Resolving human object recognition in space and time

Radoslaw Martin Cichy¹, Dimitrios Pantazis² & Aude Oliva¹

A comprehensive picture of object processing in the human brain requires combining both spatial and temporal information about brain activity. Here we acquired human magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI) responses to 92 object images. Multivariate pattern classification applied to MEG revealed the time course of object processing: whereas individual images were discriminated by visual representations early, ordinate and superordinate category levels emerged relatively late. Using representational similarity analysis, we combined human fMRI and MEG to show content-specific correspondence between early MEG responses and primary visual cortex (V1), and later MEG responses and inferior temporal (IT) cortex. We identified transient and persistent neural activities during object processing with sources in V1 and IT. Finally, we correlated human MEG signals to single-unit responses in monkey IT. Together, our findings provide an integrated space- and time-resolved view of human object categorization during the first few hundred milliseconds of vision.

The past decade has seen much progress in the unraveling of the neuronal mechanisms supporting human object recognition, with studies corroborating each other across species and methods¹⁻⁵. Object recognition involves a hierarchy of regions in the occipital and temporal lobes¹,⁴,⁶⁻⁸ and unfolds over time⁹⁻¹¹. However, comparing data quantitatively from different imaging modalities, such as magneto- and electroencephalography (MEG/EEG) and functional magnetic resonance imaging (fMRI) within and across species³,¹²⁻¹⁵ remains challenging, and we still lack fundamental knowledge about where and when in the human brain visual objects are processed.

Here we demonstrate how the processing of objects in the human brain unfolds in time, using MEG, and space, using fMRI, within the first few hundred milliseconds of neural processing¹,¹⁶. First, by applying multivariate pattern classification¹⁷⁻²⁰ to human MEG responses to object images, we show the time course with which individual images are discriminated by visual representations¹⁹,²¹⁻²³. Whereas individual images were best linearly decodable relatively early, membership at the ordinate and superordinate levels became linearly decodable later and with distinct time courses. Second, using representational similarity analysis¹⁹,²⁴,²⁵, we define correspondences between the temporal dynamics of object processing and cortical regions in the ventral visual pathway of the human brain. By comparing representational dissimilarities across MEG and fMRI responses, we distinguished MEG signals reflecting low-level visual processing in V1 from signals reflecting later object processing in IT. Further, we identify V1 and IT as two differentiable cortical sources of persistent neural activity during object vision. This suggests that the brain actively maintains representations at different processing stages of the visual hierarchy. Lastly, using previously reported single-cell recordings in macaque²⁶, we extend our approach across species, finding that human MEG responses to objects correlated with the patterns of neuronal spiking in monkey IT. This work resolved dynamic object processing with a fidelity that has, to our knowledge, previously not been shown by offering an integrated space- and time-resolved view of the occipitotemporal visual pathway during the first few hundred milliseconds of visual processing.

RESULTS

Human participants (n = 16) viewed images of 92 real-world objects³,²⁶ while we acquired MEG data (Fig. 1a and Supplementary Fig. 1a). The image set comprised images of human and nonhuman faces and bodies, as well as natural and artificial objects. Images were presented for 500 ms every 1.5–2 s. To maintain attention, participants performed an object-detection task on a paper clip image shown on average every four trials. Paper clip trials were excluded from further analysis.

We extracted and preprocessed peri-stimulus MEG signal from −100 ms to 1,200 ms (1 ms resolution) with respect to stimulus onset. For each time point, we used a support vector machine (SVM) classifier to classify pairwise between all conditions (object images) (Fig. 1b). The results of the classification (percentage decoding accuracy, 50% chance level) were stored in a 92 × 92 decoding matrix, indexed by the 92 conditions. Thus, each cell in the matrix indicates the decoding accuracy with which the classifier distinguishes between two images. This matrix is symmetric across the diagonal, with the diagonal undefined. Results were averaged across two independent sessions. Figure 1c shows example matrices averaged across participants (see also Supplementary Movie 1). For all time points, significance was determined non-parametrically at the participant level by a cluster-based randomization approach²⁷,²⁸ (cluster-defining threshold P < 0.001, corrected significance level P < 0.05). 95% confidence intervals for mean peak latencies and onsets (reported in parentheses throughout the Results) were determined by bootstrapping the participant sample.

