5-10-2017

Posterior inferotemporal cortex cells use multiple input pathways for shape encoding

Carlos R. Ponce
Harvard Medical School

Stephen G. Lomber
Western University, steve.lomber@uwo.ca

Margaret S. Livingstone
Harvard Medical School

Systems/Circuits

Posterior Inferotemporal Cortex Cells Use Multiple Input Pathways for Shape Encoding

Carlos R. Ponce,[1] **Stephen G. Lomber,**[2] **and Margaret S. Livingstone**[1]

[1]Department of Neurobiology, Harvard Medical School, Boston, Massachusetts 02115, and [2]Department of Psychology, University of Western Ontario, London, Ontario N6A 5C2, Canada

In the macaque monkey brain, posterior inferior temporal (PIT) cortex cells contribute to visual object recognition. They receive concurrent inputs from visual areas V4, V3, and V2. We asked how these different anatomical pathways shape PIT response properties by deactivating them while monitoring PIT activity in two male macaques. We found that cooling of V4 or V2|3 did not lead to consistent changes in population excitatory drive; however, population pattern analyses showed that V4-based pathways were more important than V2|3-based pathways. We did not find any image features that predicted decoding accuracy differences between both interventions. Using the HMAX hierarchical model of visual recognition, we found that different groups of simulated "PIT" units with different input histories (lacking "V2|3" or "V4" input) allowed for comparable levels of object-decoding performance and that removing a large fraction of "PIT" activity resulted in similar drops in performance as in the cooling experiments. We conclude that distinct input pathways to PIT relay similar types of shape information, with V1-dependent V4 cells providing more quantitatively useful information for overall encoding than cells in V2 projecting directly to PIT.

Key words: convolutional networks; cooling; electrophysiology; inferotemporal cortex; V2; V4

Significance Statement

Convolutional neural networks are the best models of the visual system, but most emphasize input transformations across a serial hierarchy akin to the primary "ventral stream" (V1 → V2 → V4 → IT). However, the ventral stream also comprises parallel "bypass" pathways: V1 also connects to V4, and V2 to IT. To explore the advantages of mixing long and short pathways in the macaque brain, we used cortical cooling to silence inputs to posterior IT and compared the findings with an HMAX model with parallel pathways.

Introduction

In the macaque brain, posterior IT (PIT) neurons are the penultimate stage of the ventral visual processing stream, comprising cortical areas V1 → V2 → V4 → PIT → anterior IT (AIT). This main pathway represents a serial sequence of visual areas, but PIT also receives direct feedforward projections from V3 and V2 (Distler et al., 1993) whereas V4 receives direct inputs from V1 (Kuypers et al., 1965; Yukie and Iwai, 1985; Nakamura et al., 1993; Ungerleider et al., 2008). These shorter routes to PIT (V1 → V4 → PIT and V1 → V2 → PIT) have been called bypass pathways (Serre et al., 2005) and represent a significant fraction of the inputs to PIT: 14% of all neurons in the brain projecting to PIT are located in areas V2|3 (for context, 26% of inputs to PIT arrive from V4; only 1% of inputs to V1 come from the LGN) (Markov et al., 2011, 2014). The remaining projections arise from AIT and the dorsal pathway. The goal of this study is to define the roles of these different input pathways in PIT function by lesioning different input areas while recording from PIT units.

Despite a wealth of studies describing the behavioral effects of lesioning V2, V4, and PIT (Wilson and Mishkin, 1959; Cowey and Gross, 1970; Dean, 1976; Heywood and Cowey, 1987; Desimone et al., 1990; Merigan et al., 1993; Merigan, 1996; Merigan and Pham, 1998), there are surprisingly few studies reporting the electrophysiological effects of early extrastriate input lesions on IT neurons. In one study, it was shown that AIT neurons continued to respond selectively to complex images after aspiration of ipsilateral dorsal V4 and PIT, showing the same firing rate if stimuli were presented either in the intact or deafferented visual quadrants, although the same AIT units were also impaired in the

Received Aug. 23, 2016; revised April 6, 2017; accepted April 12, 2017.

Author contributions: C.R.P. and M.S.L. designed research; C.R.P., M.S.L., and S.G.L. performed research; C.R.P. analyzed data; C.R.P. wrote the paper.

This work was supported by National Institutes of Health Grant NEI EY 16187 to M.S.L., Burroughs Wellcome Fund to C.R.P., and Core Grant for Vision Research Grant NEI EY12196. We thank Gabriel Kreiman and John Assad for comments; and Tim LaFratta and John LeBlanc for machine shop support.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Carlos R. Ponce, Department of Neurobiology, Harvard Medical School, 220 Longwood Avenue, WAB 229C, Boston, MA 02115. E-mail: crponce@gmail.com.

DOI:10.1523/JNEUROSCI.2674-16.2017

filtering of distractor stimuli (Buffalo et al., 2005). Although it may be practically impossible to deactivate all inputs to AIT, it was surprising to learn that firing rate and shape selectivity were intact in AIT despite such gross input lesions. One potential explanation for this finding is that the lesions were chronic and thus might have engaged plasticity mechanisms that compensated for the effects of input lesions on AIT shape selectivity, for example, by increasing the input weights from bypass projections. To avoid the issue of long-term plasticity, we used reversible cortical cooling to deactivate inputs to PIT. We have previously used this technique to show transient impairments in MT neuronal tuning for binocular disparity and speed tuning during V2|3 cooling (Ponce et al., 2008, 2011). In this study, we recorded from unbiased samples of PIT neurons while deactivating areas V2-V3 (together) or V4 (see Fig. 1*a*). We measured PIT firing rates before and during cooling of V4 or V2|3, and quantified changes in the representational capacity of PIT. Our goal was to define the loss in excitatory drive in PIT neurons and to characterize how this loss in spike rate would affect the representational capabilities of PIT neurons. By deactivating V4 or V2|3, we affected the main input pathway (V1 → V2 → V4 → PIT) and allowed for the possibility to highlight any specific functions relayed through different bypass pathways (V1 → V2 → PIT during V4 cooling and V1 → V4 → PIT during V2 cooling). We found a significant impairment on firing rate and shape representations on PIT during partial deactivation of these extrastriate input areas, and further characterized a quantitative advantage of V4-based inputs relative to V2|3 inputs. Finally, we used a hierarchical feedforward model (HMAX) with multiple bypass pathways to test the representational capabilities of simulated "PIT" units receiving inputs from different pathways, and found that different input pathways can sustain object decoding with limited loss in performance, similar to the loss found during our cooling experiments.

Materials and Methods

All procedures were approved by the Harvard Medical School Institutional Animal Care and Use Committee, following the *Guide for the care and use of laboratory animals* (Ed 8). This paper conforms to the ARRIVE Guidelines checklist.

Behavior. Two adult male macaques (10–17 kg) were trained to perform a fixation task. The task required them to stare at a 0.5° wide red square in the middle of the screen, keeping their gaze within ±1.3° from the fixation spot. We used an ISCAN eye monitoring system to keep track of eye movements (www.iscaninc.com). The trial timeline was as follows: at the start of each trial, the fixation target appeared and the animal had up to 8 s to direct its gaze to the fixation target. Once fixation was acquired, a small reward could be delivered to encourage the animal. Within a random period between 17 and 117 ms after fixation onset, an image appeared perifoveally for 200 ms, then disappeared for 200 ms until a new image appeared. This on-off cycle could be repeated with 3–5 different images per trial. If the animal held fixation until the end of the final on-off cycle, a reward was dispensed. The reward size increased by 25% of the initial reward size every 100 trials.

Visual stimuli. We used MonkeyLogic to control experimental workflow (http://www.brown.edu/Research/monkeylogic/). As stimuli, we used 293 different images ranging from simple to complex. The simple images were line shapes, such as straight contours (four examples), angles (8), crosses (2), curves (8), tristars (8), radial and linear gabors (8), and combinations of lines and curves (joint angles, 16). These line shapes were generated using the Cogent MATLAB toolbox (developed by John Romaya at the LON at the Wellcome Department of Imaging Neuroscience; http://www.vislab.ucl.ac.uk/cogent_graphics.php). Our choice of complex images was guided by categories used in previous IT studies (Kiani et al., 2007; Kriegeskorte et al., 2008) and included animals (20 examples), artificial gadgets (20), body parts (20), faces (21 total, 10 monkeys, 11 humans), places (20), and plants/foodstuffs (20). Most of these images were used to create scrambled counterparts (118 examples) via the Portilla and Simoncelli (2000) visual texture model, which transforms white noise images into textures that share pairwise joint statistical constraints as the original intact images. These textures convincingly replicate small shape primitives present in the original images and scatters them throughout the image (118 textures). Images measured 1.4° (for Monkey G) or 2.0° (for Monkey R) at their longest axis. The images were not normalized for luminance.

Implanted devices. We used cryoloops (Lomber et al., 1999) composed of 23-gauge hypodermic stainless steel tubing, shaped to fit the individual curvature of each animal's occipitotemporal gyri/sulci as determined by structural magnetic resonance images. The cryoloops were 3.5 mm wide and between 4 and 11 mm long. A microthermocouple sensor was attached to the stem of the cryoloop to monitor its temperature. The bodies of the cryoloops were wrapped in Teflon tubing, except at the loop. The loops contained protected inlet/outlet ports that permitted the daily connection of Teflon tubes carrying chilled methanol, as driven by FMI "Q" Pumps (model QG150; www.fluidmetering.com). The methanol was contained within the tubing system and could not cause any chemical harm to the tissue. The custom-floating microelectrode arrays were manufactured by MicroProbes for Life Sciences; each had 32 platinum/iridium electrodes per ceramic base, electrode lengths of 4–16 mm, impedances between 0.7 and 1.0 MΩ, all connected to a 36-channel Omnetics connector (allowing for two additional grounds and two reference electrodes).

Surgical procedures. Both animals were implanted with custom-made titanium headposts before fixation training. After several weeks of postsurgical recovery and fixation training, the animals underwent a second surgery for the implantation of cryoloops and floating microelectrode arrays. In each animal, we performed a craniotomy centered at the lunate sulcus and extending anterolaterally. Monkey R received three cryoloops: two placed within the left lunate sulcus and one over the prelunate gyrus. The medial lunate sulcus loop was located 20 mm from the midline, traveled 7 mm deep into the sulcus, and was 3 mm wide; the lateral lunate sulcus loop traveled 4.5 mm into the sulcus and was 3.5 mm wide; the prelunate gyrus loop was placed anteriorly to the lunate sulcus loops, was 11 mm long, and 3 mm wide. Monkey G was implanted with two cryoloops: one over the prelunate gyrus and one within the lunate sulcus. The lunate sulcus loop was placed 2.1 cm from the midline, was 11 mm long with this axis running in the mediolateral axis within the lunate sulcus, 3 mm wide, and its most dorsal edge was 1.5 mm deep. The prelunate gyrus loop was also placed 2.1 cm from the midline, anteriorly to the lunate sulcus loop, ran 10.5 mm long, and was 3 mm wide. We collected thermal images to map the spread of cooling from the tubing, and confirmed that it was limited to 1–3 mm radially, as first shown in previous publications (Carrasco et al., 2013). Two or three floating microelectrode arrays were implanted within the same intraoperative session, after placement of the cryoloops. Their insertion sites were determined using three guidelines: they had to be anterior to the inferior occipital sulcus, many millimeters away from the prelunate gyrus cryoloop, and their location had to avoid large vasculature. All arrays were implanted caudal to the posterior middle temporal sulcus. We implanted two 32-channel arrays in Monkey R, and one 32-channel plus two 16-channel arrays in Monkey G.

Experimental session workflow. We describe results from data collected in 6–8 d from each animal. Each day, the animal would be head-fixed and its implants connected to the experimental rig: first, the cryoloops were connected to the chilled-methanol-bath tubing and temperature sensors; then the microelectrode arrays were attached to their headstages. The first step each day was to calibrate our measurements of the animal's gaze using the built-in MonkeyLogic routine. We used the Plexon Multichannel Acquisition Processor Data Acquisition System to collect electrophysiological information, including high-frequency ("spike") events, local field potentials, and other experimental variables, such as eye position, reward rate, and photodiode outputs tracking monitor frame display timing. Each channel was auto-configured daily for the optimal gain and threshold; we collected all electrical events that crossed a threshold of 2.5 SDs from the mean peak height of the distribution of electrical signal amplitudes per channel. These signals included typical single-unit waveforms, multiunit waveform bursts, and visually active hash.

The animal began its fixation task while we collected responses from the arrays with the cryoloops at body temperature (36°C–37°C; "control" or "warm" condition). After ~20 min of data collection to permit ~5 repetitions of each image, we activated either the V2|3 or V4 cryoloops, bringing the temperature of the cryoloops to 9°C–11°C, which lowered the temperature of the adjacent cortex to 16°C–18°C. We waited for another 5 repetitions of each image to pass, and then turned off the cryoloop pumps and collected 1–2 more repetitions under this first rewarming session. We then paused the fixation task for 10 min to allow the tissue temperature to increase and to preserve the animal's motivation for a second round of cooling. After 10 min, the temperature reported by the cryoloops was ~34°C, and we restarted our experiment. We repeated each image presentation 3 or 4 times and then activated the second set of cryoloop(s) (~8°C), waited for 5 repetitions, and turned off the cryoloops. We then collected data until the animal was satiated. We balanced the order of the V2|3 vs V4 cryoloop activations: if on the first day we activated the V4 cryoloop first and V2|3 cryoloops second, the next day we activated the V2|3 cryoloops first and V4 cryoloop second. There was an even number of days for each cooling order.

