A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain

Alexander G. Huth,[1] Shinji Nishimoto,[1] An T. Vu,[2] and Jack L. Gallant[1,2,3,*]
[1]Helen Wills Neuroscience Institute
[2]Program in Bioengineering
[3]Department of Psychology
University of California, Berkeley, Berkeley, CA 94720, USA
*Correspondence: gallant@berkeley.edu

http://dx.doi.org/10.1016/j.neuron.2012.10.014

SUMMARY

Humans can see and name thousands of distinct object and action categories, so it is unlikely that each category is represented in a distinct brain area. A more efficient scheme would be to represent categories as locations in a continuous semantic space mapped smoothly across the cortical surface. To search for such a space, we used fMRI to measure human brain activity evoked by natural movies. We then used voxelwise models to examine the cortical representation of 1,705 object and action categories. The first few dimensions of the underlying semantic space were recovered from the fit models by principal components analysis. Projection of the recovered semantic space onto cortical flat maps shows that semantic selectivity is organized into smooth gradients that cover much of visual and nonvisual cortex. Furthermore, both the recovered semantic space and the cortical organization of the space are shared across different individuals.

INTRODUCTION

Previous fMRI studies have suggested that some categories of objects and actions are represented in specific cortical areas. Categories that have been functionally localized include faces (Avidan et al., 2005; Clark et al., 1996; Halgren et al., 1999; Kanwisher et al., 1997; McCarthy et al., 1997; Rajimehr et al., 2009; Tsao et al., 2008), body parts (Downing et al., 2001; Peelen and Downing, 2005; Schwarzlose et al., 2005), outdoor scenes (Aguirre et al., 1998; Epstein and Kanwisher, 1998), and human body movements (Peelen et al., 2006; Pelphrey et al., 2005). However, humans can recognize thousands of different categories of objects and actions. Given the limited size of the human brain, it is unreasonable to expect that every one of these categories is represented in a distinct brain area. Indeed, fMRI studies have failed to identify dedicated functional areas for many common object categories including household objects (Haxby et al., 2001), animals and tools (Chao et al., 1999), food, clothes, and so on (Downing et al., 2006).

An efficient way for the brain to represent object and action categories would be to organize them into a continuous space that reflects the semantic similarity between categories. A continuous semantic space could be mapped smoothly onto the cortical sheet so that nearby points in cortex would represent semantically similar categories. No previous study has found a general semantic space that organizes the representation of all visual categories in the human brain. However, several studies have suggested that single locations on the cortical surface might represent many semantically related categories (Connolly et al., 2012; Downing et al., 2006; Edelman et al., 1998; Just et al., 2010; Konkle and Oliva, 2012; Kriegeskorte et al., 2008; Naselaris et al., 2009; Op de Beeck et al., 2008; O'Toole et al., 2005). Some studies have also proposed likely dimensions that organize these representations, such as animals versus nonanimals (Connolly et al., 2012; Downing et al., 2006; Kriegeskorte et al., 2008; Naselaris et al., 2009), manipulation versus shelter versus eating (Just et al., 2010), large versus small (Konkle and Oliva, 2012), or hand- versus mouth- versus foot-related actions (Hauk et al., 2004).

To determine whether a continuous semantic space underlies category representation in the human brain, we collected blood-oxygen-level-dependent (BOLD) fMRI responses from five subjects while they watched several hours of natural movies. Natural movies were used because they contain many of the object and action categories that occur in daily life, and they evoke robust BOLD responses (Bartels and Zeki, 2004; Hasson et al., 2004, 2008; Nishimoto et al., 2011). After data collection, we used terms from the WordNet lexicon (Miller, 1995) to label 1,364 common objects (i.e., nouns) and actions (i.e., verbs) in the movies (see Experimental Procedures for details of labeling procedure and see Figure S1 available online for examples of typical labeled clips). WordNet is a set of directed graphs that represent the hierarchical "is a" relationships between object or action categories. The hierarchical relationships in WordNet were then used to infer the presence of an additional 341 higher-order categories (e.g., a scene containing a dog must also contain a canine). Finally, we used regularized linear regression (see Experimental Procedures for details; Kay et al., 2008; Mitchell et al., 2008; Naselaris et al., 2009; Nishimoto et al.,

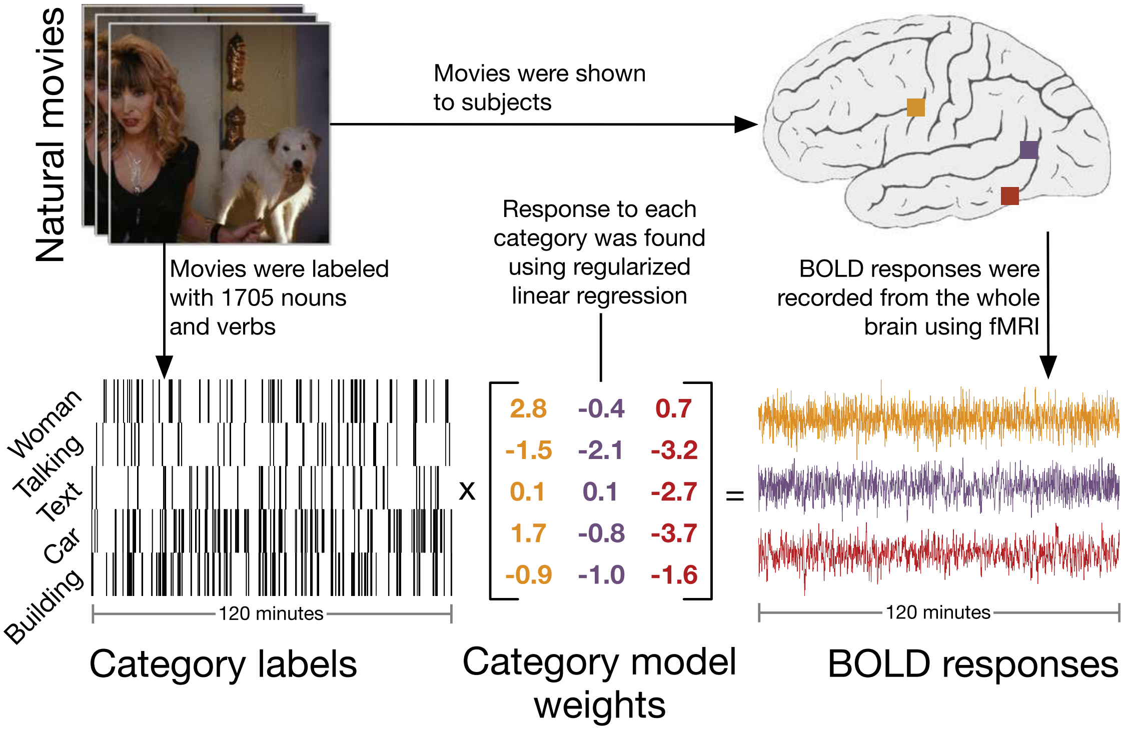

Natural movies
Movies were shown to subjects
Movies were labeled with 1705 nouns and verbs
Response to each category was found using regularized linear regression
BOLD responses were recorded from the whole brain using fMRI
Woman
Talking
Text
Car
Building
120 minutes
x
2.8 -0.4 0.7
-1.5 -2.1 -3.2
0.1 0.1 -2.7
1.7 -0.8 -3.7
-0.9 -1.0 -1.6
=
120 minutes
Category labels
Category model weights
BOLD responses

Figure 1. Schematic of the Experiment and Model

Subjects viewed 2 hr of natural movies while BOLD responses were measured using fMRI. Objects and actions in the movies were labeled using 1,364 terms from the WordNet lexicon (Miller, 1995). The hierarchical "is a" relationships defined by WordNet were used to infer the presence of 341 higher-order categories, providing a total of 1,705 distinct category labels. A regularized, linearized finite impulse response regression model was then estimated for each cortical voxel recorded in each subject's brain (Kay et al., 2008; Mitchell et al., 2008; Naselaris et al., 2009; Nishimoto et al., 2011). The resulting category model weights describe how various object and action categories influence BOLD signals recorded in each voxel. Categories with positive weights tend to increase BOLD, while those with negative weights tend to decrease BOLD. The response of a voxel to a particular scene is predicted as the sum of the weights for all categories in that scene.

2011) to characterize the response of each voxel to each of the 1,705 object and action categories (Figure 1). The linear regression procedure produced a set of 1,705 model weights for each individual voxel, reflecting how each object and action category influences BOLD responses in each voxel.

RESULTS

Category Selectivity for Individual Voxels

Our modeling procedure produces detailed information about the representation of categories in each individual voxel in the brain. Figure 2A shows the category selectivity for one voxel located in the left parahippocampal place area (PPA) of subject A.V. The model for this voxel shows that BOLD responses are strongly enhanced by categories associated with man-made objects and structures (e.g., "building," "road," "vehicle," and "furniture"), weakly enhanced by categories associated with outdoor scenes (e.g., "hill," "grassland," and "geological formation") and humans (e.g., "person" and "athlete"), and weakly suppressed by nonhuman biological categories (e.g., "body parts" and "birds"). This result is consistent with previous reports that PPA most strongly represents information about outdoor scenes and buildings (Epstein and Kanwisher, 1998).

Figure 2B shows category selectivity for a second voxel located in the right precuneus (PrCu) of subject A.V. The model shows that BOLD responses are strongly enhanced by categories associated with social settings (e.g., people, communication verbs, and rooms) and suppressed by many other categories (e.g., "building," "city," "geological formation," and "atmospheric phenomenon"). This result is consistent with an earlier finding that PrCu is involved in processing social scenes (Iacoboni et al., 2004).

A Semantic Space for Representation of Object and Action Categories

We used principal components analysis (PCA) to recover a semantic space from the category model weights in each subject. PCA ensures that categories that are represented by similar sets of cortical voxels will project to nearby points in the estimated semantic space, while categories that are represented very differently will project to different points in the space. To maximize the quality of the estimated space, we included only voxels that were significantly predicted ($p < 0.05$, uncorrected) by the category model (see Experimental Procedures for details).

Because humans can perceive thousands of categories of objects and actions, the true semantic space underlying

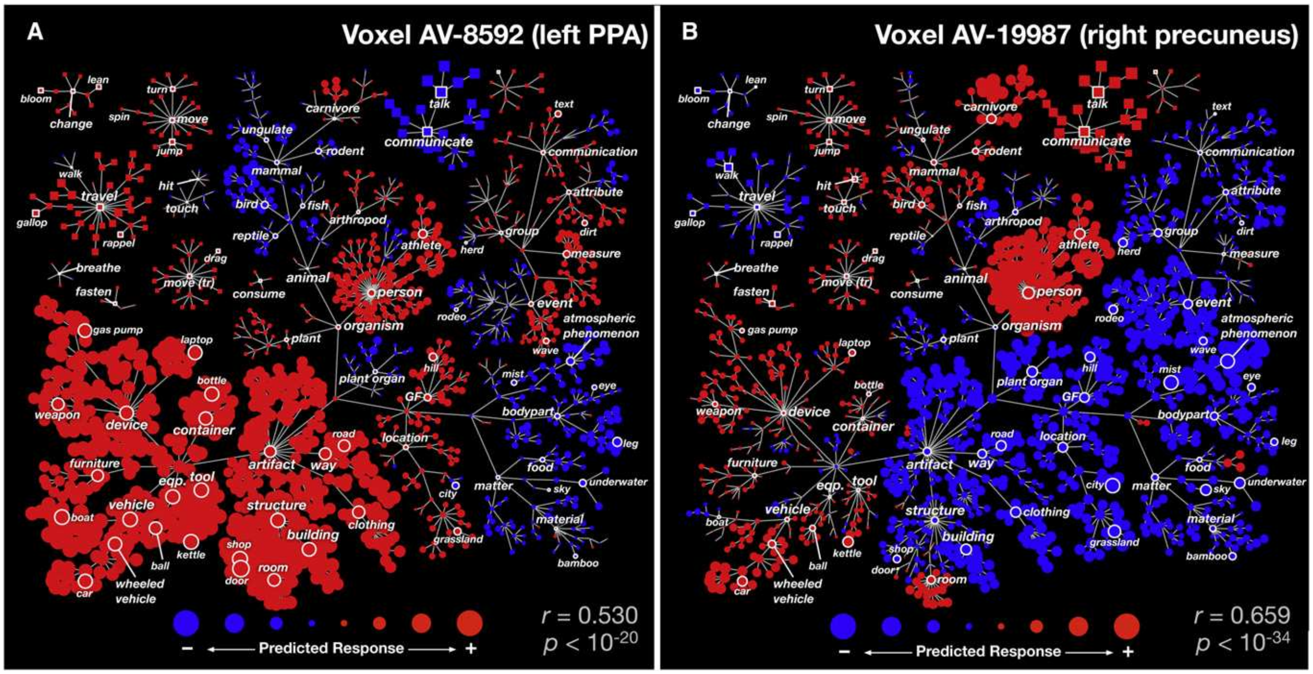


Figure 2. Category Selectivity for Two Individual Voxels

Each panel shows the predicted response of one voxel to each of the 1,705 categories, organized according to the graphical structure of WordNet. Links indicate "is a" relationships (e.g., an athlete is a person); some relationships used in the model are omitted for clarity. Each marker represents a single noun (circle) or verb (square). Red markers indicate positive predicted responses and blue markers indicate negative predicted responses. The area of each marker indicates predicted response magnitude. The prediction accuracy of each voxel model, computed as the correlation coefficient (*r*) between predicted and actual responses, is shown in the bottom right of each panel along with model significance (see Results for details).

(A) Category selectivity for one voxel located in the left hemisphere parahippocampal place area (PPA). The category model predicts that movies will evoke positive responses when "structures," "buildings," "roads," "containers," "devices," and "vehicles" are present. Thus, this voxel appears to be selective for scenes that contain man-made objects and structures (Epstein and Kanwisher, 1998).

(B) Category selectivity for one voxel located in the right hemisphere precuneus (PrCu). The category model predicts that movies will evoke positive responses from this voxel when "people," "carnivores," "communication verbs," "rooms," or "vehicles" are present and negative responses when movies contain "atmospheric phenomena," "locations," "buildings," or "roads." Thus, this voxel appears to be selective for scenes that contain people or animals interacting socially (Iacoboni et al., 2004).

category representation in the brain probably has many dimensions. However, given the limitations of fMRI and a finite stimulus set, we expect that we will only be able to recover the first few dimensions of the semantic space for each individual brain and fewer still dimensions that are shared across individuals. Thus, of the 1,705 semantic PCs produced by PCA on the voxel weights, only the first few will resemble the true underlying semantic space, while the remainder will be determined mostly by the statistics of the stimulus set and noise in the fMRI data.

To determine which PCs are significantly different from chance, we compared the semantic PCs to the PCs of the category stimulus matrix (see Experimental Procedures for details of why the stimulus PCs are an appropriate null hypothesis). First, we tested the significance of each subject's own category model weight PCs. If there is a semantic space underlying category representation in the subject's brain, then we should find that some of the subject's model weight PCs explain more of the variance in the subject's category model weights than is explained by the stimulus PCs. However, if there is no semantic space underlying category representation in the subject's brain, then the stimulus PCs should explain the same amount of variance in the category model weights as do the subject's PCs. The results of this analysis are shown in Figure 3. Six to eight PCs from individual subjects explain significantly more variance in category model weights than do the stimulus PCs ($p < 0.001$, bootstrap test). These individual subject PCs explain a total of 30%–35% of the variance in category model weights. Thus, our fMRI data are sufficient to recover semantic spaces for individual subjects that consist of six to eight dimensions.

Second, we used the same procedure to test the significance of group PCs constructed using data combined across subjects. To avoid overfitting, we constructed a separate group semantic space for each subject using combined data from the other four subjects. If the subjects share a common semantic space, then some of the group PCs should explain more of the variance in the selected subject's category model weights than do the stimulus PCs. However, if the subjects do not share a common semantic space, then the stimulus PCs should explain the same amount of variance in the category model weights as do the group PCs. The results of this analysis are also shown in Figure 3. The first four group PCs explain significantly more variance ($p < 0.001$, bootstrap test) than do the stimulus PCs in four out of five subjects.