MEG signals allow pairwise decoding of individual images

What is the time course with which individual images of objects are discriminated by visual representations? We found that MEG signals

¹Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. ²McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. Correspondence should be addressed to R.M.C. (rmcichy@mit.edu).

Received 12 October 2013; accepted 23 December 2013; published online 26 January 2014; doi:10.1038/nn.3635
could resolve brain responses on the single-image level\textsuperscript{19}, for up to 92 different objects. For every time point, we averaged across all cells of the MEG decoding matrix, yielding a time course of grand average decoding accuracy across all images (Fig. 1d). We calculated the onset of significance, the time point for which objects were first discriminated by visual representations, and the peak latency, the time point when visual representations of each individual image were most distinct in terms of linear separability. We report onset and mean peak latency with 95\% confidence intervals in parentheses (for overview with all data, see Supplementary Table 1).

Before and just after stimulus presentation, grand average decoding accuracy fluctuated around chance level (50\%). The curve rose sharply and reached significance at 48 ms (45–51 ms), followed by a peak at 102 ms (98–107 ms; horizontal error bar above peak shows 95\% confidence interval). (Fig. 1e) 51–61 ms after stimulus onset, followed by peaks at 99–112 ms (Supplementary Table 1b). Thus, multivariate analysis of MEG data revealed the temporal dynamics of visual content processing in the brain even at the level of individual images\textsuperscript{19}.

**Time course of category decoding**

To determine when visual representations discriminate object membership at superordinate, ordinate and subordinate categorization levels, we compared decoding accuracy within and between the relevant partitions of the MEG decoding matrix (Fig. 2). The resulting

---

**Figure 1** Decoding of images from MEG signals. (a) Image set of 92 images\textsuperscript{3,26} of different categories of objects. (b) Multivariate analysis of MEG data. (c) Examples of 92 × 92 MEG decoding matrices (averaged over participants, n = 16). (d) Time course of grand total decoding was significant at 48 ms (45–51 ms), with a peak at 102 ms (98–107 ms; horizontal error bar above peak shows 95\% confidence interval). (e) Time course of object decoding within subdivisions. The left panel illustrates the separately averaged sections of the MEG decoding matrix (color-coded), the right panel the corresponding decoding time courses. Peak latencies and onsets of significance are listed in Supplementary Table 1b. Rows of asterisks indicate significant time points (n = 16, cluster-defining threshold P < 0.001, corrected significance level P < 0.05). The gray vertical line indicates onset of image presentation.
The first two dimensions of the MDS (criterion: decoding matrix at peak latency) revealed by the panel illustrates the structure in the MEG and error bars same as in Figure 1. Statistical details in Supplementary Table 1c. The right panel illustrates the structure in the MEG decoding matrix at peak latency revealed by the first two dimensions of the MDS (criterion: metric stress, 0.24 for a-e, 0.27 for d). “Dec. acc.” indicates decoding accuracy.

We determined significance as before by sign permutation tests (n = 16; cluster-defining threshold P < .001, corrected significance level P < .05) and 95% confidence intervals for peak latencies and onsets by bootstrapping the participant sample (n = 16).

We found that visual representations discriminated objects by animacy,21,23 with a peak at 157 ms (152–302 ms) (Fig. 2a). Similarly, visual representations discriminated objects by naturalness with a peak at 122 ms (107–254 ms) (Fig. 2b). Multidimensional scaling (MDS)29,30 illustrated the main structure in the MEG decoding matrix at peak latency: clustering of objects into animate and inanimate, as well as natural and artificial.

Within the animate division, faces and bodies clustered separately. This suggested that membership to categorical divisions below the animate/inanimate distinction might be discriminated by visual representations.22,31–35 Indeed, we found that the distinction between faces and bodies was significant at 56 ms (46–74 ms), with a clear peak at 136 ms (131–140 ms) (Fig. 2c). MDS at the 136-ms peak (Fig. 2c) showed a division between faces and bodies, dominated by nonhuman bodies versus the other conditions. At the subordinate level, we found that visual representations distinguished bodies by species with an onset at 75 ms (64–113 ms) and a peak at 170 ms (104–252 ms), and the MDS showed a clear species-specific clustering of bodies (Fig. 2d). We also observed a significant difference in decoding accuracy for human versus nonhuman faces starting at 70 ms (54–74 ms), followed by two prominent peaks at 127 ms (122–133 ms) and 190 ms (175–207 ms), calculated on the time window starting at the trough between the two peaks at 156 ms and 1,200 ms (Fig. 2e). An MDS at the first peak illustrated the effect, with a perfect separation of faces along the species border.

