Neuron, Volume *72*

Supplemental Information

A Common, High-Dimensional Model of the Representational Space in Human Ventral Temporal Cortex

James V. Haxby, J. Swaroop Guntupalli, Andrew C. Connolly, Yaroslav O. Halchenko, Bryan R. Conroy, M. Ida Gobbini, Michael Hanke, and Peter J. Ramadge

Supplemental Data

A

$$M = VP \quad (1)$$

M		Principal components				
		pc_1	pc_2	pc_3	...	pc_{35}
Time-points	t_1	$y_{1,1}$	$y_{2,1}$	$y_{3,1}$	...	$y_{35,1}$
	t_2	$y_{1,2}$	$y_{2,2}$	$y_{3,2}$	...	$y_{35,2}$
	t_3	$y_{1,3}$	$y_{2,3}$	$y_{3,3}$	...	$y_{35,3}$
	...	...	...	...		...
	t_{2205}	$y_{1,2205}$	$y_{2,2205}$	$y_{3,2205}$	...	$y_{35,2205}$

=

V		Voxels				
		v_1	v_2	v_3	...	v_{1000}
Time-points	t_1	$x_{1,1}$	$x_{2,1}$	$x_{3,1}$	...	$x_{1000,1}$
	t_2	$x_{1,2}$	$x_{2,2}$	$x_{3,2}$	...	$x_{1000,2}$
	t_3	$x_{1,3}$	$x_{2,3}$	$x_{3,3}$	...	$x_{1000,3}$
	...	...	...	...		...
	t_{2205}	$x_{1,2205}$	$x_{2,2205}$	$x_{3,2205}$	...	$x_{1000,2205}$

x

P		Principal components				
		pc_1	pc_2	pc_3	...	pc_{35}
Voxels	v_1	$w_{1,1}$	$w_{2,1}$	$w_{3,1}$	...	$w_{35,1}$
	v_2	$w_{1,2}$	$w_{2,2}$	$w_{3,2}$	...	$w_{35,2}$
	v_3	$w_{1,3}$	$w_{2,3}$	$w_{3,3}$	...	$w_{35,3}$
	...	...	...	...		...
	v_{1000}	$w_{1,1000}$	$w_{2,1000}$	$w_{3,1000}$	...	$w_{35,1000}$

B

$$M_{\exp 2} = V_{\exp 2} P \quad (2)$$

M_{exp2}		Principal components				
		pc_1	pc_2	pc_3	...	pc_{35}
Stimuli	s_1	$y_{1,1}$	$y_{2,1}$	$y_{3,1}$	...	$y_{35,1}$
	s_2	$y_{1,2}$	$y_{2,2}$	$y_{3,2}$	...	$y_{35,2}$
	s_3	$y_{1,3}$	$y_{2,3}$	$y_{3,3}$	...	$y_{35,3}$
	...	...	...	...		...
	s_N	$y_{1,N}$	$y_{2,N}$	$y_{3,N}$	...	$y_{35,N}$

=

V_{exp2}		Voxels				
		v_1	v_2	v_3	...	v_{1000}
Stimuli	s_1	$x_{1,1}$	$x_{2,1}$	$x_{3,1}$	...	$x_{1000,1}$
	s_2	$x_{1,2}$	$x_{2,2}$	$x_{3,2}$	...	$x_{1000,2}$
	s_3	$x_{1,3}$	$x_{2,3}$	$x_{3,3}$	...	$x_{1000,3}$
	...	...	...	...		...
	s_N	$x_{1,N}$	$x_{2,N}$	$x_{3,N}$	...	$x_{1000,N}$

x

P		Principal components				
		pc_1	pc_2	pc_3	...	pc_{35}
Voxels	v_1	$w_{1,1}$	$w_{2,1}$	$w_{3,1}$	...	$w_{35,1}$
	v_2	$w_{1,2}$	$w_{2,2}$	$w_{3,2}$	...	$w_{35,2}$
	v_3	$w_{1,3}$	$w_{2,3}$	$w_{3,3}$	...	$w_{35,3}$
	...	...	...	...		...
	v_{1000}	$w_{1,1000}$	$w_{2,1000}$	$w_{3,1000}$	...	$w_{35,1000}$

C

P		Principal components				
		pc_1	pc_2	pc_3	...	pc_{35}
Voxels	v_1	$w_{1,1}$	$w_{2,1}$	$w_{3,1}$	...	$w_{35,1}$
	v_2	$w_{1,2}$	$w_{2,2}$	$w_{3,2}$	...	$w_{35,2}$
	v_3	$w_{1,3}$	$w_{2,3}$	$w_{3,3}$	...	$w_{35,3}$
	...	...	...	...		...
	v_{1000}	$w_{1,1000}$	$w_{2,1000}$	$w_{3,1000}$	...	$w_{35,1000}$

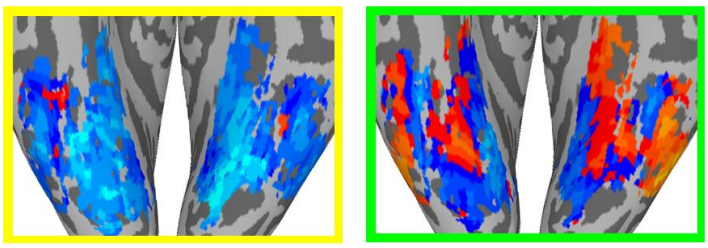


Figure S1, Parts A, B, & C

D

$$R_{topography} = M(t_i)P^T \quad (3)$$

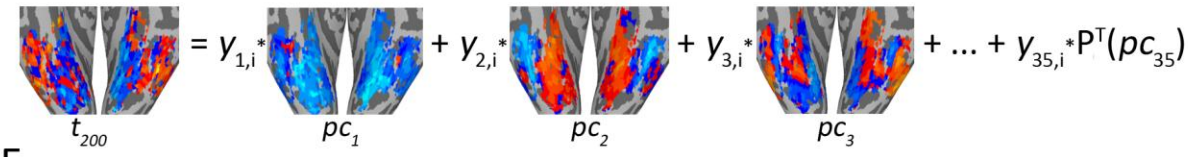
R_{topo}	Voxels				
	v_1	v_2	v_3	...	v_{1000}
t_i	w_1	w_2	w_3	...	w_{1000}

=

M	Principal components				
Time-points	pc_1	pc_2	pc_3	...	pc_{35}
t_1	$y_{1,1}$	$y_{2,1}$	$y_{3,1}$	...	$y_{35,1}$
t_2	$y_{1,2}$	$y_{2,2}$	$y_{3,2}$	...	$y_{35,2}$
t_i	$y_{1,i}$	$y_{2,i}$	$y_{3,i}$	...	$y_{35,i}$
...	...	...	...	...	...
t_{2205}	$y_{1,2205}$	$y_{2,2205}$	$y_{3,2205}$	...	$y_{35,2205}$

x

P^T	Voxels				
PCs	v_1	v_2	v_3	...	v_{1000}
pc_1	$w_{1,1}$	$w_{2,1}$	$w_{3,1}$	...	$w_{1000,1}$
pc_2	$w_{1,2}$	$w_{2,2}$	$w_{3,2}$	...	$w_{1000,2}$
pc_3	$w_{1,3}$	$w_{2,3}$	$w_{3,3}$	...	$w_{1000,3}$
...	...	...	...	...	...
pc_{35}	$w_{1,35}$	$w_{2,35}$	$w_{3,35}$	...	$w_{1000,35}$



