Neuron
Article

Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey

Nikolaus Kriegeskorte,[1,*] Marieke Mur,[1,2] Douglas A. Ruff,[1] Roozbeh Kiani,[3] Jerzy Bodurka,[1,4] Hossein Esteky,[5,6] Keiji Tanaka,[7] and Peter A. Bandettini[1,4]
[1]Section on Functional Imaging Methods, Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892-1148, USA
[2]Department of Cognitive Neuroscience, Faculty of Psychology, Maastricht University, 6229 ER Maastricht, The Netherlands
[3]Department of Neurobiology and Behavior, University of Washington, Seattle, WA 98195-7270, USA
[4]Functional Magnetic Resonance Imaging Facility, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892-1148, USA
[5]Research Group for Brain and Cognitive Sciences, School of Medicine, Shaheed Beheshti University, Tehran, Iran
[6]School of Cognitive Sciences, Institute for Studies in Theoretical Physics and Mathematics, Niavaran, Tehran, Iran
[7]Cognitive Brain Mapping Laboratory, RIKEN Brain Science Institute, Wako, Saitama 351-0198, Japan
*Correspondence: kriegeskorten@mail.nih.gov
DOI 10.1016/j.neuron.2008.10.043

SUMMARY

Inferior temporal (IT) object representations have been intensively studied in monkeys and humans, but representations of the same particular objects have never been compared between the species. Moreover, IT's role in categorization is not well understood. Here, we presented monkeys and humans with the same images of real-world objects and measured the IT response pattern elicited by each image. In order to relate the representations between the species and to computational models, we compare response-pattern dissimilarity matrices. IT response patterns form category clusters, which match between man and monkey. The clusters correspond to animate and inanimate objects; within the animate objects, faces and bodies form subclusters. Within each category, IT distinguishes individual exemplars, and the within-category exemplar similarities also match between the species. Our findings suggest that primate IT across species may host a common code, which combines a categorical and a continuous representation of objects.

INTRODUCTION

Do monkeys and humans see the world similarly? Do monkeys categorize objects as humans do? What main distinctions between objects define their cortical representation in each species? The comparison between monkey and human brains is important from an evolutionary perspective. High-level visual object representations are of particular interest in this context, because they are at the interface between perception and cognition and have been extensively studied in each species. Moreover, the monkey brain provides the major model system for understanding primate and, in particular, human brain function. Understanding the species relationship is therefore a challenge central not only to comparative neuroscience but to systems neuroscience in general.

Great progress has been made by comparing monkey and human brains with functional magnetic resonance imaging (fMRI). Previous studies have used classical activation mapping in both species and cortical-surface-based alignment to define a spatial correspondency mapping between the species. Within the visual system, this approach has revealed coarse-scale regional homologies for early visual areas and object-sensitive inferior temporal (IT) cortex (Van Essen et al., 2001; Tsao et al., 2003; Tootell et al., 2003; Denys et al., 2004; Orban et al., 2004; Van Essen and Dierker, 2007).

These studies employed classical activation mapping in which activity patterns elicited by different particular stimuli within the same class (e.g., an object category) are averaged. Moreover, in order to increase statistical power and relate individuals and species, spatial smoothing is typically applied to the data. As a result, this classical approach reveals regions involved in the processing of particular stimulus classes. It does not reveal how those regions represent particular stimuli. In order to address the questions posed above, however, we need to understand how particular real-world object images are represented in fine-grained activity patterns within each region and how their representations are related between the species. Here, we take a first step in that direction by studying IT response patterns elicited by the same 92 object images in monkeys and humans.

One way to relate the IT representations between the species would be to compare the activity patterns on the basis of a spatial correspondency mapping between monkey and human IT. However, this approach is bound to fail at some level of spatial detail even within a species: every individual primate brain is unique by nature and nurture. A neuron-to-neuron functional correspondency cannot exist. (For proof, consider that different individuals have different numbers of neurons.) However, even if a fine-grained representation is unique in each individual, like a fingerprint, the region containing the representation may be homologous, like a finger—serving the same function in both species. For example, the region may serve the function of

representing a particular kind of object information. In this study, relating the species is additionally complicated by the fact that activity was measured with single-cell recording in the monkeys and fMRI in the humans. For these reasons, we do not attempt to define a spatial correspondency mapping between monkey and human IT. Instead, we compare each response pattern elicited by a stimulus to each other response pattern in the same individual animal, so as to obtain a "representational dissimilarity matrix" (RDM) for each species. An RDM shows which distinctions between stimuli are emphasized and which are deemphasized in the representation, thus encapsulating, in an intuitive sense, the information content of the representation. Since RDMs are indexed horizontally and vertically by the stimuli, they can be directly compared between the species.

Our approach has the following key features. (1) The same particular images of real-world objects are presented to both species while measuring brain activity in IT (with electrode recording in monkeys and high-resolution fMRI in humans). (2) Stimuli are presented in random sequences; neither the experimental design nor the analysis is biased by any predefined grouping. (3) Each stimulus is treated as a separate condition, for which a response pattern is estimated without spatial smoothing or averaging (Kriegeskorte et al., 2007; Eger et al., 2008; Kay et al., 2008). (4) The analysis targets the information in distributed response patterns (Haxby et al., 2001; Cox and Savoy, 2003; Carlson et al., 2003; Kamitani and Tong, 2005; Kriegeskorte et al., 2006). (5) In order to compare the IT representations between the species and to computational models, we use the method of "representational similarity analysis" (Kriegeskorte et al., 2008), in which RDMs are visualized and quantitatively compared.

Population representations of the same stimuli have not previously been compared between monkey and human. However, our approach is deeply rooted in the similarity analyses of mathematical psychology (Shepard and Chipman, 1970). An introduction is provided by Edelman (1998), who pioneered the application of similarity analysis to fMRI activity patterns (Edelman et al., 1998) using the technique of multidimensional scaling (Torgerson, 1958; Kruskal and Wish, 1978; Shepard, 1980). Several studies have applied similarity analyses to brain activity patterns and computational models (Laakso and Cottrell, 2000; Op de Beeck et al., 2001; Haxby et al., 2001; Hanson et al., 2004; Kayaert et al., 2005; O'Toole et al., 2005; Aguirre, 2007; Lehky and Sereno, 2007; Kiani et al., 2007; Kay et al., 2008).

Beyond the species comparison, our approach allows us to address the question of categoricality. IT is thought to contain a population code of features for the representation of natural images of objects (e.g., Desimone et al., 1984; Tanaka, 1996; Grill-Spector et al., 2001; Haxby et al., 2001). Does IT simply represent the visual appearance of objects? Or are the IT features designed to distinguish categories defined independent of the visual appearance of their members?

Whether IT is optimized for the discrimination of object categories is unresolved. Human neuroimaging has investigated category-average responses for predefined conventional object categories (Puce et al., 1995; Martin et al., 1996; Kanwisher et al., 1997; Aguirre et al., 1998; Epstein and Kanwisher, 1998; Haxby et al., 2001; Downing et al., 2001; Cox and Savoy, 2003; Carlson et al., 2003; Downing et al., 2006, but see Edelman et al., 1998). This approach requires the assumption of a particular category structure and therefore cannot address whether the representation is inherently categorical. Monkey studies have reported IT responses that are correlated with categories (Vogels, 1999; Sigala and Logothetis, 2002; Baker et al., 2002; Tsao et al., 2003; Freedman et al., 2003; Kiani et al., 2005; Hung et al., 2005; Tsao et al., 2006; Afraz et al., 2006). However, more clearly categorical responses have been found in other regions (Kreiman et al., 2000; Freedman et al., 2001; Quiroga et al., 2005; Freedman and Assad, 2006), suggesting that IT has a lesser role in categorization (Freedman et al., 2003). A brief summary of the previous evidence on IT categoricality is given in the Supplemental Data.

Kiani et al. (2007) investigated monkey-IT response patterns elicited by over 1000 images of real-world objects to address whether IT is inherently categorical. The present study uses the same monkey data and a subset of the stimuli to compare the species. Cluster analysis of the monkey data revealed a detailed hierarchy of natural categories inherent to the monkey-IT representation. Will human-IT show a similar categorical structure? Our approach allows us to address the question of categoricality without the bias of predefined categories. Independent of the result, this provides a crucial piece of evidence for current theory. The question of the inherent category structure of IT is of particular interest with respect to the species comparison, because the prevalent categorical distinctions might be expected to differ between species.

Our goal is to investigate to what extent monkey and human-IT represent the same object information. In particular, we ask the following. (1) Do human-IT response patterns form category clusters as reported for monkey IT (Kiani et al., 2007)? If so, what is the categorical structure and does it match between species? (2) Is within-category exemplar information present in IT? If so, is this continuous information consistent between the species? (3) How is the representation of the objects transformed between early visual cortex and IT? (4) What computational models can account for the IT representation?

RESULTS

We presented the same 92 images of isolated real-world objects (Figure S1) to monkeys and humans while measuring IT response patterns with single-cell recording and high-resolution fMRI, respectively. Two monkeys were presented with the 92 images in rapid succession (stimulus duration, 105 ms; interstimulus interval, 0 ms) as part of a larger set while they performed a fixation task. Neuronal activity was recorded extracellularly with tungsten electrodes, one cell at a time. The cells were located in anterior IT cortex, in the right hemisphere in monkey 1 and in the left in monkey 2. The analyses are based on all cells that could be isolated and for which sufficient data were available across the stimuli. This yielded a total of 674 neurons for both monkeys combined. For each stimulus, each neuron's response amplitude was estimated as the average spike rate within a 140 ms window starting 71 ms after stimulus onset (for details on this experiment, see Kiani et al., 2007).

Four humans were presented with the same images (stimulus duration, 300 ms; interstimulus interval, 3700 ms) while they

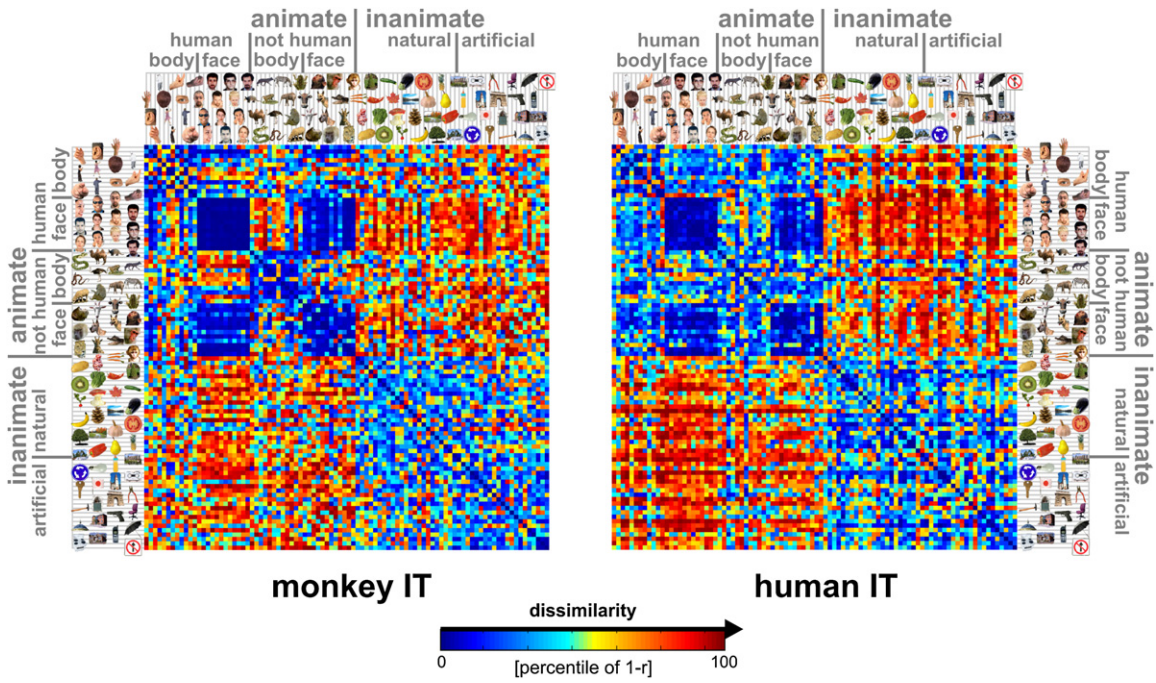


Figure 1. Representational Dissimilarity Matrices for Monkey and Human IT

For each pair of stimuli, each RDM (monkey, human) color codes the dissimilarity of the two response patterns elicited by the stimuli in IT. The dissimilarity measure is 1 − r (Pearson correlation across space). The color code reflects percentiles (see color bar) computed separately for each RDM (for 1 − r values and their histograms, see Figure 3A). The two RDMs are the product of completely separate experiments and analysis pipelines (data not selected to match). Human data are from 316 bilateral inferior temporal voxels (1.95 × 1.95 × 2 mm^3) with the greatest visual-object response in an independent data set. For control analyses using different definitions of the IT region of interest (size, laterality, exclusion of category-sensitive regions), see Figures S9–S11. RDMs were averaged across two sessions for each of four subjects. Monkey data are from 674 IT single cells isolated in two monkeys (left IT in one monkey, right in the other; Kiani et al., 2007).

performed a fixation task in a rapid event-related fMRI experiment. Each stimulus was presented once in each run in random order and repeated across runs within a given session. The amplitudes of the overlapping single-image responses were estimated by fitting a linear model. The task required discrimination of fixation-cross color changes occurring during image presentation. We measured brain activity with high-resolution blood-oxygen-level-dependent fMRI (3-Tesla, voxels: 1.95 × 1.95 × 2 mm^3, SENSE acquisition; Prüssmann, 2004; Kriegeskorte and Bandettini, 2007; Bodurka et al., 2007) within a 5 cm thick slab including all of inferior temporal and early visual cortex bilaterally. Voxels within an anatomically defined IT-cortex mask were selected according to their visual responsiveness to the images in an independent set of experimental runs.

Representational Dissimilarity Matrices: The Same Categorical Structure May Be Inherent to IT in Both Species

What stimulus distinctions are emphasized by IT in each species? Figure 1 shows the RDMs for monkey and human IT. Each cell of a given RDM compares the response patterns elicited by two stimuli. The dissimilarity between two response patterns is measured by correlation distance, i.e., 1 − r (Pearson correlation), where the correlation is computed across the population of neurons or voxels (Haxby et al., 2001; Kiani et al., 2007). An RDM is symmetric about a diagonal of zeros here, because we use a single set of response-pattern estimates.

The RDMs allow us to compare the representations between the species, although there may not be a precise correspondency of the representational features between monkey IT and human IT and although we used radically different measurement modalities (single-cell recordings and fMRI) in the two species. Our approach of representational similarity analysis requires comparisons only between response patterns within the same individual animal, obviating the need for a monkey-to-human correspondency mapping within IT.

Several important results (to be quantified in subsequent analyses) are apparent by visual inspection of the RDMs (Figure 1). First, there is a striking match between the RDMs of monkey and human IT. Two stimuli tend to be dissimilar in the human-IT representation to the extent that they are dissimilar in the monkey-IT representation, and vice versa. This is unexpected because the behaviorally relevant stimulus distinctions might be very different between the species. Moreover, single-cell recording and fMRI sample brain activity in fundamentally different ways, and it is not well understood to what extent they similarly reflect distributed representations. Second, the dissimilarity tends to be large when one of the depicted objects is animate and the other inanimate and smaller when the objects are either both animate or both inanimate. Third, dissimilarities

are particularly small between faces (including human and animal faces). These observations suggest that the IT representation reflects conventional category boundaries in the same way in both species and that there may be a hierarchical structure inherent to the representation. The categorical structure of the matching dissimilarity patterns raises the question, whether the fine-scale patterns of dissimilarities within the categories also match between the species. Alternatively, the categorical structure may fully account for the apparent match. These questions cannot be decided by visual inspection and are addressed by quantitative analysis of the RDMs in the subsequent figures.

It is important to note that the two RDMs (Figure 1) which form the basis of the subsequent interspecies analyses (Figure 2–4) are the product of completely independent experiments and analysis pipelines. In particular, voxels and cells were not selected to maximize the match in any way.

The human RDMs are averages of RDMs computed separately for each of two fMRI sessions in each of four subjects. Note that averaging RDMs for corresponding functional regions is a useful way of combining the data across subjects. As for the species comparison, a precise intraregional spatial correspondency mapping between human subjects is not required. In total, 7 hr, 24 min, and 16 s of fMRI data were used for these analyses. For Figures 1–6, 316 inferior temporal voxels (1.95 × 1.95 × 2 mm^3) with the greatest visual response were selected bilaterally in each subject and session.

Multidimensional Scaling: Category Determines Global Grouping when Stimuli Are Arranged by Representational Similarity

Figure 2A shows unsupervised arrangements of the stimuli reflecting the response-pattern similarity for monkey and human IT (multidimensional scaling, criterion: metric stress; Torgerson, 1958; Shepard, 1980; Edelman et al., 1998). In each arrangement, stimuli placed close together elicited similar response patterns; stimuli placed far apart elicited dissimilar response patterns. The arrangement is unsupervised in that it does not presuppose a categorical structure. For ease of visual comparison, the two arrangements have been scaled to equal size (matching the areas of their convex hulls) and rigidly aligned (Procrustes alignment). Stimulus arrangements computed by multidimensional scaling are data driven and serve an important exploratory function: they can reveal the properties that dominate the representation of our stimuli in the population code without any prior hypotheses. For IT in both species, the global grouping reflects the categorical distinctions between animates and inanimates, and between faces and bodies among the animates. This suggests that category is the dominant factor determining the IT response pattern in both species: if any other stimulus property were more important, it would dominate the stimulus arrangement.

Note also that neighboring stimuli within a category often differ markedly in both shape and color. The arrangements are very similar between the species. Both are characterized by a clean separation of animate and inanimate objects. Furthermore, body parts and faces occupy separate regions among the animate objects.

Note that faces appear to form a particularly tight cluster in the IT response-pattern space of both species. For the human face images this could reflect their similarity in shape and color. However, the face cluster also includes animal faces. Although human and animal faces may be somewhat separated within the face cluster (see also Figure 4), very visually dissimilar animal faces appear to group together. These and other hypotheses inspired by exploring the stimulus arrangement will need to be tested in separate experiments.

Interspecies Dissimilarity Correlation (1): Most Single Stimuli Are Consistently Represented in Both Species

Inspecting the stimulus arrangements for monkey and human separately reveals their overall similarity. However, from the arrangements alone it is not easy to see to what extent particular stimuli within a category appear in different "neighborhoods" of the representational space in the two species. The "fiber-flow" visualization of Figure 2B reproduces both stimulus arrangements and relates them by "fibers" linking dots that represent the same stimulus. This makes it easier to see how stimuli within the same category match up between species. Most fibers flow in a roughly straight line (i.e., without much displacement) from the monkey to the human representation. This reflects the within-category match of the representations, which is analyzed and tested for significance in Figure 3.

In order to reveal the species differences, we chose the thickness of the fibers in Figure 2B to reflect the extent to which each stimulus is inconsistently represented in monkey and human IT. For each stimulus *i*, its place in the high-dimensional monkey-IT response-pattern space is characterized by the vector m_i of its dissimilarities to the other 91 stimuli. Its place in the human-IT representation is characterized analogously by dissimilarity vector h_i. The interspecies correlation r_i (Pearson) between m_i and h_i reflects the consistency of placement of the stimulus in the representations of both species (see Figure S2 for details). For each stimulus *i*, the thickness of its fiber in Figure 2B is proportional to $(1 - r_i)^2$, thus emphasizing the most inconsistently represented stimuli. The prevalence of thin fibers (which tend to be straight) reflects the overall interspecies consistency.

The single-stimulus interspecies dissimilarity correlations r_i are further visualized and statistically analyzed in Figures S2 and S3. Results show significant interspecies consistency for about two-thirds of the single stimuli ($p < 0.05$, corrected for multiple tests). Furthermore human faces exhibit significantly higher interspecies correlations than the stimulus set as a whole, and several stimuli (including images of animate and inanimate objects) exhibit significantly lower interspecies correlations. The two stimuli with the lowest interspecies correlations (eggplant, back-view of human head) were the only two stimuli described as ambiguous by human subjects during debriefing (Figure S1). This is consistent with the idea that the IT representation reflects not only the visual appearance, but also the conceptual interpretation of a stimulus.

