Published in final edited form as:
Nat Neurosci. 2014 March ; 17(3): 455–462. doi:10.1038/nn.3635.

Resolving human object recognition in space and time

Radoslaw Martin Cichy[1], **Dimitrios Pantazis**[2], and **Aude Oliva**[1]

[1] Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA, USA

[2] McGovern Institute for Brain Research, MIT, Cambridge, MA, USA

Abstract

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 later. Using representational similarity analysis, we combine 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, human MEG signals were correlated 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.

INTRODUCTION

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 methods[1-5]. Object recognition involves a hierarchy of regions in the occipital and temporal lobes[1,4,6-8] and unfolds over time[9-11]. However, comparing data quantitatively from different imaging modalities, such as MEG/EEG (magneto- and electroencephalography) and fMRI (functional magnetic resonance imaging) within and across species[3,12-15], remains challenging and we still lack fundamental knowledge about *where* and *when* in the human brain visual objects are processed.

Here, we demonstrated 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 processing[1,16]. First, by applying multivariate pattern classification[17-20] to human MEG

CORRESPONDING AUTHOR Radoslaw Martin Cichy Science and Artificial Intelligence Laboratory MIT 32-D430 Cambridge, MA, USA Phone: +1 617 253 1428 rmcichy@mit.edu.

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

responses to object images, we showed the time course with which individual images are discriminated by visual representations[19,21-23] 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 analysis[19,24,25], we defined 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 primary visual cortex (V1) from signals reflecting later object processing in inferior temporal cortex (IT). Further, we identified 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 recording data in macaque [26], we extended our approach across species, and showed that human MEG responses to objects correlated with the patterns of neuronal spiking in monkey IT.

Together, this work resolved dynamic object processing with a fidelity that has previously not been shown by offering an integrated space- and time-resolved view of the occipito-ventral visual pathway during the first few hundred milliseconds of visual processing.

RESULTS

Human participants ($n = 16$) viewed images of 92 real-world objects[3,26] while MEG data was acquired (Fig. 1a and, Supplementary Fig. 1a). The image set comprised images of human and non-human faces and bodies, as well as natural and artificial objects. Images were presented for 500ms every 1.5–2s. To maintain attention, participants performed an object-detection task on a paper clip image shown on average every 4 trials. Paper clip trials were excluded from further analysis.

We extracted and preprocessed peri-stimulus MEG signal from –100ms to 1200ms (1ms resolution) with respect to stimulus onset. For each time point, we used a support vector machine (SVM) classifier pair–wise to classify between all conditions, i.e. object images (Fig. 1b). The results of the classification (% 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 reported, significance was determined non-parametrically at the Level by a cluster-based randomization approach[27,28] (cluster-defining threshold $p<0.001$, corrected significance level $p<0.05$). 95% confidence intervals for mean peak latencies and onsets were determined by bootstrapping the participant sample (reported in brackets throughout the results section).

MEG signals allow pairwise decoding of individual images

What is the time course with which individual images of objects are discriminated by visual representations? Here, we showed that MEG signals can resolve brain responses on the single-image level[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 *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 brackets (for overview of 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 48ms (45–51ms), followed by a peak at 102ms (98–107ms) and a gradual decline (Fig. 1d,). Notably, we observed significant decoding accuracy of individual images within each of the 6 subdivisions of the image set (human and non-human faces and bodies, natural and artificial objects:, Fig. 1e) 51–61ms after stimulus onset followed by peaks at 99–112ms (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[19].

Time course of category decoding

To determine when visual representations discriminate object membership at superordinate, ordinate, and subordinate different categorization levels, we compared decoding accuracy within and between the relevant partitions of the MEG decoding matrix (Fig. 2). The resulting measure (difference in decoding accuracy) indicates both linear separability and clustering of objects according to subdivision membership. Peaks in this measure represent time points at which the brain has untangled visual input such that relevant information about object membership is explicitly coded. We determined significance as before by sign permutation tests ($n = 16$, cluster-defining threshold $p<0.001$, corrected significance level $p<0.05$) 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 157ms (152–302ms) (Fig. 2a, middle). Similarly, visual representations discriminated objects by *naturalness* with a peak at 122ms (107–254ms) (Fig. 2b). Multidimensional scaling (MDS)[29,30] To 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. 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 56ms (46–74ms), with a clear peak at 136ms (131–140ms) (Fig. 2c,). MDS at the 136ms peak (Fig. 2c) showed a division between faces and bodies, dominated by non-human bodies versus the other conditions. At the subordinate level, we found that visual representations distinguished bodies by species (onset at at 75ms (64–113ms), peak at 170ms (104–252ms), (Fig. 2d, middle). tThe MDS showed a clear species-specific clustering of bodies (Fig. 2d). ,). We also observed a significant difference in decoding accuracy for human faces versus non-human faces starting at 70ms (54–74ms), followed by two prominent peaks at 127ms (122–133ms) and 190ms (175–207ms), calculated on the time window starting at the trough between the two peaks at 156ms and 1200ms) (Fig. 2e,). An MDS at the first peak illustrated the effect with a perfect separation of faces along the species border.

Note that 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 generalize across particular object images showed that the discrimination of category membership is 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 non-human body, for details see Supplementary Table 2). In a behavioral experiment, novel 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'scorrelations by bootstrapping (n=16) between peak-decoding accuracy and reaction times (R=0.53, p=0.003) as 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 and 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 sources. 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 serve the role of maintaining 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 a 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 92×92 MEG decoding matrix for every pair of time points (t_x,t_y). The process was repeated across all pairs of time points, resulting in a 4-dimensional image-image-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 ~100ms and ~200–1,000ms. As this effect was clearly circumscribed in time, and persisted beyond the offset of image presentation at 500ms, 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[36]. Second, between ~250ms and ~500ms, the classifier produced a broader diagonal. This indicated that neural activity is 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 $p<0.05$, red–colored) indicated widespread similarity of neural activity. However, the fact this that 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, we found that 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[6], 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[3], resulting in distinct patterns over time in the MEG decoding matrices. Here, we used *representational similarity analysis*[24,25] to show when representations extracted with MEG are comparable to those extracted with fMRI in human V1 and IT, by (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 pair-wisepairwise Pearson's correlation and then ordered into a 92×92 similarity matrix indexed by image condition. We converted 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, for V1 and IT, for each participant.

