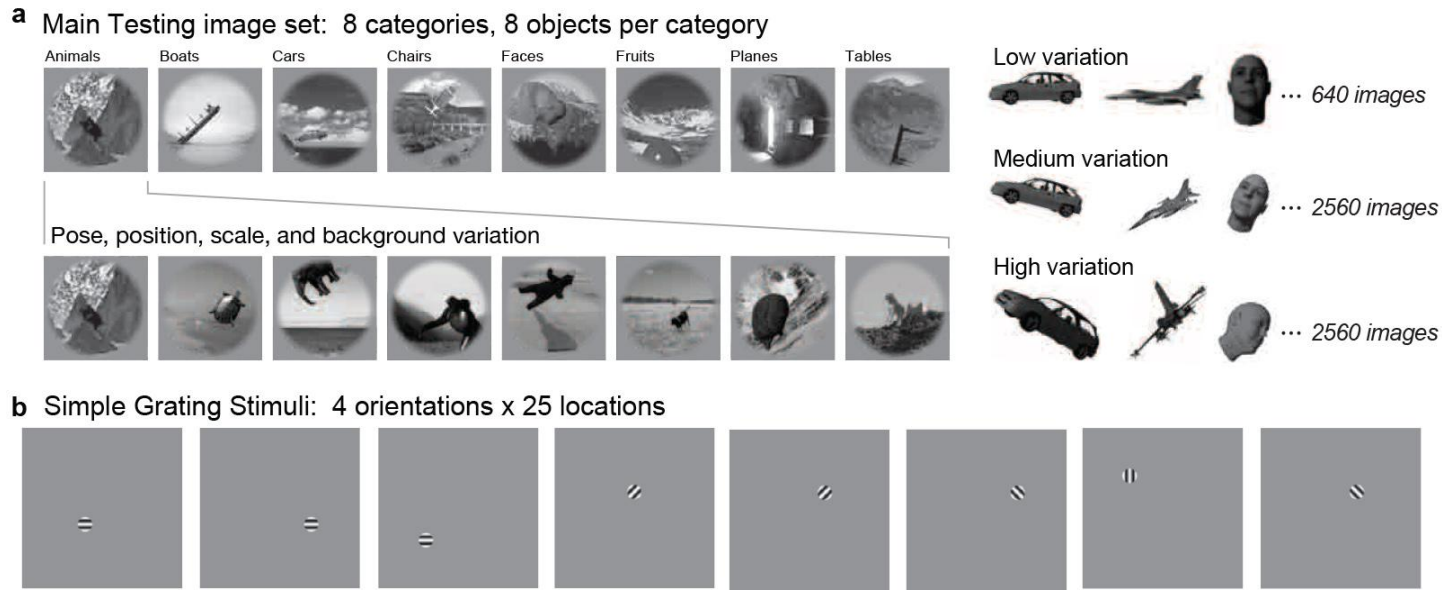
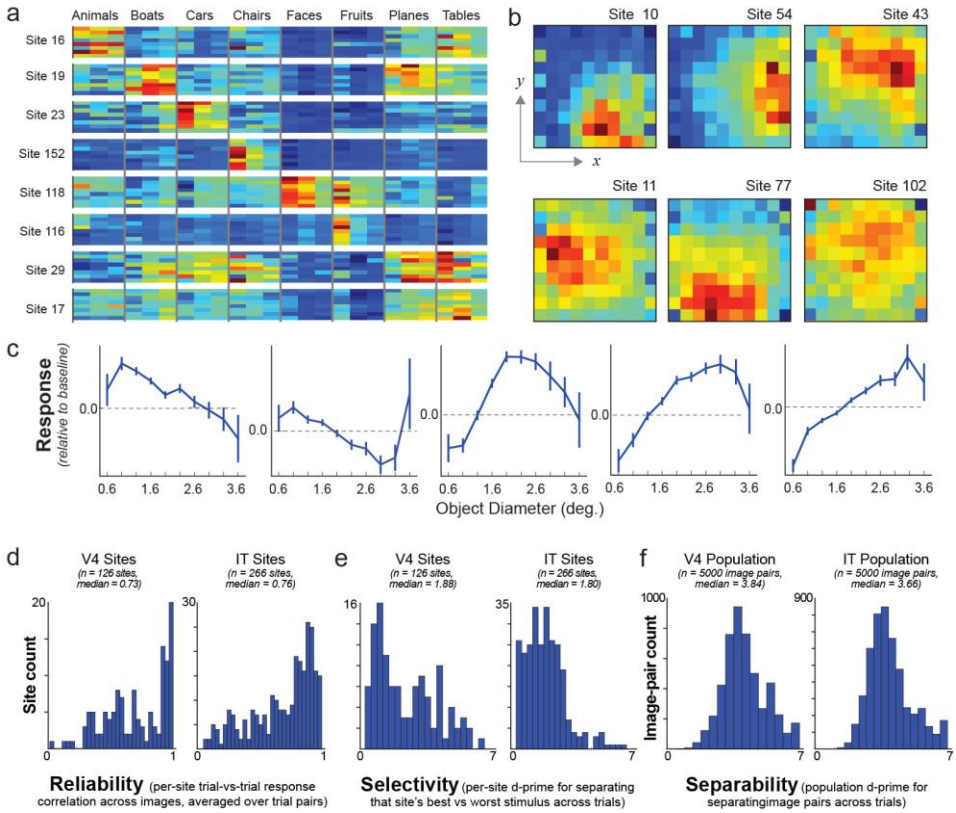


Supplementary Figure 1

Image sets.

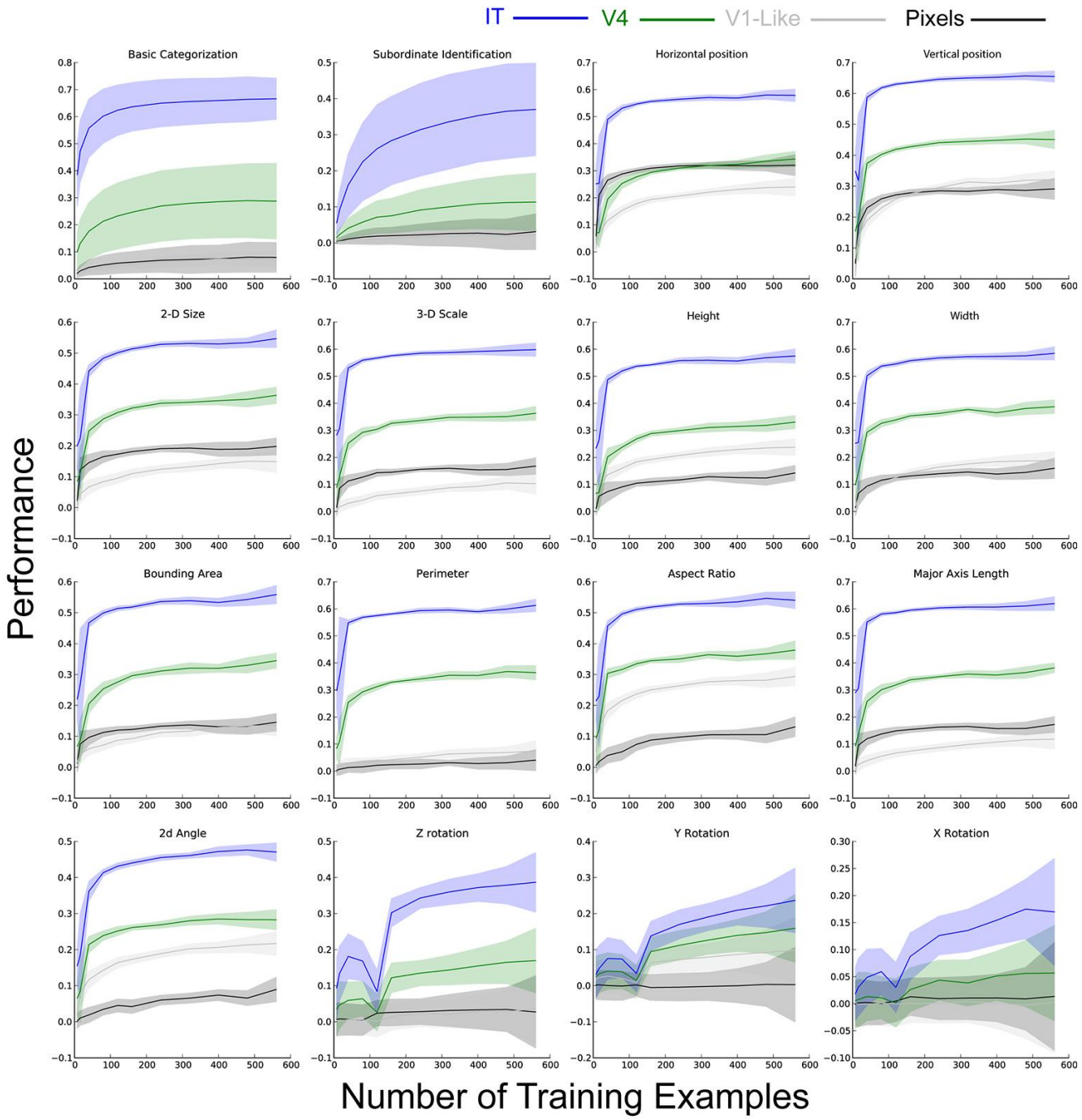
(**a**) Typical examples drawn from the complex high-variability testing image set on which we collected neural data and evaluated models. The imageset contained 5760 images of 64 three-dimensional objects in 8 common categories, including animals, boats, cars, chairs, faces, fruits, planes and tables. The overall imageset was comprised of three subsets, with object view parameters (position, size and pose) chosen randomly from within low-, medium- and high-variation ranges, respectively. Object images were then superimposed on complex real-world background photographs, which were chosen randomly to ensure there were no correlations between background content and object category identity. This dataset supported a wide range of categorical and non-categorical tasks (Fig. 2), on which we evaluated population performance of V4 and IT neural populations (Figs. 3 – 5 and 7c) as well as computational models (Figs. 6, 7c, and Supplementary Figs. 7d–10). (**b**) Low-variation simpler examples drawn from the set of simple grating stimuli. This stimulus set supported three simple tasks, including horizontal and vertical position estimation of the center of grating object, as well as grating orientation (see Fig. 7a,b).