Spike data preparation. The raw data files comprised event ("spike") times per channel for the entire experimental session (the number of channels available per day were 64, but not all provided reliable signal-to-noise values). We divided each daily dataset into thousands of raster plots defined by the onset of each image presentation and labeled each raster plot with its corresponding channel, image name, and temperature condition. We defined three windows of analysis: the baseline period lasted from 0 to 50 ms after image onset, the early period from 51 to 150 ms after image onset, the late period from 151 to 250 ms after image onset; a full image presentation window was 51–400 ms after image onset. We found that multiunit responses could last almost 400 ms, although their peak responses always occurred within the early window. Here we report responses within the early window minus the activity within the baseline window (we call these evoked responses). For all multivariate analyses, we normalized the activity of each site by transforming its evoked spike rate responses to *z* scores: all evoked responses emitted by a single site during an experimental daily session were averaged, this mean response was subtracted from all individual evoked rates, and each value was then divided by the SD of all evoked responses.

Although our full dataset contained 293 images, we did not have enough time to present all images every day and still get the minimum of >15 presentations across the control and cooling conditions. Thus, we presented more than half of the total image set each day (10 images from each complex category, such as faces and places, along with half of the scrambled textures per day, with most of the simple line shapes, rounding to ~148–177 unique images per day). The responses of each individual channel were correlated from day to day but were also statistically different by multivariate descriptors, such as multidimensional scaling. Because of these differences, we did not combine channel information across days and instead created a multiday pseudopopulation, where sets of concurrently recorded channels ($N = 50–64$) from different days were "stacked" on top of each other. Thus, the final activity space is defined by "site-days," where some dimensions represent responses from the same channel to the same image collected on different days. Because the whole image set was presented on different days, we had two pseudopopulations per animal, each containing different site-day responses to each half of the image set. Each of our pseudopopulations had between 100 and 300 multiunits.

Scotoma mapping experiments. The goal of these experiments was to identify the parts of the retinotopic field that were captured by our arrays, and the relative location of the response impairment caused by cooling. The animals fixated while we presented a single image (black-and-gray diagram face, 2.0° wide) across positions in a radial grid (angular coverage of 0°–315°, 45° steps; radial coverage of 0°–8° from the center of the screen, in 0.5° steps). Three to five positions were randomly chosen per trial. After data collection, we defined evoked responses per position as follows: first, we quantified the firing rate per site during the early window of activity (51–151 ms after stimulus onset) and then subtracted the firing rate per site during the baseline window of activity (0–50 ms after stimulus onset). We averaged these evoked responses per position within each site and used the griddata.m MATLAB function to interpolate the scattered data into a continuous map. This map was smoothed using a 1° diameter disk filter. This map represented the aggregate receptive field (RF) of each multiunit site in our arrays. To identify the overall scotoma, we averaged the receptive fields of all sites during the control condition and subtracted the average receptive fields of all sites during V2|3 or V4 deactivation. We measured the size of each scotoma by hand, using the calcArea.m function (http://www.mathworks.com/matlabcentral).

Firing rate and latency analyses. The goal of these analyses was to measure changes in the overall firing rate (excitatory drive) of PIT multiunits during input deactivation. These changes included the amplitude of peristimulus rate histograms (PSTHs) and the latency of response. To quantify the changes in evoked response magnitude, we computed the evoked responses per site as described in Spike data preparation and averaged these responses across all channels within each temperature condition. We did the same operation using *z* scores. We calculated the probability that the median responses emitted during each temperature condition (control, V4, and V2|3 cooling) were sampled from the same distribution using a Kruskal–Wallis one-way ANOVA. To determine whether there was a statistical difference between the V4 and V2|3 cooling condition responses, we used the Wilcoxon signed rank test for zero median. For the latency analyses, we obtained the mean PSTH in response to each image, per site and temperature, and then stacked all image-specific PSTHs in a matrix measuring $N_{\text{images}} \times 400$ (ms after stimulus onset). We identified the time when each PSTH exceeded 2 SDs over baseline and called this response latency, with the only acceptance criteria that a plausible response latency would only occur between 30 and 200 ms after image onset. We also computed the earliest time point when all PSTHs demonstrated the greatest variance in amplitude, as an indicator of the tuning latency.

Identification of channels with reliable visually driven activity. Many electrodes in the arrays reported electrical activity that was not visually driven, possibly because the electrodes were on the pial surface. We repeated some analyses only using channels that showed a statistical difference in mean activity between the baseline and evoked time periods. Using a cross-validation approach, we used 5% of all trials to perform a Wilcoxon signed rank test for the median rate difference during each interval. This told us which channels showed a statistical difference in rate during visual stimulation. We then used the remaining 95% of trials to compute the firing rates during baseline and evoked windows for the selected channels. Monkey R's arrays showed 38 of 64 visually responsive sites; Monkey G's arrays showed 30 of 64 visually responsive sites ($p < 0.05$, Wilcoxon signed ranked test for zero median).

Encoding accuracy analyses. We trained support vector machines (SVMs) with a linear kernel using the MATLAB function fitcecoc.m. We used a modified one-versus-one approach, with one classifier c_{ik} trained to discriminate between every image pair *i* and *k*, where *i* is the positive class and *k* the negative class. In the image identification task, each c_{ik} was trained/tested using leave-one-out cross-validation, and in the category task, with fivefold cross-validation. After testing, instead of choosing the best classifier via arg max(c_{ik}) as the final vote, we averaged the accuracy scores of all c_{ik} for each given image *i*. We interpreted this as postulating that there exists a downstream neuron for each possible image pair classification, and the performance for a given pair classification is the average over this population of neurons. To estimate the chance accuracy for each paired comparison, we concurrently trained SVMs using the same set of data vectors but with shuffled labels. To estimate the reliability of the classification accuracy values for each individual image, we used binomial tests because the leave-one-out cross-validation resulted in dichotomous scores of 0, 1. First, we used binofit.m to estimate a 95% CI for the accuracy value of the shuffled-label classifiers (using all classifiers where a given image was the positive class). We then asked how likely it was that the correct-label classifiers would show their observed accuracy value if the underlying probability is the same as the upper CI value of the shuffled classifiers (via binocdf.m).

There were 4 or 5 response vectors per class within each comparison (the data used for classification were *z* score vectors; see Spike data preparation). The number of vectors for each two-class comparison was small, and thus we found that chance accuracy values could vary between 0 and 1 across all image versus image comparisons; the median shuffled-label misclassification rates were 0.60–0.63 for Monkeys R and G. We subtracted the chance, shuffled-label accuracy classification rate from the correct-label accuracy classification rates to account for this bias. As an insight to explain this deviation from the expected chance accuracy of 0.5, we trained SVMs to distinguish between stimulus categories (listed in Visual stimuli). Each category pair comparison involved 10–20 times as many response vectors as the individual image-versus-image SVM analyses, and the dataset was otherwise identical. Here we found a reassuring shuffled-label statistical baseline of 0.50 in both animals. Both the category and image-per-image SVM accuracy analyses led subsequently to the same conclusions.

Projection analysis. The goal of this analysis was to reconcile the findings that cooling V2|3 and V4 led to different reductions in PIT population firing rates, but the same relative reductions in classification accuracy. To do this, we defined a neural activity coordinate system where each unit's activity forms one dimension (given at least $N = 100$ dimensions per pseudopopulation) and each image I_i ($i = 1$–293) is represented by a vector of coordinates $\vec{\boldsymbol{v}}_i$ with N elements. We defined the cooling trajectory traveled by each image during deactivation as the difference vector $\vec{\boldsymbol{t}}_i^{-V4} = \vec{\boldsymbol{v}}_i^{warm} - \vec{\boldsymbol{v}}_i^{-V4}$ and $\vec{\boldsymbol{t}}_i^{-V2/3} = \vec{\boldsymbol{v}}_i^{warm} - \vec{\boldsymbol{v}}_i^{-V2/3}$, where $\vec{\boldsymbol{v}}_i^{warm}$ was the population response vector for image I_i before cooling, $\vec{\boldsymbol{v}}_i^{-V4}$ is the response vector during V4 cooling, and $\vec{\boldsymbol{v}}_i^{-V2/3}$ during V2|3 cooling. Each cooling trajectory vector $\vec{\boldsymbol{t}}_i^{-V4}$ or $\vec{\boldsymbol{t}}_i^{-V2/3}$ was projected against a minimum response vector $\vec{\boldsymbol{v}}_i^{proj} = \vec{\boldsymbol{v}}_i^{warm} - \vec{\boldsymbol{v}}_{min}^{warm}$, where $\vec{\boldsymbol{v}}_{min}^{warm}$ comprised N elements, which were the lowest response values per unit in a given pseudopopulation and represented a nonspecific reduction in firing rate across all sites. Thus, each cooling trajectory $\vec{\boldsymbol{t}}_i^{-VX}$ (where $-VX$ could be −V4 or −V2|3) was separated into a projected (parallel) component and a rotational (perpendicular) component $\vec{\boldsymbol{t}}_i^{-VX} = \vec{\boldsymbol{t}}_{i,\|}^{-VX} + \vec{\boldsymbol{t}}_{i,\perp}^{-VX}$, representing a simple population gain change versus a population representational change.

*Selectivity analyses (*F *statistics).* In this analysis, we used the *F* statistic as a measure of selectivity for each cortical site. The *F* statistic is a ratio of mean squares, specifically the mean square error estimate for the variance of responses among images, divided by the mean square error estimate for the variance within each image. We computed each *F* statistic in a channel-by-channel basis using the responses to all images within each temperature condition. For each channel, one *F* statistic was computed using the warm data ($F_{control}$), another using the V4 cooling data ($F\text{-}_{V4}$), and another using the V2|3 cooling data ($F\text{-}_{V2|3}$). We plotted each $F_{control}$ against its paired $F\text{-}_{V4}$ and $F\text{-}_{V2|3}$ values. To determine whether the slope in each given scatterplot was different from unity, we used a bootstrap approach, where we computed 1000 different slopes by sampling each channel with replacement (we kept each *F* ratio trio together; we did not mix warm and cooling *F* statistics from different channels). We then asked whether the slope distribution from this bootstrap included 1.

We used randomization to determine whether there was a difference between the mean slopes computed during the V4 and V2|3 cooling conditions (i.e., whether there was a difference between the mean $F\text{-}_{V4}/F_{control}$ slope vs the mean $F\text{-}_{V2|3}/F_{control}$ slope). The null hypothesis is that the mean V4 and V2|3 slopes came from the same distribution. We created this distribution by randomly mixing the labels of the V4 and V2|3 *F* statistics 999 times. At each pass, we sampled two subsets of $F_{control}$ and $F_{cooling}$ pairs, computed their regression slopes, and took the slope difference. We then compared the observed slope difference against this distribution.

Simulation of loss of decoding accuracy by perturbing control vectors. To simulate the potential effects of cooling on the control-condition population response vectors, we manipulated the norm and/or angle of each response vector. First, we defined the matrix $\boldsymbol{R}_{warm}^{m,p}$ for monkey m and pseudopopulation p, measuring $c \times t$, where c is the number of units and t is the number of image presentations under the warm condition. To simulate gain changes, we changed the mean vector norm of each vector $\vec{\boldsymbol{r}}_{i\,warm}^{\,m,p}$ in $\boldsymbol{R}_{warm}^{m,p}$ by sampling t values from a Gaussian distribution with mean of either 0.1, 0.2, 0.3, . . . , 1 (SD of 0.025) and multiplying each vector by one of the t values. To simulate representational changes, we changed the mean vector angle of $\vec{\boldsymbol{r}}_{i\,warm}^{\,m,p}$ relative to its original position; this was more difficult because there is a large number of ways to change the angle of a vector in c-dimensional space depending on the chosen plane (there are *c-choose*-2 possibilities, given that c can be >100; i.e., at least 4950 planes of rotation), and further, some of these artificial rotations could occur outside of the natural tuning shown by the PIT population. Thus, to pick the rotation angles to transform all vectors in $\boldsymbol{R}_{warm}^{m,p}$, we used the following approach. We defined the range of rotations of the warm population by first computing the mean population response vector $\bar{\vec{\boldsymbol{r}}}_{warm}^{m,p}$ (measuring $c \times 1$) and subtracting all vectors in $\boldsymbol{R}_{warm}^{m,p}$ from $\bar{\vec{\boldsymbol{r}}}_{warm}^{m,p}$, resulting in a matrix of trajectory vectors $\boldsymbol{T}^{m,p} = \bar{\vec{\boldsymbol{r}}}_{warm}^{m,p} - \boldsymbol{R}_{warm}^{m,p}$. These trajectory vectors contained the planes of rotation that occurred naturally in the PIT population along within the warm condition. To isolate the pure rotational components of these trajectory vectors, we separated each vector in $\boldsymbol{T}^{m,p}$ into parallel and perpendicular components $\boldsymbol{T}^{m,p} = \boldsymbol{T}_{\|}^{m,p} + \boldsymbol{T}_{\perp}^{m,p}$ relative to the vector $\bar{\vec{\boldsymbol{r}}}_{warm}^{m,p} - \min(\boldsymbol{R}_{warm}^{m,p})$, where $\min(\boldsymbol{R}_{warm}^{m,p})$ is the minimum response vector (baseline-subtracted, so it is not the zero vector). The parallel vectors $\boldsymbol{T}_{\|}^{m,p}$ isolated the change in gain, and the perpendicular vectors $\boldsymbol{T}_{\perp}^{m,p}$ pointed to the rotational directions in the natural planes. To simulate the mean angular rotation during cooling, we extended the length of each perpendicular trajectory vector $\vec{\boldsymbol{t}}_{i,\perp}^{\,m,p}$ by a factor f (for each "cooling" simulation, the value could be $f = 0.1$, 0.2, 0.4, 0.8, 1.6, 3.2, 6.4, 12.8, and 25.6, which corresponded to mean angular changes of 3°, 8°, 18°, 29°, 38°, 45°, 50°, 53°, 56° across all monkeys and pseudopopulations). We then sampled these modified trajectory vectors randomly and added them to $\boldsymbol{R}_{warm}^{m,p}$ vectors to change their direction in c-dimensional space, making sure that each rotated vector was matched in norm as the original vector. We combined different mean values of angular and gain changes in our various simulations to give rise to a number of $\boldsymbol{R}_{\text{“}cooled\text{”}}^{m,p}$ populations used for linear classification analyses.