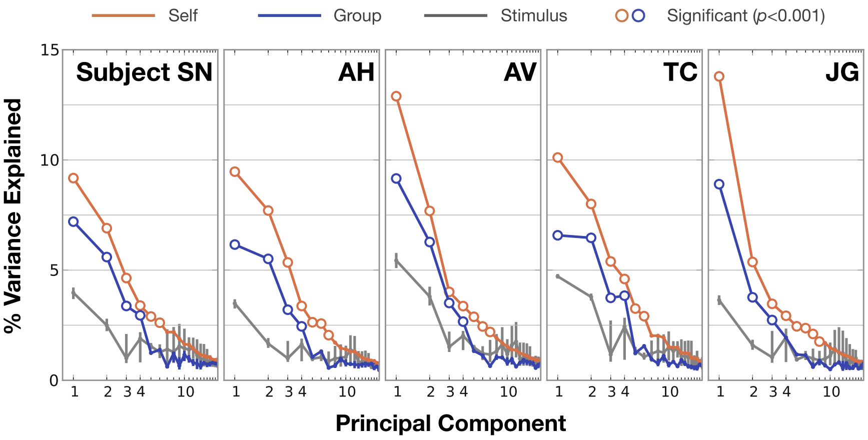


Figure 3. Amount of Model Variance Explained by Individual Subject and Group Semantic Spaces

Principal components analysis (PCA) was used to recover a semantic space from category model weights in each subject. Here we show the variance explained in the category model weights by each of the 20 most important PCs. Orange lines show the amount of variance explained in category model weights by each subject's own PCs and blue lines show the variance explained by PCs of combined data from other subjects. Gray lines show the variance explained by the stimulus PCs, which serve as an appropriate null hypothesis (see text and Experimental Procedures for details). Error bars indicate 99% confidence intervals (the confidence intervals for the subjects' own PCs and group PCs are very small). Hollow markers indicate subject or group PCs that explain significantly more variance ($p < 0.001$, bootstrap test) than the stimulus PCs. The first four group PCs explain significantly more variance than the stimulus PCs for four subjects. Thus, the first four group PCs appear to comprise a semantic space that is common across most individuals and that cannot be explained by stimulus statistics. Furthermore, the first six to nine individual subject PCs explain significantly more variance than the stimulus PCs ($p < 0.001$, bootstrap test). This suggests that while the subjects share broad aspects of semantic representation, finer-scale semantic representations are subject specific.

These four group PCs explain on average 19% of the total variance, 72% as much as do the first four individual subject PCs. In contrast, the first four stimulus PCs only explain 10% of the total variance, 38% as much variance as the individual subject PCs. This result suggests that the first four group PCs describe a semantic space that is shared across individuals.

Third, we determined how much stimulus-related information is captured by the group PCs and full category model. For each model, we quantified stimulus-related information by testing whether the model could distinguish among BOLD responses to different movie segments (Kay et al., 2008; Nishimoto et al., 2011; see Experimental Procedures for details). Models using 4–512 group PCs were tested by projecting the category model weights for 2,000 voxels (selected using the training data set) onto the group PCs. Then, the projected model weights were used to predict responses to the validation stimuli. We then tried to match the validation stimuli to observed BOLD responses by comparing the observed and predicted responses. The same identification procedure was repeated for the full category model.

The results of this analysis are shown in Figure S2. The full category model correctly identifies an average of 76% of stimuli across subjects (chance is 1.9%). Models based on 64 or more group PCs correctly identify an average of 74% of the stimuli but incorporate information that we know cannot be distinguished from the stimulus PCs. A model based on the four significant group PCs correctly identifies 49% of the stimuli, roughly two-thirds as many as the full model. These results show that the four-PC group space does not capture all of the stimulus-related information present in the full category model, indicating that the true semantic space is likely to have more than four dimensions. Further experiments will be required to determine these other semantic dimensions.

To visualize the group semantic space, we formed a robust estimate by pooling data from all five subjects (for a total of 49,685 voxels) and then applying PCA to the combined data.

Visualization of the Semantic Space

The previous results demonstrate that object and action categories are represented in a semantic space consisting of at least four dimensions and that this space is shared across individuals. To understand the structure of the group semantic space, we visualized it in two different ways. First, we projected the 1,705 coefficients of each group PC onto the graph defined by WordNet (Figure 4). The first PC (shown in Figure 4A) appears to distinguish between categories that have high stimulus energy (e.g., moving objects like "person," "vehicle," and "animal") and those that have low stimulus energy (e.g., stationary objects like "sky," "city," "building," and "plant"). This is not surprising, as the first PC should reflect the stimulus dimension with the greatest influence on brain activity, and stimulus energy is already known to have a large effect on BOLD signals (Fox et al., 2009; Nishimoto et al., 2011; Smith et al., 1998).

We then visualized the second, third, and fourth group PCs simultaneously using a three-dimensional (3D) colormap projected onto the WordNet graph. A color was assigned to each of the 1,705 categories according to the following scheme: the category coefficient in the second PC determined the value of the red channel, the third PC determined the green channel, and the fourth PC determined the blue channel (see Figure 4B; see Figure S3 for individual PCs). This scheme assigns similar colors to categories that are represented similarly in the brain. Figure 4C shows the second, third, and fourth PCs projected onto the WordNet graph. Here humans, human body parts, and communication verbs (e.g., "gesticulate" and "talk") appear in shades of green. Other animals appear yellow and green-yellow. Nonliving objects such as "vehicles" appear pink and purple, as do movement verbs (e.g., "run"), outdoor categories (e.g., "hill," "city," and "grassland"), and paths (e.g., "road"). Indoor categories (e.g., "room," "door," and "furniture") appear in blue and indigo. This figure suggests that semantically related categories (e.g., "person" and "talking") are represented more similarly than unrelated categories (e.g., "talking" and "kettle").

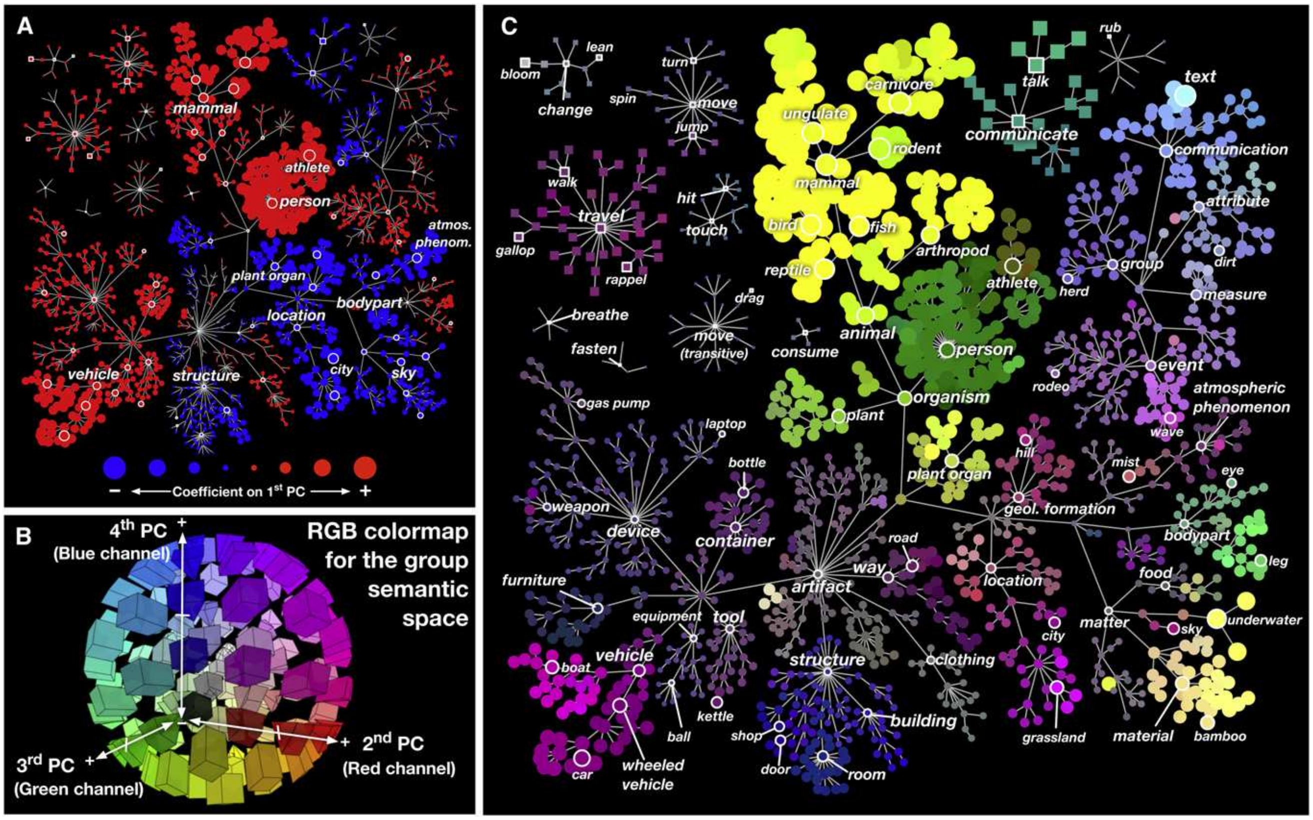


Figure 4. Graphical Visualization of the Group Semantic Space

(A) Coefficients of all 1,705 categories in the first group PC, organized according to the graphical structure of WordNet. Links indicate ''is a'' relationships (e.g., an athlete is a person); some relationships used in the model have been omitted for clarity. Each marker represents a single noun (circle) or verb (square). Red markers indicate positive coefficients and blue indicates negative coefficients. The area of each marker indicates the magnitude of the coefficient. This PC distinguishes between categories with high stimulus energy (e.g., moving objects like ''person'' and ''vehicle'') and those with low stimulus energy (e.g., stationary objects like ''sky'' and ''city'').

(B) The three-dimensional RGB colormap used to visualize PCs 2–4. The category coefficient in the second PC determined the value of the red channel, the third PC determined the green channel, and the fourth PC determined the blue channel. Under this scheme, categories that are represented similarly in the brain are assigned similar colors. Categories with zero coefficients appear neutral gray.

(C) Coefficients of all 1,705 categories in group PCs 2–4, organized according to the WordNet graph. The color of each marker is determined by the RGB colormap in (B). Marker sizes reflect the magnitude of the three-dimensional coefficient vector for each category. This graph shows that categories thought to be semantically related (e.g., ''athletes'' and ''walking'') are represented similarly in the brain.

To better understand the overall structure of the semantic space, we created an analogous figure in which category position is determined by the PCs instead of the WordNet graph. Figure 5 shows the location of all 1,705 categories in the space formed by the second, third, and fourth group PCs (Movie S1 shows the categories in 3D). Here, categories that are represented similarly in the brain are plotted at nearby positions. Categories that appear near the origin have small PC coefficients and thus are generally weakly represented or are represented similarly across voxels (e.g., ''laptop'' and ''clothing''). In contrast, categories that appear far from the origin have large PC coefficients and thus are represented strongly in some voxels and weakly in others (e.g., ''text,'' ''talk,'' ''man,'' ''car,'' ''animal,'' and ''underwater''). These results support earlier findings that categories such as faces (Avidan et al., 2005; Clark et al., 1996; Halgren et al., 1999; Kanwisher et al., 1997; McCarthy et al., 1997; Rajimehr et al., 2009; Tsao et al., 2008) and text (Cohen et al., 2000) are represented strongly and distinctly in the human brain.

Interpretation of the Semantic Space

Earlier studies have suggested that animal categories (including people) are represented distinctly from nonanimal categories (Connolly et al., 2012; Downing et al., 2006; Kriegeskorte et al., 2008; Naselaris et al., 2009). To determine whether hypothesized semantic dimensions such as animal versus nonanimal are captured by the group semantic space, we compared each of the group semantic PCs to nine hypothesized semantic dimensions. For each hypothesized dimension, we first assigned a value to each of the 1,705 categories. For example, for the dimension animal versus nonanimal, we assigned the value +1 to all animal categories and the value 0 to all nonanimal categories. Then we computed how much variance each hypothesized dimension explained in each of the group PCs. If a hypothesized dimension provides a good description of one of the group PCs, then that dimension will explain a large fraction of the variance in that PC. If a hypothesized dimension is captured by the group semantic space but does not line up exactly with one of the PCs, then that dimension will explain variance in multiple PCs.

The comparison between the group PCs and hypothesized semantic dimensions is shown in Figure 6. The first PC is best explained by a dimension that contrasts mobile categories (people, nonhuman animals, and vehicles) with nonmobile categories. The first PC is also well explained by a dimension that is an extension of a previously reported ‘‘animacy’’ continuum (Connolly et al., 2012). Our animacy dimension assigns the highest weight to people, decreasing weights to other mammals, birds, reptiles, fish, and invertebrates, and zero weight to all nonanimal categories. The second PC is best explained by a dimension that contrasts categories associated with social interaction (people and communication verbs) with all other categories. The third PC is best explained by a dimension that contrasts categories associated with civilization (people, man-made objects, and vehicles) with categories associated with nature (nonhuman animals). The fourth PC is best explained by a dimension that contrasts biological categories (animals, plants, people, and body parts) with nonbiological categories, as well as a similar dimension that contrasts animal categories (including people) with nonanimal categories. These results provide quantitative interpretations for the group PCs and show that many hypothesized semantic dimensions are captured by the group semantic space.

The results shown in Figure 6 also suggest that some hypothesized semantic dimensions are not captured by the group semantic space. The contrast between place categories (buildings, roads, outdoor locations, and geological features) and nonplace categories is not captured by any group PC. This is surprising because the representation of place categories is thought to be of primary importance to many brain areas, including the PPA (Epstein and Kanwisher, 1998), retrosplenial cortex (RSC; Aguirre et al., 1998), and temporo-occipital sulcus (TOS; Nakamura et al., 2000; Hasson et al., 2004). Our results may appear different from the results of earlier studies of place representation because those earlier studies used static images and not movies.

Another hypothesized semantic dimension that is not captured by our group semantic space is real-world object size (Konkle and Oliva, 2012). The object size dimension assigns a high weight to large objects (e.g., ‘‘boat’’), medium weight to human-scale objects (e.g., ‘‘person’’), a small weight to small

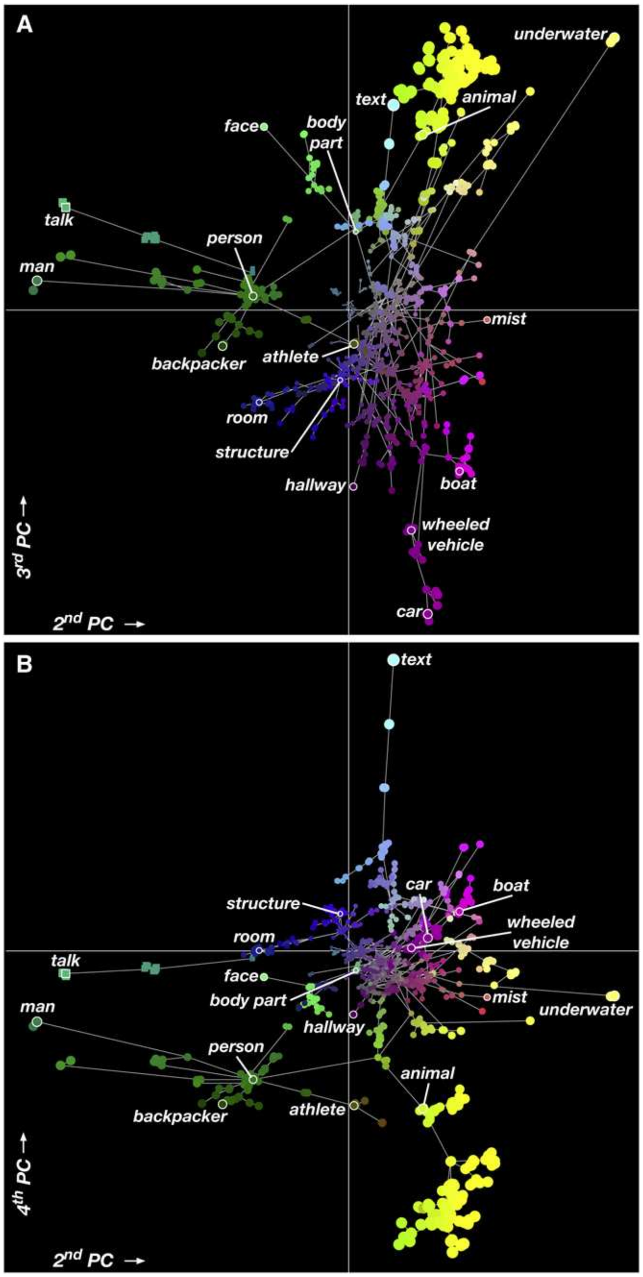


Figure 5. Spatial Visualization of the Group Semantic Space

(A) All 1,705 categories, organized by their coefficients on the second and third PCs. Links indicate ‘‘is a’’ relationships (e.g., an athlete is a person) from the WordNet graph; some relationships used in the model have been omitted for clarity. Each marker represents a single noun (circle) or verb (square). The color of each marker is determined by an RGB colormap based on the category coefficients in PCs 2–4 (see Figure 4B for details). The position of each marker is also determined by the PC coefficients: position on the x axis is determined by the coefficient on the second PC and position on the y axis is determined by the coefficient on the third PC. This ensures that categories that are represented similarly in the brain appear near each other. The area of each marker indicates the magnitude of the PC coefficients for that category; more important or strongly represented categories have larger coefficients. The categories ‘‘man,’’ ‘‘talk,’’ ‘‘text,’’ ‘‘underwater,’’ and ‘‘car’’ have the largest coefficients on these PCs.

