SUPPLEMENTARY MATERIALS

Supplementary Section 1: reductions in firing rates and latency during PIT-input cooling

Here we describe the reduction in firing rate during PIT input deactivation. Below is the table of mean firing rates before and during cooling. Values were obtained using either all array channels, or only channels that showed a statistically reliable visual response.

Supplementary Table 1. Firing rates before and during cooling

	All channels							
Monkey	**Warm (spikes/s)**		**-V4**		**-V2|3**		**-V2|3|4**	
	Baseline	Evoked	Baseline	Evoked	Baseline	Evoked	Baseline	Evoked
R	79±4	97±5	76±4	89±4	78±4	91±4	-	-
G	118±2	140±3	104±2	119±2	106±2	120±3	107±2	121±2
	Visually driven channels							
	Warm (spikes/s)		**-V4**		**-V2|3**		**-V2|3|4**	
	Baseline	Evoked	Baseline	Evoked	Baseline	Evoked	Baseline	Evoked
R	81±5	110±6	78±5	98±6	80±5	101±6	-	-
G	121±4	168±5	103±3	136±4	105±4	137±5	109±3	144±3

Table I. Firing rate changes during cooling of areas V2, V3 and/or V4. Baseline period = 0-50 ms after image onset; evoked period = 51- 151 ms.

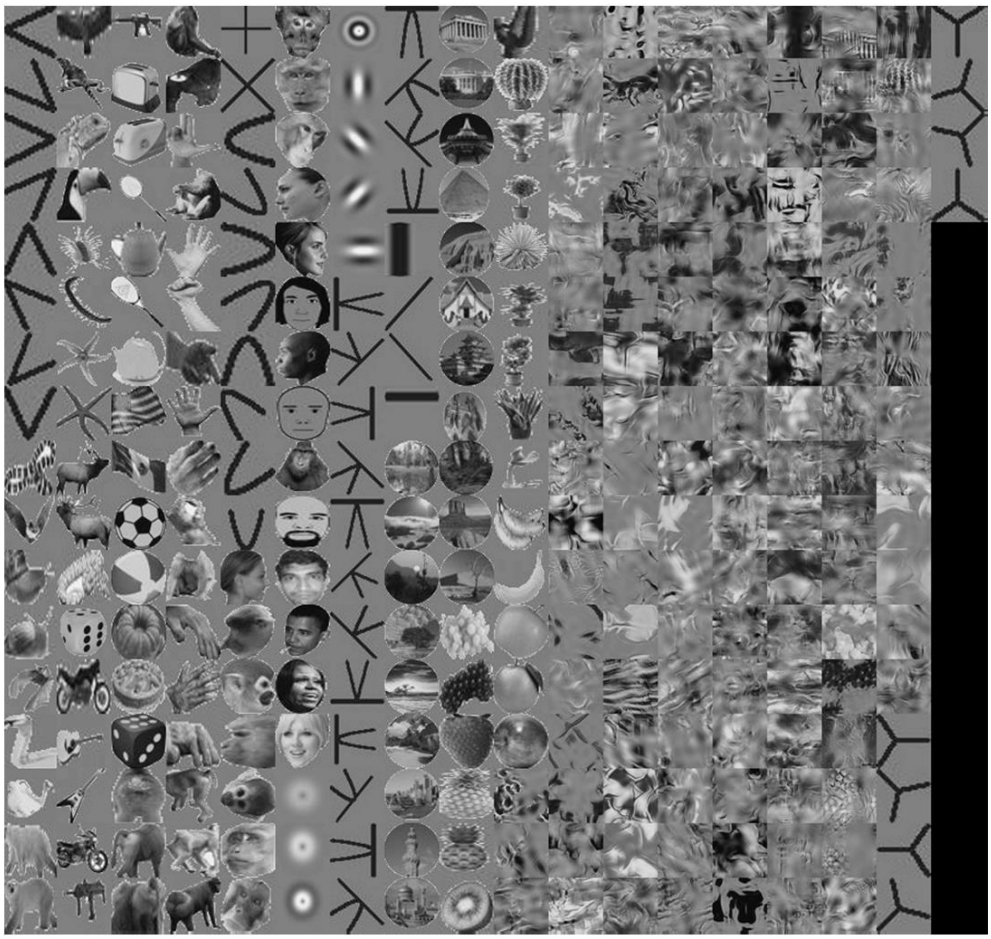
We considered two metrics for latency for each site: 1) response latency, defined as the earliest time after stimulus onset when activity rose two standard deviations above baseline, 2) tuning latency, defined as the time after stimulus onset when the tuning curve variance was highest. PIT multiunits showed no change in response latency; tuning latency was delayed by 10 ms in monkey R for both V4 and V2|3 cooling; and by 6 ms for monkey G during V2|3 deactivation)**.**

Supplementary Table 2. Latency values.

