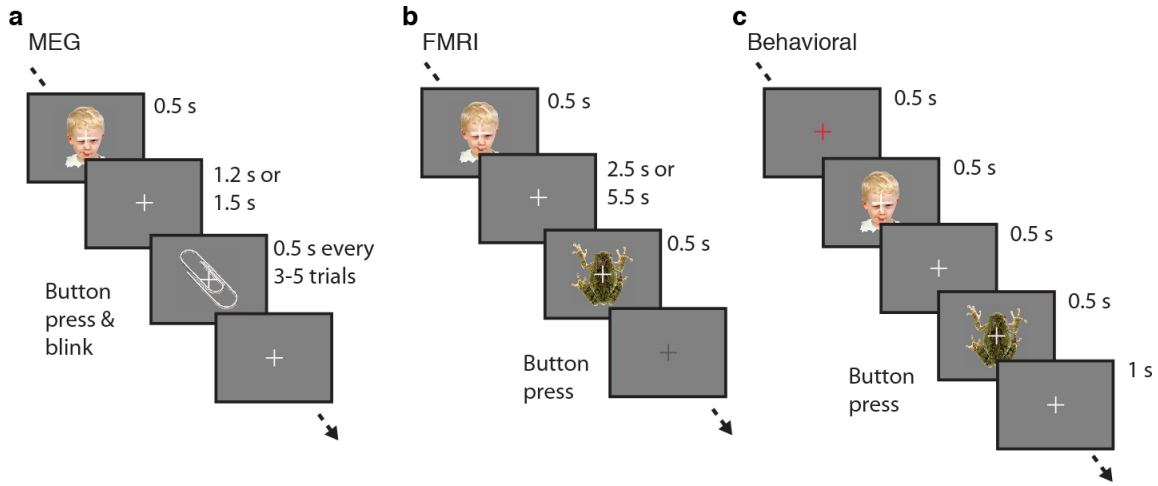
Supplementary Information

Radoslaw Martin Cichy, Dimitrios Pantazis, Aude Oliva
Resolving human object recognition in space and time

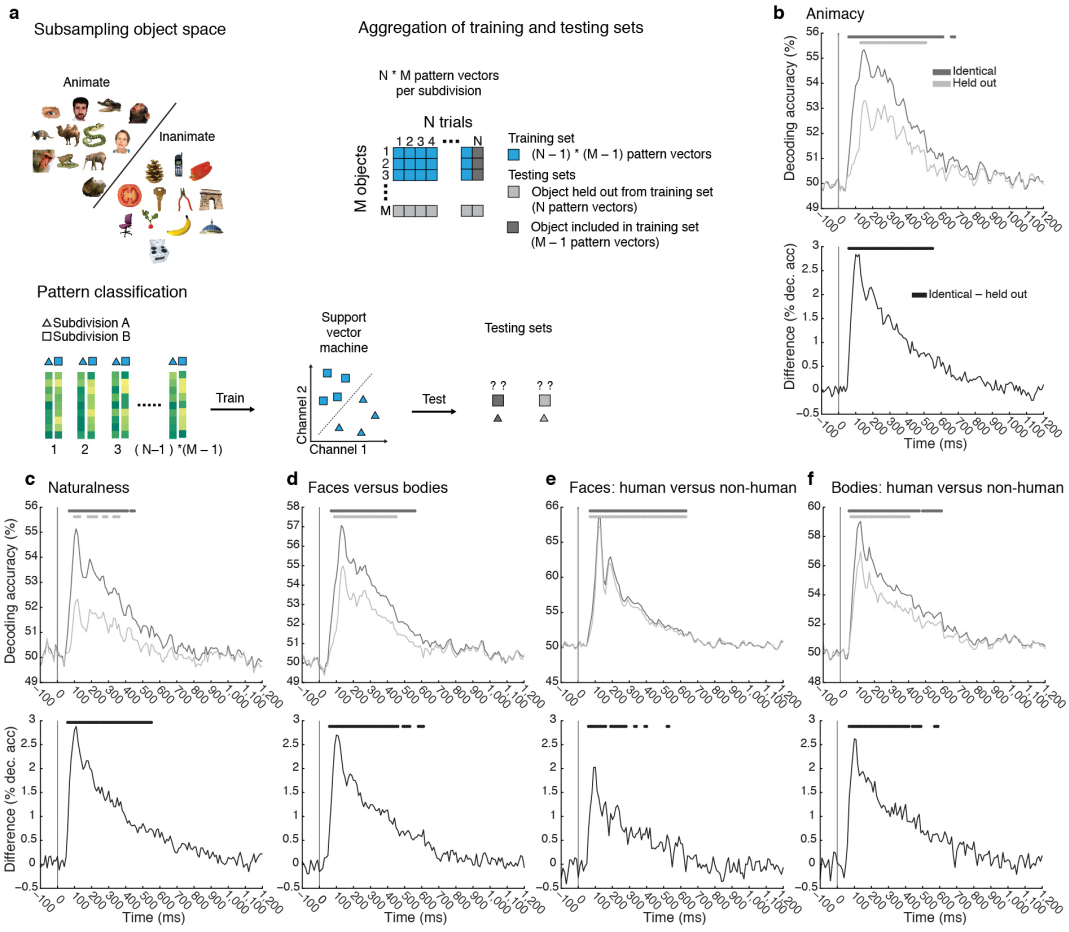
Nature Neuroscience: doi:10.1038/nn.3635