Supplementary Figure 2

Characterization of IT and V4 responses.

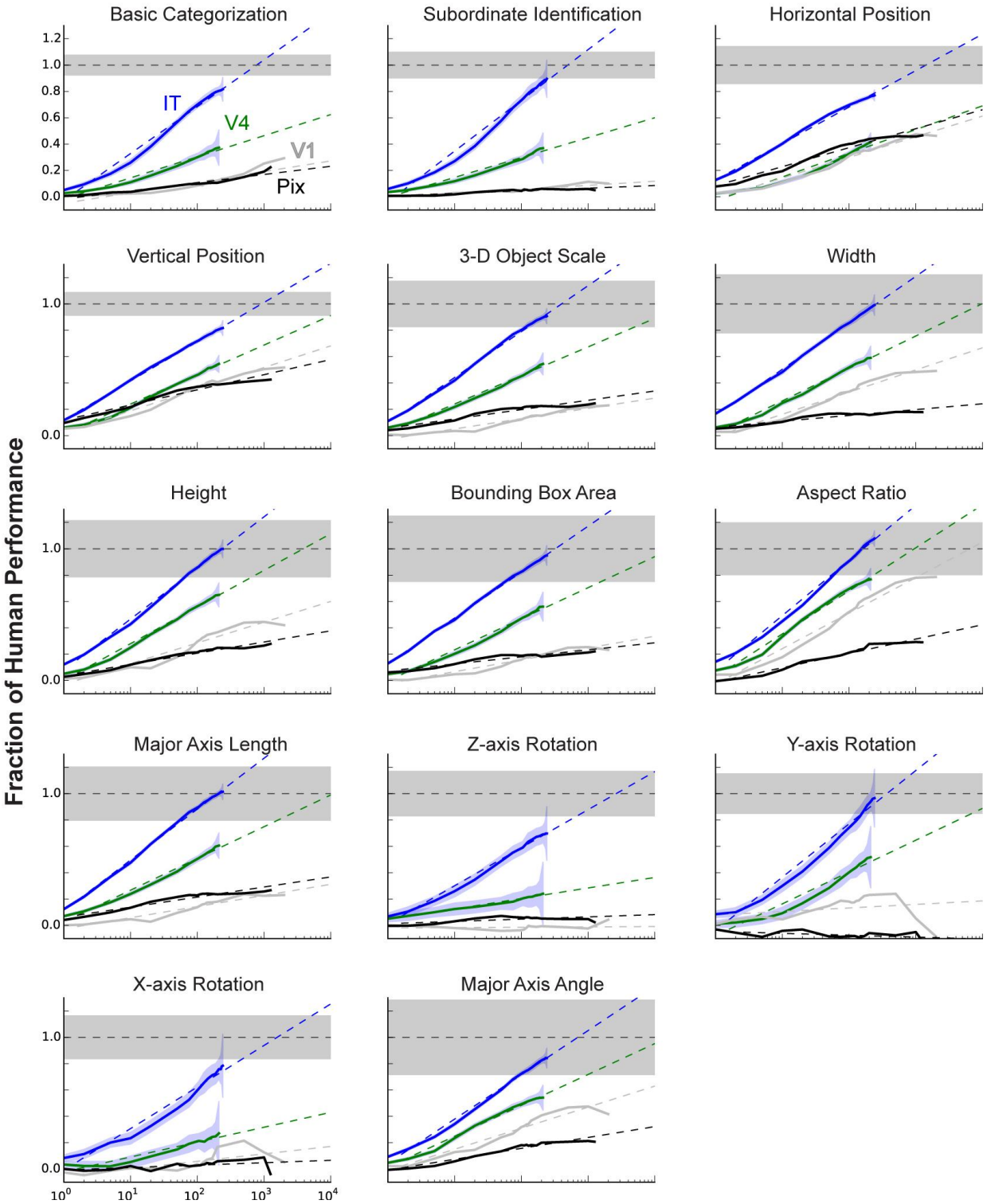
Single sites with high selectivity (panels a–c). (**a**) Category selectivity heatmaps of the single sites in our IT sample that are best at decoding each of the eight categories present in our stimulus set. Each colored bar represents the response of the indicated site relative to that site's baseline (blue=low, red=high). The colored bars represent responses averaged over images of each of the eight object exemplars in the indicated category (vertical axis), further broken down into three increasing levels of parameter variation level (horizontal axis, see Methods). (**b**) Position selectivity response heatmaps for best single sites for object position estimation. Each colored squared position in each heat map represents the average of the indicated site's activity over all images where the objects center is located in that square's position. (**c**) Size selectivity response profiles for the best single sites for object size estimation. The *x*-axis represents the object diameter in degrees as seen by the animal. The *y*-axis represents response relative to baseline of the indicated site, averaged over all images whose size falls in the indicated diameter bin. Error bars are standard errors of the mean over images within each bin. Simple image-level metrics (panels d–f). (**d**) Histogram of per-site reliabilities for V4 sites (left panel) and IT sites (right panel), as measured by trial-vs-trial correlation of each site's responses across images in high-variation image set (Supplementary Fig. 1a). (**e**) Histogram of per-site selectivity for V4 sites (left panel) and IT sites (right panel), defined for each site as d-prime of separating that site's responses for its best (most highly response-driving) stimulus versus its worst (least responsive-driving) stimulus. (**f**) Histogram of population separability for V4 population (left panel) and IT population (right panel) for 5000 image pairs, measured as d-prime of population readout for separating the two images in the pair. In both the case of V4 and IT, 126 units were used to build population decodes: in the case of V4, all 126 measured units were included; while in the case of IT, for each image pair we randomly sampled a different subset of 126 units on which to build the decoder, to equalize with the V4 population.



Supplementary Figure 3

Performance training curves.