	Warm	–V4	–V2/V3
Response latency (ms)			
Monkey R	55.1±0.9	55.0±1.0	57.6±1.1
Monkey G	56.7±1.1	60.4±1.8	57.8±1.7
Tuning latency (ms)			
Monkey R	119.1±2.4	129.0±2.4	129.3±2.3
Monkey G	111.5±2.6	110.7±2.6	118.0±2.7

Figure S1. Image set

Supplementary Figure S1. Images used for testing. Each image was sized such that its longest axis was either 1.4° (monkey G) or 2.0° (monkey R), and placed within the same location in the screen.

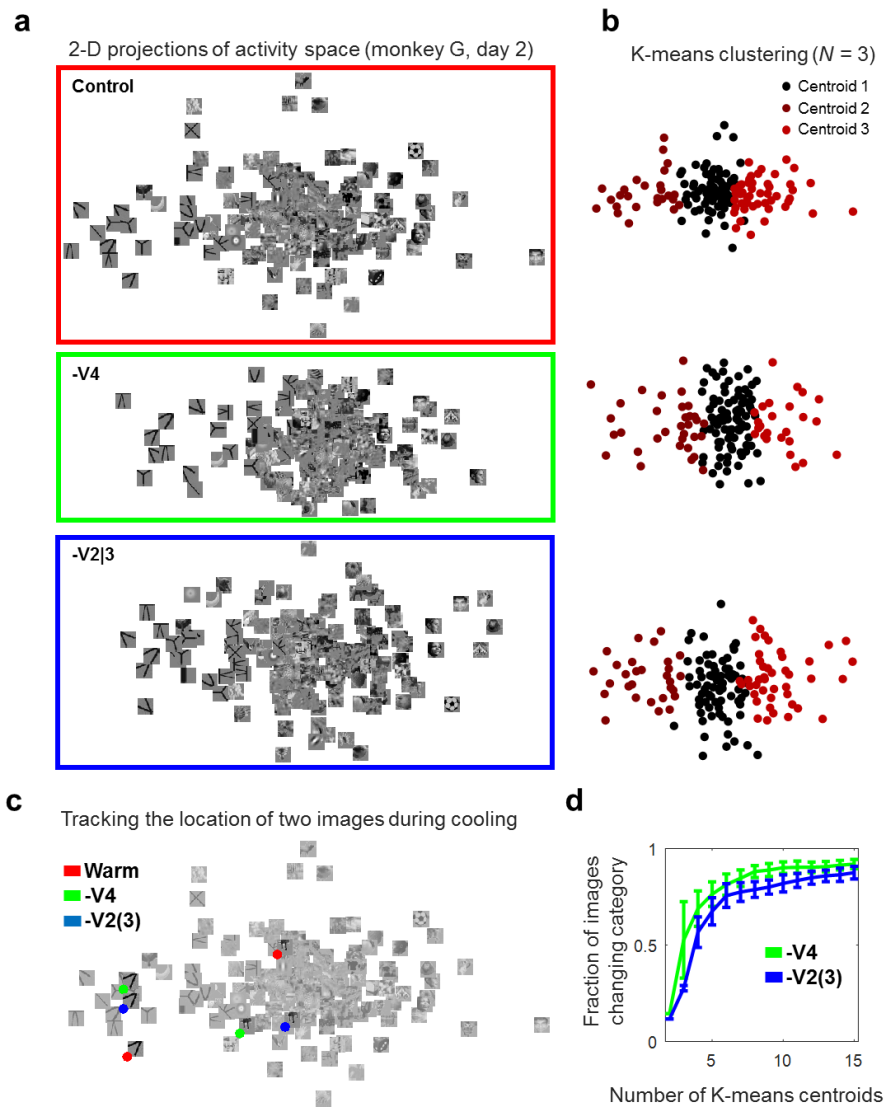


Supplementary Section 2: category classification by PIT-based classifiers during cooling