For photographs of real-world objects, category membership is often associated with differences in low-level image properties. Thus, linear separability of objects by category may be achieved on the basis of low-level image property representations. However, an analysis testing the degree to which classifiers generalized across particular object images showed that the discrimination of category...
membership was not solely determined by image-specific properties (Supplementary Fig. 2).

Comparing peak-to-peak latency differences (Figs. 1 and 2) using bootstrapping ($n = 16$, $P < 0.05$, Bonferroni-corrected), we found that images were discriminated earlier at the level of individual images than at higher categorization levels (all $P < 0.001$, except for human versus nonhuman body; for details, see Supplementary Table 2). In a behavioral experiment, new participants were asked to perform same-different image classification in the context of different categorization levels (identity, subordinate to superordinate classification) (Supplementary Fig. 1c). We found significant Pearson’s correlations by bootstrapping ($n = 16$) between peak decoding accuracy and reaction times ($R = 0.53$, $P = 0.003$) and correctness ($R = 0.49$, $P = 0.012$) (Supplementary Fig. 3).

Taken together, multivariate analysis of MEG signals revealed the time course with which membership of individual objects at different categorical levels was linearly decodable. Our results complement previous work on rapid object detection in go/no-go tasks,21,23 and provide a content-based analysis of the time course with which object information is processed.19

**Transient and persistent neuronal activity**

The dynamics of the decoding time courses above suggested highly variable and transient neural activity as their source. As neuronal signals propagate along the ventral visual stream, different image properties appear to be processed at subsequent time points. However, the previous analyses did not allow us to determine the existence of persistent neural activity during the course of object processing. Such persistent neural activity could maintain the results of a particular neural processing stage for later use.

Intuitively, if neuronal activity persists over time, MEG signals should be similar across time as well. To search for such similarities, we trained an SVM classifier at one time point ($t_x$) and tested at other time points ($t_y$) (Fig. 3). Conducting all pairwise discriminations between objects, we obtained a $92 \times 92$ MEG decoding matrix for every pair of time points ($t_x,t_y$). We then repeated the process across all pairs of time points, resulting in a four-dimensional image–time–time–time decoding matrix. Averaging across the first two dimensions yielded a time–time decoding matrix (Fig. 3b,c).

As expected, some neural activity during object processing was transient: the classifier generalized best to neighboring time points and performed poorly for distant time points. This is illustrated by the highest decoding accuracy along the diagonal and the sharp drop of decoding accuracy away from the diagonal (Fig. 3b), depicted as a high and steep crest of decoding accuracy (Fig. 3c). Notably, we also found evidence for persistent neural activity. First, the classifier generalized well for the time-point combination of $\sim 100$ ms and $\sim 200–1,000$ ms. As this effect was clearly circumscribed in time and persisted beyond the offset of image presentation at 500 ms, it is unlikely that it merely reflected constant passive influx of information during image presentation. This suggests that the brain actively maintains visual representations in early stages of the visual processing hierarchy, potentially as memory for low-level visual features.16 Second, between $\sim 250$ ms and $\sim 500$ ms, the classifier produced a broader diagonal, this indicated that neural activity was similar across these time points, suggesting that a stable representation of objects in later stages of visual processing hierarchy is kept online.

Statistical testing (Fig. 3d, sign permutation test, $n = 16$, cluster-defining threshold $P < 0.0001$, corrected significance level $P < 0.05$) indicated widespread similarity of neural activity. However, the fact that this is not limited to particular time–point combinations may indicate either neural activity actively maintained at all cortical processing levels or a passive response of the brain to the constant influx of visual information during the presence of the stimulus.