(B) All 1,705 categories, organized by their coefficients on the second and fourth PCs. Format is the same as (A). The large group of ‘‘animal’’ categories has large PC coefficients and is mainly distinguished by the fourth PC. Human categories appear to span a continuum. The category ‘‘person’’ is very close to indoor categories such as ‘‘room’’ on the second and third PCs but different on the fourth. The category ‘‘athlete’’ is close to vehicle categories on the second and third PCs but is also close to ‘‘animal’’ on the fourth PC. These semantically related categories are represented similarly in the brain, supporting the hypothesis of a smooth semantic space. However, these results also show that some categories (e.g., ‘‘talk,’’ ‘‘man,’’ ‘‘text,’’ and ‘‘car’’) appear to be more important than others. Movie S1 shows this semantic space in 3D.

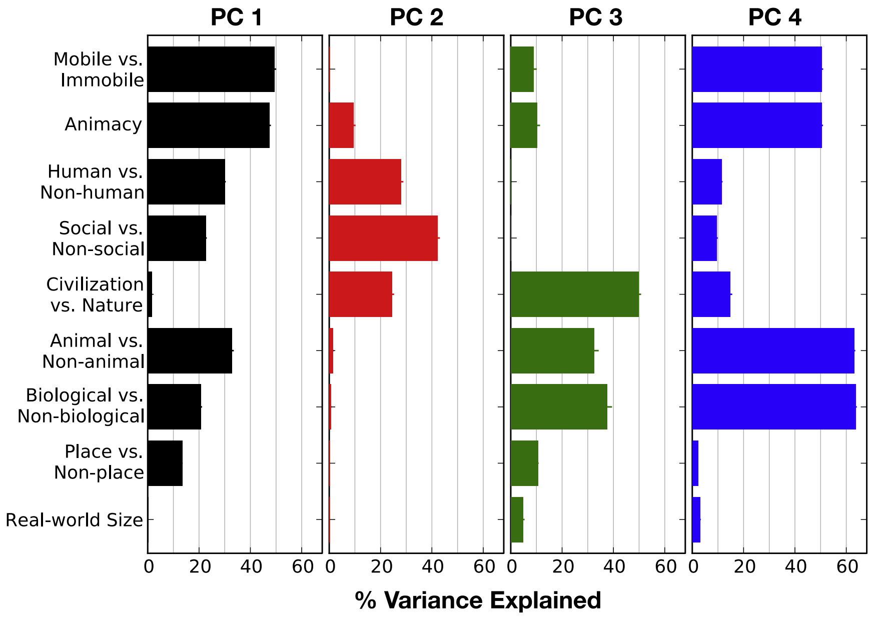


Figure 6. Comparison between the Group Semantic Space and Nine Hypothesized Semantic Dimensions

For each hypothesized semantic dimension, we assigned a value to each of the 1,705 categories (see Experimental Procedures for details) and we computed the fraction of variance that each dimension explains in each PC. Each panel shows the variance explained by all hypothesized dimensions in one of the four group PCs. Error bars indicate bootstrap SE. The first PC is best explained by a dimension that contrasts mobile categories (people, nonhuman animals, and vehicles) with nonmobile categories and an "animacy" dimension (Connolly et al., 2012) that assigns high weight to humans, decreasing weights to other mammals, birds, reptiles, fish, and invertebrates, and zero weight to other categories. The second PC is best explained by a dimension that contrasts social categories (people and communication verbs) with all other categories. The third PC is best explained by a dimension that contrasts categories associated with civilization (people, man-made objects, and vehicles) with categories associated with nature (nonhuman animals). The fourth PC is best explained by a dimension that contrasts biological categories (people, animals, plants, body parts, and plant parts) with nonbiological categories and a dimension that contrasts animals (people and nonhuman animals) with nonanimals.

objects (e.g., "glasses"), and zero weight to objects that have no size (e.g., "talking") or can be many sizes (e.g., "animal"). This object size dimension was not well captured by any of the four group PCs. However, based on earlier results (Konkle and Oliva, 2012), it appears that object size is represented in the brain. Thus, it is likely that object size is captured by lower-variance group PCs that could not be significantly discerned in this experiment.

Cortical Maps of Semantic Representation

The results of the PC analysis show that the brains of different individuals represent object and action categories in a common semantic space. Here we examine how this semantic space is represented across the cortical surface. To do this, we first constructed a separate cortical flatmap for each subject using standard techniques (Van Essen et al., 2001). Then we used the scheme described above (see Figure 4) to assign a color to each voxel according to the projection of its category model weights into the PC space (for separate PC maps, see Figure S4). The results are shown in Figures 7A and 7C for two subjects (corresponding maps for other subjects are shown in Figure S5). (Readers who wish to explore these maps in detail, and examine the category selectivity of each voxel, may do so by going to http://gallantlab.org/semanticmovies.)

These maps reveal that the semantic space is represented in broad gradients that are distributed across much of anterior visual cortex (some of these gradients are shown schematically in Figure S6). In inferior temporal cortex, regions of animal (yellow) and human representation (green and blue-green) run along the inferior temporal sulcus (ITS). Both the fusiform face area and occipital face area lie within the region of human representation, but the surrounding region of animal representation was previously unknown. In a gradient that runs from the ITS toward the middle temporal sulcus, human representation gives way to animal representation, which then gives way to representation of human action, athletes, and outdoor spaces (red and red-green). The dorsal part of the gradient contains the extrastriate body area and area MT+/V5 and also responds strongly to motion (positive on the first PC, see Figures 7B and 7D).

In medial occipitotemporal cortex, a region of vehicle (pink) and landscape (purple) representation sits astride the collateral sulcus. This region, which contains the PPA, lies at one end of a long gradient that runs across medial parietal cortex. Toward RSC and along the PrCu, the representational gradient shifts toward buildings (blue-indigo) and landscapes (purple). This gradient continues forward along the superior bank of the intraparietal sulcus as far as the posterior end of the cingulate sulcus while shifting representation toward geography (purple-red) and human action (red). This long gradient encompasses both the dorsal and ventral visual pathways (Ungerleider and Mishkin, 1982) in one unbroken band of cortex that represents a continuum of semantic categories related to vehicles, buildings, landscapes, geography, and human actions.

This map also reveals that visual semantic categories are well represented outside of occipital cortex. In parietal cortex, an anterior-posterior gradient from animal (yellow) to landscape (purple) representation is located in the posterior bank of the postcentral sulcus (PoCeS). This is consistent with earlier reports that movies of hand movements evoke responses in the PoCeS (Buccino et al., 2001; Hasson et al., 2004) and may reflect learned associations between visual and somatosensory stimuli.

In frontal cortex, a region of human action and athlete representation (red) is located at the posterior end of the superior frontal sulcus (SFS). This region, which includes the frontal eye fields (FEFs), lies at one end of a gradient that shifts toward landscape (purple) representation while extending along the SFS.

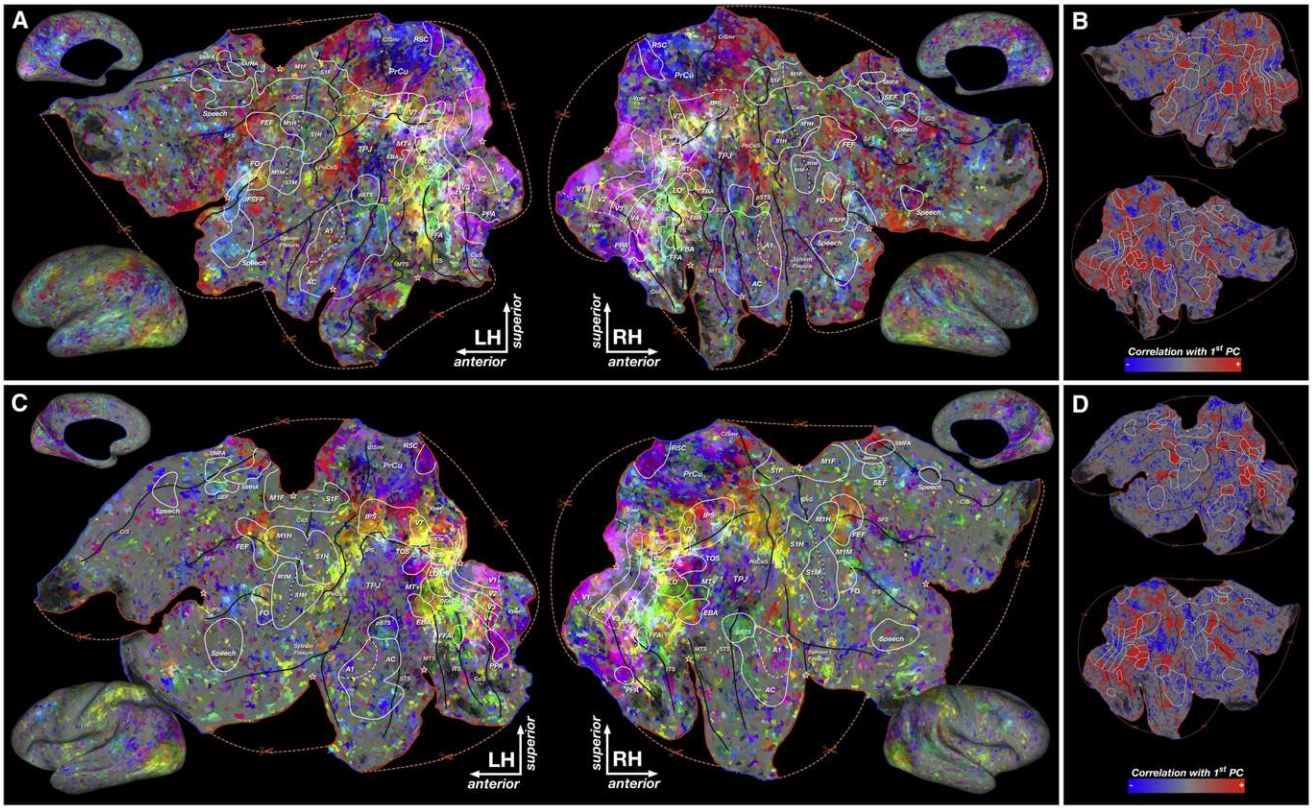


Figure 7. Semantic Space Represented across the Cortical Surface

(A) The category model weights for each cortical voxel in subject A.V. are projected onto PCs 2–4 of the group semantic space and then assigned a color according to the scheme described in Figure 4B. These colors are projected onto a cortical flat map constructed for subject A.V. Each location on the flat map shown here represents a single voxel in the brain of subject A.V. Locations with similar colors have similar semantic selectivity. This map reveals that the semantic space is represented in broad gradients distributed across much of anterior visual cortex. Semantic selectivity is also apparent in medial and lateral parietal cortex, auditory cortex, and lateral prefrontal cortex. Brain areas identified using conventional functional localizers are outlined in white and labeled (see Table S1 for abbreviations). Boundaries that have been inferred from anatomy or that are otherwise uncertain are denoted by dashed white lines. Major sulci are denoted by dark blue lines and labeled (see Table S2 for abbreviations). Some anatomical regions are labeled in light blue (abbreviations: PrCu, precuneus; TPJ, temporoparietal junction). Cuts made to the cortical surface during the flattening procedure are indicated by dashed red lines and a red border. The apex of each cut is indicated by a star. Blue borders show the edge of the corpus callosum and subcortical structures. Regions of fMRI signal dropout due to field inhomogeneity are shaded with black hatched lines.

(B) Projection of voxel model weights onto the first PC for subject A.V. Voxels with positive projections on the first PC appear red, while those with negative projections appear blue and those orthogonal to the first PC appear gray.

(C) Projection of voxel weights onto PCs 2–4 of the group semantic space for subject T.C.

(D) Projection of voxel model weights onto the first PC for subject T.C. See Figure S5 for maps of semantic representation in other subjects.

Note: explore these data sets yourself at http://gallantlab.org/semanticmovies.

Another region of human action, athlete, and animal representation (red-yellow) is located at the posterior inferior frontal sulcus (IFS) and contains the frontal operculum (FO). Both the FO and FEF have been associated with visual attention (Büchel et al., 1998), so we suspect that human action categories might be correlated with salient visual movements that attract covert visual attention in our subjects.

In inferior frontal cortex, a region of indoor structure (blue), human (green), communication verb (also blue-green), and text (cyan) representation runs along the IFS anterior to the FO. This region coincides with the inferior frontal sulcus face patch (Avidan et al., 2005; Tsao et al., 2008) and has also been implicated in processing of visual speech (Calvert and Campbell, 2003) and text (Poldrack et al., 1999). Our results suggest that visual speech, text, and faces are represented in a contiguous region of cortex.

Smoothness of Cortical Semantic Maps

We have shown that the brain represents hundreds of categories within a continuous four-dimensional semantic space that is shared among different subjects. Furthermore, the results shown in Figure 7 suggest that this space is mapped smoothly onto the cortical sheet. However, the results presented thus far are not sufficient to determine whether the apparent smoothness of the cortical map reflects the specific properties of the group semantic space, or rather whether a smooth map might result from any arbitrary four-dimensional projection of our voxel weights onto the cortical sheet. To address this issue, we tested

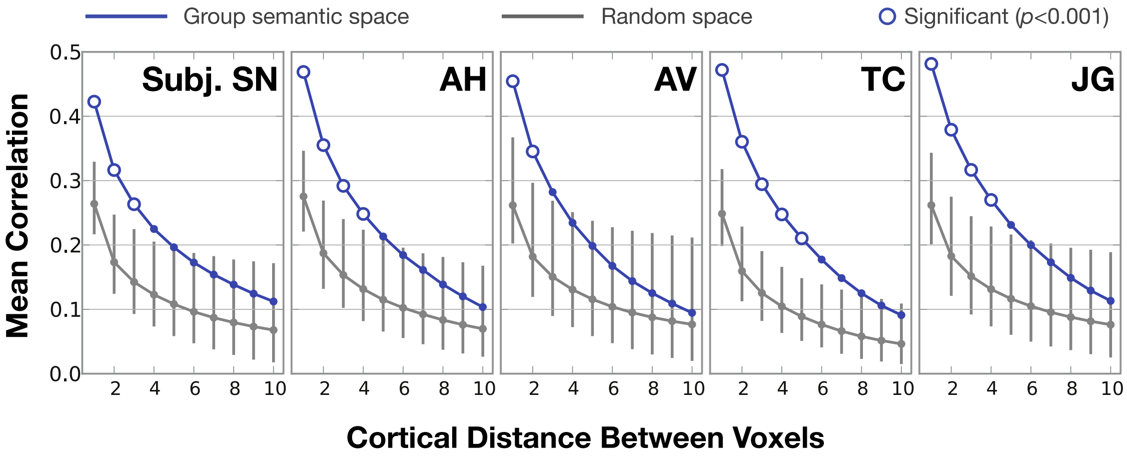


Figure 8. Smoothness of Cortical Maps under the Group Semantic Space

To quantify smoothness of cortical representation under a semantic space, we first projected voxel category model weights into the semantic space. Second, we computed the mean correlation between voxel semantic projections as a function of the distance between voxels along the cortical sheet. To determine whether cortical semantic maps under the group semantic space are significantly smoother than chance, we computed smoothness using the same analysis for 1,000 random four-dimensional spaces. Mean correlations for the group semantic space are plotted in blue, and mean correlations for the 1,000 random spaces are plotted in gray. Gray error bars show 99% confidence intervals for the random space results. Group semantic space correlations that are significantly different from the random space results ($p < 0.001$) are shown as hollow symbols. For adjacent voxels (distance 1) and voxels separated by one intermediate voxel (distance 2), correlations of group semantic space projections are significantly greater than chance in all subjects. This shows that cortical semantic maps under the group semantic space are much smoother than would be expected by chance.

whether cortical maps under the four-PC group semantic space are smoother than expected by chance.

In order to quantify the smoothness of a cortical map, we first projected the category model weights for every voxel into the four-dimensional semantic space. Then we computed the correlation between the projections for each pair of voxels. Finally, we aggregated and averaged these pairwise correlations based on the distance between each pair of voxels along the cortical sheet. To estimate the null distribution of smoothness values and to establish statistical significance, we repeated this procedure using 1,000 random four-dimensional semantic spaces (see Experimental Procedures for details).

Figure 8 shows the average correlation between voxel projections into the semantic space as a function of the distance between voxels along the cortical sheet. In all five subjects, the group semantic space projections have significantly ($p < 0.001$) higher average correlation than the random projections, for both adjacent voxels (distance 1) and voxels separated by one intermediate voxel (distance 2). These results suggest that smoothness of the cortical map is specific to the group semantic space estimated here. Because the group semantic space was constructed without using any spatial information, this finding independently confirms the significance of the group semantic space.

Importance of Category Representation across Cortex

The cortical maps shown in Figure 7 demonstrate that much of the cortex is semantically selective. However, this does not necessarily imply that semantic selectivity is the primary function of any specific cortical site. To assess the importance of semantic selectivity across the cortical surface, we evaluated predictions of the category model, using a separate data set reserved for this purpose (Kay et al., 2008; Naselaris et al., 2009; Nishimoto et al., 2011). Prediction performance was quantified as the correlation between predicted and observed BOLD responses, corrected to account for noise in the validation data (see Experimental Procedures and Hsu et al., 2004).