Inferotemporal cortex neurons cluster images into categories. More precisely, IT population response vectors represent images, and these vectors cluster together in activity space. When projected into 2-D space, these groups can be compelling, as they appear to represent “faces,” “body parts” and other groupings with semantic interpretation. There is some debate as to whether anterior IT populations encode for actual semantic categories, in addition to shape-based categories. Because our study focuses on posterior IT populations, we do not contribute to this debate. However, we were still interested whether our PIT populations encoded for any intrinsic clusters (heretofore referred to as “categories”), and if so, what happened upon input deactivation.

First, we projected population vectors from the original 100+ activity space into 2-D (via non-metric multidimensional scaling), to look for any intrinsic category representations. The map did not reveal segregated clusters but instead a spectrum with faces and line drawings at maximum separation (**Figure S2a**). During cooling, the area of this spectrum shrank, but cooling did not change its gross organization (line shapes remained in one side, faces in the other). Thus while we conclude that there were no intrinsic categories, we could still ask how images changed position during cooling. The key was to use specific centroids to divide the spectrum into territories (**Figure S2b**). We could then keep track of images as they migrated from territory to territory during cooling (**Figure S2c**).

We used the K-means algorithm to determine how many images changed territory during cooling of V4 or V2|3. Because K-means is a stochastic clustering algorithm, we ran each analysis 100 times, postulating 2 to 15 different centroids per pass (thus we ran 100 iterations per centroid number, or 1400 total passes). For each pass, we ran K-means on the control activity space first and saved the centroids. We then ran K-means on the cooling data, inputting the warm centroids, and counting how many images changed labels during cooling. We found that the fraction of individual images that changed territory varied with the number of centroids. The more centroids, the smaller the territory claimed by that centroid, and thus the more likely that an image would change label**.** When the activity space was divided into two clusters, only 12-14% of images moved enough to change clusters. With four clusters, up to 69±9% of images could change territory during V4 cooling, compared to 57±8% with V2|3 cooling (mean ± SEM). We found that generally V4 cooling induced more images to change territories compared to V2|3 cooling (**Figure S2d**).



Supplementary Figure 2. Effects of cooling on intrinsic image grouping. **a.** Two-dimensional plot of activity space for monkey G, pseudo-population 2. Top (red frame) = control (warm) data; middle (green) = V4 cooling, and bottom (blue) = V2|3 cooling.
b. K-means clustering of points in A, using three centroids (black, dark brown, red-brown). Top, middle and bottom correspond to control, -V4 and –V2|3.
c. Tracking the location of two images. Faded images show the control 2-D plot. The two saturated images are tracked before cooling (red dot), during V4 cooling (green dot) and during V2|3 cooling (blue dot).
d. Fraction of images changing centroid labels during cooling (green = V4 cooling, blue = V2|3 cooling), as a function of the number of centroids. Bars show the mean and standard error after 100 K-means repetitions.

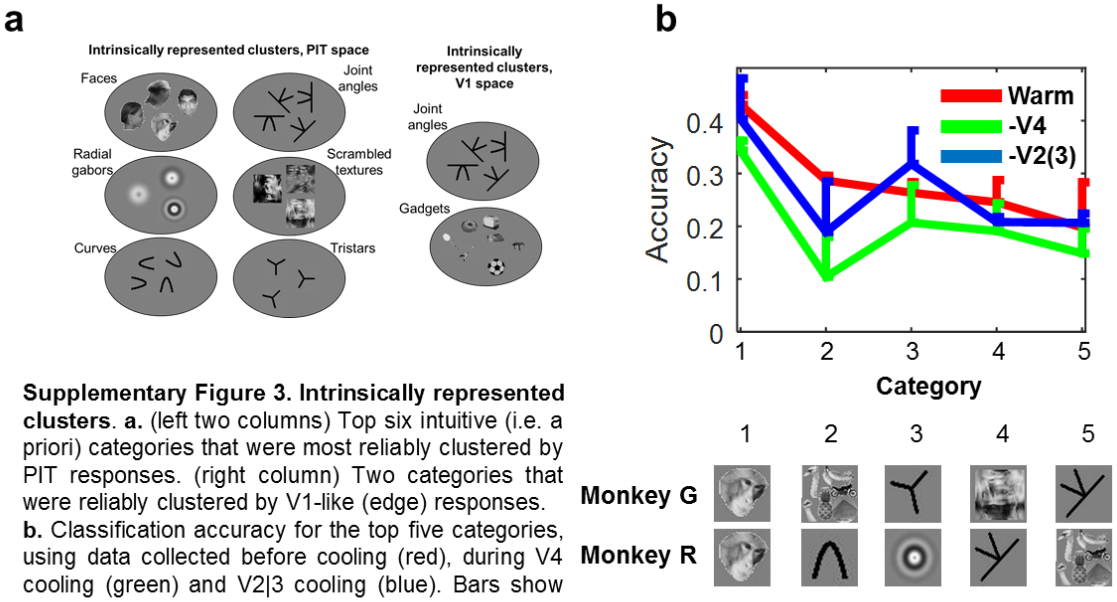
Digging deeper into clustering: group-based classification accuracy