E

$$R_{voxel-time-series} = P(v_j)M^T \quad (4)$$

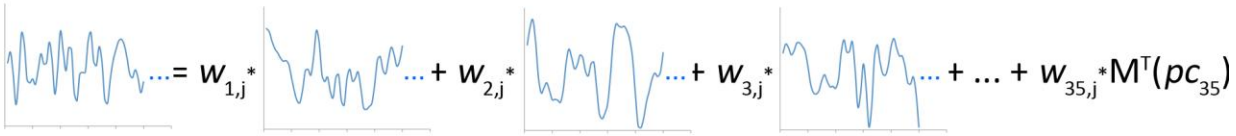
R_{vox_ts}	Time-points				
	t_1	t_2	t_3	...	t_{2205}
v_j	x_1	x_2	x_3	...	x_{2205}

=

P	Principal components				
Voxels	pc_1	pc_2	pc_3	...	pc_{35}
v_1	$w_{1,1}$	$w_{2,1}$	$w_{3,1}$	...	$w_{35,1}$
v_2	$w_{1,2}$	$w_{2,2}$	$w_{3,2}$	...	$w_{35,2}$
v_j	$w_{1,j}$	$w_{2,j}$	$w_{3,j}$	...	$w_{35,j}$
...	...	...	...	...	...
v_{1000}	$w_{1,1000}$	$w_{2,1000}$	$w_{3,1000}$	...	$w_{35,1000}$

x

M^T	Time-points				
PCs	t_1	t_2	t_3	...	t_{2205}
pc_1	$y_{1,1}$	$y_{2,1}$	$y_{3,1}$	...	$y_{2205,1}$
pc_2	$y_{1,2}$	$y_{2,2}$	$y_{3,2}$	...	$y_{2205,2}$
pc_3	$y_{1,3}$	$y_{2,3}$	$y_{3,3}$	...	$y_{2205,3}$
...	...	...	...	...	...
pc_{35}	$y_{1,35}$	$y_{1,35}$	$y_{1,35}$	...	$y_{2205,35}$



F

$$C_{topography} = CP^T \quad (5)$$

C_{topo}	Voxels				
	v_1	v_2	v_3	...	v_{1000}
	x_1	x_2	x_3	...	x_{1000}

=

C	Principal components				
C_{faces_obj}	pc_1	pc_2	pc_3	...	pc_{35}
	y_1	y_2	y_3	...	y_{35}

x

P^T	Voxels				
PCs	v_1	v_2	v_3	...	v_{1000}
pc_1	$w_{1,1}$	$w_{2,1}$	$w_{3,1}$	...	$w_{1000,1}$
pc_2	$w_{1,2}$	$w_{2,2}$	$w_{3,2}$	...	$w_{1000,2}$
pc_3	$w_{1,3}$	$w_{2,3}$	$w_{3,3}$	...	$w_{1000,3}$
...	...	...	...	...	...
pc_{35}	$w_{1,35}$	$w_{2,35}$	$w_{3,35}$	...	$w_{1000,35}$

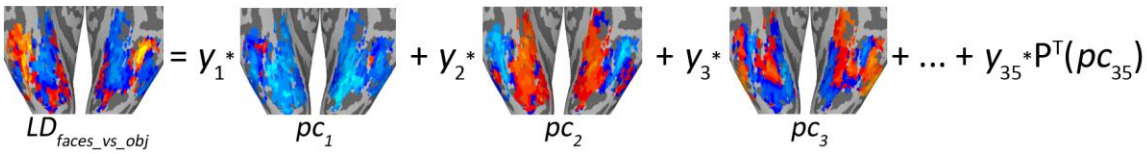


Figure S1, Parts D, E, & F

Figure S1. Related to Figure 1.

(A) The movie data from an individual subject in voxel space (V) are transformed into model space (M) by multiplying V times the reduced matrix of hyperalignment parameters (P)(equation 1). V has 1000 columns (voxels) and 2205 rows (time-points). The number of rows was reduced by half for BSC of movie time-segments, for which the model space was derived on only half of the data. P has 35 columns (PCs) and 1000 rows (voxels). M has 35 columns (PCs) and 2205 rows (time-points). We illustrate the models with the number of voxels that we present in the paper for our principal analyses and with 35 PCs, but we have recalculated BSC (both for movie-aligned and anatomically-aligned data) and WSC for a wide range of voxel set sizes, including an analysis of all VT voxels in all subjects, and we have recalculated BSC for a wide range of PC sets (see Supplemental Figure S2). Each subject's parameter matrix P is derived by calculating the Procrustean transformation that transforms a subject's voxel space data to the group mean model space data, then applying the rotation for the PCA of group mean data and reducing the number of columns to the selected number of PCs. We calculated the group mean model space data through two iterations of pairwise Procrustean transformations (see Results and Methods).

(B) Voxel data from other experiments (V_{exp2}) can be transformed into the same model space coordinate system, producing a new model space data matrix M_{exp2}, by multiplying V_{exp2} by the same reduced hyperalignment parameter matrix (P) that is used to transform movie data (equation 2). V_{exp2} can have any number of stimuli or time-points (N). We denote the rows with 's' to indicate that these response pattern vectors can be either single time-points or calculated responses to specific stimuli over multiple time-points, e.g. by averaging or deconvolution.

(C) The weights in each column of an individual subject's parameter matrix (P) are voxel weights that can be plotted in that subject's voxel space to examine the cortical topography of each PC. The cortical topographies for two PCs (PC1 and PC3) are displayed on one subject's inflated cortex (subject 1; compare with Figure 5 and Supplemental Figure S5).

(D) The cortical topography for the model space response pattern vector to any stimulus can be displayed in any individual subject's voxel space. A response pattern vector to a time-point or stimulus is modeled as a weighted sum of the patterns for each PC. The weights are in one row of a model space data matrix (M or M_{exp2}). The voxel responses (w) in the pattern of response ($R_{topography}$) are calculated by multiplying the row for that stimulus (t_i) in a model space data matrix ($M(t_i)$) by the transpose of that subject's parameter matrix (P^T)(equation 3).

(E) The time-series responses for PCs in a model data matrix (M or M_{exp2}) are basis functions for modeling voxel time-series responses in individual subjects. A voxel time-series response ($R_{voxel\text{-}time\text{-}series}$) is the product of the weights in that voxel's row (v_j) in a subject's parameter matrix (P) and the transpose of a model data matrix (M^T or $M_{exp2}{}^T$)(equation 4). Thus, the variety of individual voxel time-series is not modeled simply as a closed set of 35 types, but rather as an unlimited variety of weighted sums of the 35 PC time-series.