SUPPLEMENTARY FIGURES



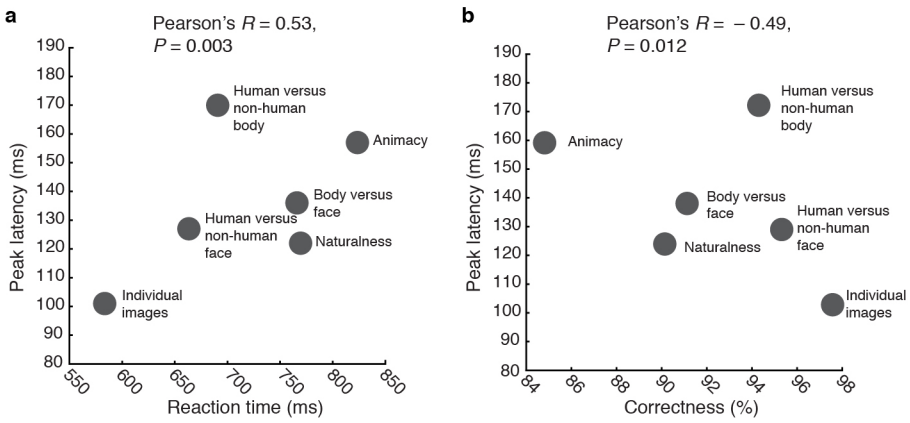
Supplementary Figure 1 Experimental design in MEG, fMRI and behavioral experiments. Participants (n = 16) viewed the same 92 images (2.9 degrees visual angle overlaid with a gray fixation cross). (**a**) For MEG, images were presented in random order every 1.5 – 2 s. Every 3 – 5 trials, a paper clip was presented prompting a button press response. (**b**) For fMRI, stimulus onset asynchrony was 3s, or 6s when a null trial (uniform gray background) was shown. During null trials the fixation cross changed to dark gray, prompting a button press response. (**c**) For behavioral testing, participants classified pairs of images either by identity (same/different image) or by category for 5 different categorizations: animacy, naturalness, face versus body, human versus non-human body, human versus non-human face in blocks of 24 trials each. Before every block, participants received instructions about the categorization task (e.g. animate versus inanimate). Each trial consisted of a red fixation cross (0.5 s) then two images (0.5 s, separating offset 0.5 s). Participants were instructed to respond as fast and accurately as possible, indicating whether the two images were same or different with respect to the instructed classification by pressing a button. Participants completed 8 runs, each consisting of a random sequence of the 6 blocks, given the 6 classification tasks. Results (reaction times for correct responses, and percent correct responses) were determined for each block and then averaged by participant.