Our choice of experimental images was inspired by categories reported to be robustly clustered in anterior IT, including faces, body parts, places, and others (see Methods)[22]. The previous analysis showed that our PIT units did not actually represent all of these categories as non-overlapping clusters in activity space. Instead, PIT placed images along a spectrum with faces on one side and line shapes on the other. We observed that cooling did not change the overall organization of this spectrum, though: faces and line shapes were polar opposite each time, in each monkey and each pseudo-population. Thus we proceeded to test whether these two categories were reliably classified using support vector machines. Our goal was not to label each of these images ourselves, but to let unsupervised algorithms “discover” them for us. Then we could test classifiers on these groups.

We started with a well-known analysis that measures the overlap between two different sets of categories[10,21]. The first set were the 15 categories we used to pick the images in the first place: angles, animals, artificial objects, curves, faces, radial and linear gabors, joint angles, plants, places, noise textures and tristars. We will call this the “intuitive” set. The second category set was unsupervised, derived using K-means. We ran K-means using 15 centroids, the same number as the original categories. We asked how much overlap there was between each of our intuitive categories and any of the K-means-derived categories.

The overlap value was defined using a metric[10,21] that counted the number of images at the intersection of a given pair of clusters (in this case, always one intuitive and one K-means-based), then divided that number by the total number of images in each cluster. The average of both ratios was defined as the overlap value. The procedure was repeated hundreds of times and objects were shuffled across categories to obtain a null distribution. PIT formed clusters with overlap values between 0.29-0.96 (monkey R, after correction with shuffling: -0.10-0.46) and 0.37-0.85 (monkey G, after correction with shuffling: -0.12-0.54). The categories that showed mean overlap values above 95% of all shuffled distributions were **faces** and **joint angles** (for both monkeys), **curves** and **radial gabors** (monkey R), **scrambled textures** and **tristars** (monkey G) (**Figure S3b**). This unsupervised clustering analysis thus confirmed the intrinsic grouping for faces and joint angles (at the extremes of the spectrum), and further added two other categories per monkey, both relatively simple in geometry. It was possible that all reliable categories were being clustered by PIT because they were physically similar at the pixel level, or at the level of a V1-like hierarchical stage. To test this hypothesis, we repeated the above analyses using grayscale pixel values (0-1) and a V1-like transformation of the pixel values. None of the pixel-based clusters showed a statistical overlap with our own categories (interestingly, artificial objects, joint angles and scrambled images showed the highest overlap values within this analysis; overlap values between 0.33-0.62, after correction with shuffling: -0.27-0.07). The V1-level responses did show statistical overlap with joint angles and artificial objects (0.49-0.88, -0.41-0.29 after correction). Thus we conclude that the intrinsic PIT clustering of simple images, like joint angles, could be explained by their physical similarity. Yet neither the pixels- or V1-like response analyses showed that faces were similar enough as a group to explain the intrinsic PIT clustering of faces.

We used support vector machines to determine if different categories showed disproportionate reductions in accuracy: as with the individual images, we saw that V4 deactivation led to larger drops in accuracy (5-18%, per-category average of both monkeys) than V2|3 deactivation (-1-10%). This difference held for all categories, and thus we saw no evidence that some intrinsic category representations are selectively affected by either V2|3 and V4 deactivation.



