A shared code for perceiving and imagining objects in human ventral temporal cortex

Authors:

V. S. Wadia[1,3], C. M. Reed[2], J. M. Chung[2], L. M. Bateman[2], A. N. Mamelak[1], U. Rutishauser[ψ1,2,3,4], D. Y. Tsao[ψ5,6]

Affiliations:

1 Department of Neurosurgery, Cedars-Sinai Medical Center, Los Angeles, CA, USA
2 Department of Neurology, Cedars-Sinai Medical Center, Los Angeles, CA, USA
3 Division of Biology and Bioengineering, California Institute of Technology, Pasadena, CA, USA
4 Center for Neural Science and Medicine, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, USA
5 Department of Neuroscience, Helen Wills Neuroscience Institute, University of California, Berkeley, CA, USA
6 Howard Hughes Medical Institute, Berkeley, CA, USA

ψ - Joint senior authors

Keywords:

Visual perception; object recognition; visual imagery; inferotemporal cortex; intracranial electrophysiology; generative model; human brain;

Abstract: Mental imagery is a remarkable phenomenon that allows us to remember previous experiences and imagine new ones. Animal studies have yielded rich insight into mechanisms for visual perception, but the neural mechanisms for visual imagery remain poorly understood. Here, we first determined that ~80% of visually responsive single neurons in human ventral temporal cortex (VTC) used a distributed axis code to represent objects. We then used that code to reconstruct objects and generate maximally effective synthetic stimuli. Finally, we recorded responses from the same neural population while subjects imagined specific objects and found that ~40% of axis-tuned VTC neurons recapitulated the visual code. Our findings reveal that visual imagery is supported by reactivation of the same neurons involved in perception, providing single neuron evidence for existence of a generative model in human VTC.

One Sentence Summary: Single neurons in human temporal cortex use feature axes to encode objects, and imagery reactivates this code.

Introduction

Mental imagery refers to our brains' capacity to generate percepts, emotions, and thoughts in the absence of external stimuli. This ability pervades many aspects of the human condition. It allows for the generation of visual art (*1–3*), musical composition (*4–7*), and creative writing (*8–10*). It subserves efficient planning (*11*, *12*) and navigation (*13–15*) via the simulation of actions and outcomes. It is also the basis for calling to mind a recent experience, person, place, or object, which is a key aspect of episodic memory (*11*, *16–20*). In a clinical setting, uncontrolled mental imagery can contribute to psychological disorders including anxiety, schizophrenia, and post-traumatic stress disorder (*21*).

Perhaps the most consistent and established finding in mental imagery research is that imagery of a given sense co-opts the neural machinery used for perception, in other words imagery is supported by activation of sensory areas. This has been shown explicitly during auditory (*22–25*), olfactory (*26*, *27*), tactile (*28*), speech (*29*), and even motor imagery (*30–34*), though most extensively in visual imagery (*35–46*); damage in or loss of either dorsal or ventral visual pathways (*47*) often leads to parallel deficits in visual imagery (*48–50*). However, these studies lack the spatial resolution to determine whether imagery reactivates the exact same neural populations that support perception, i.e., the feedforward neural code (which would imply a generative model) or whether instead imagery is subserved by activation of separate circuitry roughly located in the same cortical regions. The concept of a generative model wherein representations can be hierarchically re-activated in a top-down manner has great power computationally, explaining self-supervised learning (*51*), inference under ambiguity (*52*), and object-based attention (*53*). Yet, direct single-neuron evidence for regeneration of the sensory code during imagery has been conspicuously lacking.

Here, we attempt to shed light on the single-neuron mechanisms of visual imagery by determining the code for visual objects and then examining whether this code is reactivated during imagery. We focused our investigations on ventral temporal cortex (often referred to as inferotemporal cortex), a large swath of the primate temporal lobe dedicated to representing visual objects (*54*, *55*). We recorded single neurons in human patients implanted to localize their focal epilepsy (*56*) as they viewed and subsequently visualized carefully parametrized visual objects. First, we found that, as in non-human primates (*57–60*), human VTC neurons showed robust visual responses (*61–65*), and are well modelled by linearly combining features in deep layers of a deep network trained to perform object classification (*66*). We confirmed two consequences of such a code: each neuron had a linear null space orthogonal to its encoding axis (given by the coefficients of linear combination); furthermore, each neuron responded maximally to synthetic stimuli generated using its encoding axis. Second, we asked subjects to imagine (i.e., visualize) a diverse subset of objects that they had previously seen, while recording responses of visually-characterized VTC neurons. We found that a subset (~40%) of axis-tuned neurons reactivated during imagery, and the imagined responses of individual neurons to specific objects were proportional to the projection value of those objects onto the neurons' preferred axes.

Together these findings provide evidence for the implementation of a generative model in human VTC by neurons that represent both real and imagined stimuli.

Results

Neurons show diverse category tuning

We examined how human VTC neurons encode visual objects by recording responses of these neurons to a large set of objects with varying features using a rapid screening task. Patients sequentially viewed a series of 500 images (4 repetitions each, 2000 total trials), drawn from face, text, plant, animal, and object categories (Figure 1A, top; Figure S1A for detailed schematic; Figure S2A for stimulus examples). At random intervals a catch question pertaining to the immediately preceding image would appear on the screen. Images stayed on screen for 250 ms and the inter trial interval was randomized between 100-150 ms. Despite the rapid presentation rate, patients answered 77% of the catch questions correctly, indicating that the stimuli were carefully attended to (Figure 1B).

We recorded 714 VTC neurons in 57 sessions across 16 patients. Response onset latency of individual neurons was computed on a trial-by-trial basis using a Poisson burst metric (*67*) (see methods). Out of the 714 neurons, 456 showed a significantly increased response to the onset of a visual stimulus (Figure 1C; 1 x 5 sliding window ANOVA, bin size 50 ms, step size 5 ms, 5 consecutive significant bins with $p < 0.01$; see methods). The locations of all electrodes from which we recorded visually responsive neurons can be seen in Figure 1D (red and blue dots); supplementary tables S1 and S2 contain neuron count and patient demographic information. The mean response latency was 162 ms (Figure 1E). Human VTC neurons showed diverse tuning patterns, including neurons that were selectively silent to a given category (Figure 1F), neurons that were maximally responsive to a given category (Figure 1G, top), neurons characterized by an initial suppression of activity (Figure 1G, middle), and neurons that distinguished categories via a latency code (Figure 1G, bottom).

Neurons are tuned to specific axes in object space

We next examined whether visually responsive neurons encoded specific object features. We leveraged deep networks trained on object classification to build a low-dimensional object space that captures the shape and appearance of arbitrary objects (*68*) without relying on subjective visual descriptions. We built our object space by passing the 500 images shown to patients through AlexNet (*66*) and performing principal components analysis (PCA) on the unit activations in layer fc6 (Figure 2A). We determined that 50 dimensions explained 80.68% of the variance in fc6 responses (Figure S2B), and used these 50 dimensions in all remaining analyses. This allowed us to describe every visual object shown to patients as a point in a 50-dimensional feature space. We next investigated how neural activity mapped onto this feature space.