Interspecies Dissimilarity Correlation (2): IT Emphasizes the Same Stimulus Distinctions in Both Species within and between Categories

The RDMs of Figure 1 suggest similar representations in monkey and human IT. Figure 3A quantifies this impression. The scatter plot of the monkey-IT dissimilarities (horizontal axis) and the

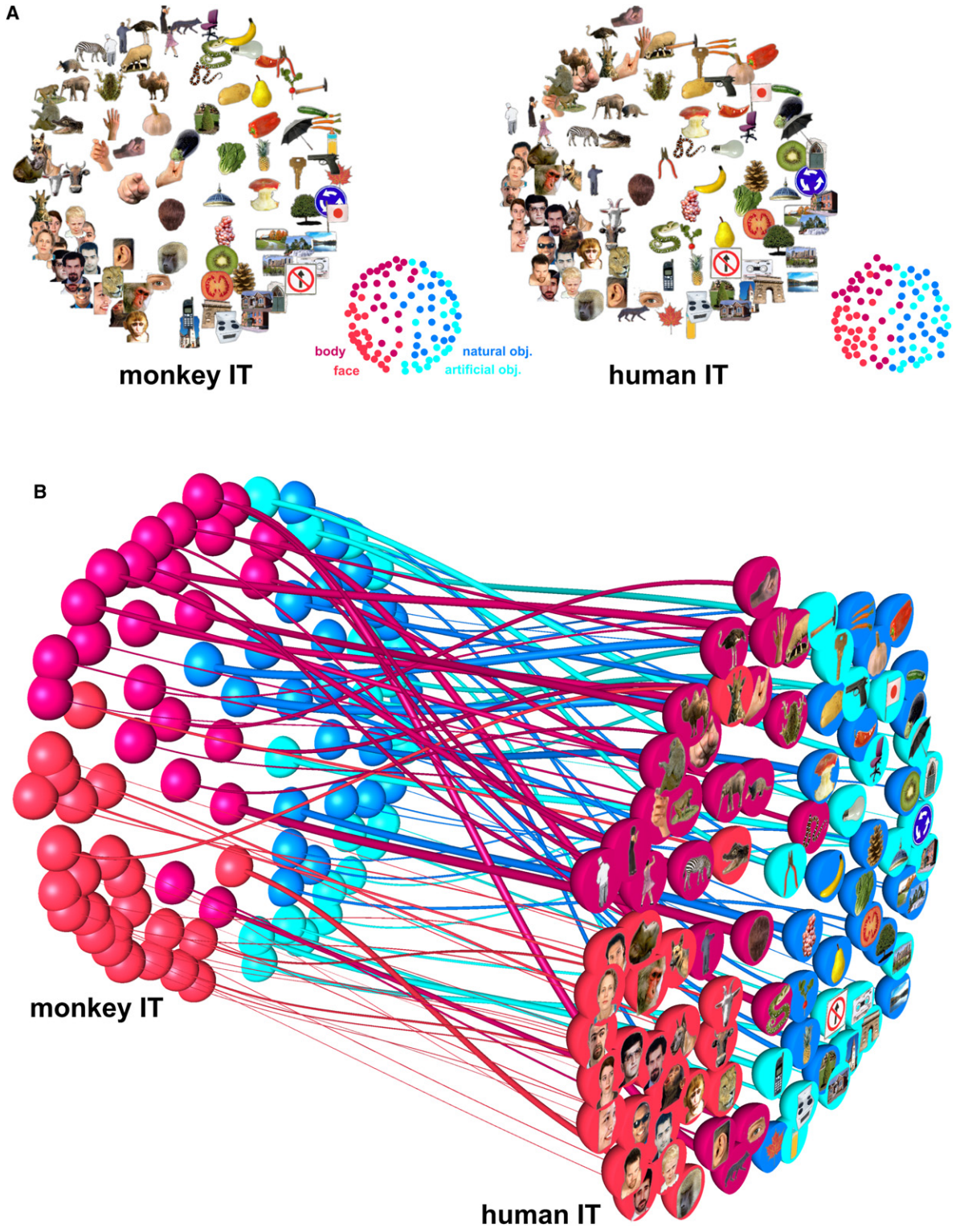


Figure 2. Stimulus Arrangements Reflecting IT Response-Pattern Similarity in Monkey and Human and the Interspecies Relationship

(A) The experimental stimuli have been arranged such that their pairwise distances approximately reflect response-pattern similarity (multidimensional scaling, dissimilarity: 1 − Pearson r, criterion: metric stress). In each arrangement, images placed close together elicited similar response patterns. Images placed far apart elicited dissimilar response patterns. The arrangement is unsupervised: it does not presuppose any categorical structure. The two arrangements have been scaled to match the areas of their convex hulls and rigidly aligned for easier comparison (Procrustes alignment). The correlations between the

human-IT dissimilarities (vertical axis) across pairs of stimuli reveals a substantial correlation ($r = 0.49$, $p < 0.0001$ estimated by means of 10,000 randomizations of the stimulus labels).

Does the matching categorical structure fully explain the interspecies correlation of dissimilarities? Figure 3A (colored subsets) shows that the correlation is substantial also within animates ($r = 0.51$, $p < 0.0001$) and, to a lesser extent, within inanimates ($r = 0.20$, $p < 0.0001$), as well as across pairs of stimuli crossing the animate-inanimate boundary ($r = 0.19$, $p < 0.0001$). The monkey-to-human correlation is also present (Figure 3B) within images of humans ($r = 0.66$, $p < 0.0001$), within images of nonhuman animals ($r = 0.35$, $p < 0.0001$), within images of faces (including human and animal faces, $r = 0.31$, $p < 0.0058$), within images of bodies (including human and animal bodies, $r = 0.31$, $p < 0.0001$; not shown in Figure 3), within images of human bodies ($r = 0.53$, $p < 0.0001$; not shown in Figure 3), within images of natural objects ($r = 0.19$, $p < 0.0039$), and within images of artificial objects ($r = 0.23$, $p < 0.0139$). These within-category dissimilarity correlations between the species indicate that the continuous variation of response patterns within each category cluster is not noise, but distinguishes exemplars within each category in a way that is consistent between monkey and human.

We did not find a significant monkey-to-human dissimilarity correlation either within images of human faces ($r = -0.05$, $p < 0.605$) or within images of animal faces ($r = 0.12$, $p < 0.234$; Figure 3B, bottom left). We also did not find a significant correlation within images of nonhuman bodies ($r = 0.12$, $p < 0.21$; not shown in Figure 3).

The negative findings all occurred for small subsets of images (12 images, 66 pairs) for which we have reduced power. Note, however, that the effect sizes (r) were also smaller for the insignificant correlations than for all significant correlations. That the correlation is significant within human bodies, but not within nonhuman bodies, could reflect the fact that the human-body images included whole bodies as well as body parts, whereas the nonhuman body images were all of whole bodies and 9 of the 12 images were of four-limbed animals (Figure S1), which may constitute a separate subordinate category (Kiani et al., 2007).

Regarding the absence of a significant interspecies dissimilarity correlation within human faces and within nonhuman faces, one interpretation of particular interest is that human and monkey differ in how they represent individual human faces as well as individual nonhuman faces. For example, within each species the representation of images of its own members may have a special status.

Species-Specific Face Analysis: IT Might Better Distinguish Conspecific Faces in Each Species

We observed greater dissimilarities in the human representation of human faces than in the monkey representation of human faces (Figure S8). Figure S4 explores the possibility of a species-specific face representation. We selectively analyzed the representation of monkey, ape, and human faces in monkey and human IT. The dissimilarities among human faces are significantly larger in human IT than in monkey IT ($p = 0.009$). The dissimilarities among monkey-and-ape faces are larger in monkey IT than in human IT in our data, but the effect is not significant ($p = 0.12$). The difference between the two effects is significant ($p = 0.02$). This analysis provides an interesting lead for future studies designed to address species-specific face representation (for details, see Figure S4).

Hierarchical Clustering: A Nested Categorical Structure Matching between Species Is Inherent to IT

The stimulus arrangements of Figure 2 suggest that the categories correspond to contiguous regions in IT response-pattern space. However, it is not apparent from Figure 2 whether the response patterns form clusters corresponding to the categories. Contiguous category regions in response-pattern space could exist for a unimodal response-pattern distribution. Category clusters would imply separate modes in response-pattern space, separated by boundaries of lower probability density. We therefore ask whether the category boundaries can be determined without knowledge of the category labels.

Figure 4 shows hierarchical cluster trees computed for the IT response patterns in monkey and human. Unlike the unsupervised stimulus arrangements (Figure 2), hierarchical cluster analysis (Johnson, 1967) assumes the existence of some categorical structure, but it does not assume any particular grouping into categories.

We find very similar cluster trees for both species. The top-level distinction is that between animate and inanimate objects. Faces and body parts form subclusters within the animate objects. Note that the clustering conforms closely, though not perfectly, to the these human-conventional categories. The deviating placements could be a consequence of inaccurate response-pattern estimation: because of the large number of conditions (92) in these experiments, our response-pattern estimates are noisier than they would be for a small number of conditions based on the same amount of data.

Results Similar between Hemispheres and Robust to Exclusion of Category-Sensitive Regions and to Varying the Number of Voxels

The results we describe for bilateral IT are similar when IT is restricted to either cortical hemisphere (for details, see Figure S9). Results are also robust to changes of the number of voxels selected. The categorical structure is present already at 100 voxels and decays only when thousands of voxels are selected (for details, see Figure S10). Finally, the categorical

high-dimensional response-pattern dissimilarities ($1 - r$) and the two-dimensional Euclidean distances in the figure are 0.67 (Pearson) and 0.69 (Spearman) for monkey IT and 0.78 (Pearson) and 0.78 (Spearman) for human IT.

(B) Fiber-flow visualization emphasizing the interspecies differences. This visualization combines all the information from (A) and links each pair of dots representing a stimulus in monkey and human IT by a "fiber." The thickness of each fiber reflects to what extent the corresponding stimulus is inconsistently represented in monkey and human IT. The interspecies consistency r_i of stimulus i is defined as the Pearson correlation between vectors of its 91 dissimilarities to the other stimuli in monkey and human IT. The thickness of the fiber for stimulus i is proportional to $(1 - r_i)^2$, thus emphasizing the most inconsistently represented stimuli. The analysis of single-stimulus interspecies consistency is pursued further in Figures S2 and S3.

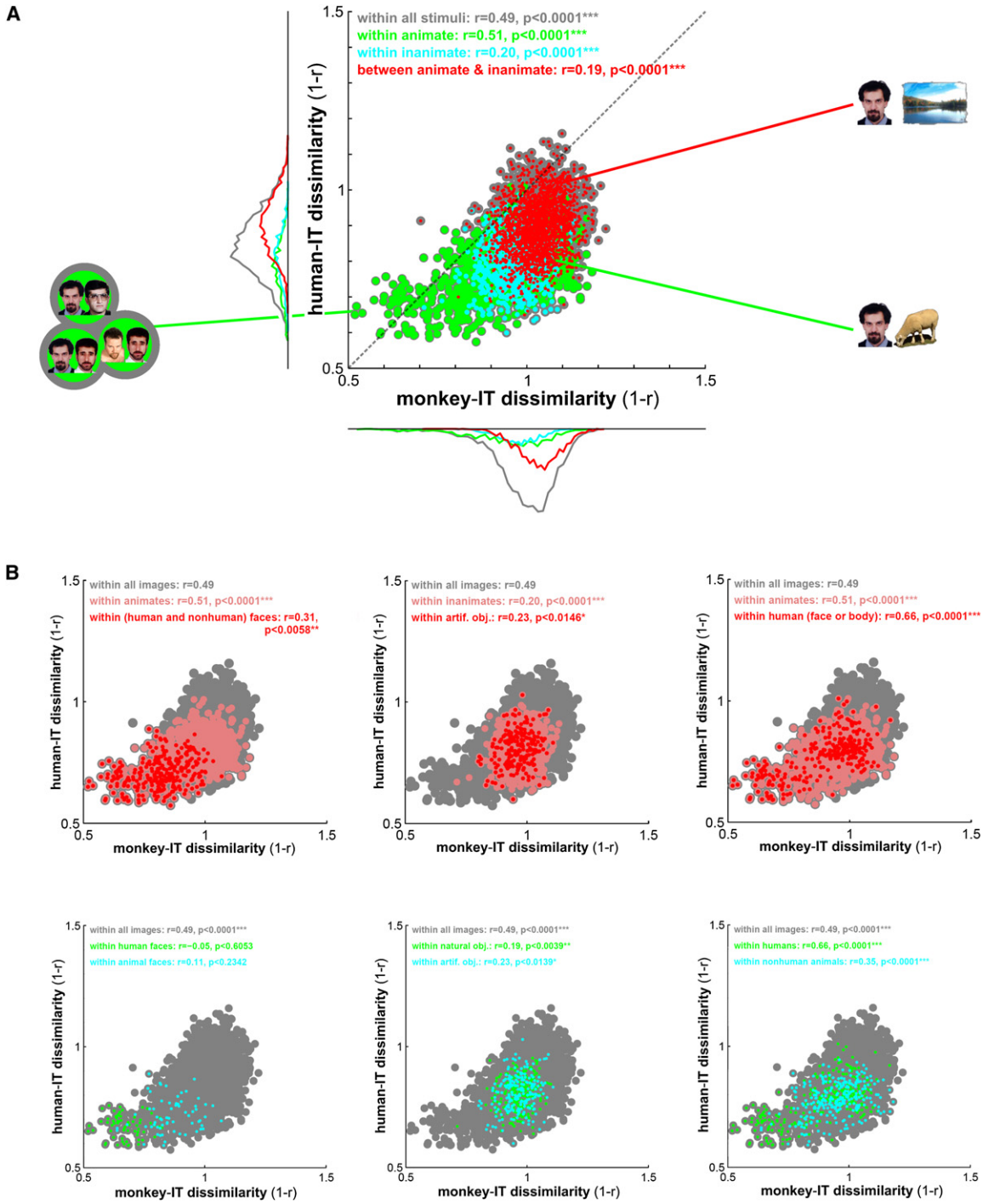


Figure 3. Correlation of Representational Dissimilarities between Monkey and Human IT

(A) For each pair of stimuli, a dot is placed according to the IT response-pattern dissimilarity in monkey (horizontal axis) and human (vertical axis). As before, the dissimilarity between the two response patterns elicited by each stimulus pair is measured as 1 − r (Pearson correlation). Dot colors correspond to all pairs of stimuli (gray), pairs within the animate objects (green), pairs within the inanimate objects (cyan), and pairs crossing the animate-inanimate boundary (red). Marginal histograms of dissimilarities are shown for the three subsets of pairs using the same color code. For detailed exploratory analysis of the species differences, Figures S13 and S14 show the stimulus pairs corresponding to the dots for the three apical regions of the scatter plot.

(B) The same analysis as in (A), but for within-category correlations between human and monkey-IT object dissimilarities. Colored dots correspond to all pairs of stimuli (gray) or pairs within stimulus-category subsets (colors). In the top row, each panel shows the whole set (gray), a subset (pink), and a subset nested within that subset (red), as indicated in the colored legend of each panel. In the bottom row, each panel shows the whole set (gray) and two disjointed subsets (green and cyan), as indicated in the colored legend of each panel.

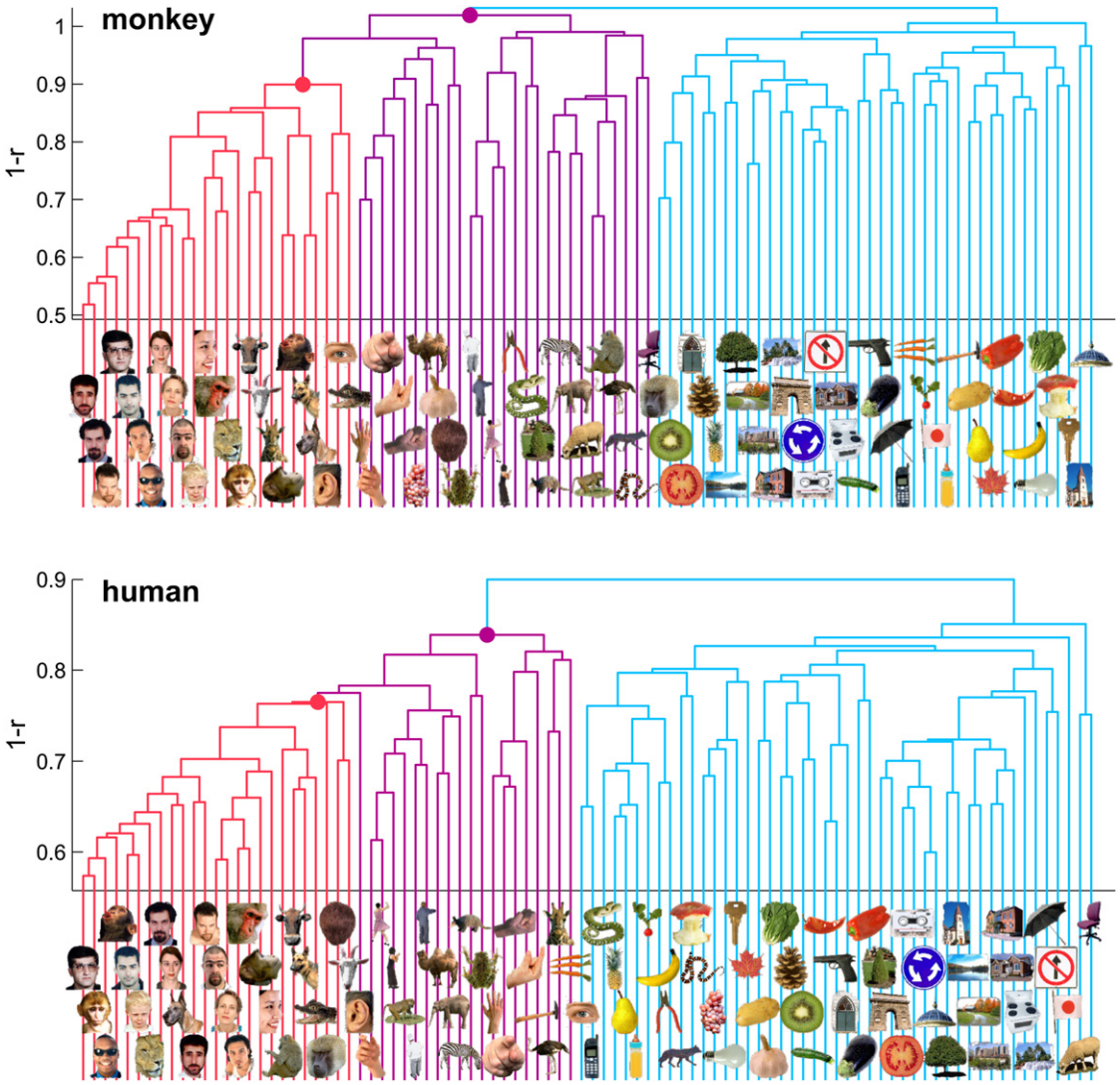


Figure 4. Hierarchical Clustering of IT Response Patterns

In order to assess whether IT response patterns form clusters corresponding to natural categories, we performed hierarchical cluster analysis for human (top) and monkey (bottom). This analysis proceeds from single-image clusters (bottom of each panel) and successively combines the two clusters closest to each other in terms of the average response-pattern dissimilarity, so as to form a hierarchy of clusters (tree structure in each panel). The vertical height of each horizontal link indicates the average response-pattern dissimilarity (the clustering criterion) between the stimuli of the two linked subclusters (dissimilarity: $1 - r$). The cluster trees for monkey and human are the result of completely independent experiments and analysis pipelines. This data-driven technique reveals natural-category clusters that are consistent between monkey and human. For easier comparison, we colored subcluster trees (faces, red; bodies, magenta; inanimate objects, light blue). Early visual cortex (Figures 5, 6, and S5) and low-level computational models (Figures S6 and S7) did not reveal such category clusters.

structure appears unaffected when the fusiform face area (FFA) (Kanwisher et al., 1997) and the parahippocampal place area (PPA) (Epstein and Kanwisher, 1998) are bilaterally excluded from the selected voxel set (for details, see Figure S11). After exclusion of FFA and PPA, the region of interest has most of its voxels in the lateral occipital complex, but also includes more anterior object-sensitive voxels within IT.

Category-Boundary Effect: Weak in Early Visual Cortex and Strong in IT

The human fMRI data allowed us to compare the representations between early visual cortex and IT (Figures 5, 6, and S5). Visual inspection of the RDMs (Figure 5) suggests a categorical representation in IT, but not in early visual cortex. The multidimensional scaling arrangements and hierarchical cluster trees also

In both (A) and (B), each panel's color legend (top inset) also states the correlations (r, Pearson) between monkey and human-IT dissimilarities and their significance (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). The dissimilarity correlations are tested by randomization of the stimulus labels. This test correctly handles the dependency structure within each RDM. All p values < 0.0001 are stated as $p < 0.0001$ because the randomization test terminates after 10,000 iterations.