(F) The model space can be explored further to examine the cortical topography associated with differential responses to specific stimuli. The dimension that is defined by a contrast is calculated as a linear discriminant, a line between the mean response vectors for two stimuli in the model space. The voxel weights (w) for the cortical topography associated with a contrast ($C_{topography}$) are calculated by multiplying the transpose of the vector of PC weights for the linear discriminant (C) times the transpose of a subject's parameter matrix (P^T)(equation 5). Here we illustrate this method by showing the topography associated with the contrast between responses to faces and objects (compare with Figure 6 and Supplemental Figure S5).

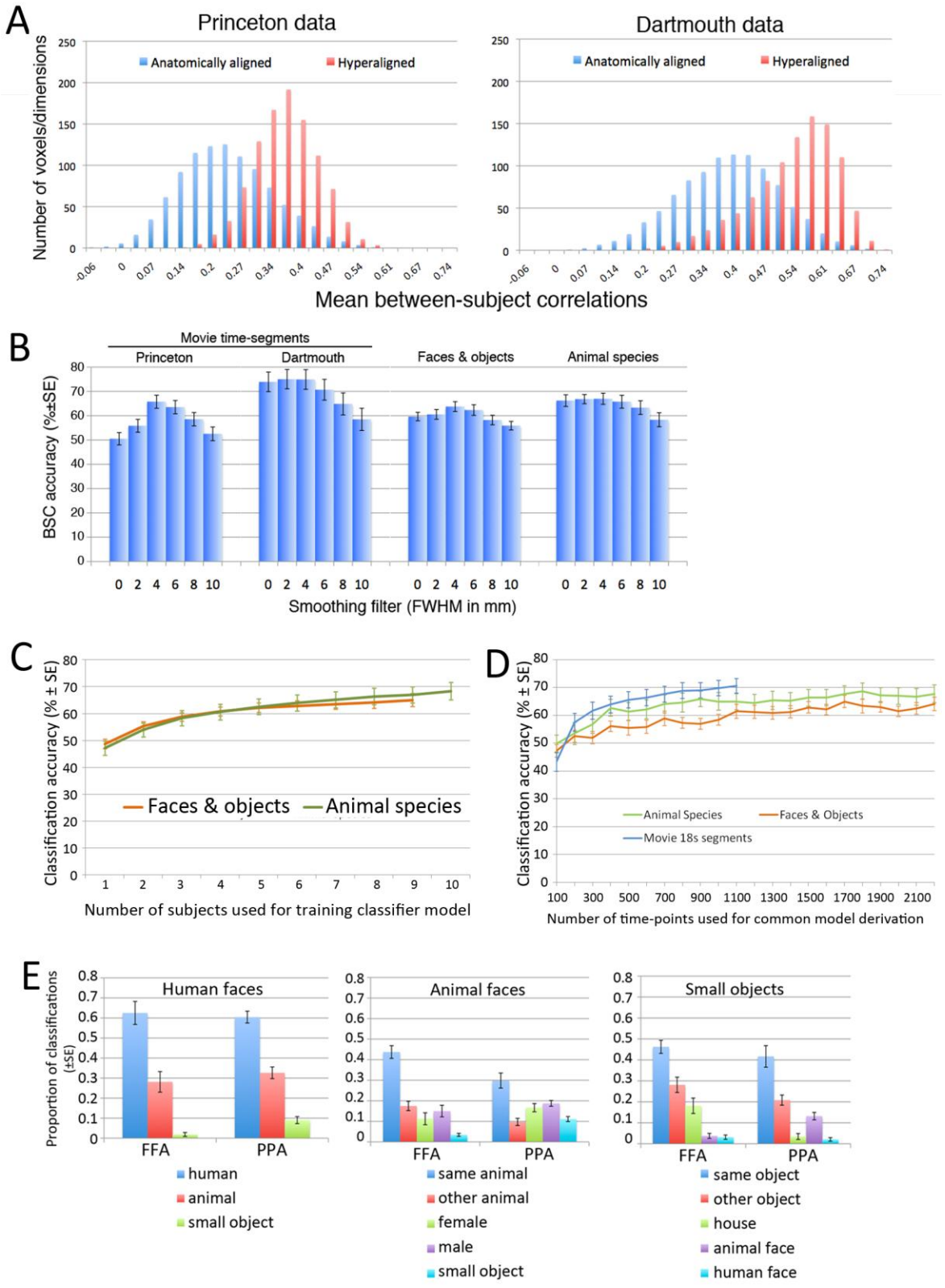

A
Princeton data
Dartmouth data
Anatomically aligned
Hyperaligned
Number of voxels/dimensions
Mean between-subject correlations
B
Movie time-segments
Princeton
Dartmouth
Faces & objects
Animal species
BSC accuracy (%±SE)
Smoothing filter (FWHM in mm)
C
Classification accuracy (% ± SE)
Faces & objects
Animal species
Number of subjects used for training classifier model
D
Classification accuracy (% ± SE)
Animal Species
Faces & Objects
Movie 18s segments
Number of time-points used for common model derivation
E
Proportion of classifications (±SE)
Human faces
Animal faces
Small objects
FFA
PPA
human
animal
small object
same animal
other animal
female
male
small object
same object
other object
house
animal face
human face

Figure S2

Figure S2. Related to Figure 2.

(A) Histograms of between-subject correlations of movie time-series for each voxel, in anatomically-aligned data, and each dimension, in hyperaligned data. For each voxel or dimension, we calculated the mean correlation for each subject with other subjects' time-series for the same voxel or dimension. Note that correlations were higher overall for the Dartmouth data as compared to the Princeton data, reflecting higher-signal-to-noise presumably due to scanner differences, such as the use of an 8-channel phased array coil as compared to a standard 'bird-cage' head coil. The median correlations for anatomically-aligned data were 0.23 (Princeton) and 0.40 (Dartmouth), and for hyperaligned data were 0.37 (Princeton) and 0.55 (Dartmouth). Mean values of r-squared for anatomically-data were 0.069 (Princeton) and 0.167 (Dartmouth) and for hyperaligned data were 0.141 and 0.293. This nearly two-fold increase in variance shared across subjects was due to both individual-specific voxel selection and hyperalignment. Smoothing had a similar effect on WSC of the category experiment data, with maximal accuracy at 4 mm (FWHM) smoothing for both experiments.

(B) The effect of smoothing on BSC of hyperaligned data (1000 voxels). The optimal level of smoothing was 4 mm (FWHM) for all data sets. The advantage of 4 mm smoothing over no smoothing was greater for the Princeton data than for the Dartmouth data, presumably because the Princeton data had more noise, which smoothing suppresses. Note also that the deleterious effect of excessive smoothing (> 4 mm) was greatest for BSC of movie time-segments, presumably reflecting greater dependence of this more demanding MVP classification on high spatial frequency features in patterns of response.

(C) Effect of number of subjects used for training the classifier. BSC of category perception experiments with varying number subjects used for training the classifier model. We repeated the BSC of data in the common model space with sets of 1 to 9 subjects for face & object classification and of 1 to 10 subjects for animal species classification to train the classifier. Increasing the number of subjects for training increases between subject classification accuracies, and both accuracy curves didn't asymptote for the maximum number of subjects tested, suggesting that we might get even better BSC accuracies using more subjects.