Linear regression model. The goal of this analysis was to determine whether the change in classification accuracy during V4 cooling or during V2|3 cooling could be predicted using different image features. The regression matrix had dimensions of 293 × 87 (images × visual features). The features were luminance (defined as the mean pixel value transformed by the monitor's gamma function), contrast (variance of the pixel values transformed by the monitor's gamma function), horizontal versus vertical power (obtained via a wavelet decomposition analysis using the MATLAB function wavedec2.m), curvature (defined by the variance of each image's discrete Fourier transform spectral power around all orientations), 50-pixel-based principal components as defined by the pca.m function), 30 spatial frequency principal components (pca.m applied to the discrete Fourier transformed images), categorical membership (i.e., angles, animal, artificial, bodies, cross, curve, face, gabors, radial gabors, joint angles, line, places, plants, scrambled, tristar), the mean population control firing rate per image, and control classification accuracy per image. Values within each feature group were *z* scored before fitting. The dependent variables were as follows: (1) accuracy loss during V4 cooling (control accuracy per image minus −V4 accuracy); (2) accuracy loss during V2|3 cooling (control accuracy per image minus −V2|3 accuracy); or (3) the difference in accuracy loss during V4 minus V2|3 cooling ([control accuracy per image minus −V4 accuracy] − [control accuracy per image minus −V2|3 accuracy]). The probability that the linear model differed from the constant model was obtained two ways: first, we used the *t* statistic provided by the "fitglm.m" function; second, we used a randomization test where the dependent variable was fit with a regression vectors made up of random numbers, sampled from a flat distribution. The table had the same dimensions as the true data matrix table. The R^2 values of 1000 randomization tests were compared with the R^2 from the regular regression table. To identify the most interesting predictors, we used the regression weights with the highest *t* statistics.

Standard model of visual recognition. Our computational model was based on an implementation by Serre et al. (2007b), available at http://cbcl.mit.edu/software-datasets/standardmodel/index.html. This model belongs to the HMAX family, inaugurated by Riesenhuber and Poggio (1999) and developed over subsequent publications (Serre et al., 2005, 2007a, c). The model represents the visual object recognition system as a series of convolutional and pooling operations, which transform an image from pixels into neuronal responses. These responses can be used in a statistical classifier to decode their abstracted representational content.

The architecture of our network was four layers deep and contained three pathways: one main pathway and two bypass pathways. The main pathway had four layers: layer 1 (representing V1), layer 2 (V2|3), layer 3 (V4), and layer 4 (PIT units receiving inputs from the main pathway). The second pathway had three layers: layer V1, layer V4b (representing units in V4 receiving direct input from V1), and layer PITb (PIT units receiving input solely from the V1 → V4 inputs). The third pathway also had three layers: layer V1, layer V2|3, and layer PITc (units in PIT receiving input directly from V2|3). These three types of "PIT" neurons showed different kinds of activation patterns, which we could decode using SVMs. Each layer represented a stereotypical set of operations: a convolution/tuning operation and a pair of max operations. The tuning operation is equivalent to a simple cell, which convolves the input with a filter bank via the tuning function $r = \exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^{Ncomb}(w_j - x_j)^2\right)$, where σ = sharpness parameter, N_{comb} is the number of filters to combine, w = filter weight, and x = input image or activity. This simple cell operation describes the Euclidean distance between the RF shape and the incoming input. Different simple cells are characterized by different shapes and sizes of their RF patches. There is more than one RF sizes at each layer, and each filter-size convolution is performed in parallel. Several outputs of this tuning operation are then combined in a complex-cell-like operation. Complex cells perform a pooling operation: they receive inputs from N_s simple cells with different RF sizes and compute the maximum response emitted by the set. Thus, the output of a complex cell layer is sparser than the output of a simple cell layer because maximum values are repeated across limited areas of response space. These complex layer responses are finally subsampled, imitating the decreasing number of cells that can cover visual space as one moves down the visual pathway.

Building the model required two major implementation stages: (1) we had to create RF patches for each layer; and (2) we had to use these RF patches to compute responses to our experimental images. As in the 2007 publication, the patches were imprinted using experience-dependent activity. Each layer contained a set of up to 200 unique filters: V1 layer filters were Gabors at four orientations and eight sizes (3–10 pixels wide, or 0.12°-0.39° wide given our monitor distance). Subsequent layer filters were imprinted using random samples of activity from the preceding layer. For example, after randomly selecting an image from the Caltech database, we processed it through the V1 layer, and the 2-D response image was randomly sampled to create a smaller patch that represented weights for V2|3. After repeating this process hundreds of times, we selected new images from the database, processed them through V1 and V2|3, and used the resulting activity to shape the V4 filters. This was repeated up to the second-to-last layer. To make sure that the RF shapes would match the statistics of natural images presented close to and far from the fovea, we also imprinted using differently sized variants of the same images (1°, 2°, and 4° wide versions of the same image). In the first layer, filter sizes were 0.1°–0.4° in width (in our experimental setup, 1° = 26 pixels) and doubled at every hierarchical step, with the exception of the bypass pathways, where filter sizes quadrupled in width at the skip level (e.g., "V4" filters in the "V1 → V4" pathway were 4 times the size of "V1" filters; PIT filters in the "V1 → V2|3" pathway were 4 times the size of "V2|3" filters). To obtain the responses for the decoders, we transformed our 293 experimental images into PIT responses using the fully assembled network and tested each PIT population using SVMs, as in the experiments above. To introduce variability into the model's response vectors during presentations of the same image, we created six variations of each image by adding random changes in position to simulate fixational eye movements. These fixational eye movements were simulated by measuring the distribution of each monkey's eye position during image presentation across trials, fitting with Gaussian models, then randomly sampling this distribution to change the center of each object within the image frame. The Gaussian models had mean 0° and 0.15°–0.18° SDs (corresponding to Monkeys G and R).

The key contrast involved the relative performance between all simulated PIT units and the subsets of simulated PIT units receiving inputs from each bypass pathway only: we considered the full output population to represent our control temperature condition, and the smaller populations to represent the cooling conditions. Each population was used to train SVMs that tested the linear classifiability of each image against each other image. As we did with the biological units, we defined accuracy as the percentage of correct choices over the shuffled-label accuracy.

Results

Cooling affected portions of PIT receptive fields

We implanted floating microelectrode arrays in PIT of two adult male monkeys (2 arrays in Monkey R and 3 arrays in Monkey G) along with cryoloops in retinotopically corresponding dorsal V2, V3, and V4 (Gattass et al., 1981, 1988). The arrays were placed anteriorly to the inferior occipital sulcus, the cryoloops were placed within the lunate sulcus and over the predorsal gyrus (Fig. 1*b*). We activated the cryoloops intraoperatively, using thermal imaging to plot the extent of cooling and found that the lower thermal region was limited to 1–1.5 mm around and within the cryoloop (Fig. 1*c*). The electrode arrays were at least 5 mm anterior to the prelunate cryoloop, and anterior to the inferior occipital sulcus. After postsurgical recovery of the animals, we visualized the population RF of the arrays by measuring the mean spike rate of all electrodes while flashing a 2° diameter image randomly within a 16 × 16° radial grid. We found that the mean PIT population receptive fields covered by the arrays were biased toward the upper contralateral hemifield but also included the lower perifoveal hemifield (Fig. 1*d*). We then collected spike data using the same stimulus positions during deactivation of V2/V3 and V4. By subtracting the population receptive fields collected during cooling from the receptive fields collected during the control condition, we were able to estimate the retinotopic extent of the impaired inputs. If cooling was solely affecting PIT, we would have expected that all PIT responses would decrease uniformly. Alternatively, if cooling indeed affected V2|3 and V4, we would expect that the core region of impairment would be biased to the lower perifoveal hemifield, as predicted by the retinotopy of V2 and V4 in the dorsal brain, and this is what we found (Fig. 1*e*). Deactivation of V4 resulted in scotomas with an estimated size of $6.7°^2$ and $7.5°^2$ (Monkeys R and G), while deactivation of V2|3 resulted in scotomas with an estimated size of $9.1°^2$ and $5.7°^2$ (Monkeys R and G), all centered in the same location. In subsequent experiments, stimuli were sized to fit within the overlapping region of both scotomas (1.4° wide images for Monkey G, 2.0° wide images for Monkey R).

Cooling reduced firing rates in PIT units

For all following experiments, we showed the fixating animals 293 images belonging to 15 different categories (angles, animals, artificial objects, curves, faces, radial and linear gabors, joint angles, plants, places, noise textures, and tristars). When we cooled either set of cryoloops, PIT multiunits showed reduced visually evoked responses (Fig. 2*a*). We defined visual responses as the rate of spikes within 50–150 ms after image onset − the rate of spikes within the first 50 ms after image onset. Over all electrodes in our arrays, PIT multiunits showed a mean visual response of 18 ± 1 (Monkey R) and 22 ± 2 (Monkey G) spikes/s (range of −3 to 120 spikes per seconds for Monkey R; −12 to 106 spikes per seconds for Monkey G). When the V2/V3 loops were cooled, the overall average rate was reduced to 13 ± 1 and 14 ± 1 spikes/s (Monkeys R and G). When the V4 cryoloops were cooled, the overall rate was reduced to 12 ± 1 and 15 ± 1 spikes/s (Monkeys R and G; for all values, see Table 1). The probability that the cooled mean responses arose from the same distribution as the precooling responses was 1×10^{-3} per randomization one-way

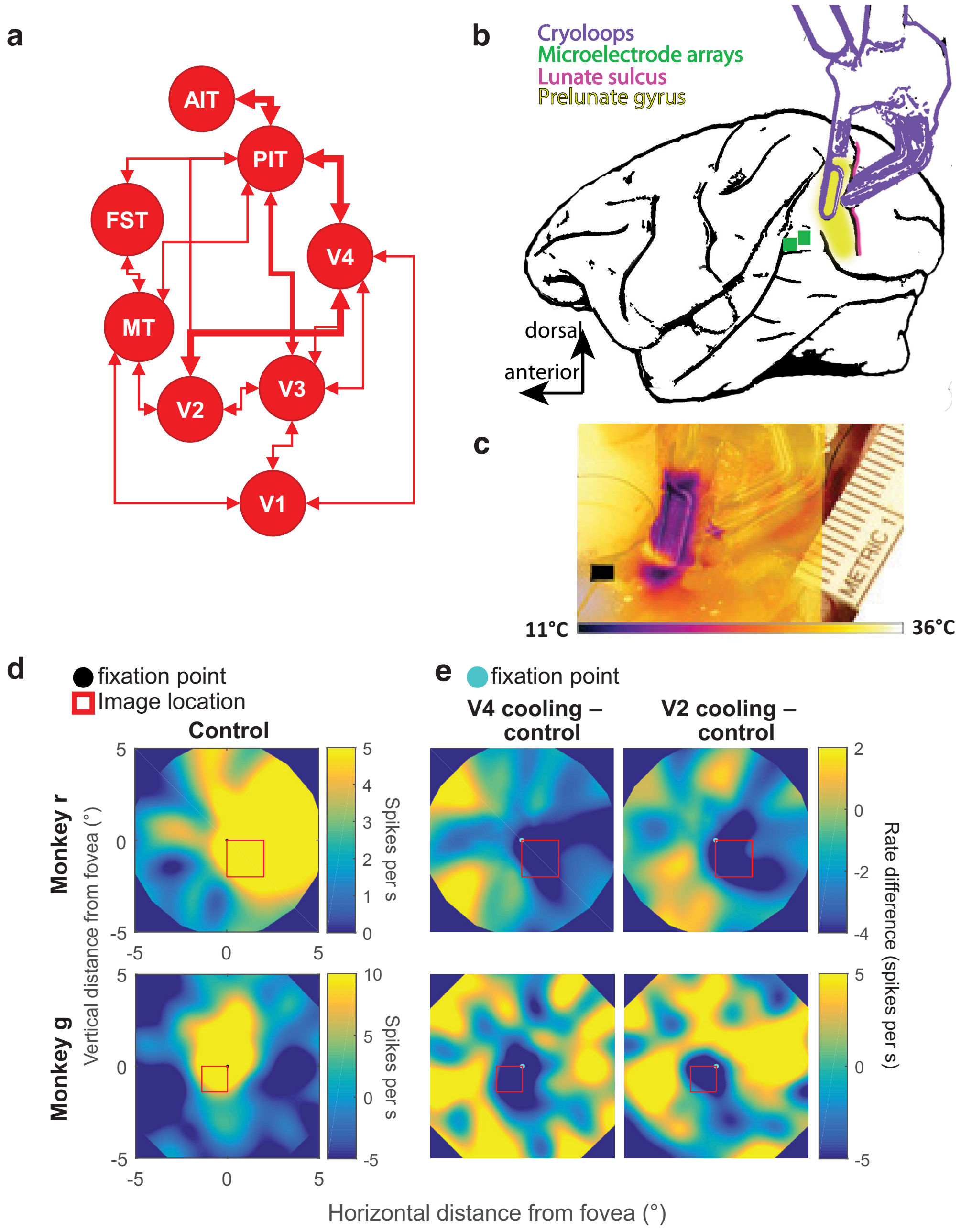

Figure 1. Cooling affected portions of the aggregate PIT receptive fields. ***a***, Partial input network to PIT. ***b***, Schematic showing location of cryoloops (purple) and microelectrode arrays (green) relative to the prelunate gyrus (yellow) and lunate sulcus (pink). Monkey R had two short loops inside the lunate sulcus. Monkey G had one long loop. ***c***, Composite image showing the superimposed thermal and visible light images taken intraoperatively while the prelunate gyrus loop is active. Black square represents the approximate location of the closest array. ***d***, Average firing rate for all units, evoked by flashing stimuli within a 8° × 8° grid, in the control (warm) condition. Red boxes represent the regions of stimulus presentation in subsequent experiments. Central dot indicates the center of the screen. ***e***, Difference in activity during cooling of area V4 (left column) or V2|3 (right column).