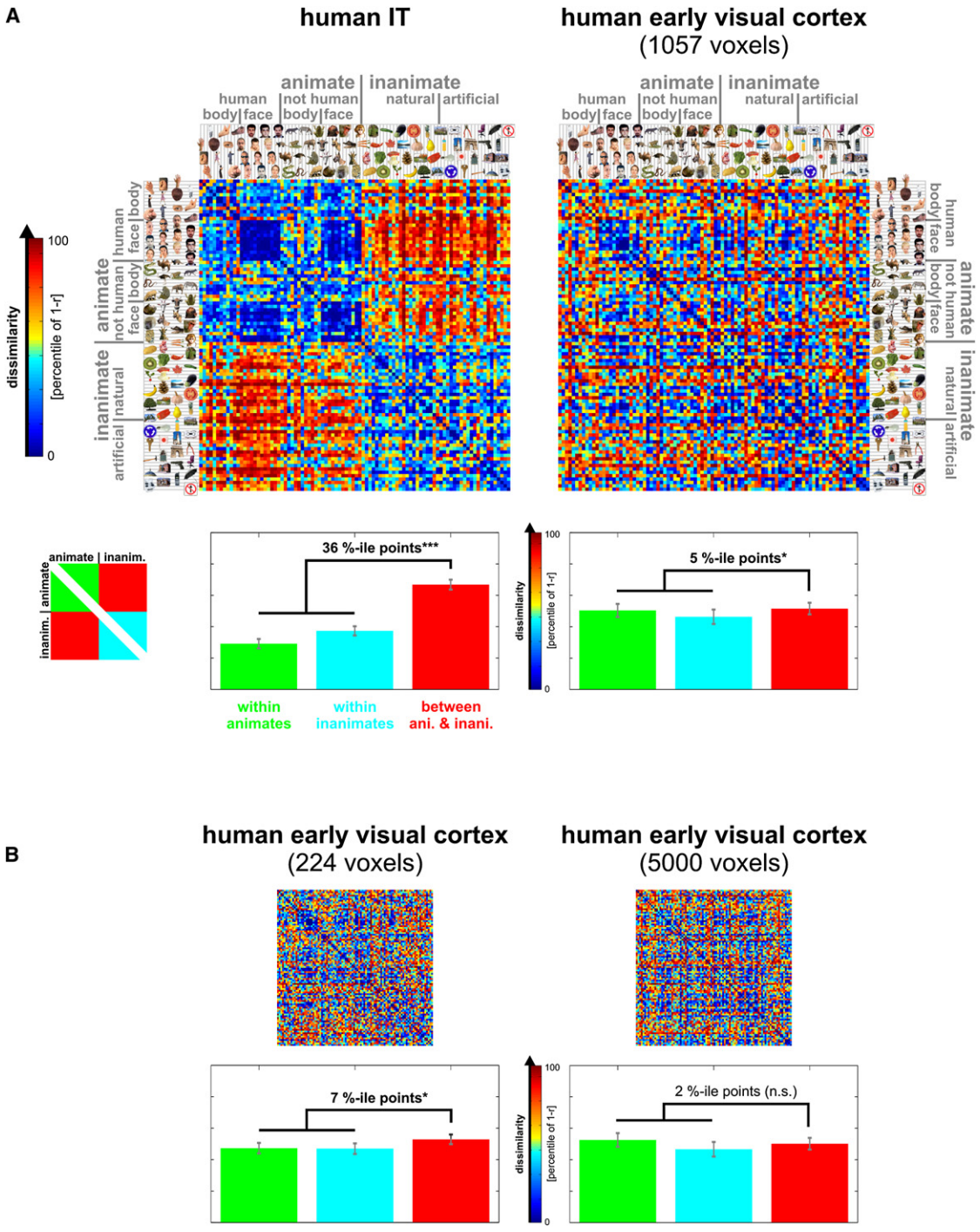


Figure 5. Early Visual Cortex and IT in the Human: Representational Dissimilarity Matrices and Category-Boundary Effects

(A) RDMs for human IT (top left, same as in Figure 1) and human early visual cortex (top right). As in Figure 1, the color code reflects percentiles (see color bar) computed separately for each RDM (for $1 - r$ values and their histograms, see Figure 6). The bar graph below each RDM shows the average dissimilarity (percentile of $1 - r$) within the animates (green bars), within the inanimates (cyan bars), and for pairs crossing the category boundary (red bars). Error bars indicate the standard error of the mean estimated by bootstrap resampling of the stimulus set. We define the category-boundary effect as the difference (in percentile points) between the mean dissimilarity for between-animate-and-inanimate pairs and the mean dissimilarity for within-animate and within-inanimate pairs. The zeros on the diagonal are excluded in computing these means. The category-boundary effect sizes are given above the bars in each panel with significant effects marked by stars ($p \geq 0.05$ indicated by n.s. for "not significant," *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). The p values are from a bootstrap test; a randomization test yields the same pattern of significant effects (see Results). Here, as in Figures 1–4, human IT has been defined at 316 voxels (for IT at 100–10,000 voxels, see Figure S10) and human early visual cortex at 1057 voxels.

(B) The same analyses for smaller and larger definitions of human early visual cortex (224 and 5000 voxels, respectively).

do not support that early visual cortex contains an inherently categorical representation (Figure S5). Note, however, that a subset of the human faces appears to be associated with somewhat lower dissimilarities in early visual cortex (Figure 5). This could be caused by similarities in shape and color among these stimuli.

Although the early visual representation does not exhibit a categorical structure as observed for IT, the top-level animate-inanimate distinction might be reflected in the early visual responses in more subtle ways. To test this possibility for the top-level animate-inanimate distinction, we analyze the category-boundary effect, which we define as the difference between the mean dissimilarity for between-category pairs (i.e., one is animate, the other inanimate) and the mean dissimilarity for within category pairs (i.e., both are animate or both are inanimate). As in Figure 1, the dissimilarities are correlation distances between spatial response patterns, converted to percentiles for each RDM separately (for histograms of the original correlation distances, see Figure 6). The analysis indicates that the category-boundary effect is strong in IT and weak, but present, in early visual cortex. The category-boundary effect estimates in percentile points are 36% ($p < 0.001$) for IT (defined at 316 voxels as before) and 7% ($0.01 < p < 0.05$), 5% ($0.01 < p < 0.05$), and 2% (not significant) for early visual cortex, defined at 224, 1057, and 5000 voxels, respectively. We further investigated the category distinction by a linear decoding analysis (Figure S12), which suggested linear separability of animates and inanimates in IT, but not early visual cortex.

The p values for the category-boundary effect were computed by bootstrap resampling of the stimulus set, thus simulating the distributions of mean dissimilarities expected if the experiment were to be repeated with different stimuli from the same categories and with the same subjects. We also tested the category-boundary effect by a randomization test in which the stimulus labels are randomly permuted, thus simulating the null hypothesis of no difference between the response patterns elicited by the stimuli, but not generalizing to different stimuli from the same categories. This test yields the same pattern of significant results as the bootstrap test ($p < 0.001$ for IT and early visual cortex defined at 224 or 1057 voxels, $p \geq 0.05$ for human early visual cortex defined at 5000 voxels). In addition, both bootstrap and randomization tests show a significantly larger category-boundary effect in IT than in early visual cortex ($p < 0.0001$ for each of the three sizes of the early visual region of interest).

Representational Connectivity Analysis: Early Visual Cortex and IT Share Visual-Similarity Information

We have compared monkey and human IT by correlating representational dissimilarities across pairs of stimuli. The same approach can serve to characterize the relationship between the representations in two brain regions of a given species. In analogy to the concept of functional connectivity, we refer to the correlation of representational dissimilarities between two brain regions as their "representational connectivity." Like functional connectivity, representational connectivity does not imply an anatomical connection or a directed influence. Unlike functional connectivity, representational connectivity is a multivariate, nonlinear, and design-dependent connectivity measure.

Visual comparison of the RDMs for early visual cortex and IT does not suggest a strong correlation, because the categorical structure dominating IT appears absent in early visual cortex.

However, the representational-connectivity scatter plot (Figure 6) reveals a substantial correlation of the dissimilarities ($r = 0.38$, $p < 0.0001$). Pairs of stimuli eliciting more dissimilar response patterns in early visual cortex also tend to elicit more dissimilar response patterns in IT. The analysis for between- and within-category subsets of pairs reveals what drives this effect (see diagram in Figure 6B). First, the between-category distribution (red) is shifted relative to the within-category distributions, but only along the IT axis, not along the early-visual axis. This is evident in the marginal dissimilarity histograms framing the scatter plot (Figure 6) and reflects the category-boundary effect, which is strong in IT and weak in early visual cortex (Figure 5). Second, in addition to its category-boundary effect, IT reflects the dissimilarity structure of the early visual representation for within- as well as between-category pairs (diagonally elongated distributions, $p < 0.0001$ for all correlations, tested by randomization of the stimulus labels). These results do not depend on the size of the early-visual region of interest (Figure 6C).

Early visual response patterns are likely to reflect shape similarity in this experiment, because all stimuli were presented at the same retinal location (fovea) and size (2.9° visual angle). Shape similarity as reflected in the early visual representation (see also Kay et al., 2008) may therefore carry over to the IT representation, even if IT is more tolerant to changes of position and size.

Computational Modeling: A Range of Low- and Intermediate-Level Representations Cannot Account for the Categorical Structure Observed in IT

Can low-level feature similarity account for our results? We compared the IT representation to several low-level model representations. The low-level models included the color images themselves (in CIELAB color space), simple processed versions of the images (low-resolution color image, grayscale image, low-resolution grayscale image, spatial low- and high-pass-filtered grayscale image, binary silhouette image), CIELAB joint color histograms, and a computational model of V1 (including simple and complex cells). We also tested an intermediate-level computational model corresponding approximately to the level of V4 and posterior IT, the HMAX-C2 representation based on natural image patches. (For details on these models, see Supplemental Data.) The RDMs, multidimensional scaling arrangements, and hierarchical cluster trees for these models (Figures S6 and S7) suggest that none of them can account for the category clustering we observed in IT cortex.

DISCUSSION

Matching Information in Monkey and Human IT

IT is thought to contain a high-level representation of visual objects at the interface between perception and cognition.

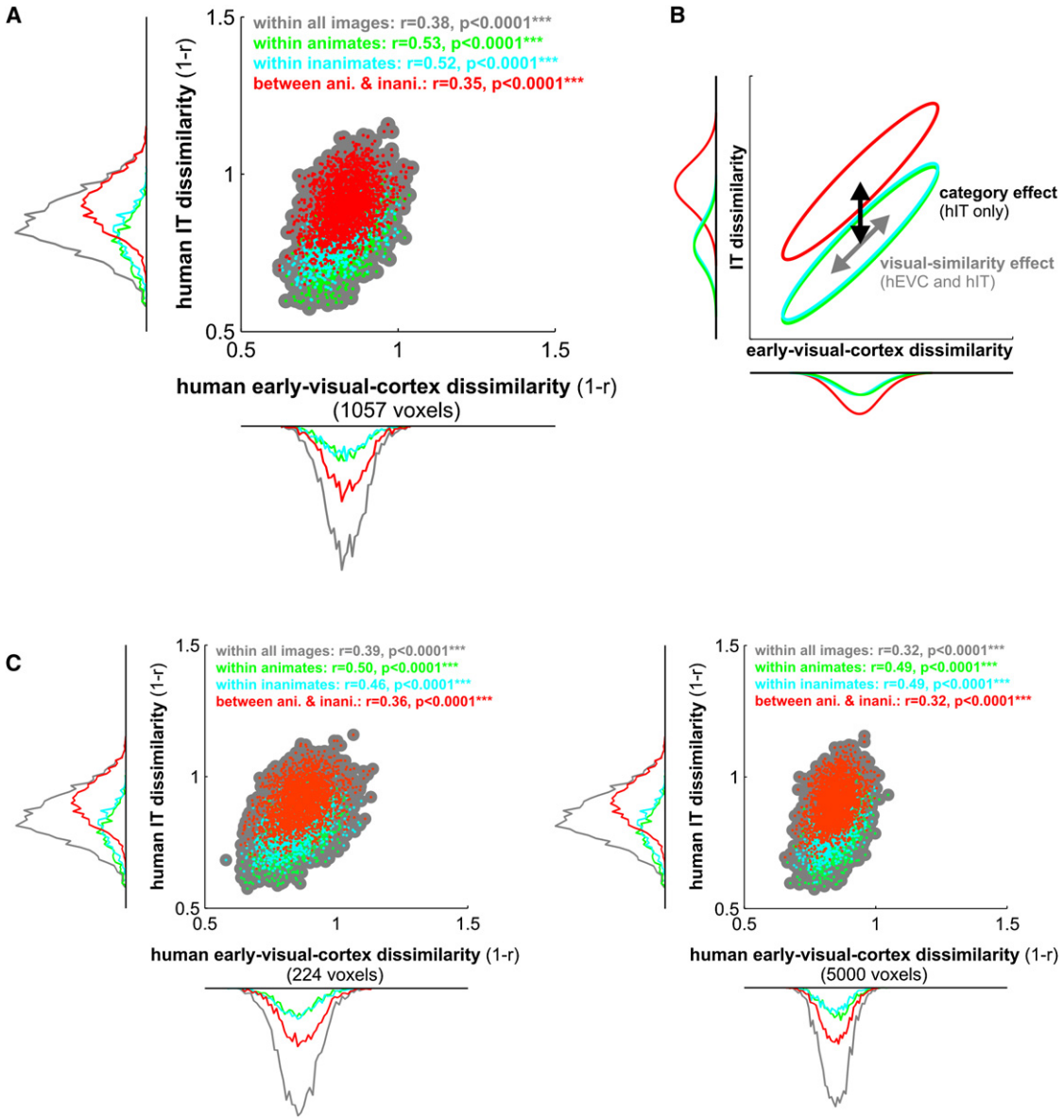


Figure 6. Representational Connectivity between Early Visual Cortex and IT in the Human

(A) For each pair of stimuli, we plot a dot with horizontal position reflecting early visual response-pattern dissimilarity and vertical position reflecting IT response-pattern dissimilarity. Scatter plots and correlation analyses (insets) show that pairs of stimuli eliciting more dissimilar response patterns in early visual cortex also tend to elicit more dissimilar response patterns in IT. This suggests that visual similarity as reflected in the early visual representation carries over into the IT representation. However, IT additionally exhibits a strong category-boundary effect: when a stimulus pair crosses the animate-inanimate boundary (red) the two response patterns tend to be more dissimilar than when both stimuli are from the same category (green, cyan). The category-boundary effect is evident in the marginal dissimilarity histograms framing the scatter plot (for statistical analysis, see Figure 5).

(B) In this conceptual diagram, the distributions from the scatter plots are depicted as ellipsoids (iso-probability-density contours) with the same color code. The visual-similarity effect is shared between early visual and IT representations (each distribution diagonally elongated), whereas the category-boundary effect is only present in IT (red distribution vertically, but not horizontally shifted with respect to the within-category distributions).

(C) The same analyses for smaller and larger definitions of human early visual cortex (224 and 5000 voxels, respectively) show that the findings above do not depend on the size of the early visual region of interest.

Our results show that monkey and human IT emphasize very similar distinctions among objects. To answer the questions posed at the end of the Introduction: (1) IT response patterns elicited by object images appear to cluster according to the same categorical structure in monkey and human. (2) Within each category, primate IT appears to represent more fine-grained object information. This information as well is remarkably consistent across species and may reflect subordinate

categorical distinctions as well as a high-level form of visual similarity (Op de Beeck et al., 2001; Eger et al., 2008). (3) Categoricality appears to arise in IT; it is largely absent in human early visual cortex. (4) A range of low- and intermediate-level computational models did not reproduce the categorical structure observed for human and monkey IT.

A Hierarchical Category Structure Inherent to IT

The categorical structure inherent to IT in both species appears hierarchical: animate and inanimate objects form the two major clusters; faces and bodies form subclusters within the animate cluster. The hierarchy observed for the present set of 92 stimuli has only two levels. However, the previous study by Kiani et al. (2007), using over 1000 stimuli, reported a hierarchy for monkey IT, which is consistent with our findings here, but extends into finer distinctions. This raises the question, whether finer categorical distinctions are also present in the human and, if so, if they match between the species.

Relationship between Category-Sensitive Regions and IT Pattern Information

Human neuropsychology has described category-specific deficits resulting from temporal brain damage and suggested a special status for the living/nonliving distinction (Martin, 2007; Capitani et al., 2003; Humphreys and Forde, 2001; Martin et al., 1996). Our results support the view that this distinction has a special status. We note that fruit and vegetables fall into the inanimate category in the IT cluster structure we observed (Figure 4). Our findings are also consistent with the idea that IT contains specialized features or processing mechanisms for faces and bodies (Puce et al., 1995; Kanwisher et al., 1997; Downing et al., 2001; but see also Gauthier et al., 2000).

Can the previous findings on human-IT regions sensitive to these categories explain the IT response-pattern clustering we report? Let us assume that the FFA responds with a similar overall activation to each individual face (Kriegeskorte et al., 2007). Including the FFA in the region of interest will then render IT response patterns to faces more similar, thus contributing to their clustering. More generally, a sufficiently category-sensitive feature set will exhibit categorical clustering of the response patterns.

However, the existence of the category-sensitive regions does not predict (1) that the category effects will dominate the representation such that response patterns form category clusters that are separable without prior knowledge of the categories (alternatively, the category-sensitive component could be too weak—in relation to the total response-pattern variance—to form clusters), (2) that the cluster structure will be hierarchical, or (3) what categorical distinction is at the top of the hierarchy (explaining most response-pattern variance). Moreover, our human results hardly changed when FFA and PPA were excluded (Figure S11), leaving the lateral occipital complex as the main focus within the IT region of interest. (Voxels were selected by their average response to objects versus fixation using independent data.) In the monkeys, the IT recordings did not target category-sensitive regions (Tsao et al., 2003); nevertheless, the population exhibited a complex categorical clustering of its response patterns (Figure 4), and for each pair of categories, discrimination by the cell population was robust to exclusion of cells responding maximally to either category (Figure 10 of Kiani et al., 2007).

If FFA and PPA can be excluded without a qualitative change to the representational dissimilarity structure (Figure S11), the remaining portion of human IT must have similarly category-sensitive features. One interpretation is that the prominent category regions are just particularly conspicuous concentrations of related features within a larger category-sensitive feature map (Haxby et al., 2001). Such large, consistently localized foci of features may only exist for a few categories (Downing et al., 2006). Their number is limited by the available space in the brain. Beyond discovering those regions, our larger goal should be to understand the representation as a whole, including the contribution of its less prevalent—or perhaps just more scattered—features. After all, IT categoricality is inherently a population phenomenon: Step-function-like categorical responses as reported for cells in the medial temporal lobe (Kreiman et al., 2000) and prefrontal cortex (Freedman et al., 2001) are not typically observed in either single IT cells (Vogels, 1999; Freedman et al., 2003; Kiani et al., 2007; but see Tsao et al., 2006) or category-sensitive fMRI responses (Haxby et al., 2001). Categorical clustering of response patterns indicates that the categorical distinctions explain a lot of variance across the population. It does not imply that any single cell exhibits a step-function-like response.

Explaining the IT Representational Similarity Structure

Can low-level features explain the IT representational similarity structure? The categorical cluster structure observed in IT was absent in the fMRI response patterns in human early visual cortex (Figures 5 and S5) and also in several low-level model representations of the images (luminance pattern, color pattern, color histogram, silhouette pattern, V1 model representation; Figures S6 and S7). The possibility that our findings can be explained by low-level features can never be formally excluded, because the space of models to be tested is infinite. However, our results suggest that the categorical clustering in IT does not reflect only low-level features.

Can more complex natural-image features explain the IT representational similarity structure? Categorical clustering was not evident in the intermediate-complexity HMAX-C2 model based on natural image fragments (Figure S7; Serre et al., 2005). In addition, a high-level representation composed of shape-tuned units adapted to real-world object images in the HMAX framework has previously been shown not to exhibit categorical clustering (Kiani et al., 2007).

Our interpretation of the current evidence is that evolution and development leave primate IT with features optimized not only for representing natural images (as the features of the models described above), but also for discriminating between object categories. This suggests that an IT model should acquire category-discriminating features by supervised learning (Ullman, 2007). A recent study suggests that human IT responds preferentially to such category-discriminating features (Lerner et al., 2008).

Does IT categoricality arise from feedforward or feedback processing? Our tasks (in both species) minimize the top-down

component by withdrawing attention from briefly presented stimuli. Although this does not abolish local recurrent processing, it minimizes feedback from higher regions, suggesting that IT categoricality is not a product of top-down influences. One interpretation is that IT categoricality arises from feedforward connectivity. Rapid feedforward animate-inanimate discrimination would explain reports that humans can perform animal detection at latencies allowing for limited recurrent processing (Thorpe et al., 1996; Kirchner and Thorpe, 2006).

Serre et al. (2007) proposed a feedforward model of rapid categorization (see also Riesenhuber and Poggio, 2002), which summarizes a wealth of neuroscientific findings. Their architecture may be able to account for our findings. However, these authors associate the category-discrimination stage with prefrontal cortex. Our results suggest that features at the stage of IT already are optimized for category discrimination.

Beyond visual features optimized for categorization, could IT represent more complex semantic information? So far we have considered the features a means to the end of categorization. Instead, we could argue, more generally, that the features serve to infer nonvisual properties from the visual input. The features, then, are the end, and category clusters may arise as a consequence of the feature set. It has been suggested, for example, that IT represents action-related properties (Mahon et al., 2007). This perspective relates our findings to the literature on semantic representations (Tyler and Moss, 2001; McClelland and Rogers, 2003; Patterson et al., 2007). In order to test semantic-feature hypotheses along with computational models, we could predict the IT representational similarity from semantic property descriptions of the stimuli.

To find a model that reproduces the empirical representational similarity structure of IT (Figure 1) would constitute a substantial theoretical advance. The reader is invited to join us in testing additional models by exposing them to our stimuli and comparing the RDMs of the model representations to our empirical RDMs from IT. If the models have parameters fitted, so as best to predict the empirical RDMs, independent stimulus sets will be needed for fitting and testing. We will provide both stimuli and RDMs of monkey and human IT upon request.