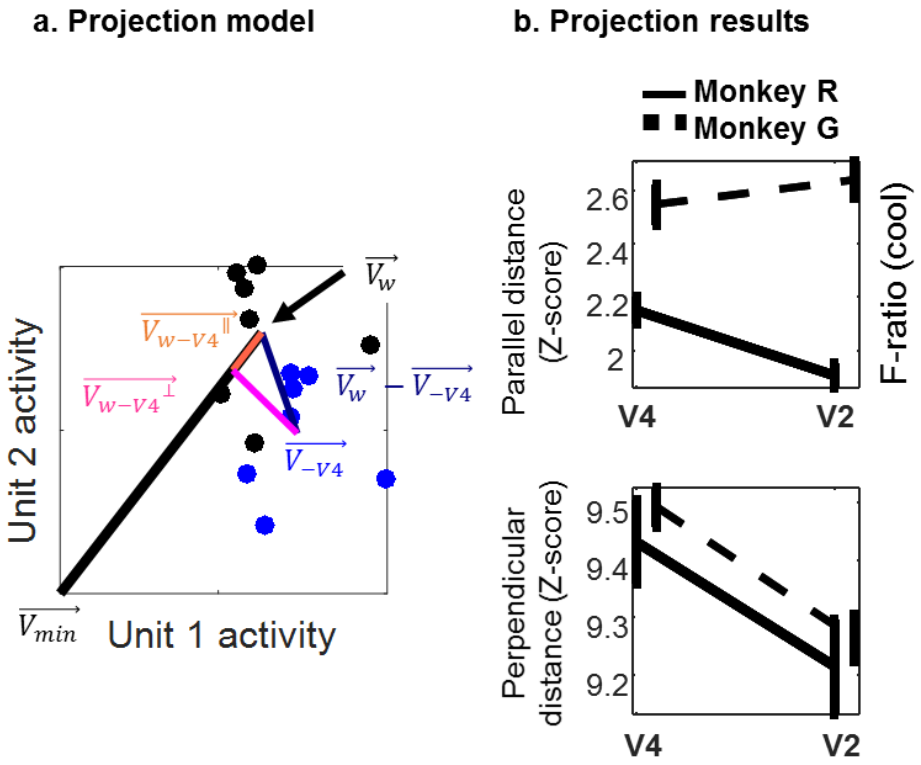
Supplementary Figure 3. Intrinsically represented clusters. **a.** (left two columns) Top six intuitive (i.e. a priori) categories that were most reliably clustered by PIT responses. (right column) Two categories that were reliably clustered by V1-like (edge) responses. **b.** Classification accuracy for the top five categories, using data collected before cooling (red), during V4 cooling (green) and V2|3 cooling (blue). Bars show mean ± SEM.

Methods on the above overlap analysis. The overlap values were defined using a commonly used metric[10,21]: briefly, for each intuitive cluster (e.g. "bodies") and each K-means-based cluster (e.g. "cluster 2"), we counted the number of images present in both ($n_{overlap}$), divided this number by the total number of images in the intuitive cluster ($n_{intuitive}$) and by the total number of images in the K-means cluster ($n_{unsupervised}$). Finally, we took the average of each ratio as the overlap score and used the maximum score per intuitive category. We repeated this analysis 500 times, to account for the fact that K-means algorithm is not deterministic and for statistical reliability: at each pass, we shuffled objects across categories and calculated randomized overlap scores to create a null hypothesis distribution.

Supp. Section 3: Cooling effects described as trajectories in neural activity space

Cooling V4 reduced decoding accuracy in PIT more than cooling V2|3 even though overall firing rate reductions were similar for V4 versus V2/3 cooling. We therefore explored neural state trajectories in multivariate activity space to explore the larger effect of V4 cooling on decoding accuracy. The activity space comprised the concurrent activity of all *N* PIT sites (where *N* = 100-300 sites depending on the monkey pseudo-population, see ***Methods***). We can visualize the neuronal representation of an image as a point (mean vector) in this multidimensional coordinate space[8,9]. Different mean vectors in the space represent different images. Excitatory drive was decreased during input cooling, which should reduce the length of the image vectors – moving them towards a minimum response vector. The difference in the control vs. cooling positions for a given image describes a *cooling* vector. If the main effect of cooling was to randomly remove spikes from the population, this cooling vector should be parallel to the minimum response vector. If the main effect of cooling was to alter the representational identity of each image, then the cooling vector should be more perpendicular (or at least *non*-parallel) to the minimum response vector. We therefore measured the *parallel* and *perpendicular* components of each image's V4 and V2|3 cooling trajectory vector, relative to the minimum response vector (**Supplementary Figure 4a**).

