100$  to  $+700$  ms, preprocessed MEG was extracted and arranged as 306 dimensional measurement vectors (corresponding to the 306 MEG sensors), yielding  $M$  pattern vectors per time point and condition (image). We then used supervised learning with a leave-one-trial out cross-validation to train a SVM classifier (LibSVM implementation, [www.csie.ntu.edu.tw/~cjlin/libsvm](http://www.csie.ntu.edu.tw/~cjlin/libsvm)) to pairwise discriminate any 2 conditions. For each time point and condition pair,  $M-1$  measurement vectors were assigned to the training set and used to train a support vector machine. The left out  $M$ th vectors for the 2 conditions were

assigned to a testing set and used to assess the classification performance of the trained classifier (percent decoding accuracy). The training and testing process was repeated 100 times with random assignment of trials to the training and testing set. Decoding results were averaged across iterations, and the average decoding accuracy was assigned to a matrix of size  $C \times C$  ( $C = 92$  for Experiment 1, 118 for Experiment 2), with rows and columns indexed by the classified conditions. The matrix was symmetric and the diagonal was undefined. This procedure yielded  $1 \times C$  matrix of decoding accuracies for every time point, referred to as MEG representational dissimilarity matrix (MEG RDM). For Experiment 2, measurement vectors were averaged by 5 before entering multivariate analysis to reduce computational load.

### fMRI Acquisition

MRI scans were conducted on a 3 T Trio scanner (Siemens, Erlangen, Germany) with a 32-channel head coil. In both experiments and each session, we acquired structural images using a standard  $T_1$ -weighted sequence (192 sagittal slices,  $\text{FOV} = 256 \text{ mm}^2$ ,  $\text{TR} = 1900 \text{ ms}$ ,  $\text{TE} = 2.52 \text{ ms}$ , flip angle =  $9^\circ$ ).

Functional data were collected with 2 different protocols. In Experiment 1 (partial brain coverage), data had high spatial resolution, but covered the ventral visual brain only. In each of 2 sessions, we acquired 10–14 runs of 192 volumes each for each participant (gradient-echo EPI sequence:  $\text{TR} = 2000 \text{ ms}$ ,  $\text{TE} = 31 \text{ ms}$ , flip angle =  $80^\circ$ ,  $\text{FOV read} = 192 \text{ mm}$ ,  $\text{FOV phase} = 100\%$ , ascending acquisition, gap = 10%, resolution = 2 mm isotropic, slices = 25). The acquisition volume covered the occipital and temporal lobe and was oriented parallel to the temporal cortex.

The second data set (Experiment 2) had lower spatial resolution, but covered the whole brain. We acquired 9–11 runs of 648 volumes for each participant (gradient-echo EPI sequence: TR = 750 ms, TE = 30 ms, flip angle = 61°, FOV read = 192 mm, FOV phase = 100% with a partial fraction of 6/8, through-plane acceleration factor 3, bandwidth 1816 Hz/Px, resolution = 3 mm<sup>3</sup>, slice gap 20%, slices = 33, ascending acquisition).

### fMRI Analysis

We preprocessed fMRI data using SPM8 (<http://www.fil.ion.ucl.ac.uk/spm/>). Analysis was identical for the 2 experiments. For each participant, fMRI data were realigned and co-registered to the T<sub>1</sub> structural scan acquired in the first MRI session. Then, MRI data were normalized to the standard MNI template. We used a general linear model (GLM) to estimate the fMRI response to the 92 (Experiment 1) or 118 (Experiment 2) image conditions. Image onsets and duration entered the GLM as regressors and were convolved with a hemodynamic response function. Movement parameters were included as nuisance parameters. Additional regressors modeling the 2 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.

We then analyzed fMRI data in a spatially unbiased approach using a searchlight analysis method (Kriegeskorte et al. 2006; Haynes et al. 2007). We processed each subject separately. For each voxel  $v$ , we extracted condition-specific t-value patterns in a sphere centered at  $v$  with a radius of 4 voxels (searchlight at  $v$ ) and arranged them into fMRI t-value pattern vectors. For each pair of conditions, we calculated the pairwise dissimilarity between pattern vectors by 1 minus Spearman's  $R$ , resulting in a  $92 \times 92$  (Experiment 1) or  $118 \times 118$  (Experiment 2) fMRI representational dissimilarity matrix (fMRI RDM) indexed in columns and rows by the compared conditions. fMRI RDMs were symmetric across the diagonal, and entries were bounded between 0 (no dissimilarity) and 2 (complete dissimilarity). This procedure resulted in 1 fMRI RDM for each voxel in the brain.

### MEG–fMRI Fusion with Representational Similarity Analysis

To relate neuronal temporal dynamics observed in MEG with their spatial origin as indicated by fMRI, we used representational similarity analysis (Kriegeskorte 2008; Kriegeskorte and Kievit 2013; Cichy et al. 2014). The basic idea is that if 2 images are similarly represented in MEG patterns, they should also be similarly represented in fMRI patterns. Comparing similarity relations in this way allows linking particular locations in the brain to particular time points, yielding a spatio-temporally resolved view of the emergence of visual representations in the brain.

We compared fMRI RDMs with MEG RDMs. For Experiment 1, analysis was conducted within subject, comparing subject-specific fMRI and MEG RDMs. Since only 15 subjects completed the fMRI task, out of the 16 that completed the MEG task,  $N$  was 15. For Experiment 2, only one MEG session was recorded, and to increase power, we first averaged MEG RDMs across participants before comparing the subject-averaged MEG RDMs with the subject-specific fMRI RDMs.

Further analysis was identical for both experiments and independent for each subject. For each time point, we computed the similarity (Spearman's  $R$ ) between the MEG RDM and the fMRI RDM of each voxel. This yielded a 3-D map of representational similarity correlations, indicating locations in the brain at which neuronal processing emerged at a particular time point.

Repeating this analysis for each time point yielded a spatio-temporally resolved view of neural activity in the brain during object perception.

### Spatio-Temporally Unbiased Analysis of Category-Related Spatial Effects

We investigated the spatio-temporal neural dynamics with which neural activity indicated category membership of objects. The analysis was based on data from Experiment 1, where the image set showed a clear categorical structure at supra-category, category, and sub-category level. We investigated 5 categorical divisions: 1) animacy (animate vs. inanimate), 2) naturalness (natural vs. man-made), 3) faces versus bodies, and 4,5) human versus animal faces and bodies.

To determine the spatial extent of category-related neural activity (Fig. 6A), we conducted a searchlight analysis comparing (Spearman's  $R$ ) fMRI RDMs to theoretical model RDMs capturing category membership, that is, higher dissimilarity between patterns for objects across category boundary than within category boundary (modeled as RDM cell entries 1 and 0, respectively). Analysis was conducted independently for each subject (comparing subject-specific fMRI and MEG RDMs) and yielded 5 categorical division-specific 3-D maps.

To visualize the spatio-temporal extent of category-related activity, we repeated the MEG–fMRI fusion analysis for each categorical division limited to the parts of the RDMs related to the division. We also masked the results of this category-related MEG–fMRI fusion approach with the results of the category-related fMRI analysis. This yielded a spatio-temporally resolved view of neural activity in the brain during object perception of a particular categorical division.

### Definition of Regions of Interest

To characterize the spatial extent of the MEG–fMRI fusion results and enable region-of-interest-based MEG–fMRI analysis, we obtained visual regions in occipital, parietal, and temporal cortex from a probabilistic atlas (Wang et al. 2015). Since visual stimuli were presented at 3°–4° visual angle, they fell largely into the foveal confluence confounding the distinction between V1, V2, and V3. We thus combined regions V1 to V3 into a single ROI termed early visual cortex (EVC), also easing visualization. Additional regions included 6 ROIs in the dorsal visual stream (intraparietal sulcus: IPS0, IPS1, IPS2, IPS3, IPS4, and IPS5) and 5 ROIs in the ventral visual stream (h4, ventral occipital cortex VO1&2, temporal occipital cortex TO1&2, lateral-occipital cortex LO1&2 and parahippocampal cortex PHC1&2). Unique ROIs were created from the probabilistic map by assigning each voxel to the ROI of highest probability, conditional the aggregate probability over all ROIs was  $\geq 33\%$ . Due to limited fMRI coverage, ROIs for parietal regions could only be defined for Experiment 2.

### Region-of-Interest-Based MEG–fMRI Fusion

We conducted the MEG–fMRI fusion analysis using fMRI voxel patterns restricted to specific ROIs. For each subject and ROI, we constructed an fMRI RDM by comparing condition-specific voxel activation patterns in dissimilarity (1–Spearman's  $R$ ), equivalent to the searchlight-based analysis. This yielded 1 fMRI RDM for each ROI and subject.

The fusion of MEG with fMRI in the ROI approach followed the approach in Cichy et al. (2014). In detail, for each ROI we averaged the subject-specific fMRI RDMs and calculated the similarity between the average-fMRI RDM and the subject-specific MEG RDMs

for each time point (Spearman's  $R$ ), yielding time courses of MEG–fMRI similarity for each subject ( $n = 15$ ), in each ROI.

### Statistical Testing

For the searchlight-based analyses, we used permutation tests for random-effects inference and corrected results for multiple comparisons with a cluster-level correction (Nichols and Holmes 2002; Pantazis et al. 2005; Maris and Oostenveld 2007). In detail, to determine a cluster-definition threshold, we averaged the 4-dimensional (4-D) subject-specific data across subjects (3-D space  $\times$  1-D time), yielding a mean 4-D result matrix. We then constructed an empirical baseline voxel distribution by aggregating voxel values across space and time points from  $-100$  to  $0$  ms of the mean result matrix and determined the right-sided 99.9% threshold. This constituted a baseline-based cluster-definition threshold at  $P = 0.001$ .