Figure 9 shows prediction performance projected onto cortical flat maps for two subjects (corresponding maps for other subjects are shown in Figure S7). The category model accurately predicts BOLD responses in occipitotemporal cortex, medial parietal cortex, and lateral prefrontal cortex. On average, 22% of cortical voxels are predicted significantly ($p < 0.01$ uncorrected; 19% in subject S.N., 20% in A.H., 26% in A.V., 26% in T.C., and 21% in J.G.). The category model explains at least 20% of the explainable variance (correlation > 0.44) in an average of 8% of cortical voxels (5% in subject S.N., 7% in A.H., 10% in A.V., 12% in T.C., and 7% in J.G.). These results show that category representation is broadly distributed across the cortex. This result is inconsistent with the results of previous fMRI studies that reported only a few category-selective regions (Schwarzlose et al., 2005; Spiridon et al., 2006). (Note, however, that the category selectivity of individual brain areas reported in these previous studies is consistent with our results.) We suspect that previous studies have underestimated the extent of category representation in the cortex because they used static images and tested only a handful of categories.

Figure 9 also shows that some regions of cortex that appeared semantically selective in Figure 7 are predicted poorly. This suggests that the semantic selectivity of some brain regions is inconsistent or nonstationary. These inconsistent regions include the middle precuneus, temporoparietal junction, and medial prefrontal cortex. All of these regions are thought to be components of the default mode network (Raichle et al.,

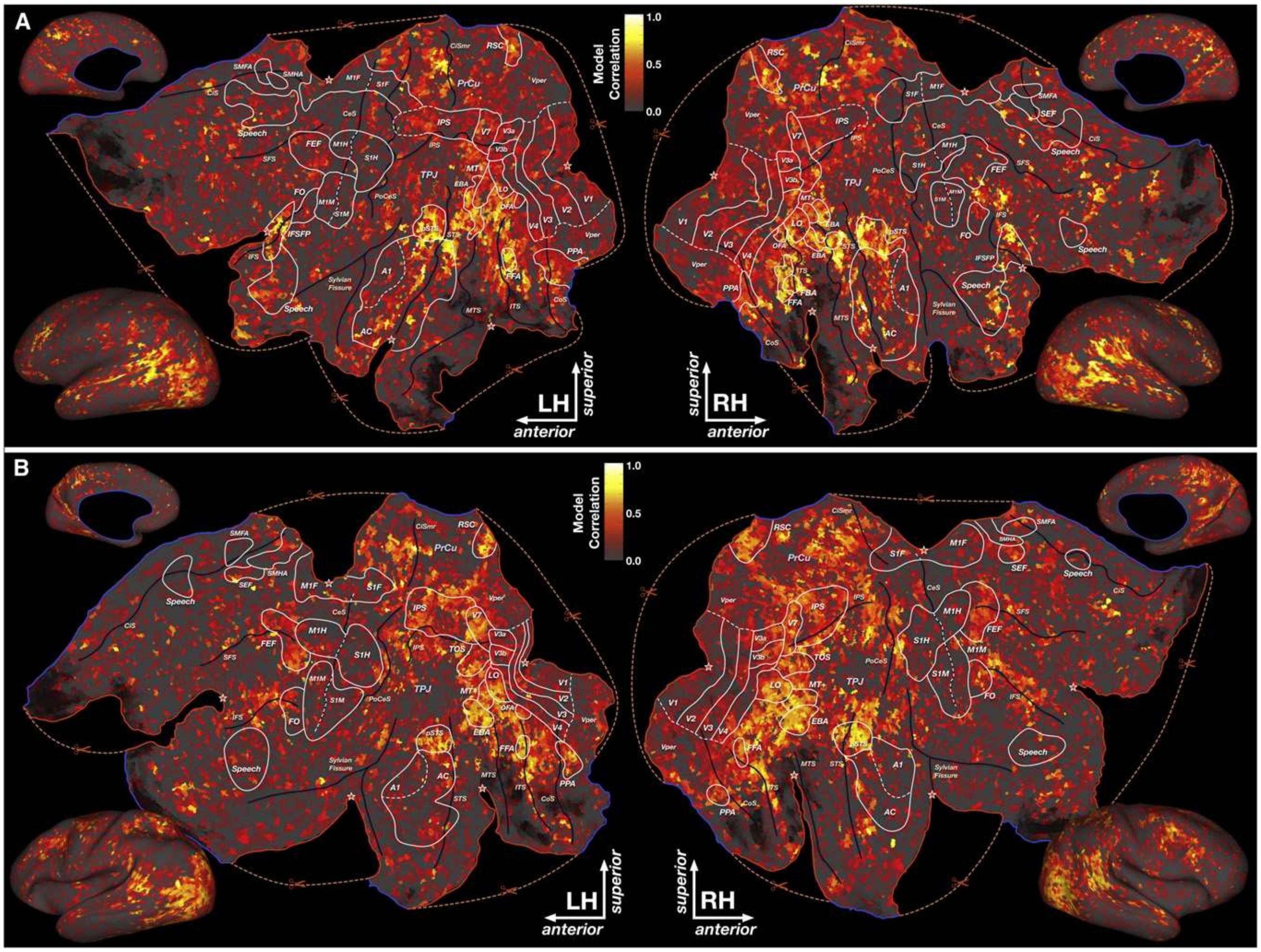


Figure 9. Model Prediction Performance across the Cortical Surface

To determine how much of the response variance of each voxel is explained by the category model, we assessed prediction performance using separate validation data reserved for this purpose.

(A) Each location on the flat map represents a single voxel in the brain of subject A.V. Colors reflect prediction performance on the validation data. Well-predicted voxels appear yellow or white, and poorly predicted voxels appear gray. The best predictions are found in occipitotemporal cortex, the posterior superior temporal sulcus, medial parietal cortex, and inferior frontal cortex.

(B) Model performance for subject T.C. See Figure S7 for model prediction performance in other subjects. See Table S3 for model prediction performance within known functional areas.

2001) and are known to be strongly modulated by attention (Downar et al., 2002). Because we did not control or manipulate attention in this experiment, the inconsistent semantic selectivity of these regions may reflect uncontrolled attentional effects. Future studies that control attention explicitly could improve category model predictions in these regions.

DISCUSSION

We used brain activity evoked by natural movies to study how 1,705 object and action categories are represented in the human brain. The results show that the brain represents categories in a continuous semantic space that reflects category similarity. These results are consistent with the hypothesis that the brain efficiently represents the diversity of categories in a compact space, and they contradict the common hypothesis that each category is represented in a distinct brain area. Assuming that semantically related categories share visual or conceptual features, this organization probably minimizes the number of neurons or neural wiring required to represent these features.

Across the cortex, semantic representation is organized along smooth gradients that seem to be distributed systematically. Functional areas defined using classical contrast methods are merely peaks or nodal points within these broad semantic gradients. Furthermore, cortical maps based on the group semantic space are significantly smoother than expected by chance. These results suggest that semantic representation is analogous to retinotopic representation, in which many smooth gradients of visual eccentricity and angle selectivity tile the cortex (Engel et al., 1997; Hansen et al., 2007). Unlike retinotopy, however, the relevant dimensions of the space underlying semantic representation are not known a priori and so must be derived empirically.

Previous studies have shown that natural movies evoke widespread, robust BOLD activity across much of the cortex (Bartels and Zeki, 2004; Hasson et al., 2004, 2008; Haxby et al., 2011; Nishimoto et al., 2011). However, those studies did not attempt to systematically map semantic representation or discover the underlying semantic space. Our results help explain why natural movies evoke widely consistent activity across different individuals: object and action categories are represented in terms of a common semantic space that maps consistently onto cortical anatomy.

One potential criticism of this study is that the WordNet features used to construct the category model might have biased the recovered semantic space. For example, the category "surgeon" only appears four times in these stimuli, but because it is a descendent of "person" in WordNet, surgeon appears near person in the semantic space. It is possible (however unlikely) that surgeons are represented very differently from other people but that we are unable to recover that information from these data. On the other hand, categories that appeared frequently in these stimuli are largely immune to this bias. For example, among the descendents of "person," there is a large difference between the representations of "athlete" (which appears 282 times in these stimuli) and "man" (which appears 1,482 times). Thus, it appears that bias due to WordNet only affects rare categories. We do not believe that these considerations have a significant effect on the results of this study.

Another potential criticism of the regression-based approach used in this study is that some results could be biased by stimulus correlations. For example, we might conclude that a voxel responds to "talking" when in fact it responds to the presence of a "mouth." In theory, such correlations are modeled and removed by the regression procedure as long as sufficient data are collected, but our data are limited and so some residual correlations may remain. However, we believe that the alternative—bias due to preselecting a small number of stimulus categories—is a more pernicious source of error and misinterpretation in conventional fMRI experiments. Errors due to stimulus correlation can be seen, measured, and tested. Errors due to stimulus preselection are implicit and largely invisible.

The group semantic space found here captures large semantic distinctions such as mobile versus stationary categories but misses finer distinctions such as "old faces" versus "young faces" (Op de Beeck et al., 2010) and "small objects" versus "large objects" (Konkle and Oliva, 2012). These fine distinctions would probably be captured by lower-variance dimensions of the shared semantic space that could not be recovered in this experiment. The dimensionality and resolution of the recovered semantic space are limited by the quality of BOLD fMRI and by the size and semantic breadth of the stimulus set. Future studies that use more sensitive measures of brain activity or broader stimulus sets will probably reveal additional dimensions of the common semantic space. Further studies using more subjects will also be necessary in order to understand differences in semantic representation between individuals.

Some previous studies have reported that animal and nonanimal categories are represented distinctly in the human brain (Downing et al., 2006; Kriegeskorte et al., 2008; Naselaris et al., 2009). Another study proposed an alternative: that animal categories are represented using an animacy continuum (Connolly et al., 2012), in which animals that are more similar to humans have higher animacy. Our results show that animacy is well represented on the first, and most important, PC in the group semantic space. The binary distinction between animals and nonanimals is also well represented but only on the fourth PC. Moreover, the fourth PC is better explained by the distinction between biological categories (including plants) and nonbiological categories. These results suggest that the animacy continuum is more important for category representation in the brain than is the binary distinction between animal and nonanimal categories.

A final important question about the group semantic space is whether it reflects visual or conceptual features of the categories. For example, people and nonhuman animals might be represented similarly because they share visual features such as hair, or because they share conceptual features such as agency or self-locomotion. The answer to this question probably depends upon which voxels are used to construct the semantic space. Voxels from occipital and inferior temporal cortex have been shown to have similar semantic representation in humans and monkeys (Kriegeskorte et al., 2008). Therefore, these voxels probably represent visual features of the categories and not conceptual features. In contrast, voxels from medial parietal cortex and frontal cortex probably represent conceptual features of the categories. Because the group semantic space reported here was constructed using voxels from across the entire brain, it probably reflects a mixture of visual and conceptual features. Future studies using both visual and nonvisual stimuli will be required to disentangle the contributions of visual versus conceptual features to semantic representation. Furthermore, a model that represents stimuli in terms of visual and conceptual features might produce more accurate and parsimonious predictions than the category model used here.

EXPERIMENTAL PROCEDURES

MRI Data Collection

MRI data were collected on a 3T Siemens TIM Trio scanner at the UC Berkeley Brain Imaging Center using a 32-channel Siemens volume coil. Functional scans were collected using a gradient echo-EPI sequence with repetition time (TR) = 2.0045 s, echo time (TE) = 31 ms, flip angle = 70°, voxel size = 2.24 × 2.24 × 4.1 mm, matrix size = 100 × 100, and field of view = 224 × 224 mm. We prescribed 32 axial slices to cover the entire cortex. A custom-modified bipolar water excitation radio frequency (RF) pulse was used to avoid signal from fat.

Anatomical data for subjects A.H., T.C., and J.G. were collected using a T1-weighted MP-RAGE sequence on the same 3T scanner. Anatomical data for subjects S.N. and A.V. were collected on a 1.5T Philips Eclipse scanner as described in an earlier publication (Nishimoto et al., 2011).

Subjects

Functional data were collected from five male human subjects, S.N. (author S.N., age 32), A.H. (author A.G.H., age 25), A.V. (author A.T.V., age 25), T.C. (age 29), and J.G. (age 25). All subjects were healthy and had normal or corrected-to-normal vision. The experimental protocol was approved by the

Committee for the Protection of Human Subjects at University of California, Berkeley.

Natural Movie Stimuli

Model estimation data were collected in 12 separate 10 min scans. Validation data were collected in nine separate 10 min scans, each consisting of ten 1 min validation blocks. Each 1 min validation block was presented ten times within the 90 min of validation data. The stimuli and experimental design were identical to those used in Nishimoto et al. (2011), except that here the movies were shown on a projection screen at 24 × 24 degrees of visual angle.

fMRI Data Preprocessing

Each functional run was motion corrected using the FMRIB Linear Image Registration Tool (FLIRT) from FSL 4.2 (Jenkinson and Smith, 2001). All volumes in the run were then averaged to obtain a high-quality template volume. FLIRT was also used to automatically align the template volume for each run to the overall template, which was chosen to be the template for the first functional movie run for each subject. These automatic alignments were manually checked and adjusted for accuracy. The cross-run transformation matrix was then concatenated to the motion-correction transformation matrices obtained using MCFLIRT, and the concatenated transformation was used to resample the original data directly into the overall template space.

Low-frequency voxel response drift was identified using a median filter with a 120 s window and this was subtracted from the signal. The mean response for each voxel was then subtracted and the remaining response was scaled to have unit variance.

Flatmap Construction

Cortical surface meshes were generated from the T1-weighted anatomical scans using Caret5 software (Van Essen et al., 2001). Five relaxation cuts were made into the surface of each hemisphere and the surface crossing the corpus callosum was removed. The calcarine sulcus cut was made at the horizontal meridian in V1 using retinotopic mapping data as a guide. Surfaces were then flattened using Caret5.

Functional data were aligned to the anatomical data for surface projection using custom software written in MATLAB (MathWorks).

Stimulus Labeling and Preprocessing

One observer manually tagged each second of the movies with WordNet labels describing the salient objects and actions in the scene. The number of labels per second varied between 1 and 14, with an average of 4.2. Categories were tagged if they appeared in at least half of the 1 s clip. When possible, specific labels (e.g., "priest") were used instead of generic labels (e.g., "person"). Label assignments were spot checked for accuracy by two additional observers. For example labeled clips, see Figure S1.

The labels were then used to build a category indicator matrix, in which each second of movie occupies a row and each category occupies a column. A value of 1 was assigned to each entry in which that category appeared in that second of movie and all other entries were set to zero. Next, the WordNet hierarchy (Miller, 1995) was used to add all the superordinate categories entailed by each labeled category. For example, if a clip was labeled with "wolf," we would automatically add the categories "canine," "carnivore," "placental mammal," "mammal," "vertebrate," "chordate," "organism," and "whole." According to this scheme the predicted BOLD response to a category is not just the weight on that category but the sum of weights for all entailed categories.

The addition of superordinate categories should improve model predictions by allowing poorly sampled categories to share information with their WordNet neighbors. To test this hypothesis, we compared prediction performance of the model with superordinate categories to a model that used only the labeled categories. The number of significantly predicted voxels is 10%–20% higher with the superordinate category model than with the labeled category model. To ensure that the PCA results presented here are not an artifact of the added superordinate categories, we performed the same analysis using the labeled categories model. The results obtained using the labeled categories model were qualitatively similar to those obtained using the full model (data not shown).

The regression procedure also included one additional feature that described the total motion energy during each second of the movie. This regressor was added in order to explain away spurious correlation between responses in early visual cortex and some categories. Total motion energy was computed as the mean output of a set of 2,139 motion energy filters (Nishimoto et al., 2011), in which each filter consisted of a quadrature pair of space-time Gabor filters (Adelson and Bergen, 1985; Watson and Ahumada, 1985). The motion energy filters tile the image space with a variety of preferred spacial frequencies, orientations, and temporal frequencies. The total motion energy regressor explained much of the response variance in early visual cortex (mainly V1 and V2). This had the desired effect of explaining away correlations between responses in early visual cortex and categories that feature full-field motion (e.g., "fire" and "snow"). The total motion energy regressor was used to fit the category model but was not included in the model predictions.

Voxelwise Model Fitting and Testing

The category model was fit to each voxel individually. A set of linear temporal filters was used to model the slow hemodynamic response inherent in the BOLD signal (Nishimoto et al., 2011). To capture the hemodynamic delay, we used concatenated stimulus vectors that had been delayed by two, three, and four samples (4, 6, and 8 s). For example, one stimulus vector indicates the presence of "wolf" 4 s earlier, another the presence of "wolf" 6 s earlier, and a third the presence of "wolf" 8 s earlier. Taking the dot product of this delayed stimulus with a set of linear weights is functionally equivalent to convolution of the original stimulus vector with a linear temporal kernel that has nonzero entries for 4, 6, and 8 s delays.

For details about the regularized regression procedure, model testing, and correction for noise in the validation set, please see the Supplemental Experimental Procedures.

All model fitting and analysis was performed using custom software written in Python, which made heavy use of the NumPy (Oliphant, 2006) and SciPy (Jones et al., 2001) libraries.