Using these two fMRI matrices, we first successfully reproduced a main previous finding[3]: 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 in human V1 and human IT with different time courses (Fig. 4c, sign-permutation test, cluster-defining threshold $p<0.0001$, corrected significance level $p<0.05$, 95% confidence intervals by bootstrapping). The V1 correlation time course peaked early at 101ms (84–109ms), whereas the IT time course peaked later at 132ms (129–290ms) (for onset of significance see Supplementary Table 1d). The difference in peak-to-peak latency was significant ($n=16$, sign-permutation test, $p=0.016$,). Importantly, comparing the V1 and IT time course directly (Fig. 4d), we found that MEG signals correlated more with V1 than with IT early (peak at 93ms (79–102ms)), and more with IT than with V1 later (peak at 284ms (152– 303ms)).

Notably, the correlation of MEG with human IT was also present within each of the six subdivisions of the image set (Supplementary Fig. 5). Correlating MEG with previously reported fMRI data of human IT[3] yielded comparable effects (Supplementary Fig. 6), i.e. a peak at 158ms (152–300ms), reinforcing the validity and generalizability of our results. Additionally, the correlation of MEG to V1 was specific to the stimulated portion of V1: an immediately adjacent V1 region-of-interest corresponding to an unstimulated portion of the visual field (3–6° visual angle) showed significantly weaker correlation than central V1 (Supplementary Fig. 7).

In summary, we demonstrated for the first time that temporal dynamics as measured by MEG can be mapped onto distinct early and late human cortical regions along the ventral visual stream using representational similarity analysis.

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_x,t_y) (Figs 3a and, 5a). The results (Fig. 5 b,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 ~100ms and ~200– 1,000ms correlated with V1. In contrast, neural activity between ~250ms and ~500ms (marked by the striped white ellipse) correlated with IT. Importantly, this is 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

Prior research[3] 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 electro-physiologically for the same image set[26]) (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 66ms (56–71ms), and peaking at 141ms (132–292ms) (Fig. 6b). MDS at peak latency (Fig. 6c) revealed an arrangement strongly dominated by human faces. However, significant correlations were also present within all subdivisions of the image set (Fig 6d, Supplementary Table 1g). These results corroborated and extended the evidence for a common representational space for objects in monkeys and humans[3].

DISCUSSION

Using multivariate pattern classification methods[18–20] and representational similarity analysis[3,24,25] on combined human MEG-fMRI data, we demonstrated how object recognition proceeds in space and time in the human ventral visual cortex. First, we showed that whereas individual images werare discriminated early, membership to ordinate and superordinate levels was discriminated later[19]. Notably, we identified neural activity that is 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 early visual areaprimary visual cortex (V1), and later MEG responses and inferior -temporal (IT) cortex. 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 [3,25] 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 temporal domain.

The time course of object processing

Applying multivariate pattern classification to MEG data, we showed that visual representations discriminated individual images[19] (peak at 102 ms) and then proceeded to classify them into larger categories[37,38]. We found the peak-latencies for classification of naturalness (122ms) and animacy (157ms) , to previous reports of latency of neural responses latencies hu man and monkey IT[2,11,23,37]. 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[21,23]. At the subordinate level, the body-specific peak (170ms) and the two face-specific peaks (127ms and 190ms) concur with previous work (for bodies, 170– 260ms[31,38,39] and for faces, first peak at 100–120ms[35] and second peak at 170ms[22,40]). Note that it remains controversial whether the two face peaks have different cortical sources [44-46], a question that future studies comparing representational dissimiliarities in MEG and fMRI may resolve. While our MEG results confirm a body of work[19], they generalize prior 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, by 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 are discriminated by visual representations early, whereas ordinate and superordinate levels emerge relatively later. However, we found that onset latencies of significance for the various categorization levels were all early (48–70ms, except naturalness at 93ms). 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[44,45]. Note that our results might seem to be at odds with prior research suggesting clear multi-stage processing in IT, with a stage of global processing being followed by local processing[35,46,47]. 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 at a specific time 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 [4,5]. 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 compared to monkey electrophysiology[48]. Similarly, most MEG/EEG studies based on ERP analysis investigating content-specific modulation of brain activity by cognitive factors like memory or attention have to rely on a handful of categorical markers in the ERP waveform, e.g. 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[19] might play a similarly fruitful role in the future study of object recognition as the introduction of these techniques did in fMRI[18,20].

Resolving object processing in time, space and species

Relating signals measured in different imaging modalities and combining the methods' respective advantages, are current challenges in systems neuroscience[25]. Using representational similarity analysis, we showed a content-selective correspondence between early MEG signals and V1, and later MEG signals and. Our results match previously reported average onset latencies in the literature, ranging between 50–80ms in V1[9], and from 80–200ms in IT[11,37]. Thus, representational similarity analysis combining MEG and fMRI is a promising method to relate cortical activations across space and time.

Comparing visual representations to themselves 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 clearly limited temporal extent. They might form the basis of memory of images in representational formats that make explicit different properties of these images, e.g. low-level features versus category membership.

An integrated theory of object recognition requires to quantitatively bridge 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[3]. Here, we 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 71–210ms after stimulus onset[26]. Future studies might compare the dynamics in human and monkey IT based on 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 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[25]. Here we provided key advance 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 hundreds 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.

ONLINE METHODS

Participants and experimental design

16 right-handed, healthy volunteers with normal or corrected-to-normal vision (mean ± std. = 25.87 ± 5.38, 10 female) participated in the experiment. The study was conducted according to the Declaration of Helsinki and is approved by the local ethics committee (Institutional Review Board of the Massachusetts Institute of Technology). 15 participants completed two MRI and MEG sessions, and one participant participated in the MEG experiment only. The sample size is comparable to that used in previous fMRI and MEG studies. All participants provided written consent for each of the sessions. During the experiment participants sawee images of 92 different objects presented at the center of the screen (2.9 degrees visual angle, 500ms duration) overlaid with a dark gray fixation cross. We chose this particular dataset for two reasons. First, it allowed assessment of distinctions at three levels: superordinate-, ordinate, and subordinate category. Second, it enabled direct comparison of our MEG and fMRI results with previous experiments utilizing the same date set in monkey electrophysiology and human MRI[3,26]. The presentation parameters were adapted to the specific requirements of each acquisition technique (Supplementary Fig. 1). In detail, for each MEG session, participants completed 10 to 15 runs, each having duration

420s. Each image wais presented twice in each MEG run in random order, with a trial onset asynchrony (TOA) of 1.5 or 2s. Participants were instructed to press a button and blink their eyes in response to a paperclip that was shown randomly every 3 to 5 trials (average 4). For each fMRI session, participants completed 10 to 14 runs, each having duration 384s. Each image wais presented once in each run in random order, with the restriction of not displaying the same condition on consecutive trials. 30 null trials with no stimulus presentation were randomly interspersed, during which the fixation cross turneds darker for 100ms and participants reported the change with a button press. TOA was 3s or 6s in the presence of a null trial.