To obtain a permutation distribution of maximal cluster size, we randomly shuffled the sign of subject-specific data points (1000 times), averaged data across subjects, and determined 4-D clusters by spatial and temporal contiguity at the cluster-definition threshold. Storing the maximal cluster statistic (size of cluster with each voxel equally weighed) for each permutation sample yielded a distribution of the maximal cluster size under the null hypothesis. We report clusters as significant if they were greater than the 99% threshold constructed from the maximal cluster size distribution (i.e., cluster size threshold at  $P = 0.01$ ).

For the analyses of category-related spatial signals in Experiment 1, we conducted 5 independent statistical analyses, one for each categorical subdivision. We randomly shuffled the sign of subject-specific data points (10 000 times) and determined 3-D clusters by spatial contiguity at a cluster-definition threshold at  $P = 0.001$  1-sided on the right tail. We report clusters as significant with a cluster size threshold at  $P = 0.01$ .

For the region-based MEG–fMRI fusion analysis, we used sign-permutation tests and cluster-based correction for multiple comparisons. In detail, we randomly shuffled the sign of subject-specific data points (1000 times) and determined 1-D clusters by temporal contiguity at a cluster-definition threshold at  $P = 0.05$  1-sided. We report clusters as significant with a cluster size threshold at  $P = 0.05$ .

To calculate 95% confidence intervals for the peak latency and onsets of the first significant cluster in the ROI-based MEG–fMRI fusion analysis, we created 1000 bootstrapped samples with replacement. For each sample, we recalculated the peak latency and onset of the first significant cluster, yielding empirical distributions for both. This resulted in bootstrap estimates of both statistics and thus their 95% confidence intervals.

To determine significance in peak latency differences for MEG–fMRI time courses of different ROIs, we used bootstrapping. For this, we created 1000 bootstrapped samples by sampling participants with replacement. This yielded an empirical distribution of mean peak latency differences. Setting  $P < 0.05$ , we rejected the null hypothesis if the 95% confidence interval did not include 0. Results were FDR-corrected for multiple comparisons.

## Results

### MEG–fMRI Fusion in Representational Similarity Space

In 2 independent experiments, participants viewed images of real-world objects: 92 objects on a gray background in Experiment 1 (Fig. 2A), and 118 objects on a real-world background in Experiment 2 (Fig. 4A). Each experiment consisted of separate recording sessions presenting the same stimulus material during MEG or fMRI data collection. Images were presented in random

order in experimental designs optimized for each imaging modality. Subjects performed orthogonal vigilance tasks.

To integrate MEG with fMRI data, we abstracted from the sensor space of each imaging modality to a common similarity space, defined by the similarity of condition-specific response patterns in each modality. Our approach relies on representational similarity analysis (Kriegeskorte 2008; Cichy et al. 2014) and effectively constructs MEG and fMRI representational similarity relations that are directly comparable.

In detail, for MEG we analyzed data in a time point-specific fashion for each millisecond from  $-100$  to  $+1000$  ms with respect to stimulus onset (Fig. 1). To make full use of the information encoded at each time point across MEG sensors, we used multivariate pattern analysis (Carlson, Tovar, et al. 2013; Cichy et al. 2014; Clarke et al. 2015; Isik et al. 2014). Specifically, for each time point, we arranged the measurements of the entire MEG sensor array into vectors, resulting in different pattern vectors for each condition, stimulus repetition, and time point. These pattern vectors were the input to a support vector machine procedure designed to classify different pairs of conditions (images). Repeating the classification procedure for all pairs of conditions yielded a condition  $\times$  condition ( $C \times C$ ) matrix of classification accuracies, one per time point. Interpreting classification accuracies as a representational dissimilarity measure, the  $C \times C$  matrix summarizes for each time point which conditions are represented similarly (low decoding accuracy) or dissimilarly (high decoding accuracy). The matrix is thus termed MEG representational dissimilarity matrix (MEG RDM).

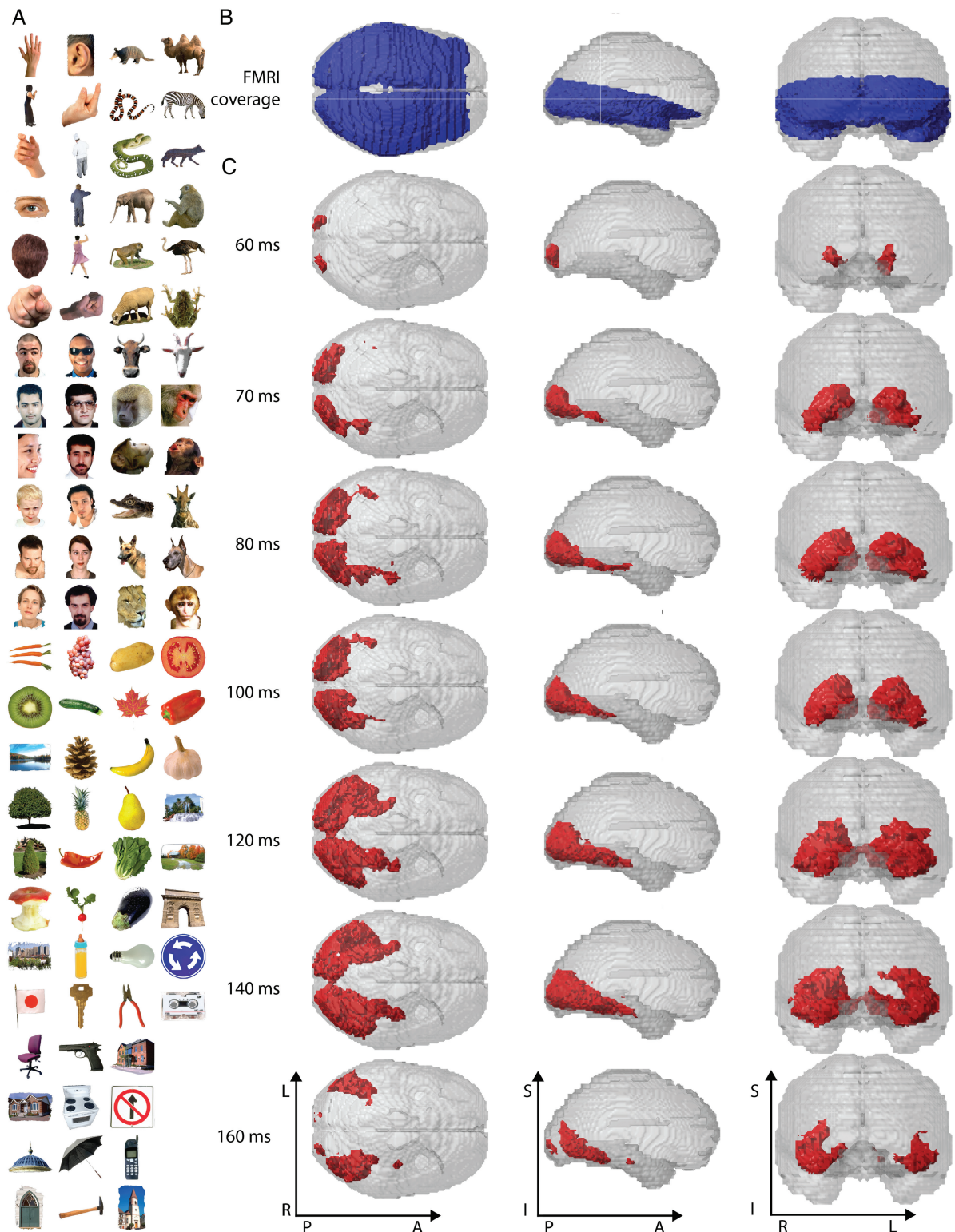
For fMRI, we used a searchlight approach (Haynes and Rees 2005; Kriegeskorte et al. 2006) in combination with multivariate pattern similarity comparison (Kriegeskorte 2008; Kriegeskorte et al. 2008; Kriegeskorte and Kievit 2013) to extract information stored in local neural populations resolved with the high spatial resolution of fMRI. For every voxel in the brain, we extracted condition-specific voxel patterns in its vicinity and calculated pairwise dissimilarity (1 minus Pearson's  $R$ ) between voxel patterns of different conditions. Repeating this procedure for all pairs of conditions yielded a  $C \times C$  fMRI representational dissimilarity matrix (fMRI RDM) summarizing for each voxel in the brain which conditions are represented similarly or dissimilarly.

Thus, MEG and fMRI data became directly comparable to each other in similarity space via their RDMs, allowing integration of their respective high temporal and spatial resolution. For a given time point, we compared (Spearman's  $R$ ) the corresponding MEG RDM with each searchlight-specific fMRI RDM, storing results in a 3-D volume at the location of the center voxel of each searchlight. Each time-specific 3-D volume indicated where in the brain the fMRI representations were similar to the ongoing MEG representations. Repeating this procedure for all time points yielded a set of brain volumes that unraveled the spatio-temporal neural dynamics underlying object recognition. We corrected for multiple comparisons across space and time by 4-D cluster correction (cluster-definition threshold  $P < 0.001$ , cluster size threshold  $P < 0.05$ ).