Nature Neuroscience: doi:10.1038/nn.3635



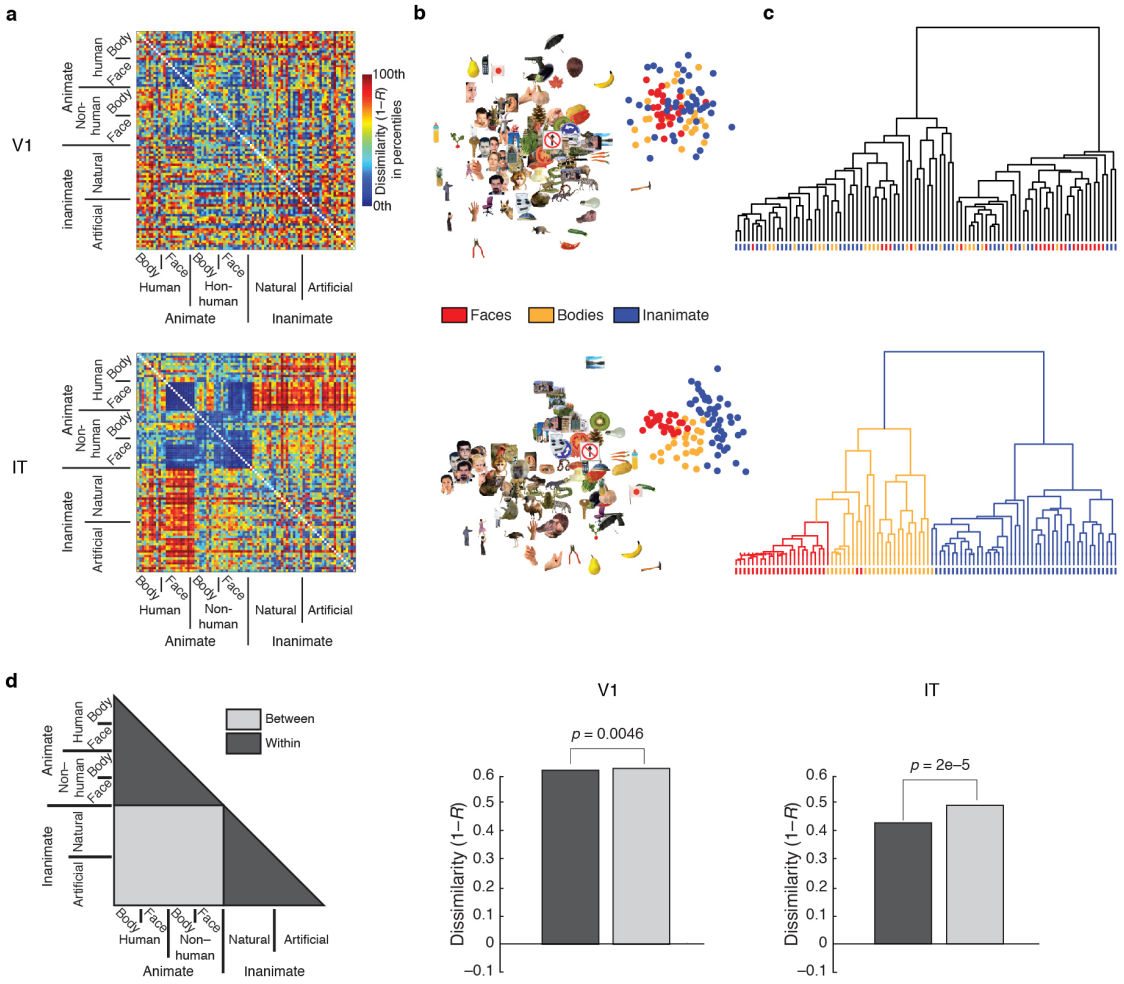
Supplementary Figure 2 Linear separability of categorical subdivisions. (**a**) We determined whether the membership of an image to a category (here shown for animacy) can be linearly discriminated by visual representation directly. Analysis was conducted independently for each participant and session, and for each time point from –100 to 1200 ms in 10ms steps. For each category subdivision, we subsampled the set of objects by randomly drawing *M* (12) objects. Each object was presented N times. We assigned (*N*–1) × (*M*–1) trials to a training set of a linearized SVM (liblinear, http://www.csie.ntu.edu.tw/~cjlin/liblinear/) in the L2-regularized L2-loss SVM (primal) configuration. We tested the SVM on independent trials in two ways: from objects included in the training set ('identical' condition, dark gray), or held out from the training set ('held-out' condition, light gray). We repeated the above procedure 100 times, using different subsamples of objects and random assignment of trials to training and testing sets. Decoding accuracy was averaged across repetitions. **(b–f)** The upper panel shows the decoding accuracy time courses for objects included or held-out from the training set (color-coded as in (a)). The lower panel illustrates the difference of decoding accuracy between identical and held-out objects. Stars indicate time points with significant effects (sign-permutation test, $n = 16$, cluster-defining threshold $P < 0.001$, corrected significance level $P < 0.05$). For details see Supplementary Table 1e. Abbreviations: dec. acc. = decoding accuracy. Vertical gray line indicates image onset.