Human MEG

We acquired continuous MEG signals from 306 channels (204 planar gradiometers, 102 magnetometers, Elektra Neuromag TRIUX, Elekta, Stockholm) at a sampling rate of 1,000Hz, filtered between 0.03 and 330Hz. Raw data was preprocessed using spatiotemporal filters (maxfilter software, Elekta, Stockholm) and then analyzed using Brainstorm[49]. MEG trials were extracted with 100ms baseline and 1,200ms post-stimulus (i.e. 1301ms length), the baseline mean of each channel was removed, and data was temporally smoothed with a 20ms sliding window. A total of 20–30 trials were obtained for each condition, session and participant.

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: (www.csie.ntu.edu.tw/~cjlin/libsvm)[50]. The SVM analysis was conducted independently for each participant and session (Fig. 1a,b). For each time point (100ms before to 1,200ms 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 to train the SVM classifier to pairwise decode any two conditions, with a leave-one-out cross- validation approach. Namely, for each time point and pair of conditions, N–1 pattern vectors comprised the training set and the remaining Nth pattern vector 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, 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 wasis difficult to visualize. To reveal any underlying patterns, we used multidimensional scaling (MDS)[29,30] to plot the data into a 2-D 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 dipping[51] 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 (MRI) 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 = 256mm^2, TR = 1,900ms, TE = 2.52ms, flip angle = 9°). For fMRI, we conducted 10ten–14 to fourteen runs in which 192 volumes were acquired for each participant (gradient-echo EPI sequence: TR = 2,000ms, TE = 31ms, flip angle = 80°, FOV read = 192 mm, FOV phase = 100%, ascending acquisition, gap = 10%, resolution = 2mm isotropic, slices = 25). The acquisition volume covered the occipital and temporal lobe and is oriented parallel to the temporal cortex.

Human fMRI analysis

FMRI data was processed using SPM8 (www.fil.ion.ucl.ac.uk/spm). For each participant and session separately, data was 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 3 runs of each session, and one comprising the remaining runs. The onsets and durations of each image presentation as well as 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 based on an anatomical eccentricity template[52]. The cortical surface of each participant was reconstructed with FreeSurfer based on the T1 structural scan[53]. The right hemisphere was mirror-reversed and registered to the left hemisphere. This allowed us to register a V1 eccentricity template[52] 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 Atlas[54]). 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, for each participant and session we restricted the definition of IT: 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 3 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 V1, peripheral V1, or 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 the result was stored in a 92 × 92 symmetric matrix. We converted the correlations into a dissimilarity measure 1–R[3], 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 elsewhere[26]. 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 (105ms 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 71–210ms 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 Nikolaus Kriegeskorte and Roozbeh Kiani[3,26].

Significance testing

We used non-parametric statistical inference[27,28] that does not make assumptions on 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) on 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 differeds 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 wasis 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 couldan permute the condition labels of the MEG data, which effectively corresponds to a sign permutation test that randomly multiplies the participant-specific data (e.g. 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 (e.g. MEG decoding time series, time-time decoding matrices, etc.) into 1-D or 2-D 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 1D and $p<0.0001$ for 2D to define supra-threshold clusters. These supra-threshold 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 supra-threshold 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 (i.e. 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 yieldeds 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This work was funded by NEI grant EY020484 (to A.O.), NSF grant BCS-1134780 (to D.P.), Humboldt Scholarship (to R.M.C), and was conducted at the Athinoula A. Martinos Imaging Center at McGovern Institute for Brain Research, MIT.

REFERENCES

1. Grill-Spector K, Malach R. The Human Visual Cortex. Annu. Rev. Neurosci. 2004; 27:649–677. [PubMed: 15217346]
2. Hung CP, Kreiman G, Poggio T, DiCarlo JJ. Fast Readout of Object Identity from Macaque Inferior Temporal Cortex. Science. 2005; 310:863–866. [PubMed: 16272124]

3. Kriegeskorte N, et al. Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. Neuron. 2008; 60:1126–1141. [PubMed: 19109916]
4. Kourtzi Z, Connor CE. Neural Representations for Object Perception: Structure, Category, and Adaptive Coding. Annu. Rev. Neurosci. 2011; 34:45–67. [PubMed: 21438683]
5. DiCarlo JJ, Zoccolan D, Rust NC. How Does the Brain Solve Visual Object Recognition? Neuron. 2012; 73:415–434. [PubMed: 22325196]
6. Felleman DJ, Van Essen DC. Distributed Hierarchical Processing in the Primate Cerebral Cortex. Cereb. Cortex. 1991; 1:1–a–47. [PubMed: 1822724]
7. Ungerleider, LG.; Mishkin, M. Analysis of Visual Behavior. MIT Press; 1982. Two visual systems.; p. 549-586.
8. Milner, AD.; Goodale, MA. The visual brain in action. Oxford University Press; 2006.
9. Schmolesky MT, et al. Signal Timing Across the Macaque Visual System. J. Neurophysiol. 1998; 79:3272–3278. [PubMed: 9636126]
10. Luck, SJ. An Introduction to the Event-Related Potential Technique. Mit Press; 2005.
11. Mormann F, et al. Latency and Selectivity of Single Neurons Indicate Hierarchical Processing in the Human Medial Temporal Lobe. J. Neurosci. 2008; 28:8865–8872. [PubMed: 18768680]
12. Baillet S, Mosher JC, Leahy RM. Electromagnetic brain mapping. IEEE Sign. Proc. Mag. 2001; 18:14–30.
13. Hari R, Salmelin R. Magnetoencephalography: From SQUIDs to neuroscience: Neuroimage 20th Anniversary Special Edition. Neuroimage. 2012; 61:386–396. [PubMed: 22166794]
14. Dale AM, et al. Dynamic Statistical Parametric Mapping: Combining fMRI and MEG for High-Resolution Imaging of Cortical Activity. Neuron. 2000; 26:55–67. [PubMed: 10798392]
15. Debener S, Ullsperger M, Siegel M, Engel AK. Single-trial EEG–fMRI reveals the dynamics of cognitive function. Trends Cogn. Sci. 2006; 10:558–563. [PubMed: 17074530]
16. Logothetis NK, Sheinberg DL. Visual object recognition. Annu. Rev. Neurosci. 1996; 19:577–621. [PubMed: 8833455]
17. Carlson TA, Hogendoorn H, Kanai R, Mesik J, Turret J. High temporal resolution decoding of object position and category. J. Vis. 11. 2011
18. Haynes J-D, Rees G. Decoding mental states from brain activity in humans. Nat. Rev. Neurosci. 2006; 7:523–534. [PubMed: 16791142]
19. Carlson T, Tovar DA, Alink A, Kriegeskorte N. Representational dynamics of object vision: The first 1000 ms. J. Vis. 2013; 13
20. Tong F, Pratte MS. Decoding Patterns of Human Brain Activity. Ann. Rev. Psychol. 2012; 63:483–509. [PubMed: 21943172]
21. Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system. Nature. 1996; 381:520–522. [PubMed: 8632824]
22. Bentin S, Allison T, Puce A, Perez E, McCarthy G. Electrophysiological Studies of Face Perception in Humans. J. Cogn. Neurosci. 1996; 8:551–565. [PubMed: 20740065]
23. VanRullen R, Thorpe SJ. The time course of visual processing: from early perception to decision-making. J. Cogn. Neurosci. 2001; 13:454–461. [PubMed: 11388919]
24. Edelman S. Representation is representation of similarities. Behav. Brain. Sci. 1998; 21:449–467. discussion 467–498. [PubMed: 10097019]
25. Kriegeskorte N. Representational similarity analysis – connecting the branches of systems neuroscience. Front. Sys. Neurosci. 2008; 2:4.
26. Kiani R, Esteky H, Mirpour K, Tanaka K. Object Category Structure in Response Patterns of Neuronal Population in Monkey Inferior Temporal Cortex. J. Neurophysiol. 2007; 97:4296–4309. [PubMed: 17428910]
27. Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: A primer with examples. Hum. Brain. Mapp. 2002; 15:1–25. [PubMed: 11747097]
28. Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods. 2007; 164:177–190. [PubMed: 17517438]
29. Kruskal, JB.; Wish, M. Multidimensional Scaling. SAGE; 1978.