### MEG–fMRI Fusion Reveals Successive Cortical Region Activation in the Ventral and Dorsal Visual Stream During Visual Object Recognition

We applied the MEG–fMRI fusion method to data from Experiment 1, that is, visual responses to a set of  $C = 92$  images of real-world objects (Fig. 2A). fMRI data had only a partial brain coverage of occipital and temporal cortex (Fig. 2B). Results revealed the spatio-temporal dynamics in the ventral visual pathway in the first few hundred milliseconds of visual object



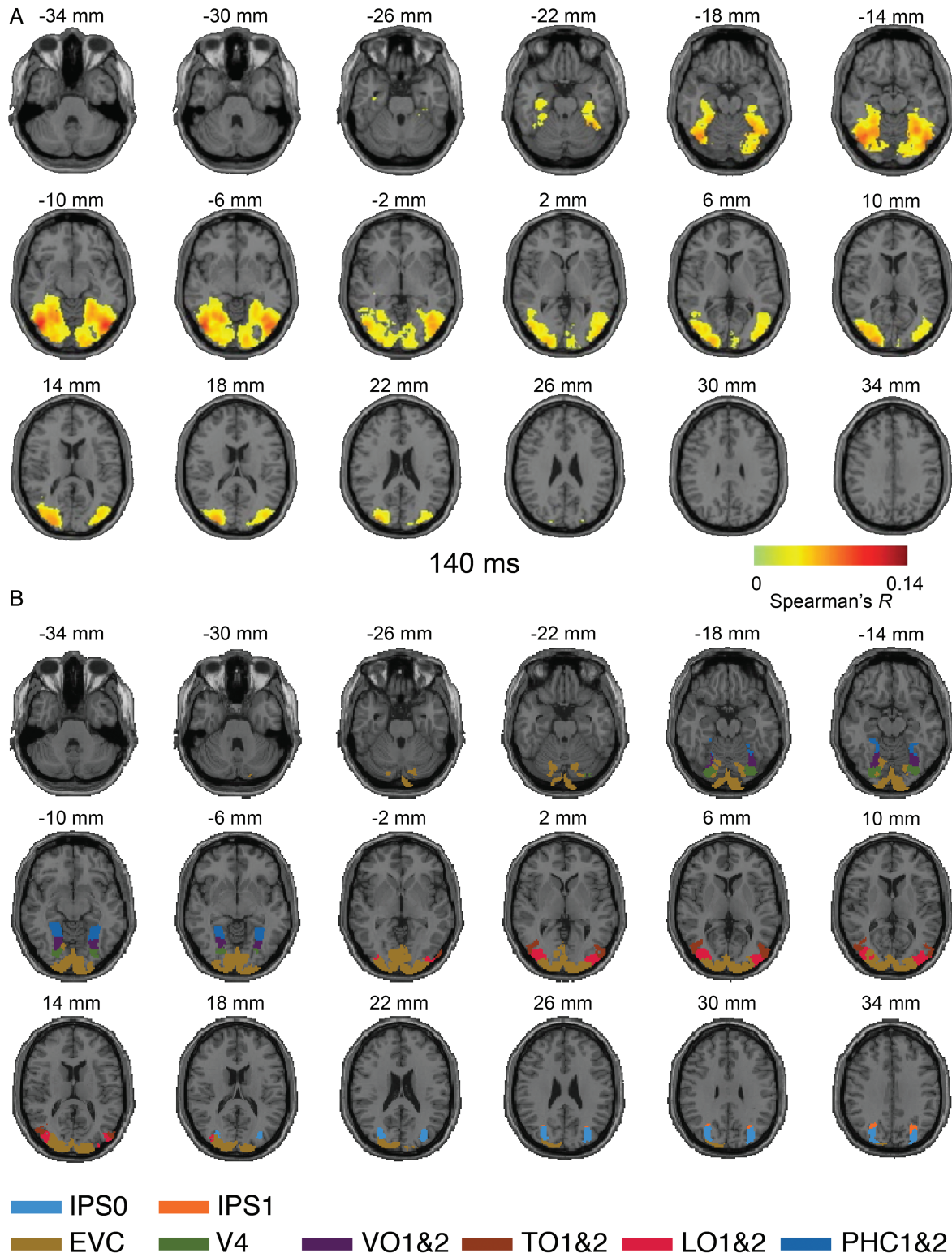


**Figure 2.** Stimulus material, fMRI brain coverage, and significant MEG–fMRI fusion results for Experiment 1. (A) The stimulus set consisted of 92 cropped objects, including human and animal bodies and faces, as well as natural and artificial objects. During the experiment, images were shown on a uniform gray background (shown on white background here for visibility). (B) We recorded fMRI with partial brain coverage (blue regions), reflecting the trade-off between high-resolution (2 mm isovoxel), extent of coverage, and temporal sampling ( $TR = 2s$ ). (C) MEG–fMRI fusion revealed spatio-temporal neural dynamics in the ventral visual stream, originating at the occipital lobe and extending rapidly in anterior direction along the ventral visual stream. Red voxels indicate statistical significance ( $n = 15$ , cluster-definition threshold  $P < 0.001$ , cluster threshold  $P < 0.01$ ). A millisecond resolved movie is available as [Supplementary Movie 1](#).



perception. We observed the emergence of neuronal activity with onset around approximately 50–60 ms at the occipital pole, succeeded by rapid activation in the anterior direction into the temporal lobe (Fig. 2C, for ms-resolved, see [Supplementary Movie 1](#)).

A projection of the results onto axial slices at 140 ms after stimulus onset exemplifies the results (Fig. 3, for ms-resolved, see [Supplementary Movie 2](#)). Supplementary MEG–fMRI fusion with MEG RDMs averaged across subjects, rather than subject specific as



**Figure 3.** Axial slices exemplifying the MEG–fMRI fusion results for Experiment 1 at 140 ms. (A) MEG–fMRI representation correlations were projected onto axial slices of a standard  $T_1$  image in MNI space. The results indicate neural activity reaching far into the ventral visual stream in temporal cortex. Color-coded voxels indicate strength of MEG–fMRI representation correlations (Spearman's  $R$ , scaled between 0 and maximal observed value,  $n = 15$ , cluster-definition threshold  $P < 0.001$ , cluster threshold  $P < 0.01$ ). A millisecond resolved movie is available as [Supplementary Movie 2](#). (B) Rendering of ventral and dorsal regions of interest from a probabilistic atlas ([Wang et al. 2015](#)).

above, yielded qualitatively similar results (for ms-resolved results, see [Supplementary Movies 3 and 4](#)).

Due to the limited MRI coverage in Experiment 1, spatio-temporal dynamics beyond the ventral visual cortex, and in particular the dorsal visual stream, could not be assessed ([Andersen et al. 1987](#); [Serenio and Maunsell 1998](#); [Serenio et al. 2002](#); [Denys et al. 2004](#); [Lehky and Sereno 2007](#); [Janssen et al. 2008](#); [Konen and Kastner 2008](#)). To provide a full-brain view of spatio-temporal neuronal dynamics during object recognition and to assess the robustness of the MEG–fMRI fusion method to changes in experimental parameters such as image set, timing, recording protocol, we conducted Experiment 2 with full brain MRI coverage (Fig. 4B) and a different, more extended, set of object images ( $C = 118$ ; Fig. 4A). We again observed early neuronal activity in the occipital pole, though with a somewhat slower significant onset (~70–80 ms), and rapid successive spread in the anterior direction across the ventral visual stream into the temporal lobe (Fig. 4C; see [Supplementary Movie 5](#)), reproducing the findings of Experiment 1. Crucially, we also observed a progression of neuronal activity along the dorsal visual pathway into the inferior parietal cortex from approximately 140 ms onward. Figure 5 shows example results at 180 ms projected onto axial slices (for ms-resolved, see [Supplementary Movie 6](#)).

Together, our results describe the spatio-temporal dynamics of neural activation underlying visual object recognition during the first few hundred milliseconds of vision as a cascade of cortical region activations in both the ventral and dorsal visual pathways.

### Spatio-Temporally Unbiased Analysis of Category-Related Effects

The results of the MEG–fMRI fusion approach indicated the overall spatio-temporal neural dynamics during object recognition, but offered no information on the content of the emerging neural activity. To investigate, we determined object category-related spatio-temporal dynamics using data from Experiment 1, which had an image set with a clear categorical structure at supra-category, category, and sub-category level in 5 subdivisions: 1) animacy (animate vs. inanimate), 2) naturalness (natural vs. man-made), 3) faces versus bodies, and 4,5) human versus animal faces and bodies.

We first determined the spatial extent of category-related neural activity. For this, we created theoretical model RDMs capturing category membership, that is, higher dissimilarity between patterns for objects across category boundary than within category boundary (modeled as RDM cell entries 1 and 0, respectively). We conducted a searchlight analysis, determining the similarity between each model RDM and the searchlight RDMs. This yielded five 3-D maps indicating where category-related signals emerged in the brain. Across-subject results (sign-permutation test, cluster-definition threshold  $P < 0.001$ , cluster threshold  $P < 0.01$ ) are reported in volume in Figure 6 and projected onto axial slices in Figure 7.

We found significant clusters for all categorical subdivisions, except naturalness. All clusters extended from early visual cortex up to parahippocampal cortex for animacy and into lateral-temporal-, and ventral-occipital cortex for the remaining categories.

To visualize the spatio-temporal extent of category-related neural activity, we repeated the MEG–fMRI fusion analysis for each categorical subdivision, limiting the comparison of MEG and fMRI RDMs to the parts relevant to the categorical division. We then masked the results of the category-related MEG–fMRI fusion with the results of the category-related fMRI analysis. This

analysis showed the dynamics for category-related activity for the animacy, the face versus body, and the human versus animal face subdivisions, and is visualized in [Supplementary Movie 7](#) in volume, and [Supplementary Movie 8](#) projected onto axial slices. In sum, these results describe the spatio-temporal dynamics of neural activation related to object category in the ventral visual stream.

### Parietal and Ventral Cortices Exhibited MEG–fMRI Correspondence in Representations Significantly Later than Early Visual Cortex

The results of the MEG–fMRI fusion analysis indicated a progression of activity over time in the dorsal and ventral stream. To further investigate the parietal and ventral activity progression, that is, its precise anatomical extent and relation to activity in early visual cortex, we conducted a region-of-interest-based analyses. For this, we determined the time course with which ROI-specific fMRI RDMs correlated with MEG RDMs over time (Fig. 8A). We determined statistical significance by sign-permutation tests ( $P < 0.05$  cluster-definition threshold,  $P < 0.05$  cluster threshold). For both experiments, corroborating the findings of the searchlight-based MEG–fMRI analysis, significant clusters were found in EVC (Fig. 8B) and all investigated ventral visual areas (Fig. 8C, TO1&2, VO1&2, LO1&2, PHC1&2) for Experiments 1 and 2 (Fig. 6C,D), as well as IPS0, 1, and 2, but not IPS3, 4, or 5 (Fig. 8D). 95% confidence intervals on the onset of the first significant peak and the peak latency for each ROI are reported in Table 1.