For each cell, we computed a 'preferred axis' given by the coefficients c in the equation $\vec{r} = c \cdot \vec{f} + \vec{c}_0$, where $\vec{r}$ is the response of the neuron to a given image, $\vec{f}$ is the 50D object feature vector of that image, and $\vec{c}_0$ is a constant offset (see methods). The axis tuning of an example neuron is shown in Figure 2B. This neuron showed a monotonically increasing response to stimuli with higher projection values onto its preferred axis (see methods) while showing no change in tuning along an axis orthogonal to its preferred axis. We confirmed that the correlation between projection value and firing rate was significant when compared to a shuffled distribution along the preferred axis ($p < 0.01$, Figure 2B, top distribution) and not significant along the principal orthogonal axis ($p = 0.496$, Figure 2B, bottom distribution). This "axis code" emphasizes the geometric picture that neurons are projecting incoming stimuli, formatted as vectors in the complex feature space, onto their specific preferred axes in the space (*57*, *58*, *69*).

Axis tuning also clarified the response pattern of some neurons with complex tuning. The example neuron in Figure 2C responded most strongly to the plant category, but also responded strongly to specific stimuli in other categories (see raster, Figure 2C). Indeed, no semantic label could obviously delineate between the stimuli that elicited the strongest and weakest response from the neuron (Figure 2D). However, this neuron showed significant axis tuning (Figure 2E), and sorting the stimuli according to their projection value onto the preferred axis shows that the top stimuli were those with the highest projection values and vice versa (Figure 2F). Across the population, we found that a majority (367/456, ~80%) of visually responsive neurons were significantly axis tuned (Figure 2G, H), showing a significant positive correlation between projection value and response along the preferred axis with no such correlation along the orthogonal axis (Figure 2I; r_{pref} = 0.54, r_{ortho} = -1.3833e-10, p = 1.43e-121, Wilcoxon ranksum test).

We subsequently compared the variance explained by the axis model to that explained by two alternative models: a category label model and an exemplar model. The category label model was chosen for comparison due to the robust category responses seen in VTC (Figure 1F, G). The exemplar model is a well-known alternative to the axis model (*70–72*) which posits that neurons have maximal responses to specific exemplars (i.e., specific points in object space) and decaying responses to objects with increasing distance from the exemplar (see methods). Only 18/456 neurons had good fits (positive explained variance) across all models (34/456 exemplar model, 343/456 category label, 339/456 axis model) and even in these neurons the axis model explained significantly more variance than either alternative (39.62% axis, 17.04% category, p = 9.42e-04, Wilcoxon ranksum test; 39.62% axis, 10.37% exemplar, p = 4.78e-05, Wilcoxon ranksum test; Figure 2J, K).

We built our low-dimensional object space by leveraging AlexNet, a deep network trained to perform object classification. However, there now exist a plethora of deep convolutional neural network models that are capable of performing object recognition at or beyond human capability (*73*). A comparison of several such models (VGG-16/19/Face and CORNet models; see supplementary methods) revealed that axis tuning in feature spaces built from the fully connected layers of these models explained a large amount of the explainable neural variance (~45%) in the 367 axis-tuned VTC neurons (Figure S3, S4). Moreover, the proportion of explainable

variance explained was roughly the same for all models except VGG-Face and an 'eigen-model' where PCA is conducted directly on the pixels; implying that the axis code is independent of the specific convolutional network used to build the feature space.

Reconstructing objects using human VTC responses

We next investigated the richness of the VTC representation by attempting to reconstruct objects using VTC responses. A consequence of the linear relationship between VTC neuron responses and object features is the ease of learning a linear decoder that predicts object feature values from the population activity (*74–76*). The responses of the neurons can be approximated as a linear combination of object features, with the slopes of the ramps corresponding to the weights (*69–71*). Then for a population of neurons, $\vec{R} = C\vec{F} + \vec{C}_0$, where $\vec{R}$ is the response vector of the different neurons, C is the weight matrix, $\vec{F}$ is the vector of object feature values, and $\vec{C}_0$ is the offset vector. Inverting this equation to solve for $\vec{F}$ provides a linear decoder that predicts object feature values from the population activity (*74–76*) (see methods). We used this approach coupled with leave-one-out cross-validation to learn the linear transform that maps responses to features (Figure 3A).

Figure 3B shows the correlation between actual feature values and model predictions for the first two dimensions of object space; both correlations were significant positive values ($r_{\dim 1}$ = 0.46, p = 2.6e-28, paired t-test, $r_{\dim 2}$ = 0.69, p = 0, paired t-test); Figure S2C shows correlations for all 50 dimensions. To reconstruct objects, we searched a large auxiliary object database for the object with the feature vector closest to that decoded from the neural activity. A normalized distance in feature space between the best possible and actual reconstruction was computed to quantify the decoding accuracy across all stimuli (Figure 3C, see methods). Side-by-side comparisons of the original images with the 'reconstructions' chosen from the auxiliary database showed striking visual correspondence (Figure 3D). We also confirmed that VTC responses could support well-above-chance decoding of the correct object from distractors (Figure 3E).

Predicting responses to synthetic stimuli

To further validate the axis model, we attempted to predict the responses of VTC neurons to images not shown to our patients, in a "closed loop" experiment. We used an initial screening session to compute axes for all neurons recorded. We then used a generative adversarial network (GAN) trained to invert the responses of AlexNet (*77*, *78*) to systematically generate images corresponding to evenly spaced points along both the preferred and orthogonal axes. We then returned to the patient room for a second session and rescreened with the synthetic images added to the original stimulus set. We performed this experiment in 8 sessions across 4 patients which yielded 16 axis tuned neurons.

We predicted that images with projection values greater than the maximum projection value of the original stimulus images should serve as 'super-stimuli' driving the neuron to a higher response than any of the original stimulus images (*78*). Figure 3F shows an example neuron tested

with the synthetic stimuli. This neuron's most preferred stimulus was a t-shirt, while the least preferred stimulus was a ladder (Figure 3F); these two stimuli were also the ones that projected most and least strongly onto the cell's preferred axis. Synthetic stimuli were sampled evenly along both preferred (yellow dots) and orthogonal axes (green dots) such that the extremes had larger projection values than any of the original stimulus images (Figure 3F). Responses to these synthetic stimuli demonstrate the expected increase along the preferred axis and no increase along the orthogonal axis (Spearman's rank correlation between projection value and firing rate for preferred axis 0.8303, p = 4e-3 when compared to shuffled distribution, orthogonal axis = -0.1037, p = 0.604; Figure 3F). Moreover, the firing rates to the subset of synthetic stimuli with larger projection values along the preferred axis than the original stimulus images were substantially higher than those to the original stimulus images. Interestingly, the most and least effective synthetic stimuli showed striking resemblance to the most and least effective original stimuli. Figure 3G shows responses of the same cell to synthetic stimuli evenly sampled in a 2D grid spanned by the preferred and orthogonal axes (purple square, Figure 3F); responses showed the expected changes in tuning only along directions parallel to the preferred axis. Figure 3H shows a second example neuron which was face selective (r_{pref} = 0.89, p = 5.53e-04; r_{ortho} = 0.52, p = 0.12).