30. Shepard RN. Multidimensional Scaling, Tree-Fitting, and Clustering. Science. 1980; 210:390–398. [PubMed: 17837406]

31. Allison T, et al. Face recognition in human extrastriate cortex. J. Neurophysiol. 1994; 71:821–825. [PubMed: 8176446]

32. Kanwisher N, McDermott J, Chun MM. The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. J. Neurosci. 1997; 17:4302–4311. [PubMed: 9151747]

33. McCarthy G, Puce A, Belger A, Allison T. Electrophysiological Studies of Human Face Perception. II: Response Properties of Face-specific Potentials Generated in Occipitotemporal Cortex. Cereb. Cortex. 1999; 9:431–444. [PubMed: 10450889]

34. Downing PE, Jiang Y, Shuman M, Kanwisher N. A Cortical Area Selective for Visual Processing of the Human Body. Science. 2001; 293:2470–2473. [PubMed: 11577239]

35. Liu J, Harris A, Kanwisher N. Stages of processing in face perception: an MEG study. Nat. Neurosci. 2002; 5:910–916. [PubMed: 12195430]

36. Harrison SA, Tong F. Decoding reveals the contents of visual working memory in early visual areas. Nature. 2009; 458:632–635. [PubMed: 19225460]

37. Liu H, Agam Y, Madsen JR, Kreiman G. Timing, Timing, Timing: Fast Decoding of Object Information from Intracranial Field Potentials in Human Visual Cortex. Neuron. 2009; 62:281–290. [PubMed: 19409272]

38. Stekelenburg JJ, de Gelder B. The neural correlates of perceiving human bodies: an ERP study on the body-inversion effect. Neuroreport. 2004; 15:777–780. [PubMed: 15073513]

39. Thierry G, et al. An event-related potential component sensitive to images of the human body. Neuroimage. 2006; 32:871–879. [PubMed: 16750639]

40. Jeffreys DA. Evoked Potential Studies of Face and Object Processing. Visual Cognition. 1996; 3:1–38.

41. Halgren E, Raij T, Marinkovic K, Jousmäki V, Hari R. Cognitive Response Profile of the Human Fusiform Face Area as Determined by MEG. Cereb. Cortex. 2000; 10:69–81. [PubMed: 10639397]

42. Sadeh B, Podlipsky I, Zhdanov A, Yovel G. Event-related potential and functional MRI measures of face-selectivity are highly correlated: A simultaneous ERP-fMRI investigation. Mapp. 2010; 31:1490–1501.

43. Tsao DY, Freiwald WA, Tootell RBH, Livingstone MS. A Cortical Region Consisting Entirely of Face-Selective Cells. Science. 2006; 311:670–674. [PubMed: 16456083]

44. Mack ML, Palmeri TJ. The timing of visual object categorization. Front. Psychol. 2011; 2:165. [PubMed: 21811480]

45. Kravitz DJ, Saleem KS, Baker CI, Ungerleider LG, Mishkin M. The ventral visual pathway: an expanded neural framework for the processing of object quality. Trend. Cogn. Sci. 2013; 17:26–49.

46. Sugase-Miyamoto Y, Matsumoto N, Kawano K. Role of Temporal Processing Stages by Inferior Temporal Neurons in Facial Recognition. Front. Psychol. 2011; 2

47. Brincat SL, Connor CE. Dynamic Shape Synthesis in Posterior Inferotemporal Cortex. Neuron. 2006; 49:17–24. [PubMed: 16387636]

48. Freiwald WA, Tsao DY. Functional Compartmentalization and Viewpoint Generalization Within the Macaque Face-Processing System. Science. 2010; 330:845–851. [PubMed: 21051642]

49. Tadel F, Baillet S, Mosher JC, Pantazis D, Leahy RM. Brainstorm: A User-Friendly Application for MEG/EEG Analysis. Comp. Intell. Neurosci. 2011; 2011:1–13.

50. Müller K, Mika S, Rätsch G, Tsuda K, Schölkopf B. An introduction to kernel-based learning algorithms. IEEE Trans. Neural. Netw. 2001; 12:181–201. [PubMed: 18244377]

51. Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. Nat. Neurosci. 2009; 12:535–540. [PubMed: 19396166]

52. Benson NC, et al. The Retinotopic Organization of Striate Cortex Is Well Predicted by Surface Topology. Curr. Biol. 2012; 22:2081–2085. [PubMed: 23041195]

53. Dale AM, Fischl B, Sereno MI. Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. Neuroimage. 1999; 9:179–194. [PubMed: 9931268]

54. Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage. 2003; 19:1233–1239. [PubMed: 12880848]