(D) Effect of number of time points used for deriving the hyperalignment. We derived the hyperalignment parameters using only a subset of continuous time-points in the movie varying in size from 100 to 2100 time-points and performed BSC of category perception experiments and 18s movie segments in the hyperaligned space using those parameters. Movie segment classification has been performed on each half of the movie data using at most 1100 time-points from the other half for deriving the hyperalignment parameters. BSC accuracies for all three experiments increase with increasing number of time points.

(E) BSC results of face & object categories using hyperaligned data from within the FFA or PPA. We derived the hyperalignment parameters to align voxel spaces within the FFA and PPA separately based on the movie data for 10 subjects who had these functionally-define ROIs and then applied them to the data from the face & object experiment. In each ROI, we used the subject with most voxels as our initial reference and skipped the voxel selection step. Average BSC accuracies for the hyperaligned data were 42.0%±1.6% and 42.3%±2.2% for the FFA and PPA, respectively. BSC for comparably small set of voxels from all of VT (100 top ranking voxels in our voxel selection method) was 44.6%±2.6%. Between-subject pattern classification analyses of hyperaligned data from only the FFA and from only the PPA both significantly discriminated the responses to faces from each other and the responses to man-made objects from each other, as shown in these figures. Each panel indicates the category predictions of the classifier model for particular target categories. Subordinate categories of faces and small objects could be classified within FFA and PPA indicating finer-scale information than just a simple discrimination between faces and objects/houses. Hyperalignment captures these subtle discriminations that are

based on voxel response biases that vary on a spatial scale finer than that of these category-selective regions.

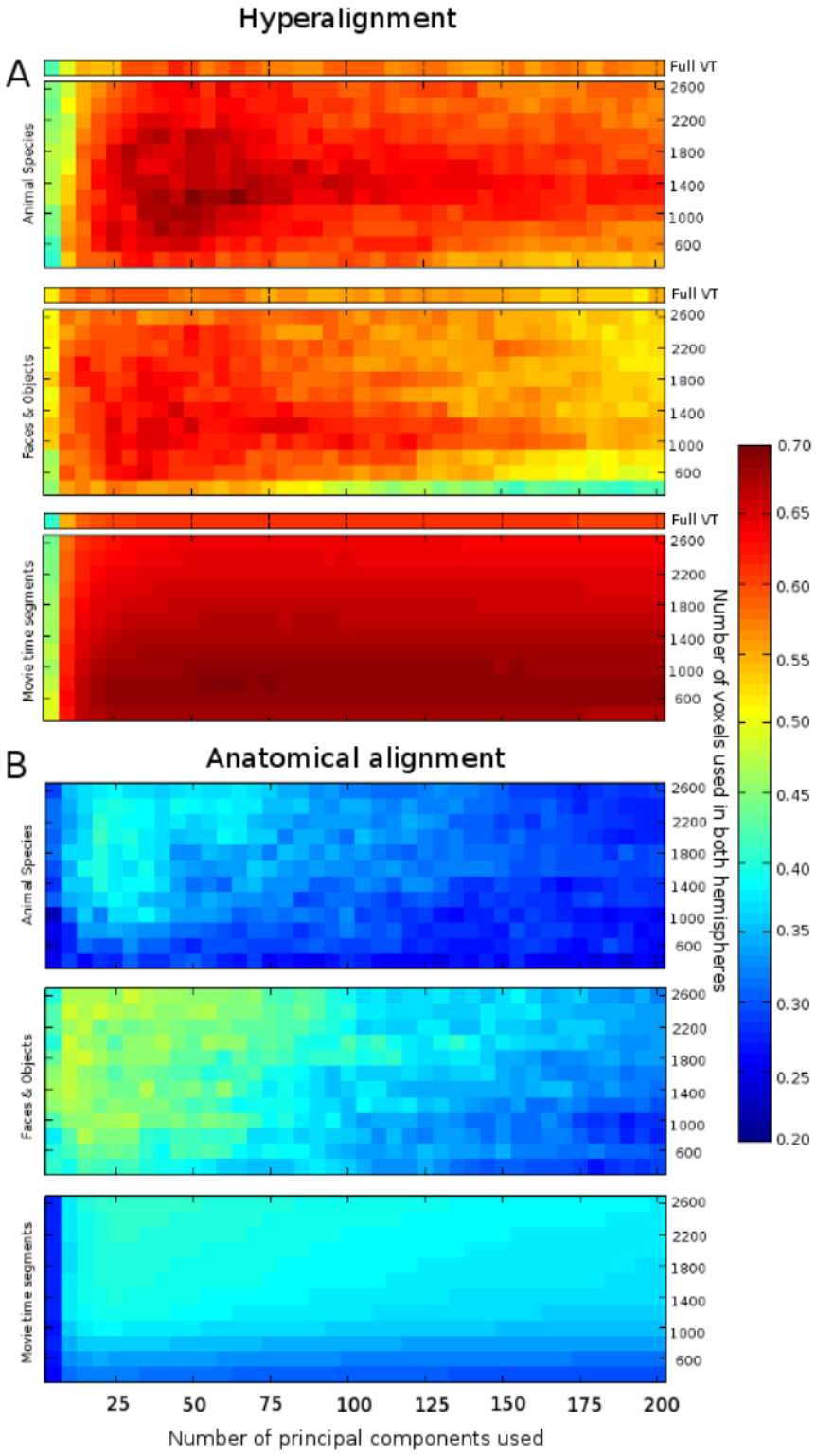
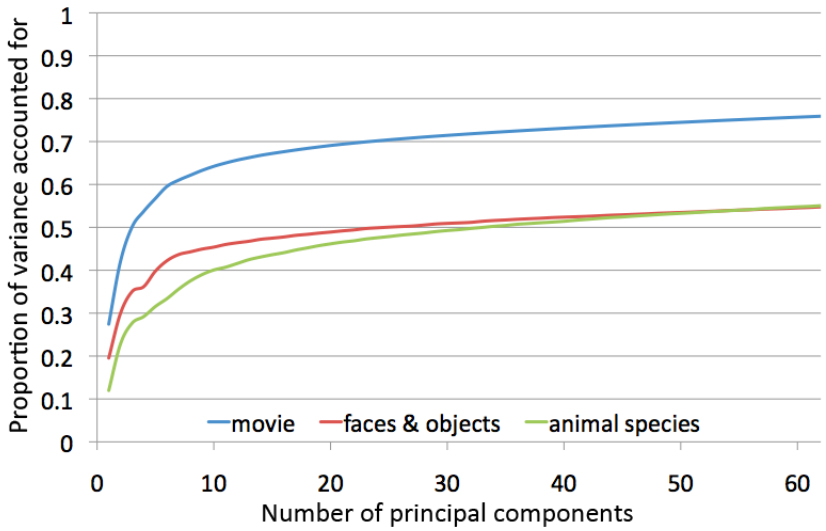