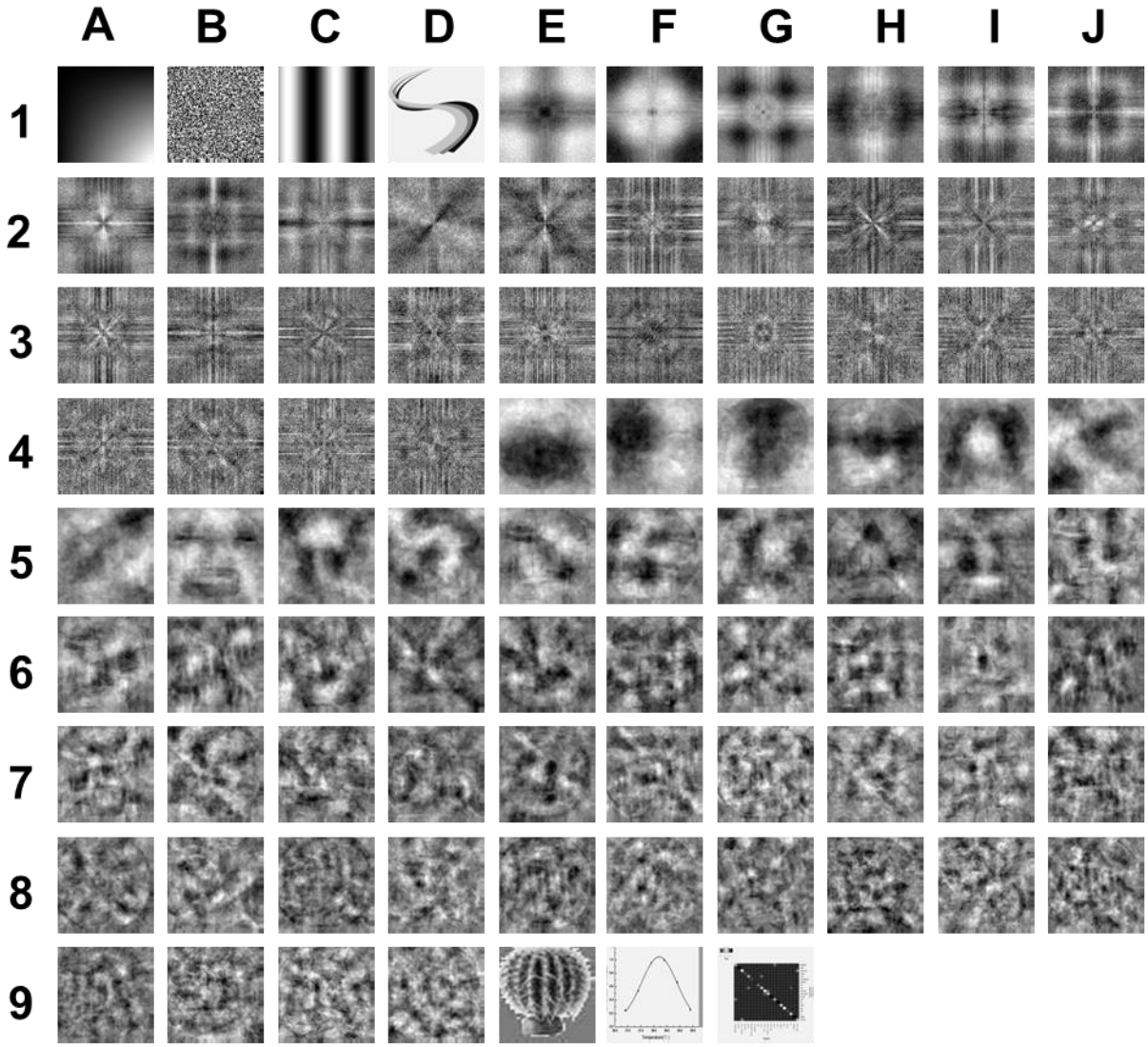
First, the largest component of the cooling vector was perpendicular to the minimum response vector. Second, the mean parallel component of the cooling vector was proportional to the firing rate changes: for monkey R, the mean parallel component was larger during V4 deactivation compared to V2|3 deactivation; for monkey G, the parallel components of both deactivations were roughly the same (Monkey r, norm of parallel component: **-V4:** 2.2±0.05, **-V2|3:** 1.9±0.05 | P = 6.6×10^{-8} per Wilcoxon sign rank test; Monkey g, **-V4:** 2.6±0.08, **-V2|3:** 2.6±0.07, P = 0.07). In contrast, for both monkeys, the mean perpendicular component of the V4 deactivation was larger than that of the V2|3 deactivation (Monkey r, norm of perpendicular vector component: **-V4:** 9.4±0.07, **-V2|3:** 9.2±0.08, P = 3.84e-08 per Wilcoxon sign rank; Monkey g, **-V4:** 9.5±0.03, **-V2|3:** 9.3±0.04, P = 4.08e-08). This indicates that V4 deactivation redirected image representations to further locations in the activity space than V2|3 deactivation, and this difference was independent of the observed reduction in spike rate (**Supplementary Figure 4b**).