In sum, across-time analysis of the dynamics in visual representations revealed both transient and persistent neuronal processing of objects. The presence of persistent neuronal activity at well-delineated time point combinations may indicate active maintenance of visual representations at different processing stages.
Resolving object recognition in space and time

What are the cortical sources of the MEG signals that discriminate objects? Given that V1 and IT process different aspects of the images, we expected MEG signals originating from these two cortical areas to differ: in other words, V1 and IT responses to individual objects should differ in their dissimilarity relations, resulting in distinct patterns over time in the MEG decoding matrices. Here we used representational similarity analysis to show when representations extracted with MEG were comparable to those extracted with fMRI in human V1 and IT (Fig. 4a).

After adapting the MEG stimulation protocol to the specifics of fMRI (Supplementary Fig. 1b), we repeated the experiment with the same images in the same participants while acquiring fMRI data. We estimated individual (92) object image-related brain responses by fitting a general linear model. We then extracted voxel values from a region of interest (V1 or IT) to form a pattern vector (Fig. 4a). The resulting 92 pattern vectors were subjected to pairwise Pearson’s correlation and then ordered into a 92 × 92 similarity matrix indexed by image condition. We computed matrix elements from R (similarity) to 1 – R (dissimilarity) to make the matrices directly comparable to the MEG decoding accuracy matrices. The above process produced two fMRI dissimilarity matrices, one for V1 and one for IT, for each participant.

Using these two fMRI matrices, we first successfully reproduced a main previous finding: stronger representation of animacy in IT than in V1, shown by MDS, hierarchical clustering and quantitative testing (Supplementary Fig. 4). For further analysis, we computed participant-averaged fMRI matrices for both human V1 and IT. We then evaluated the extent of similar representations between fMRI and MEG by computing Spearman’s rank-order correlations between fMRI dissimilarity matrices (separately for V1 and IT) and participant-specific MEG decoding accuracy matrices (separately for each time point) (Fig. 4b). We found that MEG signals correlated with fMRI dissimilarity matrices in human V1 and IT with different time courses (Fig. 4c).

We estimated individual object image-related brain responses by the same images in the same participants while acquiring fMRI data. We selected voxel activation values, yielding 92 pattern vectors. Then we calculated the pairwise Pearson’s correlation (R) for all combinations of experimental conditions (i, j). The dissimilarity measure 1 – R was assigned to a 92 × 92 fMRI dissimilarity matrix indexed by the experimental conditions (i, j). This analysis was conducted independently for each region of interest.

For each time point t, we correlated (Spearman’s rank-order correlation) the MEG decoding matrix to the fMRI dissimilarity matrices of V1 and IT (Fig. 4a). (c) MEG signals correlated with the fMRI dissimilarity matrix of central V1 earlier than with the fMRI dissimilarity matrix of IT. Blue and red asterisks indicate significant time points for V1 and IT. (d) Difference between the two curves in c. MEG correlated early more with V1 than with IT, and late more with IT than with V1. Blue and red asterisks in the plots indicate significant time points for positive and negative clusters respectively. For details, see Supplementary Table 1a. n = 16; gray line and statistical procedure as in Figure 1.

Relating MEG and fMRI object signals across time

The above MEG-fMRI representational similarity analysis naturally extends to include the MEG time-time decoding matrices constructed earlier (Fig. 3). This analysis allows identifying the cortical sources that have persistent neural activity. We therefore correlated the fMRI dissimilarity matrices of V1 and IT with the MEG 92 × 92 decoding matrices obtained for each pair of time points (t, t) (Figs. 3a and 5a). The results (Fig. 5b,c, sign-permutation test, n = 16, cluster-defining threshold P < 0.0001, corrected significance level P < 0.05) demonstrated that neural activity for the time point combinations of ~100 ms
and -200–1,000 ms correlated with V1 dissimilarity matrices. In contrast, neural activity between -250 ms and -500 ms correlated with IT dissimilarity matrices. Notably, this was also true when comparing the correlations directly (Fig. 5d).

In sum, by combining fMRI and MEG, we identified V1 and IT as distinct cortical sources of persistent neural activity during visual object perception. This suggests that the visual system actively maintains neural activity at different levels of visual processing.