We further characterized the nature of the neural activity in the ventral and parietal cortex by determining whether parietal and ventral cortex showed MEG–fMRI correspondence in representations at the same or different time points compared with activity in EVC, and to each other, by comparing peak-onset latencies. Compared with EVC, MEG–fMRI correspondence in representations emerged significantly later in IPS0 and 2 (all  $P < 0.01$ , details in Table 2A). Peak latency in IPS1 was not significantly different from EVC, likely due to the low SNR. For the ventral stream, corroborating [Cichy et al. \(2014\)](#), convergent results across experiments showed that all regions (except LO1&2 for Experiment 1) peaked later than EVC (all  $P < 0.01$ , details in Table 2B). However, we did not find a difference in peak latency between any of the ventral regions and IPS0, 1, or 2 except IPS2 and TO1&2 ( $P = 0.011$ , all other  $P > 0.12$ ). This indicates that neural activity in the dorsal and ventral stream occurred later than in early visual cortex during object vision.

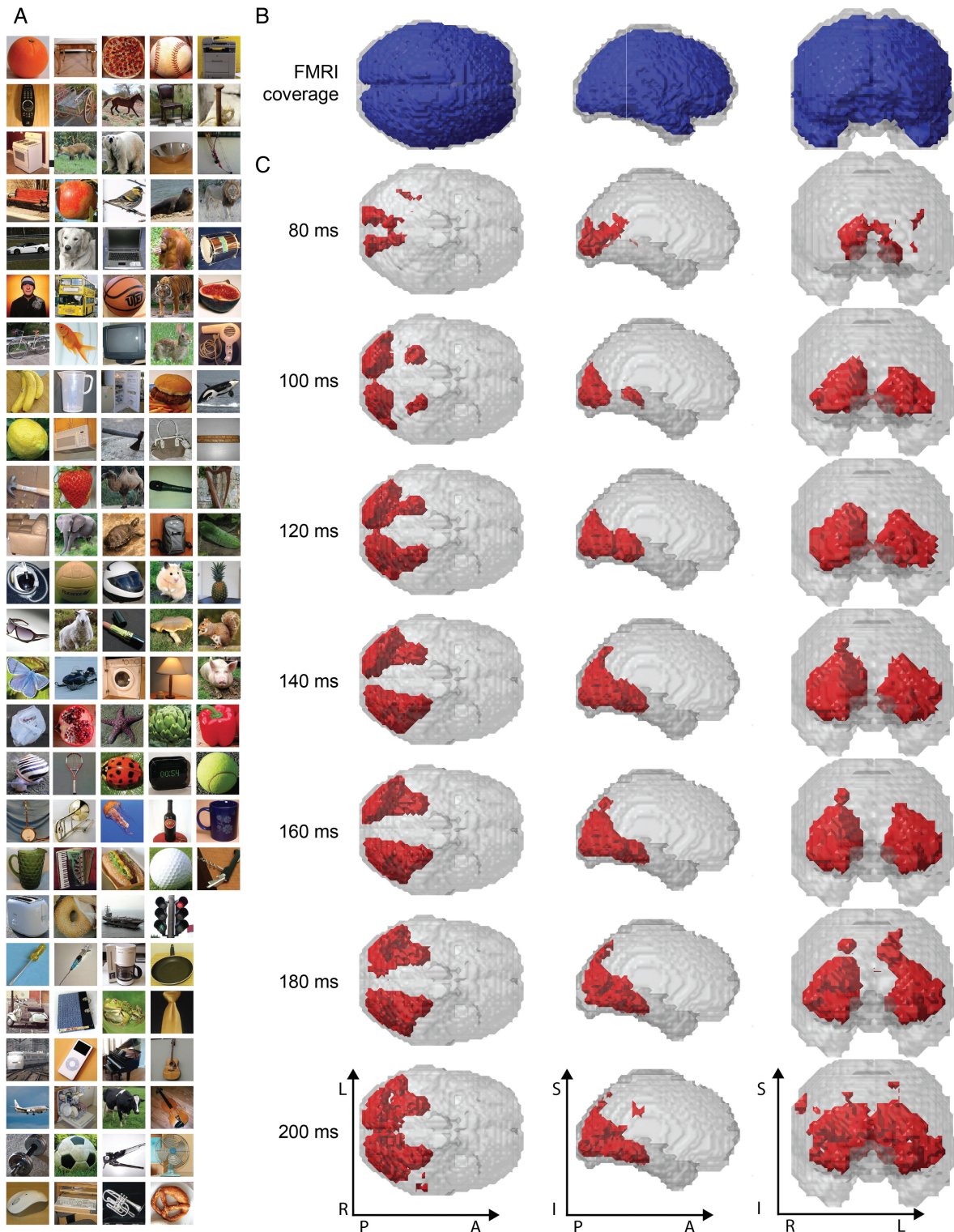
Finally, visually the temporal duration of significant clusters seemed to be larger for ROIs in the ventral compared with the dorsal progression, possibly indicating different temporal dynamics. However, quantitative analysis by bootstrapping the participant pool (1000 times) did not show statistical differences in the temporal duration of the ventral and dorsal progression (in 2 definitions: either the length of the first significant cluster or temporal duration from the first to the last significant time point for all clusters).

## Discussion

### Summary

To resolve spatio-temporal neural dynamics in the human brain during visual object recognition, we proposed a MEG–fMRI fusion method based on representational similarities and multivariate pattern analysis. The locus of activation onset was at the occipital pole, followed by rapid and progressive activation along the

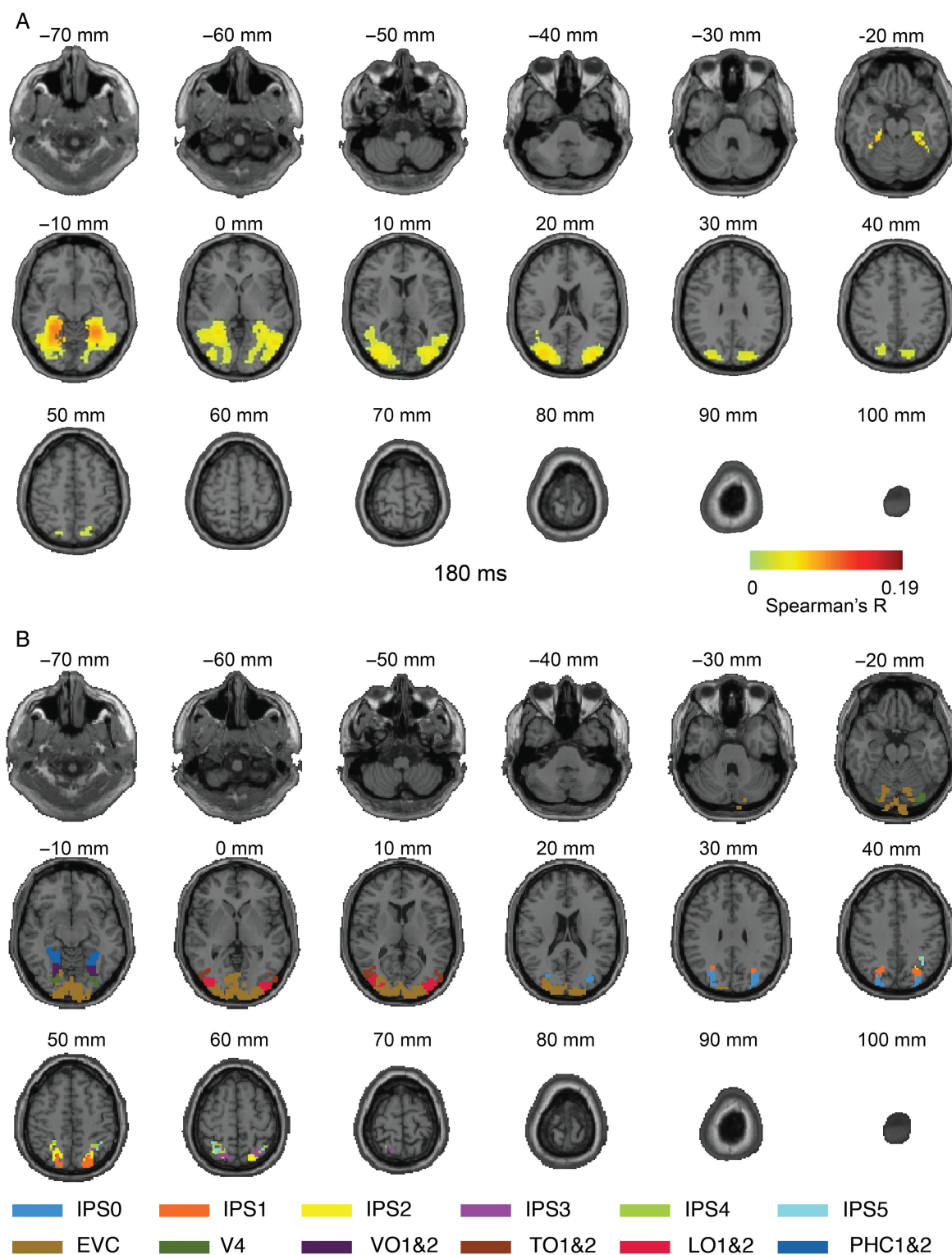




**Figure 4.** Stimulus material, brain coverage, and significant MEG–fMRI fusion results for Experiment 2. (A) The stimulus set consisted of 118 objects on natural backgrounds. (B) We recorded fMRI with a whole-brain coverage (blue regions). (C) Whole-brain analysis revealed neural dynamics along the dorsal visual stream into parietal cortex in addition to the ventral visual stream. Red voxels indicate statistical significance ( $N = 15$ , cluster-definition threshold  $P < 0.001$ , cluster threshold  $P < 0.01$ ). A millisecond resolved movie is available as [Supplementary Movie 5](#).