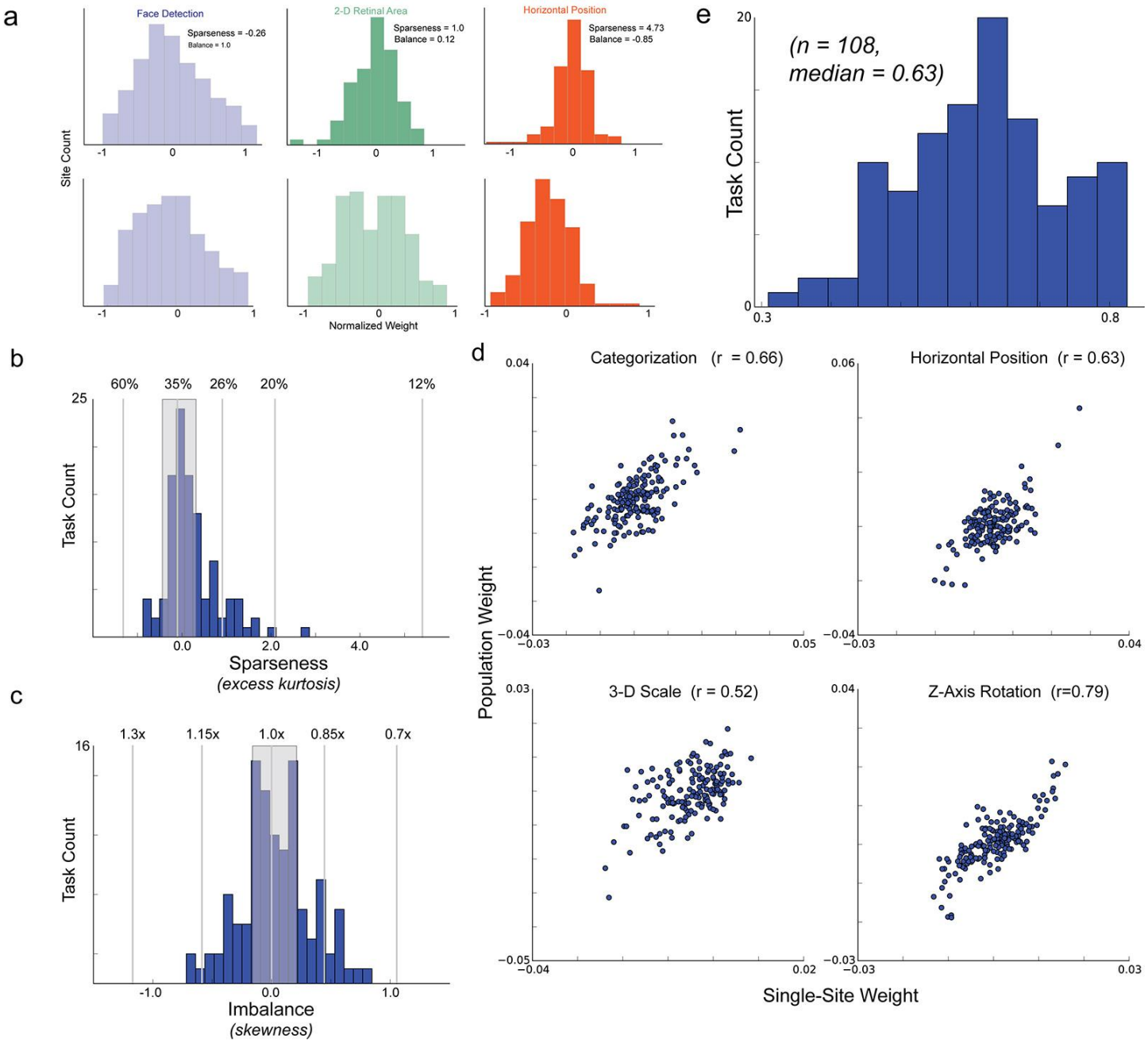
For each task, performance of population decodes as a function of number of training examples used to train linear classifiers and regressors. Error bars are over samples of units and image splits. Blue lines are for IT neural population, green lines are V4 neural population, gray is V1-like model population, and black is pixel control.



Supplementary Figure 4

Performance extrapolation.

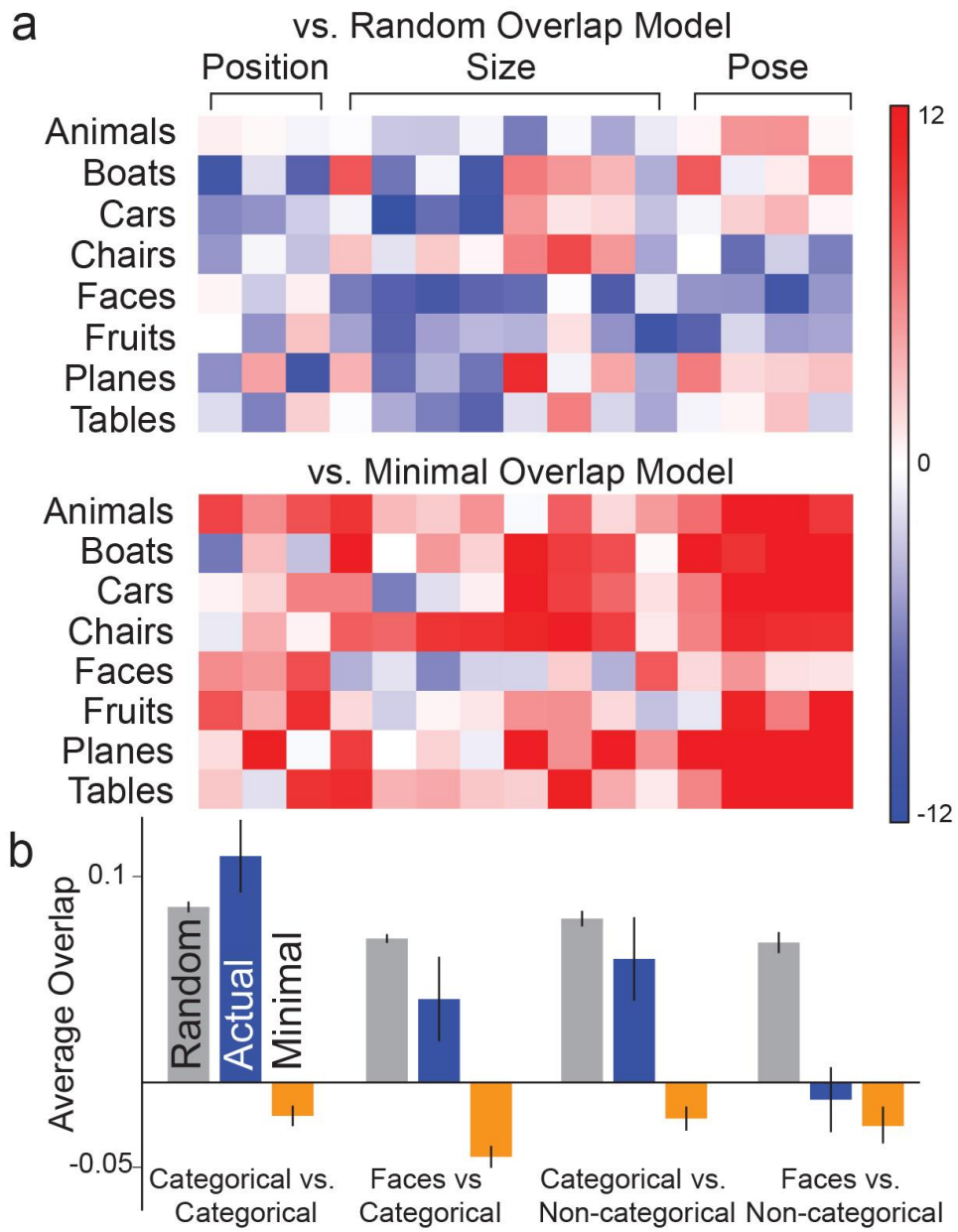
Extrapolations of neural decoding performance as a fraction of human performance, for all tasks on which we measured human performance, including those shown in 4a. All figure definitions, including the *x* and *y*-axes, are as in Figure 4a.



Supplementary Figure 5

Task performance distributions for single IT sites.