Representational Similarity Analysis

Studying a brain region's pairwise response-pattern dissimilarities for a sizable set of stimuli reveals what distinctions are emphasized and what distinctions are abstracted from by the representation. Representational similarity analysis allows us to make comparisons between brain regions (Figure 6), between species (Figure 3), between measurement modalities (Figure 3, confounded with the species-effect here), and between biological brains and computational models (Figures S6 and S7). An RDM usefully combines the evidence across the patterns of response within a functional region (thus allowing us to see the forest), but it requires no averaging of activity across space, time, or stimuli (thus honoring the trees). The RDM has a very intricate structure ($(n^2 - n)/2$ dissimilarities, where n is the number of stimuli), thus providing a rich characterization of the representation.

In order to understand a population code, representational similarity analysis must be complemented with a wide range of methods. For example, we need to quantify the pairwise stimulus information, address how the representation can be read out (e.g., is a given distinction explicit in the sense of linear decodability?; Figure S12), how it relates to other brain representations (Figure 6) and behavior, and how the activity patterns are organized in space and time.

Implications for the Relationship between fMRI and Single-Cell Data

A single voxel in blood-oxygen-level-dependent fMRI reflects the activity of tens of thousands of neurons (Logothetis et al., 2001). We therefore expect to find somewhat different stimulus information in hemodynamic and neuronal response patterns. fMRI patterns may contain more information about fine-grained neuronal activity patterns than voxel size would suggest (Kamitani and Tong, 2005). But to what extent neuronal pattern information is reflected in fMRI pattern information is not well understood, because a voxel's signal does not provide us simply with the average activity within its boundaries, but rather reflects the complex spatiotemporal transform of the hemodynamic response. The close match we report here between the RDMs from single-cell recording and fMRI provides some hope that data from these two modalities, for all their differences, may somewhat consistently reveal neuronal representations when subjected to massively multivariate analyses of activity-pattern information (Kriegeskorte and Bandettini, 2007).

A Common Code in Primate IT

Taken together, our results suggest that evolution and individual development leave primate IT with representational features that emphasize behaviorally important categorical distinctions. The major distinctions, animate-inanimate and face-body, are so basic that their conservation across species appears plausible. However, the IT representation is not purely categorical. Within category clusters, object exemplars are represented in a continuous object space, which may reflect a form of visual similarity. The categorical and continuous aspects of the representation are both consistent between man and monkey, suggesting that a code common across species may characterize primate IT.

EXPERIMENTAL PROCEDURES

This section describes the experimental designs and brain-activity measurements in monkey and human. The monkey experiments have previously been described in detail (Kiani et al., 2007), so we only give a brief summary here. Detailed descriptions of the statistical analysis and human localizer experiments are in the Supplemental Data.

Stimuli Presented to Humans and Monkeys

The stimuli presented to monkeys and humans were 92 color photographs (175 × 175 pixels) of isolated real-world objects on a gray background (Figure S1). The objects included natural and artificial inanimate objects as well as faces and bodies of humans and nonhuman animals. No predefined stimulus grouping was implied in either the experimental design or the core analyses for either species.

Monkey Experiments

Experimental Design and Task

Two alert monkeys were presented with the 92 images in rapid succession (stimulus duration, 105 ms; interstimulus interval, 0 ms) as part of a larger

set of over 1000 similar images while they performed a fixation task. Fixation was monitored with an infra-red eye-tracking system. Stimuli were presented in a pseudorandom order. The stimulus sequence started after the monkey maintained fixation for 300 ms. Stimuli spanned a visual angle of about 7°. Each stimulus lasted for 105 ms and was followed by another stimulus without intervening interstimulus interval.

Brain-Activity Measurements

Neuronal activity was recorded extracellularly with tungsten electrodes, one cell at a time. The cells were located in anterior IT cortex (anterior 13–20 mm, distributed over the ventral bank of the superior temporal sulcus and the ventral convexity up to the medial bank of the anterior middle temporal sulcus), in the right hemisphere in monkey 1 and in the left in monkey 2. On average, the stimulus set was repeated 9 ± 2 (median, 10) times for each recording site. A different random stimulus sequence was used on each repetition for each recording site in order to avoid consistent interactions between successively presented stimuli.

Human Experiments

Experimental Design and Task

We presented the 92 images to subjects in a "quick" event-related fMRI experiment, which balances the need for separable hemodynamic responses (suggesting a slow event-related design) and the need for presenting many stimuli in the limited time-span of the fMRI experiment (suggesting a rapid event-related design). The experiment included four additional images, which were excluded from the interspecies analyses because of insufficient monkey data (Figure S1). Stimuli spanned a visual angle of 2.9° and were presented foveally for a duration of 300 ms on a constantly visible uniform gray background. Stimuli were centered with respect to a fixation cross superimposed to them.

Each stimulus was presented exactly once in each run. The sequence also included 40 null trials with no stimulus presented (4 of them at the beginning, 4 of them at the end, and 32 randomly interspersed in the sequence). The trial-onset asynchrony was 4 s; the stimulus-onset asynchrony was either 4 s or a multiple of that duration when null trials occurred in the sequence. The trials (including 96 stimulus presentations and 32 interspersed null trials) occurred in random order (no sequence optimization). We used a different random sequence on each of up to 14 runs (spread over two fMRI sessions) per subject. A run lasted 9 min and 4 s (4 + 96 + 32 + 4 = 136 trials, each 4 s long).

Subjects continually fixated a fixation cross superimposed to the stimuli and performed a color-discrimination task. During stimulus presentation the fixation cross turned from white to either green or blue and the subject responded with a right-thumb button press for blue and a left-thumb button press for green. The fixation-cross changes to blue or green were chosen according to an independent random sequence.

Brain-Activity Measurements

Blood-oxygen-level-dependent fMRI was performed at high spatial resolution using a 3T GE HDx MRI scanner. For signal reception, we used a receive-only whole-brain surface-coil array (16 elements, NOVA Medical Inc., Wilmington, MA). Twenty-five 2 mm axial slices (no gap) were acquired, covering the occipital and temporal lobe, using single-shot interleaved gradient-recalled Echo Planar Imaging (EPI) with a sensitivity-encoding sequence (SENSE, acceleration factor: 2, Prüssmann 2004). Imaging parameters were as follows: EPI matrix size: 128 × 96, voxel size: 1.95 × 1.95 × 2 mm^3, echo time (TE): 30 ms, repetition time (TR): 2 s. Each functional run consisted of 272 volumes (9 min and 4 s per run). Four subjects were scanned in two separate sessions each, resulting in 11 to 14 runs per subject, yielding a total of 49 runs (equivalent to 7 hr, 24 min, and 16 s of fMRI data). As an anatomical reference, we acquired high-resolution T1-weighted whole-brain anatomical scans with a Magnetization Prepared Rapid Gradient Echo (MPRAGE) sequence. Imaging parameters were as follows: matrix size: 256 × 256, voxel size: 0.86 × 0.86 × 1.2 mm^3, 124 slices.

SUPPLEMENTAL DATA

The Supplemental Data can be found with this article online at http://www.neuron.org/supplemental/S0896-6273(08)00943-4.

ACKNOWLEDGMENTS

We thank Chris Baker, James Haxby, Alex Martin, Nancy Kanwisher, Dwight Kravitz, Leslie Ungerleider, and Vina Vo for helpful discussions. This research was supported in part by the Intramural Program of the National Institute of Mental Health.

Accepted: October 13, 2008
Published: December 24, 2008

REFERENCES

Afraz, S.R., Kiani, R., and Esteky, H. (2006). Microstimulation of inferotemporal cortex influences face categorization. Nature *442*, 692–695.

Aguirre, G.K. (2007). Continuous carry-over designs for fMRI. Neuroimage *35*, 1480–1494.

Aguirre, G.K., Zarahn, E., and D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. Neuron *21*, 373–383.

Baker, C.I., Behrmann, M., and Olson, C.R. (2002). Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. Nat. Neurosci. *5*, 1210–1216.

Bodurka, J., Ye, F., Petridou, N., Murphy, K., and Bandettini, P.A. (2007). Mapping the MRI voxel volume in which thermal noise matches physiological noise – Implications for fMRI. Neuroimage *34*, 542–549.

Capitani, E., Laiacona, M., Mahon, B., and Caramazza, A. (2003). What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. Cogn. Neuropsychol. *20*, 213–261.

Carlson, T.A., Schrater, P., and He, S. (2003). Patterns of activity in the categorical representations of objects. J. Cogn. Neurosci. *15*, 704–717.

Cox, D.D., and Savoy, R.L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage *19*, 261–270.

Denys, K., Vanduffel, W., Fize, D., Nelissen, K., Peuskens, H., Van Essen, D., and Orban, G.A. (2004). The processing of visual shape in the cerebral cortex of human and nonhuman primates: a functional magnetic resonance imaging study. J. Neurosci. *24*, 2551–2565.

Desimone, R., Albright, T.D., Gross, C.G., and Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. J. Neurosci. *4*, 2051–2062.

Downing, P.E., Jiang, Y., Shuman, M., and Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. Science *293*, 2470–2473.

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

Edelman, S. (1998). Representation is representation of similarities. Behav. Brain Sci. *21*, 449–467.

Edelman, S., Grill-Spector, K., Kushnir, T., and Malach, R. (1998). Towards direct visualization of the internal shape space by fMRI. Psychobiology *26*, 309–321.

Eger, E., Ashburner, J., Haynes, J.D., Dolan, R.J., and Rees, G. (2008). FMRI activity patterns in human LOC carry information about object exemplars within category. J. Cogn. Neurosci. *20*, 356–370.

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

Freedman, D.J., and Assad, J.A. (2006). Experience-dependent representation of visual categories in parietal cortex. Nature *443*, 85–88.

Freedman, D.J., Riesenhuber, M., Poggio, T., and Miller, E.K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. Science *291*, 312–316.

Freedman, D.J., Riesenhuber, M., Poggio, T., and Miller, E.K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. J. Neurosci. *23*, 5235–5246.

Gauthier, I., Skudlarski, P., Gore, J.C., and Anderson, A.W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. Nat. Neurosci. *3*, 191–197.

Grill-Spector, K., Kourtzi, Z., and Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. Vision Res. *41*, 1409–1422.

Hanson, S.J., Matsuka, T., and Haxby, J.V. (2004). Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a "face" area? Neuroimage *23*, 156–166.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science *293*, 2425–2430.

Humphreys, G.W., and Forde, E.M.E. (2001). Hierarchies, similarity, and interactivity in object recognition: "Category-specific" neuropsychological deficits. Behav. Brain Sci. *24*, 453–509.

Hung, C.P., Kreiman, G., Poggio, T., and DiCarlo, J.J. (2005). Fast readout of object identity from macaque inferior temporal cortex. Science *310*, 863–866.

Johnson, S.C. (1967). Hierarchical clustering schemes. Psychometrika *32*, 241–254.

Kamitani, Y., and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. Nat. Neurosci. *8*, 679–685.

Kanwisher, N., McDermott, J., and Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. *17*, 4302–4311.

Kay, K.N., Naselaris, T., Prenger, R.J., and Gallant, J.L. (2008). Identifying natural images from human brain activity. Nature *452*, 352–355.

Kayaert, G., Biederman, I., and Vogels, R. (2005). Representation of regular and irregular shapes in macaque inferotemporal cortex. Cereb. Cortex *15*, 1308–1321.

Kiani, R., Esteky, H., and Tanaka, K. (2005). Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. J. Neurophysiol. *94*, 1587–1596.

Kiani, R., Esteky, H., Mirpour, K., and Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. J. Neurophysiol. *97*, 4296–4309.

Kirchner, H., and Thorpe, S.J. (2006). Ultra-rapid object detection with saccadic eye-movements: visual processing speed revisited. Vision Res. *46*, 1762–1776.

Kreiman, G., Koch, C., and Fried, I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. Nat. Neurosci. *3*, 946–953.

Kriegeskorte, N., and Bandettini, P. (2007). Analyzing for information, not activation, to exploit high-resolution fMRI. Neuroimage *38*, 649–662.

Kriegeskorte, N., Goebel, R., and Bandettini, P. (2006). Information-based functional brain mapping. Proc. Natl. Acad. Sci. USA *103*, 3863–3868.

Kriegeskorte, N., Formisano, E., Sorger, B., and Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. Proc. Natl. Acad. Sci. USA *104*, 20600–20605.

Kriegeskorte, N., Mur, M., and Bandettini, P.A. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. Front. Syst. Neurosci. *2*, in press. Published online November 24, 2008. 10.3389/neuro.06.004.2008.

Kruskal, J.B., and Wish, M. (1978). Multidimensional Scaling (Beverly Hills, CA: Sage University Series).

Laakso, A., and Cottrell, G.W. (2000). Content and cluster analysis: Assessing representational similarity in neural systems. Philos. Psychol. *13*, 47–76.

Lehky, S.R., and Sereno, A.B. (2007). Comparison of shape encoding in primate dorsal and ventral visual pathways. J. Neurophysiol. *97*, 307–319.

Lerner, Y., Epshtein, B., Ullman, S., and Malach, R. (2008). Class information predicts activation by object fragments in human object areas. J. Cogn. Neurosci. *20*, 1189–1206.

Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. Nature *412*, 150–157.

Mahon, B.Z., Milleville, S.C., Negri, G.A.L., Rumiati, R.I., Caramazza, A., and Martin, A. (2007). Action-related properties shape object representations in the ventral stream. Neuron *55*, 507–520.

Martin, A. (2007). The representations of object concepts in the brain. Annu. Rev. Psychol. *58*, 25–45.

Martin, A., Wiggs, C.L., Ungerleider, L.G., and Haxby, J.V. (1996). Neural correlates of category-specific knowledge. Nature *379*, 649–652.

McClelland, J.L., and Rogers, T.T. (2003). The parallel distributed processing approach to semantic cognition. Nat. Rev. Neurosci. *4*, 310–322.

O'Toole, A., Jiang, F., Abdi, H., and Haxby, J.V. (2005). Partially distributed representation of objects and faces in ventral temporal cortex. J. Cogn. Neurosci. *17*, 580–590.

Op de Beeck, H., Wagemans, J., and Vogels, R. (2001). Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. Nat. Neurosci. *4*, 1244–1252.

Orban, G.A., Van Essen, D., and Vanduffel, W. (2004). Comparative mapping of higher visual areas in monkeys and humans. Trends Cogn. Sci. *8*, 315–324.

Patterson, K., Nestor, P.J., and Rogers, T.T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. Nat. Rev. Neurosci. *8*, 976–987.

Prüssmann, K.P. (2004). Parallel imaging at high field strength: Synergies and joint potential. Top. Magn. Reson. Imaging *15*, 237–244.

Puce, A., Allison, T., Gore, J.C., and McCarthy, G. (1995). Face-sensitive regions in human extrastriate cortex studied by functional MRI. J. Neurophysiol. *74*, 1192–1199.

Quiroga, R.Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. Nature *435*, 1102–1107.

Riesenhuber, M., and Poggio, T. (2002). Neural mechanisms of object recognition. Curr. Opin. Neurobiol. *12*, 162–168.

Serre, T., Wolf, L., and Poggio, T. (2005). Object recognition with features inspired by visual cortex. In: Computer Vision and Pattern Recognition (CVPR 2005), San Diego, CA, USA.

Serre, T., Oliva, A., and Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. Proc. Natl. Acad. Sci. USA *104*, 6424–6429.

Shepard, R.N. (1980). Multidimensional scaling, tree-fitting, and clustering. Science *210*, 390–398.

Shepard, R.N., and Chipman, S. (1970). Second-order isomorphism of internal representations: Shapes of states. Cognit. Psychol. *1*, 1–17.

Sigala, N., and Logothetis, N.K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. Nature *415*, 318–320.

Tanaka, K. (1996). Inferotemporal cortex and object vision. Annu. Rev. Neurosci. *19*, 109–139.

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. Nature *381*, 520–522.

Tootell, R.B., Tsao, D., and Vanduffel, W. (2003). Neuroimaging weighs in: humans meet macaques in "primate" visual cortex. J. Neurosci. *23*, 3981–3989.

Torgerson, W.S. (1958). Theory and Methods of Scaling (New York: Wiley).

Tsao, D.Y., Freiwald, W.A., Knutsen, T.A., Mandeville, J.B., and Tootell, R.B.H. (2003). Faces and objects in macaque cerebral cortex. Nat. Neurosci. *6*, 989–995.

Tsao, D.Y., Freiwald, W.A., Tootell, R.B.H., and Livingstone, M.S. (2006). A cortical region consisting entirely of face-selective cells. Science *311*, 670–674.

Tyler, L.K., and Moss, H.E. (2001). Towards a distributed account of conceptual knowledge. Trends Cogn. Sci. *5*, 244–252.

Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. Trends Cogn. Sci. *11*, 58–64.

Van Essen, D.C., and Dierker, D.L. (2007). Surface-based and probabilistic atlases of primate cerebral cortex. Neuron *56*, 209–225.

Van Essen, D.C., Lewis, J.W., Drury, H.A., Hadjikhani, N., Tootell, R.B.H., Bakircioglu, M., and Miller, M.I. (2001). Mapping visual cortex in monkeys and humans using surface-based atlases. Vision Res. *41*, 1359–1378.

Vogels, R. (1999). Categorization of complex visual images by rhesus monkeys. Part 2: single-cell study. Eur. J. Neurosci. *11*, 1239–1255.

Supplemental Material; Neuron Volume 60

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA

Matching categorical object representations in inferior temporal cortex of man and monkey

---

Supplemental Figures

Overview

Fig.	Title
S1	Stimuli
S2	Interspecies correlation of IT object dissimilarities related to single stimuli
S3	Statistical analyses of single-stimulus interspecies dissimilarity correlations
S4	Species-specific face representation
S5	Representation in human early visual cortex defined at 224-5000 voxels
S6	Model representations (1)
S7	Model representations (2)
S8	Unsupervised stimulus-quartet arrangements for monkey and human IT
S9	Representation in left and right human IT
S10	Representation in human IT defined at 100-10,000 voxels
S11	Representation in human IT without FFA and PPA
S12	Linear discriminant analysis for human IT and early visual cortex
S13	Scatterplot of stimulus-pairs relating monkey- and human-IT representations
S14	Scatterplot of stimulus-pairs relating monkey- and human-IT representations (continued)

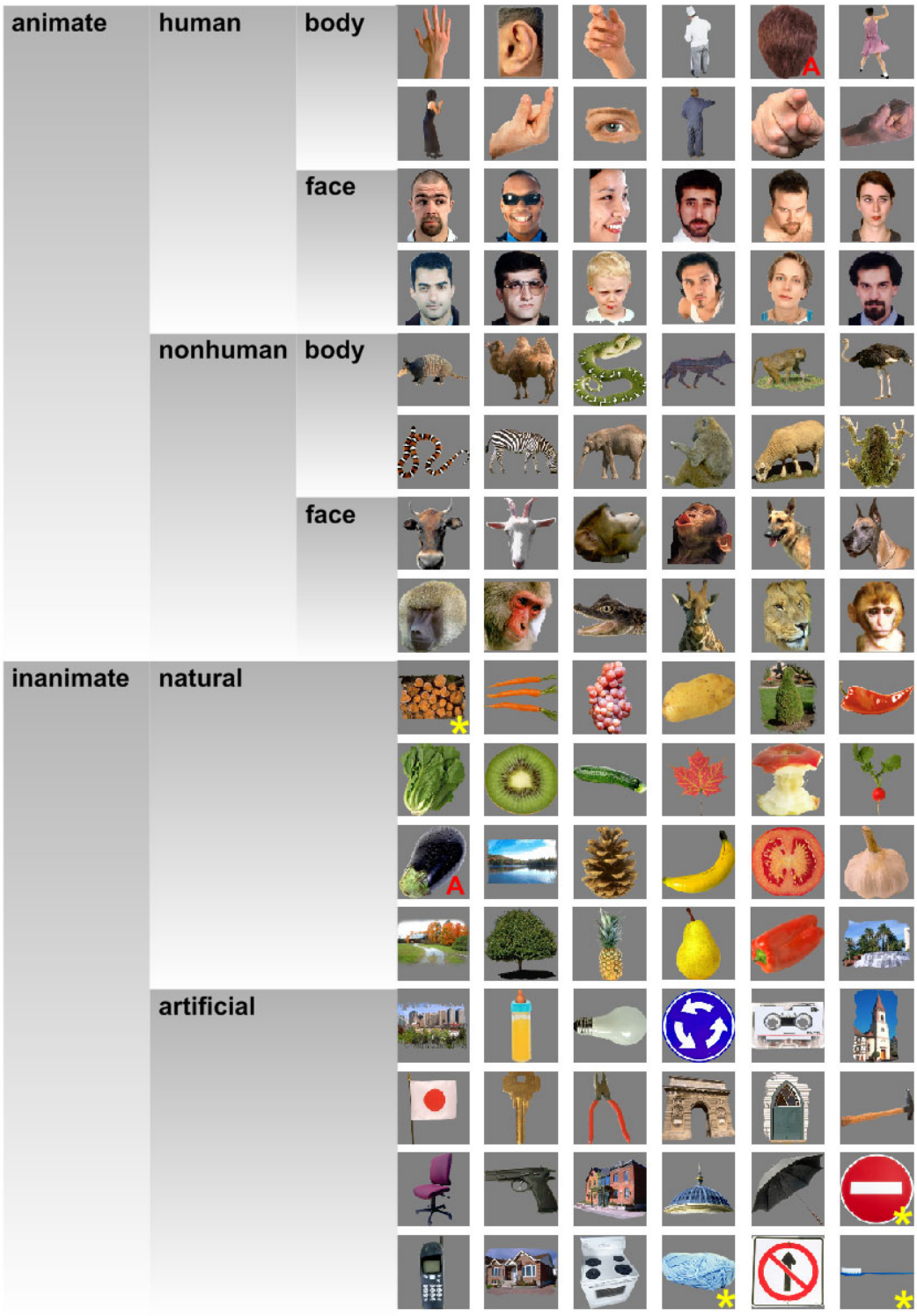


Fig. S1. Stimuli. The object images presented to monkeys and humans. The four images marked by yellow stars were excluded from the analysis because of insufficient data in the monkey experiments. Responses to the remaining 92 form the basis of all analyses. Several of our human subjects described two of the stimuli as ambiguous during debriefing. These two stimuli (egg plant, back of a human head) are marked by a red “A”.