ANOVA (comparing the mean firing rates before cooling, during V2|3, and during V4 cooling, $N = 300$ values per temperature condition; mean F value distribution from shuffled-label test was 0.98 ± 1.0 [±SE], 999 iterations, whereas the F value from the experimental distributions was 282.94). We repeated the analysis using only channels showing statistically reliable visual activity, and this showed similar results (see Materials and Methods; Table 1, bottom half): during V4 cooling, mean response amplitude was reduced by 33% and 34% (Monkeys R and G); during V2|3 cooling, mean response amplitude was reduced by 26% and 35%.

In one animal, we cooled both V4 and V2|3 loops concurrently, measuring a similar reduction in firing rate (38%); cooling both sets of loops did not silence PIT. Another measure of input strength is response latency, and here we similarly observed little difference between V2/V3 cooling and V4 cooling. We considered two metrics for latency for each site: (1) response latency, defined as the earliest time after stimulus onset when activity rose 2 SDs above baseline; and (2) tuning latency, defined as the time after stimulus onset when the tuning curve variance was highest. PIT multiunits showed a −3 ms difference in response latency between V4 and V2|3 cooling in Monkey R ($\text{latency}_{V4} - \text{latency}_{V2|3} = 55–58$ ms) and a 2 ms difference in Monkey G (60–58 ms); 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 (Table 2).

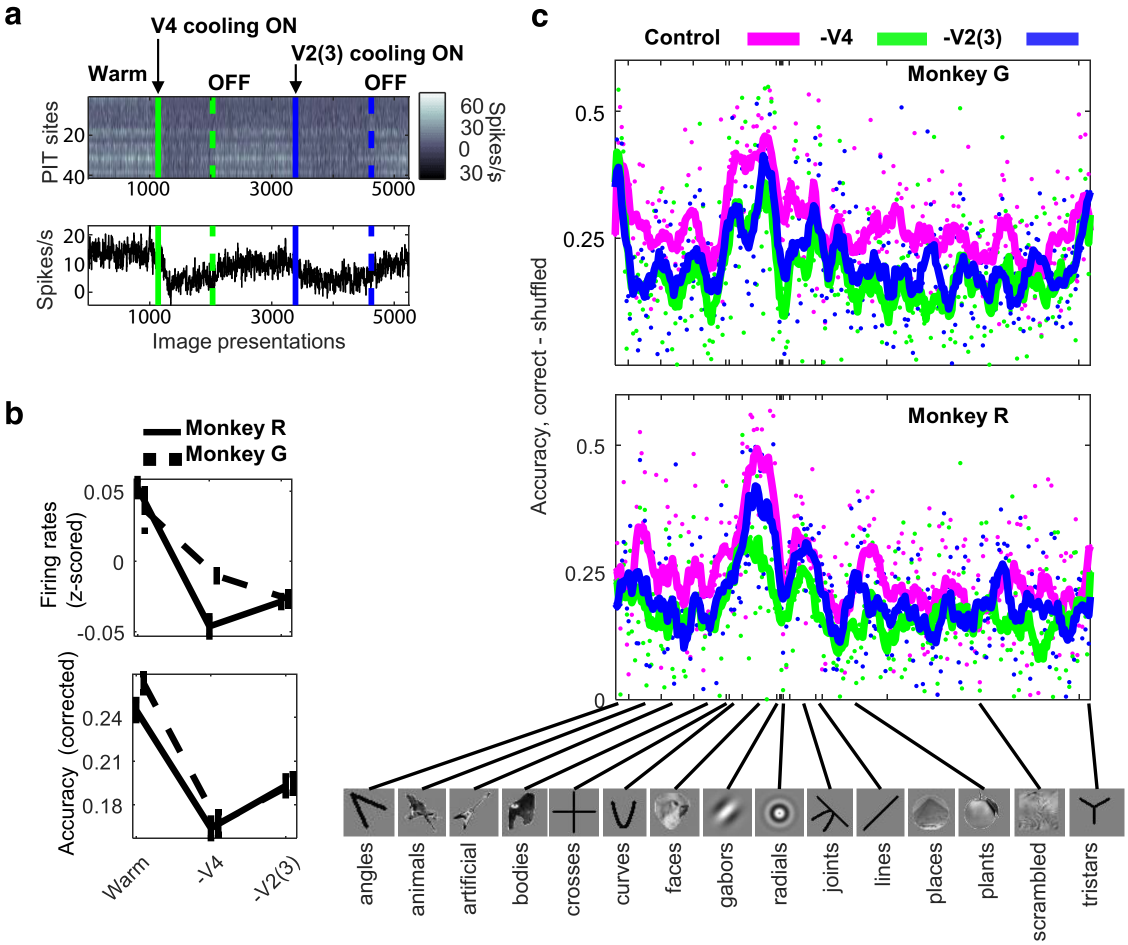

Figure 2. Effects of cooling on firing rate and classification accuracy. ***a***, Top, Data from one cooling session (Monkey R, day 1). The evoked spike rates from 41 visually responsive PIT sites (rows) recorded concurrently before, during, and after cooling of V4 and V2|3. Each column represents one image presentation. Solid lines indicate the onset of each cooling condition. Broken lines indicate the onset of the rewarming periods. Bottom, Mean firing rate across each temperature condition. ***b***, Top, Average firing rate activity (*z* scored) for all channels during each temperature condition. Bottom, Median classification accuracy for all images during each temperature condition. ***c***, Mean accuracy for each image before and during cooling (baseline-corrected; magenta represents control, green represents V4 cooling, blue represents V2|3 cooling). The *x*-axis indicates all 293 images listed within their category.

Table 1. Firing rate changes during cooling of areas V2, V3, and/or V4[a]

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

[a] Baseline period = 0–50 ms after image onset; evoked period = 51–151 ms.

In summary, PIT multiunits lost approximately one-third of visually driven activity during deactivation of either subset of their inputs, with no reliable difference between cooling V2-V3 or V4 across monkeys. PIT multiunits also showed increased response latency during input deactivation but showed no consistent differences between the effects of V2/V3 or V4 cooling across monkeys either.

Cooling reduced decoding accuracy by linear classifiers

Next, we used pattern analysis to quantify the encoding capacity of PIT during V4 or V2|3 cooling. We trained statistical classifiers (SVMs with a linear kernel) using data from each experimental condition (before cooling, during V4 or V2|3 cooling). SVMs were used in a one-versus-one approach: for a given image *A*, we trained SVMs to perform a simple classification against a second image *B*, then against a third image *C*, until all other images had been compared. Then we took image *B* and repeated the process. We had few trials per image (4–6 repetitions per temperature condition); thus, we used leave-one-out cross-validation for each paired comparison. To control for statistical bias in chance performance due to small sample number, we also trained SVMs using the same data but shuffling the image labels. Thus, accuracy was defined as the mean of all cross-validation cycles

Table 2. Latency values[a]

	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

[a] Response and tuning latency values measure before cooling, during V4 cooling, and during V2|3 cooling.

using the correct labels minus the mean of cross-validation cycles using the shuffled labels, so a baseline-subtracted accuracy score of 0.5 should be close to perfect accuracy.

The decoding analysis showed that faces elicited the highest classification accuracy in both animals, which was noteworthy because we did not preselect the array implantation sites by functional response (i.e., proximity to fMRI face patches; of note, SVMs did not show that this subset of faces was more easily classified at the pixel level relative to other categories). Before cooling, SVMs showed a median accuracy value of 0.24 ± 0.01 and 0.27 ± 0.01 above baseline (Monkeys R and G, SE of the median). During V4 deactivation, median accuracy dropped to 0.16 ± 0.01 in both animals; during V2|3 deactivation, median accuracy dropped to 0.19 ± 0.01 and 0.20 ± 0.01. These median accuracy values were statistically different at the group level ($p < 10^{-11}$, one-way Kruskal–Wallis ANOVA comparing median accuracy values obtained during control, V4 and V2|3 cooling, one test applied to each monkey). The differences in median values between V4 and V2|3 cooling were also statistically reliable in both animals ($p < 2 \times 10^{-4}$, Wilcoxon sign rank test, comparing median accuracy values during V4 and V2|3 cooling, one test applied per monkey; Fig. 2*b*, bottom, *c*). We also estimated how many individual images led to classification values that were not statistically likely to occur ($p < 0.05$) given the shuffled accuracy values (i.e., null probability values), using binomial tests (see Materials and Methods). During the control condition, classifiers led to statistically reliable classification scores for 280 of 293 images (96%, Monkey R) and 293 of 293 (100%, Monkey G). During V4 cooling, the number of images with reliable scores were 246 of 293 (84%) and 248 of 293 (85%), and during V2|3 cooling, 273 of 293 (94%) and 279 of 293 (95%). In summary, while populations of PIT multiunits showed inconsistent overall mean firing rate reductions during V2|3 or V4 deactivation (see previous section), in both monkeys, SVM-based encoding accuracy was reduced more by V4 deactivation. We repeated the previous section's firing rate comparisons using their *z* scored transformations because SVMs were trained and tested using *z* scores and we wanted to make sure that this transformation did not cause any discrepancies. We found that the *z* scored firing rates had the same pattern as the raw firing rates (reductions in mean *z* score: warm −V4: Monkey R, 0.13, Monkey G, 0.15, warm −V2|3: Monkey R, 0.11; Monkey G, 0.16, probability that all median *z* score values arose from the same distribution $p < 10^{-3}$, Kruskal–Wallis ANOVA, temperature conditions as groups). Thus, we conclude that V4-based inputs are more important for image identification and categorization compared with the V2|3 inputs; further, this difference is best detected using pattern analysis.

Categorization of images in PIT activity space during input cooling

To estimate the image-by-image similarity of our visual images in neural activity space before and during cooling, we computed the pairwise distance of the PIT population response vectors at each temperature condition. This approach has previously revealed intriguing structure in AIT activity space, with clusters of image-response vectors belonging to different categories (e.g., faces or body parts) occupying different parts of activity space (Kiani et al., 2007; Baldassi et al., 2013). We wanted to define the structure of PIT activity space given our image set, and then to determine how input deactivation changed this structure. First, we visualized this neural activity space by projecting the multidimensional population vectors (warm condition) into two dimensions via nonmetric multidimensional scaling. We found that this mapping did not reveal segregated clusters as previously observed in AIT, but instead showed a spectrum with faces and line drawings at maximum separation (Fig. 3*a*). During cooling of either V4 and V2|3, the perimeter of this spectrum shrank, but the interventions did not change its gross organization (line shapes remained in one side, faces in the other). To quantify this observation, we asked how individual images changed position during cooling of V4 or V2|3 using K-means (in the original multidimensional space). In this K-means analysis, we first chose a number of multiple centroids to divide the spectrum into territories during control conditions (leaving the algorithm to locate the most efficient locations of these centroids; Fig. 3*b*). Then we kept track of images as they migrated from territory to territory during cooling (Fig. 3*c*) in the following manner: for each pass, we ran K-means on the control data first and saved the centroids. We then ran K-means on the cooling data, inputting the warm-condition centroids, and counted how many images changed labels during cooling. Because K-means is stochastic, we ran each analysis 100 times, postulating 2–15 different centroids per pass (thus we ran 100 iterations per centroid number, or 1400 total passes). We found that the fraction of individual images that changed territory varied with the number of centroids: not surprisingly, the more centroids used, the smaller the territory claimed by that centroid, and thus the more likely that an image would change label. For example, considering Monkey R, when its PIT activity space was divided into four clusters, 31 ± 4% of images changed membership during V4 cooling and 24 ± 6% during V2|3 cooling (mean ± SEM). With eight clusters, 60 ± 6% of images changed cluster membership during V4 cooling and 52 ± 5% during V2|3 cooling. In both monkeys, we found that V4 cooling consistently induced more images to change territories compared with V2|3 cooling (Fig. 3*d*). To determine whether the percentage difference between the −V4 and −V2|3 conditions was likely to arise from the same underlying distribution, we used a randomization test; for each of the 100 K-means passes, we shuffled all −V4 and −V2|3 cooling response vectors twice, ran the K-means analysis above, and subtracted the percentage value of one shuffled vector distribution from the percentage value of the second shuffled vector distribution. We considered the difference between the observed −V4 minus −V2|3 percentages to be statistically reliable if this experimentally observed difference was >95% than the values from the shuffled distribution. We found that the great majority of V4 percentage values were statistically larger than the −V2|3 values in both animals, confirming our observation that V4 cooling was more likely than V2|3 cooling to change the positions of individual images in activity space, and thus their potential category membership.