Across all 16 neurons the correlation between projection value and firing rate for the synthetic images along both was significantly higher along the preferred axis (Figure 3I, r_{pref} = 0.58, r_{ortho} = 0.08, p = 1.34e-05, Wilcoxon ranksum test). In this experiment the axes were computed using the 500 original stimulus images only (see methods). The axes were also used to predict the firing rate responses to the synthetic stimuli. The distribution of correlation values between the responses predicted using the axes computed in the first session and the recorded responses in the second session are shown in Figure 3J (mean = 0.58).

Neuronal activity during imagery

So far, we have established that human VTC uses an axis code to represent objects and confirmed this code through stimulus reconstruction and generation of super-stimuli. Armed with this understanding of the sensory code, we could now tackle the single neuron mechanisms of visual imagery. In 12 sessions across 6 patients, we tested patients on a cued imagery task following visual screening (Figure 1A, bottom; Figure S1B for detailed schematic). In the imagery task, patients viewed and subsequently visualized from memory 6-8 objects out of the original 500 used for screening. Each trial required alternating, cued visualization of two object stimuli after an initial encoding period during which patients passively viewed the two stimuli. During the cued imagery period patients would close their eyes and imagine the 2 objects in the trial in alternating 5 s periods until each stimulus had been imagined 4 times, being informed to switch images at the end of a given 5 s period verbally by the experimenter. Each image appeared in 2 separate trials for a total of 8 imagery trials per image (see methods). We recorded 231 VTC neurons in the imagery task, of which 131 were visually responsive and 107 were axis tuned.

We found robust activation of neurons in human VTC during imagery, with 66/231 neurons (~30%) showing activation to at least one object (Figure 4A, 1 x n sliding window ANOVA or sliding window ttest, n = number of stimuli, bin size 1.5 s, step size 300 ms, 6 consecutive bins with $p < 0.05$ – see methods). As in vision, neurons showed a diverse range of response profiles during imagery, some activating sparsely during imagery of a single specific stimulus (Figure 4B), others activating to imagery of multiple stimuli in a graded manner (Figure 4C), and a small number (15/231) activating only during imagery and not during viewing (Figure 4D).

VTC neurons recapitulate the visual code during imagery

A central goal of the current study was to clarify whether imagery reactivates visual neurons in a way that respects their perceptual code, or whether an alternative code (possibly implemented by a distinct population of neurons) is recruited during imagery. The former would constitute strong evidence for the existence of a generative model in the human brain. To establish whether VTC neurons reactivate in a manner that respects the axis code, the rapid screening session was conducted using the 500 object images described earlier and axes were computed for the axis-tuned neurons. Then, 6-8 stimuli that were spread along the preferred and orthogonal axes were chosen for use in the cued imagery task. Figures 5A&B show two example neurons: both showed significant axis tuning and responded most strongly to the image with the largest projection value onto the preferred axes during both encoding and recall (piano and mirror respectively).

Examining the population of axis tuned neurons that reactivated during imagery (43/107, ~40%) revealed a significant correlation between projection value onto the neurons' preferred axes (computed using screening responses) and responses during imagery (r_{pref} = 0.20, Figure 5E left, p = 0.001 as compared to a shuffled distribution, Figure 5F top) with no such correlation along the orthogonal axes (r_{ortho} = -0.08, Figure 5E right, p = 0.929, Figure 5F bottom). Lastly, the distribution of Spearman's rank correlation coefficients between viewing and imagery of the same images in reactivated axis tuned neurons showed a high median value significantly larger than 0 (0.18 ± 0.05, p = 2.4e-03, one sample ttest; Figure 5G). Taken together these findings indicate that neurons in human VTC support visual imagery by reinstating similar network activity to viewing.

Discussion

In this paper, we explore the long-standing hypothesis that mental imagery is supported by reactivation of sensory areas in the visual domain. We focused our investigations on VTC, a region long known to harbor representations of complex visual objects (*57*, *58*, *64*). We first mapped out the feedforward code for visual objects, and then measured responses of the same population during imagery. We find that, as in non-human primates (*57*, *58*, *69*), human VTC neurons (367 out of 456 visually responsive neurons recorded) represented visual objects via linear projection of incoming object vectors onto a specific 'preferred' axis in a high dimensional feature space built using the unit activations of a deep network. Confirming the axis model, we could reconstruct viewed objects with high accuracy using a linear decoder (Figure 3A-E), and generate

synthetic super-stimuli for cells using the axis mapped to real stimuli (Figure 3F-J). These results indicate that the VTC neurons reported in this study serve to recognize objects by simply measuring their features not identifying them semantically, and that axis tuning is a powerful quantitative way to conceptualize the response of a substantial portion of human VTC neurons. Thus it would appear that in macaque, human, and deep neural network (DNN) architectures, an essential stage of object processing relies on a meaning-agnostic distributed shape representation (*57–60*, *68*, *79*).

A subset of neurons (66/231 total, 43/107 axis tuned) reactivated during imagery in a manner that respected the axis code: imagined responses were significantly correlated to projection value onto the preferred axis but not the orthogonal axis (Figure 5F), and viewed and imagined responses were positively correlated (Figure 5G). No previous single neuron study has examined visual imagery while recording from neurons in human VTC whose sensory code was characterized in detail. A single neuron study of imagery in the medial temporal lobe and another of spoken free recall (a related behavior) in human VTC demonstrated reactivation of a few neurons in both areas (*62*, *80*), but in both of these studies the sensory code was unknown. Our findings demonstrate for the first time that neurons in VTC support visual imagery by reactivating in a structured manner that respects the visual code.

The source of the top-down signal driving VTC reactivation during imagery remains an open question. Candidates for this source include the hippocampus and prefrontal cortex, given their involvement in various forms of memory (*81–87*), their dense connections to VTC (*88*, *89*), and the known ability of human hippocampal neurons to be selectively reactivated by free recall (*19*, *90*). Another question is the relationship between the VTC signals during imagery and those previously reported in primary visual areas (V1/V2) (*39*, *91*, *92*). Given hierarchically organized feedback connections (*88*), we hypothesize that the VTC signals may be driving the imagery-related signals in earlier visual areas. Future work is required to investigate the response characteristics of reactivated neurons across the brain, including those both upstream and downstream of VTC during imagery.

This work is the first to reveal a detailed understanding of the neural codes underlying visual object perception and imagery in human VTC. In particular, the results provide evidence for the existence of a generative model in the human brain—a mechanism capable of synthesizing detailed sensory contents from an abstract, semantic representation (*93–95*), effectively inverting the classic feedforward pathway. Generative models derive incredible computational power by transforming challenges in perception and cognition, such as inference under ambiguity (*52*, *93*, *94*) and object-based attention (*94*), into closed-loop feedback systems (*96*). The existence of a generative model in the human brain may even explain creative artistic processes that have so far remained out of reach of neuroscientific understanding.

Methods

Participants

The study participants were 16 adult patients who were implanted with depth electrodes for seizure monitoring as part of an evaluation for treatment of drug-resistant epilepsy (see Table S2 for demographic data). All patients provided informed consent and volunteered to participate in this study. Research protocols were approved by the institutional review board of Cedars-Sinai Medical Center (Study 572). The tasks were conducted while patients stayed in the epilepsy monitoring unit following implantation of depth electrodes. The location of the implanted electrodes was solely determined by clinical needs. The neural results were analyzed across all 16 patients. Each of the 16 patients included in this study had at least one depth electrode targeting the ventral temporal cortex.