Estimating Predicted Category Response

In the semantic category model used here, each category entails the presence of its superordinate categories in the WordNet hierarchy. For example, "wolf" entails the presence of "canine," "carnivore," etc. Because these categories must be present in the stimulus if "wolf" is present, the model weight for "wolf" alone does not accurately reflect the model's predicted response to a stimulus containing only a "wolf." Instead, the predicted response to "wolf" is the sum of the weights for "wolf," "canine," "carnivore," etc. Thus, to determine the predicted response of a voxel to a given category, we added together the weights for that category and all categories that it entails. This procedure is equivalent to simulating the response of a voxel to a stimulus labeled only with "wolf."

We used this procedure to estimate the predicted category responses shown in Figure 2, to assign colors and positions to the category nodes shown in Figures 4 and 5, and to correct PC coefficients before comparing them to hypothetical semantic dimensions as shown in Figure 6.

Principal Components Analysis

For each subject, we used PCA to recover a low-dimensional semantic space from category model weights. We first selected all voxels that the model predicted significantly, using a liberal significance threshold ($p < 0.05$ uncorrected for multiple comparisons). This yielded 8,269 voxels in subject S.N., 8,626 voxels in A.H., 11,697 voxels in A.V., 11,187 voxels in T.C., and 9,906 voxels in J.G. We then applied PCA to the category model weights of the selected voxels, yielding 1,705 PCs for each subject. (In additional tests, we found that varying the voxel selection threshold does not strongly affect the PCA results.) Partial scree plots showing the amount of variance accounted for by each PC are shown in Figure 3. The first four PCs account for 24.1% of variance in subject S.N., 25.9% of variance in A.H., 28.0% of variance in A.V., 25.8% of variance in T.C., and 25.6% of variance in J.G.

Second, we tested whether the recovered PCs were different from what we would expect by chance. For details of this procedure, please see the Supplemental Experimental Procedures.

In this paper, we present semantic analyses using PCA, but PCA is only one of many dimensionality reduction methods. Sparse methods such as independent components analysis and nonnegative matrix factorization can also be used to recover the underlying semantic space. We found that these methods produced qualitatively similar results to PCA on the data presented here. In this paper, we present only PCA results because PCA is commonly used, easy to understand, and the results are highly interpretable.

Stimulus Identification Using Category Model and Models Based on Group PCs

To quantify the relative amount of information that can be represented by the full category model and the models based on group PCs, we used the validation data to perform an identification analysis (Kay et al., 2008; Nishimoto et al., 2011). For the full category model, we calculated log likelihoods of the observed responses given predicted responses to the validation stimuli and the fitted category model (Nishimoto et al., 2011). Here we declare correct identification if the highest likelihood for aggregated 18 s (9 TR) chunks of responses can be associated with the correct timings for the matched stimulus chunks within ±1 volume (TR). In order to minimize the potential confound due to nonsemantic stimulus features, we subtracted the prediction of the total motion energy regressor from responses before the analysis.

To perform the identification analysis for models based on the group PCs, we repeated the same procedures as above but using group PC models. We obtained these models by voxelwise regression using the category stimuli projected into the group PC space (see voxelwise model fitting and principal component analysis in Experimental Procedures). In order to assess variability in the performance measurements, we performed the identification analysis ten times, based on group PCs obtained using bootstrap voxel samples.

To reduce noise, the identification analyses used only the 2,000 most predictable voxels. Prediction performance was assessed using 10% of the training data that we reserved from the regression for this purpose. Voxel selection was performed separately for each model and subject.

Comparison between Group Semantic Space and Hypothesized Semantic Dimensions

To compare the dimensions of the group semantic space to hypothesized semantic dimensions, we first defined each hypothesized dimension as a vector with a value for each of the 1,705 categories. We then computed the variance that each hypothesized dimension explains in each group PC as the squared correlation between the PC vector and hypothesized dimension vector. To find confidence intervals on the variance explained in each PC, we bootstrapped the group PCA by sampling with replacement 100 times from the pooled voxel population.

We defined nine semantic dimensions based on previous publications and our own hypotheses. These dimensions included mobile versus immobile, animacy, humans versus nonhumans, social versus nonsocial, civilization versus nature, animal versus nonanimal, biological versus nonbiological, place versus nonplace, and object size. For the mobile versus immobile dimension, we assigned positive weights to mobile categories such as animals, people, and vehicles, and zero weight to all other categories. For the animacy dimension based on Connolly et al. (2012), we assigned high weights to people and intermediate and low weights to other animals based on their phylogenetic distance from humans: more distant animals were assigned lower weights. For the human versus nonhuman dimension, we assigned positive weights to people and zero weights to all other categories. For the social versus nonsocial dimension, we assigned positive weights to people and communication verbs and zero weights to all other categories. For the civilization versus nature dimension, we assigned positive weights to people, man-made objects (e.g., ‘‘buildings,’’ ‘‘vehicles,’’ and ‘‘tools’’), and communication verbs and negative weights to nonhuman animals. For the animal versus nonanimal dimension, we assigned positive weights to nonhuman animals, people, and body parts and zero weight to all other categories. For the biological versus nonbiological dimension, we assigned positive weights to all organisms (e.g., ‘‘people,’’ ‘‘nonhuman animals,’’ and ‘‘plants’’), plant organs (e.g., ‘‘flower’’ and ‘‘leaf’’), body parts, and body coverings (e.g., ‘‘hair’’). For the place versus nonplace dimension, we assigned positive weights to outdoor categories (e.g., ‘‘geological formations,’’ ‘‘geographical locations,’’ ‘‘roads,’’ ‘‘bridges,’’ and ‘‘buildings’’) and zero weight to all other categories. For the real-world size dimension based on Konkle and Oliva (2012), we assigned a high weight to large objects (e.g., ‘‘boat’’), medium weight to human-scale objects (e.g., ‘‘person’’), a small weight to small objects (e.g., ‘‘glasses’’), and zero weight to objects that have no size (e.g., ‘‘talking’’) and those that can be many sizes (e.g., ‘‘animal’’).

Smoothness of Cortical Maps under Group Semantic Space

Projecting voxel category model weights onto the group semantic space produces semantic maps that appear spatially smooth (see Figure 7). However, these maps alone are insufficient to determine whether the apparent smoothness of the cortical map is a specific property of the four-PC group semantic space. If the categorical model weights are themselves smoothly mapped onto the cortical sheet, then any four-dimensional projection of these weights might appear equally as smooth as the projection onto the group semantic space. To address this issue, we tested whether cortical maps under the four-PC group semantic space are smoother than expected by chance.

First, we constructed a voxel adjacency matrix based on the fiducial cortical surfaces. The cortical surface for each hemisphere in each subject was represented as a triangular mesh with roughly 60,000 vertices and 120,000 edges. Two voxels were considered adjacent if there was an edge that connects a vertex inside one voxel to a vertex inside the other. Second, we computed the distance between each pair of voxels in the cortex as the length of the shortest path between the voxels in the adjacency graph. This distance metric does not directly translate to physical distance, because the voxels in our scan are not isotropic. However, this affects all models that we test and thus will not bias the results of this analysis.

Third, we projected the voxel category weights onto the four-dimensional group semantic space, which reduced each voxel to a length 4 vector. We then computed the correlation between the projected weights for each pair of voxels in the cortex. Fourth, for each distance up to ten voxels, we computed the mean correlation between all pairs of voxels separated by that distance. This procedure produces a spatial autocorrelation function for each subject. These results are shown as blue lines in Figure 8.

To determine whether cortical map smoothness is specific to the group semantic space, we repeated this analysis 1,000 times using random semantic spaces of the same dimension as the group semantic space. Random orthonormal four-dimensional projections from the 1,705-dimensional category space were constructed by applying singular value decomposition to randomly generated 4 × 1,705 matrices. One can think of these spaces as uniform random rotations of the group semantic space inside the 1,705-dimensional category space.

We considered the observed mean pairwise correlation under the group semantic space to be significant if it exceeded all of the 1,000 random samples, corresponding to a p value of less than 0.001.

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, three tables, Supplemental Experimental Procedures, and one movie and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2012.10.014.

ACKNOWLEDGMENTS

The work was supported by grants from the National Eye Institute (EY019684) and from the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-0939370. A.G.H. was also supported by the William Orr Dingwall Neurolinguistics Fellowship. We thank Natalia Bilenko and Tolga Çukur for helping with fMRI data collection, Neil Thompson for assistance with the WordNet analysis, and Tom Griffiths and Sonia Bishop for discussions regarding the manuscript. A.G.H., S.N., and J.L.G. conceived and designed the experiment. A.G.H., S.N., and A.T.V. collected the fMRI data. A.T.V. and Tolga Çukur customized and optimized the fMRI pulse sequence. A.T.V. did brain flattening and localizer analysis. A.G.H. tagged the movies. S.N. and A.G.H. analyzed the data. A.G.H. and J.L.G. wrote the paper.

Accepted: October 8, 2012
Published: December 19, 2012

REFERENCES

Adelson, E.H., and Bergen, J.R. (1985). Spatiotemporal energy models for the perception of motion. J. Opt. Soc. Am. A *2*, 284–299.

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.

Avidan, G., Hasson, U., Malach, R., and Behrmann, M. (2005). Detailed exploration of face-related processing in congenital prosopagnosia: 2. Functional neuroimaging findings. J. Cogn. Neurosci. *17*, 1150–1167.

Bartels, A., and Zeki, S. (2004). Functional brain mapping during free viewing of natural scenes. Hum. Brain Mapp. *21*, 75–85.

Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R.J., Zilles, K., Rizzolatti, G., and Freund, H.J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. Eur. J. Neurosci. *13*, 400–404.

Büchel, C., Josephs, O., Rees, G., Turner, R., Frith, C.D., and Friston, K.J. (1998). The functional anatomy of attention to visual motion. A functional MRI study. Brain *121*, 1281–1294.

Calvert, G.A., and Campbell, R. (2003). Reading speech from still and moving faces: the neural substrates of visible speech. J. Cogn. Neurosci. *15*, 57–70.

Chao, L.L., Haxby, J.V., and Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. Nat. Neurosci. *2*, 913–919.

Clark, V.P., Keil, K., Maisog, J.M., Courtney, S., Ungerleider, L.G., and Haxby, J.V. (1996). Functional magnetic resonance imaging of human visual cortex during face matching: a comparison with positron emission tomography. Neuroimage *4*, 1–15.

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.A., and Michel, F. (2000). The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. Brain *123*, 291–307.

Connolly, A.C., Guntupalli, J.S., Gors, J., Hanke, M., Halchenko, Y.O., Wu, Y.-C., Abdi, H., and Haxby, J.V. (2012). The representation of biological classes in the human brain. J. Neurosci. *32*, 2608–2618.

Downar, J., Crawley, A.P., Mikulis, D.J., and Davis, K.D. (2002). A cortical network sensitive to stimulus salience in a neutral behavioral context across multiple sensory modalities. J. Neurophysiol. *87*, 615–620.

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

Engel, S.A., Glover, G.H., and Wandell, B.A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cereb. Cortex *7*, 181–192.

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

Fox, C.J., Iaria, G., and Barton, J.J.S. (2009). Defining the face processing network: optimization of the functional localizer in fMRI. Hum. Brain Mapp. *30*, 1637–1651.

Halgren, E., Dale, A.M., Sereno, M.I., Tootell, R.B.H., Marinkovic, K., and Rosen, B.R. (1999). Location of human face-selective cortex with respect to retinotopic areas. Hum. Brain Mapp. *7*, 29–37.

Hansen, K.A., Kay, K.N., and Gallant, J.L. (2007). Topographic organization in and near human visual area V4. J. Neurosci. *27*, 11896–11911.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. Science *303*, 1634–1640.

Hasson, U., Yang, E., Vallines, I., Heeger, D.J., and Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. J. Neurosci. *28*, 2539–2550.

Hauk, O., Johnsrude, I., and Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. Neuron *41*, 301–307.

Haxby, J.V., Guntupalli, J.S., Connolly, A.C., Halchenko, Y.O., Conroy, B.R., Gobbini, M.I., Hanke, M., and Ramadge, P.J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. Neuron *72*, 404–416.

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.

Hsu, A., Borst, A., and Theunissen, F.E. (2004). Quantifying variability in neural responses and its application for the validation of model predictions. Network *15*, 91–109.

Iacoboni, M., Lieberman, M.D., Knowlton, B.J., Molnar-Szakacs, I., Moritz, M., Throop, C.J., and Fiske, A.P. (2004). Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. Neuroimage *21*, 1167–1173.

Jenkinson, M., and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. Med. Image Anal. *5*, 143–156.

Jones, E., Oliphant, T. E., and Peterson, P. (2001). SciPy: Open Source Scientific Tools for Python. http://www.scipy.org.

Just, M.A., Cherkassky, V.L., Aryal, S., and Mitchell, T.M. (2010). A neurosemantic theory of concrete noun representation based on the underlying brain codes. PLoS ONE *5*, e8622.

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.

Konkle, T., and Oliva, A. (2012). A real-world size organization of object responses in occipitotemporal cortex. Neuron *74*, 1114–1124.

Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., and Bandettini, P.A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron *60*, 1126–1141.

McCarthy, G., Puce, A., Gore, J.C., and Allison, T. (1997). Face-specific processing in the human fusiform gyrus. J. Cogn. Neurosci. *9*, 605–610.

Miller, G.A. (1995). WordNet: a lexical database for English. Commun. ACM *38*, 39–41.

Mitchell, T.M., Shinkareva, S.V., Carlson, A., Chang, K.-M., Malave, V.L., Mason, R.A., and Just, M.A. (2008). Predicting human brain activity associated with the meanings of nouns. Science *320*, 1191–1195.

Nakamura, K., Kawashima, R., Sato, N., Nakamura, A., Sugiura, M., Kato, T., Hatano, K., Ito, K., Fukuda, H., Schormann, T., and Zilles, K. (2000). Functional delineation of the human occipito-temporal areas related to face and scene processing. A PET study. Brain *123*, 1903–1912.

Naselaris, T., Prenger, R.J., Kay, K.N., Oliver, M., and Gallant, J.L. (2009). Bayesian reconstruction of natural images from human brain activity. Neuron *63*, 902–915.

Nishimoto, S., Vu, A.T., Naselaris, T., Benjamini, Y., Yu, B., and Gallant, J.L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. Curr. Biol. *21*, 1641–1646.

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

Oliphant, T.E. (2006). Guide to NumPy (Provo, UT: Brigham Young University).

Op de Beeck, H.P., Haushofer, J., and Kanwisher, N.G. (2008). Interpreting fMRI data: maps, modules and dimensions. Nat. Rev. Neurosci. *9*, 123–135.

Op de Beeck, H.P., Brants, M., Baeck, A., and Wagemans, J. (2010). Distributed subordinate specificity for bodies, faces, and buildings in human ventral visual cortex. Neuroimage *49*, 3414–3425.

Peelen, M.V., and Downing, P.E. (2005). Selectivity for the human body in the fusiform gyrus. J. Neurophysiol. *93*, 603–608.

Peelen, M.V., Wiggett, A.J., and Downing, P.E. (2006). Patterns of fMRI activity dissociate overlapping functional brain areas that respond to biological motion. Neuron *49*, 815–822.

Pelphrey, K.A., Morris, J.P., Michelich, C.R., Allison, T., and McCarthy, G. (2005). Functional anatomy of biological motion perception in posterior temporal cortex: an FMRI study of eye, mouth and hand movements. Cereb. Cortex *15*, 1866–1876.

Poldrack, R.A., Wagner, A.D., Prull, M.W., Desmond, J.E., Glover, G.H., and Gabrieli, J.D.E. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. Neuroimage *10*, 15–35.

Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., and Shulman, G.L. (2001). A default mode of brain function. Proc. Natl. Acad. Sci. USA *98*, 676–682.

Rajimehr, R., Young, J.C., and Tootell, R.B.H. (2009). An anterior temporal face patch in human cortex, predicted by macaque maps. Proc. Natl. Acad. Sci. USA *106*, 1995–2000.

Schwarzlose, R.F., Baker, C.I., and Kanwisher, N.G. (2005). Separate face and body selectivity on the fusiform gyrus. J. Neurosci. *25*, 11055–11059.

Smith, A.T., Greenlee, M.W., Singh, K.D., Kraemer, F.M., and Hennig, J. (1998). The processing of first- and second-order motion in human visual cortex assessed by functional magnetic resonance imaging (fMRI). J. Neurosci. *18*, 3816–3830.

Spiridon, M., Fischl, B., and Kanwisher, N. (2006). Location and spatial profile of category-specific regions in human extrastriate cortex. Hum. Brain Mapp. *27*, 77–89.

Tsao, D.Y., Moeller, S., and Freiwald, W.A. (2008). Comparing face patch systems in macaques and humans. Proc. Natl. Acad. Sci. USA *105*, 19514–19519.

Ungerleider, L.G., and Mishkin, M. (1982). Two cortical visual systems. In Analysis of Visual Behavior, D.J. Ingle, M.A. Goodale, and R.J.W. Masfield, eds. (Cambridge, MA: MIT Press), pp. 549–596.