processing hierarchies of the dorsal and ventral visual streams into temporal and parietal cortex, respectively. In the ventral visual stream, further analysis revealed the spatio-temporal extent of category-related neural signals. Further, we found that dorsal

and ventral regions exhibited MEG–fMRI correspondence in representations later than early visual cortex. This result provides a comprehensive and refined view of the complex spatio-temporal neural dynamics underlying visual object recognition



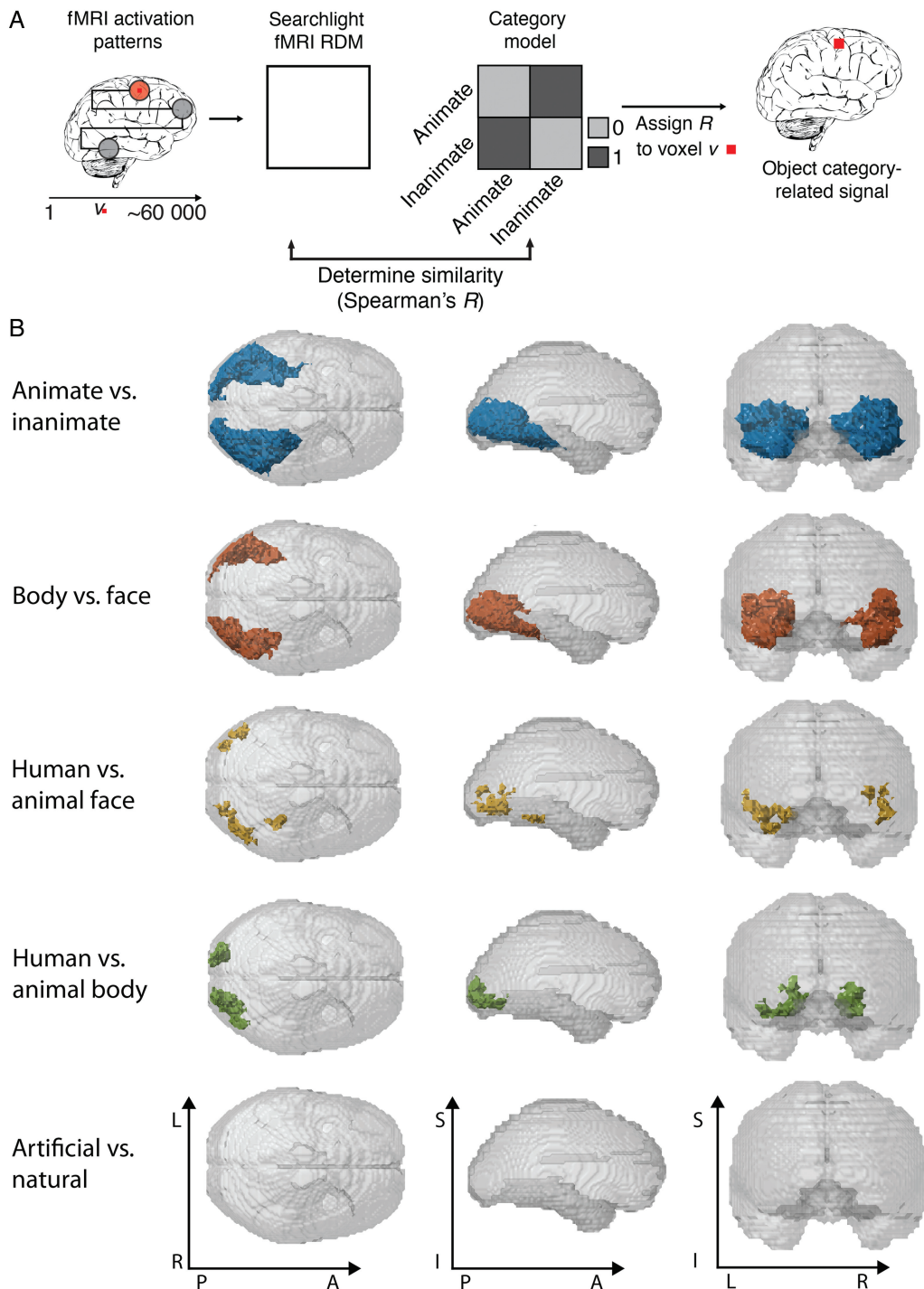
**Figure 5.** Axial slices exemplifying the MEG-fMRI fusion results for Experiment 2 at 180 ms. (A) MEG-fMRI representation correlations were projected onto axial slices of a standard T<sub>1</sub> image in MNI space. In addition to ventral cortex activation, reproducing Experiment 1, we also observed activation far up the dorsal visual stream in inferior parietal cortex. Color-coded voxels indicate strength of MEG-fMRI representation correlations (Spearman's R, scaled between 0 and maximal value observed,  $n = 15$ , cluster-definition threshold  $P < 0.001$ , cluster threshold  $P < 0.01$ ). A millisecond resolved movie is available as [Supplementary Movie 6](#). (B) Rendering of ventral and dorsal regions of interest from a probabilistic atlas.

in the human brain. It further suggests that the proposed methodology is a strong analytical tool for the study of any complex cognitive function.

#### Tracking Neural Activity During Visual Object Recognition

Our results revealed the spatio-temporal neural dynamics during visual object recognition in the human brain, with neuronal



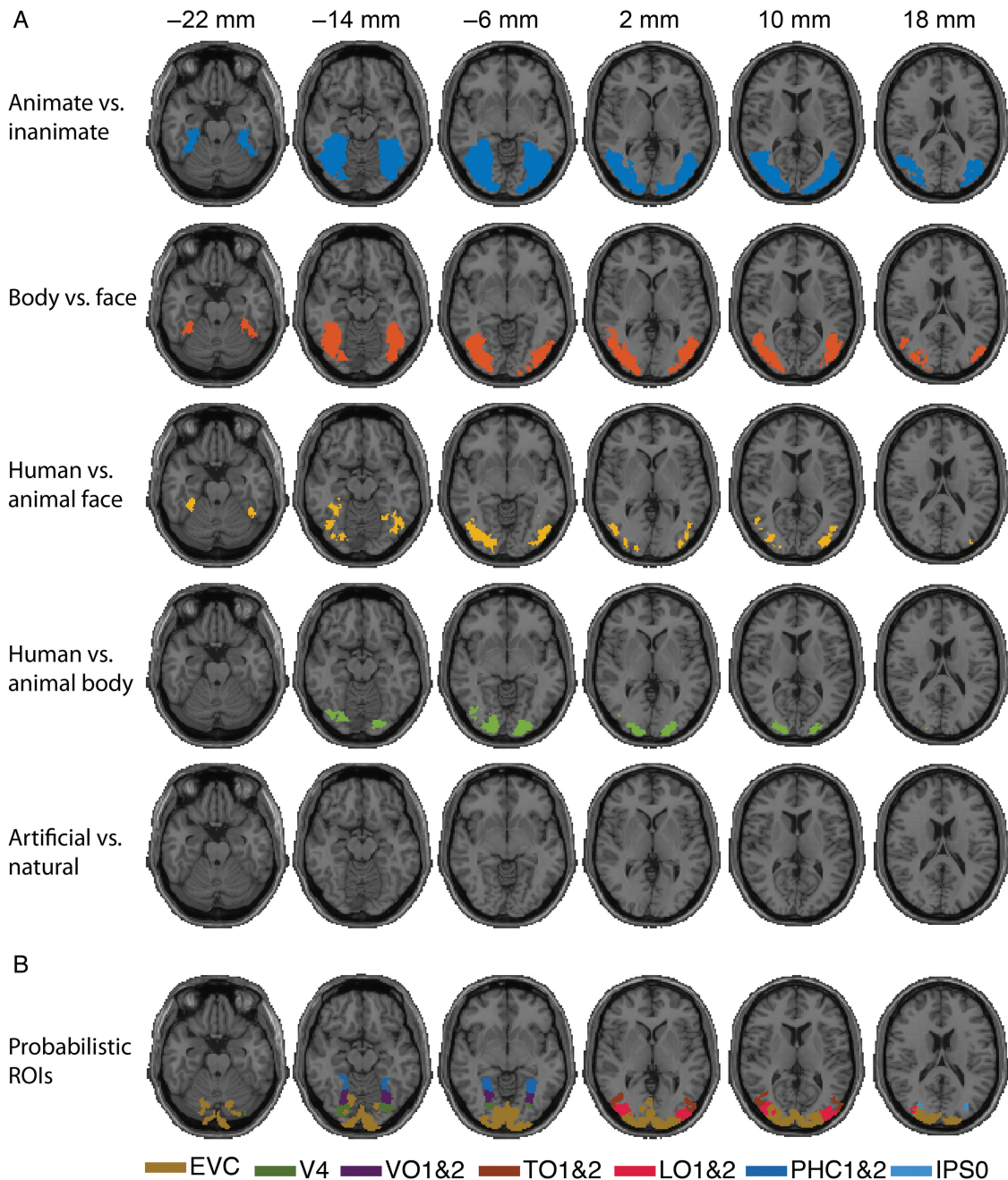


**Figure 6.** Analysis scheme and results of fMRI-category model fusion for Experiment 1. (A) Analysis scheme for fMRI-categorical model fusion. For each investigated category subdivision, we created a model RDM capturing category membership, with cell entries 1 and 0 for objects between- and within-category, respectively (here shown for animacy). For each voxel, we determined the similarity between the corresponding searchlight RDM and the model RDM, yielding 3-D maps of category-related signals in the brain. (B) fMRI-category model fusion results for each of the 5 categorical divisions investigated. ( $n = 15$ , cluster-definition threshold  $P < 0.001$ , cluster threshold  $P < 0.01$ ).

activation emerging in the occipital lobe and rapidly spreading along the ventral and dorsal visual pathways. Corroborating results in 2 independent experiments differing in stimulus material, experimental protocol, and fMRI acquisition parameters, among others, demonstrated the robustness of the MEG-fMRI fusion approach. In particular, the finding of a temporal progression from

early visual areas to high-level ventral visual areas over time is a direct reproduction across experiments, and corroborates a previous study (Cichy et al. 2014).