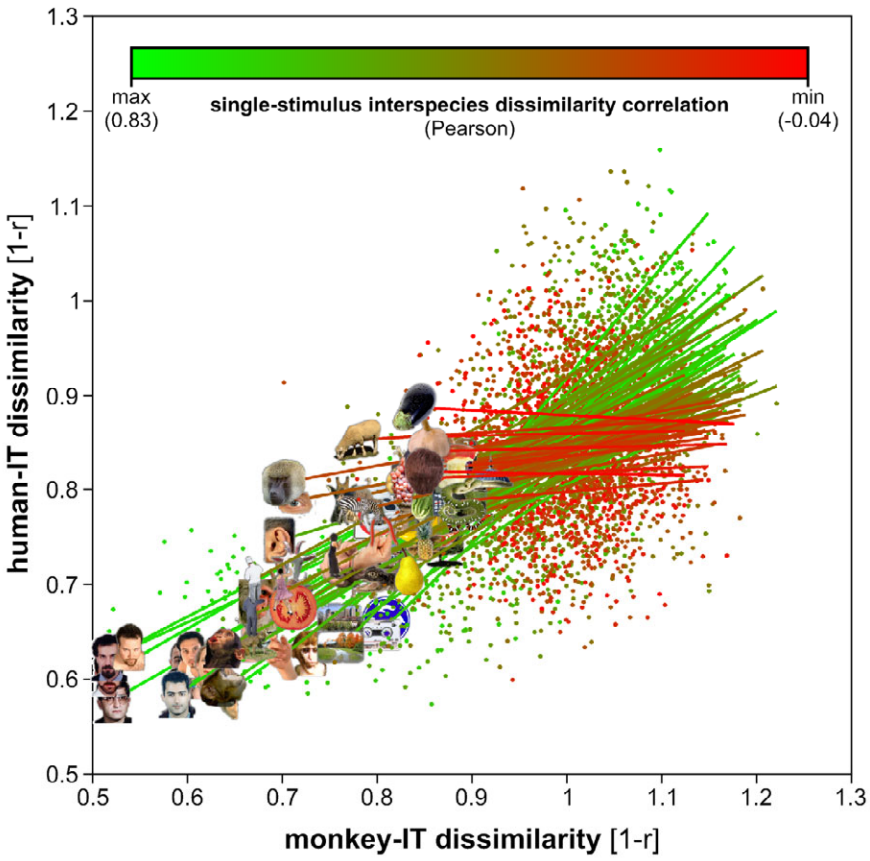
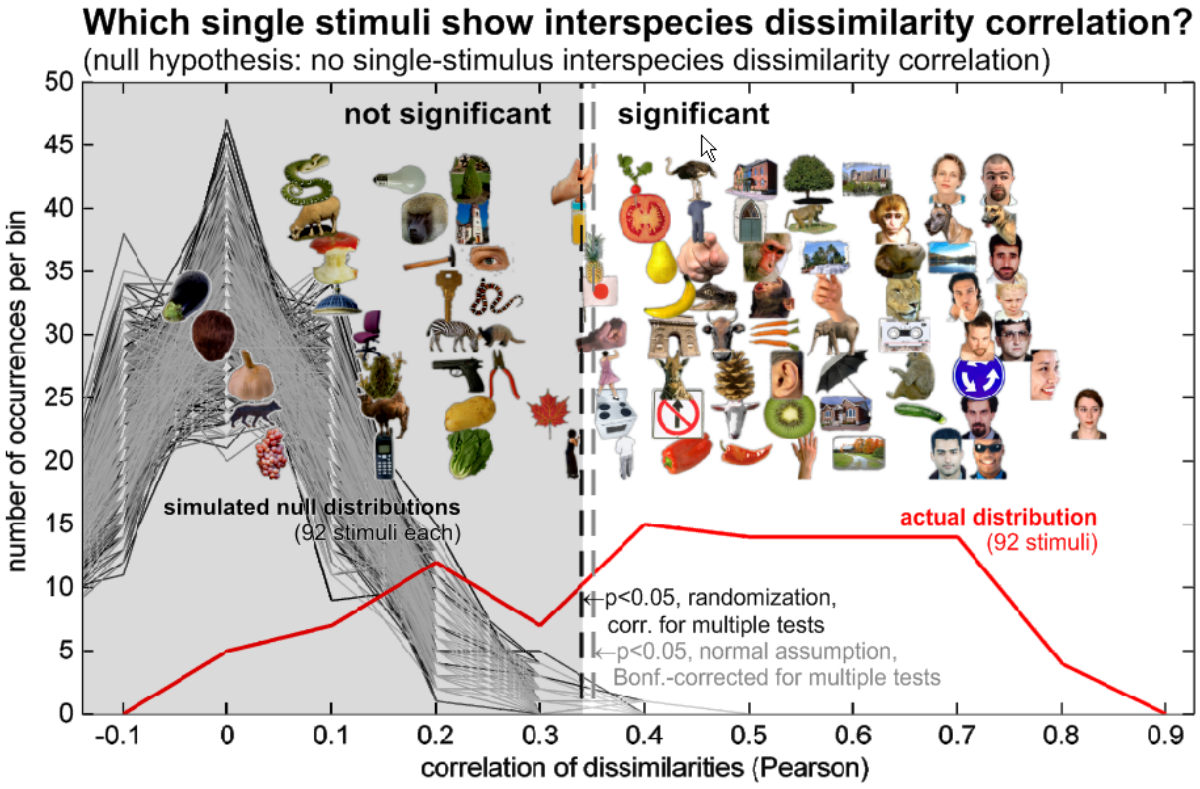


Fig. S2. Interspecies correlation of IT object dissimilarities related to single stimuli. This figure addresses the question to what extent each of the stimuli is similarly represented in both species. A stimulus is considered "similarly represented" if its pattern of representational dissimilarities to the other 91 stimuli is correlated between human and monkey IT. For each stimulus, we consider its row (or, equivalently, its column) in the representational dissimilarity matrix in each species (Fig. 1) and plot the monkey-IT dissimilarities against the human-IT dissimilarities (analogously to Fig. 3). In addition, we plot a straight line for each stimulus, which is obtained as the least-squares fit to the 91 human-IT dissimilarities (vertical axis) of that stimulus and extends horizontally along the range of the corresponding 91 monkey-IT dissimilarities. The stimulus itself is plotted on the left end of the line. In order to highlight the stimuli most inconsistently represented in monkey and human, the scatterplots, fit lines, and stimuli are overplotted in the order of their interspecies representational-dissimilarity correlation, starting from the most highly correlated (green scatterplot and fit line, thin fibers in Fig. 2b) and progressing to the least interspecies-correlated stimulus (red scatterplot and fit line, thick fibers in Fig. 2b). The scatterplots and fit lines for intermediate stimuli are plotted in intermediate colors ranging from green to red, which linearly reflect the interspecies correlation (see colorbar).

a


Which single stimuli show interspecies dissimilarity correlation?
(null hypothesis: no single-stimulus interspecies dissimilarity correlation)
not significant
significant
number of occurrences per bin
simulated null distributions
(92 stimuli each)
actual distribution
(92 stimuli)
p<0.05, randomization,
corr. for multiple tests
p<0.05, normal assumption,
Bonf.-corrected for multiple tests
correlation of dissimilarities (Pearson)

b

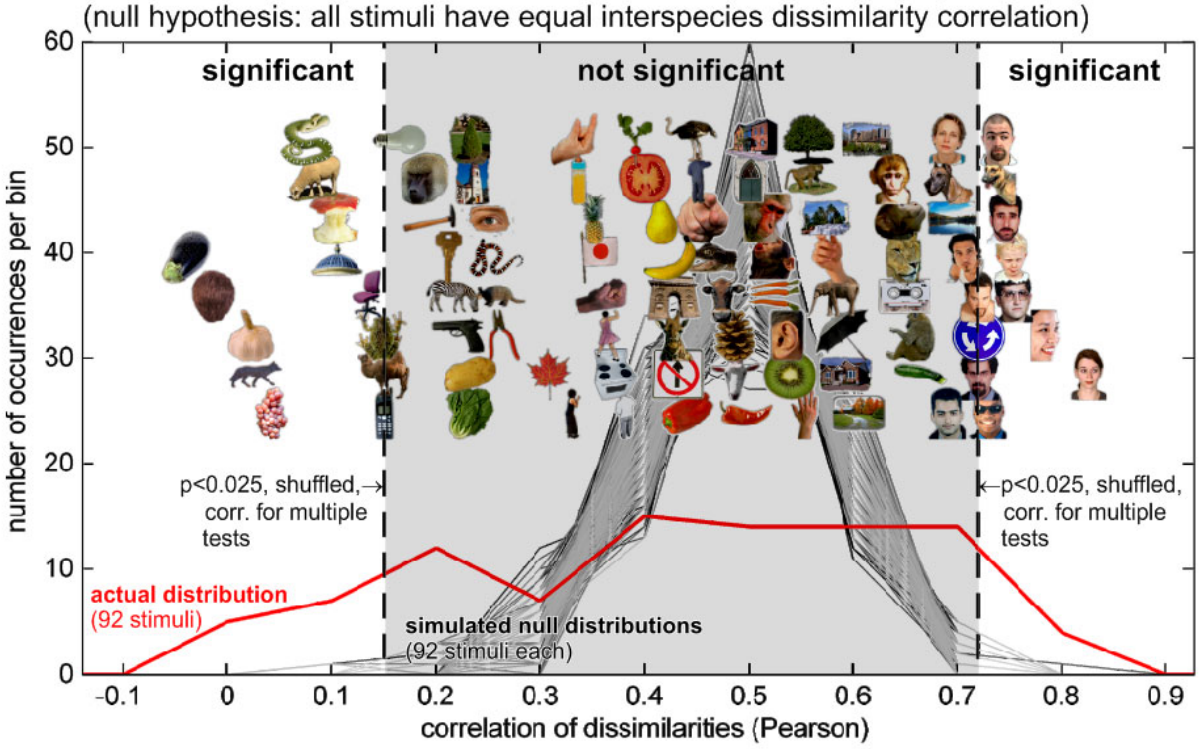

Do stimuli show different interspecies dissimilarity correlation?
(null hypothesis: all stimuli have equal interspecies dissimilarity correlation)
significant
not significant
significant
number of occurrences per bin
p<0.025, shuffled,
corr. for multiple
tests
p<0.025, shuffled,
corr. for multiple
tests
actual distribution
(92 stimuli)
simulated null distributions
(92 stimuli each)
correlation of dissimilarities (Pearson)

Fig. S3. Statistical analyses of single-stimulus interspecies dissimilarity correlations. In both panels (a and b), each stimulus is placed along a horizontal axis according to its interspecies dissimilarity correlation (see previous figure for details). The stimuli are spaced out vertically so that they could be displayed in a larger size. In each panel (a and b), the red line shows the interspecies-correlation histogram for the 92 stimuli. (**a**) Statistical analysis addressing the question, which single stimuli show interspecies dissimilarity correlation. As in Fig. 3, we use randomization of stimulus labels to test the interspecies dissimilarity correlation. Here we assess, which single-stimulus interspecies correlations are significant. Each gray line shows a histogram obtained by randomizing the stimulus labels for one species before computing the interspecies dissimilarity correlations. Each of 1000 such randomizations simulates the null hypothesis that there are no interspecies correlations. We define an interspecies correlation threshold (dashed black line) that is exceeded by even a single stimulus in only 5% of the 1000 null simulations (i.e. by thresholding the randomization distribution of maxima among the 92 interspecies correlations obtained in each null simulation). This threshold limits the family-wise false-positives rate at $p<0.05$. A similar threshold (dashed gray line) is obtained by (incorrectly) assuming normality and independence, and using the Bonferroni method to control the family-wise false-positives rate. Using either method, about 61 of the 92 stimuli, including all faces, exhibit significant interspecies correlation. (**b**) Statistical analysis addressing the question, whether stimuli vary in terms of interspecies correlation. The analysis in (a) highlights some stimuli and not others as significantly consistently represented in IT of both species. However, this does not mean that there are significant differences between single-stimulus interspecies correlations. A mixture of significant and insignificant interspecies correlations as obtained in (a) could result from an interspecies correlation constant across all stimuli in conjunction with noise. We therefore tested the null hypothesis that all single-stimulus interspecies correlations are equal. We simulated the null distribution of equal interspecies correlation across all stimuli by shuffling interspecies pairs of dissimilarites across stimuli (without replacement). This conserves the overall interspecies correlation and yields single-stimulus interspecies correlations that differ only because of the noise and limited data points (91 interspecies dissimilarity pairs for each stimulus). Each of 1000 null simulations yielded an interspecies correlation for each stimulus (1000 histograms shown in gray). We define a lower interspecies correlation threshold (dashed black line on the left), such that lower interspecies correlations occur for even a single stimulus in only 2.5% of the 1000 null simulations (i.e. by thresholding the randomization distributions of minima). Analogously, we define an upper threshold (dashed black line on the right), such that higher interspecies correlations occur for even a single stimulus in only 2.5% of the 1000 null simulations (i.e. by thresholding the randomization distributions of maxima). These two thresholds limit the family-wise false-positives rate at $p<0.05$. Results show that human faces exhibit significantly higher interspecies correlations than the stimulus set as a whole and several stimuli (including images of animate and inanimate objects) exhibit significantly lower interspecies correlations. The two stimuli with the lowest interspecies correlation (eggplant, back-view of human head) were the only two stimuli described as ambiguous by human subjects during debriefing (Fig. S1). Their significantly low interspecies correlation is consistent with the idea that the IT representation reflects not only the visual appearance, but also the conceptual interpretation of a stimulus.

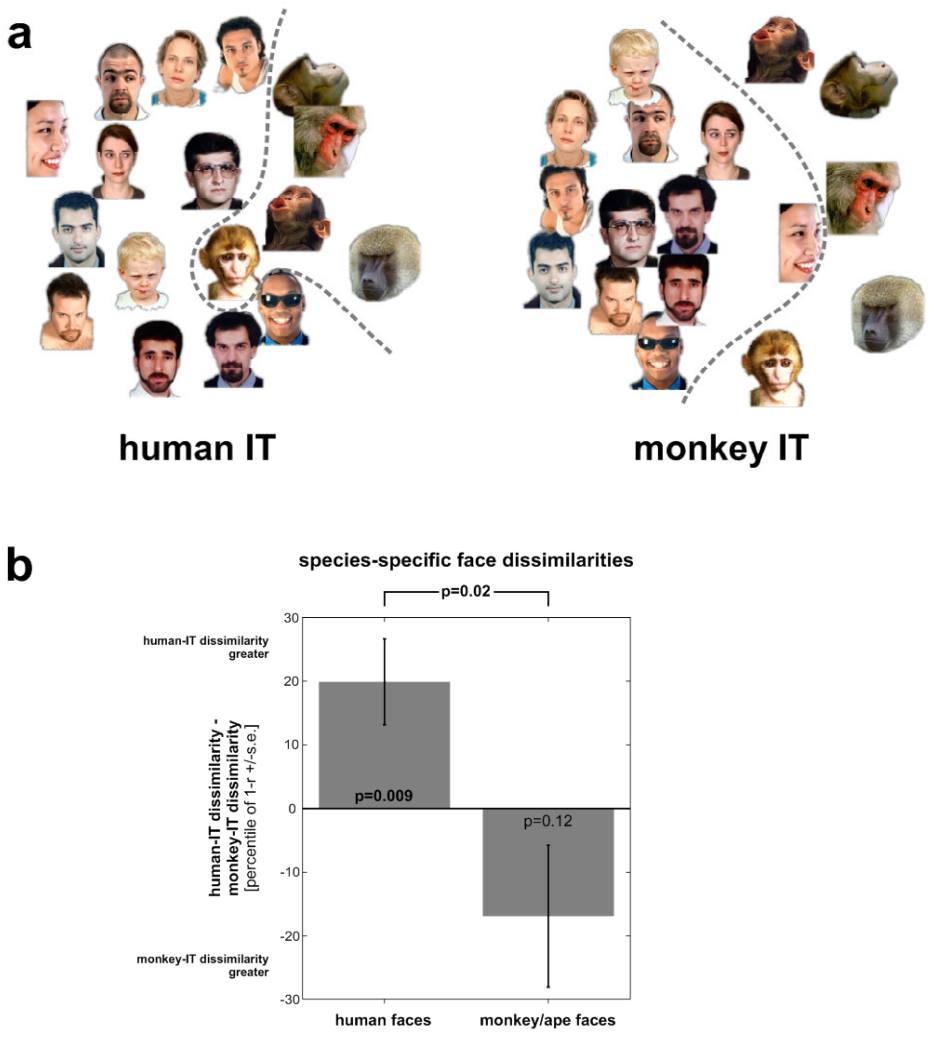


Fig. S4. Species-specific face representation. Here we selectively analyzed the representation of monkey, ape, and human faces in monkey and human IT. (**a**) The face stimuli have been arranged such that their pairwise distances approximately reflect response-pattern similarity. The arrangement was computed by multidimensional scaling with the same settings as in Fig. 2 and elsewhere in this paper (dissimilarity: 1 - Pearson r, criterion: metric stress, arrangements scaled to match the areas of their convex hulls and rigidly aligned for easier comparison with the Procrustes method). A line (dashed gray) separating the monkey/ape faces from the human faces has been manually added. Visual inspection suggests that human IT may better discriminate the human faces than the monkey faces and that the converse may hold for monkey IT. (**b**) Statistical analysis comparing human- and monkey-IT mean dissimilarities for human faces (left) and for monkey/ape faces (right). The left bar shows that dissimilarities among human faces are significantly larger in human IT than in monkey IT (p=0.009). The right bar shows that dissimilarities among monkey/ape faces are larger in monkey IT than in human IT in our data, although the effect is not significant (p=0.12). The difference between the two effects is significant (p=0.02). Because the dissimilarities are not independent or normal, the statistical tests and error bars (indicate +/- 1 standard error) are based on bootstrap resampling of the stimulus set. Note that our stimulus set is not well-suited for comparing the representation of human and monkey/ape faces, because faces were a small subset of our stimuli and because the monkey/ape faces were few and varied in species, pose, and view more than the human faces. The comparison in (b) of the representations of a given set of stimuli (either human faces or monkey faces) between human and monkey IT nevertheless provides an interesting lead for future studies designed to address this question.