Supp. Figure 4. Excitatory drive vs. classification accuracy analysis. a. Schematic explaining our projection analysis. The axes represent the responses of two IT units in response to multiple presentations of a single image before cooling (black dots) and during cooling (blue dots). The thick black line shows the axis between the mean control response and the minimum response vector $\mathbf{V}_{min}$. The blue line shows the trajectory between $\mathbf{V}_w$ and the mean cooling response $\mathbf{V}_{-v4}$. The orange and pink lines show the parallel and perpendicular components of the cooling trajectory vector.

b. Mean parallel and perpendicular components of the cooling trajectory vectors for each deactivation condition and monkeys (monkey R, solid lines, monkey G, broken lines).

Supplementary Figure 5: identifying features that can predict reductions in accuracy



159

Supplementary Figure 5. Features used for multiple linear regression model. Each cell represents a specific predictor, illustrated as either a graphical sketch or an actual principal component. Features include luminance (1A), contrast (1B), vertical power (1C), curvature (1D) , category (9E), control firing rate (9F), and control classification accuracy (9G). Features also included 50 principal components derived from the 293-image set (4E-9D, labeled by prefix “Pix”). Among these components, there were luminance contrasts between up and down (4E), left and right (4F), center and surround (4G), and further contrasts across smaller regions within the picture (4H-9D). Finally, we included 30 principal components derived from the discrete Fourier transforms of each image (1E-4D, labeled by prefix “FFT”). These components were contrasts between high and low spatial frequencies at all orientations (1E), band-pass frequencies against high and low (1G), and so on (1H-4D). The label of each cell is as follows:

'Luminance'	'Contrast‘	'Horizontal‘	'Curvature‘	'FFT1‘	'FFT2‘	'FFT3‘	'FFT4'	'FFT5'	'FFT6'
'FFT7‘	'FFT8'	'FFT9'	'FFT10'	'FFT11'	'FFT12'	'FFT13'	'FFT14'	'FFT15‘	'FFT16'
'FFT17'	'FFT18'	'FFT19'	'FFT20'	'FFT21'	'FFT22'	'FFT23'	'FFT24'	'FFT25'	'FFT26'
'FFT27'	'FFT28'	'FFT29‘	'FFT30'	'Pix1'	'Pix2'	'Pix3'	'Pix4'	'Pix5'	'Pix6'
'Pix7'	'Pix8'	'Pix9‘	'Pix10'	'Pix11'	'Pix12'	'Pix13'	'Pix14'	'Pix15'	'Pix16'
'Pix17'	'Pix18'	'Pix19‘	'Pix20'	'Pix21'	'Pix22'	'Pix23'	'Pix24'	'Pix25'	'Pix26'
'Pix27'	'Pix28'	'Pix29‘	'Pix30'	'Pix31'	'Pix32'	'Pix33'	'Pix34'	'Pix35'	'Pix36'
'Pix37‘	'Pix38'	'Pix39‘	'Pix40'	'Pix41'	'Pix42'	'Pix43'	'Pix44'	'Pix45'	'Pix46'
'Pix47'	'Pix48'	'Pix49‘	'Pix50'	'Category'	‘Rate'	‘Accuracy'			