To probe the efficiency of category encoding by PIT neurons during V4 and V2|3 deactivation, we also trained SVMs to classify between pairs of categories. Each category was defined a priori and comprised between 2 and 118 unique images of faces and other complex objects, line shapes, or scrambled textures (me-

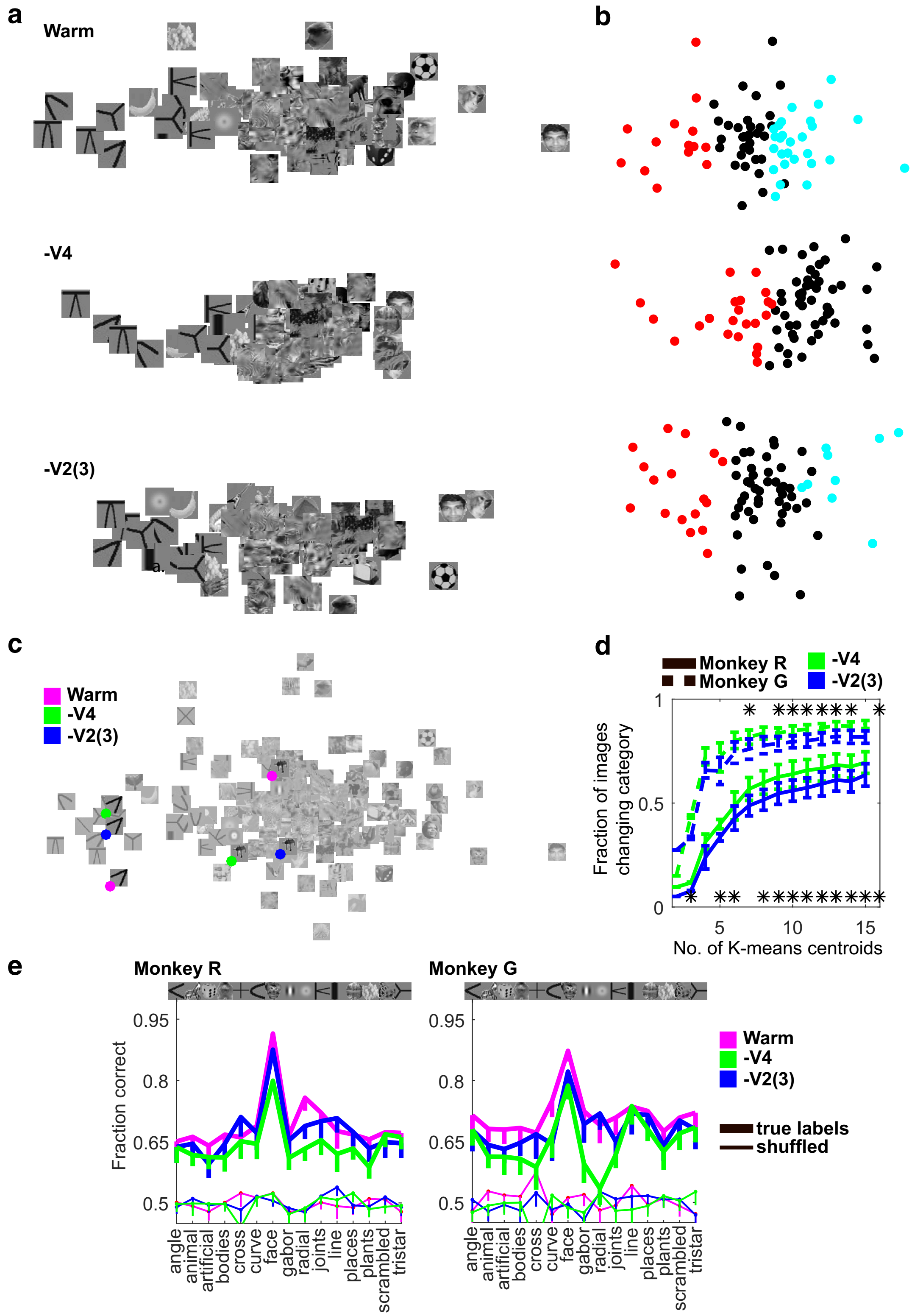

Figure 3. Effects of cooling on categorization. ***a***, Two-dimensional plot of activity space for Monkey G, pseudopopulation 2. Top, Control (warm) data. Middle, V4 cooling. Bottom, V2|3 cooling. For clarity, only a subset of images is shown. ***b***, K-means clustering of points in ***a***, using three centroids. Top, Control. Middle, −V4. Bottom, −V2|3. ***c***, Tracking the location of two images. Faded images represent 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 represents V4 cooling; blue represents V2|3 cooling; solid lines indicate Monkey R; dashed lines indicate Monkey G), as a function of the number of centroids. Error bars indicate mean ± SE after 100 K-means repetitions. Asterisks indicate if the difference in fractions between −V4 and −V2|3 conditions have <0.05 probability of arising from the same distribution (randomization test; top row of asterisks, Monkey G; bottom row, Monkey R). ***e***, Category classification accuracy before cooling (magenta), during V4 cooling (green), and V2|3 cooling (blue). Each point indicates the mean accuracy (±SEM). Thick lines indicate true-label classification scores. Thin lines indicate shuffled-label scores. Images were generated in the laboratory, and a small subset was obtained from Google Images under the filter "Labeled for reuse with modification."

dian = 16; see Materials and Methods). SVMs were trained and tested using five-fold cross-validation, and chance performance was defined using label shuffling. We found that, before cooling, SVMs showed a median (baseline-corrected) accuracy of 0.18 ± 0.02 and 0.20 ± 0.01 (±SE, Monkeys R and G); during V4 cooling, 0.15 ± 0.01 and 0.14 ± 0.02, and during V2|3 cooling, 0.16 ± 0.01 and 0.18 ± 0.01; Fig. 3*e*). We tested the hypothesis that V4 deactivation led to a deeper reduction in category classification accuracy compared with V2|3 cooling, and found that this reduction was statistically reliable in both animals ($p = 1.9 \times 10^{-3}$ and 2.7×10^{-2}, Wilcoxon signed rank test, $N = 30$ scores per temperature condition). We conclude that PIT neurons are less efficient at categorization during V4 cooling compared with V2|3 cooling.

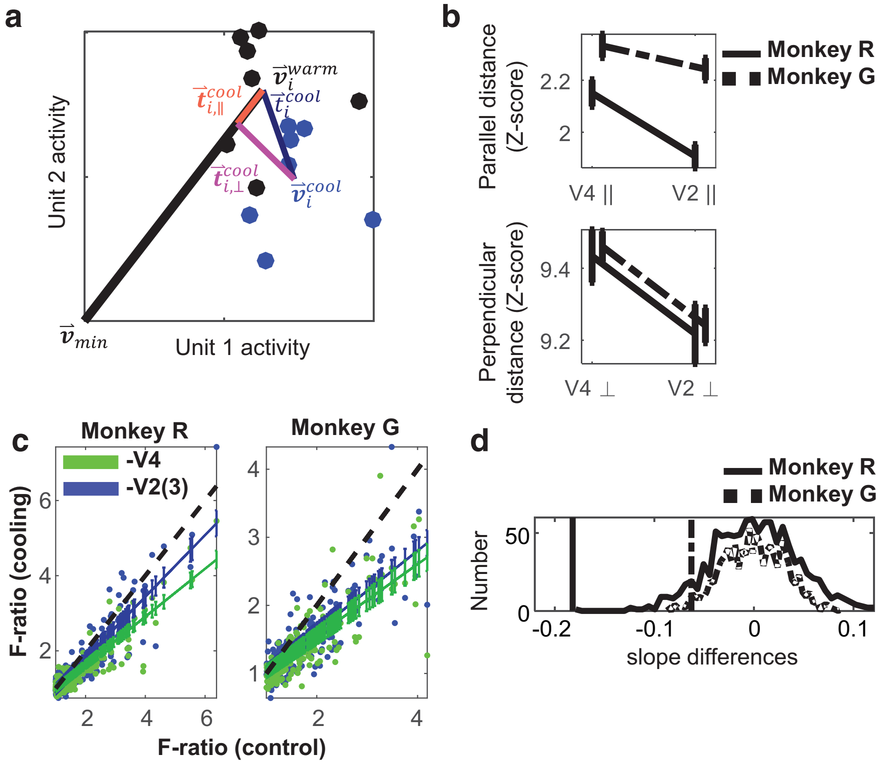

Figure 4. ***a***, Projection analysis. Schematic. The axes represent the hypothetical responses of two units in response to multiple presentations of a single image *i* before cooling (black dots) and during cooling (blue dots). The mean responses to the image are noted by $\vec{v}_i^{warm}$ (control) and $\vec{v}_i^{cool}$ (cooling). Thick black line indicates the axis between the mean control response and the minimum response vector $\vec{v}_{min}$. Blue line indicates the trajectory vector $\vec{t}_i^{cool}$ between $\vec{v}_i^{warm}$ and the mean cooling response $\vec{v}_i^{cool}$. Orange and purple lines indicate the parallel ($\vec{t}_{i,\|}^{cool}$) and perpendicular ($\vec{t}_{i,\perp}^{cool}$) components, respectively, of the cooling trajectory vector. ***b***, Mean parallel and perpendicular components of the cooling trajectory vectors for each deactivation condition and monkey. Solid lines indicate Monkey R, dashed lines, Monkey G. ***c***, Scatterplots of *F* statistics (green represents warm vs −V4; blue represents warm vs −V2|3). Each point indicates the paired *F* ratios for a given site, measured before and during cooling. Solid colored lines indicate the mean slope describing the control and cooling *F* ratio distributions. Error bars around each slope indicate SE. Dashed black line indicates unity. ***d***, Differences in slopes expected given a mixed temperature distribution. Solid curve indicates distribution from Monkey R data. Broken line indicates distribution from Monkey G data. Vertical lines indicate the experimental difference.

Cooling effects as trajectories in neural state space

Cooling V4 reduced decoding accuracy in PIT more than cooling V2|3 even though overall firing rate reductions were not consistently lower during V4 versus V2/3 cooling across monkeys. We therefore explored neural state trajectories in multivariate activity space to clarify 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 pseudopopulation; see Materials and Methods). We can visualize the neuronal representation of an image as a coordinate point in this multidimensional coordinate space (DiCarlo and Cox, 2007; Rust and DiCarlo, 2010). Different mean vectors in the space represent different images. Excitatory drive was decreased during input cooling, which should reduce the length of all vectors, moving them toward a minimum response vector $\vec{v}_{min}$. This $\vec{v}_{min}$ would be the zero vector if each unit's output was defined as raw spike counts, but because we are using *z* scored, baseline-subtracted spike rates, the actual $\vec{v}_{min}$ must be defined using each unit's lowest *z* score value. The difference in the control versus cooling coordinates for a given image *i* describes a cooling trajectory $\vec{t}_i^{cooling} = \vec{v}_i^{warm} + \vec{v}_i^{cooling}$. If the main effect of cooling was to randomly remove spikes from the population, this cooling trajectory should be parallel to the minimum response vector $\vec{v}_{min}$. 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 nonparallel) to $\vec{v}_{min}$. We therefore measured the parallel and perpendicular components of each image's V4 and V2|3 cooling trajectory vectors as $\vec{t}_i^{-V4} = \vec{t}_{i,\|}^{-V4} + \vec{t}_{i,\perp}^{-V4}$ and $\vec{t}_i^{-V2/3} = \vec{t}_{i,\|}^{-V2/3} + \vec{t}_{i,\perp}^{-V2/3}$ for each image *i* (Fig. 4*a*). We found that the perpendicular component of trajectory vector was usually larger than the parallel component. The mean parallel component of the cooling vector behaved like the population firing rate changes: for Monkey R, the mean parallel component was larger during V4 deactivation compared with V2|3 deactivation; for Monkey G, the parallel components of both deactivations were approximately the same (Monkey R, norm of parallel component: −V4: 2.2 ± 0.05, −V2|3: 1.9 ± 0.05 | $p = 4.9 \times 10^{-8}$ per Wilcoxon sign rank test; Monkey G, −V4: 2.3 ± 0.05, −V2|3: 2.2 ± 0.04, $p = 0.01$; Fig. 4*b*). In contrast, for both monkeys, the mean perpendicular component of the V4 deactivation was consistently 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, $p = 3.9 \times 10^{-8}$ per Wilcoxon sign rank; Monkey G, −V4: 9.5 ± 0.03, −V2|3: 9.2 ± 0.05, $p = 1.5 \times 10^{-14}$). This showed that V4 deactivation redirected image representations to further locations in the activity space than V2|3 deactivation. Like multivariate linear classifiers, this projection technique was more reliable in highlighting differences between V4 and V2|3 cooling than simply comparing mean population responses across individuals.

Cooling reduced selectivity of individual PIT sites

Cooling inputs to PIT reduced classification accuracy at the population level. To examine accuracy at the level of individual sites, we measured selectivity using an *F* test. Let us say that PIT sites were selective to specific images if the mean variance of their spike counts to different images was greater than the mean variance of their spike counts to each image; this can be estimated using the F statistic. The F statistic is the ratio of between-group variance divided by within-group variance, where the value 1 suggests no selectivity; the greater the value, the more selectivity. We called the F test statistic per channel before cooling $F_{control}$, during V4 cooling $F\text{-}_{V4}$ and during V2|3 cooling, $F\text{-}_{V2|3}$. If the distributions of $F\text{-}_{V4}$ and $F\text{-}_{V2|3}$ are closer to the nonselectivity value of 1 compared with the $F_{control}$ distribution, this would suggest that PIT sites become less selective during cooling.