Relating human MEG to spiking activity in monkey IT

Previous research has shown that object representations in IT are comparable in monkeys and humans. Here, using representational similarity analysis, we related the dynamics in human MEG to the pattern of activity in monkey IT (as measured electrophysiologically for the same image set) (Fig. 6a). Brain responses in human MEG and monkey IT were significantly correlated (sign permutation test, n = 16, cluster-defining threshold P < 0.001, corrected significance level P < 0.05, 95% confidence intervals by bootstrapping), first at 66 ms (56–71 ms) and peaking at 141 ms (132–292 ms) (Fig. 6b). MDS at peak latency (Fig. 6c) revealed an arrangement strongly dominated by human faces. However, significant correlations (sign permutation test, n = 16, cluster-defining threshold P < 0.001, corrected significance level P < 0.05) were
also present within all subdivisions of the image set (Fig. 6d and Supplementary Table 1g). These results corroborated and extended the evidence for a common representational space for objects in monkeys and humans.

**DISCUSSION**

Using multivariate pattern classification methods and representational similarity analysis on combined human MEG-fMRI data, we have demonstrated how object recognition proceeds in space and time in the human ventral visual cortex. First, we found that whereas individual images were discriminated early, membership to ordinate and superordinate levels was discriminated later. Notably, we identified neural activity that was either persistent or transient during the first few hundred milliseconds of object processing. Second, using representational similarity analysis, we combined human fMRI and MEG to show content-specific correspondence between early MEG responses and V1 fMRI responses, and later MEG responses and IT fMRI responses. Extending this analysis, we located the sources of differentiable persistent neural activity in V1 and IT. Last, we extended the representational similarity analysis across species by showing that human MEG signals can be related to spiking activity in macaque IT. We thus extended the evidence for a common representational space for objects in monkeys and humans to the time domain.

**The time course of object processing**

Applying multivariate pattern classification to MEG data, we found that visual representations discriminated individual images (peak at 102 ms) and then proceeded to classify them into categories. We found the peak latencies for classification of naturalness (122 ms) and animacy (157 ms) to match previous reports of neural response latencies in human and monkey IT. The broad confidence intervals for peak latency for animacy and naturalness may indicate that object information is sustained online for more in-depth analysis after discrimination is first possible. At the subordinate level, the body-specific peak (170 ms) and the two face-specific peaks (127 ms and 190 ms) concur with previous work (for bodies, 170–260 ms; for faces, first peak at 100–120 ms and second peak at 170 ms). It remains controversial whether the two face peaks have different cortical sources, a question that future studies comparing representational dissimilarities in MEG and fMRI may resolve. While our MEG results confirm a body of work, they generalize previous findings to a large image set and show when neural activity is transient or persistent during object analysis. This goes beyond what was previously possible with standard analysis techniques of the evoked response, allowing us to dissect the evoked response into functionally distinct neural components.

When comparing peak latencies of decoding accuracy at different levels of categorization, we observed that individual images were discriminated by visual representations early, whereas ordinate and superordinate levels emerged relatively late. However, onset latencies of significance for the various categorization levels were early (48–70 ms, except naturalness at 93 ms). Therefore, our results support models of object categorization that suggest processing of object information to begin at all levels of categorization simultaneously, with differential time courses of evidence accumulation for different levels of categorization. Our results might seem to be at odds with previous research suggesting clear multistage processing in IT, with a stage of global processing followed by local processing. However, our approach captures signals from the whole of the human brain simultaneously, not only IT. Thus, although we can determine which region predominates in shaping the MEG response at a specific time, we cannot distinguish between early and late phases of a response of a particular region.

The method and results of this work provide a gateway to resolving the time course of visual processing to a variety of other visual stimuli. In effect, it may permit a denser sampling of object space than previously achieved. For example, a description of the temporal dynamics of face representation in humans might be feasible with a large and rich parametrically modulated stimulus set and comparison to monkey electrophysiology. Similarly, most MEG or EEG studies based on event-related potential analysis investigating content-specific modulation of brain activity by cognitive factors like memory or attention must rely on a handful of categorical markers in the event-related potential waveform—for example, the M100 and N170 for faces. In contrast, by applying multivariate methods, potentially any kind of content and the modulation of its representation by cognitive factors may be tractable, increasing experimental flexibility and generalizability of results. Thus, the application of multivariate analysis techniques to MEG might be as fruitful in the future study of object recognition as the introduction of these techniques was in fMRI.