Psychophysical tasks

Patients participated in three different tasks: an initial screening, cued imagery, and a final re-screening. The initial screening session was conducted to identify axis tuned neurons. Then 6-8 stimuli were chosen for the imagery task that had some spread along both the preferred and orthogonal axes. After axis tuned neurons were identified and the stimuli chosen we then conducted the cued imagery task, followed by another screening session immediately after. The second screening was used to match the neurons from the first and second sessions.

Screening

Patients viewed a set of 500 object stimuli (grayscale, white background, size: 224x224) with varying features (taken from www.freepngs.com) 4 times each for a total of 2000 trials in a shuffled order. Images were displayed on laptop computer with a 15.5” screen placed 1 meter away and subtended 6-7 visual degrees. Each image stayed on screen for 250 ms and the inter-trial interval consisting of a blank screen was jittered between 100-150 ms. The task was punctuated with yes/no ‘catch’ questions pertaining to the image that came right before the question requiring the patients to pay close attention in order to answer them correctly (Figure S1A). Catch questions occurred between 2 to 80 trials after the previous one. Patients responded to the questions using an RB-844 response pad (https://cedrus.com/rb_series/). During the synthetic image screens, the synthetic images would be added to the original stimulus set and the task parameters remained unchanged.

Cued Imagery

Patients viewed a set of 6-8 object stimuli chosen from the 500 used for screening (taken from www.freepngs.com). Each trial focused on 2 images and had an encoding period, a visual search

distraction period, and a cued imagery period. During encoding patients would see the 2 images 4 times each in a shuffled order. Each image stayed on screen for 1.5 s and the inter-image interval was 1.5-2 s. After this encoding period a visual search puzzle was presented (puzzles created by artist Gergely Dudas https://thedudolf.blogspot.com/) and stayed on screen for 30 s. After reporting via button press whether or not they were able to find the object in the puzzle, patients began the cued imagery period. During cued imagery patients would close their eyes and imagine the stimuli in the trial in an alternating fashion for 4 repetitions of 5 s each (40 s continuous imagery period). Patients were cued to switch the image they were imagining every 5 s by verbal cue (Figure S1B). After the imagery period patients would begin the next trial via button press when ready. Every image was present in 2 trials, leading to 8 repetitions of both encoding and imagery for each image.

Electrophysiology

The data in this paper was recorded from left and/or right VTC in addition to the other clinically relevant targets (unique to each patient) using Behnke-Fried micro-macro electrodes (Ad-Tech Medical Instrument Corporation) (*95*). All analyses in this paper are based on the signals recorded from the 8 microwires protruding from the end of the electrode. Recordings were performed with an FDA-approved electrophysiology system, and sampled at 32khz (ATLAS, Neuralynx Inc.) (*97*).

Spike sorting and quality metrics

Signals were bandpass filtered offline in the range of 300-3000hz with a zero phase lag filter before spike detection. Spike detection and sorting was carried out via the semiautomated template matching algorithm Osort (*98*). The properties of clusters identified as putative neurons and subsequently used for analysis were documented using a suite of spike quality metrics (Figure S5).

Electrode localization

Electrode localization was based on postoperative imaging using either MRI or computed tomography (CT) scans. We co-registered postoperative and preoperative MRIs using Freesurfer's mri_robust_register (*99*). To summarize recording locations across participants we aligned each participant's preoperative MRI to the CITI168 template brain in MNI152 coordinates (*100*) using a concatenation of an affine transformation and symmetric image normalization (SyN) diffeomorphic transform (*101*). The MNI coordinates of the microwires from a given electrode shank were marked as one location. MNI coordinates of microwires with putative neurons detected from all participants were plotted on a template brain for visualization (Figure 1D).

Data Analyses

Visual responsiveness classification

To assess whether a neuron was visually responsive we used a 1 x 5 sliding window ANOVA with the factor visual category (face, plant, animal, text, object). We counted spikes in an 80-400 ms period relative to stimulus onset, using a bin size of 50 ms and a step size of 5 ms. Beginning at each time point, the average response in each 50 ms trial snippet was computed and the vector of responses, labeled by their stimulus identity, were fed into the ANOVA. The time point was then incremented by 5 ms and the ANOVA was re-computed. A neuron was considered visually responsive if the ANOVA was significant ($p < 0.01$) for 6 consecutive time points. These parameters were chosen to ensure that the probability of selecting a neuron by chance was less than 0.05 (compared to bootstrap distribution with 1000 repetitions, Figure S6A).

Response latency computation

We computed such a single trial onset latency by using a Poisson spike-train analysis for all visually responsive neurons. This method detects points of time in which the observed inter-spike intervals (ISI) deviate significantly from that assumed by a constant-rate Poisson process. This is done by maximizing a Poisson surprise index (*83*). The mean firing rate of the neuron during the inter-trial interval was used to set the baseline rate for the Poisson process. Spikes from a window of 80-300 ms after stimulus onset were included. For a given burst of spikes, if the probability that said burst was produced by a constant-rate Poisson process - where the rate parameters are specified by the baseline firing rate - was less than 0.001, we took the timepoint of the first spike as the onset latency. The response latency of the neuron was taken to be the average latency across all trials.

Building an object space using a deep network

We built a high dimensional object space by feeding our 500 stimulus images into the pre-trained MATLAB implementation of AlexNet (*90*) (Deep Learning Toolbox, command: 'net = alexnet'). The responses of the 4096 nodes in fc6 were extracted to form a 500 x 4096 matrix (using the 'activations' function). PCA was then performed on this matrix yielding 499 PCs, each of length 4096. To reduce the dimensionality of this space we retained only the first 50PCs which captured 80.68% of the response variance across fc6 units (Figure S2B). The first two dimensions accounted for 20.17% of the response variance across fc6 units.

Axis computation

Preferred axis

The preferred axis of each neuron was computed using the spike triggered average (STA). The neural response vector was computed by binning spikes elicited by each stimulus in a 250ms window starting from the response latency of the neuron — necessarily restricting analysis to visually responsive neurons.

Once the neural response vector was computed the STA is defined as:

$$\vec{P}_{sta} = (\vec{r} - \bar{r})^T F \tag{1}$$

where $\vec{r}$ is the n x 1 neural response vector to n objects, $\bar{r}$ is the mean firing rate, and F is an n x d matrix of features with each row corresponding to the features for a given object that is computed via PCA on deep network activations (see above). The projection value of the stimulus objects onto the preferred axis is given by:

$$Proj_{sta} = \left(\frac{\vec{P}_{sta}}{\|\vec{P}_{sta}\|}\right)F^T \tag{2}$$

Principal orthogonal axis

The orthogonal axis seen in all plots is the principal orthogonal axis. This is defined as the axis orthogonal to the preferred axis along which there is the most variation. For each neuron, the preferred axis was computed, the component along the preferred axis ($\vec{P}$) was subsequently subtracted from all object feature vectors in F leaving a matrix of orthogonal feature vectors. Succinctly, for a given feature vector $\vec{f}_d$ in feature space we computed:

$$\vec{f}_{d-1} = \vec{f}_d - \vec{f}_d^{\,T}\vec{P}\frac{\vec{P}}{\|\vec{P}\|^2} \tag{3}$$

Then principal component analysis was performed on this set of n vectors $\vec{f}_{d-1}$, and the first principal component is chosen as the principal orthogonal axis.