Figure S3, parts A and B

C



D

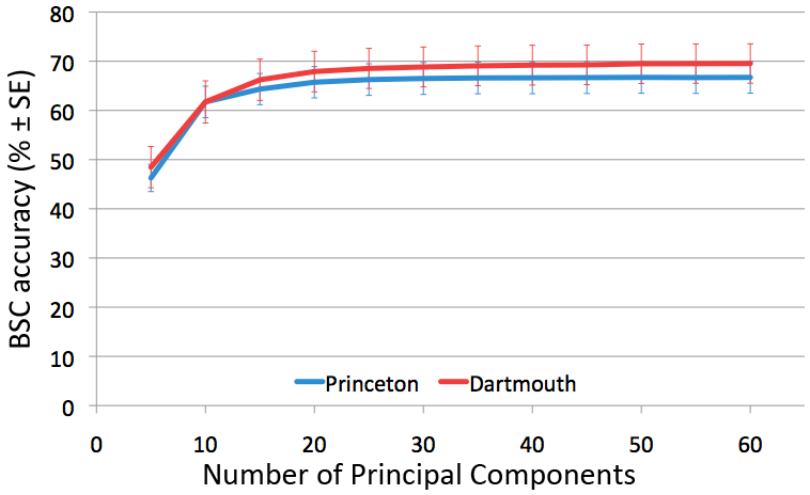


Figure S3, parts C & D

Figure S3. Related to Figure 3.

(A) BSC accuracies of movie time segments, face & object categories, and animal species categories using the data that is hyperaligned and mapped into the common model space derived from the movie data using variable number of voxels, including all VT voxels (top line), and PCs. Each square represents the accuracy using a particular number of top-ranked voxels and most-variance-explaining PCs. For a given number of voxels and PCs, we derived the hyperalignment parameters with that many top-ranked voxels and used that many PCs in our common model space. For BSC of movie segments, we used only half of the movie to derive our common model space. For BSC of category perception experiments, we used the data from full movie. Between-subject classification accuracies for all three studies remained stable from voxel-set sizes of 600 to 2400 and from 30 PCs to 175 PCs. Accuracies peak at 69.3% for 800 voxels and 65 PCs for the movie time segment classification, 65.71% for 1400 voxels and 45 PCs for the face & object classification and at 69.7% for 1200 voxels and 65 PCs for the animal species classification.

(B) BSC accuracies of movie time segments, face & object categories, and animal species categories using the data that is anatomically aligned and mapped into the PC space derived from the anatomically aligned movie data. Each square represents the accuracy using a particular number of top-ranked voxels and most-variance-explaining PCs. For a given numbers of voxels and PCs, we used that many top-ranked voxels to perform the PCA and used that many PCs. For BSC of movie segments, we used only half of the movie to derive our common model space. For BSC of category perception experiments, we used the data from full movie. Between subject classification accuracies tops at 40.82% for 2600 voxels and 35 PCs for the movie time segments, at 48.57% for 2000 voxels and 10 PCs for the face & object categories and at 41.52% for 2000 voxels and 20 PCs for the animal species categories. BSC accuracies only become stable after 1000 voxels and between 10 to 70PCs. Since anatomical alignment only captures coarse topographies, information is coded in fewer PCs compared to hyperaligned data in the common model space.

(C) Proportion of variance accounted for as a function of the number of top PCs from PCA of movie data. For analysis of variance accounted for in the movie data, the PCA was performed on half of the movie and variance accounted for was calculated from the other half of the movie.

(D) BSC accuracy for movie time-segment data that was modeled within research centers (Princeton and Dartmouth). Note that the shapes of the curves are very similar, suggesting that over 20 dimensions are required to capture the fine distinctions among stimuli (movie time-segments), and that BSC is consistently higher for Dartmouth, suggesting that those data have better signal-to-noise characteristics (see also Supplemental Figure 2A and Figure 4).

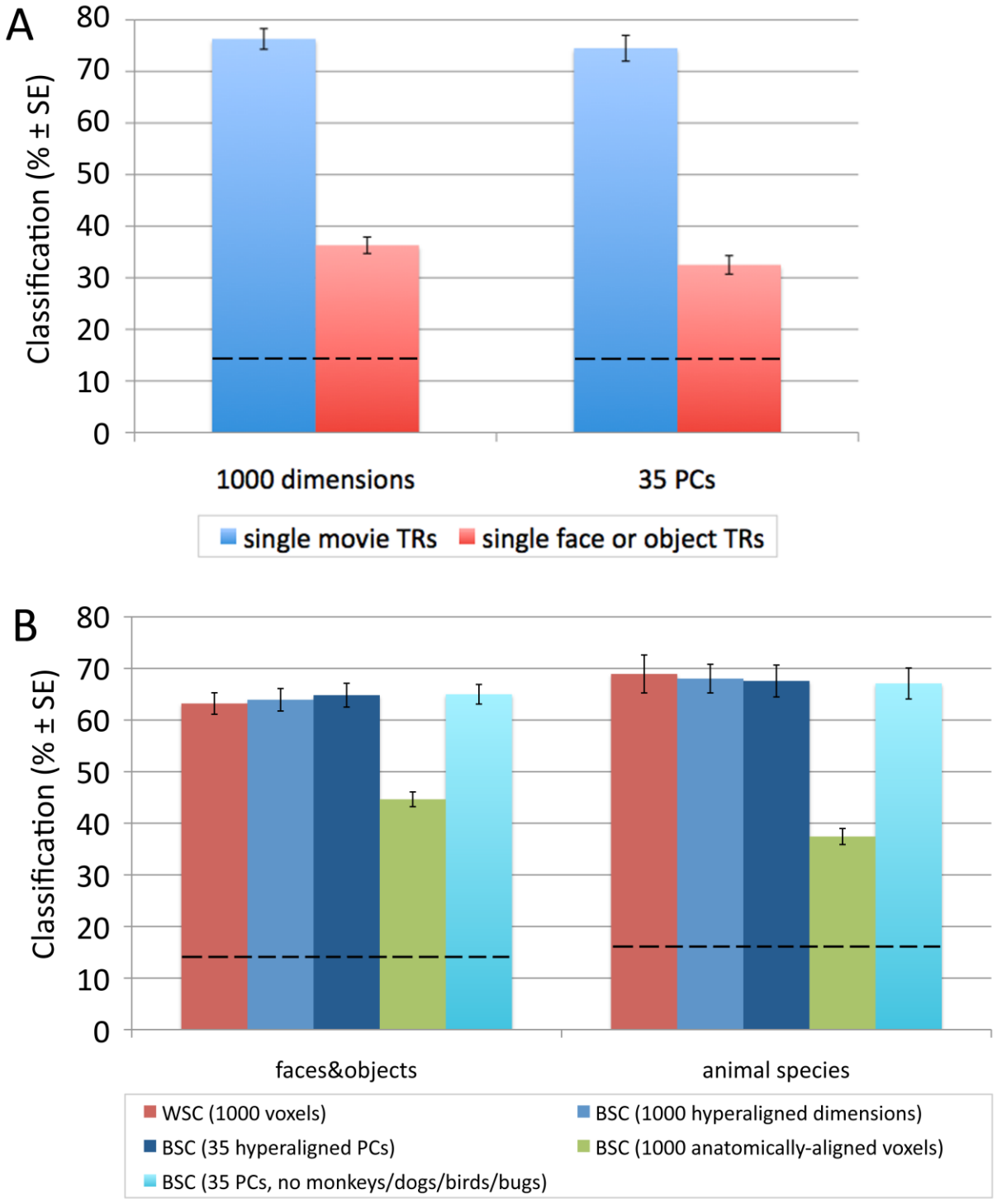

A
Classification (% ± SE)
80
70
60
50
40
30
20
10
0
1000 dimensions
35 PCs
single movie TRs
single face or object TRs
B
Classification (% ± SE)
80
70
60
50
40
30
20
10
0
faces&objects
animal species
WSC (1000 voxels)
BSC (1000 hyperaligned dimensions)
BSC (35 hyperaligned PCs)
BSC (1000 anatomically-aligned voxels)
BSC (35 PCs, no monkeys/dogs/birds/bugs)

Figure S4

Figure S4. Related to Figure 4.