Beyond providing corroborative evidence and reproduction, the results presented here go further in both novel insight and methodological approach. Cichy et al. (2014) limited analysis to



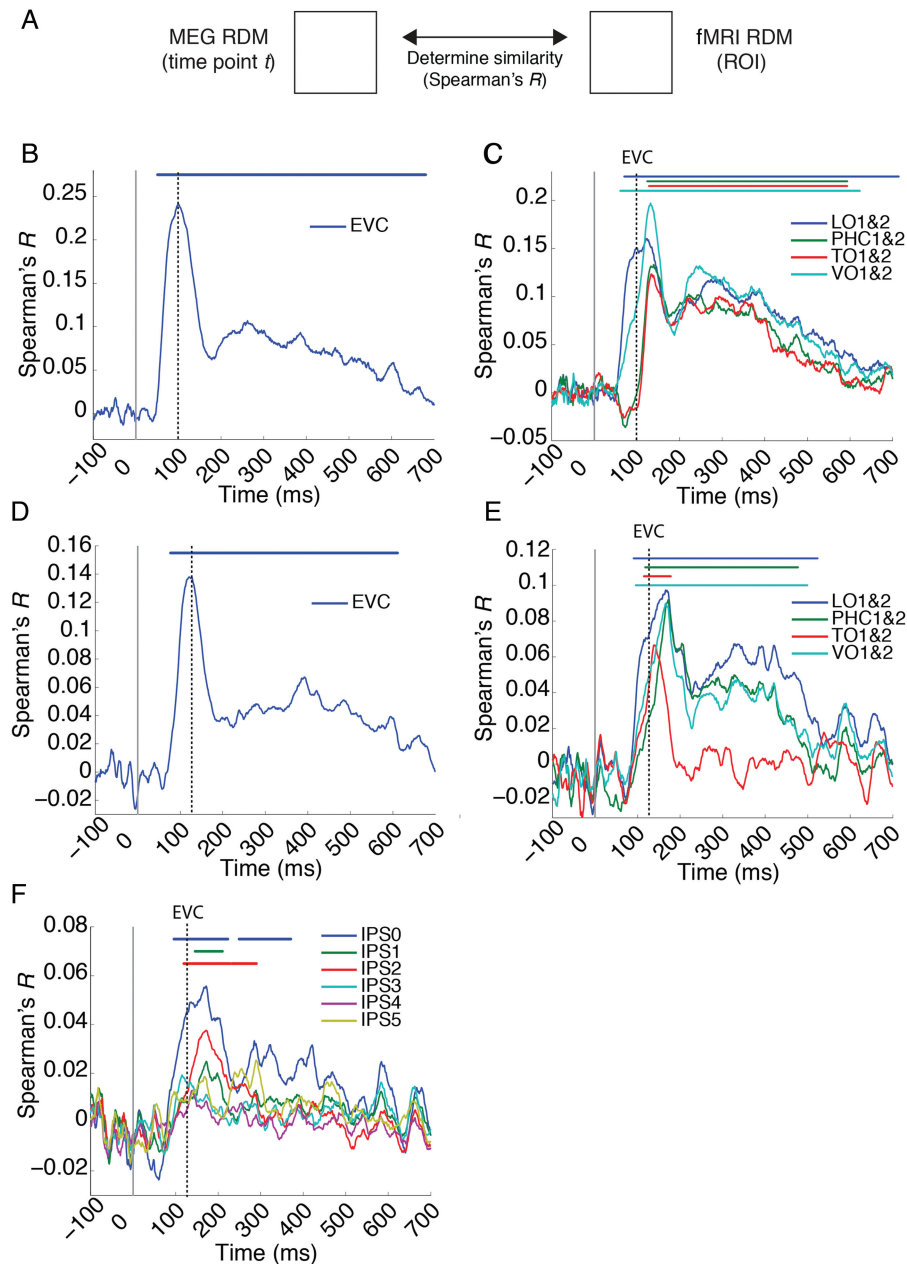
**Figure 7.** Axial slices displaying the results of fMRI-category model fusion for Experiment 1 and its relation to a probabilistic atlas of visual regions. (A) Same as Figure 6 in axial slices view. Colored voxels (same color code as in Fig. 6) indicate statistical significance ( $n = 15$ , cluster-definition threshold  $P < 0.001$ , cluster threshold  $P < 0.01$ ). (B) Visualization of ventral and dorsal regions from a probabilistic atlas (Wang et al. 2015).

the ventral visual cortex as the major pathway involved in object recognition, in particular regions V1 and IT. Here, adding an additional data set, we extended the analysis to the study of the dorsal visual pathway and elucidated the relation between object representations in ventral and dorsal pathway. Second, whereas Cichy et al. (2014) investigated the emergence of category-related activity over time using MEG only, here we extend the analysis to the spatial dimension and the spatio-temporal dynamics using MEG-fMRI fusion.

Concerning methodological advance, here we showed that MEG-fMRI fusion by representational similarity is not restricted to a spatially limited region-of-interest analysis, but extendable

to a spatially unbiased analysis. This strongly widens the applicability of MEG-fMRI fusion to the pursuit of scientific questions without spatial priors.

Together, our results provide evidence that object recognition is a hierarchical process unfolding over time (Robinson and Rugg 1988; Schmolesky et al. 1998; Bullier 2001; Cichy et al. 2014) and space (Ungerleider and Mishkin 1982; Felleman and Van Essen 1991; Grill-Spector and Malach 2004; Wandell et al. 2007; Op de Beeck et al. 2008; Kravitz et al. 2011; Grill-Spector and Weiner 2014) in both dorsal and ventral streams (Ungerleider and Mishkin 1982; Milner and Goodale 2008; Kravitz et al. 2011).



**Figure 8.** Relationship between dorsal and ventral visual stream activity. (A) Region-of-interest-based MEG–fMRI fusion analysis. For each ROI, we constructed an fMRI RDM and compared it (Spearman's  $R$ ) with time point-specific MEG RDMs for Experiment 1 (B and C) and Experiment 2 (D–F). This yielded time courses of MEG–fMRI correspondence in (B,D) early visual cortex (EVC) (C,E) ventral stream regions LO1&2, TO1&2, VO1&2, and PHC1&2, and (F) parietal regions IPS0 to 5. Solid gray line indicates stimulus onset. Dashed gray lines indicate EVC peak latency for Experiment 1 (B and C) and 2 (D and E). Significant time points (horizontal lines above plots) were identified for all cases other than the parietal regions IPS3 to 5.

### Tracking Neural Activity During Visual Object Recognition: The Ventral Visual Stream

The time course of neural activity revealed by MEG–fMRI fusion largely concurred with previously reported increases of response latencies along the posterior–anterior gradient in the ventral pathway in humans (Mormann et al. 2008; Liu et al. 2009) and in monkeys (Schmolecky et al. 1998; Bullier 2001). This provides further evidence that object recognition is a hierarchical process unfolding over time (Robinson and Rugg 1988; Schmolecky et al. 1998; Bullier 2001; Cichy et al. 2014) and space (Ungerleider and Mishkin 1982; Felleman and Van Essen 1991; Grill-Spector and

Malach 2004; Wandell et al. 2007; Op de Beeck et al. 2008; Kravitz et al. 2011; Grill-Spector and Weiner 2014).

To elucidate the contents of the spatio-temporal neural dynamics uncovered by the MEG–fMRI approach in the ventral visual stream, we investigated whether the reported signals indicated category membership of objects. Our analysis revealed extended spatio-temporal effects for the categorical division of animacy, faces versus bodies, and human versus animal faces. This complements previous findings about the temporal dynamics of category processing in the human brain for the same categorical divisions (Carlson, Tovar, et al. 2013; Cichy et al. 2014) by providing a view that is both spatial and temporal of the neural dynamics.



**Table 1** Mean peak and onset latency of the first significant cluster in (A) early visual cortex, (B) dorsal, and (C) ventral regions for Experiments 1 and 2

Region of interest	Experiment 1		Experiment 2	
	Peak latency (ms)	Onset of significance (ms)	Peak latency (ms)	Onset of significance (ms)
(A) Early visual regions	100 (82–119)	51 (46–55)	127 (110–134)	77 (68–84)
(B) Dorsal regions				
IPS0			172 (117–204)	96 (78–212)
IPS1			172 (154–246)	146 (98–276)
IPS2			172 (154–204)	119 (100–151)
(C) Ventral regions				
LO1&2	112 (79–127)	56 (51–62)	166 (147–326)	91 (83–96)
VO1&2	120 (128–285)	111 (100–119)	166 (155–203)	96 (155–302)
TO1&2	134 (128–355)	114 (107–119)	139 (136–155)	115 (88–124)
PHC1&2	139 (128–285)	111 (100–119)	172 (165–207)	119 (97–147)

Note: Dorsal regions only reported for Experiment 2 due to limited fMRI coverage in Experiment 1. All values are averages across subjects ( $n = 15$ ), with 95% confidence intervals in brackets.

**Table 2** Comparison of peak latencies of EVC versus ROIs in (A) dorsal and (B) ventral regions for Experiments 1 and 2

Comparing peak latency in EVC to	Experiment 1		Experiment 2	
	Latency difference (ms)	Significance (P value)	Latency difference (ms)	Significance (P value)
(A) Dorsal regions				
IPS0			45	0.010
IPS1			45	0.146
IPS2			45	0.001
(B) Ventral regions				
LO1&2	12	0.25	39	0.011
VO1&2	20	0.001	39	0.001
PHC1&2	39	0.001	45	0.001
TO1&2	34	0.001	12	0.008

Note: Dorsal regions only reported for Experiment 2 due to limited fMRI coverage in Experiment 1. The table reports the latency difference and P values determined by bootstrapping the sample of participants (1000 samples).