Quantifying significance of axis tuning

For each neuron after the preferred axis was computed we examined the correlation between the firing rate response to the stimuli and their projection value along the preferred axis. This correlation value was recomputed after shuffling the features (1000 repetitions) and the original value was compared to this bootstrap distribution. If the original value was greater than 99% of the shuffled values the neuron was considered axis tuned.

Explained variance computation

Axis model

The axis model assumes a linear relationship between the projection value of an incoming object onto the neurons preferred axis and its response. Therefore, to quantify the explained variance for each neuron, we fit a linear regression model between the PCs of the features and the responses of the neuron. A leave-one-out cross validation approach was used i.e. the responses to 499 objects were used to fit the model, and the responses of the neuron to the left-out object was predicted using the same linear transform. In this manner we could produce a predicted response for all images. Note that the computation of variance explained by the category label was done in this manner as well, replacing the PC features with the vector of category labels.

Exemplar model

The exemplar model assumes that each neuron has a maximal response to a specific exemplar in object space and that the response of the neuron to an incoming object decays as a function of the distance from the object to this exemplar. We used a previous implementation of the exemplar model (*92*) in which the response of the neuron which has an exemplar $\vec{e}$ to an incoming object $\vec{x}$ is:

$$f(d) = C_0 + C_1 d + C_2 d^2 + C_3 d^3 \quad (4)$$

where d is the Euclidean distance between the exemplar object and the incoming object:

$$d = \sqrt{\sum_{i=1}^{N} (e_i - x_i)^2} \quad (5)$$

and N is the dimensionality of the object space, which in our analyses was 50. In such an implementation the coefficients of the polynomial C and the features of the exemplar $\vec{e}$ are considered free parameters. They were adjusted iteratively to minimize the error of fit using the MATLAB function *lsqcurvefit*.

To set an upper bound for the explained variance different trials of responses to the stimuli were randomly split into two halves. The Pearson correlation (r) between the average responses from two half-splits across images was calculated and corrected using the Spearman-Brown correction:

$$r' = \frac{2r}{(1+r)} \quad (6)$$

The square of r' was considered the upper bound or explainable variance. The reported results are the ratio of explained to explainable variance.

Decoding analysis

We find that neurons in human VTC are performing linear projection onto specific preferred axis in object space. As such, their responses can be well modeled as a population by the equation:

$$\vec{R} = \left[C\ \vec{C}_o\right] \begin{bmatrix} \vec{F} \\ 1 \end{bmatrix} \tag{7}$$

where $\vec{R}$ is the n x 1 population response vector to a given image (n = number of neurons), C is the n x d weight matrix for different neurons (d = number of dimensions i.e. 50), $\vec{F}$ is the d x 1 vector of object feature values, and $\vec{C}_O$ is the n x 1 offset vector. Thus, the decoding analysis was performed by inverting (7), yielding:

$$\begin{bmatrix} \vec{F} \\ 1 \end{bmatrix} = \left[C\ \vec{C}_o\right]^{+} \vec{R} = K\vec{R} \tag{8}$$

Where C^{+} indicates the Moore-Penrose pseudoinverse, and K is the $d+1$ x n matrix that transforms measured firing rates $\vec{R}$ into predicted features $\vec{F}$. We used the responses of all but one of the objects (500 – 1 = 499) to determine K using the MATLAB function *regress*. These were then substituted into (8) to predict the feature vector of the last object. For the ith image

$$\vec{F}_i = \sum_{d=1}^{N_{dim}} \left(k_{o_d} + \sum_{n=1}^{N_{cells}} r_{i,n} k_{n,d}\right) \tag{9}$$

Where $r_{i,n}$ is the nth element of $\vec{R}$, i.e. the response of neuron n to the image i, $k_{n,d}$ is the (n, d) element of the weight (regression coefficient) matrix, and k_{o_d} is the dth element of the offset vector. Decoding accuracy was quantified by randomly selecting a subset of object images that included the actual feature vector of the decoded object from the total set of 500 and compared their feature vectors to the predicted feature vector of the decoded object by Euclidean distance. If the actual feature vector closest to the predicted feature vector is of the object being decoded ('target') the decoding is considered correct. This procedure is repeated 1000 times for each of the 500 images with a varying number of distractors to get an aggregate measure of decoding accuracy (Figure 3E).

Object 'reconstruction'

To generate images that reflect the features encoded in the neural responses we gathered images from an auxiliary database and passed 15,901 background free images through AlexNet. The images were then projected into the space built by the 500 stimulus objects. None of these ~16k images had been shown to the patients. For each stimulus image the feature vector decoded from the neural activity was compared to the feature vectors of the large stimulus set. The object in the large image set with the smallest Euclidean distance to the decoded feature vector was considered the 'reconstruction' of that stimulus image (*94*).

To account for the fact that the large object set did not contain any images shown to the patients, which sets a limit on how good the reconstruction can be, we computed a 'normalized distance' to quantify the reconstruction accuracy for each object. We defined the normalized reconstruction distance for an image as

$$Normalized\ distance\ = \frac{|V_{recon} - V_{original}|}{|V_{best\ possible\ recon} - V_{original}|} \tag{10}$$

where V_{recon} is the feature vector reconstructed from neuronal responses, $V_{original}$ is the feature vector of the image presented to the patients, and $V_{best\ possible\ recon}$ is the feature vector of the best possible reconstruction (image in the large set with the closest distance to $V_{original}$). A normalized distance of 1 means the decoded image is the best reconstruction possible. The median normalized distance value for our data was 2.256. For a vast majority (482/500, ~96%) of the images the normalized distance was small (< 5, Figure 3C), with the distance of the worst performing image being 10.471, implying that the neural responses captured many of the fine feature details of the original objects.

Generation of synthetic stimuli

The axis model provides a very clear relationship between images and responses for individual neurons. In essence, images with increasing projection values onto a neuron's preferred axis will show increasing firing rates. This implies that if one computes a neuron's preferred axis and then evenly samples points along it and generates images from those points, those images will elicit systematically increasing responses from the neuron. This also implies that if one generates an image from a point further along the axis than any of the stimulus images used to compute the neurons axis, that image will act as a super stimulus and drive the neuron to a higher firing rate than any of the stimulus images.

To test these predictions we ran the screening task in one session, computed the axes for the neurons recorded, sampled points along the preferred and orthogonal axes and fed those vectors back into a pre-trained GAN (*90*) to generate the synthetic stimuli. We then went back to the patient room and re-ran the screening task with the synthetic images added in. The neurons from the first and second sessions were matched (see below) and the responses of the neurons to the synthetic stimuli were recorded.

Computation of predicted responses to synthetic images

Predicted responses to the synthetic images were computed by fitting a linear regression model between the PCs of the features and the responses of the neuron during the first session to the 500 stimulus images. That linear transform was then used to predict the responses to the synthetic images. These predicted responses were then compared to the responses recorded in matching neuron during the second session.