Each PIT site led to one $F_{control}$, one $F\text{-}_{V4}$, and one $F\text{-}_{V2|3}$ value. We plotted each control F value against its counterparts and found that cooling F statistics were lower than the warm distribution (Fig. 4*c*; Monkey R, median F ± SE, Warm: 1.13 ± 0.03, −V4: 1.10 ± 0.01, −V2|3: 1.14 ± 0.02; Monkey G, Warm: 1.21 ± 0.05, −V4: 1.13 ± 0.02, −V2|3: 1.14 ± 0.02), although there was no statistical difference between the medians of the warm, −V4 and −V2|3 temperature groups ($p < 0.20$ for both monkeys, one-way Kruskal–Wallis test, comparing all temperature conditions). However, many of the PIT sites were not that selective to start with, having precooling F statistics already bottomed out at 1. We noticed that units with higher precooling F values showed greater changes during cooling (Fig. 4*c*). To quantify this observation, we asked whether the slope describing the relationship between the precooling and cooling F values was statistically different from unity. We used a bootstrap approach. For 1000 iterations, we resampled sites with replacement and used their $F_{control}$, $F\text{-}_{V4}$, and $F\text{-}_{V2|3}$ values to fit linear regression lines between control and −V4 values, and then between control and −V2|3 values. This analysis resulted in 1000 slopes describing the $F_{control}$ and $F\text{-}_{V4}$ relationship, and another 1000 slopes describing the $F_{control}$ and $F\text{-}_{V2|3}$ relationship. None of these slope values overlapped the line of unity (Monkey R, mean slope ± SEM, control vs −V4: 0.64 ± 0.03, control vs −V2|3: 0.82 ± 0.04; Monkey G, control vs −V4: 0.53 ± 0.04, control vs −V2|3: 0.59 ± 0.03). We also noticed that the mean $F_{control}/F_{cooling}$ slope was shallower during V4 cooling than V2|3 cooling in both animals. This suggested that PIT multiunits become less selective during V4 cooling than during V2|3 cooling. To determine whether the slope between the warm −V4 conditions was lower than that of the warm −V2|3 condition, we used a randomization test where we shuffled the V4 and V2|3 F values. The null hypothesis is that the mean V4 and V2|3 slopes came from the same distribution, so we created this null distribution as follows: in each of 999 passes, we randomly mixed the labels between the V4 and V2|3 F statistics for each channel and computed an $F_{cooling}/F_{control}$ slope. We did that twice per pass, and then subtracted the two slopes. After all passes, we had 999 slope differences (not including the experimentally observed difference) that we then compared with the experimental slope difference. We found that these null difference distributions were defined by 5th and 95th percentile values of −0.07 to 0.07 (Monkey R) and −0.05 to 0.05 (Monkey G). The observed differences in mean cooling slopes were −0.18 and −0.06 (Monkeys R and G). The probability that the experimental differences in V4 and V2|3 slopes came from such mixed distributions was 0.001 and 0.02, respectively.

We further asked whether there was any relationship between the retinotopic location of a PIT receptive field relative to the cooling scotomas, and its subsequent change in selectivity (F statistic). The images were presented at the intersection of the population-wide V4 and V2|3 scotomas. Therefore, some individual multiunit PIT receptive fields (RFs) would by chance "see" more of the stimulus than other PIT RFs. For each PIT site, we measured the fraction of its RF that overlapped the stimulus location/scotoma, and correlated this value against the subsequent change in selectivity (change in F statistic). The RF overlap measure was computed using data from different recording days (see first section of Results). For each site its RF overlap was defined as the average number of spikes emitted in response to stimuli presented in the stimulus/scotoma overlap region, divided by the total number of spikes emitted in the central 8 × 8°. The mean RF overlap value was 0.12 ± 0.01 and 0.15 ± 0.01 (Monkeys R and G). There was a small but statistically reliable correlation of RF overlap with selectivity change (selectivity change was defined as $F_{control} - F_{cooling}$): during V4 cooling, the Pearson correlation coefficient was 0.19 and 0.32 ($p = 1.2 \times 10^{-3}$ and 1.3×10^{-7}, Monkeys R and G). During V2|3 cooling, the correlation coefficient was 0.11 and 0.26 ($p = 0.06$ and 2.3×10^{-5}). The stimuli were placed in the same overlapping region between both −V4 and −V2|3 scotomas, so the lower correlation values for V2|3 were not due to differences in scotoma overlap; rather, it was because the selectivity change is less pronounced for V2|3 cooling (if there was no selectivity change, the correlation would be zero). We conclude that PIT multiunits lost selectivity across images as a function of RF location.

In summary, individual PIT multiunits became less selective during V4 and V2|3 cooling, as determined by a variance test. This loss of tuning was more pronounced during V4 cooling than V2|3 cooling and was a function of distance from the scotoma. Both the population decoding accuracy change and the projection analysis results similarly indicate that V4-based inputs are more important for overall image coding than V2|3-based inputs.

Loss of decoding accuracy simulated by perturbing control vectors

We wanted to explore how losses in decoding accuracy could result from random reductions in the magnitude and direction of population firing rate vectors. This would illustrate the range of potential decoding accuracy losses that could be incurred by (1) reducing each unit's excitatory drive without changing their tuning, (2) changing their tuning without reducing response magnitude, and (3) both mechanisms acting at once. To do this simulation, we transformed the experimentally measured warm-condition population rate vectors into "cooling" vectors by multiplying their norms by decreasing fractions (simulating lower firing rates) and/or by adding increasingly larger angular rotations in activity space (to simulate representational changes). After these transformations, these simulation-cooling response vectors were put through the same linear decoding algorithms as above (SVMs), creating a decoding accuracy map as a function of gain and angular change. To simulate changes in excitatory drive, for each simulation, we created nine normal distributions of gain changes with means of 0.1 to 1 (in steps of 0.1) and a SD of 0.025, then multiplied each warm population response vector by randomly sampled gain values from a given normal distribution. Within that same simulation, we simulated representational changes by rotating each vector in along different planes through angles between 3° and 56° in the original multidimensional activity space. Because most multidimensional neuronal activity really stays within lower-dimensional hyperplanes, we wanted to make sure that our artificial rotations stayed close to those hyperplanes. Thus, we rotated each vector along directions observed during normal fluctuations within the warm condition: first, we computed the vector trajectories shown by individual population vectors relative to the grand mean (Fig. 5*ai*); we then computed the parallel components of those trajectories to the minimum re-

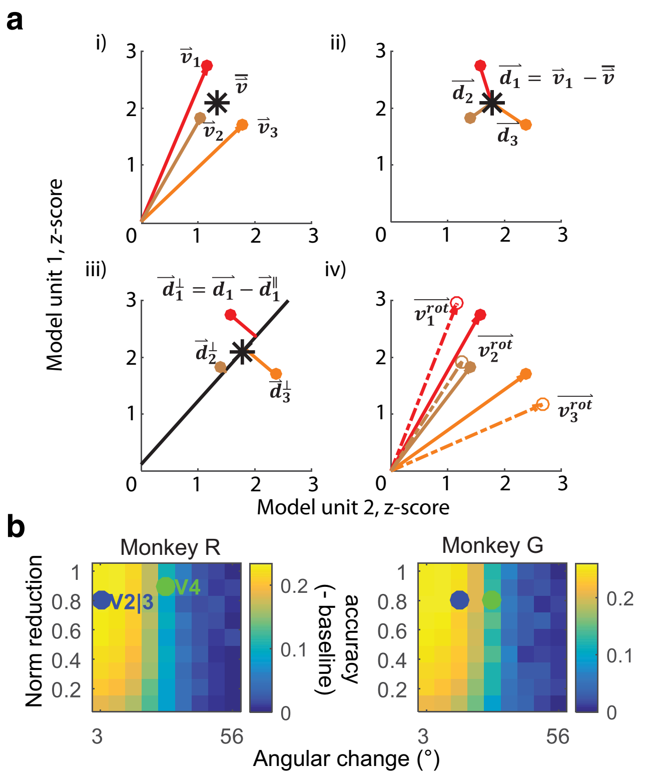

Figure 5. Cooling simulations. ***a***, Vector rotations in N-dimensional space. ***ai***, Hypothetical responses of two units (axes) to three different image presentations ($\vec{v}_i$, where $i = 1–3$) during control conditions; $\bar{\vec{v}}$ = mean response vector (asterisk). ***aii***, Difference vectors $\vec{d}_i$ between mean response vector $\bar{\vec{v}}$ and each individual image representation $\vec{v}_i$. ***aiii***, Perpendicular components $\vec{d}_i^{\perp}$ of difference vectors $\vec{d}_i$. ***aiv***, Rotated vectors $\overrightarrow{v_i^{rotated}}$ obtained after adding $\vec{d}_i^{\perp}$ multiplied by scalar factors to $\vec{v}_i$ and dividing by a factor to keep $\overrightarrow{v_i^{rotated}} = \vec{v}_i$. In the actual simulations, modified $\vec{d}_i^{\perp}$ were shuffled before adding to $\vec{v}_i$. ***b***, Decoding accuracy gradient. Mean accuracy values (baseline-subtracted, ±SE) computed after transforming warm-condition response vectors by various norm (gain) and angular changes. Green and blue circles represent the experimentally measured accuracy values during V4 and V2|3 deactivation.

sponse vector, which is in the direction of simple gain changes. We used the perpendicular components of trajectories as directions for rotation (Fig. 5*aii*,*aiii*), multiplying each perpendicular components by scalar values to control the magnitude of rotation. These modified trajectory vectors were randomly added to the warm-condition population vectors, and matched in vector norm, resulting in "cooling-condition" vectors that stayed closely to the hyperplanes. After performing 81 simulations (9 mean gain values × 9 rotation values), we found that this approach could successfully reduce SVM decoding accuracy from warm-condition values (0.24 ± 0.01 and 0.27 ± 0.01, Monkeys R and G) down to shuffled-label baseline (difference in decoding accuracy relative to shuffled baseline of −0.0015). We noticed that decoding accuracy was resilient to changes in gain: reducing the norm of all spike rate vectors to values as low as 10% of the original norm only lowered decoding accuracy from 23% and 26% to 18% and 21% (Monkeys R and G). This is in contrast to angular changes, which lowered accuracy down to 0% with 56° rotations. For this reason, when we mapped the experimental-cooling decoding accuracy values onto this simulation-cooling map, we found that V4 cooling values consistently required larger rotation values than V2|3 values but were not drastically different along the gain gradient: to achieve V4 cooling decoding accuracy, the model gain values required were 0.9 and 0.8 (Monkeys R and G) and the angular change was 25.6° (Monkeys R and G). In contrast, achieving the observed V2|3 cooling decoding accuracy required 0.8 gain changes (Monkeys R and G) and only 0.4° and 12.8° of angular changes (Monkeys R and G). Thus, this simulation suggested that affecting decoding accuracy in PIT was primarily a function of perturbing tuning representational mechanisms, not simple reductions in excitatory drive, and that V4 deactivation induced more representational changes than V2|3 deactivation.

Cooling did not reveal shape-specific deficits

PIT input deactivation led to a reduction in image classification accuracy both at the population level and also at the individual multiunit level. We wanted to determine whether there were any features of the images that predicted the consequent loss in decoding accuracy during cooling, and we used a series of regression analyses to explore this issue. We came up with a comprehensive list of 87 different quantitative and categorical descriptions for each of our 293 images, such that each image was described by values corresponding to their luminance, contrast, horizontal and vertical orientation content, curvature, and categorical membership (e.g., "faces," "body parts," "tristars"). To include implicit features within our image set, so we also applied principal component analysis to the image set and to their discrete Fourier transforms, deriving 50 spatial principal components derived from the raw images (amounting to 90% of image variance) and 30 principal components from the images' Fourier transforms (amounting to 80% of discrete Fourier transform variance). In addition to those 85 descriptors, we added two additional predictors: the mean population rate evoked per image (before cooling), and the mean decoding accuracy achieved per image (also before cooling). These 87 descriptors were used as regression variables in a general linear model where the dependent variable was the change in decoding accuracy per image: we fit three linear models of the form $\boldsymbol{\Delta accuracy_i^{condA-condB}} = \boldsymbol{w_0} * x_i^{lum} + \boldsymbol{w_1} * x_i^{contrast} + \ldots + \boldsymbol{w_6} * x_i^{prin.comp.1} + \ldots + \boldsymbol{w_{86}} * x_i^{firing\ rate,warm} + \boldsymbol{w_{87}} * x_i^{accuracy,warm}$, where i = index of image ranging from 1 to 293, and the dependent variable $\boldsymbol{\Delta accuracy_i^{condA-condB}}$ could represent $\boldsymbol{\Delta accuracy_i^{warm-V4}} = accuracy_i^{warm} - accuracy_i^{-V4}$, $\boldsymbol{\Delta accuracy_i^{warm-V2|3}} = accuracy_i^{warm} - accuracy_i^{-V2|3}$ or $\boldsymbol{\Delta accuracy_i^{(warm-V4)-(warm-V2|3)}} = \boldsymbol{\Delta accuracy_i^{warm-V4} - \Delta accuracy_i^{warm-V2|3}}$. We found that the only consistent predictor of V4- or V2|3-cooling accuracy loss was the magnitude of classification accuracy before deactivation: the larger the classification accuracy for each image before deactivation, the larger the subsequent reduction in accuracy. We identified this predictor using two methods: (1) in the main linear regression analysis of each monkey, we saw that this variable had the highest t statistic (8.6–8.7); and (2) we also used regularized linear regression (see Materials and Methods). To ensure that this was not simple regression to the mean, we also divided our control trials such that the control classification tuning curve used for regression was not the same as the control classification tuning curve used to calculate the cooling difference in classification accuracy (the estimated correlation between these cross-validated datasets were 0.33–0.45 for Monkeys R and G, $p < 10^{-8}$, Student's t test). During V4 cooling, the percentage of variation explained by each model was 45%–55% ($R^2 = 0.45$ and 0.55, Monkeys R and G, $p = 4 \times 10^{-6}$ and $p = 1.74 \times 10^{-7}$); during V2 cooling, the R^2 values were 0.48 and 0.56 ($p = 1 \times 10^{-7}$ and 3×10^{-8}). In contrast, when trying to account for differences in decoding accuracy between V4 and V2|3 cooling ($\boldsymbol{\Delta accuracy_i^{(warm-V4)-(warm-V2|3)}}$), the model was a poor fit, as no linear combination of the image features above could account for >29% of the variation ($p = 0.91$).