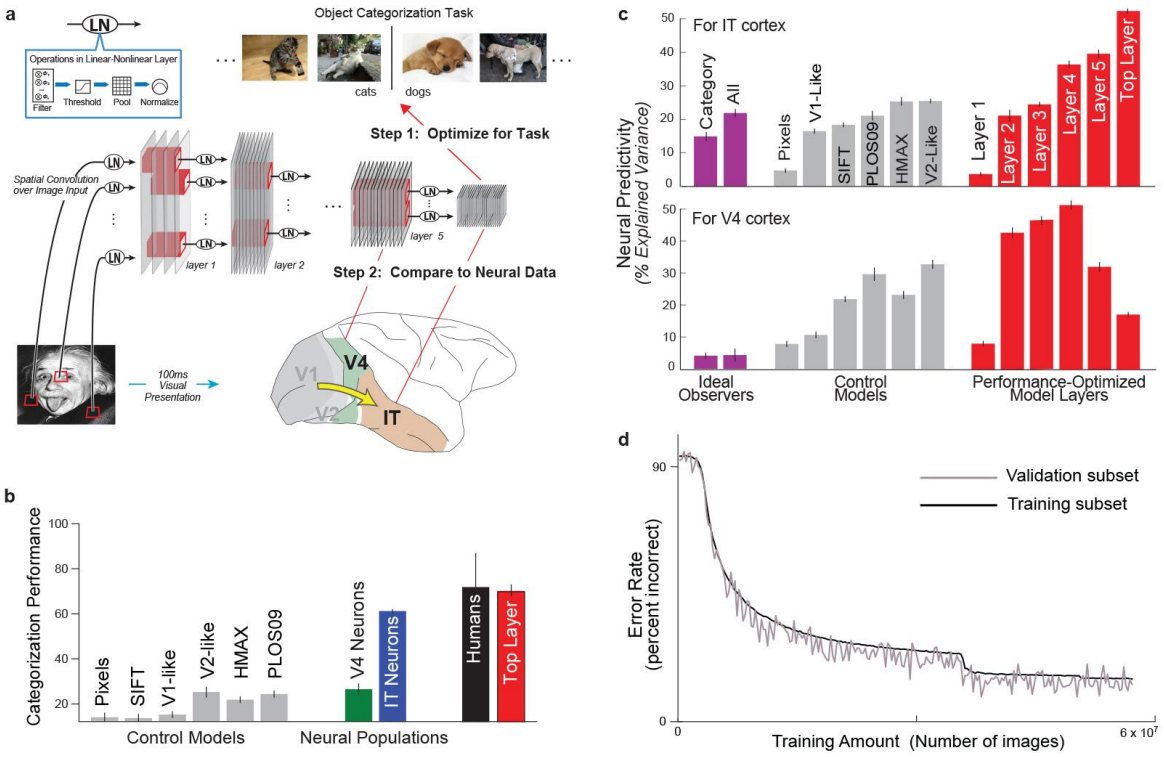
(**a**) Representative examples of distributions of classifier weights across neural sites for several tasks. *x*-axis represents normalized decoder weight, and *y*-axis is site bincount. Top row shows distributions for weights determined via linear SVM (and as used in Fig. 5); bottom row shows distributions for weights determined on a single-site basis. (**b**) and (**c**) Sparsity and imbalance measures for single-site weights (compare Fig 5a-c). (**d**) Scatter plots of single-site weights versus weights on the same units from SVM classifiers/regressors, for selected tasks. Each point is a distinct IT site. (**e**) Quantification of correlations of plots in panel d, for all tasks. *x*-axis measures Pearson's R correlation between single-site and population decoding weights for each individual task.



Supplementary Figure 6

Comparison of task overlap to random and minimal control models.

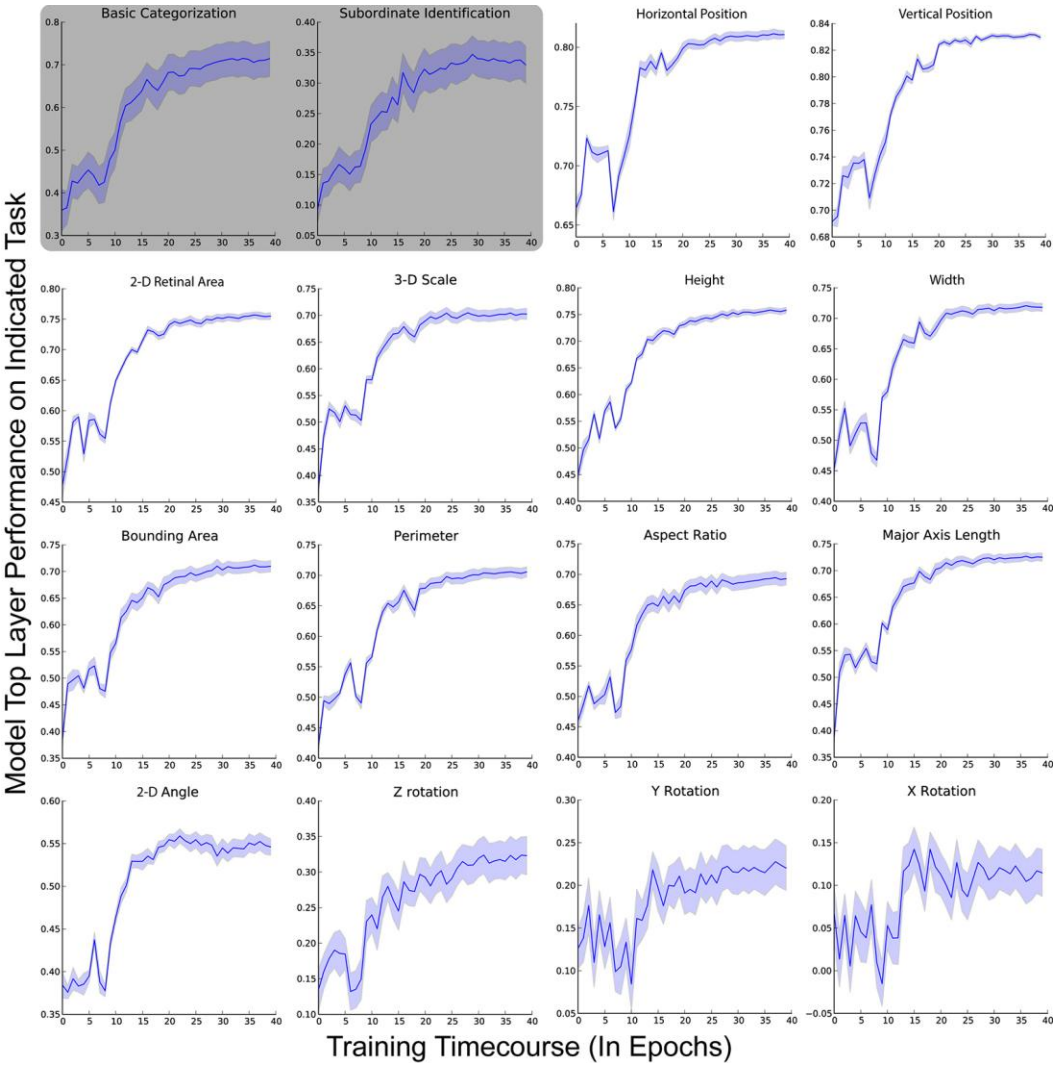
(**a**) Comparison of weight overlap for categorical vs non-categorical tasks, relative to random overlap (top) and minimal overlap (bottom) models. Color represents *t*-statistic of separation of each pairwise overlap from the indicated control model (random in the top panel, minimal overlap in lower panel), with red color representing more overlap than random and blue representing less. (**b**) Average overlap for (i) (non-face) categorical tasks, (ii) faces-vs-non-face categorical tasks, (iii) non-face categorical tasks vs non-categorical tasks and (iv) faces vs non-categorical tasks. Shown are actual neural overlap (blue bars) in comparison to random overlap (gray) and minimal overlap (orange) models. Error bars for neural data overlap are due to variation in unit sampling and classifier training split. Error bars in model overlaps are due to variation of model input data (per-task and per-unit weight constrains) due to unit sampling and classifier training split, as well as random initial conditions of model weights.