Matching neurons across experiment sessions

Given the nature of recording neurons in a clinical setting wherein you can only record a few neurons at a time it is common practice to run an initial screening task in order to determine the stimuli that drive the neurons being recorded before using those stimuli in other tasks. In such cases it is generally assumed that the neurons recorded a few hours apart are the same but it is important to provide evidence.

One method of verification is to re-run the same screen in a subsequent session and compare the selectivity of the neurons in both sessions. However, if there are multiple neurons having roughly similar selectivity (which was sometimes the case in our data) matching individual neurons is difficult. In order to meet this challenge, we examined multiple features of the neurons recorded in the first and second sessions. Our algorithm would compute the selectivity vector (rank ordered list of stimulus number for the neuron), the waveform, the burst index which is a measure of how many bursts per unit time the neuron discharges (*98*), the computed response latency of the neuron, and whether or not the neuron was axis-tuned (binary variable). The selectivity vectors were compared using cosine distance (MATLAB function *pdist*), and the waveforms by Euclidean distance.

Each axis-tuned neuron in the initial session was compared to all ipsilateral axis-tuned neurons in the subsequent session (same procedure for non-axis tuned neurons). The session two neurons were then rank ordered in every category with first place in a given category giving a session two neuron a score of x where x = number of session two neurons being compared to the one session one neuron, second place receiving a score of $x - 1$ all the way until the last place neuron in a given category receives a score of 1. The scores of all session two neurons were then summed and the algorithm would assign the session two neuron with the max score to be the session one neuron's 'match'. This procedure is then repeated in the reverse direction, i.e. each session two neuron is compared to all session one neurons. Pairs that were bijective were automatically returned as 'matches' and all others were marked out for manual curation. Manual curation was carried out by examining the used metrics in addition to shape of the peri-stimulus time histogram of the category response of all potential neuron match pairs.

Reactivation metric

To assess whether a neuron was active during imagery we used a combination of a 1 x N (N = number of images) sliding window ANOVA and a sliding window ttest during the cued imagery period. We counted spikes in a 0-5 s window relative to imagery onset (i.e. the entire cued imagery period), using a bin size of 1.5 s and a step size of 300 ms. Beginning at each time point, the trial average was computed using spikes in a 1.5 s window and the vector of trials was fed into the ANOVA or ttest. For the ANOVA the trials were labeled by their stimulus identity. The time point was incremented by 300 ms and the ANOVA was re-computed. A neuron was considered active during imagery if either the ANOVA or the ttest was significant ($p < 0.05$) for 6

consecutive bins (or 5 consecutive steps). These parameters were chosen to ensure that the probability of selecting a neuron by chance (p_{ANOVA} + p_{ttest}) was less than 0.05 (compared to bootstrap distribution with 1000 repetitions, Figure S6B&C).

Correlation of viewed and imagined responses

For neurons that were active during imagery we computed the correlation between viewed and imagined responses. To compute the viewed response to each stimulus, we collected spikes in a 1 s window of the encoding period starting from the response latency of the neuron (computed during screening) and averaged across repetitions of a given stimulus. To compute the imagined response for all neurons active during imagery, we collected spikes in a 2 s window starting from the first significant time bin (1 x N sliding window ANOVA or sliding window ttest, N = number of stimuli, $p < 0.05$) and averaged across repetitions of a given stimulus. The Spearman rank correlation (r) was then computed between these 2 vectors.

Acknowledgements: We thank the staff of the epilepsy monitoring unit (neuromonitoring staff, nursing staff, and physicians) and of the Biomedical Imaging Research Institute at Cedars-Sinai Medical Center for patient care and support. We thank Emily Choe for help implementing the GAN used in Figure 3. We thank members of the Tsao and Rutishauser labs namely Janis Hesse, Hristos Courellis, and Francesco Lanfranchi for helpful comments throughout all stages of this project. We thank the patients for all their patience and perseverance.

Funding: This work was funded by the BRAIN initiative through the NIH Office of the Director (U01NS117839), the Howard Hughes Medical Institute, the Simons Foundation Collaboration on the Global Brain, and the Chen Center for Systems Neuroscience at Caltech.

Author contributions: V.W., U.R., and D.Y.T designed the study. V.W. collected and analyzed the data. V.W., U.R., and D.Y.T. wrote the paper with input from C.M.R, J.M.C., L.M.B., and A. N. M. C.M.R, J.M.C., and L.M.B. provided patient care and facilitated experiments. A. N. M. performed surgery.

Competing interests: Authors declare no competing interests.

Data and materials availability: Data and code will be made publicly available upon acceptance.

References

1. A. Schlegel, P. Alexander, S. V. Fogelson, X. Li, Z. Lu, P. J. Kohler, E. Riley, P. U. Tse, M. Meng, The artist emerges: visual art learning alters neural structure and function. *Neuroimage* **105**, 440–451 (2015).

2. J. F. Christensen, A. Gomila, Introduction: Art and the brain: From pleasure to well-being. *Prog. Brain Res.* **237**, xxvii–xlvi (2018).

3. M. Ellamil, C. Dobson, M. Beeman, K. Christoff, Evaluative and generative modes of thought during the creative process. *Neuroimage* **59**, 1783–1794 (2012).

4. S. S. H. Wong, S. W. H. Lim, Mental imagery boosts music compositional creativity. *PLoS One* **12**, e0174009 (2017).

5. L. Taruffi, M. B. Küssner, A review of music-evoked visual mental imagery: Conceptual issues, relation to emotion, and functional outcome. *Psychomusicology: Music, Mind, and Brain* **29**, 62–74 (2019).

6. S. Liu, H. M. Chow, Y. Xu, M. G. Erkkinen, K. E. Swett, M. W. Eagle, D. A. Rizik-Baer, A. R. Braun, Neural correlates of lyrical improvisation: an FMRI study of freestyle rap. *Sci. Rep.* **2**, 834 (2012).

7. P. E. Keller, Mental imagery in music performance: underlying mechanisms and potential benefits. *Ann. N. Y. Acad. Sci.* **1252**, 206–213 (2012).

8. K. Erhard, F. Kessler, N. Neumann, H.-J. Ortheil, M. Lotze, Professional training in creative writing is associated with enhanced fronto-striatal activity in a literary text continuation task. *Neuroimage* **100**, 15–23 (2014).

9. C. Shah, K. Erhard, H.-J. Ortheil, E. Kaza, C. Kessler, M. Lotze, Neural correlates of creative writing: an fMRI study. *Hum. Brain Mapp.* **34**, 1088–1101 (2013).

10. S. Liu, M. G. Erkkinen, M. L. Healey, Y. Xu, K. E. Swett, H. M. Chow, A. R. Braun, Brain activity and connectivity during poetry composition: Toward a multidimensional model of the creative process. *Hum. Brain Mapp.* **36**, 3351–3372 (2015).

11. D. L. Schacter, D. R. Addis, R. L. Buckner, Episodic simulation of future events: concepts, data, and applications. *Ann. N. Y. Acad. Sci.* **1124**, 39–60 (2008).

12. S. T. Moulton, S. M. Kosslyn, Imagining predictions: mental imagery as mental emulation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364**, 1273–1280 (2009).