In summary, we found that before-cooling decoding accuracy was the strongest predictor of the loss of decoding accuracy. None of our 87 features showed a statistical dependence with the difference in −V4 and −V2|3 accuracy loss, suggesting that a yet-undiscovered image property is differentially represented among the pathways, or that image feature encoding does not differ between them. This latter interpretation is consistent with the usual implementations of the standard model of visual recognition, which do not handcraft any special shape selectivity roles among bypass pathways (Serre et al., 2005). To explore the theoretical implications of encoding impairment in PIT during input deactivation, we then turned to this standard model of visual recognition.

Cooling parallels in the standard model of visual recognition

The most decisive way to define the advantages of concurrent, parallel pathways would be to trace the input history to every PIT neuron from V1 through V2, V3, or V4. This is not yet technically feasible. Thus, we pursued this option in a computational model, specifically the standard model of visual recognition (Serre et al., 2007b). Our goal was to isolate subpopulations of simulated "PIT" units in a network with concurrent pathways and to compare the subpopulations' ability to encode for individual objects. We hoped that, by comparing model PIT units that skipped input from different layers (the analogs of V2 or V4), we would complement our experimental observations during cooling of different input regions to PIT. We designed the model so that its top layer would have a similar number of units as our microelectrode arrays (58 units). As a brief summary of our results, when tested with our experimental image set, the models showed that parallel pathways to simulated "PIT" units delivered comparably useful information for image classification, and that eliminating up to 45% (32 of 58) of active units in "PIT" resulted in 5%–6% reductions in decoding accuracy, compared with 5%–8% reductions in the PIT data. We conclude that parallel pathways transmit equally useful information for object decoding both in convolutional network models and in the brain.

The standard model is a hierarchical, feedforward-only model inspired by the visual system (Hubel and Wiesel, 1962; Fukushima, 1975). This model comprised multiple layers (areas), each with many filters (RFs) of different sizes. Each layer performed three serial operations: a convolutional tuning operation, a pooling (invariance) operation, and normalization of output responses. At the highest layer of the model, there emerged a sparse population of units, whose activations encoded an abstract representation of the original pixel-space image. This vector was used in a final classification step (via SVM) to measure the accuracy of representation of the original image versus every other image. As in the published 2007 model, our first layer consisted of gabors and subsequent layer filter weights were trained on images from a separate custom set (Serre et al., 2007b). Our version of this model included three alternative pathways that could provide input to each PIT unit: one long four-layer pathway and two short three-layer pathways. The long pathway represented V1 → V2 → V4 → PIT; the first bypass pathway skipped V2 (V1 → V4 → PIT), and a second bypass pathway skipped V4 (V1 → V2|3 → PIT; Fig. 6*a*). We asked how PIT units at the endpoint of each parallel pathway differed in their representation of the same visual image. Each layer had 200 filters of different sizes. In the first layer, filter sizes were 0.1°–0.4° in width (in our experimental setup, 1° = 26 pixels) and doubled at each layer, but for the bypass pathways, RF size quadrupled at the bypass layer. This ensured that V4 filters were the same size, regardless of their inputs arising from V1 or V2. In the electrophysiology experiments, we presented each image multiple times and obtained a distribution of correlated but nonidentical response vectors. In contrast, the model is not stochastic; and so to induce variability across presentations of the same image, we created six variations of each of the 293 images by simulating fixational eye movements (see Materials and Methods). We processed all 293 × 6 images through the model and used SVMs to measure the classification accuracy in the model PIT units for each image in a one-versus-one approach, with leave-one-out cross-validation and shuffled-label control. Similar to the number of sites sampled by our microelectrode arrays, there were 58 units total at the final layer: 4 long-pathway units, 25 V1 → V4 → PIT units, 25 V2 → PIT units, and 4 units that received mixed inputs from long and short pathways.

We found that SVMs performed best at classifying each image when using all long- and short-pathway units ($N = 58$); SVMs achieved 46 ± 1% accuracy after baseline subtraction (mean of six models ± SEM). SVMs performed worse when relying only on short-pathway units: PIT units without input from the third layer ("V4") led to an SVM performance of 41 ± 1%, and PIT units without input from the second layer ("V2") led to 40 ± 1%. Thus, reducing the activity of the output layer by 43% (32 of 58) of all units worsened SVM performance by 5%–6%. As comparison, SVMs trained on monkey PIT data showed an average accuracy of 26 ± 2% (over baseline) before cooling, 18 ± 1% during V4 cooling, and 21 ± 1% during V2|3 cooling, thus reducing the performance of SVMs by 5%–8% (Fig. 6*b*,*c*). In summary, we found two common effects across the brain and the models: first, removing activity in PIT (achieved via input cooling) reduced SVM accuracy by an amount comparable with removing activity in the output layer of a HMAX model (achieved by querying smaller populations). Second, simulated PIT subpopulations that lacked second-layer ("V2") input versus third-layer ("V4") input performed comparably, allowing SVMs to achieve similar classification accuracy. This is consistent with our observation that there were no shape-specific deficits to V2|3 versus V4 cooling. We had also found that cooling V4 led to deeper reductions in classification accuracy in PIT, and the models did not show this reliably. This is consistent with our interpretation that the biological relevance of V4 to shape-encoding in PIT depends on its numerous anatomical projections, which we did not model here.

Discussion

There have been almost no studies describing the effects of early extrastriate area deactivation on shape encoding on IT neurons. We investigated how posterior inferotemporal cortex cells combine information from areas V2, V3, and V4 by implanting microelectrode arrays in PIT while cooling areas V2 and V3 (together) or area V4. We used linear classifiers to decode the information contained in PIT before and during cooling of each input area and found that cooling any of these areas resulted in a similar reduction in firing rate activity across PIT, but that cooling V4 led to a deeper reduction in SVM classification accuracy in both animals. This suggests that the strength of the connections predicted by anatomical projection maps was best reflected in the pattern of responses, not in the magnitude of the population firing rate. We also modified an HMAX model by adding different bypass projections and a similar number of units at the top layer as our arrays, and found that this model architecture was very robust to "lesions," in a range comparable with that observed during cooling of PIT inputs. We also found that simu-

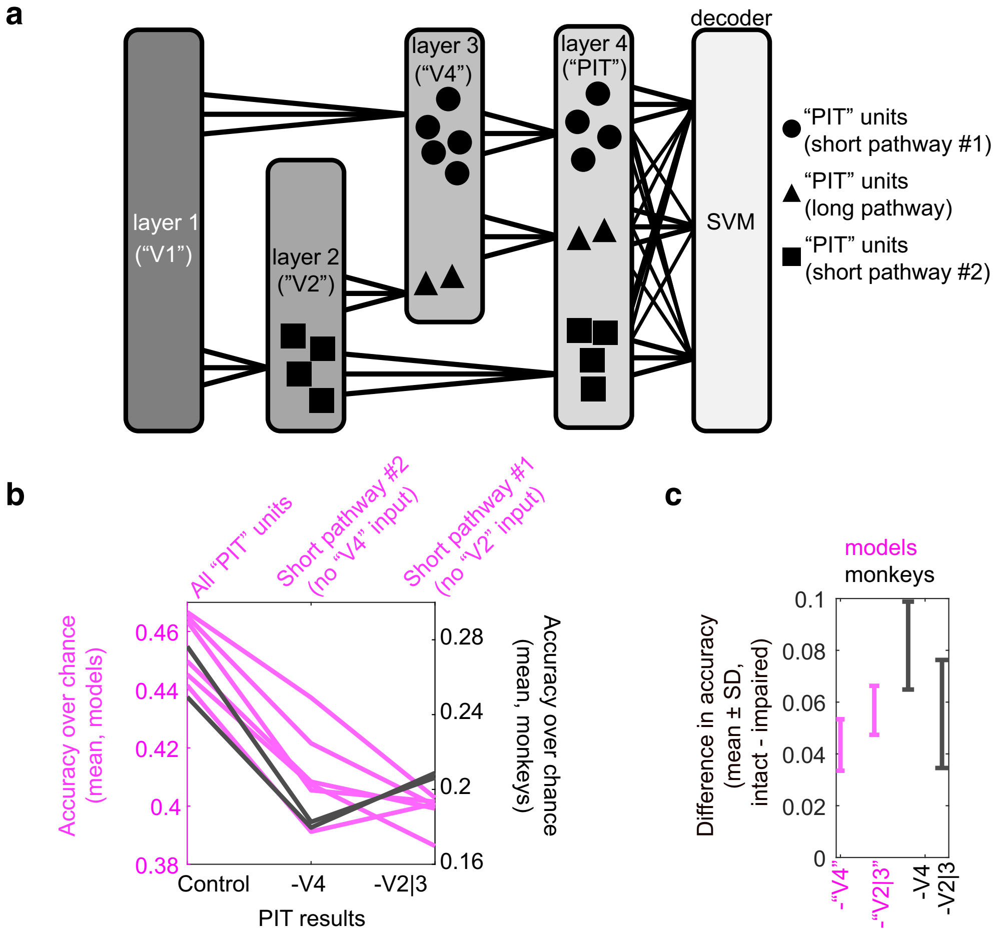

Figure 6. Computational approach to testing units with different input histories. ***a***, Architecture of our HMAX implementation, which included three parallel pathways resulting in three top layer units with different input histories (circles, triangles, and squares). SVMs were used to query each population independently or as a whole group. Six different models were trained with the same architecture but different RF shapes. ***b***, Mean classification accuracy over chance (±SEM) for both animals during the warm, V4 and V2|3 deactivation (black) and for all simulated PIT populations in six models (pink). The label "All 'PIT' units" describes results from querying the full population of short- and long-pathway PIT cells. ***c***, Differences in classification accuracy shown in ***b***, for different implementations of the model and monkeys. "Intact" refers to the "All 'PIT' units" condition in the model implementations and to the precooling condition in the monkeys. "Impaired" refers to the queried subpopulations in the model and to the cooling conditions in the monkeys.

lated "PIT" cells with different input histories (skipping either V2 or V4) were very similar at encoding image identity.

Validity of cooling manipulations

We used sulcal landmarks to position the cryoloops: one loop was placed over the lateral prelunate gyrus and others within the lunate sulcus. We have treated these landmarks as being equivalent to areas V4, V2, and V3. This assumption is strong because there is no variation in the location of these visual areas relative to these anatomical landmarks, as documented through decades of electrophysiology and imaging articles (Essen and Zeki, 1978; Gattass et al., 1981, 1988; Kennedy and Bullier, 1985; Boussaoud et al., 1991; Distler et al., 1993; Nakamura et al., 1993; Levitt et al., 1994; Gegenfurtner et al., 1997; Brewer et al., 2002; Fize et al., 2003; Ungerleider et al., 2008). We have also defined the spatiotemporal cooling properties of our cryoloops. These devices can reliably cool cortical tissue up to ~3 mm away from the metal tubing within minutes, and our intraoperative imaging is consistent with previous descriptions of more detailed thermocline information (Lomber et al., 1999, 2010; Lomber, 1999). Finally, our placement of the microelectrode arrays led to results consistent with previous descriptions of the retinotopic organization of PIT: Boussaoud et al. (1991) showed that the posterior border of PIT lies at the anterior lip of the inferior temporal sulcus and its anterior border at the posterior middle temporal sulcus. PIT runs as a band of cortex anterior to V4, sharing posterior-anterior isoeccentricity bands with V2, V3, and V4. In contrast to eccentricity, polar angle is poorly organized in PIT, such that the most reliable distinction in PIT is between superior and inferior visual fields (neurons with superior field RFs are located adjacent to V4, neurons with inferior RFs closer to AIT). Nearly all fields at the foveal and perifoveal representations of PIT are large enough to cross the horizontal meridian (Boussaoud et al., 1991; Yasuda et al., 2010). Our results confirm these observations. We placed our arrays at the convexity of the inferior temporal gyrus, anterior to the inferior temporal sulcus, and we observed that the aggregate RFs were located superiorly, biased toward the fovea, and extending into the lower hemifield. The effects of V2|3 and V4 deactivation manifested in the lower hemifield portion of the recorded RFs, as predicted by the retinotopy of V2, V3, and V4 (Gattass et al., 1981, 1988).

Why did input cooling not silence PIT activity entirely?

One might expect that disrupting the V1-V2|3-V4 input pathway would extinguish nearly all activity within the PIT scotoma. Indeed, the mean population firing rate fell by 38% when deactivating V2, V3, and V4. First, we did not cool all of V2, V3, or V4, the scotomas covered only a few degrees of central vision. Second, PIT receives inputs from at least 60 cortical regions, including V3A, V4t, 7A, 7B, LIP, DP, MT, MST, FST, anterior IT, insula, and prefrontal areas 9/46 (Distler et al., 1993; Markov et al., 2014), as well as subcortical structures, such as the pulvinar (Chow, 1950; Gross et al., 1974; Baizer et al., 1993). Markov et al. (2014) injected a retrograde marker in PIT and estimated the weight of each input source to PIT as the number of cell bodies stained in that input area, and they found that these numbers could be as few as a handful of neurons (e.g., seven cells in the insular pathway, or 0.004% of all PIT-projecting neurons) and as many as tens of thousands (39,000 in V4, or 26% of all projecting neurons). Per these results, the projection weight of area V2 is 2% (3782 cells), the weight of V3 is 12% (19,116 cells), and the weight of V4 is 26% (39,911 cells), for a total of 40% of all inputs to PIT. This is similar to the overall reduction in firing rate we observed (38% during V2/V3/V4 cooling). In the context of these many alternative inputs, we thus believe it may be practically impossible to silence PIT without radical interventions, such as bilateral V1 resection, silencing of lateral connections, and feedback. This is one possible reason why, in a different study, anterior IT cells showed no significant changes in overall firing rate after surgical resection of areas V4 and PIT (Buffalo et al., 2005).