Nature Neuroscience: doi:10.1038/nn.3635



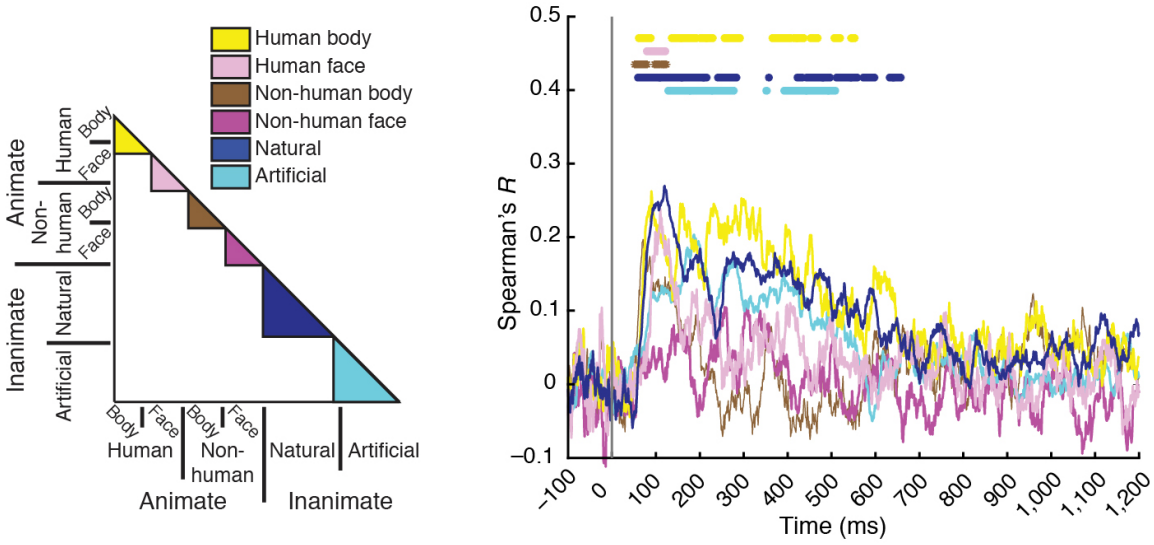
Supplementary Figure 3 Relation of behavior to peak latency of decoding accuracy. We determined whether (**a**) reaction time and (**b**) correctness are linearly related to peak latency of decoding accuracy (Pearson's R). We assessed significance by bootstrapping the sample of participants ($n = 16$, $P < 0.05$). Reaction time shows a positive relationship ($R = 0.53$, $P = 0.003$); correctness a negative relationship ($R = –0.49$, $P = 0.012$).