Van Essen, D.C., Drury, H.A., Dickson, J., Harwell, J., Hanlon, D., and Anderson, C.H. (2001). An integrated software suite for surface-based analyses of cerebral cortex. J. Am. Med. Inform. Assoc. *8*, 443–459.

Watson, A.B., and Ahumada, A.J., Jr. (1985). Model of human visual-motion sensing. J. Opt. Soc. Am. A *2*, 322–341.

Neuron, Volume *76*

Supplemental Information

A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain

Alexander G. Huth, Shinji Nishimoto, An T. Vu, and Jack L. Gallant

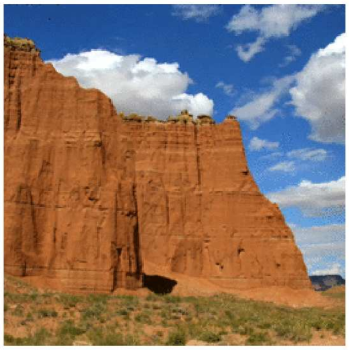
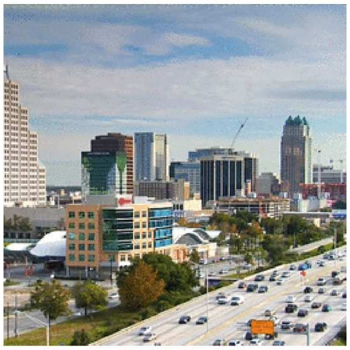
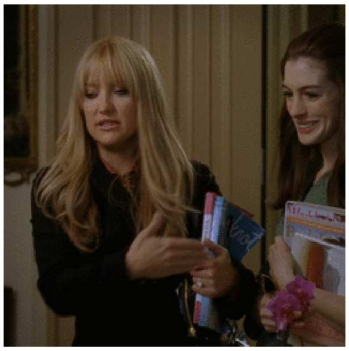
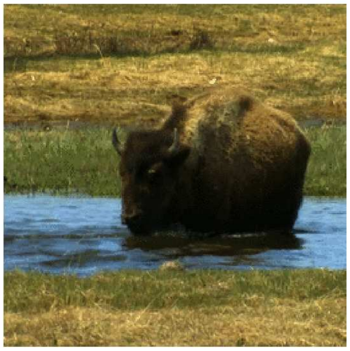
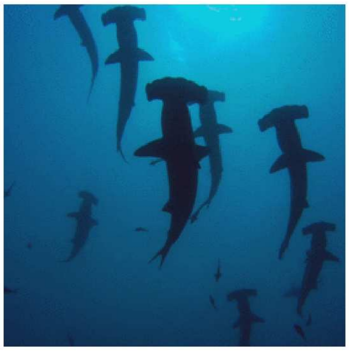
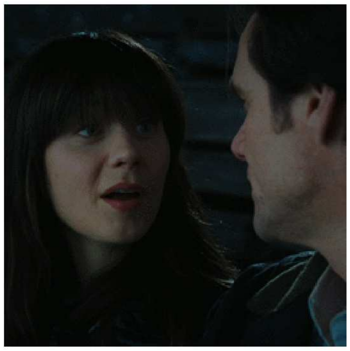
Movie Clip	Labels	Movie Clip	Labels
	butte.n.01 desert.n.01 sky.n.01 cloud.n.01 brush.n.01		city.n.01 expressway.n.01 skyscraper.n.01 traffic.n.01 sky.n.01
	woman.n.01 talk.v.02 gesticulate.v.01 book.n.01		bison.n.01 walk.v.01 grass.n.01 stream.n.01
	hammerhead.n.01 swim.v.01 water.n.01		woman.n.01 man.n.01 talk.v.02

Figure S1 (related to Figure 1). Example stimuli with category labels. Representative frames from six movie clips that were used as stimuli in this experiment, along with the labels that were assigned to those clips. The WordNet lexicon (Miller, 1995) was used to label salient objects and actions in each second of the movies. Each WordNet label (e.g. *bison.n.01*) has a name (*bison*), a part-of-speech (*n* for noun or *v* for verb), and a number (*01*, indicating the first meaning of bison).

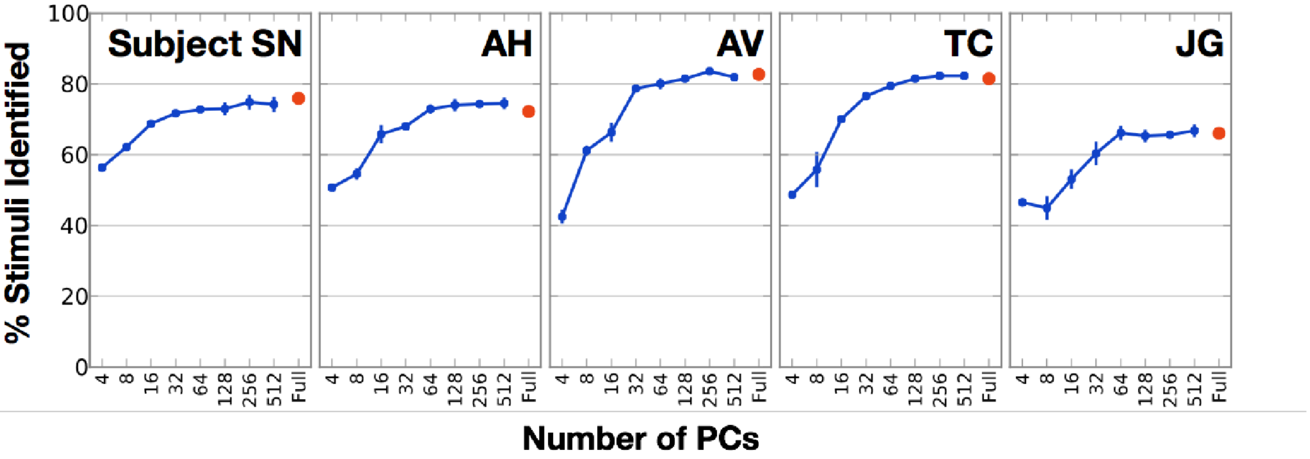


Figure S2 (related to Figure 3). Stimulus identification accuracy using different numbers of group PCs. We used an identification analysis (Kay et al., 2008; Nishimoto et al., 2011) to quantify the amount of information retained in the semantic space defined by the group PCs. Identification accuracy measures how well a model can associate BOLD responses observed across a voxel population with the stimuli that evoked them (see Experimental Procedures for details). Here we show the identification accuracy for each subject in a separate panel. Identification accuracy for the full category model is shown as a red marker. Identification accuracy for models based on different numbers of group PCs (4 – 512) are shown as blue markers. The voxel set used to construct the group PC space was bootstrapped 10 times, and error bars here show the minimum and maximum identification accuracy over the 10 bootstrap samples, while the markers show the average identification accuracy across the 10 samples. The full category model can correctly identify an average of 76% of stimuli across subjects (chance is 1.9%), while the model based on the 4-PC group space can correctly identify 49% of the stimuli, roughly two thirds as many as the full model. These results show that the 4-PC group space does not capture all of the stimulus-related information present in the full category model, indicating that the true semantic space likely has many more than four dimensions. Further experiments will be required to significantly resolve the other semantic dimensions. Note that models based on 64 or more group PCs have approximately the same identification accuracy as the full category model.

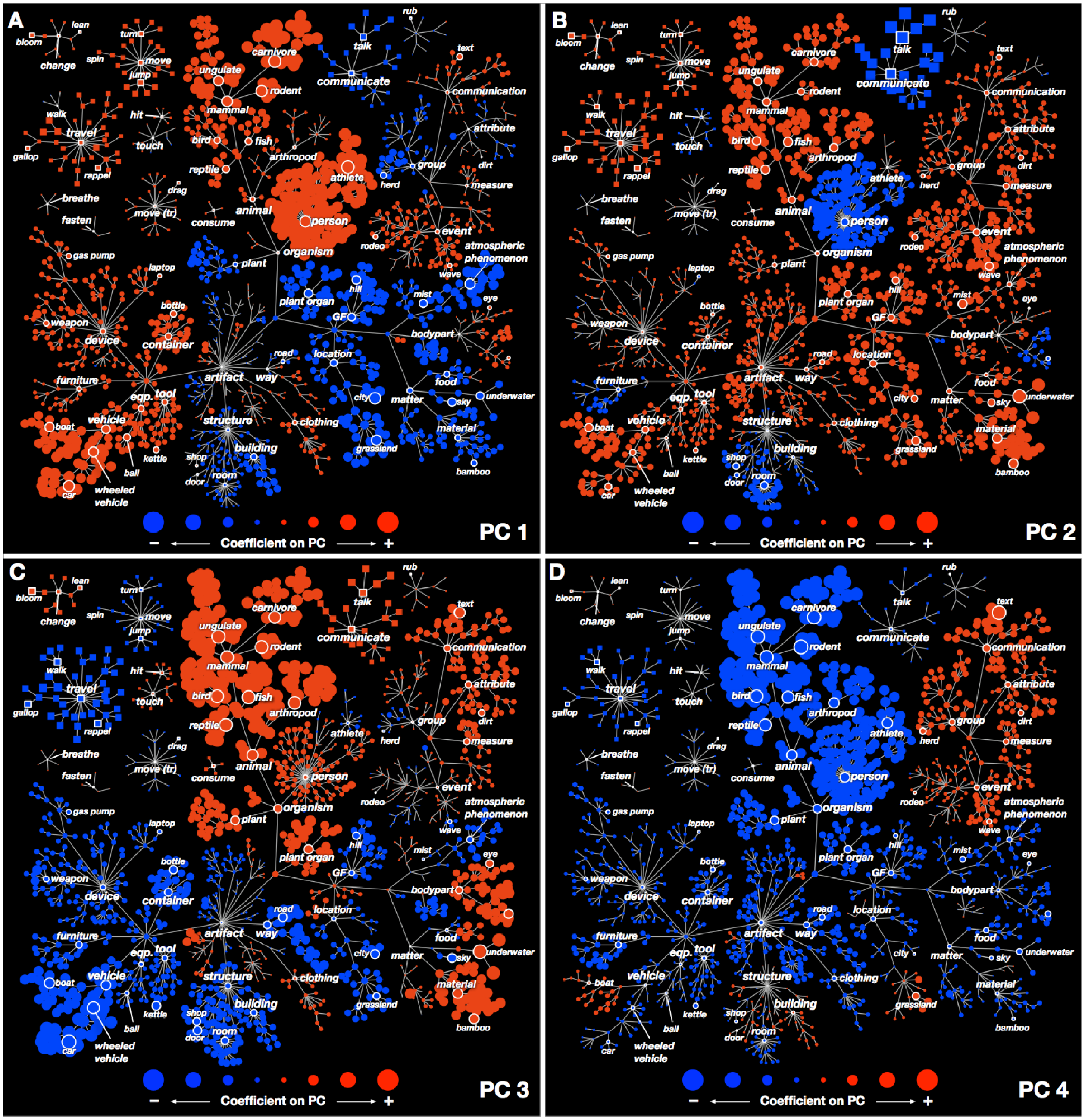


Figure S3 (related to Figure 4). Coefficients of the 1st-4th group semantic PCs. Each panel shows all 1705 categories, organized according to the graphical structure of WordNet. Each marker represents a single noun (circle) or verb (square). Red markers indicate positive coefficients and blue negative. The size of each marker indicates the magnitude of its coefficient on the PC. (**A**) The first PC distinguishes between categories with high stimulus energy (e.g. moving objects like *person* and *vehicle)* and low stimulus energy (e.g. stationary objects like *sky* and *city*). (**B**) The second PC distinguishes between categories involved in social interaction (e.g., *people*, communication verbs, *indoor spaces*, *body parts*, and *furniture*) and those involved in outdoor activities and actions (e.g., *animals*, *vehicles*, *outdoor events*, *geological formations*, movement verbs, and *landscapes*). (**C**) The third PC distinguishes between categories associated with nature (e.g., *animals*, *plants*, *body parts*, and *materials*) and categories associated with civilization (e.g., *vehicles*, *roads*, and *indoor spaces*). (**D**) The fourth PC is difficult to interpret, but roughly distinguishes between living categories (e.g., *animals*, *people*, and *plants*) and most other categories (particularly *text*).

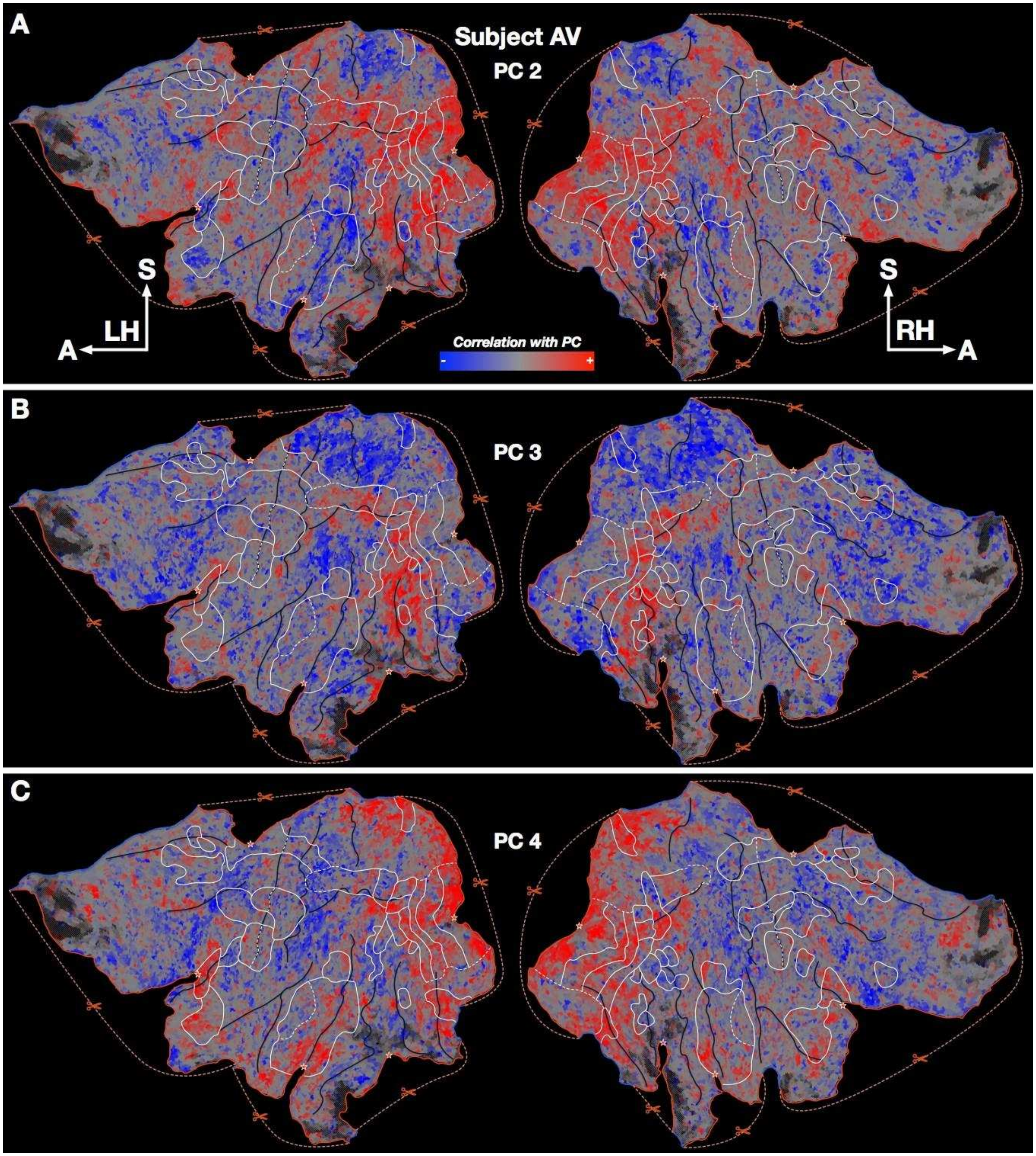


Figure S4 (related to Figure 7). Cortical flat maps with individual PC projections. Similar to Figure 7 in the main text, this figure shows the second, third, and fourth group semantic PCs projected on the cortical flat map constructed specially for subject AV. (**A**) Each location on the map represents a single voxel in subject AV. The color reflects the projection of the voxel category model weights on the second PC (shown in Figure S3B). Voxels that are positively correlated with the second PC appear red, while negatively correlated voxels appear blue. Voxels orthogonal to the second PC appear gray. (**B**) Same, for third PC. (**C**) Same, for fourth PC.