13. C. Guariglia, L. Pizzamiglio, “The Role of Imagery in Navigation: Neuropsychological Evidence” in *Spatial Processing in Navigation, Imagery and Perception*, F. Mast, L. Jäncke, Eds. (Springer US, Boston, MA, 2007), pp. 17–28.

14. C. M. Bird, J. A. Bisby, N. Burgess, The hippocampus and spatial constraints on mental imagery. *Front. Hum. Neurosci.* **6**, 142 (2012).

15. A. Bocchi, M. Carrieri, S. Lancia, V. Quaresima, L. Piccardi, The Key of the Maze: The role of mental imagery and cognitive flexibility in navigational planning. *Neurosci. Lett.* **651**, 146–150 (2017).

16. E. Tulving, Others, Episodic and semantic memory. *Organization of memory* **1**, 1 (1972).

17. D. R. Addis, A. T. Wong, D. L. Schacter, Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia* **45**, 1363–1377 (2007).

18. D. L. Schacter, D. R. Addis, R. L. Buckner, Remembering the past to imagine the future: the prospective brain. *Nat. Rev. Neurosci.* **8**, 657–661 (2007).

19. H. Gelbard-Sagiv, R. Mukamel, M. Harel, R. Malach, I. Fried, Internally generated reactivation of single neurons in human hippocampus during free recall. *Science* **322**, 96–101 (2008).

20. A. P. Vaz, J. H. Wittig Jr, S. K. Inati, K. A. Zaghloul, Replay of cortical spiking sequences during human memory retrieval. *Science* **367**, 1131–1134 (2020).

21. A. Mary, J. Dayan, G. Leone, C. Postel, F. Fraisse, C. Malle, T. Vallée, C. Klein-Peschanski, F. Viader, V. de la Sayette, D. Peschanski, F. Eustache, P. Gagnepain, Resilience after trauma: The role of memory suppression. *Science* **367** (2020).

22. T. D. Griffiths, Musical hallucinosis in acquired deafness. Phenomenology and brain substrate. *Brain* **123 (Pt 10)**, 2065–2076 (2000).

23. S. Kumar, W. Sedley, G. R. Barnes, S. Teki, K. J. Friston, T. D. Griffiths, Neural bases of musical hallucinations. *J. Neurol. Neurosurg. Psychiatry* **85**, e3–e3 (2014).

24. R. J. Zatorre, A. R. Halpern, D. W. Perry, E. Meyer, A. C. Evans, Hearing in the mind’s ear: A PET investigation of musical imagery and perception. *J. Cogn. Neurosci.* **8**, 29–46 (1996).

25. A. R. Halpern, R. J. Zatorre, When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. *Cereb. Cortex* **9**, 697–704 (1999).

26. R. J. Stevenson, T. I. Case, Olfactory imagery: a review. *Psychon. Bull. Rev.* **12**, 244–264 (2005).

27. J. A. Gottfried, A. P. R. Smith, M. D. Rugg, R. J. Dolan, Remembrance of odors past. *Neuron* **42**, 687–695 (2004).

28. S.-S. Yoo, D. K. Freeman, J. J. McCarthy III, F. A. Jolesz, Neural substrates of tactile imagery: a functional MRI study. *Neuroreport* **14**, 581–585 (2003).

29. X. Tian, J. M. Zarate, D. Poeppel, Mental imagery of speech implicates two mechanisms of perceptual reactivation. *Cortex* **77**, 1–12 (2016).

30. W. Richter, R. Somorjai, R. Summers, M. Jarmasz, R. S. Menon, J. S. Gati, A. P. Georgopoulos, C. Tegeler, K. Ugurbil, S. G. Kim, Motor area activity during mental rotation studied by time-resolved single-trial fMRI. *J. Cogn. Neurosci.* **12**, 310–320 (2000).

31. R. N. Shepard, J. Metzler, Mental rotation of three-dimensional objects. *Science* **171**, 701–703 (1971).

32. L. M. Parsons, P. T. Fox, J. H. Downs, T. Glass, T. B. Hirsch, C. C. Martin, P. A. Jerabek, J. L. Lancaster, Use of implicit motor imagery for visual shape discrimination as revealed by PET. *Nature* **375**, 54–58 (1995).

33. J. Decety, The neurophysiological basis of motor imagery. *Behav. Brain Res.* **77**, 45–52 (1996).

34. K. J. Miller, G. Schalk, E. E. Fetz, M. den Nijs, J. G. Ojemann, R. P. N. Rao, Cortical activity during motor execution, motor imagery, and imagery-based online feedback. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 4430–4435 (2010).

35. S. M. Kosslyn, A. Pascual-Leone, O. Felician, S. Camposano, J. P. Keenan, W. L. Thompson, G. Ganis, K. E. Sukel, N. M. Alpert, The role of area 17 in visual imagery: convergent evidence from PET and rTMS. *Science* **284**, 167–170 (1999).

36. A. Spagna, D. Hajhajate, J. Liu, P. Bartolomeo, Visual mental imagery engages the left fusiform gyrus, but not the early visual cortex: A meta-analysis of neuroimaging evidence. *Neurosci. Biobehav. Rev.* **122**, 201–217 (2021).

37. J. Pearson, T. Naselaris, E. A. Holmes, S. M. Kosslyn, Mental Imagery: Functional Mechanisms and Clinical Applications. *Trends Cogn. Sci.* **19**, 590–602 (2015).

38. M. B. Bone, M. St-Laurent, C. Dang, D. A. McQuiggan, J. D. Ryan, B. R. Buchsbaum, Eye Movement Reinstatement and Neural Reactivation During Mental Imagery. *Cereb. Cortex* **29**, 1075–1089 (2019).

39. S. D. Slotnick, W. L. Thompson, S. M. Kosslyn, Visual mental imagery induces retinotopically organized activation of early visual areas. *Cereb. Cortex* **15**, 1570–1583 (2005).

40. S.-H. Lee, D. J. Kravitz, C. I. Baker, Disentangling visual imagery and perception of real-world objects. *Neuroimage* **59**, 4064–4073 (2012).

41. A. Ishai, J. V. Haxby, L. G. Ungerleider, Visual imagery of famous faces: effects of memory and attention revealed by fMRI. *Neuroimage* **17**, 1729–1741 (2002).

42. N. Dijkstra, S. E. Bosch, M. A. J. van Gerven, Shared Neural Mechanisms of Visual Perception and Imagery. *Trends Cogn. Sci.* **23**, 423–434 (2019).

43. A. Ishai, D. Sagi, Common mechanisms of visual imagery and perception. *Science* **268**, 1772–1774 (1995).

44. C. Hofstetter, A. Achaibou, P. Vuilleumier, Reactivation of visual cortex during memory retrieval: content specificity and emotional modulation. *Neuroimage* **60**, 1734–1745 (2012).

45. K. M. O'Craven, N. Kanwisher, Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *J. Cogn. Neurosci.* **12**, 1013–1023 (2000).

46. J. Pearson, The human imagination: the cognitive neuroscience of visual mental imagery. *Nat. Rev. Neurosci.* **20**, 624–634 (2019).

47. L. G. Ungerleider, J. V. Haxby, "What" and "where" in the human brain. *Curr. Opin. Neurobiol.* **4**, 157–165 (1994).