Supplementary Figure 7

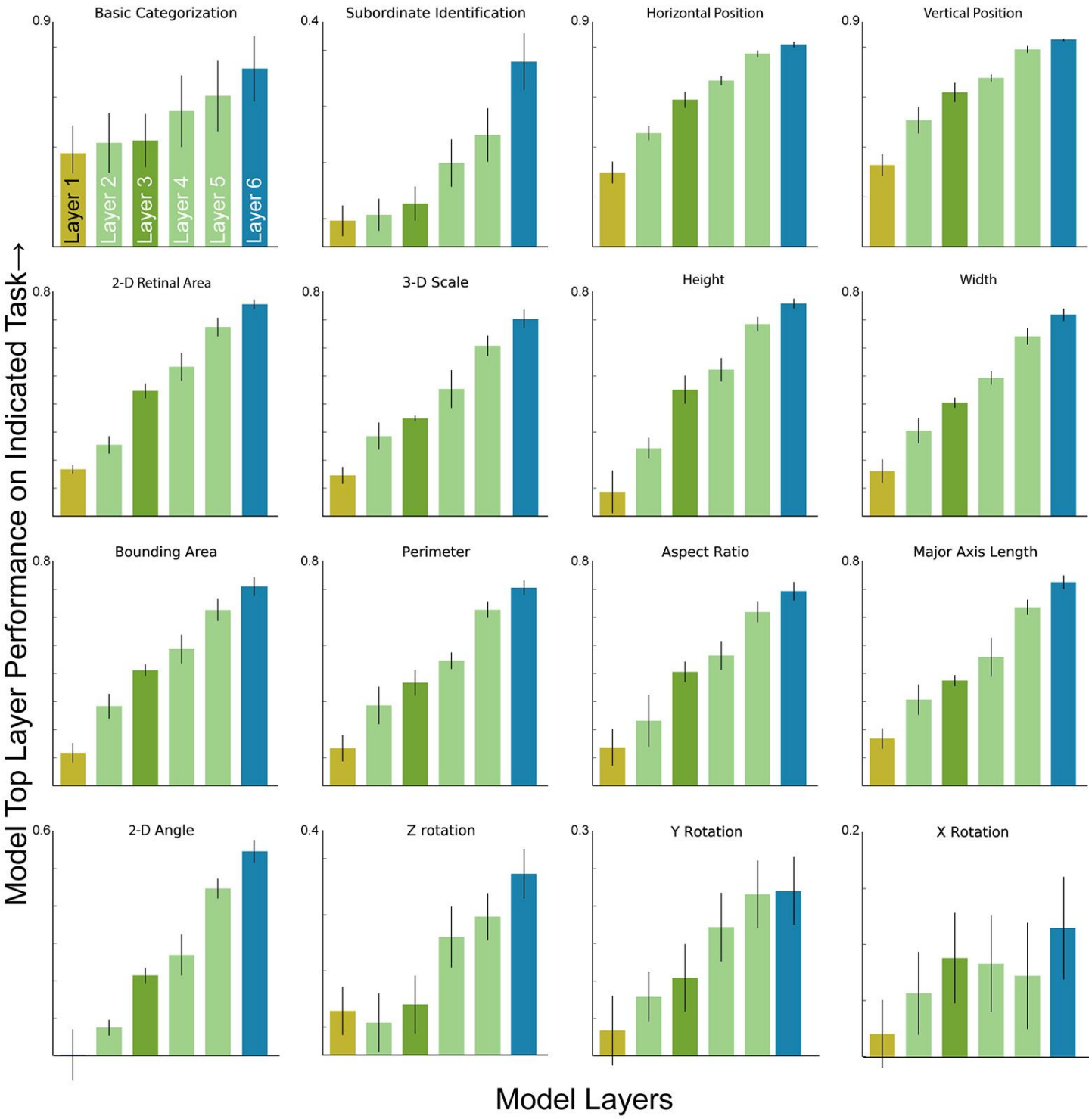
Computational Modeling approach.

(**a**) In previous work (Yamins, D., and Hong, H. et al. (2014) *Proc Natl Acad Sci USA* **111**, 8619–24), we developed a model of areas in higher ventral cortex. Our approach was based on optimizing model parameters to maximize performance of the model's output layer on a challenging object recognition task involving real-world images in a large number of semantically distinct categories. (Note: The cat-vs-dog task shown in the figure is just for illustration purposes; the actual task on which the model was optimized contained no animal images.) This model was then used to predict neural responses in macaque V4 and IT cortex, recorded on the images shown in Supplementary Fig. 1. (**b**) The top layer of the performance-optimized model generalizes from the photographic training set (red bar), and significantly outperforms control models (gray bars) and the V4 neural population on the 8-way object categorization task (Animals vs Boats vs Cars vs Chairs vs Faces vs Fruits vs Planes vs Tables) in the images shown in Supplementary Fig. 1a. Model performance is comparable to IT neural population (blue bar) and human performance measured via psychophysical experiments (black bar). (**c**) The performance-optimized model is then used predict neural response in IT cortex (top panel) and V4 cortex (bottom bars). Ability to predict IT neural patterns is better with each subsequent model layer, peaking at the top layer (red bars), whereas ability to predict V4 neurons peaks in the middle layers. For both V4 and IT, the performance-optimized model's most predictive layer is significantly better than other control models, including ideal observers that perform perfectly on categorization tasks (purple bars) as well as control models that are also in the general class of neural networks (gray bars). All panels in this figure are related to previously published results (Yamins, D., and Hong, H., et al. (2014) *Proc Natl Acad Sci USA* **111**, 8619–24). (**d**) Computational model training set performance timecourses. *x* axis represents amount of training data during model training, for the computational model used in Figs. 6 and Supplementary Figs. 8–10. *y* axis represents error rate, measured in percent incorrect. Gray line is performance on a held-out subset of the images in the overall training image set; black line is performance on actual subset of images on which the model was trained. (See Methods section for details of training.)



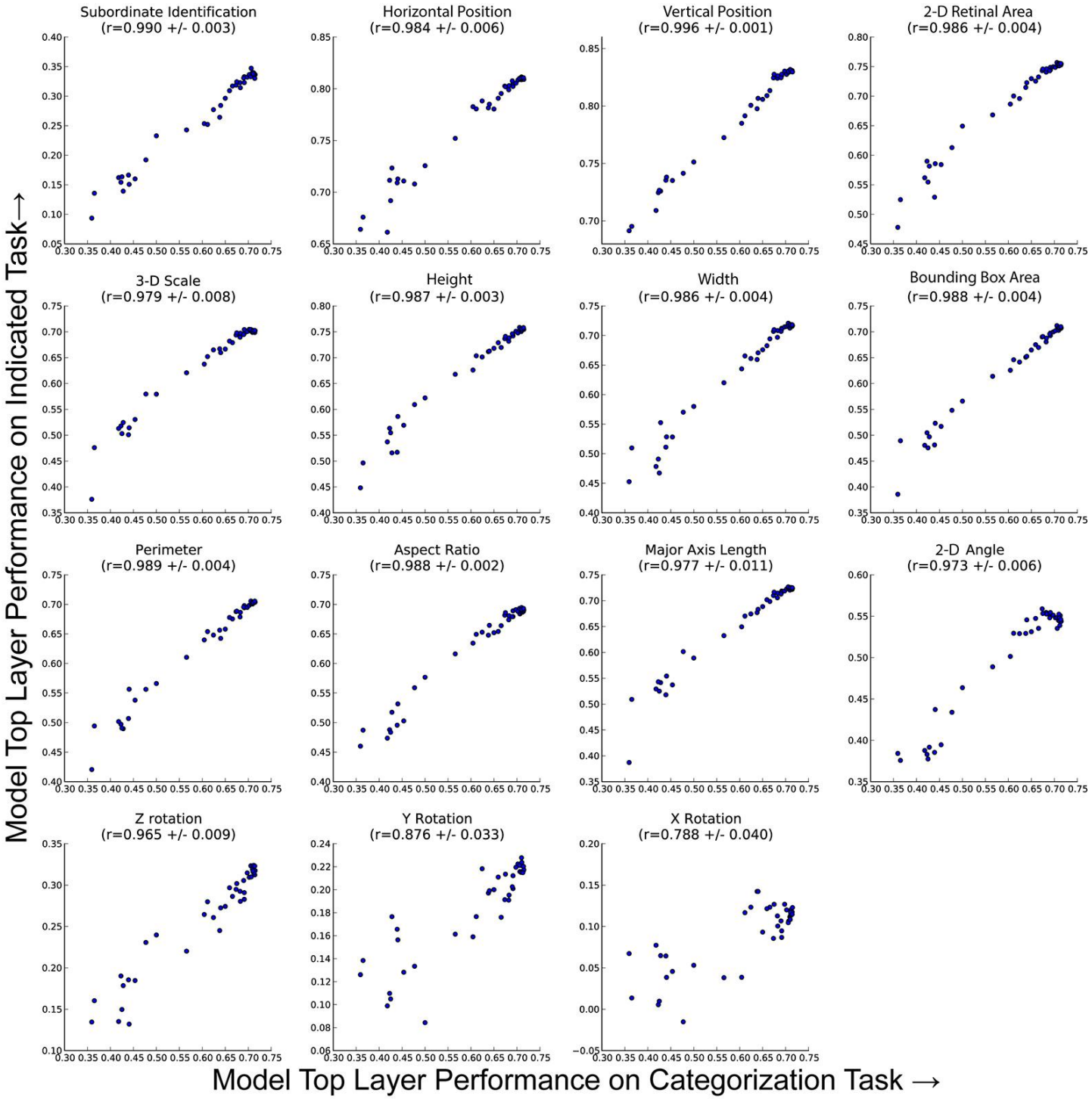
Supplementary Figure 8

Performance time-courses for computational model on test set tasks.
y-axis represents performance of linear regressors and classifiers trained on the top hidden layer of the computational model, on each task defined on the testing set (see Supplementary Fig. 1a). *x*-axis represents timepoints taken during training for categorization on the ImageNet dataset (as described in Methods). Performance was estimated by building top-level regularized classifiers and regressors (as described in the methods text) separately at each time step. Note that the *x*-axis is the same for all panels, representing the same training trajectory; the various *y*-axis panels are all based on the single feature set produced by the categorization training. The first two panels, with gray backgrounds, indicate categorical tasks (8-way basic categorization and subordinate category identifications); the remaining white-background panels indicate non-categorical tasks.