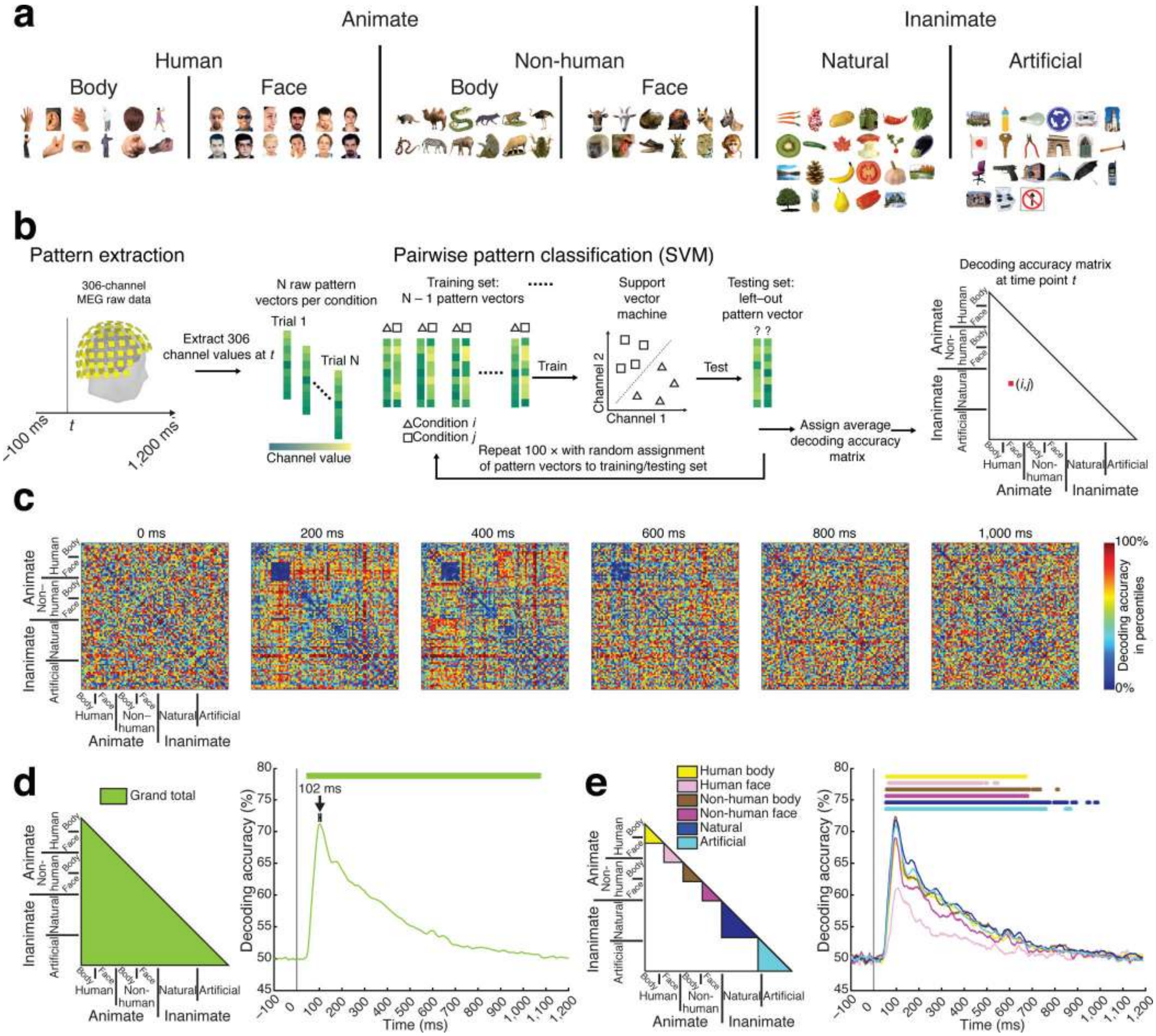

Figure 1. Decoding of images from MEG signals

(a) Image set of. 92 images[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 48ms (45–51ms), with a peak at 102ms (98–107ms; 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. Stars 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.

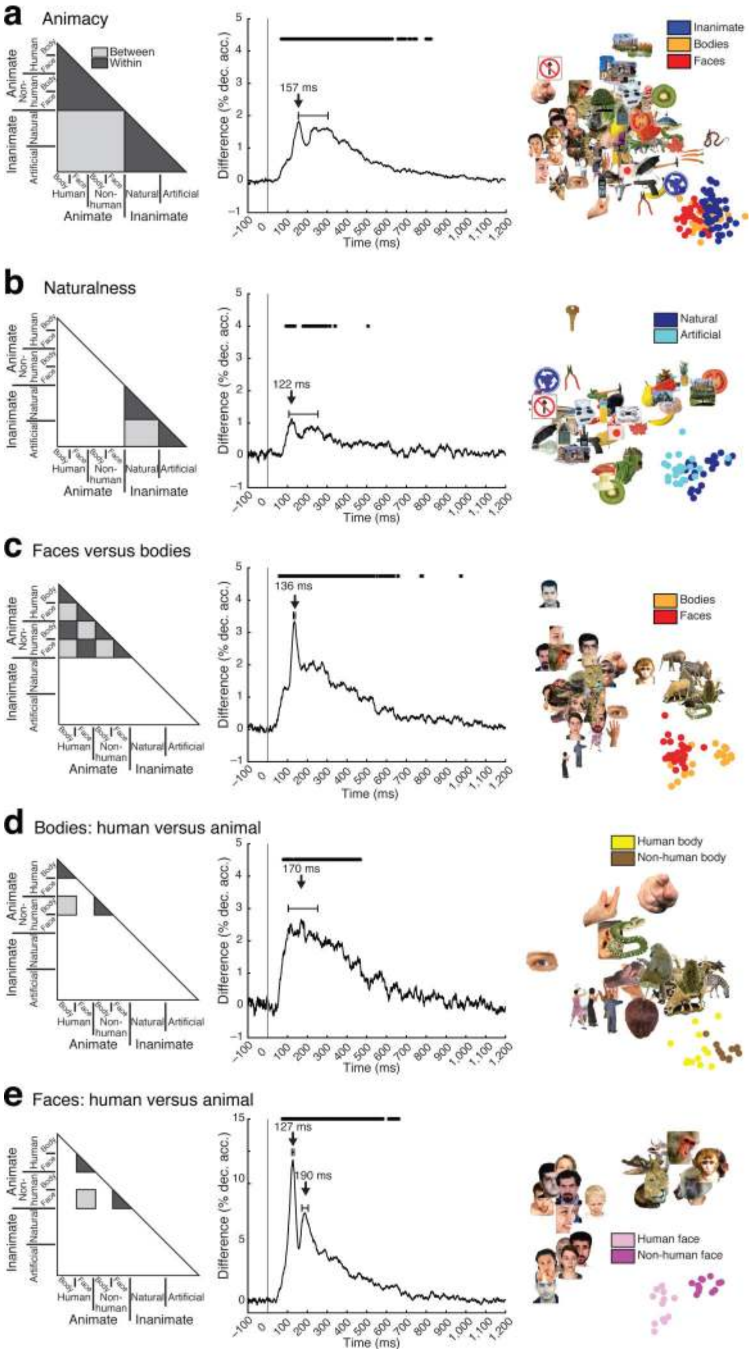

Figure 2.