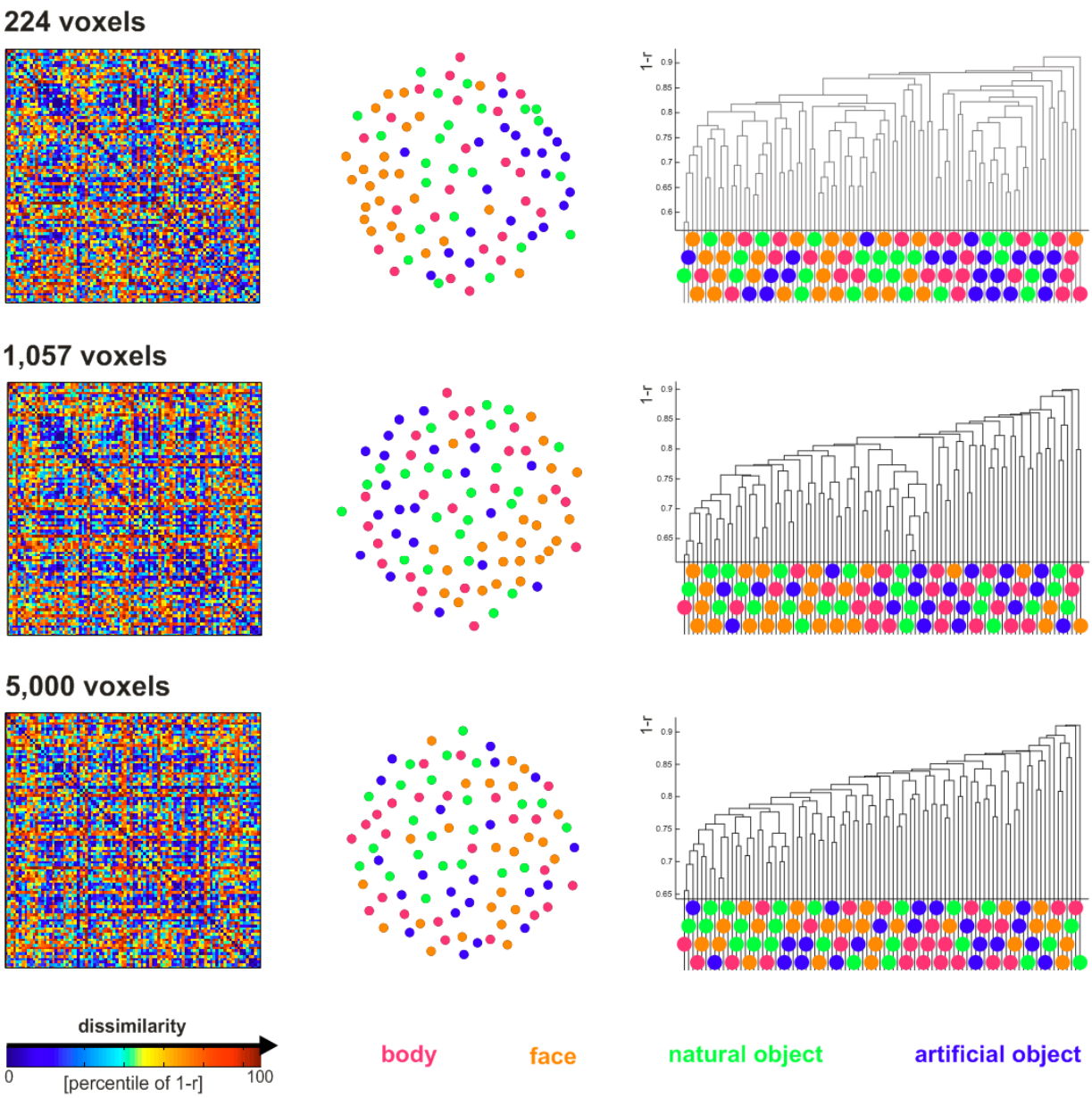


Fig. S5. Representation in human early visual cortex defined at 224-5000 voxels. Human early visual cortex shows no evidence of categorical clustering in the fMRI data. This result is independent of the number of voxels included in the region of interest (rows). Early visual cortex was defined by selecting the most visually responsive voxels within a manually drawn anatomical mask in each subject. As for human IT, independent data were used for voxel selection. Dissimilarity matrices (left), multidimensional scaling arrangements (middle), and hierarchical clustering trees (right) were computed with the same parameters as for IT (Figs. 1, 2, 4, respectively).

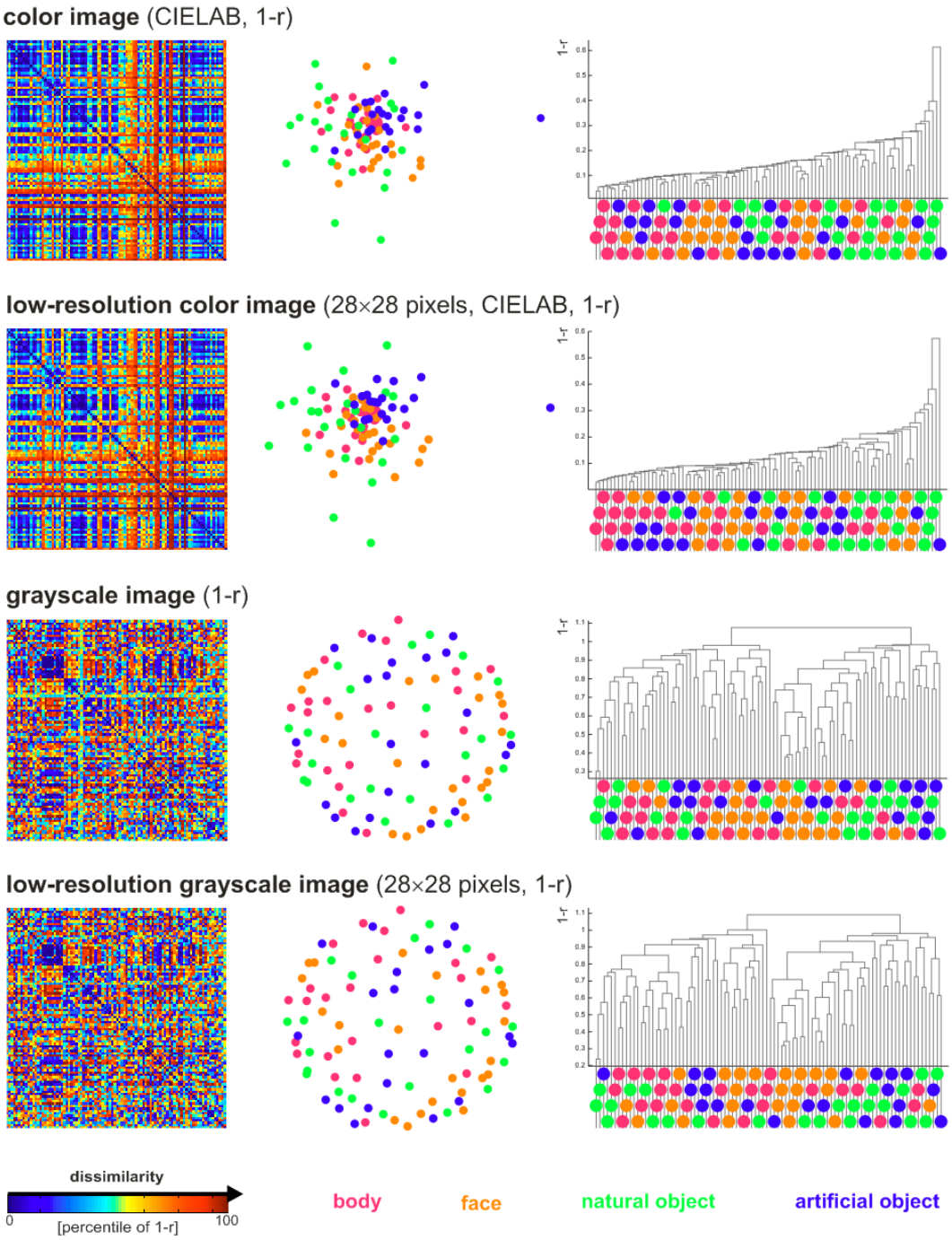


Fig. S6. Model representations (1). We processed our stimuli to obtain their representations in a number of low-level models (rows, continued in Fig. S7). We analyzed these model representations in the same way as the brain-activity data from early visual cortex and IT. None of the models could account for the categorical clustering found in monkey and human IT. The models are described in the section *Model representations* in the Supplemental Material. Dissimilarity matrices (left), multidimensional scaling arrangements (middle), and hierarchical clustering trees (right) were computed with the same parameters as for IT (Figs. 1, 2, 4, respectively).

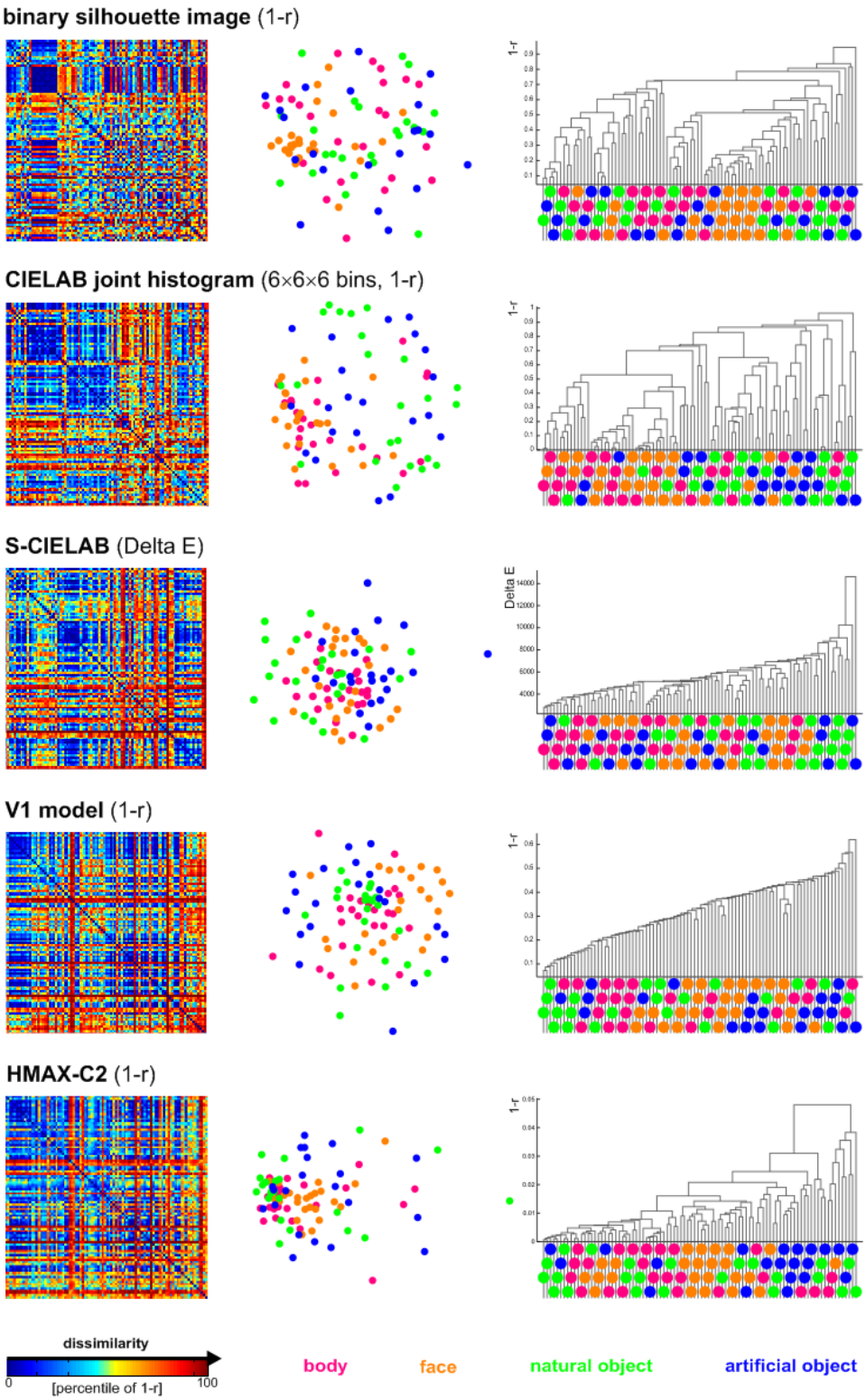


Fig. S7. Model representations (2). Continuation of Fig. S6. See legend of Fig. S6.

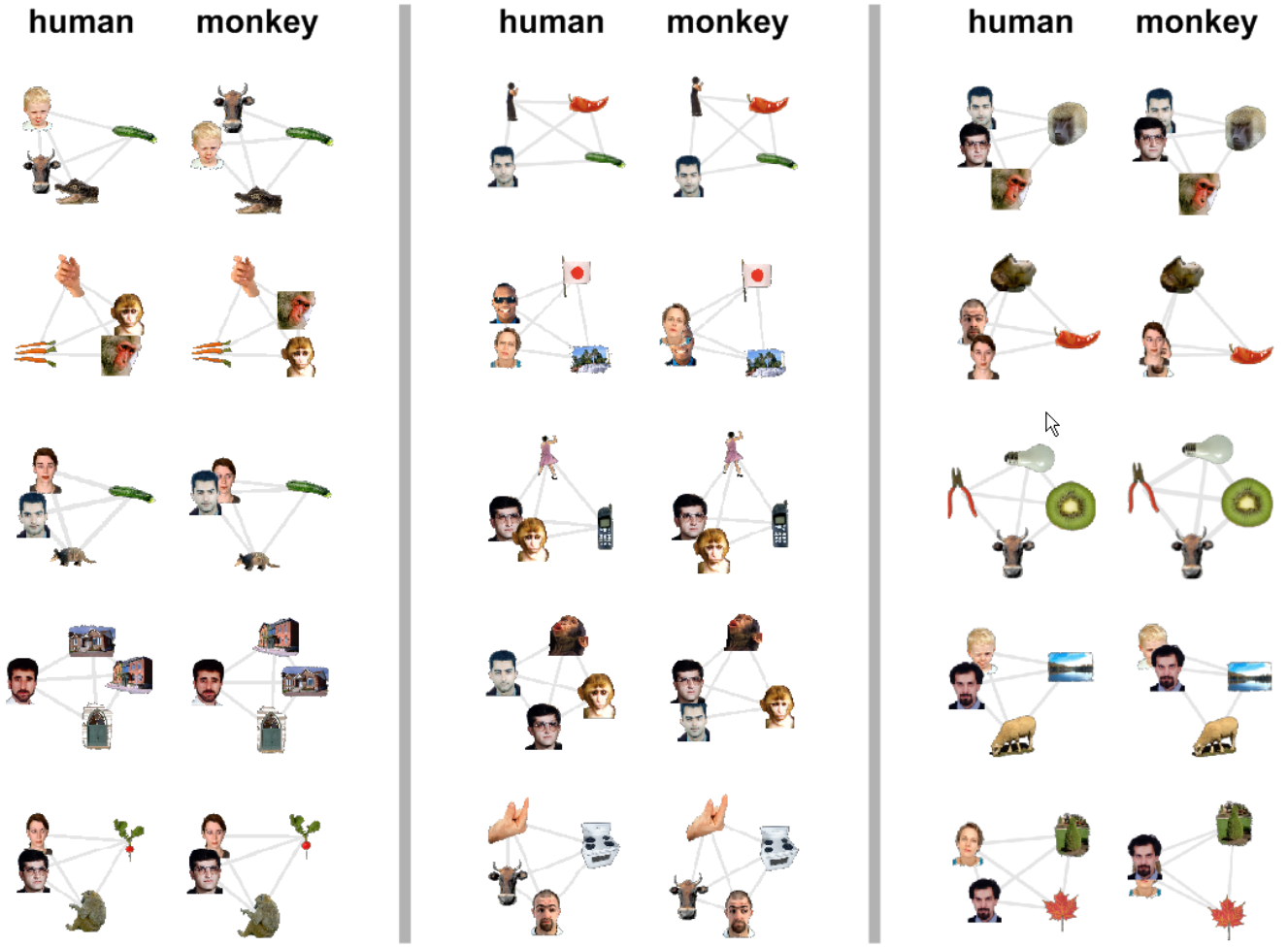


Fig. S8. Unsupervised stimulus-quartet arrangements for monkey and human IT. For 15 representative stimulus quartets, this figure depicts the representation in human and monkey IT in terms of unsupervised arrangements reflecting response-pattern similarity. As in Fig. 2, images placed close together elicited similar response patterns; images placed far apart elicited dissimilar response patterns. There is no special significance to the choice of 4 as the number of stimuli. However, considering quartets allows us to appreciate the underlying dissimilarity relationships at a glance. Moreover, the inevitable distortion of the original dissimilarities in the 2-dimensional arrangement is very small when only 4 stimuli are considered at a time. The arrangements were computed using multidimensional scaling with the same settings as for Fig. 2 (dissimilarity: 1-Pearson correlation, criterion: metric stress). For each stimulus quartet, the two arrangements (human, monkey) have been rigidly aligned for easier comparison (Procrustes alignment) and scaled to the same approximate size. We introduce “rubberband graphs” (gray lines connecting the stimuli) to depict the residual distortion: the gray lines behave like rubberbands, thinning when stretched beyond the length they are to represent and thickening when compressed. More precisely, the actual dissimilarity equals the area of the rubberband connection (dissimilarity = line thickness × line length).

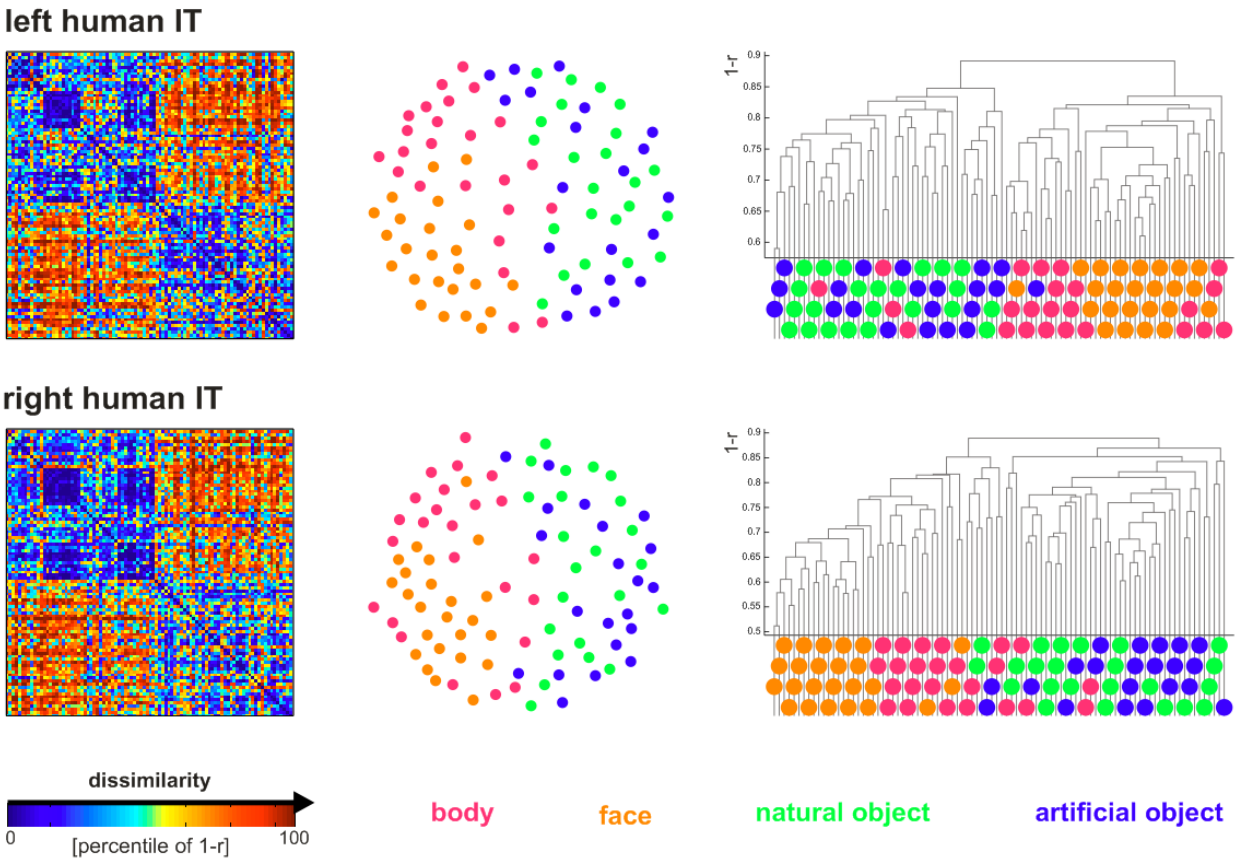


Fig. S9. Representation in left and right human IT. The categorical clustering in human IT is only weakly dependent on the cortical hemisphere (left human IT in top row, right human IT in bottom row). Here we selected 266 voxels according to their visual responsiveness (independent data) within each hemisphere's manually defined IT mask. Dissimilarity matrices (left), multidimensional scaling arrangements (middle), and hierarchical clustering trees (right) were computed with the same parameters as for Figs. 1, 2, 4.

.