**Resolving object processing in time, space and species**

Relating signals measured in different imaging modalities and combining the methods’ respective advantages are challenges in systems neuroscience. Using representational similarity analysis, we showed a content-selective correspondence between early MEG signals and fMRI responses in V1 and between later MEG signals and fMRI responses in IT. Our results match previously reported average onset latencies in the literature, ranging between 50 and 80 ms in V1 (ref. 9) and 80 and 200 ms in IT. Thus, representational similarity analysis combining MEG and fMRI is a promising method for relating cortical activations across space and time.

Comparing visual representations across time, we differentiated transient from persistent neural activity during object processing and found evidence for persistent activity in both V1 and IT. Thus, during object viewing the brain maintained both low- and high-level feature representations. These effects are most likely actively controlled processes, as indicated by their limited temporal extent. They might form the basis of memory of images in representational formats that make explicit different properties of these images—for example, low-level features versus category membership.

An integrated theory of object recognition requires quantitative bridging of the gap not only across imaging modalities but also across species. Using representational similarity analysis, it has been shown that human and monkey IT share a similar object coding scheme. Here we have extended this finding by taking first steps in linking the dynamics in human MEG to single-cell activity in monkey IT. Note that in this experiment all temporal variance came from MEG: the dissimilarity matrix of monkey IT is based on averaged activity in IT cells 71–210 ms after stimulus onset. Future studies might compare the dynamics in human and monkey IT using monkey data resolved in time, thus potentially complementing spatial homologies with temporal ones. In the meantime, the linkage between the time course of individual object coding in humans and coding of these same objects in monkey IT, although predictable by previous research, is shown here to our knowledge for the first time.

**Conclusion**

Progress in understanding how object recognition is implemented in the brain is likely to come from the combination of advances in data analyses suitable for different imaging techniques and comparison
across species.\(^\text{25}\) Here we have provided key advances on two fronts. First, by applying multivariate pattern classification to human MEG signals, we showed the dynamics with which the brain processes objects at different levels of categorization and actively maintains visual representation. Second, we proposed an integrated space- and time-resolved view of the human brain during the first few hundred milliseconds of visual object processing and showed that representational similarity analysis allows brain signals in space, time and species to be understood in a common framework.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This work was funded by US National Eye Institute grant EY020484 (to A.O.), US National Science Foundation grant ECS-1357380 (to D.P.) and a Humboldt Scholarship (to R.M.C.), and was conducted at the Aihlona A. Martinos Imaging Center at the McGovern Institute for Brain Research, Massachusetts Institute of Technology.

AUTHOR CONTRIBUTIONS

R.M.C., D.P. and A.O. designed the research. R.M.C. and D.P. performed experiments and analyzed the data. R.M.C. and D.P. and A.O. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.
Multivariate analysis of MEG data. To determine the amount of object image information contained in MEG signals, we employed multivariate analysis in the form of linear support vector machines (SVMs; libsvm: http://www.csie.ntu.edu.tw/~cjlin/libsvm/). The SVM analysis was conducted independently for each participant and session (Fig. 1a,b). For each time point (100 ms before to 1,200 ms after image onset), MEG data were arranged in the form of 306 dimensional measurement vectors, yielding N pattern vectors per time point and condition (image). We used supervised learning, with a leave-one-out cross-validation approach, to train the SVM classifier to pairwise decode any two conditions. Namely, for each time point and pair of conditions, N – 1 pattern vectors comprised the training set and the remaining Nth pattern vectors the testing set, and the performance of the classifier to separate the two conditions was evaluated. The process was repeated 100 times with random reassignment of the data to training and testing sets, yielding an overall decoding accuracy of the classifier. The decoding accuracy was then assigned to a decoding accuracy matrix of size 92 × 92, with rows and columns indexed by the conditions classified. The matrix is symmetric across the diagonal, with the diagonal undefined. This procedure yielded one 92 × 92 matrix of decoding accuracies for every time point.