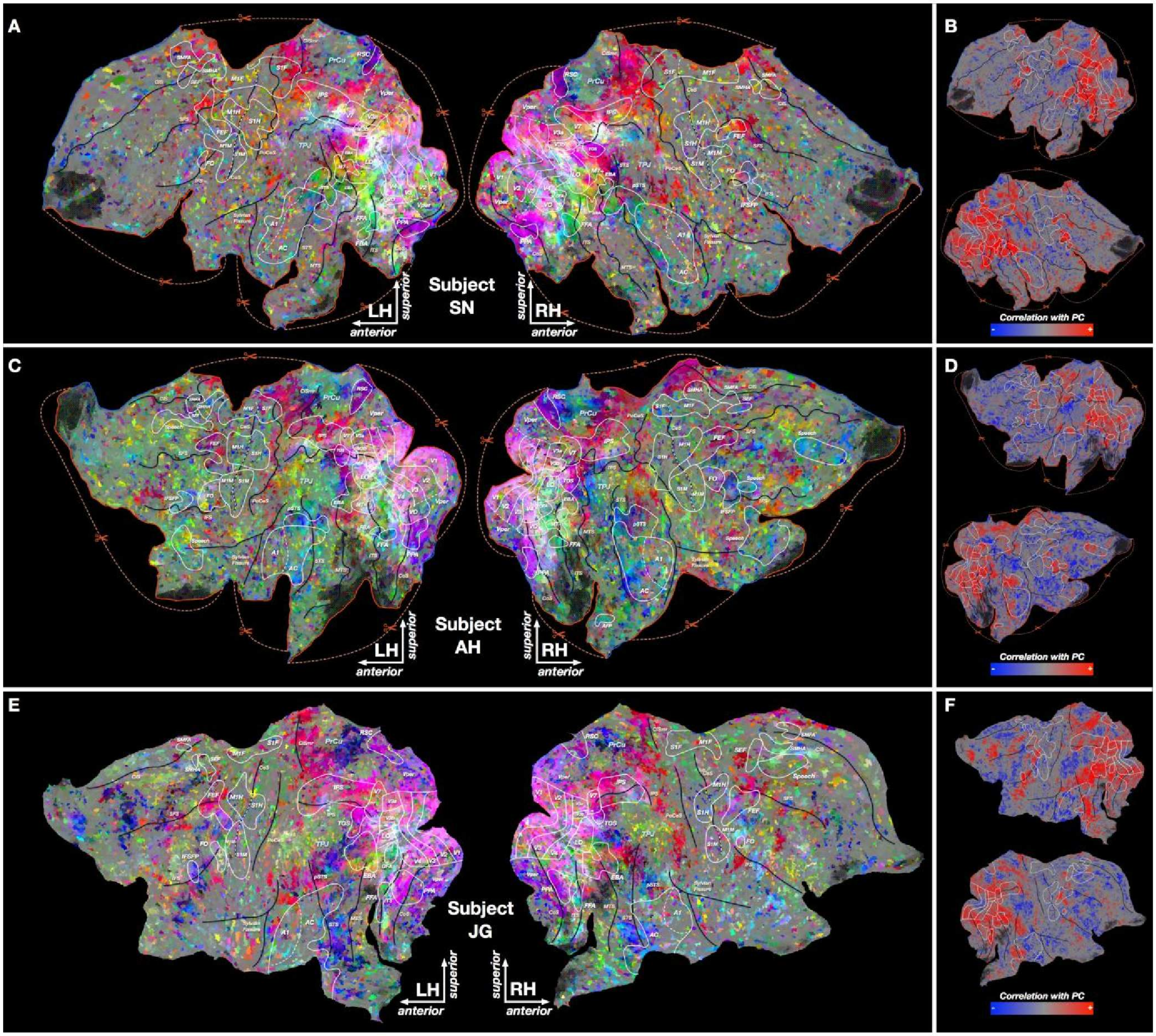


Figure S5 (related to Figure 7). Cortical maps of semantic representation for subjects SN, AH, and JG. Similar to Figure 6 in the main text, this figure shows the semantic space projected onto the cortical surface for subjects SN, AH, and JG. (**A,C,E**) Category model weights for each cortical voxel are projected onto PCs 2-4 of the group semantic space, and then assigned a color according to the scheme described in Figure 4B. These colors are projected onto cortical flat maps constructed separately for each subject. Each location on these maps represents a single voxel. Locations with similar colors have similar semantic selectivity. Brain areas identified using functional localizers are outlined in white and labeled (see Table S1 for details). Uncertain anatomical boundaries are shown as dashed white lines. Major sulci are denoted by dark blue lines and labeled (see Table S2). Some anatomical regions are labeled in light blue (*Abbreviations: PrCu = precuneus; TPJ = temporoparietal junction*). Cuts made to the cortical surface during flattening are indicated by dashed red lines and a red border. Blue borders show the edge of the corpus callosum and subcortical structures. Regions of fMRI signal dropout are shaded with black hatched lines. (**B,D,F**) Projection of voxel model weights onto the first PC. Voxels that are positively correlated with the first PC appear red, while negatively correlated voxels appear blue. Voxels orthogonal to the first PC appear gray.

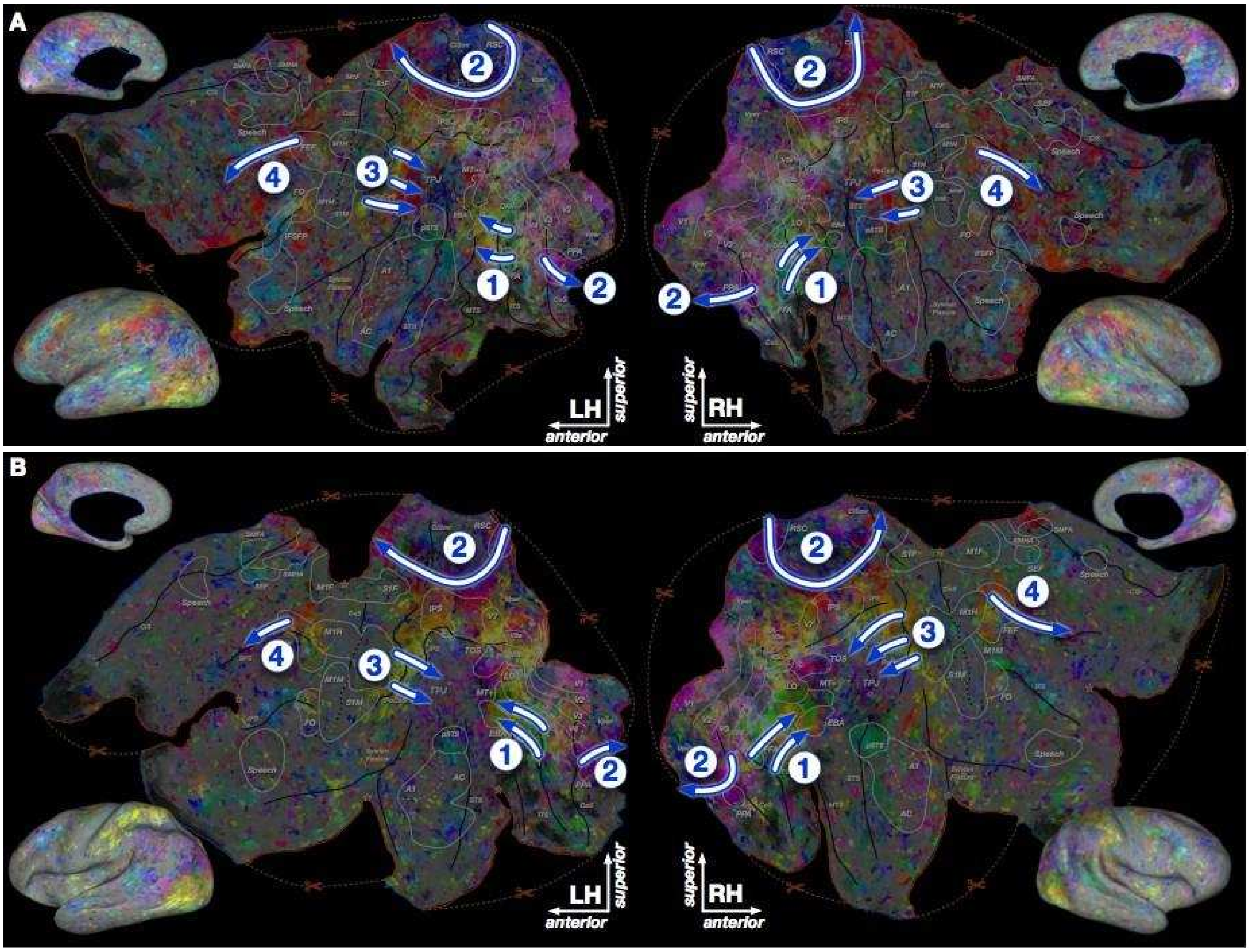


Figure S6 (related to Figure 7). Schematic of semantic gradients. This figure is similar to Figure 7 in the main text, but shows schematically the semantic gradients that appear consistently across subjects. These gradients are described in detail in the main text. (**A**) The cortical flat map constructed for subject AV. Semantic gradients are indicated by blue and white arrows and numbered. **Gradient 1** starts in the posterior inferior temporal sulcus (ITS), which is selective for humans (green and blue-green), continues through a region of animal selectivity (yellow), and ends at the posterior middle temporal sulcus (MTS), which is selective for human actions, athletes, and outdoor spaces (red and green-red). **Gradient 2** starts in a region of the collateral sulcus that is selective for vehicles and landscapes (pink and purple), continues superiorly along the medial wall to retrosplenial cortex (RSC), which is selective for buildings and landscapes (blue-indigo and purple), then continues anteriorly along the superior bank of the intraparietal sulcus (IPS), which is selective for geography and human actions (red-purple and red). Note that this gradient crosses one of the relaxation cuts made to the cortical surface, and so appears discontinuous. **Gradient 3** starts in the inferior postcentral sulcus (PoCeS), which is animal selective (yellow and yellow-green), and ends in the anterior part of the temporoparietal junction (TPJ), which is landscape selective (purple). **Gradient 4** starts in the posterior superior frontal sulcus (SFS), which is selective for human actions (red), and ends in the anterior SFS, which is selective for landscapes (purple). (**B**) The cortical flat map constructed for subject TC with semantic gradients.

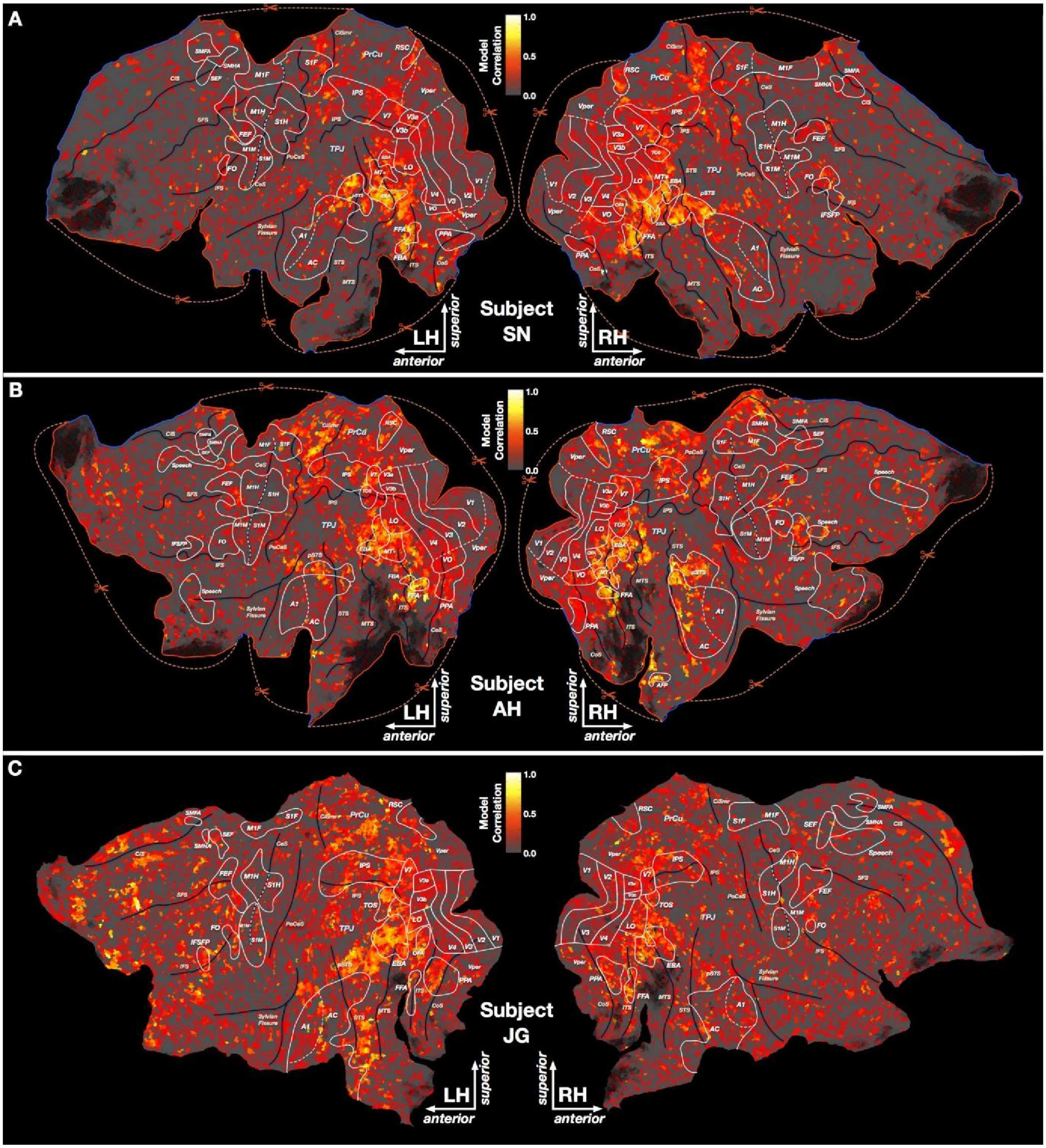


Figure S7 (related to Figure 9). Model prediction performance across the cortical surface for subjects SN, AH, and JG. To determine how much response variance for each voxel is explained by the category model, prediction performance was assessed using the validation data. (**A**) Each location on these maps represents a single voxel in subject SN. Well predicted voxels appear yellow or white, poorly predicted voxels appear gray. Borders and notations on the graph are as described in the earlier Figures. The best predictions are found in occipitotemporal cortex, the posterior superior temporal sulcus, medial parietal cortex, and inferior frontal cortex. (**B**) Model performance for subject AH. (**C**) Model performance for subject JG.

Table S1 (related to Figure 7). Abbreviations, localizers, and references for known functional areas

Name	*Anatomical Location*	*Localizer*	*References*
FFA (fusiform face area)	Posterior fusiform gyrus	Faces – objects contrast	Kanwisher et al., 1997; McCarthy et al., 1997
OFA (occipital face area)	Just anterior to V4v/VO	Faces – objects contrast	Halgren et al., 1999; Kanwisher et al., 1997
IFSFP (inferior frontal sulcus face patch)	IFS anterior to precentral sulcus	Faces – objects contrast	Avidan et al., 2005; Tsao et al., 2008
ATFP (anterior temporal face patch)	Temporal pole	Faces – objects contrast	Rajimehr et al., 2009
EBA (extrastriate body area)	Anterior to MT+ on the medial temporal gyrus	Human bodies – objects contrast	Downing et al., 2001
FBA (fusiform body area)	Fusiform sulcus/gyrus anterior to FFA	Human bodies – objects contrast	Peelen & Downing, 2005; Schwarzlose et al., 2005
PPA (parahippocampal place area)	Collateral fissure	Scenes – objects contrast	Epstein & Kanwisher, 1998
TOS (transverse occipital sulcus)	Just inferior to/overlapping with V7	Scenes – objects contrast	Hasson et al., 2003; K. Nakamura et al., 2000
RSC (retrosplenial cortex)	Medial wall just superior to PPA	Scenes – objects contrast	Aguirre, Zarahn, & D'esposito, 1998
FEF (frontal eye fields)	Precentral sulcus and superior frontal sulcus	Self generated saccades – fixation contrast	Paus, 1996
FO (frontal operculum)	Inferior portion of precentral sulcus	Self generated saccades – fixation contrast	Corbetta et al., 1998
SEF (supplementary eye fields)	Dorsal-medial frontal cortex	Self generated saccades – fixation contrast	Grosbras et al., 1999
Vper (visual periphery, including V1-V4)	Surrounding mapped retinotopic visual cortex	Self generated saccades – fixation contrast	
MT+ (middle temporal)	Posterior inferior temporal sulcus	Coherent – scrambled motion contrast	Tootell et al., 1995
pSTS (posterior superior temporal sulcus)	As it sounds	High auditory and visual repeatability	
A1 (primary auditory cortex)	Heschl's gyri	Auditory repeatability and anatomical location	
AC (auditory cortex)	Superior temporal gyrus	Auditory repeatability	
S1F/M1F (primary somatosensory and motor cortex, foot)	Superior-medial central sulcus	Foot motion – rest contrast, S1F and M1F split at fundus of central sulcus	Penfield & Boldrey, 1937
S1H/M1H (primary somatosensory and motor cortex, hand)	Central sulcus	Hand motion – rest contrast, S1H and M1H split at fundus of central sulcus	Penfield & Boldrey, 1937
S1M/M1M (primary somatosensory and motor cortex, mouth)	Inferior central sulcus	Mouth motion – rest contrast, S1M and M1M split at fundus of central sulcus	Penfield & Boldrey, 1937