(A) BSC of single time-points (TRs) from the movie and from the face & object perception study using 1000 dimensions and 35 PCs. We performed a seven-way BSC analysis of single time-points from the movie data and the face & object perception experiment data, in which we carefully matched the probability of correct classifications. Response patterns from one TR in each stimulus block of a run in face & object experiment were extracted from all subjects to perform a seven-way between subject classification using a correlation-based classifier (as described in the movie time segment classification section of Methods but with one TR instead of six). This was repeated for all 8 runs. Response patterns of TRs during the movie presentation at the same acquisition time as selected for the face & object were extracted from all subjects to perform a similar seven-way between subject classification using the same classifier. This was repeated 8 times to match with the face & object TR classification. Results showed that BSC accuracy for movie time-points was more than twice that for time-points in the face & object perception experiment indicating that the complex and dynamic visual stimuli in the movie evoke patterns of response in VT cortex that are much more distinctive than the responses evoked by still images of single faces or objects. Dashed lines indicate chance classification (one out of seven).

(B) BSC of category perception experiments in the common model space derived using only those parts of the movie that do not contain monkeys, dogs, birds, insects, or spiders. To establish the general validity of common model space derived from the movie data, we repeated the main analysis with the movie data after removing the time-points that contained monkeys, birds, or bugs (there were no dogs) and 30s following such episodes to account for any delayed hemodynamic responses. We removed 212 time points in total. This did not affect the between-subject classification accuracies suggesting that the hyperalignment is not specific to the stimuli present in the movie. We included the WSC accuracies using 1000 top-ranked voxels, BSC accuracies using 1000 hyperaligned dimensions, using 35 PC model derived from the full movie data, and using anatomical alignment for comparison. Dashed lines indicate chance classification levels.