Time course of decoding category membership of individual objects. We decoded object category membership for, **(a)** animacy, **(b)** naturalness, **(c)** faces versus bodies, **(d)** human bodies versus non-human bodies and **(e)** human versus non-human faces. The difference of within-subdivision (dark gray, left panel) minus between-subdivision (light gray, left panel) Peaks in decoding accuracy differences indicate time points at which the ratio of dissimilarity within a subdivision to dissimilarity across subdivision is smallest. *n*=16, stars, vertical gray line, and error bars same as in Figure 1. Statistical details are 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 multidimensional scaling solution (MDS, criterion: metric stress, 0.24 for (a,b,c,e), 0.27 for (d)). Abbreviations: dec. acc. = decoding accuracy.

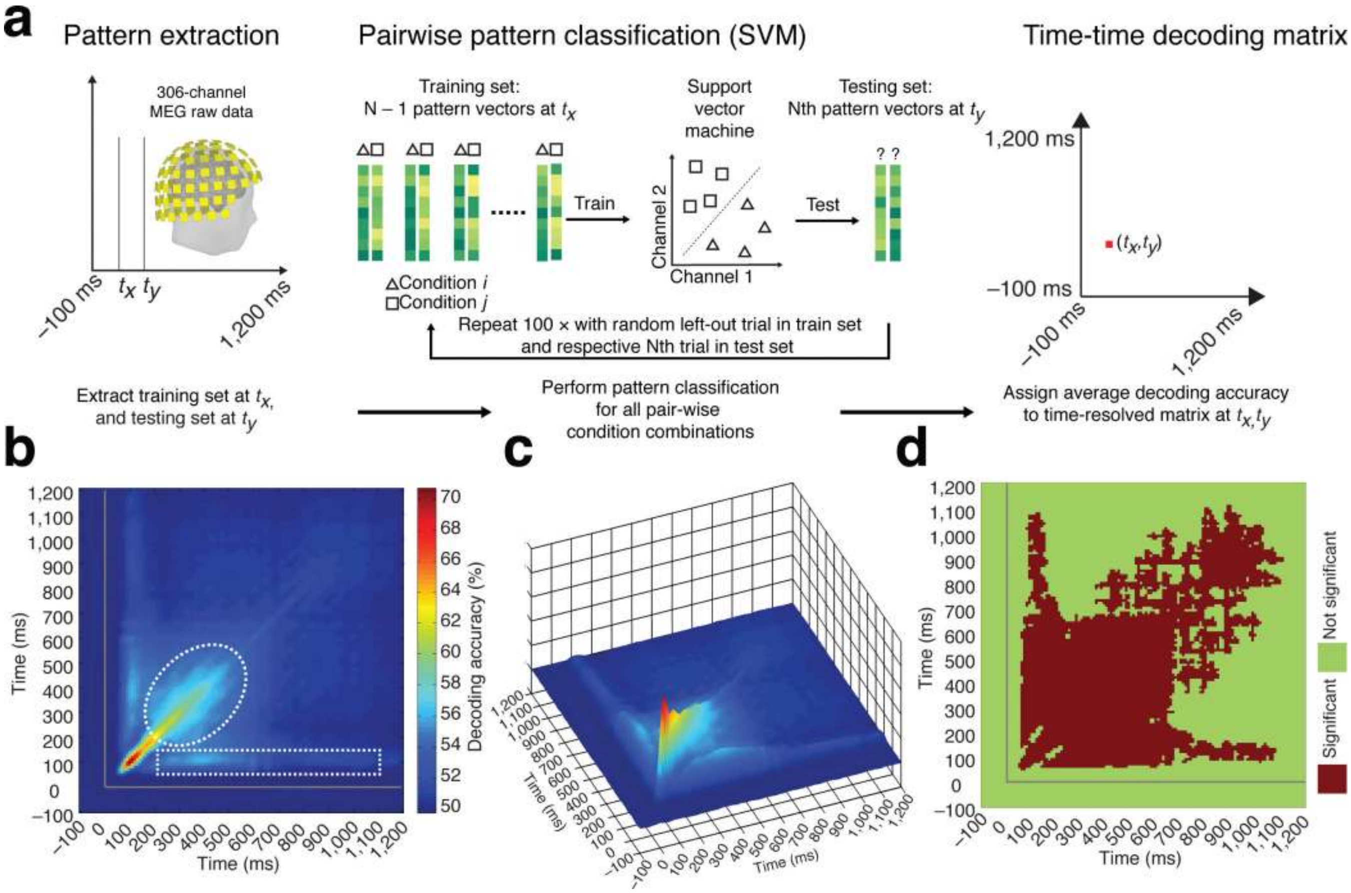

Figure 3.

Dynamics of visual representations across time. **(a)** MEG brain responses were extracted for time points t_x and t_y after stimulus onset. SVM wasis trained to distinguish between images by visual representations at time point t_x, and tested on brain responses to the same images at a different time point t_y. We conducted all object classifications and averaged the overall decoding accuracy. Last, the averaged decoding accuracy was stored in the element (t_x,t_y) of a time-time MEG decoding matrix. The process was repeated for all pairs of time points. **(b,c)** Time-time decoding matrix averaged across participants. The gray lines indicate onset of image presentation. The white dotted rectangle indicates classifier generalization for the time-point combination ~100ms and 200-1,000ms, the dotted ellipse indicates classifier generalization by the broadened diagonal. **(d)** Significance was assessed by sign-permutation tests (*n*=16, cluster-defining threshold p<0.0001, corrected significance level p<0.05). Dark red indicates elements within the significant cluster.