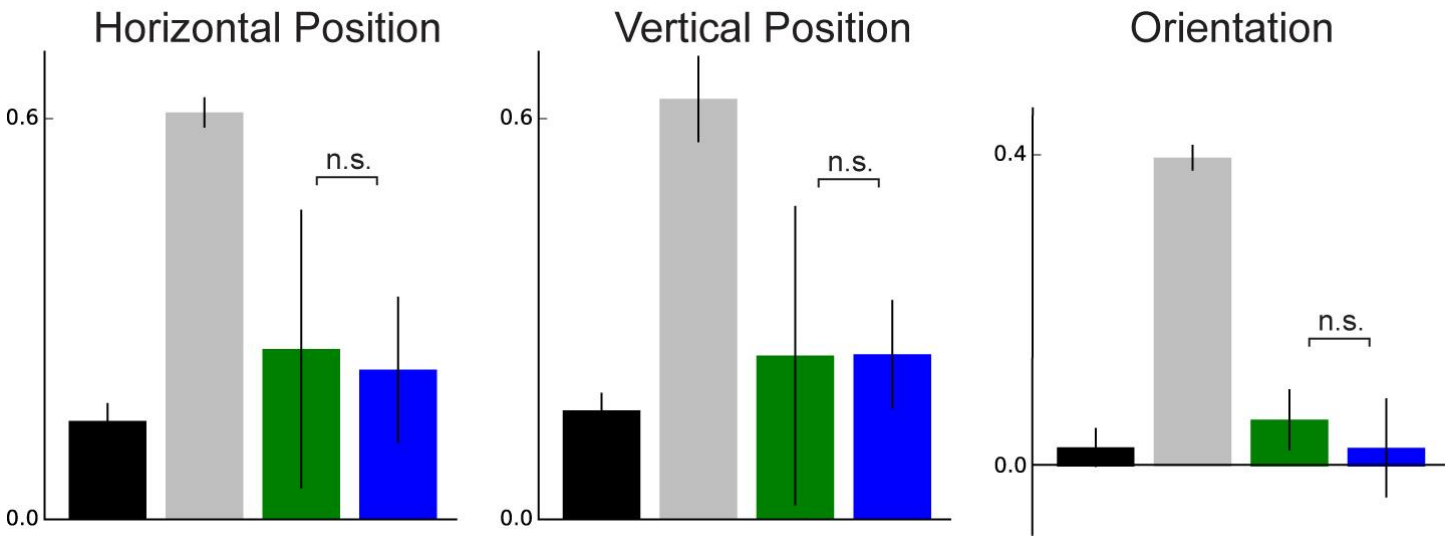
Supplementary Figure 9

Computational model performance as a function of layer.
This figure is analogous to main text Fig. 6a, but shows results for all tasks measured in the high-variation stimulus set (as in the neural data shown in Fig. 3b).



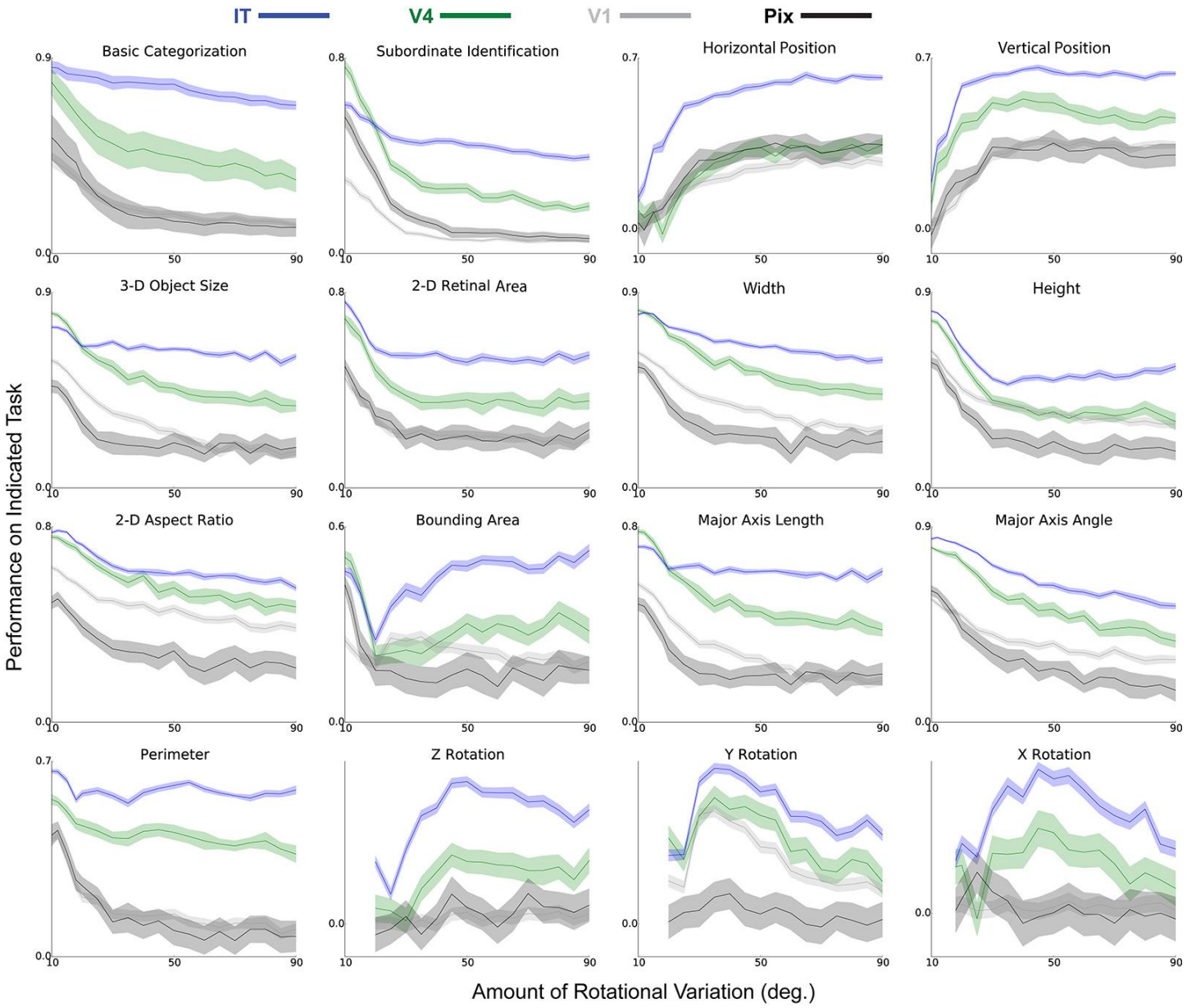
Supplementary Figure 10

Computational model categorization performance Vs non-categorical task performance.
y-axes in each panel are same as in Supplementary Fig. 8. *x*-axis is performance on test set categorization task (e.g., the *y*-axis of the upper-left-most panel in Supplementary Fig. 8). Each dot represents a distinct timepoint as shown on the *x* axis in Supplementary Fig. 8. Performance is shown for the top hidden layer of the model, which in every task achieved the highest performance level.