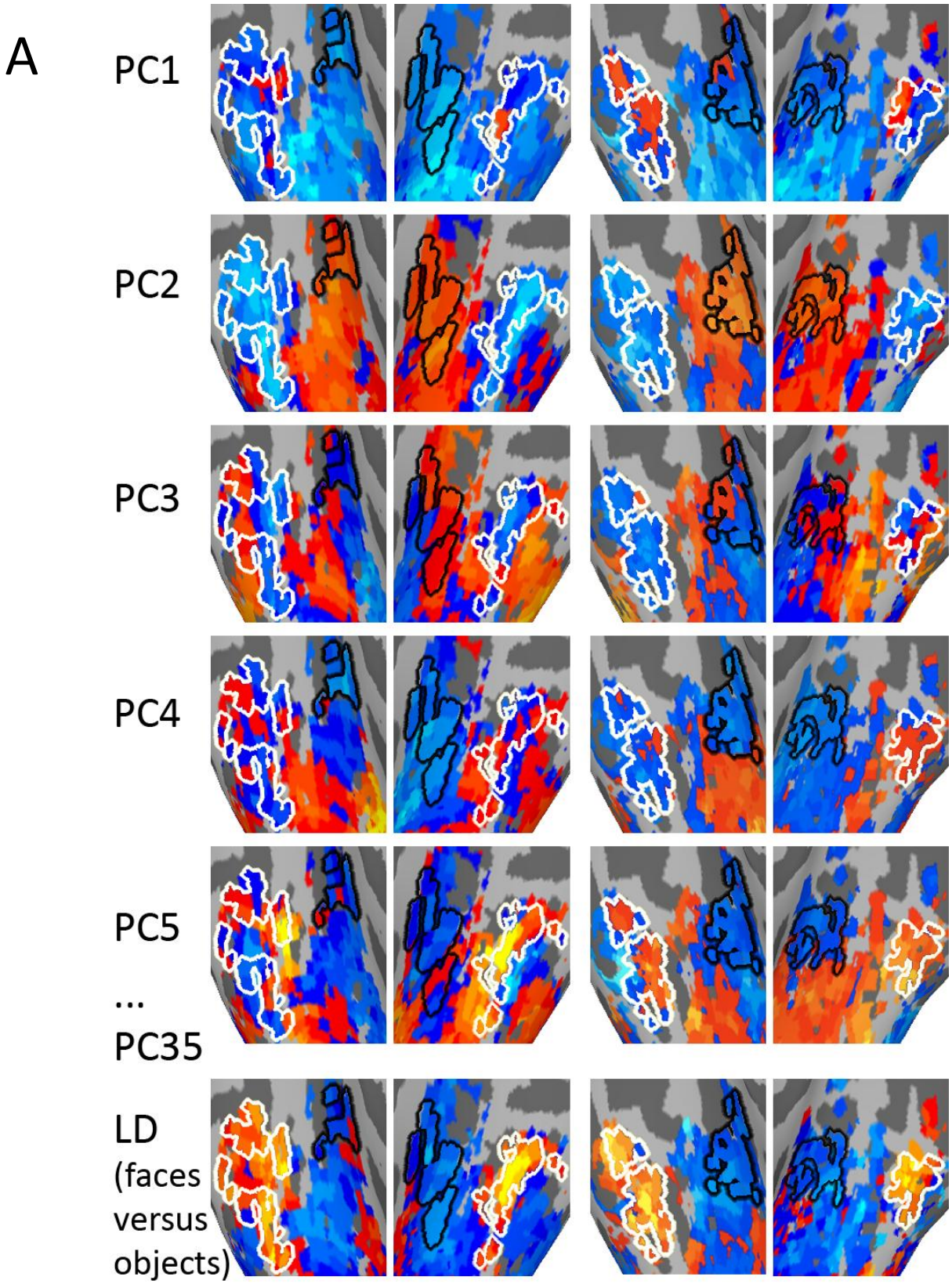


Figure S5, part A

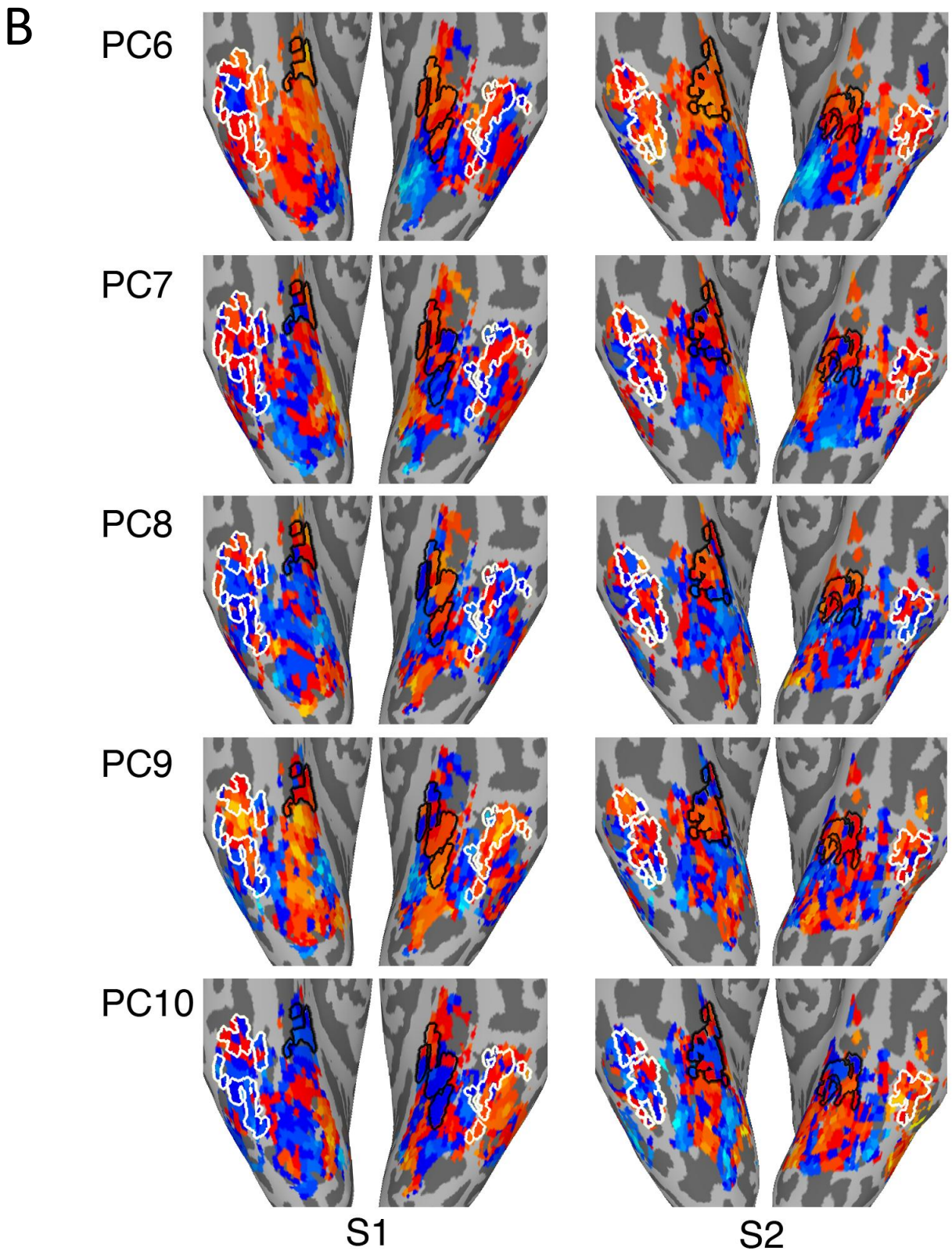
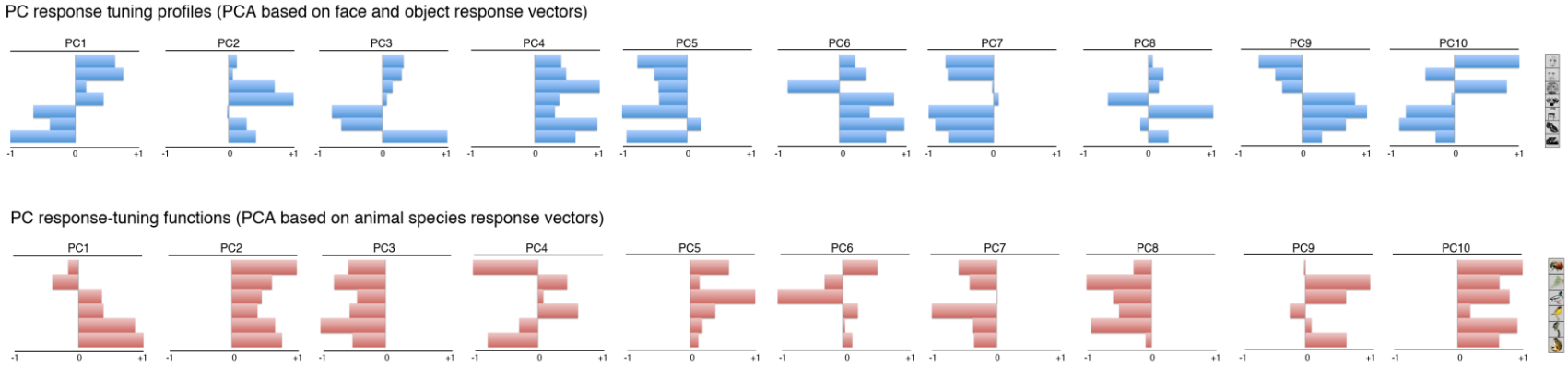


Figure S5, part B

C



Supplementary Figure 5, Part C

Figure S5. Related to Figure 5.

(A) Enlargement of the cortical topographies associated with the first 5 Principal Components and the Linear Discriminant contrasting responses to faces and objects in and around the individually-defined FFAs and PPAs (FFA: white outline; PPA: black outline) for two representative subjects. Cortical topography of the Linear Discriminant vector closely matches the outline of FFA in both the subjects.

(B) Cortical topographies for PCs 6 through 10. Note that the between-subject consistency of coarse topography that was evident in the higher order PCs (Figure 5 and Supplemental Figure 5A) is less apparent for lower order PCs.

(C) Response tuning functions for the top 10 PCs based on PCA of the face and object response vectors (blue bar graphs on top) and on PCA of animal species response vectors (red bar graphs below). The scale for response-tuning profiles is centered on zero, corresponding to the mean response to the movie, and scaled so that the maximum deviation from zero (positive or negative) is set to one. Note that the high order PCs reflect the most prominent conceptual distinctions (face and object PCA: PC1 - faces versus objects, PC2 – human versus nonhuman faces, PC3 – houses versus objects; animal species PCA: PC1 – primates versus birds versus insects), and the lower order PCs reflect finer distinctions (e.g. face and object PCA: PC4 and PC6 – monkey faces versus other faces, PC5 – shoes versus chairs; animal species PCA: PC4 – moths versus ladybugs, PC5 and P6 – mallards versus warblers, PC8 – monkey versus lemurs).