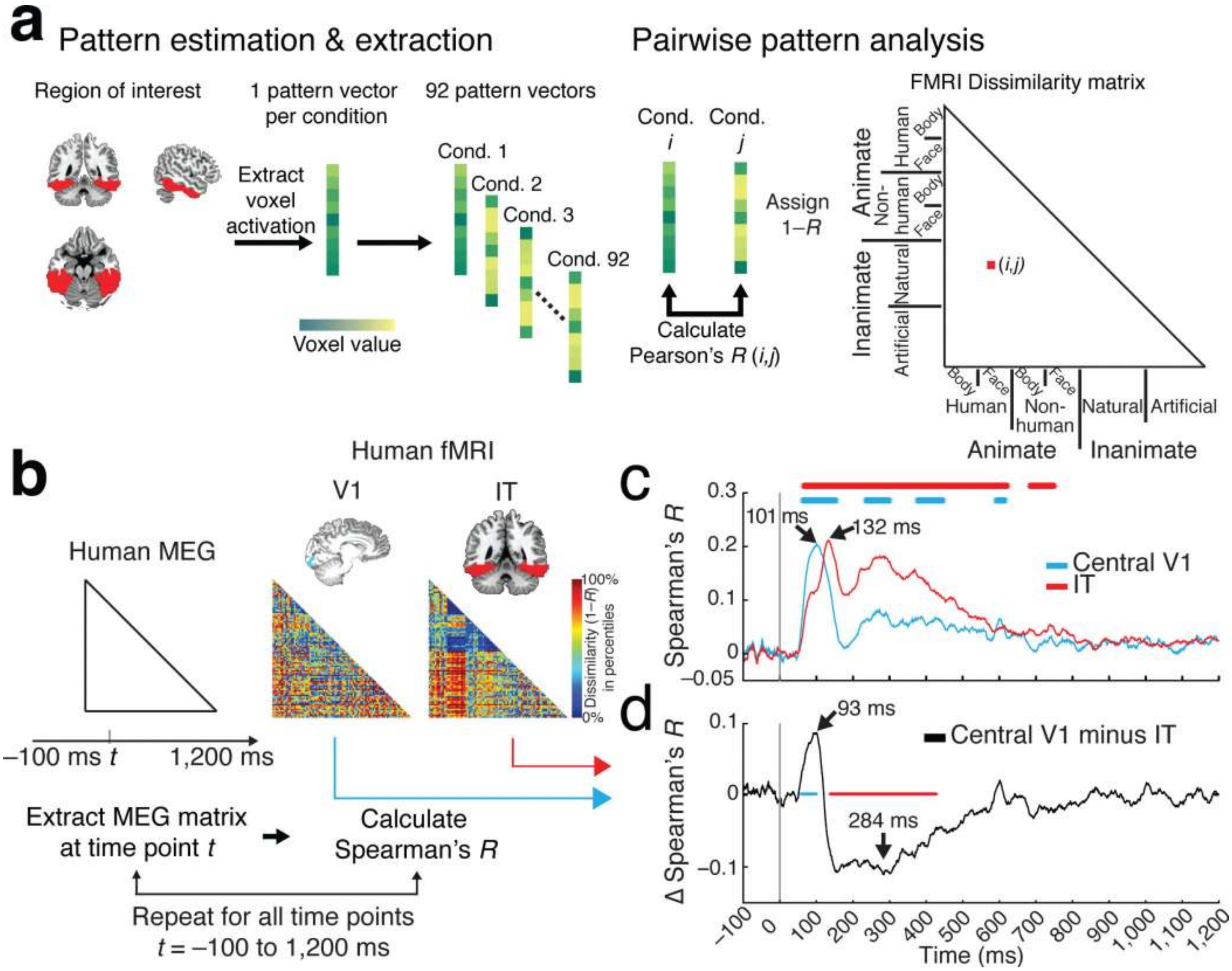

Figure 4.

Relating MEG and fMRI signals in V1 and IT. **(a)** We select voxels in two regions-of-interest (ROI): V1 and . IT, For each condition, we extracted voxel activation values, yielding 92 pattern vectors., we calculated 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 ROI. **(b)**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. **(c)**MEG signals correlated with V1 earlier than with IT. Blue and red stars indicate significant time points for V1 and IT. **(d)** Difference curve between the two curves as in (**c**). MEG correlated early more with V1 than with IT, and later more with IT than with V1. Green and red stars in the plots indicate significant time points for positive and negative clusters respectively. For details see Supplementary Table 1d. *n*=16, Gray line and statistical procedure same as in Fig. 1.