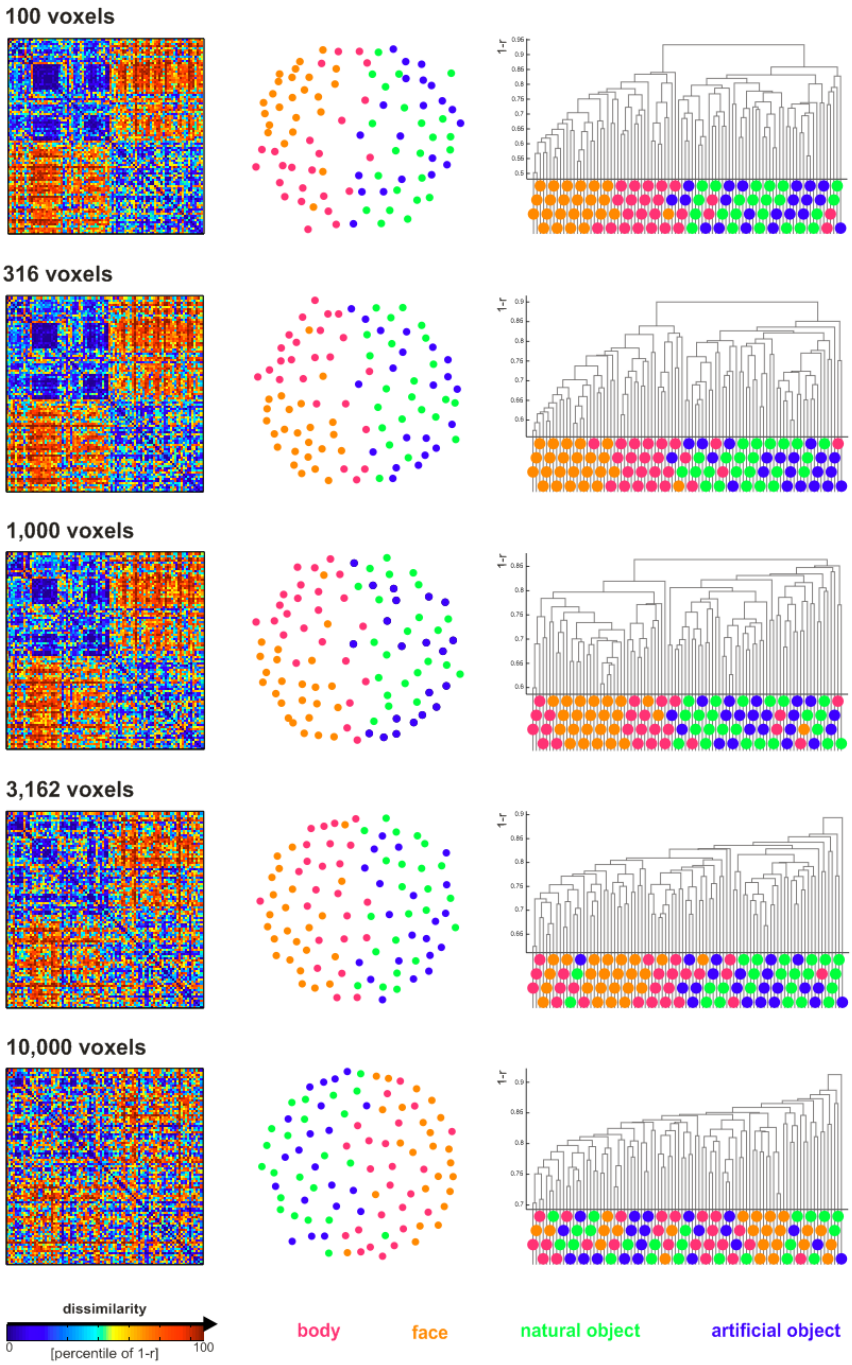


Fig. S10. Representation in human IT defined at 100-10,000 voxels. The similarity structure and categorical clustering characteristic of human IT is only weakly dependent on the number of voxels selected for inclusion in the region of interest. Voxels were selected according to their visual responsiveness as assessed with independent data. The human-IT region shown in the second row (316 voxels) is that used for Figs. 1-6. When thousands of voxels are included in the region of interest, the categorical structure becomes less distinct. Nevertheless multidimensional scaling still separates animate and inanimate objects at 10,000 voxels (bottom row, middle panel). Dissimilarity matrices (left), multidimensional scaling arrangements (middle), and hierarchical clustering trees (right) were computed with the same parameters as for Figs. 1, 2, 4.

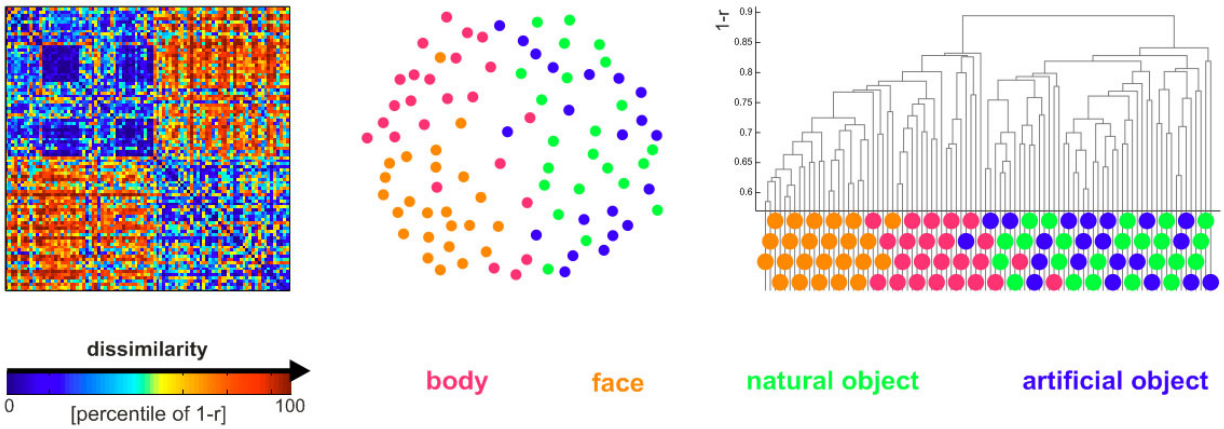


Fig. S11. Representation in human IT without FFA and PPA. Excluding FFA and PPA bilaterally from the voxels selected to define human IT did not qualitatively change the similarity structure or categorical clustering. FFA and PPA were defined in each hemisphere at 1141 mm^3 (150 voxels) and 1521 mm^3 (200 voxels), respectively, by means of an independent block-design localizer experiment. Human IT was defined bilaterally at 316 voxels as for Figs. 1-4, but FFA and PPA were first excluded from the cortex mask in both hemispheres. The dissimilarity matrix (left), multidimensional scaling arrangement (middle), and hierarchical clustering tree (right) were computed with the same parameters as for Figs. 1, 2, 4.

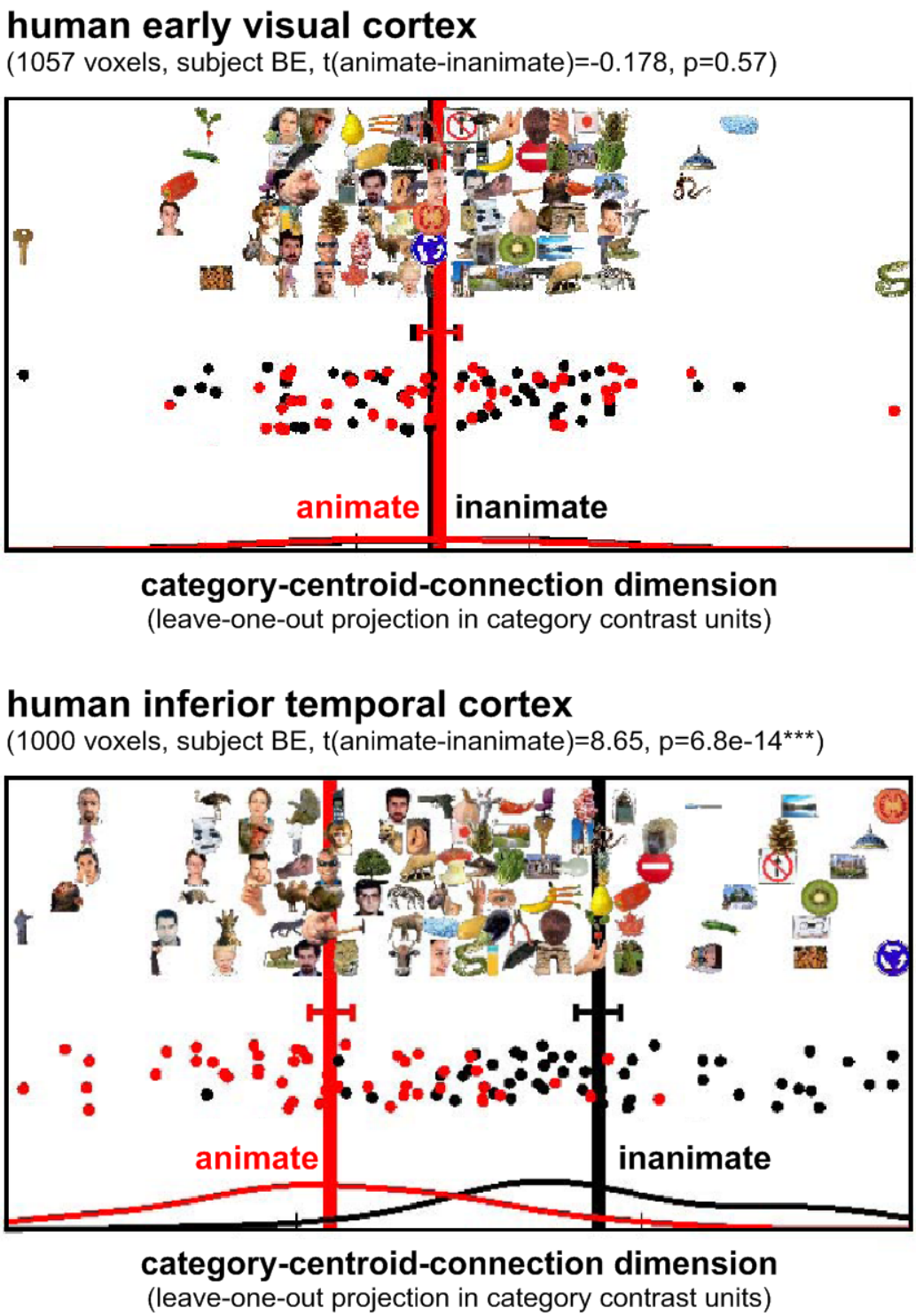


Fig. S12. Response patterns in human IT, but not human early visual cortex, allow linear discrimination between animates and inanimates. Single-subject data show highly significant linear discriminability of animates and inanimates in human IT (t(animate-inanimate)=8.65, p=6.8e-14), whereas linear discriminability of these categories is not evident in human early visual cortex (t(animate-inanimate)=-0.178, p=0.57). The single-image response patterns in early visual cortex (top) and IT (bottom) have been projected onto a dimension defined to discriminate animates from inanimates. The dimension used is the category-centroid-connection dimension (equivalent to the Fisher linear discriminant computed with the assumption of isotropic,

homoscedastic noise). To avoid circularity, the discriminant is computed using a leave-one-out procedure: In order to determine the location of a given single-image response pattern on the discriminant dimension, the other 95 single-image response patterns are used to compute the category centroids defining the discriminant dimension (the thin tick marks indicate the centroid locations defining the discriminant; left for animates, right for inanimates). Note that this approach uses not only independent response estimates, but also different images (the other 95) for defining the discriminant. The thick vertical lines indicate the category means (red for animate, black for inanimate) on the discriminant dimension. Error bars indicate the +/- 1 standard error of the mean. The 96 stimulus images (Fig. S1) have been located along the discriminant (upper portion of each panel) with vertical scattering to allow a larger image size. Colored dots (red for animate, black for inanimate) show the two category distributions (middle portion of each panel), again with vertical random scattering. Probability-density estimates (kernel-smoother method) are shown in the lower portion of each panel (red for animate, black for inanimate).

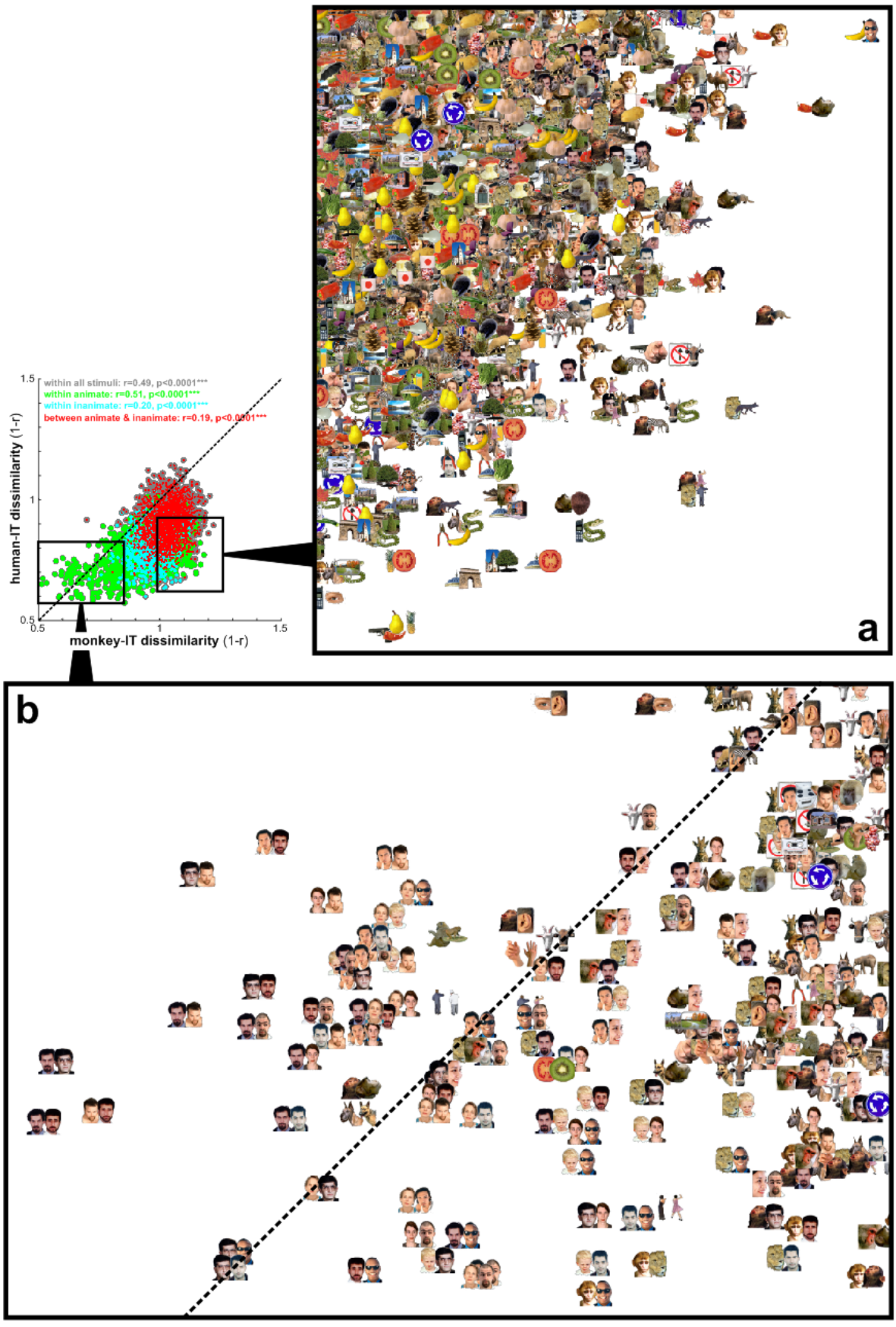

within all stimuli: r=0.49, p<0.0001***
within animate: r=0.51, p<0.0001***
within inanimate: r=0.20, p<0.0001***
between animate & inanimate: r=0.19, p<0.0001***
human-IT dissimilarity (1-r)
monkey-IT dissimilarity (1-r)
0.5
1
1.5
a
b

Fig. S13. Scatterplot of stimulus-pairs relating monkey- and human-IT representations. The scatterplot in Fig. 3a, containing a dot for each stimulus pair, relates monkey- and human-IT dissimilarities. Here we zoom in on two regions of Fig. 3a, in order to make space for plotting each pair of stimuli, whose dissimilarity in monkey and human IT determines the horizontal and vertical location of the corresponding dot in the scatterplot. For each pair, the two stimuli are placed side by side, centered on the location of the dot that represents the pair in Fig. 3a. (**a**) This region contains the stimulus pairs that are furthest from the line of identity. They have the greatest difference between monkey- and human-IT dissimilarity (1 - Pearson r) with the monkey dissimilarity greater than the human dissimilarity. (Note that the corresponding region with greater human- than monkey-IT dissimilarities is unpopulated.) An attractive interpretation would be that these pairs are more similar to humans than to monkeys. Note, however, that the relationship between human and monkey representational dissimilarities may not be linear. Plausible interpretations suggest themselves for many of the pairs, but remain speculative. (**b**) This region contains the stimulus pairs eliciting the most similar activity patterns in both monkey and human IT. This region is dominated by pairs of human faces.

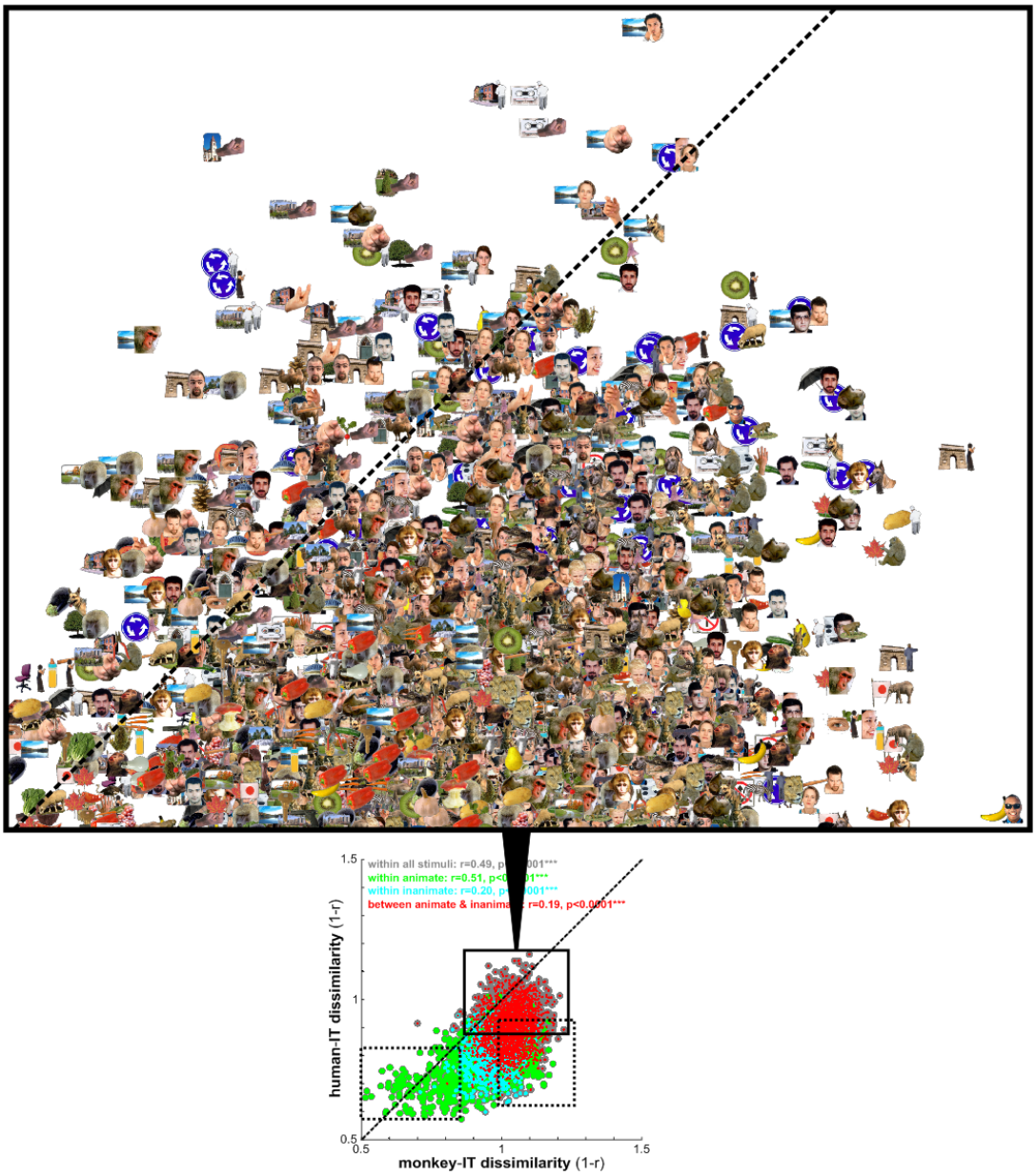


Fig. S14. Scatterplot of stimulus-pairs relating monkey- and human-IT representations (continued). This figure follows the same logic as the previous one, but zooms in on a third region of Fig. 3a. This region contains the stimulus pairs eliciting the most dissimilar activity patterns in both monkey and human IT. This region is dominated by pairs of stimuli crossing the animate-inanimate category boundary. (Dotted rectangles indicate the two regions zoomed in on in the previous figure.)

Supplemental text

Summary of the previous evidence on IT categoricality

Ever since neuropsychology described object-category-related deficits following brain damage (Humphreys and Forde, 2001; Capitani et al., 2003; Martin, 2007), it has been generally accepted that there is some relationship between conventional categories and human-IT representations. Human neuroimaging has investigated category-average responses, showing that IT contains focal regions whose activation is correlated with conventional categories (Puce et al., 1995; Martin et al., 1996; Kanwisher et al., 1997; Aguirre et al., 1998; Epstein and Kanwisher, 1998; Downing et al., 2001; Downing et al., 2006) and that the category can be read out from the IT response pattern with a linear classifier (Haxby et al., 2001; Cox and Savoy, 2003; Carlson et al., 2003). However, these studies investigated responses averaged across many different stimuli within predefined categories. This approach requires the assumption of a particular category structure and therefore cannot address whether the representation is inherently categorical. IT features might respond to natural image fragments that happen to be correlated with categories, without being optimized to distinguish categories.

Monkey studies, as well, have reported IT responses correlated with natural categories (Vogels, 1999; Tsao et al., 2003; Kiani et al., 2005; Hung et al., 2005; Tsao et al., 2006; Afraz et al., 2006) and novel, experimentally defined, categories (Sigala and Logothetis, 2002; Baker et al., 2002; Freedman et al., 2003). However, step-function-like categorical responses as reported for cells in medial temporal (Kreiman et al., 2000; Quiroga et al., 2005), prefrontal (Freedman et al., 2001), and parietal regions (Freedman and Assad, 2006) are not typically observed in either single IT cells (Vogels, 1999; Freedman et al., 2003; Kiani et al., 2007; but see Tsao et al., 2006) or category-sensitive fMRI responses (Haxby et al., 2001), suggesting that IT may have a lesser role in categorization (Freedman et al., 2003). Consistent with this perspective, one influential computational model of primate object recognition (Riesenhuber and Poggio, 2002; Serre et al., 2007) employs categorization training (i.e. supervised learning) to optimize the stage thought to correspond to prefrontal cortex. The model's IT stage is not optimized for categorization.