Supplemental Experimental Procedures

Subjects

Twenty one healthy young subjects (12 men, mean age = 23.9 yrs) with normal or corrected-to-normal vision participated in the main movie study. A subset of 10 subjects (5 men, mean age = 23 years), who were scanned at Princeton University, also participated in the face and object study. The remaining 11 subjects (7 men, mean age = 24.8 years) were scanned at Dartmouth College and participated in the animal species study. All subjects gave written, informed consent to participate in the study, and all experimental procedures were approved by the appropriate Institutional Review Boards.

MRI Acquisition

Princeton

Brain images were acquired using a 3T Siemens Allegra scanner with a standard 'bird-cage' head coil. For the movie study, whole brain volumes of 48 3 mm thick sagittal images (TR = 3 s, TE = 30 ms, Flip angle = 90°, 64 x 64 matrix, FOV = 192 mm x 192 mm) were obtained continuously through each half of the movie (1101 volumes for part 1, 1112 volumes for part 2). For the face and object study, we obtained images of brain volumes consisting of 32 3 mm thick axial images (TR = 2 s, TE = 30 ms, Flip angle = 90°, 64 x 64 matrix, FOV = 192 mm x 192 mm) that included all of the occipital and temporal lobes and all but the most dorsal parts of the frontal and parietal lobes. 192 volumes were obtained in each of eight runs. High resolution T1-weighted images of the entire brain were obtained in each imaging session (MPRAGE, TR = 2.5 s, TE = 4.3 ms, flip angle = 8°, 256 x 256 matrix, FOV = 256 mm x 256 mm, 172 1 mm thick sagittal images).

Dartmouth

Brain images were acquired using a 3T Philips Intera Achieva scanner with an eight-channel head coil. For the movie study, whole brain volumes of 41 3 mm thick sagittal images (TR = 2.5 s, TE = 35 ms, Flip

angle = 90°, 80 x 80 matrix, FOV = 240 mm x 240 mm) were obtained continuously through each part of the movie (entire movie is presented in 8 parts). For the animal species study, whole brain volumes consisting of 42 3 mm thick axial images (TR = 2 s, TE = 35 ms, Flip angle = 90°, 80 x 80 matrix, FOV = 240 mm x 240 mm) were acquired. 164 volumes were obtained in each of ten runs. High resolution T1-weighted images of the entire brain were obtained in each imaging session (MPRAGE, TR = 9.85 s, TE = 4.53 ms, flip angle = 8°, 256 x 256 matrix, FOV = 240 mm x 240 mm, 160 1 mm thick sagittal images). The voxel resolution was 0.938 mm x 0.938 mm x 1.0 mm. All functional images were acquired in an interleaved slice order.

Data Analysis

Data were preprocessed using AFNI (Cox, 1996)(http://afni.nimh.nih.gov). Surfaces were derived using FreeSurfer (Dale et al. 1999) and maps were generated using SUMA (Saad et al. 2004). All further analyses were performed using Matlab (version 7.8, The MathWorks Inc, Natick, MA) and PyMVPA (Hanke et al. 2009)(http://www.pymvpa.org).

Preprocessing

FMRI data for all studies were preprocessed in the same way unless otherwise specified. Preprocessing began with correcting slice time order and between scan head movements by spatially aligning all volumes to the volume closest to the anatomical scan. Data were then *despiked* to reduce extreme values. The movie data were then low- and high-pass filtered to remove temporal variation with frequencies higher than 0.1 Hz and lower than 0.00667 Hz (3dFourier in AFNI). High-pass filtering removes low temporal frequency changes with periods longer than 150 s. Low-pass filtering temporally smoothes fluctuations with frequencies higher than the hemodynamic response function. In both category perception experiments linear and quadratic trends were removed from each time-series

instead of Fourier filtering. Temporal components proportional to motion parameters obtained from the head movement correction step, and whole volume mean activation were regressed out (3dDetrend in AFNI). Additionally, movie data acquired at Dartmouth were resampled in time from 2.5 s TR to 3 s TR to match the temporal resolution of the movie data acquired at Princeton. Data acquired during the overlapping parts of the movie in the Dartmouth data were discarded. For each subject, data from all sessions (multiple movie sessions and category experiment sessions) were spatially aligned by applying the rotation parameters derived from aligning the mean motion-corrected EPI data of all sessions to the movie session. This spatial alignment established the voxel correspondence across sessions and studies for each subject. Data were then lightly spatially smoothed using a 4 mm full width at half-maximum Gaussian blur (see Supplemental Figure S2B). For analysis of between-subject classification based on anatomical alignment, the unsmoothed data from all experiments were resampled into Talairach atlas space (Talairach & Tournoux 1988) using AFNI using the same voxel grid as for the original data (3 mm x 3 mm x 3 mm voxels) and spatially smoothed using 4 mm full width at half-maximum Gaussian blur. Data from all three experiments were loaded into Matlab for further analyses and z-scored per session per subject separately (two movie sessions and one perception experiment).

Anatomical and Functional ROIs

A mask of ventral temporal cortex was hand-drawn on each subject's high-resolution anatomical scan as an anatomical region of interest using AFNI. The region extended from -70 to -30 on the y-axis in Talairach atlas coordinates (Talairach & Tournoux 1988). The region was drawn to include the inferior temporal, fusiform, and lingual/parahippocampal gyri. Within VT cortex, we defined functionally-defined regions of interest in the ten subjects scanned at Princeton that responded selectively to faces – the fusiform face area (FFA) – and to houses – the parahippocampal place area (PPA). The FFA was defined as all contiguous clusters of 10 or more voxels that each responded more to faces than to objects at $p < 0.001$. Voxels in the inferior temporal and occipital gyri were excluded. The PPA was defined as all

contiguous clusters of 10 or more voxels that each responded more to houses than to faces at $p < 0.001$ and more to houses than to small objects at $p < 0.05$.

Surface generation and mapping

Cortical surfaces for all subjects were generated using FreeSurfer (recon-all, Dale et. al 1999) from high-resolution anatomical scans acquired during multiple study sessions. Surfaces were then converted into SUMA format (Saad et. al 2004). Surface mapping was performed using default interpolation of 3D data from AFNI to 2D cortical surface in SUMA. Cortical surfaces were used for illustration only.

Voxel selection

Voxels from the VT mask of each subject were partitioned into left and right hemisphere voxels. For each voxel in a subject, the correlation of its time-series with the time-series of each voxel in the same hemisphere of each other subject was calculated. The highest among these correlations was considered as that voxel's correlation-score. The sum of the correlation-scores for a voxel over all twenty subjects was considered its total-correlation-score. For each subject, we then ranked voxels in each hemisphere based on their total-correlation-scores (voxels with highest scores ranked the best). These ranks formed the basis for selecting a certain number of voxels from each subject's left and right hemispheres. For BSC of movie time-segments, voxel selection and derivation of the common model space was based only on data from the half of the movie that was not used for BSC. For BSC of the category perception experiments, voxel selection and derivation of the common model space was based on the full movie.

Supplemental References

Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. NeuroImage *9*, 179-194.

Saad, Z. S., Reynolds, R. C., Argall, B., Japee, S., and Cox, R. W. (2004). SUMA: an interface for surface-based intra- and inter-subject analysis with AFNI. In Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on, pp. 1510-1513 Vol. 2.