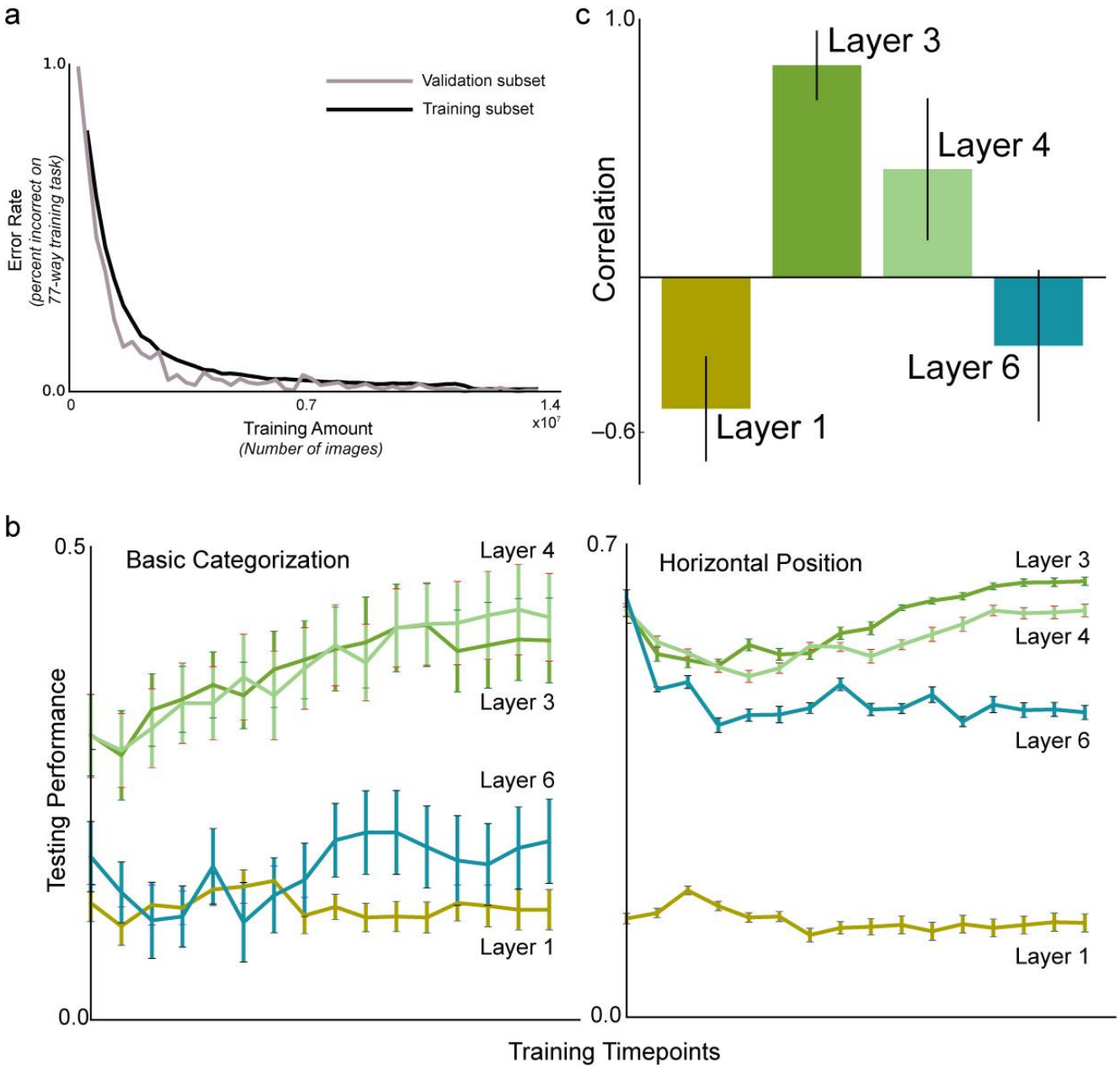
Supplementary Figure 11

Performance of populations on simple grating stimuli, using RBF classifier.
As in Fig. 7a, this figure shows population decoding results for position and orientation tasks defined on a simpler stimulus set consisting of grating patches placed on gray backgrounds. However, these results were evaluated using a non-linear Radial Basis Function (RBF) classifier with Gaussian kernels, in place of the simpler linear classifier as in Fig. 7a. As discussed in the text, the classifier parameters, including regularization constant C and kernel size σ were determined using standard cross-validation procedures. y-axis, bar colors, and error bars are as in Fig. 7a.



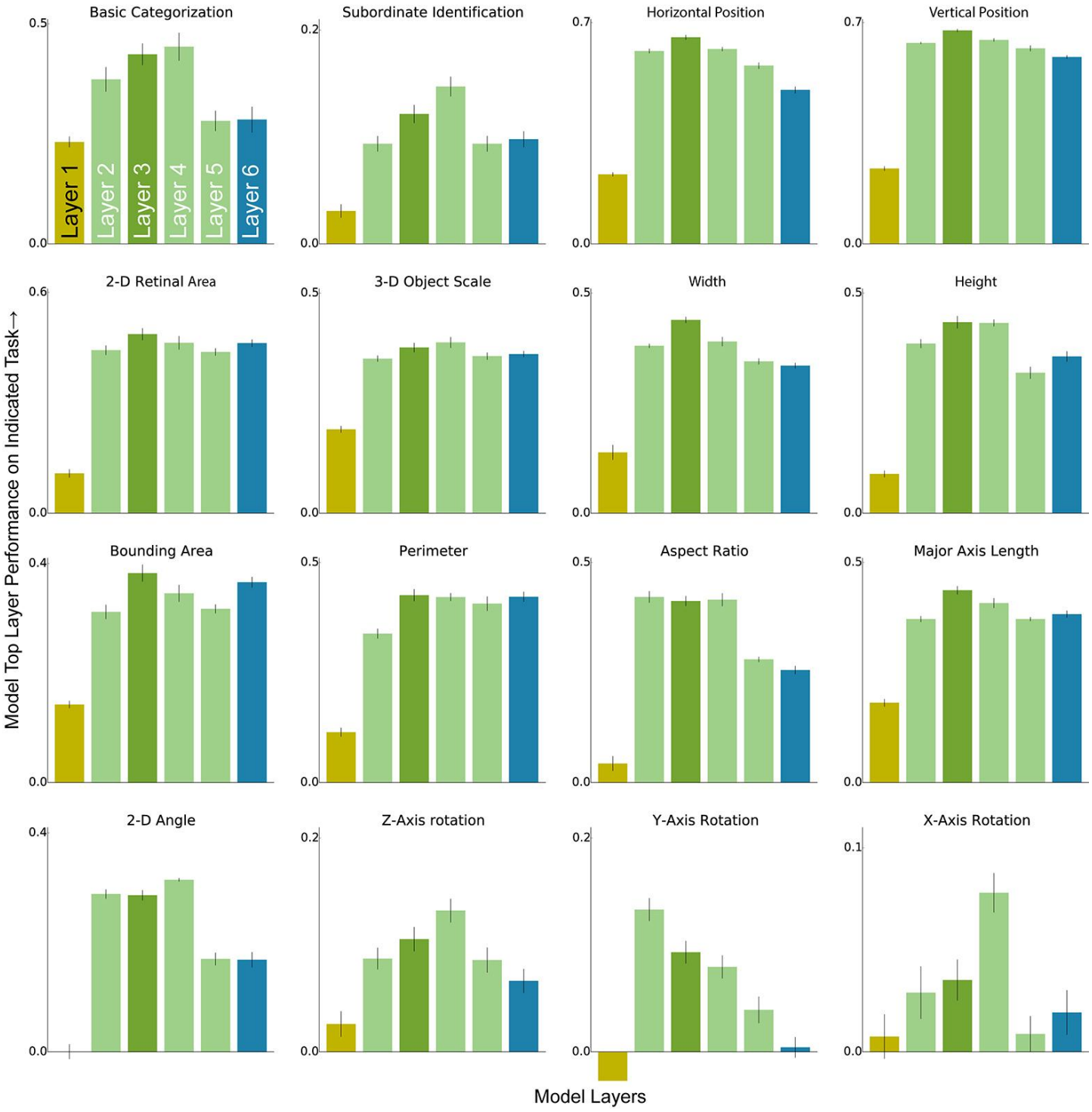
Supplementary Figure 12

Performance of neural populations as a function of increasing rotational variation.
This figure shows the same metrics as in main text Fig. 7c, for all tasks in the high-variation dataset task battery.