Name	***Anatomical Location***	***Localizer***	***References***
SMHA (supplementary motor hand area)	Middle cingulate gyrus	Hand motion – rest contrast	Fried et al., 1991
SMFA (supplementary motor foot area)	Middle cingulate gyrus/sulcus	Foot motion – rest contrast	Fried et al., 1991
IPS (intraparietal sulcus)	Lateral parietal cortex	Retinotopy	
V1-V4, V3A, V3B	Occipital cortex	Retinotopy	Hansen et al., 2007
LO (lat. occipital complex)	Anterior to V4	Retinotopy	Hansen et al., 2007
VO	Inferior to V4v	Retinotopy	Brewer et al., 2005
V7	Anterior to V3A/V3B	Retinotopy	Hansen et al., 2007

Table S2 (related to Figure 7). Abbreviations for anatomical features

Abbreviation	*Full Name*
CoS	Collateral sulcus
ITS	Inferior temporal sulcus
MTS	Middle temporal sulcus
STS	Superior temporal sulcus
IPS	Intraparietal sulcus
CiSmr	Marginal ramus of the cingulate sulcus
PoCeS	Postcentral sulcus
CeS	Central sulcus
IFS	Inferior frontal sulcus
SFS	Superior frontal sulcus
CiS	Cingulate sulcus
PrCu	Precuneus
TPJ	Temporoparietal junction

Table S3 (related to Figure 9). Category model performance within known functional areas

Area	*Hemi.*	*N*	*Avg. Corr.*	*Signif. / Total Voxels*
A1	L	4	0.088	14 / 152
	R	4	0.100	20 / 148
AC (incl. A1)	L	5	0.152	124 / 633
	R	5	0.166	139 / 572
ATFP	L	0	–	–
	R	1	0.298	3 / 8
EBA	L	5	0.449	146 / 172
	R	5	0.430	123 / 158
FBA	L	2	0.367	4 / 7
	R	1	0.525	8 / 10
FEF	L	5	0.164	65 / 216
	R	5	0.165	47 / 167
FFA	L	5	0.387	39 / 61
	R	5	0.441	58 / 72
FO	L	5	0.143	24 / 108
	R	5	0.191	24 / 74
IFSFP	L	3	0.254	21 / 53
	R	3	0.230	37 / 89
IPS	L	5	0.219	163 / 364
	R	5	0.224	147 / 316
LO	L	5	0.286	87 / 117
	R	5	0.311	120 / 165
MT+	L	4	0.378	68 / 82
	R	4	0.429	53 / 59
OFA	L	3	0.322	15 / 24
	R	3	0.312	19 / 26
PPA	L	5	0.263	70 / 123
	R	5	0.298	46 / 68
RSC	L	5	0.239	32 / 75
	R	5	0.244	46 / 107
SEF	L	4	0.110	3 / 29
	R	4	0.110	3 / 34
TOS	L	3	0.327	32 / 58
	R	4	0.322	53 / 75
V1	L	5	0.109	60 / 248
	R	5	0.111	75 / 279
V2	L	5	0.115	92 / 294
	R	5	0.125	111 / 332
V3	L	5	0.125	82 / 243
	R	5	0.147	113 / 281
V3A	L	5	0.205	40 / 69
	R	5	0.236	47 / 74
V3B	L	5	0.208	37 / 67
	R	5	0.236	50 / 77

Area	***Hemi.***	***N***	***Avg. Corr.***	***Signif. / Total Voxels***
V4	L R	5 5	0.160 0.184	85 / 197 101 / 211
V7	L R	5 5	0.292 0.266	79 / 110 95 / 137
VO	L R	2 2	0.204 0.237	16 / 26 27 / 35
Foot (S1/M1)	L R	5 5	0.102 0.083	30 / 220 21 / 209
Hand (S1/M1)	L R	5 5	0.081 0.091	29 / 365 25 / 290
Mouth (S1/M1)	L R	5 5	0.093 0.091	31 / 280 22 / 229
pSTS	L R	4 4	0.348 0.385	63 / 107 95 / 137

Prediction performance of the 1705-category model within known functional areas. For each area in each hemisphere, this table shows the number of subjects in which the area was identified (N), the average correlation coefficient within the area (corrected to account for noise in the validation dataset), the average number of voxels whose activity was predicted significantly (p<0.05 uncorrected), and the average total number of voxels within the area (each voxel is 20.7mm^3 in volume).

Supplemental Experimental Procedures

Functional localizers

Retinotopic localizer. Retinotopic mapping data were collected in four 9-minute scans. Two scans used clockwise and counterclockwise rotating polar wedges, and two used expanding and contracting rings.

Motor localizer. Motor localizer data were collected during one 10-minute scan. The subject was cued to perform six different motor tasks in a random order in 20-second blocks. For the hand, mouth, foot, speech, and rest blocks the stimulus was simply a word at the center of the screen (e.g. "Hand"). For the saccade block the subject was shown a pattern of saccade targets.

For the “Hand” cue the subject was instructed to make small finger-drumming movements with both hands for as long as the cue remained on the screen. Similarly for the “Foot” cue the subject was instructed to make small toe movements for the duration of the cue. For the “Mouth” cue the subject was instructed to make small mouth movements approximating the nonsense syllable *balabalabala* for the duration of the cue--this requires movement of the lips, tongue, and jaw. For the “Speak” cue the subject was instructed to continuously subvocalize self-generated sentences for the duration of the cue. For the saccade condition the written cue was replaced with a fixed pattern of twelve saccade targets, and the subject was instructed to make frequent saccades between the targets.

Area MT+ localizer. Area MT+ localizer data were collected in four 90-second scans consisting of alternating 16-second blocks of continuous and temporally scrambled natural movies.

Visual category localizers. Visual category localizer data were collected in six 4.5-minute scans consisting of 16 blocks, each 16 seconds long. During each block, 20 images of either places, faces, human body parts, non-human animals, household objects, or spatially scrambled household objects were displayed. Each image was displayed for 300 ms followed by a 500 ms blank. Occasionally the same image was displayed twice in a row, in which case the subject was asked to respond with a button press.

Auditory localizer. Auditory cortex localizer data were collected in one 10 minute scan. The subject listened to 10 repeats of a 1-minute auditory stimulus, which consisted of 20-second segments of music, speech, and natural sounds. To determine whether a voxel was responsive to auditory stimuli, the repeatability of the voxel response across the 10 stimulus repeats was calculated using an *F*-statistic.

RGB colormap construction

Principal components analysis (PCA) produces a low-dimensional orthogonal space. While each dimension is not necessarily interpretable on its own, the space as a whole is highly interpretable (see Fig. 5). In order to visualize projections of voxel models and categories in a three-dimensional space we constructed a trivariate colormap. In this colormap each location in the 3-D unit cube is mapped to a unique color. We used this scheme to visualize both category coefficients and model projections into the PC space. However, we do not expect these data to map uniformly onto a cube. Instead, the distribution of model projections tends to be spherical. Mapping spherical data onto the RGB cube is inefficient, as the corners of the color space (where the strongest colors reside) are underutilized. Thus we devised a modified RGB colormap that could efficiently map spherical data onto unique RGB values.

Our modified RGB map can be thought of as a 3-dimensional sphere in which each point represents a unique RGB value. This sphere is derived from the original RGB cube (a unit cube centered at the origin) by the following procedure: the coordinates of each point within the cube are first normalized by their L-infinity norm (the maximum value of the three coordinates) and then multiplied by their L-2, or Euclidean norm. This procedure maps the unit cube onto a unit sphere.

Voxel-wise model fitting and testing

L2-penalized linear least square regression (also known as ridge regression) was used to find weights on the feature channels that best predicted responses on the model estimation data, which consisted of 7200 seconds of stimuli and responses. The following procedure was repeated 15 times: we fit the model using a range of regularization coefficients on all but a randomly selected 500 seconds of model estimation data (for a total of 6700 seconds of training data). Using the weights found for each regularization coefficient we predicted the responses to the held-out 500 seconds of data and computed the correlation between actual and predicted responses separately for each voxel. Once this was done 15 times, the test correlations for each voxel and regularization coefficient were averaged across the 15 repetitions. The best regularization coefficient was then selected for each voxel.

Finally, we used all 7200 seconds of training data and the selected regularization coefficient for each voxel to fit the model. To obtain a single weight for each category and each voxel, the weights for the three delays were averaged. The resulting weights were used in all subsequent analyses.

To determine model performance we generated predictions for the 540 seconds (270 samples) of model validation data that were not used for model estimation. We then found the correlation between predicted and mean response for each voxel across the ten repetitions of the validation stimuli. To test whether a voxel was predicted significantly above chance level we used a bootstrap procedure. The 270 time points in the validation data were resampled with replacement 10,000 times and the correlation between resampled predicted and resampled actual responses was computed for each sample. The *p*-value of the correlation was computed as the fraction of samples for which the correlation was less than zero; under the null hypothesis of no correlation this would yield *p*=0.5. The voxels shown in Figure 2 have very high correlations (0.530 and 0.659) and *p*-values too small to probe using this bootstrap method. To assign *p*-values to these correlations we used an exact formula to compute the *p*-value of the correlation between two Gaussian random variables.

While the correlation between predicted response and actual mean response is an appropriate metric for assessing significance, it is biased downward due to noise in the validation data (David & Gallant, 2005; Hsu et al., 2004; Sahani & Linden, 2003). Because the actual mean response is calculated using a finite number of repetitions (in this case 10) it contains residual noise in addition to signal. This noise level is likely to vary across voxels due to vascularization and magnetic field inhomogeneity. We accounted for noise in the validation data using the method developed in Hsu et al., 2004, in which the raw correlation is divided by the expected maximum possible model correlation (called the *noise ceiling*) for each voxel. For very noisy voxels, however, this method led to divergent correlation estimates. To correct this issue we limited voxel noise ceilings to be above some value *k*. For *k*=1, the estimated actual correlation is the observed correlation between response and prediction, and for *k*=0 the estimated actual correlation is the original divergent estimate. We set *k* to 0.10, which is the *p*<0.05 significance threshold for the correlation of two gaussian variables of length 270.

Significance testing of principal components

If there is no structured semantic space underlying the true model weights (i.e. the weights for each voxel are independent from the other voxels) then the PCs of the estimated model weights will be identical to the PCs of the stimulus matrix. This bias in the estimated weight PCs is due to the regularized regression procedure, which trades a small increase in bias for a large decrease in error (Hoerl & Kennard, 1970). Thus in order to appropriately evaluate statistical significance of the estimated model weight PCs we compared them to the stimulus PCs. This significance criterion ensures that the semantic structure that we observe in the PCs is due primarily to the fMRI data and not the statistics of category co-occurrence in the stimuli.

We first tested whether each individual-subject model weight PC accounted for more variance than would be expected by chance. To find confidence intervals on the variance accounted for by each PC we bootstrapped the model weight PCA by sampling with replacement from the voxel population 1000 times. Similarly, confidence intervals on the variance in model weights accounted for by each stimulus PC were obtained by bootstrapping the stimulus PCA 1000 times. The amount of variance accounted for in the model weights by each of the model weight PCs (orange lines) and stimulus PCs (gray lines) is shown in Figure 3, along with error bars denoting 99% confidence intervals. To test the hypothesis that a model weight PC accounts for more variance than the corresponding stimulus PC we counted the number of times in the 1000 bootstrap samples that the stimulus PC accounted for more variance than the model weight PC. The null hypothesis for this analysis is that the stimulus PC and the model weight PC account for the same amount of variance. We rejected the null hypothesis if the stimulus PC never accounted for more variance than the voxel weight PC across the 1000 bootstrap samples (corresponding to $p<0.001$).

Because lower-variance PCs are more sensitive to noise and thus more likely to yield false positives, we tested the PCs sequentially and stopped testing after encountering the first non-significant PC. This procedure revealed that subject SN has 6 significant PCs (which account for 29.6% of variance), AH has 7 significant PCs (which account for 33.1% of variance), AV has 7 significant PCs (which account for 35.5% of variance), TC has 7 significant PCs (which account for 34.3% of variance), and JG has 8 significant PCs (which account for 34.2% of variance).

Next we tested PCs constructed using combined data from many subjects. For each subject we constructed a set of group PCs using combined data from the other four subjects, leaving out the selected subject. For example, to test subject SN we performed PCA on combined model weights from subjects AH, AV, TC, and JG. We then computed the amount of variance accounted for in the model weights for the left out subject by each of the group PCs.

As with the individual subject PCs and stimulus PCs, confidence intervals on the variance explained by the group PCs were found using the bootstrap. We then tested whether each group PC explained more variance than the corresponding stimulus PC using the statistical procedure described above. We found that subjects SN, AH, AV, and TC were significantly explained by 4 group PCs (which accounted for 19.1%, 17.3%, 21.6%, and 20.6% of variance, respectively), and subject JG was significantly explained by 3 group PCs (which accounted for 15.4% of variance).

This analysis suggests that the five subjects share a common semantic space consisting of at least the first three group PCs, and four of the five subjects share four group PCs. We estimated the full group semantic space using PCA on combined data from all five subjects (49685 voxels total). The data were never spatially averaged across subjects, and never transformed into a standard functional brain space.

Supplemental References

Aguirre, G. K., Zarahn, E., & D'esposito, M. (1998). An Area within Human Ventral Cortex Sensitive to Building Stimuli: Evidence and Implications. *Neuron*, *21*(2), 373-383.

Avidan, Hasson, U., Malach, R., & Behrmann, M. (2005). Detailed exploration of face-related processing in congenital prosopagnosia: 2. Functional neuroimaging findings. *J Cogn Neurosci, 17(7),* 1150-1167.

Brewer, A. a, Liu, J., Wade, A. R., & Wandell, B. a. (2005). Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nat Neurosci*, *8*(8), 1102-9. doi:10.1038/nn1507

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., Linenweber, M. R., et al. (1998). A common network of functional areas for attention and eye movements. *Neuron*, *21*(4), 761-773.

David, S. V., & Gallant, J. L. (2005). Predicting neuronal responses during natural vision. *Network: Computation in Neural Systems*, *16*(2-3), 239-260.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*(5539), 2470.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598-601.

Fried, I., Katz, A., McCarthy, G., Sass, K., Williamson, P., Spencer, S., & Spencer, D. (1991). Functional Organization of Human Supplementary Motor Cortex Studied by Electrical Stimulation Motor Cortex. *J Neurosci*, *11*(November), 3656.

Grosbras, M. H., Lobel, E., de Moortele, P. F. V., LeBihan, D., & Berthoz, A. (1999). An anatomical landmark for the supplementary eye fields in human revealed with functional magnetic resonance imaging. *Cereb Cortex*, *9*(7), 705.

Halgren, E., Dale, A. M., Sereno, M. I., Tootell, R. B. H., Marinkovic, K., & Rosen, B. R. (1999). Location of human face-selective cortex with respect to retinotopic areas. *Hum Brain Mapp*, *7*(1), 29-37.

Hansen, K. a, Kay, K. N., & Gallant, J. L. (2007). Topographic organization in and near human visual area V4. *J Neurosci*, *27*(44), 11896-911. doi:10.1523/JNEUROSCI.2991-07.2007

Hasson, U., Harel, M., Levy, I., & Malach, R. (2003). Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron*, *37*(6), 1027-1041.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, *12*(1), 55-67.

Hsu, A., Borst, A., & Theunissen, F. (2004). Quantifying variability in neural responses and its application for the validation of model predictions. *Network: Computation in Neural Systems*, *15*(2), 91-109. doi:10.1088/0954-898X/15/2/002

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J Neurosci*, *17*(11), 4302-4311.

McCarthy, Gregory, Puce, A., Gore, J. C., & Allison, T. (1997). Face-Specific Processing in the Human Fusiform Gyrus. *J Cogn Neurosci*, *9*(5), 605-610. MIT Press. doi:10.1162/jocn.1997.9.5.605

Nakamura, K., Kawashima, R., Sato, N., Nakamura, A., Sugiura, M., Kato, T., Hatano, K., et al. (2000). Functional delineation of the human occipito-temporal areas related to face and scene processing. *Brain*, *123*(9), 1903.

Paus, T. (1996). Location and function of the human frontal eye-field: a selective review. *Neuropsychologia*, *34*(6), 475-483.

Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *J Neurophysiol*, *93*(1), 603.

Penfield, W., & Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain*, *60*(4), 389.

Rajimehr, R., Young, J. C., & Tootell, R. B. H. (2009). An anterior temporal face patch in human cortex, predicted by macaque maps. *Proc Natl Acad Sci U S A*, *106*(6), 1995.

Sahani, M., & Linden, J. F. (2003). How linear are auditory cortical responses. *Advances in Neural Information Processing Systems*, *15*, 109-116.

Schwarzlose, R., Baker, C., & Kanwisher, N. G. (2005). Separate face and body selectivity on the fusiform gyrus. *J Neurosci*, *25*(47), 11055-11059.

Tootell, R. B., Reppas, J. B., Kwong, K. K., Malach, R., Born, R. T., Brady, T. J., Rosen, B. R., et al. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J Neurosci*, *15*(4), 3215.

Tsao, D. Y., Moeller, S., & Freiwald, W. A. (2008). Comparing face patch systems in macaques and humans. *Proc Natl Acad Sci U S A*, *105*(49), 19514.