Nature Neuroscience: doi:10.1038/nn.3635



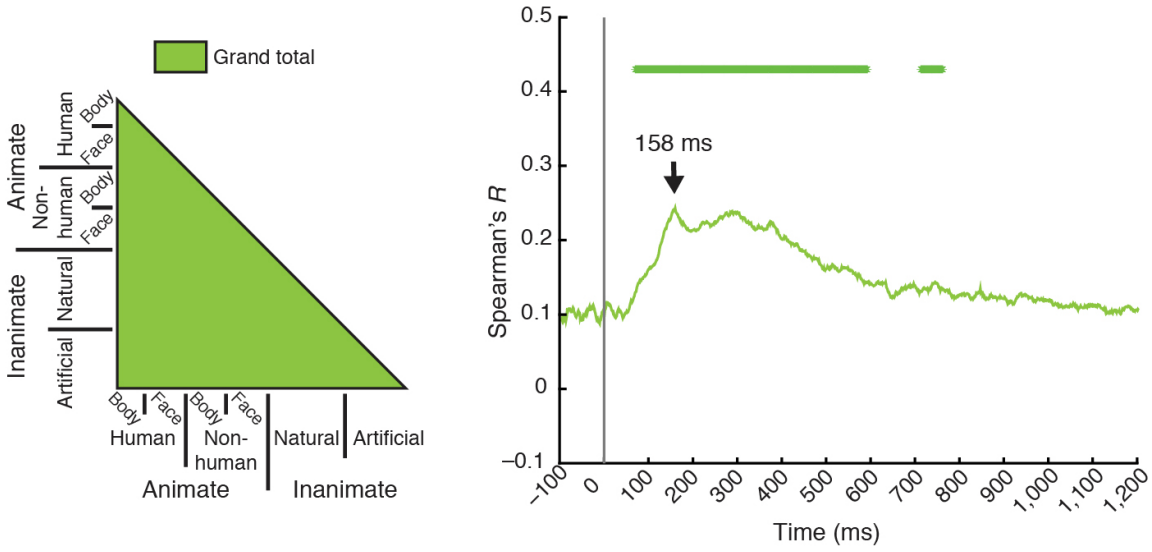
Supplementary Figure 4 Representational similarity analysis of fMRI responses in human V1 and IT. Our analyses corroborated previous major findings[3] by a random-effects analysis. (**a**) Representational dissimilarity matrices for human V1 and IT. Dissimilarity between fMRI pattern responses is color-coded as percentiles of dissimilarity (1– Spearman's *R*). (**b**) MDS and (**c**) hierarchical clustering of fMRI responses. MDS (criterion: metric stress) showed a grouping of images into inanimate objects, faces, and bodies in IT (stress = 0.24), but not in V1 (stress = 0.20). Unsupervised hierarchical clustering (criterion: average fMRI response pattern dissimilarity) revealed a nested hierarchical structure dividing animate and inanimate objects, and animates into faces and bodies in IT, but not in V1. (**d**) We compared dissimilarity (1 – Spearman's *R*) within versus between the subdivision of animate and inanimate objects. A large animacy effect was observed in IT, and a small effect in V1. A sign permutation test (n = 15, 50,000 iterations) showed that the effect was significant both in IT (P = 2e – 5) and in V1 (P = 0.0046), and significantly larger in IT (P = 2e – 5).