Representational similarity analysis

Estimation of single-image response patterns in the monkeys. The analyses are based on all cells that could be isolated and for which sufficient data was available across the stimuli. This yielded a total of 674 neurons for both monkeys combined (322 in monkey 1 and 352 in monkey 2). For each stimulus, each neuron's response amplitude was estimated as the average spike rate within a 140-ms window starting 71 ms after stimulus onset (for details, see Kiani et al., 2007).

Estimation of single-image response patterns in the humans. Single-image response patterns were estimated by univariate linear modeling. We concatenated the runs within a session along the temporal dimension. For each voxel, we performed a single univariate linear model fit to obtain a response-amplitude estimate for each of the 96 stimuli. The model included a hemodynamic-response predictor for each of the 96 stimuli. Since each stimulus occurred once in each run, each of the 96 predictors had one hemodynamic response per run and extended across all within-session runs included (i.e. all runs except those used for region-of-interest definition). The predictor time courses were computed using a linear model of the hemodynamic response (Boynton et al., 1996) and assuming an instant-onset rectangular neural response during each condition of visual stimulation. For each run, the design matrix included these stimulus-response predictors along with six head-motion-parameter time courses, a linear-trend predictor, a 6-predictor Fourier basis for nonlinear trends (sines and cosines of up to 3 cycles per run) and a confound-mean predictor. Trends were, thus, modeled by a separate set of predictors for each run. The trend predictors for a particular run had zero entries for all other runs along time. For head-motion models and confound means as well, separate predictors accounted for each run. For each stimulus, we converted the response-amplitude (beta) estimate map into a t map. The resulting t maps (one for each stimulus) were used for representational similarity analysis.

Computation of representational dissimilarity matrices. For each pair of stimuli, the dissimilarity between the associated response patterns is measured as 1 minus the Pearson linear correlation across cells or voxels within a region of interest (0 for perfect correlation, 1

for no correlation, 2 for perfect anticorrelation). The resulting dissimilarities for all pairs of object images are assembled in a representational dissimilarity matrix (RDM; Fig. 1). Each cell of the RDM, thus, compares the response patterns elicited by two stimuli. As a consequence, an RDM is symmetric about a diagonal of zeros.[1]

Testing relatedness of two representational dissimilarity matrices by randomization of condition labels. We use the Pearson correlation coefficient r to assess the relatedness of two RDMs (e.g. Figs. 3, S2). The correlation is restricted to the upper (or equivalently the lower) triangle of each RDM. In order to decide whether two RDMs are related, we perform statistical inference on the RDM correlation. The classical method for testing correlations assumes independent pairs of measurements for the two variables. For RDMs such independence cannot be assumed, because each dissimilarity is dependent on two response patterns, each of which also codetermines the dissimilarities of all its other pairings in the RDM. We therefore test the relatedness of RDMs by randomization (e.g. Nichols and Holmes 2002). In particular, we use randomization of the condition labels to rearrange the rows and columns of an RDM. We choose a random permutation of the conditions (i.e. of the 92 stimuli), reorder rows and columns of one of the two RDMs to be compared according to this permutation, and compute the correlation. By repeating this step 10,000 times, we obtain a distribution of correlations simulating the null hypothesis that the two RDMs are unrelated. If the actual correlation (i.e. the one for consistent labeling between the two RDMs) falls within the top 5% of the simulated null distribution of correlations, we reject the null hypothesis of unrelated RDMs. More generally, we estimate the p value as the percent rank/100 of the actual correlation in the randomization distribution. The percent rank is conservatively estimated, such that $p<0.0001$ indicates that the actual correlation was higher than any of the 10,000 correlations obtained after randomization of the condition labels.

[1] Alternatively, two separate data sets can be used, such that the vertical dimension of the RDM indexes pattern estimates from data set 1 and the horizontal dimension indexes pattern estimates from data set 2. The diagonal then contains dissimilarity estimates for replications of the same condition. Moreover, for each pair of conditions, two entries symmetrical about the diagonal of the RDM contain separate estimates of the pattern dissimilarity. We use the single-pattern-set approach here, because it provides more data along time for each fMRI pattern estimate. The overlapping hemodynamic responses to the stimuli are more precisely estimated when more runs are available not only by a factor of sqrt(n), where n is the number of runs, but by a larger factor, because longer random sequences more closely approximate the ideal of uncorrelated hemodynamic-response predictors.

The condition-label randomization test is justified by the random assignment of conditions (i.e. stimuli) to experimental trials. Under the null hypothesis of no relation between the RDMs, the conditions are exchangeable, i.e. the true labeling and each random relabeling of one RDM yield correlations drawn from the same distribution (i.e. from the null distribution). More generally, condition-label randomization can be used to test various RDM statistics against the null hypothesis that the condition labels are interchangeable (in the sense of not affecting the true test statistic). Note that the central results of this paper all rely on the condition-label randomization test (Figs. 3, 5, 6, S3a), but condition-bootstrap resampling has been used to test additional hypotheses (Figs. 5, S4).

Testing statistics of representational dissimilarity matrices by bootstrap resampling of the set of conditions. Not all hypotheses about RDMs can be tested by randomization. For example, the randomization test cannot be used to assess whether the mean of an RDM is greater than some constant, because the mean will be the same for each relabeling. Moreover, by condition-label randomization we test the null hypothesis that the condition labels are interchangeable. This may not be the desired null hypothesis. A less rigorous, but more versatile approach is bootstrap resampling (Efron and Tibshirani, 1993), which we apply here to the set of experimental conditions (i.e. the stimuli), in order to simulate a distribution of RDMs. Like the randomization test, the bootstrap test is appropriate for RDMs in that it does not rely on either distributional assumptions or the assumption of independence of the dissimilarity estimates. As mentioned above, tests assuming independent data may not be valid for RDMs, because dissimilarities within an RDM have a complex dependency structure. Like the condition-label randomization described above, the bootstrap resampling here operates at the level of the experimental conditions and, thus, simulates the dependency structure of RDMs.

The condition bootstrap test proceeds by resampling the set of conditions with replacement: If there are n_c condition labels, we draw n_c times from the whole set, to obtain a set of n_c labels. The bootstrap set of condition labels may include multiple instances of some labels and exclude others altogether. We construct a bootstrap RDM by resampling the original RDM according to the bootstrap sample of condition labels. We then compute the statistic of interest (e.g. the mean of all dissimilarities). We repeat this process many times (e.g. 10,000 times) to obtain a bootstrap distribution for the statistic.

The bootstrap resampling can be stratified in order to compare sets of conditions. For example, in order to test whether between-category dissimilarities are greater than within-category dissimilarities (Fig. 5), we separately bootstrap resampled the animates set and the inanimates set, recomputing mean B of between-category dissimilarities and the mean W of within-category dissimilarities to obtain the test statistic B minus W.

The bootstrap distribution of the statistic allows us to obtain error bars on arbitrary RDM statistics: the standard deviation of the bootstrap distribution is the standard error of the estimate of the statistic. In addition, we can define a 95% confidence interval by excluding 5% of the extreme values in the distribution (either on one side for a one-sided test or on both sides symmetrically for a two-sided test). If the value assigned to the test statistic by the null hypothesis falls outside the confidence interval, the null hypothesis is rejected. This is a valid test, because the confidence interval will include the null value with 95% probability given that the null hypothesis is true and assuming that the bootstrap resampling is an accurate simulation. The latter assumption is questionable, therefore the bootstrap procedures we describe here should be considered rough, approximate methods of inference.

A further complication of this method is that the bootstrap resampling of the set of condition labels moves zeros from the diagonal into the off-diagonal parts of the RDM whenever a condition is selected multiple times in the bootstrap resampling. (For 96 conditions, this is a small proportion of the entries: on the order of 1%.) In order to prevent these zeros from biasing the statistic, we exclude them before computing the statistic. For the purpose of condition bootstrapping, it may be preferable to use two data sets for computing the RDM (as suggested above), such that each diagonal value reflects a dissimilarity between two replications of the same response pattern.

The rationale for bootstrap resampling of a set of experimental conditions is to simulate the distribution of the statistic of interest that we expect to obtain for repetitions of the experiment performed with the same subjects but with different experimental conditions (e.g. stimuli) drawn from the same population of possible conditions that could have been used for the experiment (e.g. stimuli from the same categories). An interesting feature of this approach is its potential to generalize from the set of conditions actually used in the experiment to the hypothetical population of conditions, of which the actually chosen conditions can be considered a random sample. Note, however, that a sufficient number of conditions is required for this and the accuracy of the simulation is not guaranteed. For caveats and advanced bootstrap methods, see Hesterberg (2007).

Human localizer experiments and definition of regions of interest

Definition of regions of interest. All regions of interest (ROIs) were defined on the basis of independent experimental data and restricted to a cortex mask manually drawn on each subject's fMRI slices. Human IT was defined by selecting a variable number of voxels (316 voxels in Figs. 1-6; 100-10,000 voxels in Fig. S11) within the inferior temporal portion of the bilateral cortex mask according to their visual responsiveness. Visual responsiveness was assessed using the t map for the average response to the 96 object images. The t map was computed on the basis of one third of the runs of the main experiment within each session. The remaining runs were used to perform all further analyses. To define early visual cortex, we selected the most visually responsive voxels, as for IT, but within a manually defined anatomical region around the calcarine sulcus within the cortex mask (Figs. 5, 6, S5). For control analyses (Fig. S11), we defined the FFA (Kanwisher et al., 1997) using the contrast faces minus objects, and the PPA (Epstein and Kanwisher, 1998) using the contrast places minus objects in analyzing the separate localizer block-design experiment described below.

Localizer block-design experiment. We performed a functional localizer experiment using the same fMRI sequence as for the main experiment and a separate set of stimuli. Subjects viewed grayscale photos of faces, places, and objects (spanning a visual angle of about 5.7°) in category blocks. Each block lasted 30 s (stimulus-onset asynchrony: 1 s; stimulus duration: 700 ms), alternating with 20-s fixation blocks. Three blocks were presented for each stimulus category (face, place, object), resulting in a total run duration of 7 min and 50 s. Stimuli were presented on a constantly visible uniform black background. Subjects continually fixated a central white cross and performed a one-back repetition-detection task on the images, responding with a left-thumb button press for each consecutive repetition (3 to 5 repetitions per block). Each stimulus was only presented once, except for the immediate repetitions to be detected in the one-back task. Stimuli were centered with respect to the fixation cross.

Model representations

We processed our stimuli to obtain their representations in a number of low-level models. We then analyzed these model representations (Figs. S6, S7) in the same way as the brain-activity data from early visual cortex and IT (Figs. 1, 2, 4, S5, S9-S11). Each image was converted to a representational vector as described below for each model. As for the brain-activity data, each representational vector was then compared to each other representational vector by means of 1-r as the dissimilarity measure (where r is the Pearson linear correlation; only for the S-CIELAB representation, the conventional Delta E measure was used instead of 1-r) to obtain a representational dissimilarity matrix, on which further analyses (Figs. S6, S7) were based.

Color image (CIELAB). The RGB color images (175×175 pixels) were converted to the CIELAB color space, which approximates a linear representation of human perceptual color space. Each CIELAB image was then converted to a pixel vector (175×175×3 numbers).

Low-resolution color image (28×28 pixels, CIELAB). The RGB color images (175×175 pixels) were downsampled to 28×28 pixels (with bicubic interpolation) and subsequently converted to the CIELAB color space. Each 28×28 CIELAB image was then converted to a pixel vector (28×28×3 numbers).

Grayscale image. The RGB color images (175×175 pixels) were converted to grayscale. Each grayscale image was then converted to a pixel vector (175×175 numbers).

Low-resolution grayscale image (28×28 pixels). The RGB color images (175×175 pixels) were converted to grayscale and subsequently downsampled to 28×28 pixels (with bicubic interpolation). Each grayscale image was then converted to a pixel vector (28×28 numbers).

Binary silhouette image. The RGB color images (175×175 pixels) were converted to binary silhouette images, in which all background pixels had the value 0 and all figure pixels had the value 1. Each binary silhouette image was then converted to a pixel vector (175×175 binary numbers).

CIELAB joint histogram (6×6×6 bins). The RGB color images (175×175 pixels) were converted to the CIELAB color space. The three CIELAB dimensions (L, a, b), were then divided into 6 bins of equal width. The joint CIELAB histogram was computed by counting the number of figure pixels (gray background left out) falling into each of the 6×6×6 bins. The joint histogram was converted to a vector (6×6×6 numbers).

S-CIELAB (Delta E). The RGB color images (175×175 pixels, 2.9° visual angle) were compared (each to each other, to obtain a representational dissimilarity matrix) by means of the S-CIELAB Delta-E dissimilarity measure (Zhang and Wandell, 1997), which models the perceptual similarity of color images and, unlike CIELAB Delta E, accounts for pattern-color sensitivity results (Poirson and Wandell, 1993) by separating the image into components corresponding to different spatial-frequency bands.

V1 model. The RGB color images (175×175 pixels, 2.9° visual angle) were converted to grayscale and given as input to population of modeled V1 simple and complex cells (Lampl et al., 2004; Riesenhuber and Poggio, 2002; Kiani et al., 2007). The receptive fields (RFs) of simple cells were simulated by Gabor filters of 4 different orientations (0°, 90°, -45° and 45°) and 12 sizes (7-29 pixels). Cell RFs were distributed over the stimulus image at 0.017° intervals in a cartesian grid (for each image pixel there was a simple and a complex cell of each selectivity that had its RF centered on that pixel). Negative values in outputs were rectified to zero. The RFs of complex cells were modeled by the MAX operation performed on outputs of neighboring simple cells with similar orientation selectivity. The MAX operation consists in selecting the strongest (maximum) input to determine the output. This renders the output of a complex cell invariant to the precise location of the stimulus feature that drives it. Simple cells were divided into four groups based on their RF size (7-9 pixels, 11-15 pixels, 17-21 pixels, 23-29 pixels) and each complex cell pooled responses of neighboring simple cells in one of these groups. The spatial range of pooling varied across the four groups (4×4, 6×6, 9×9, and 12×12 pixels for the four groups, respectively). This yielded 4 (orientation selectivities) × 12 (RF sizes) = 48 simple-cell maps and 4 (orientation selectivities) × 4 (sets of simple-cell RF sizes pooled) = 16 complex-cell maps of 175×175 pixels. All maps of simple and complex cell outputs were vectorized and concatenated to obtain a representational vector for each stimulus image.

HMAX-C2 model based on natural image fragments. This model representation developed by Serre et al. (2005) builds on the complex-cell outputs of the V1 model described above (implemented by the same group). The C2 features used in the analysis (Fig. S7) may be comparable to those found in primate V4 and posterior IT. The model has four sequential stages: S1-C1-S2-C2. The first two stages correspond to the simple and complex cells described above, respectively. Stages S2 and C2 use the same pooling mechanisms as stages S1 and C1, respectively. Each unit in stage S2 locally pools information from the C1 stage by a linear filter and behaves as a radial basis function, responding most strongly to a particular prototype input pattern. The prototypes correspond to random fragments extracted from a set of natural images (stimuli independent of those used in the present study). S2 outputs are locally pooled by C2 units utilizing the MAX operation for a degree of position and scale tolerance. A detailed description of the model (including the parameter settings and map sizes we used here) can be found in Serre et al. (2005). The model, including the natural image fragments, was downloaded from the author's website in January 2007 (for the current version, see http://cbcl.mit.edu/software-datasets/standardmodel/index.html).

Additional model representations. In addition to the models described above and analyzed for our stimulus set in Figs. S6 and S7, we tried a number of additional models (not shown). These included (1) low-passed and high-passed grayscale representations, (2) a version of the V1 model described above, in which we averaged all simple and complex cell responses representing the same retinal location (averaging also across orientation selectivities and RF sizes) in order to mimic the effect of downsampling by population averaging within fMRI voxels, and (3) higher-level shape-tuned units created within the HMAX model framework (Riesenhuber and Poggio, 2002) as described in Kiani et al. (2007). None of these model representations exhibited categorical clustering.

References

Afraz, S. R., Kiani, R., and Esteky, H. (2006). Microstimulation of inferotemporal cortex influences face categorization. Nature *442*(7103), 692-695.

Aguirre, G. K., Zarahn, E., and D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. Neuron *21,* 373-383.

Baker, C. I., Behrmann, M., and Olson, C. R. (2002). Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. Nat Neurosci *5*(11), 1210-6.

Boynton, G. M., Engel, S. A., Glover, G. H., and Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. J. Neurosci. *16*, 4207-4221.

Capitani, E., Laiacona, M., Mahon, B., and Caramazza, A. (2003). What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. Cogn. Neuropsychol. *20*, 213-261.

Carlson, T. A., Schrater, P., and He, S. (2003). Patterns of activity in the categorical representations of objects. J Cogn Neurosci *15*(5), 704-17.

Cox, D. D., and Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage *19*, 261-270.

Downing, P. E., Chan, A. W.-Y., Peelen, M. V., Dodds, C. M., and Kanwisher, N. (2006). Domain specificity in visual cortex. Cereb. Cortex *16*(10), 1453-1461.

Downing, P. E., Jiang, Y., Shuman, M., and Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. Science *293*, 2470-2473.

Edelman, S., Grill-Spector, K., Kushnir, T., and Malach, R. (1998). Towards direct visualization of the internal shape space by fMRI. Psychobiology *26*, 309-321.

Efron, B. and Tibshirani, R. J. (1993). An Introduction to the Bootstrap. Chapman and Hall.

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

Freedman, D. J., Riesenhuber, M., Poggio, T., and Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. J. Neurosci. *23*(12), 5235–5246.

Freedman, D. J., Riesenhuber, M., Poggio, T., and Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. Science *291*, 312-316.

Freedman, D. J. and Assad, J. A. (2006). Experience-dependent representation of visual categories in parietal cortex. Nature. *443*(7107), 85-88.

Haxby, J. V. et al. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science *293*, 2425-2430.

Hesterberg, T. C. (2007) Bootstrap, http://home.comcast.net/~timhesterberg/articles/tech-encyclopedia.pdf) under review.

Humphreys, G. W., and Forde, E. M. E. (2001). Hierarchies, similarity, and interactivity in object recognition: “Category-specific” neuropsychological deficits. Behav. Brain Sci. *24*, 453-509.

Hung, C. P., Kreiman, G., Poggio, T., and DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. Science *310*, 863-866.

Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. *17*, 4302-4311.

Kiani, R., Esteky, H., and Tanaka, K. (2005). Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. J. Neurophysiol. *94*, 1587-1596.

Kiani, R., Esteky, H., Mirpour, K., and Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. J. Neurophysiol. *97*, 4296-4309.

Kreiman, G., Koch, C., and Fried, I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. Nat. Neurosci. *3*, 946-953.

Lampl, I., Ferster, D., Poggio, T., and Riesenhuber, M. (2004). Intracellular measurements of spatial integration and the MAX operation in complex cells of the cat primary visual cortex. J. Neurophysiol. *92*, 2704–2713.

Martin, A. (2007). The representations of object concepts in the brain. Annu. Rev. Psychol. *58*, 25-45.

Martin, A., Wiggs, C. L., Ungerleider, L. G., and Haxby, J. V. (1996). Neural correlates of category-specific knowledge. Nature *379*(6566), 649-52.

Nichols, T. E. and Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp. *15*(1), 1-25.

Poirson, A. B., and Wandell, B. A. (1993). Appearance of colored patterns: pattern-color separability. J. Opt. Soc. Am. A. *10*, 2458-2470.

Prüssmann, K. P. (2004). Parallel imaging at high field strength: Synergies and joint potential. Top Magn. Reson. Imaging *15*, 237-244.

Puce, A., Allison, T., Gore, J. C., and McCarthy, G. (1995). Face-sensitive regions in human extrastriate cortex studied by functional MRI. J. Neurophysiol. *74*, 1192-1199.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., and Fried I. (2005). Invariant visual representation by single neurons in the human brain. Nature. *435*(7045), 1102-7.

Riesenhuber, M., and Poggio, T. (2002). Neural mechanisms of object recognition. Curr. Opin. Neurobiol. *12*,162-168.

Serre, T., Oliva, A., and Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. Proc. Natl. Acad. Sci. U.S.A. *104*, 6424-6429.

Serre, T., Wolf, L., and Poggio, T. (2005). Object recognition with features inspired by visual cortex. In: Computer Vision and Pattern Recognition (CVPR 2005), San Diego, USA, June 2005.

Sigala, N., and Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. Nature *415*, 318-320.

Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., and Tootell R. B. H. (2003). Faces and objects in macaque cerebral cortex. Nat. Neurosci. *6*, 989-995.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H., and Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. Science *311*, 670-674.

Vogels, R. (1999). Categorization of complex visual images by rhesus monkeys. Part 2: single-cell study. Eur. J. Neurosci. *11*, 1239-1255.

Zhang, X., and Wandell, B. A. (1997). A spatial extension of CIELAB for digital color image reproduction. SID Journal.