A resulting question is which part of the estimated neural activity supports perceptual decisions such as categorization (Haxby et al. 2001; Williams et al. 2007; Haushofer et al. 2008). For example, it remains debated whether behavioral tasks on faces and bodies categorization (Spiridon and Kanwisher 2002; Reddy and Kanwisher 2006; Pitcher et al. 2009; Cichy et al. 2011, 2012) engage activity in local regions or on widely distributed representations in the ventral visual stream. The findings here extend this question into the temporal domain, that is, at which spatio-temporal coordinates does neural activity determine behavior? Future studies that constrain the results of the MEG–fMRI fusion approach by behavioral measures, for example, reaction times in categorization tasks (Carlson, Ritchie, et al. 2013; Ritchie et al. 2015), are necessary to resolve this issue.

### Tracking Neural Activity During Visual Object Recognition: The Dorsal Visual Stream and Its Relation to the Ventral Visual Stream

Our results provide novel evidence for 2 questions concerning neural activity in dorsal cortex: the timing of activity in dorsal cortex and its relation to activity in other portions of visual cortex.

Concerning timing, to our knowledge, no gold-standard intracranial response profiles for parietal cortex have been determined in humans due to typically scarce electrode coverage and low visual responsiveness in that part of the cortex (Liu et al. 2009). In monkey, 2 measures of neuronal timing in

parietal cortex have been assessed: onset and peak response latencies. While onset latencies were short and varied between 50 and 90 ms (Robinson and Rugg 1988; Colby et al. 1996; Lehky and Sereno 2007; Janssen et al. 2008; Romero et al. 2012), peak latencies occurred later at 100–200 ms, often followed by a long plateau of high firing rates (Colby et al. 1996; Shikata et al. 1996; Lehky and Sereno 2007; Janssen et al. 2008; Romero et al. 2012). The similarity of peak latency in electrophysiological recordings and the results observed here suggest that our approach reveals spatio-temporal neural dynamics mostly matching peak rather than onset firing rates in the dorsal stream. Future human intracranial studies and studies combining intracranial recordings with fMRI in primates are necessary to further resolve the relationship between the observed MEG–fMRI correspondence in representations and neuronal activity.

Concerning the relation of other portions of visual cortex, we found that MEG–fMRI correspondence in representations in dorsal regions peaked later in time than in EVC, but no evidence for latency differences between ventral and parietal cortex. A supplementary analysis investigating the representational similarity between dorsal cortex and other visual regions using partial correlation provided further evidence that representations in dorsal regions are different from representations in EVC (see [Supplementary Analysis 1](#) and [Fig. 2](#)). This suggests that dorsal cortex processes objects in a hierarchical fashion, too (Konon and Kastner 2008).



## Relation of the Approach to the Standard Taxonomy of fMRI-M/EEG Integration Approaches

How does our approach fit into the taxonomy of previous approaches for integrating fMRI and M/EEG, and in effect how does it relate to its advantages and limitations? Typically 4 distinct categories are proposed: asymmetric approaches where one modality constrains the other, that is, 1) fMRI-constrained M/EEG and 2) M/EEG constrained fMRI analysis, and symmetric approaches (fusion) in which data from either modality is weighted equally, combined in a 3) model- or 4) data-driven way (for review, see Rosa et al. 2010; Huster et al. 2012; Sui et al. 2012).

Our approach does not fit the asymmetric constraint taxa. fMRI does not constrain the solution of M/EEG source reconstruction algorithms, nor do M/EEG features enter into the estimation of voxel-wise activity. Instead, fMRI and MEG constrain each other symmetrically via representational similarity. Thus, our approach evades the danger of asymmetric approaches in giving an unwarranted bias to either imaging modality.

By exclusion, this suggests that our approach belongs to a symmetric approach taxon. However, our approach also differs critically from previous symmetric approaches, both model- and data-based. Whereas model-based approaches typically depend on an explicit generative forward model from neural activity to both fMRI and M/EEG measurements, our approach does not. The advantage of this is that our approach evades the computational complexity of solutions to such models, as well as the underlying necessary assumptions about the complex relationship between neuronal and BOLD activity (Logothetis and Wandell 2004; Logothetis 2008).

Data-driven symmetric approaches, in turn, typically utilize unsupervised methods such as independent component analysis for integration, whereas our approach relies on the explicit assumption of similarity to constrain the solution space. Our approach is data-driven, because it uses correlations to extract and compare representational patterns. However, it has the advantage that, whereas understanding results of most data-driven approaches need additional interpretational steps, our results are directly interpretable.

In summary, our approach is fundamentally different than existing fMRI and M/EEG integration approaches and offers the distinct advantages: while providing interpretational ease due to the explicit constraint of representational similarity, it evades the bias of asymmetric approaches as well as the assumptions and complexities of symmetric approaches.

## Sensitivity, Specificity, and Potential Ambiguity of the Proposed Approach

Our approach has several characteristics that provide high sensitivity and specificity with a limited risk of ambiguity in the results. The first advantage is its high sensitivity to detect information in either modality through the use of multivariate analyses. Multivariate analyses of both fMRI activity patterns (Haxby et al. 2001; Haynes and Rees 2006; Kriegeskorte et al. 2006; Norman et al. 2006; Tong and Pratte 2012) and MEG sensor patterns (Carlson, Tovar, et al. 2013; Cichy et al. 2014; Clarke et al. 2015; Isik et al. 2014) have been shown to provide increased informational sensitivity over mass univariate approaches.

While our integration approach epitomizes on the increased sensitivity provided by multivariate analysis in each modality, the multivariate treatment is not complete. Our approach does not allow estimation of the relation between any set of voxels and MEG data but is constrained in that only local response

fMRI patterns are considered in the searchlight approach. We thus trade the ability to account for widely distributed information across the whole brain for precise localization and computational tractability.

The second advantage of the integration approach is the high specificity of the integration results. By seeking shared MEG and fMRI similarity patterns across a large set of conditions, that is, covariance in similarity relations, our approach provides rich constraints on integration: for a set of any  $C$  conditions, there are  $((C \times C) - C)/2$  different values in the representation matrices being compared. In contrast, previous approaches relied on smaller sets, that is, co-variance in activation across few steps of parametric modulation (Mangun et al. 1997; Horowitz et al. 2004; Schicke et al. 2006) or co-variance across subjects (Sadeh et al. 2011). However, the high specificity of our approach is limited to experimental contexts where a large number of conditions are available. For best results, specifically optimized experimental paradigms are thus required.

A third advantage of technical nature is that M/EEG and fMRI data need not be acquired simultaneously (Eichele et al. 2005; Debener et al. 2006; Huster et al. 2012) but can be acquired consecutively (for the case of EEG-fMRI). This avoids mutual interference during acquisition and allows imaging modality-specific experimental design optimization. Note this is also a disadvantage: currently our approach does not use the information provided by single-trial-based analysis of simultaneously acquired EEG-fMRI data. However, our entire approach extends directly to EEG without modification and can thus be applied to simultaneously acquired EEG-fMRI data. Future research may address the combination of information from single-trial-based analysis with representational similarity analysis as used here.

Finally, a specific limitation of our approach is a particular case of ambiguity: if 2 regions share similar representational relations but are in fact activate at different time points, we cannot ascribe a unique activation profile to each. While this case cannot be excluded, it is highly unlikely due to the strong nonlinear transformations of neural information across brain regions, and if observed, it is an interesting finding by itself worthy of further investigation.

## Summary Statement

In conclusion, our results provide a spatio-temporal integrated view of neural dynamics underlying visual object recognition in the first few hundred milliseconds of vision. The MEG-fMRI fusion approach bears high sensitivity and specificity, promising to be a useful tool in the study of complex brain function in the healthy and diseased brain.

## Supplementary Material

Supplementary material can be found at: <http://www.cercor.oxfordjournals.org/>.

## Funding

This work was funded by National Eye Institute grant EY020484 and a Google Research Faculty Award (to A.O.), the McGovern Institute Neurotechnology Program (to D.P. and A.O.), a Feodor Lynen Scholarship of the Humboldt Foundation and an Emmy Noether grant of the DFG CI 241/1-1 (to R.M.C), and was conducted at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research, Massachusetts Institute of Technology. Funding to pay the Open Access publication charges for this article was provided by NEI grant EY020484 to Aude Oliva.

## Notes

We thank Chen Yi and Carsten Allefeld for helpful comments on methodological issues, and Seyed-Mahdi Khaligh-Razavi and Santani Teng for helpful comments on the manuscript. *Conflict of Interest*: None declared.