Visualization and exploration using multidimensional scaling. The 92 × 92 MEG decoding matrices contained complex high-dimensional structure that was difficult to visualize. To reveal any underlying patterns, we used multidimensional scaling (MDS)29,30 to plot the data into a two-dimensional space of the first two dimensions of the solution, such that similar conditions were grouped together and dissimilar conditions far apart. MDS is an unsupervised method to visualize the level of similarity of individual objects contained in a distance matrix (here the decoding matrix), whereby objects are automatically assigned coordinates in space so that distances between objects are preserved. MDS was applied to the whole or part of the decoding matrix, depending on the conditions explored. To avoid double-dipping51 the data, we computed MDS at peak-latency time points with a leave-one-participant-out approach as follows: all but one participant were used to identify the peak latency time and the remaining participant provided the decoding matrix. The decoding matrix was averaged across all permutations, and only the overall decoding matrix was subjected to MDS.

Human fMRI acquisition. Magnetic resonance imaging (fMRI) was conducted on a 3T Trio scanner (Siemens, Erlangen, Germany) with a 32-channel head coil. We acquired structural images using a standard T1-weighted sequence (192 sagittal slices, FOV = 256 mm², TR = 1,900 ms, TE = 2.52 ms, flip angle = 9°). For fMRI, we conducted 10–14 runs in which 192 volumes were acquired for each participant (gradient-echo EPI sequence: TR = 2,000 ms, TE = 31 ms, flip angle = 80°, FOV read = 192 mm, 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.

Human fMRI analysis. fMRI data were processed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/). For each participant and session separately, data were realigned and slice-time corrected, and then co-registered to the T1 structural scan acquired in the first MRI session. We neither normalized nor smoothed fMRI data. We then modeled the fMRI responses to the 92 images with a general linear model (GLM) in two independent models: one comprising only the first three runs of each session and one comprising the remaining runs. The onsets and durations of each image presentation, as well as those of the null trials, were entered into the GLM as regressors and convolved with a hemodynamic response function. Movement parameters entered the GLM as nuisance regressors. For each of the 92 image conditions, we converted GLM parameter estimates into t-values by contrasting each condition estimate against the explicitly modeled baseline. In addition, we assessed the effect of visual stimulation irrespective of condition in a separate t-contrast by contrasting the parameter estimates for all 92 images against baseline.

fMRI region of interest definition. We defined V1 separately for each participant on the basis of an anatomical eccentricity template52. The cortical surface of each participant was reconstructed with FreeSurfer on the basis of the T1 structural scan53. The right hemisphere was mirror-reversed and registered to the left hemisphere. This allowed us to register a V1 eccentricity template52 to participant-specific surfaces and to define surface-based regions of interest (ROIs) corresponding to 0–3° and 3–6° visual angle, termed here central V1 and peripheral V1. These surface-based ROIs were resampled to the space of EPI volumes and combined in a common ROI for both cortical hemispheres.

For human IT, we used a mask consisting of bilateral fusiform and inferior temporal cortex (WFU Pickatlas, IBASPM116 Atlas26). Anatomical masks were reverse-normalized from MNI-space to single-participant space. To match the size of IT to the average size of central V1, we restricted the definition of IT for each participant and session: we considered only the 361 most strongly activated voxels in the t-contrast of all conditions against baseline in the GLM based only on the first three runs; we used only these voxels to extract t-values for each of the 92 images from the remaining runs for further analysis.

fMRI pattern analysis. We used a correlation-based method to determine the relations between brain fMRI responses to the 92 images (Fig. 4). Observations were formed from each ROI (central and peripheral V1, and IT), by extracting and concatenating the corresponding voxel fMRI activation values into pattern vectors. For every pair of the 92 conditions, we then computed the Spearman’s rank-order correlation coefficient R between the corresponding pattern vectors of a given ROI and stored the result in a 92 × 92 symmetric matrix. We converted the correlations into a dissimilarity measure1–1 − R, which is bounded between 0 (no dissimilarity) and 2 (complete dissimilarity). For further analyses, we averaged the dissimilarity matrices across sessions and participants, resulting in one matrix for each ROI.