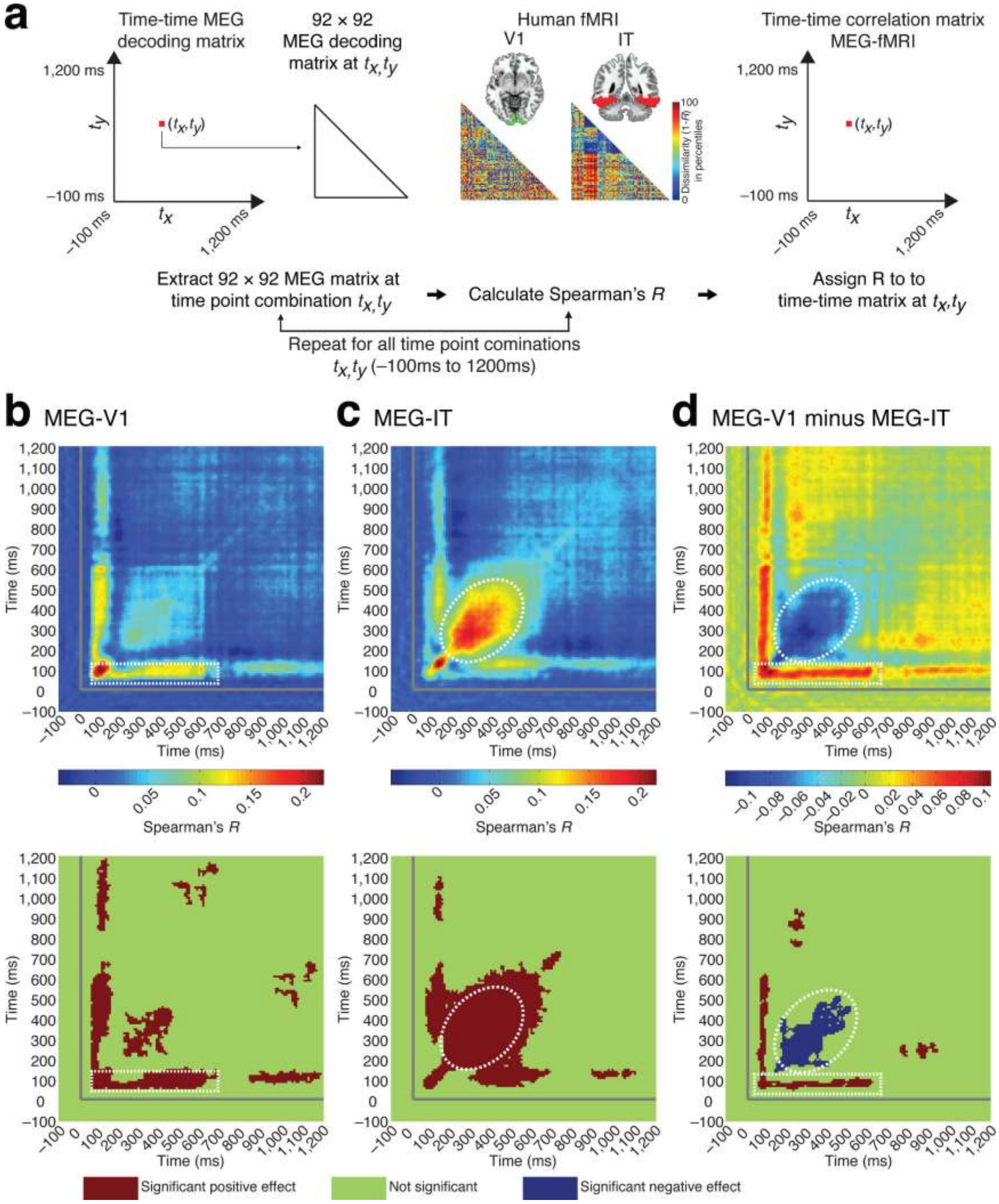

Figure 5.

Relating MEG and fMRI signals across time. **(a)** 92 × 92 decoding matrices were extracted for each combination of time points (t_x, t_y), and were correlated (Spearman's rank-order correlation) with the fMRI dissimilarity matrices for V1 and IT. The resulting correlation was assigned to a time-time MEG-fMRI correlation matrix at t_x,t_y. (**b,c)** Top panel displays the time-time MEG and fMRI correlation matrix for V1 and IT. Bottom panel shows significant cluster results (*n*=16, cluster-defining threshold p<0.0001, corrected significance level p<0.05). Neural activity for the time point combinations of ~100ms and ~200–1,000ms (marked by white dotted rectangle) correlated with V1. **(c)** Neural activity between ~250ms and ~500ms (marked by the striped white ellipse) correlated with IT. **(d)** Difference between V1 and IT. Gray lines as in Figure 1.

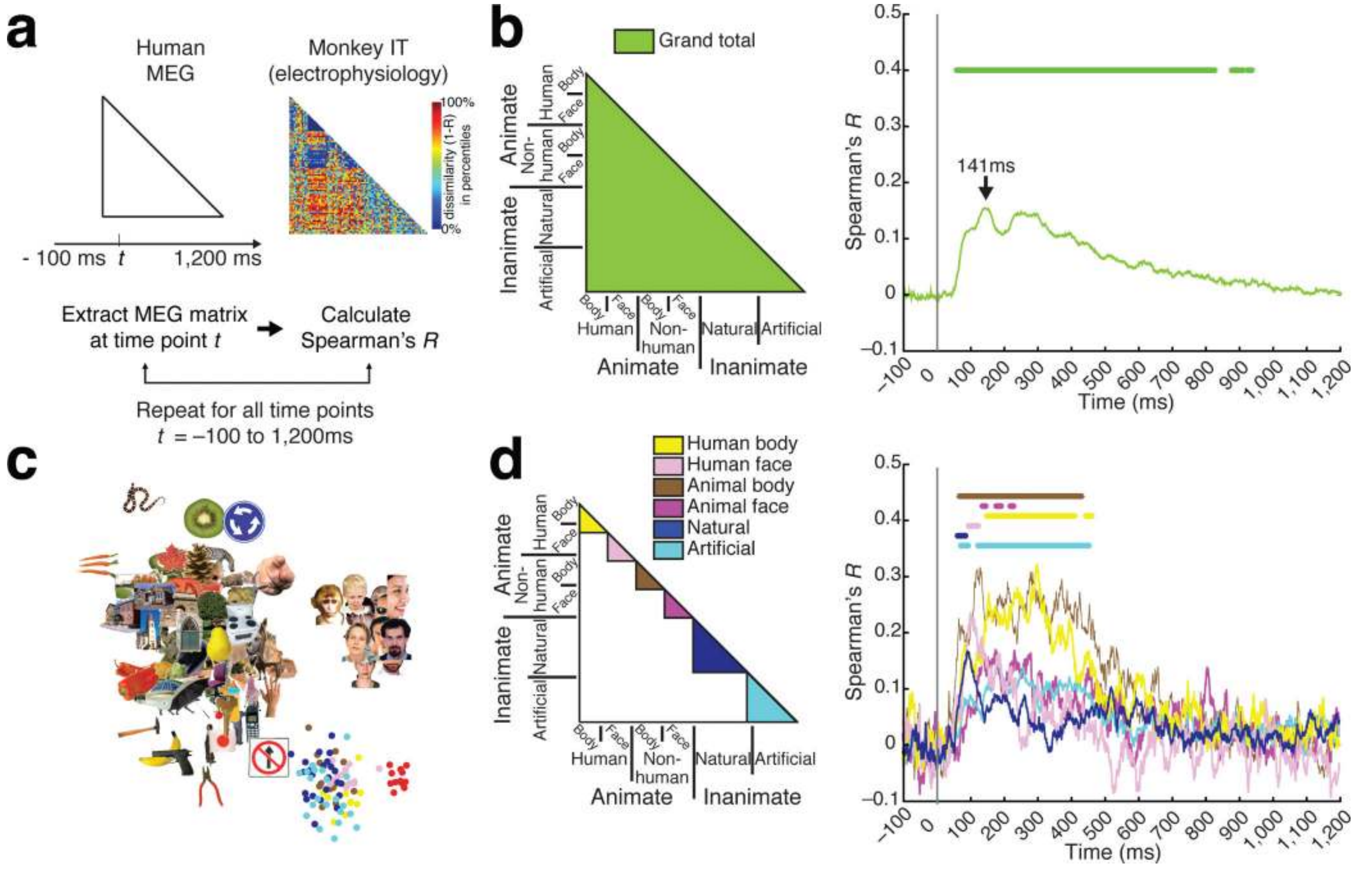

Figure 6.

Relating human MEG and electrophysiological signals in monkey IT. **(a)** The MEG decoding matrix at time *t* was compared (Spearman's rank-order correlation *R*) against the monkey dissimilarity matrix in IT,. The lower form of the monkey IT matrix is shown as percentiles of 1–*R*. - **(b)** Representational dissimilarities inhuman MEG and monkey IT correlated significantly starting at 54ms (52–64ms), with a peak latency of 141ms (132–292ms). **c)** MDS at peak-latency. **d)** Results at the fine-grained level of the image set. Representational dissimilarities were similar across species and methods even for the finest categorical subdivision of the image set. *n*=16, stars and gray line same as in Figure 1.