## References

- Andersen RA, Essick GK, Siegel RM. 1987. Neurons of area 7 activated by both visual stimuli and oculomotor behavior. *Exp Brain Res.* 67:316–322.
- Bullier J. 2001. Integrated model of visual processing. *Brain Res Rev, The Brain in Health and Disease—from Molecules to Man.* Swiss National Foundation Symposium NRP 38. 36:96–107.
- Carlson T, Tovar DA, Alink A, Kriegeskorte N. 2013. Representational dynamics of object vision: the first 1000 ms. *J Vis.* 13:1–19.
- Carlson TA, Ritchie JB, Kriegeskorte N, Durvasula S, Ma J. 2013. RT for object categorization is predicted by representational distance. *J Cogn Neurosci.* 26:1–11.
- Cichy RM, Chen Y, Haynes J-D. 2011. Encoding the identity and location of objects in human LOC. *NeuroImage.* 54:2297–2307.
- Cichy RM, Heinze J, Haynes J-D. 2012. Imagery and perception share cortical representations of content and location. *Cereb Cortex.* 22:372–380.
- Cichy RM, Pantazis D, Oliva A. 2014. Resolving human object recognition in space and time. *Nat Neurosci.* 17:455–462.
- Cichy RM, Sterzer P, Heinze J, Elliott LT, Ramirez F, Haynes J-D. 2013. Probing principles of large-scale object representation: category preference and location encoding. *Hum Brain Mapp.* 34:1636–1651.
- Clarke A, Devereux BJ, Randall B, Tyler LK. 2015. Predicting the time course of individual objects with MEG. *Cereb Cortex.* 25:3602–3612.
- Colby CL, Duhamel JR, Goldberg ME. 1996. Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. *J Neurophysiol.* 76:2841–2852.
- Dale AM, Halgren E. 2001. Spatiotemporal mapping of brain activity by integration of multiple imaging modalities. *Curr Opin Neurobiol.* 11:202–208.
- Debener S, Ullsperger M, Siegel M, Engel AK. 2006. Single-trial EEG–fMRI reveals the dynamics of cognitive function. *Trends Cogn Sci.* 10:558–563.
- Denys K, Vanduffel W, Fize D, Nelissen K, Sawamura H, Georgieva S, Vogels R, Van Essen D, Orban GA. 2004. Visual activation in prefrontal cortex is stronger in monkeys than in humans. *J Cogn Neurosci.* 16:1505–1516.
- Eichele T, Specht K, Moosmann M, Jongsma MLA, Quiroga RQ, Nordby H, Hugdahl K. 2005. Assessing the spatiotemporal evolution of neuronal activation with single-trial event-related potentials and functional MRI. *Proc Natl Acad Sci USA.* 102:17798–17803.
- Felleman DJ, Van Essen DC. 1991. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex.* 1:1–47.
- Grill-Spector K, Malach R. 2004. The human visual cortex. *Annu Rev Neurosci.* 27:649–677.
- Grill-Spector K, Weiner KS. 2014. The functional architecture of the ventral temporal cortex and its role in categorization. *Nat Rev Neurosci.* 15:536–548.
- Haushofer J, Livingstone MS, Kanwisher N. 2008. Multivariate patterns in object-selective cortex dissociate perceptual and physical shape similarity. *PLoS Biol.* 6:e187.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science.* 293:2425–2430.
- Haynes J-D, Rees G. 2006. Decoding mental states from brain activity in humans. *Nat Rev Neurosci.* 7:523–534.
- Haynes J-D, Rees G. 2005. Predicting the stream of consciousness from activity in human visual cortex. *Curr Biol.* 15:1301–1307.
- Haynes J-D, Sakai K, Rees G, Gilbert S, Frith C, Passingham RE. 2007. Reading hidden intentions in the human brain. *Curr Biol.* 17:323–328.
- Horowitz SG, Rossion B, Skudlarski P, Gore JC. 2004. Parametric design and correlational analyses help integrating fMRI and electrophysiological data during face processing. *NeuroImage.* 22:1587–1595.
- Huster RJ, Debener S, Eichele T, Herrmann CS. 2012. Methods for simultaneous EEG–fMRI: an introductory review. *J Neurosci.* 32:6053–6060.
- Isik L, Meyers EM, Leibo JZ, Poggio TA. 2014. The dynamics of invariant object recognition in the human visual system. *J Neurophysiol.* 111:91–102.
- Janssen P, Srivastava S, Ombelet S, Orban GA. 2008. Coding of shape and position in macaque lateral intraparietal area. *J Neurosci.* 28:6679–6690.
- Jorge J, van der Zwaag W, Figueiredo P. 2014. EEG–fMRI integration for the study of human brain function. *NeuroImage.* 102 (Part 1):24–34.
- Kiani R, Esteky H, Mirpour K, Tanaka K. 2007. Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *J Neurophysiol.* 97:4296–4309.
- Konen CS, Kastner S. 2008. Two hierarchically organized neural systems for object information in human visual cortex. *Nat Neurosci.* 11:224–231.
- Konkle T, Oliva A. 2012. A real-world size organization of object responses in occipitotemporal cortex. *Neuron.* 74:1114–1124.
- Kravitz DJ, Saleem KS, Baker CI, Mishkin M. 2011. A new neural framework for visuospatial processing. *Nat Rev Neurosci.* 12:217–230.
- Kreiman G. 2011. Visual population codes: toward a common multivariate framework for cell recording and functional imaging. 1st ed. Cambridge (MA): The MIT Press.
- Kriegeskorte N. 2008. Representational similarity analysis – connecting the branches of systems neuroscience. *Front Syst Neurosci.* 2:4.
- Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. *Proc Natl Acad Sci USA.* 103:3863–3868.
- Kriegeskorte N, Kievit RA. 2013. Representational geometry: integrating cognition, computation, and the brain. *Trends Cogn Sci.* 17:401–412.
- Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA. 2008. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron.* 60:1126–1141.
- Lehky SR, Sereno AB. 2007. Comparison of shape encoding in primate dorsal and ventral visual pathways. *J Neurophysiol.* 97:307–319.
- Liu H, Agam Y, Madsen JR, Kreiman G. 2009. Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron.* 62:281–290.
- Logothetis NK. 2008. What we can do and what we cannot do with fMRI. *Nature.* 453:869–878.
- Logothetis NK, Wandell BA. 2004. Interpreting the BOLD signal. *Annu Rev Physiol.* 66:735–769.

- Mangun GR, Hopfinger JB, Kussmaul CL, Fletcher EM, Heinze H-J. 1997. Covariations in ERP and PET measures of spatial selective attention in human extrastriate visual cortex. *Hum Brain Mapp.* 5:273–279.
- Maris E, Oostenveld R. 2007. Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods.* 164:177–190.
- Meyers EM, Freedman DJ, Kreiman G, Miller EK, Poggio T. 2008. Dynamic population coding of category information in inferior temporal and prefrontal cortex. *J Neurophysiol.* 100:1407–1419.
- Milner AD, Goodale MA. 2008. Two visual systems re-viewed. *Neuropsychologia.* 46:774–785.
- Mormann F, Kornblith S, Quiroga RQ, Kraskov A, Cerf M, Fried I, Koch C. 2008. Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe. *J Neurosci.* 28:8865–8872.
- Nichols TE, Holmes AP. 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp.* 15:1–25.
- Norman KA, Polyn SM, Detre GJ, Haxby JV. 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci.* 10:424–430.
- Op de Beeck HP, Haushofer J, Kanwisher NG. 2008. Interpreting fMRI data: maps, modules and dimensions. *Nat Rev Neurosci.* 9:123–135.
- Pantazis D, Nichols TE, Baillet S, Leahy RM. 2005. A comparison of random field theory and permutation methods for the statistical analysis of MEG data. *NeuroImage.* 25:383–394.
- Pasupathy A, Connor CE. 2002. Population coding of shape in area V4. *Nat Neurosci.* 5:1332–1338.
- Pitcher D, Charles L, Devlin JT, Walsh V, Duchaine B. 2009. Triple dissociation of faces, bodies, and objects in extrastriate cortex. *Curr Biol.* 19:319–324.
- Reddy L, Kanwisher N. 2006. Coding of visual objects in the ventral stream. *Curr Opin Neurobiol.* 16:408–414.
- Ritchie JB, Tovar DA, Carlson TA. 2015. Emerging object representations in the visual system predict reaction times for categorization. *PLoS Comput Biol.* 11:e1004316.
- Robinson DL, Rugg MD. 1988. Latencies of visually responsive neurons in various regions of the rhesus monkey brain and their relation to human visual responses. *Biol Psychol.* 26:111–116.
- Romero MC, Van Dromme I, Janssen P. 2012. Responses to two-dimensional shapes in the macaque anterior intraparietal area. *Eur J Neurosci.* 36:2324–2334.
- Rosa MJ, Daunizeau J, Friston KJ. 2010. EEG-fMRI integration: a critical review of biophysical modeling and data analysis approaches. *J Integr Neurosci.* 09:453–476.
- Sadeh B, Pitcher D, Brandman T, Eisen A, Thaler A, Yovel G. 2011. Stimulation of category-selective brain areas modulates ERP to their preferred categories. *Curr Biol.* 21:1894–1899.
- Schicke T, Muckli L, Beer AL, Wibral M, Singer W, Goebel R, Rösler F, Röder B. 2006. Tight covariation of BOLD signal changes and slow ERPs in the parietal cortex in a parametric spatial imagery task with haptic acquisition. *Eur J Neurosci.* 23:1910–1918.
- Schmolsky MT, Wang Y, Hanes DP, Thompson KG, Leutgeb S, Schall JD, Leventhal AG. 1998. Signal timing across the macaque visual system. *J Neurophysiol.* 79:3272–3278.
- Sereno AB, Maunsell JHR. 1998. Shape selectivity in primate lateral intraparietal cortex. *Nature.* 395:500–503.
- Sereno ME, Trinath T, Augath M, Logothetis NK. 2002. Three-dimensional shape representation in monkey cortex. *Neuron.* 33:635–652.
- Shikata E, Tanaka Y, Nakamura H, Taira M, Sakata H. 1996. Selectivity of the parietal visual neurones in 3D orientation of surface of stereoscopic stimuli. *Neuroreport.* 7:2389–2394.
- Spiridon M, Kanwisher N. 2002. How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron.* 35:1157–1165.
- Sui J, Adali T, Yu Q, Chen J, Calhoun VD. 2012. A review of multivariate methods for multimodal fusion of brain imaging data. *J Neurosci Methods.* 204:68–81.
- Tadel F, Baillet S, Mosher JC, Pantazis D, Leahy RM. 2011. Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput Intell Neurosci.* 2011:1–13.
- Taulu S, Kajola M, Simola J. 2004. Suppression of interference and artifacts by the Signal Space Separation Method. *Brain Topogr.* 16:269–275.
- Taulu S, Simola J. 2006. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Phys Med Biol.* 51:1759.
- Tong F, Pratte MS. 2012. Decoding patterns of human brain activity. *Annu Rev Psychol.* 63:483–509.
- Ungerleider LG, Mishkin M. 1982. Two cortical visual systems. In: Ingle DJ, Goodale MA, Mansfield RJW, editors. *Analysis of visual behavior.* Cambridge, MA: MIT Press. p. 549–586.
- Wandell BA, Dumoulin SO, Brewer AA. 2007. Visual field maps in human cortex. *Neuron.* 56:366–383.
- Wang L, Mruczek REB, Arcano MJ, Kastner S. 2015. Probabilistic maps of visual topography in human cortex. *Cereb Cortex.* 25:3911–3931.
- Williams MA, Dang S, Kanwisher NG. 2007. Only some spatial patterns of fMRI response are read out in task performance. *Nat Neurosci.* 10:685–686.