Nature Neuroscience: doi:10.1038/nn.3635



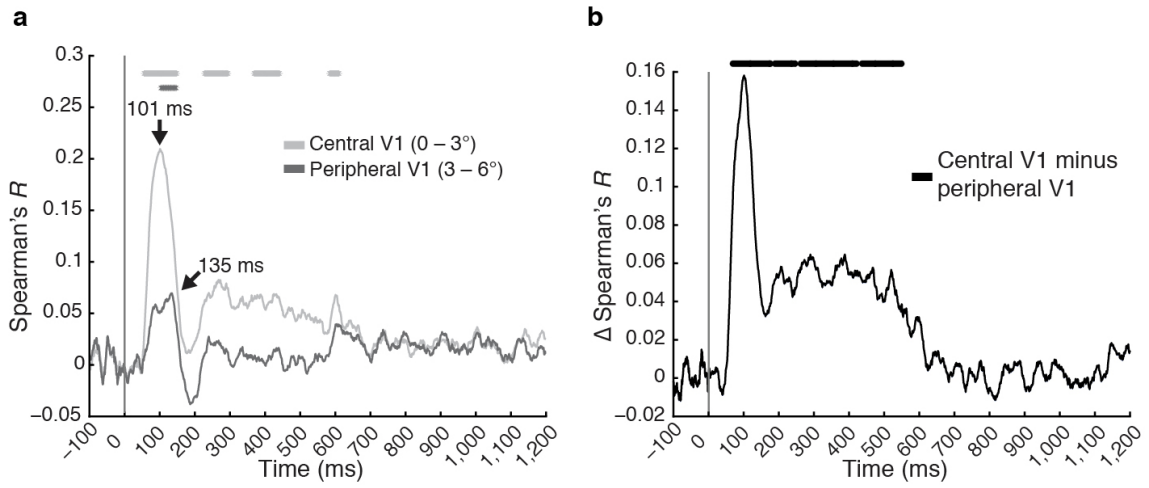
Supplementary Figure 5 Representational similarity analysis related MEG and fMRI responses in IT for the six subdivisions of the image set. Representational dissimilarities were similar for all subdivisions except non-human faces. Stars above the time course indicate time points of statistical significance (sign permutation test, $n = 16$, cluster-defining threshold $P < 0.001$, corrected significance level $P < 0.05$). For details see Supplementary Table 1f.

Nature Neuroscience: doi:10.1038/nn.3635



Supplementary Figure 6 Representational similarity analysis related MEG and fMRI responses in human IT based on previously reported fMRI data. MEG correlated significantly with human IT: onset at 68 ms (57 – 71 ms), peak at 158 ms (152 – 300 ms), showing reproducibility of effects across distinct data sets[3]. Stars above the time course indicate time points of statistical significance (sign-permutation test, $n = 16$, cluster-defining threshold $P < 0.001$, corrected significance level $P < 0.05$).

Nature Neuroscience: doi:10.1038/nn.3635



Supplementary Figure 7 Representational similarity analysis related MEG and fMRI for central and peripheral V1. (**a**) fMRI signals in both central and peripheral V1 correlated with early MEG signals (for details see Supplementary Table 1d). **(b)** MEG signals correlated more strongly with fMRI signals in central than peripheral V1, demonstrating the refined spatial specificity achieved by combining MEG and fMRI by representational similarity analysis. Stars above the time course indicate time points of statistical significance (sign-permutation test, $n = 16$, cluster-defining threshold $P < 0.001$, corrected significance level $P < 0.05$).

Nature Neuroscience: doi:10.1038/nn.3635

SUPPLEMENTARY TABLES

			Subordinate		Ordinate	Superordinate	
		Individual images	Human versus non-human body	Human versus non-human face	Body versus face	Naturalness	Animacy
	Individual images		0.00638	2e – 5*	2e – 5*	0.0002*	0.0001*
Sub-ordinate	Human versus non-human body			0.6938	0.6684	0.6017	0.3528
	Human versus non-human face				0.0006	0.6559	0.0001*
Ordi-nate	Body versus Face					0.7815	0.0010*
Super-ordinate	Naturalness						0.1251
	Animacy						

Supplementary Table 2 Comparison of peak latencies for discrimination of individual images at different levels of categorization. The table reports *P*-values determined by bootstrapping the sample of participants (50,000 samples). Significant comparisons are indexed with a star ($P < 0.05$, Bonferroni corrected). Latency differences between the classifications of 'Human versus non-human body' and 'Individual images' were in line with predictions, but did not pass Bonferroni correction.

Nature Neuroscience: doi:10.1038/nn.3635

SUPPLEMENTARY MOVIE

Supplementary Movie 1 Decoding accuracy matrices and accompanying MDS solutions. To allow a temporally unbiased and complete view of the MEG decoding accuracy data, we generated a movie from –100 to +1,000 ms in 1 ms steps, showing the averaged decoding accuracy across participants and the respective MDS solution (first two dimensions). To allow comparison of the common structure in the MDS across time, we used Procrustes alignment between the first two dimensions of the MDS solutions at neighboring time points.

Nature Neuroscience: doi:10.1038/nn.3635