Did we miss any classes of shape features that might be differentially represented between the V1-V4-PIT and V2-PIT input paths?

Although possible, there are no strong theoretical candidates for such features. Hegde and Van Essen (2007) compared the relative shape selectivities in neurons from V1, V2, and V4 and showed that these cells responded similarly to the same set of simplex and complex images, offering little qualitative diversity (Hegdé and Van Essen, 2007). Of course, one important difference between these visual areas is RF size, which suggests that these concurrent pathways may convey the same types of geometric primitives but at different scales: individual V4 cell inputs could carry information about spatially larger fragments than V2 cell inputs. We tried to measure changes in size tuning during each cooling condition in one monkey, but the relatively small size of our scotoma did not allow a sufficient experimental variation in stimulus size and position. Investigators have proposed other theoretical roles for these shortened pathways: the coarse, low-pass, fast transmission of color (Yukie and Iwai, 1985; Nakamura et al., 1993) and form (Nakamura et al., 1993; Serre et al., 2005); insurance against brain damage (Distler et al., 1993; Nakamura et al., 1993); and utility in fine recognition tasks (Serre et al., 2005). We found no evidence that lower spatial frequency information was differentially impaired during input deactivation, nor did we find that response latencies were reduced, as might be expected if information arriving from the shorter pathways got to PIT faster.

Another question raised by the cooling deactivations is that areas V2, V3, and V4 are themselves tightly interconnected, and thus removing either of these nodes nulled the primary input stream (V1 → V2 → V4 → PIT) and could have consequently induced the same effect. Of course, this is part of the logic of the experiment, which was to expose differences in the contributions of the smaller bypass pathways by removing the main pathway as a constant: cooling V2 would expose the known V1 → V4 → PIT anatomical inputs (Kuypers et al., 1965; Yukie and Iwai, 1985; Nakamura et al., 1993; Ungerleider et al., 2008) while cooling V4, the anatomical V1 → V2 → PIT inputs (Distler et al., 1993). Our results do show a significant effect overlap between both V2|3 and V4 cooling interventions, best exemplified by the reductions in overall firing rate in PIT. However, other analyses exposed partial differences in the input pathways, such as the fact that the V4-based inputs were more helpful in classification accuracy than V2|3-based inputs. Our multivariate analyses suggested that the effects of overall input magnitude are separable from the effects of shape-encoding mechanisms in PIT. It is important to make this distinction in future lesion studies and to use multivariate pattern analyses.

Finally, we found it interesting that the loss in decoding performance of the HMAX model under "impaired" versus "intact" conditions compared well with our data. Deep convolutional networks are becoming the best explanatory framework for studies of visual recognition, a framework that grows more diverse with the addition of residual networks ("ResNets"), which use shortcut connections that skip one or more layers (He et al., 2015). ResNets are intriguing because their shortcut connections are biologically relevant, allow the networks to be very deep yet trainable, and provide resilience to "lesions," a feature consistent with results in this study (He et al., 2016; Veit et al., 2016). However, it is difficult to say that current ResNets are the best models for the visual system, as the best performing versions are very deep (100–1000 layers) compared with the visual system (~7–8), AlexNet (8), and HMO models (3) (Krizhevsky et al., 2012; Yamins et al., 2014), and it is not yet clear whether ResNets can better account for IT response variance than other models. This is worthy of further study.

References

Baizer JS, Desimone R, Ungerleider LG (1993) Comparison of subcortical connections of inferior temporal and posterior parietal cortex in monkeys. Vis Neurosci 10:59–72. CrossRef Medline

Baldassi C, Alemi-Neissi A, Pagan M, Dicarlo JJ, Zecchina R, Zoccolan D (2013) Shape similarity, better than semantic membership, accounts for the structure of visual object representations in a population of monkey inferotemporal neurons. PLoS Comput Biol 9:e1003167. CrossRef Medline

Boussaoud D, Desimone R, Ungerleider LG (1991) Visual topography of area TEO in the macaque. J Comp Neurol 306:554–575. CrossRef Medline

Brewer AA, Press WA, Logothetis NK, Wandell BA (2002) Visual areas in macaque cortex measured using functional magnetic resonance imaging. J Neurosci 22:10416–10426. Medline

Buffalo EA, Bertini G, Ungerleider LG, Desimone R (2005) Impaired filtering of distracter stimuli by TE neurons following V4 and TEO lesions in macaques. Cereb Cortex 15:141–151. CrossRef Medline

Carrasco A, Brown TA, Kok MA, Chabot N, Kral A, Lomber SG (2013) Influence of core auditory cortical areas on acoustically evoked activity in contralateral primary auditory cortex. J Neurosci 33:776–789. CrossRef Medline

Chow KL (1950) A retrograde cell degeneration study of the cortical projection field of the pulvinar in the monkey. J Comp Neurol 93:313–340. CrossRef Medline

Cowey A, Gross CG (1970) Effects of foveal prestriate and inferotemporal lesions on visual discrimination by rhesus monkeys. Exp Brain Res 11: 128–144. Medline

Dean P (1976) Effects of inferotemporal lesions on the behavior of monkeys. Psychol Bull 83:41–71. CrossRef Medline

Desimone R, Lehky S, Ungerleider L, Mishkin M (1990) Effects of V4 lesions on visual discrimination performance and on responses of neurons in inferior temporal cortex. Soc Neurosci Abstr 16:Abstract 260.7.

DiCarlo JJ, Cox DD (2007) Untangling invariant object recognition. Trends Cogn Sci 11:333–341. CrossRef Medline

Distler C, Boussaoud D, Desimone R, Ungerleider LG (1993) Cortical connections of inferior temporal area TEO in macaque monkeys. J Comp Neurol 334:125–150. CrossRef Medline

Essen DC, Zeki SM (1978) The topographic organization of rhesus monkey prestriate cortex. J Physiol 277:193–226. CrossRef Medline

Fize D, Vanduffel W, Nelissen K, Denys K, Chef d'Hotel C, Faugeras O, Orban

GA (2003) The retinotopic organization of primate dorsal V4 and surrounding areas: a functional magnetic resonance imaging study in awake monkeys. J Neurosci 23:7395–7406. Medline

Fukushima K (1975) Cognitron: a self-organizing multilayered neural network. Biol Cybern 20:121–136. CrossRef Medline

Gattass R, Gross CG, Sandell JH (1981) Visual topography of V2 in the macaque. J Comp Neurol 201:519–539. CrossRef Medline

Gattass R, Sousa AP, Gross CG (1988) Visuotopic organization and extent of V3 and V4 of the macaque. J Neurosci 8:1831–1845. Medline

Gegenfurtner KR, Kiper DC, Levitt JB (1997) Functional properties of neurons in macaque area V3. J Neurophysiol 77:1906–1923. Medline

Gross C, Bender DB, Rocha-Miranda CE (1974) Inferotemporal cortex: a single unit analysis. In: The neurosciences: a third study program (Schmitt F, Worden F, eds). Cambridge, MA: Massachusetts Institute of Technology.

He K, Zhang X, Ren S, Sun J (2015) Deep residual learning for image recognition. arXiv:1512.03385

He K, Zhang X, Ren S, Sun J (2016) Identity mappings in deep residual networks. arXiv:1603.05027

Hegdé J, Van Essen DC (2007) A comparative study of shape representation in macaque visual areas v2 and v4. Cereb Cortex 17:1100–1116. CrossRef Medline

Heywood CA, Cowey A (1987) On the role of cortical area V4 in the discrimination of hue and pattern in macaque monkeys. J Neurosci 7:2601–2617. Medline

Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160:106–154. CrossRef Medline

Kennedy H, Bullier J (1985) A double-labeling investigation of the afferent connectivity to cortical areas V1 and V2 of the macaque monkey. J Neurosci 5:2815–2830. Medline

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

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

Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 1097–1105. Available online at: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.

Kuypers HG, Szwarcbart MK, Mishkin M, Rosvold HE (1965) Occipitotemporal corticortical connections in the Rhesus monkey. Exp Neurol 11:245–262. CrossRef Medline

Levitt JB, Kiper DC, Movshon JA (1994) Receptive fields and functional architecture of macaque V2. J Neurophysiol 71:2517–2542. Medline

Lomber SG (1999) The advantages and limitations of permanent or reversible deactivation techniques in the assessment of neural function. J Neurosci Methods 86:109–117. CrossRef Medline

Lomber SG, Payne BR, Horel JA (1999) The cryoloop: an adaptable reversible cooling deactivation method for behavioral or electrophysiological assessment of neural function. J Neurosci Methods 86:179–194. CrossRef Medline

Lomber SG, Meredith MA, Kral A (2010) Cross-modal plasticity in specific auditory cortices underlies visual compensations in the deaf. Nat Neurosci 13:1421–1427. CrossRef Medline

Markov NT, Misery P, Falchier A, Lamy C, Vezoli J, Quilodran R, Gariel MA, Giroud P, Ercsey-Ravasz M, Pilaz LJ, Huissoud C, Barone P, Dehay C, Toroczkai Z, Van Essen DC, Kennedy H, Knoblauch K (2011) Weight consistency specifies regularities of macaque cortical networks. Cereb Cortex 21:1254–1272. CrossRef Medline

Markov NT, Ercsey-Ravasz MM, Ribeiro Gomes AR, Lamy C, Magrou L, Vezoli J, Misery P, Falchier A, Quilodran R, Gariel MA, Sallet J, Gamanut R, Huissoud C, Clavagnier S, Giroud P, Sappey-Marinier D, Barone P, Dehay C, Toroczkai Z, Knoblauch K, et al. (2014) A weighted and directed interareal connectivity matrix for macaque cerebral cortex. Cereb Cortex 24:17–36. CrossRef Medline

Merigan WH (1996) Basic visual capacities and shape discrimination after lesions of extrastriate area V4 in macaques. Vis Neurosci 13:51–60. CrossRef Medline

Merigan WH, Pham HA (1998) V4 lesions in macaques affect both single- and multiple-viewpoint shape discriminations. Vis Neurosci 15:359–367. Medline

Merigan WH, Nealey TA, Maunsell JH (1993) Visual effects of lesions of cortical area V2 in macaques. J Neurosci 13:3180–3191. Medline

Nakamura H, Gattass R, Desimone R, Ungerleider LG (1993) The modular organization of projections from areas V1 and V2 to areas V4 and TEO in macaques. J Neurosci 13:3681–3691. Medline

Ponce CR, Lomber SG, Born RT (2008) Integrating motion and depth via parallel pathways. Nat Neurosci 11:216–223. CrossRef Medline

Ponce CR, Hunter JN, Pack CC, Lomber SG, Born RT (2011) Contributions of indirect pathways to visual response properties in macaque middle temporal area MT. J Neurosci 31:3894–3903. CrossRef Medline

Portilla J, Simoncelli EP (2000) A parametric texture model based on joint statistics of complex wavelet coefficients. Int J Comput Vis 40:49–70. CrossRef

Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. Nat Neurosci 2:1019–1025. CrossRef Medline

Rust NC, DiCarlo JJ (2010) Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. J Neurosci 30:12978–12995. CrossRef Medline

Serre T, Kouh M, Cadieu C, Knoblich U, Kreiman G, Poggio T (2005) MIT AI Memo 2005–036/CBCL Memo 259. ftp://publications.ai.mit.edu/ai-publications/2005/AIM-2005-036.pdf.

Serre T, Kreiman G, Kouh M, Cadieu C, Knoblich U, Poggio T (2007a) A quantitative theory of immediate visual recognition. Prog Brain Res 165:33–56. CrossRef Medline

Serre T, Oliva A, Poggio T (2007b) A feedforward architecture accounts for rapid categorization. Proc Natl Acad Sci U S A 104:6424–6429. CrossRef Medline

Serre T, Wolf L, Bileschi S, Riesenhuber M, Poggio T (2007c) Robust object recognition with cortex-like mechanisms. IEEE Trans Pattern Anal Mach Intell 29:411–426. CrossRef Medline

Ungerleider LG, Galkin TW, Desimone R, Gattass R (2008) Cortical connections of area V4 in the macaque. Cereb Cortex 18:477–499. CrossRef Medline

Veit A, Wilber M, Belongie S (2016) Residual networks behave like ensembles of relatively shallow networks. arXiv:1605.06431

Wilson WA Jr, Mishkin M (1959) Comparison of the effects of inferotemporal and lateral occipital lesions on visually guided behavior in monkeys. J Comp Physiol Psychol 52:10–17. CrossRef Medline

Yamins DL, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc Natl Acad Sci U S A 111:8619–8624. CrossRef Medline

Yasuda M, Banno T, Komatsu H (2010) Color selectivity of neurons in the posterior inferior temporal cortex of the macaque monkey. Cereb Cortex 20:1630–1646. CrossRef Medline

Yukie M, Iwai E (1985) Laminar origin of direct projection from cortex area V1 to V4 in the rhesus monkey. Brain Res 346:383–386. CrossRef Medline