Monkey electrophysiology. The details of monkey electrophysiological recordings and representational similarity analysis are described elsewhere28. In short, two awake macaque monkeys were presented with the same stimulus set as the...
one used in our MEG and fMRI human experiments. Images spanned 7° visual angle and were presented in a rapid design (105 ms on, ISI = 0s) among a larger set of images while the monkeys maintained fixation. Single-cell responses in 674 neurons were recorded extracellularly from anterior inferior temporal cortex. Cell responses to each image were estimated as average spike rate in a 71–210 ms time window after stimulus onset. Representational dissimilarity matrices were created by pairwise-correlating responses to images across the 674 neurons (Pearson’s product-moment correlation) and subtracting the resulting value from 1. The representational dissimilarity matrices were generously supplied to us by N. Kriegeskorte and R. Kiani1,26.

Significance testing. We used non-parametric statistical inference27,28, which does not make assumptions about the distribution of the data, for random-effects inference. Permutation tests were used for cluster-size inference, and bootstrap tests for confidence intervals on (1) maxima and cluster onsets/offsets and (2) peak-to-peak latency differences. The sample size (N) was always 16, and all tests were two-sided.

Permutation tests. The null hypothesis of no experimental effect differed throughout the paper depending on the analysis of interest: the MEG decoding time series was equal to 50% chance level; the within-subdivision minus between-subdivision portions of an MEG decoding matrix was equal to 0; the correlation of the MEG decoding matrices and fMRI (or monkey spiking activity) dissimilarity matrix was equal to 0. In all cases, under the null hypothesis we could permute the condition labels of the MEG data, which effectively corresponds to a sign permutation test that randomly multiplies the participant-specific data (for example, MEG decoding accuracies or correlations) with +1 or −1. For each MEG permutation sample, we recomputed the statistic of interest. Repeating this procedure 50,000 times, we obtained an empirical distribution of the data, which allowed us to convert our statistics (for example, MEG decoding time series, time-time decoding matrices, etc.) into one-dimensional or two-dimensional P-value maps.

Familywise error rate was then controlled with cluster-size inference. The P-value maps of the original data were thresholded at P < 0.001 for one dimension and P < 0.0001 for two dimensions to define suprathreshold clusters. These suprathreshold clusters were reported significant only if their size exceeded a threshold, estimated as follows: the previously computed permutation samples were also converted to P-value maps (relying on the same empirical distribution as the original data) and also thresholded to define resampled versions of suprathreshold clusters. These clusters were used to construct an empirical distribution of maximum cluster size and to estimate a threshold at 5% of the right tail of this distribution (that is, the corrected P-value is P < 0.05).

Bootstrap tests. We calculated 95% confidence intervals for the onsets of the first significant cluster and the peak latency of the observed effects. To achieve this, we created 1,000 bootstrapped samples by sampling the participants with replacement. For each bootstrap sample, we repeated the exact data analysis as the original data (including the permutation tests), resulting in bootstrap estimates of onsets and peak latencies and thus the determination of their 95% confidence intervals.

To calculate confidence intervals on mean peak-to-peak latency differences, we created 50,000 bootstrapped samples by sampling the participant-specific latencies with replacement. This yielded an empirical distribution of mean peak-to-peak latencies. We set P < 0.05, Bonferroni-corrected. If the 95% confidence interval did not include 0, we rejected the null hypothesis of no peak-to-peak latency differences.

49. Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D. & Leahy, R.M. Brainstorm: a user-friendly application for MEG/EEG analysis. Comput. Intell. Neurosci. 2011, 879716 (2011).
50. Möller, K.R., Mika, S., Rätsch, G., Tsuda, K. & Schölkopf, B. An introduction to kernel-based learning algorithms. IEEE Trans. Neural Netw. 12, 181–201 (2001).
51. Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S.F. & Baker, C.I. Circular analysis in systems neuroscience: the dangers of double dipping. Nat. Neurosci. 12, 535–540 (2009).
52. Berison, N.C. et al. The retinotopic organization of striate cortex is well predicted by surface topology. Curr. Biol. 22, 2081–2085 (2012).
53. Dale, A.M., Fischl, B. & Sereno, M.I. Cortical surface-based analysis: I. Segmentation and surface reconstruction. Neuroimage 9, 179–194 (1999).
54. Maldjian, J.A., Laurienti, P.J., Kraft, R.A. & Burdette, J.H. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage 19, 1233–1239 (2003).