Supplementary Figure 13

Alternative computational model trained with lower-variation training data.
(**a**) Training performance curve for an alternative computational model, trained using a dataset containing a large number of categories but less overall background and object pose variation that the original model shown in Supplementary Fig. 7 (see Online Methods for more information). Axes and labels are as in Supplementary Fig. 7d. (**b**) Performance of alternative model on the high-variation testing set on which neural data was collected, for categorization task (left panel) and horizontal position estimation task (right panel). Performance for layers 1, 3, 4, 6 are shown. Axes are as in Supplementary Fig. 8. (**c**) Correlation between performance on testing-set categorization and horizontal position estimation tasks, for each of four model layers shown in panel b. The *y* axis and error bars are as in Fig. 6e.



Supplementary Figure 14

Alternative computational model performance by layer and task.
For the alternative computational model trained using less object pose variation, this figure shows performance on the testing set (as in Supplementary Fig. 13) as a function of model layer and task. Axes, error bars, and metrics are as in Supplementary Fig. 9.

Figure Panel	Test	Hypothesis	Details	P-value	Test statistic (t-test unless otherwise noted)
3a	Bootstrap	That values are different form 0, for both IT and V4 populations, performances of most informative site for each of several tasks, as indicated in the figure labeling.		< 10e-4	N/A
3a	Paired t-test	That the value of the difference between most informative IT and V4 sites is greater than 0. Pairs were constructed using sets of images used to determine most informative sites	Y-Axis position	0.012	2.63
3a	Paired t-test	That the value of the difference between most informative IT and V4 sites is greater than 0. Pairs were constructed using sets of images used to determine most informative sites	Y-Axis Size	0.032	2.215
3a	Paired t-test	That the value of the difference between most informative IT and V4 sites is greater than 0. Pairs were constructed using sets of images used to determine most informative sites	2-D Retinal Area	0.07	2.069
3a	Paired t-test	That the value of the difference between most informative IT and V4 sites is greater than 0. Pairs were constructed using sets of images used to determine most informative sites	Major Axis Angle	0.045	2.069
3a	Paired t-test	That the value of the difference between most informative IT and V4 sites is greater than 0. Pairs were constructed using sets of images used to determine most informative sites	X-axis Rotation	0.366	< 2
3a	Paired t-test	That the value of the difference between most informative IT and V4 sites is greater than 0. Pairs were constructed using sets of images used to determine most informative sites	Y-Axis rotation	0.135	< 2
3a	Paired t-test	That the value of the difference between most informative IT and V4 sites is greater than 0. Pairs were constructed using sets of images used to determine most informative sites	Z-Axis Rotation	0.241	< 2
3a	Paired t-test	That the value of the difference between most informative IT and V4 sites is greater than 0. Pairs were constructed using sets of images used to determine most informative sites	All other properties tested.	< 10e-4	> 3

3b	Bootstrap	That values are different from chance, for population decoding performances of V4 and IT populations.		< 10e-4	N/A
3b	Paired t-test	That the value of the difference in performance between population pairs is greater than 0. Pairs were constructed from sets of images used to build the linear estimators.	All IT - V4 comparisons	< 10e-4	> 3
3b	Paired t-test	That the value of the difference in performance between population pairs is greater than 0. Pairs were constructed from sets of images used to build the linear estimators.	V4 - V1 X-Axis position	0.0625	< 2
3b	Paired t-test	That the value of the difference in performance between population pairs is greater than 0. Pairs were constructed from sets of images used to build the linear estimators.	V4 - V1 Aspect Ratio	0.0817	< 2
3b	Paired t-test	That the value of the difference in performance between population pairs is greater than 0. Pairs were constructed from sets of images used to build the linear estimators.	V4 - V1 Major Axis Angle	0.0437	2.382
3b	Paired t-test	That the value of the difference in performance between population pairs is greater than 0. Pairs were constructed from sets of images used to build the linear estimators.	V4 - V1 Y-Axis Size	0.0243	2.783
3b	Paired t-test	That the value of the difference in performance between population pairs is greater than 0. Pairs were constructed from sets of images used to build the linear estimators.	V4 - V1 Y-Axis Rotation	0.0415	2.173
3b	Paired t-test	That the value of the difference in performance between population pairs is greater than 0. Pairs were constructed from sets of images used to build the linear estimators.	V4 - V1 X-Axis Rotation	0.396	< 2
3b	Paired t-test	That the value of the difference in performance between population pairs is greater than 0. Pairs were constructed from sets of images used to build the linear estimators.	All other V4 - V1 comparisons	< 10e-4	> 3
7a	Bootstrap	Same as bootstrap test in Fig 3a	All IT, V4, V1, and pixels task tests except as noted	< 10e-4	N/A
7a	Bootstrap	Same as bootstrap test in Fig 3a	V4 X-position	0.046	N/A
7a	Bootstrap	Same as bootstrap test in Fig 3a	V4 Y-position	0.038	N/A

7a	Bootstrap	Same as bootstrap test in Fig 3a	V4 Orientation	0.413	N/A
7a	Bootstrap	Same as bootstrap test in Fig 3a	Pixels Orientation	0.75	N/A
7a	Paired t-test	Same as t-test in Fig 3a	All IT/V4 comparisons except as noted	< 10 e-3	> 3
7a	Paired t-test	Same as t-test in Fig 3a	IT - V4 X-position	0.318	< 2
7a	Paired t-test	Same as t-test in Fig 3a	IT - V4 Y-position	0.092	< 2
4b	Bootstrap	That the population consistency is different from the human pool.	IT	0.0716	N/A
4b	Bootstrap	That the population consistency is different from the human pool.	V4	0.0043	N/A
4b	Bootstrap	That the population consistency is different from the human pool.	V1	< 10e-4	N/A
4b	Bootstrap	That the population consistency is different from the human pool.	Pixels	< 10e-4	N/A
4b	1-way ANOVA	That IT is consistent with the human population while the other populations are not		< 10e-4	F=164.52
6c	Bootstrap	That the correlation between categorization performance during training and performance on indicated task is greater than 0.		< 10e-10	N/A
6e	Bootstrap	That the model-neural performance correlation is greater than 0		< 10e-5	N/A
7b	Bootstrap	That performance is greater than 0	All non-pixel population tests	< 10e-3	N/A
7b	Paired t-test	That differences in performances between each model layer and the next is greater than 0, with the exception of the pixels.	All comparisons except as noted	< 10e-3	> 3
7b	Paired t-test	That differences in performances between each model layer and the next is greater than 0, with the exception of the pixels.	Layer1 - layer 3 X-Axis position	0.0334	2.301
7b	Paired t-test	That differences in performances between each model layer and the next is greater than 0, with the exception of the pixels.	Layer1 - Layer 3 Y-Axis Position	0.101	< 2
7b	Paired t-test	That differences in performances between each model layer and the next is greater than 0, with the exception of the pixels.	Layer1 - Layer 3 Orientation	0.493	< 2

7c: categorization	Paired t-test	That the difference in performance between IT-V4 and model layer6-layer3 at the left-most end of the spectrum is greater than 0.	IT-V4	0.441	< 2
7c: categorization	Paired t-test	That the difference in performance between IT-V4 and model layer6-layer3 at the left-most end of the spectrum is greater than 0.	Model3-Model6	0.285	< 2
7c: subordinate identification	Paired t-test	That the difference in performance between IT-V4 and model layer6-layer3 at the left-most end of the spectrum is greater than 0.	V4 - IT	< 10 e-3	> 3
7c: subordinate identification	Paired t-test	That the difference in performance between IT-V4 and model layer6-layer3 at the left-most end of the spectrum is greater than 0.	Model3-Model6	< 10 e-3	> 3
7c: Y-axis position	Paired t-test	That the difference in performance between IT-V4 and model layer6-layer3 at the left-most end of the spectrum is greater than 0.	IT - V4	< 10 e-3	> 3
7c: Y-axis position	Paired t-test	That the difference in performance between IT-V4 and model layer6-layer3 at the left-most end of the spectrum is greater than 0.	Model6-Model3	< 10 e-3	> 3
7c: object scale	Paired t-test	That the difference in performance between IT-V4 and model layer6-layer3 at the left-most end of the spectrum is greater than 0.	V4 - IT	0.0023	2.94
7c: object scale	Paired t-test	That the difference in performance between IT-V4 and model layer6-layer3 at the left-most end of the spectrum is greater than 0.	Model3-Model6	0.0774	< 2

Supplementary Table 1
Summary of statistical details.