48. M. J. Farah, D. N. Levine, R. Calvanio, A case study of mental imagery deficit. *Brain Cogn.* **8**, 147–164 (1988).

49. M. J. Riddoch, Loss of visual imagery: A generation deficit. *Cogn. Neuropsychol.* **7**, 249–273 (1990).

50. S. Thorudottir, H. M. Sigurdardottir, G. E. Rice, S. J. Kerry, R. J. Robotham, A. P. Leff, R. Starrfelt, The architect who lost the ability to imagine: The cerebral basis of visual imagery. *Brain Sci.* **10**, 59 (2020).

51. R. P. Rao, D. H. Ballard, Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **2**, 79–87 (1999).

52. Y. Huang, J. Gornet, S. Dai, Z. Yu, T. Nguyen, D. Y. Tsao, A. Anandkumar, Neural networks with recurrent generative feedback, *arXiv [cs.LG]* (2020). http://arxiv.org/abs/2007.09200.

53. P. Cavanagh, G. P. Caplovitz, T. K. Lytchenko, M. Maechler, P. U. Tse, D. Sheinberg, Object-Based Attention, *PsyArXiv* (2022). https://doi.org/10.31234/osf.io/2bsn7.

54. C. G. Gross, C. E. Rocha-Miranda, D. B. Bender, Visual properties of neurons in inferotemporal cortex of the Macaque. *J. Neurophysiol.* **35**, 96–111 (1972).

55. R. Desimone, T. D. Albright, C. G. Gross, C. Bruce, Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci.* **4**, 2051–2062 (1984).

56. I. Fried, C. L. Wilson, N. T. Maidment, J. Engel Jr, E. Behnke, T. A. Fields, K. A. MacDonald, J. W. Morrow, L. Ackerson, Cerebral microdialysis combined with single-neuron and

electroencephalographic recording in neurosurgical patients. Technical note. *J. Neurosurg.* **91**, 697–705 (1999).

57. L. Chang, D. Y. Tsao, The Code for Facial Identity in the Primate Brain. *Cell* **169**, 1013-1028.e14 (2017).

58. P. Bao, L. She, M. McGill, D. Y. Tsao, A map of object space in primate inferotemporal cortex. *Nature* **583**, 103–108 (2020).

59. Y. Shi, D. Bi, J. K. Hesse, F. F. Lanfranchi, S. Chen, D. Y. Tsao, Rapid, concerted switching of the neural code in inferotemporal cortex. *bioRxivorg*, 2023.12. 06.570341 (2023).

60. L. She, M. K. Benna, Y. Shi, S. Fusi, D. Y. Tsao, Temporal multiplexing of perception and memory codes in IT cortex. *Nature* **629**, 861–868 (2024).

61. V. Axelrod, C. Rozier, T. S. Malkinson, K. Lehongre, C. Adam, V. Lambrecq, V. Navarro, L. Naccache, Face-selective neurons in the vicinity of the human fusiform face area. *Neurology* **92**, 197–198 (2019).

62. S. Khuvis, E. M. Yeagle, Y. Norman, S. Grossman, R. Malach, A. D. Mehta, Face-Selective Units in Human Ventral Temporal Cortex Reactivate during Free Recall. *J. Neurosci.* **41**, 3386–3399 (2021).

63. V. Axelrod, C. Rozier, T. S. Malkinson, K. Lehongre, C. Adam, V. Lambrecq, V. Navarro, L. Naccache, Face-selective multi-unit activity in the proximity of the FFA modulated by facial expression stimuli. *Neuropsychologia* **170**, 108228 (2022).

64. R. Quian Quiroga, M. Boscaglia, J. Jonas, H. G. Rey, X. Yan, L. Maillard, S. Colnat-Coulbois, L. Koessler, B. Rossion, Single neuron responses underlying face recognition in the human midfusiform face-selective cortex. *Nat. Commun.* **14**, 5661 (2023).

65. H. S. Courellis, J. Minxha, A. R. Cardenas, D. L. Kimmel, C. M. Reed, T. A. Valiante, C. D. Salzman, A. N. Mamelak, S. Fusi, U. Rutishauser, Abstract representations emerge in human hippocampal neurons during inference. *Nature* **632**, 841–849 (2024).

66. A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25** (2012).

67. D. P. Hanes, K. G. Thompson, J. D. Schall, Relationship of presaccadic activity in frontal eye field and supplementary eye field to saccade initiation in macaque: Poisson spike train analysis. *Exp. Brain Res.* **103**, 85–96 (1995).

68. D. L. K. Yamins, H. Hong, C. F. Cadieu, E. A. Solomon, D. Seibert, J. J. DiCarlo, Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8619–8624 (2014).

69. L. She, M. K. Benna, Y. Shi, S. Fusi, D. Y. Tsao, The neural code for face memory. *BioRxiv* (2021).

70. T. Valentine, A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Q. J. Exp. Psychol. A* **43**, 161–204 (1991).

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

72. D. Y. Tsao, W. A. Freiwald, What's so special about the average face? *Trends Cogn. Sci.* **10**, 391–393 (2006).

73. Y. Taigman, M. Yang, M. Ranzato, L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification" in *2014 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2014), pp. 1701–1708.

74. K. N. Kay, T. Naselaris, R. J. Prenger, J. L. Gallant, Identifying natural images from human brain activity. *Nature* **452**, 352–355 (2008).

75. A. S. Cowen, M. M. Chun, B. A. Kuhl, Neural portraits of perception: reconstructing face images from evoked brain activity. *Neuroimage* **94**, 12–22 (2014).

76. A. Nestor, D. C. Plaut, M. Behrmann, Feature-based face representations and image reconstruction from behavioral and neural data. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 416–421 (2016).

77. A. Dosovitskiy, T. Brox, Generating images with perceptual similarity metrics based on deep networks. *Adv. Neural Inf. Process. Syst.* **29** (2016).

78. C. R. Ponce, W. Xiao, P. F. Schade, T. S. Hartmann, G. Kreiman, M. S. Livingstone, Evolving Images for Visual Neurons Using a Deep Generative Network Reveals Coding Principles and Neuronal Preferences. *Cell* **177**, 999-1009.e10 (2019).

79. J. Kubilius, M. Schrimpf, A. Nayebi, D. Bear, D. L. K. Yamins, J. J. DiCarlo, CORnet: Modeling the Neural Mechanisms of Core Object Recognition, *bioRxiv* (2018)p. 408385.

80. G. Kreiman, C. Koch, I. Fried, Imagery neurons in the human brain. *Nature* **408**, 357–361 (2000).

81. J. Daume, J. Kaminski, A. G. P. Schjetnan, Y. Salimpour, U. Khan, C. Reed, W. Anderson, T. A. Valiante, A. N. Mamelak, U. Rutishauser, Control of working memory maintenance by theta-gamma phase amplitude coupling of human hippocampal neurons. *bioRxivorg*, doi: 10.1101/2023.04.05.535772 (2023).

82. J. Kamiński, S. Sullivan, J. M. Chung, I. B. Ross, A. N. Mamelak, U. Rutishauser, Persistently active neurons in human medial frontal and medial temporal lobe support working memory. *Nat. Neurosci.* **20**, 590–601 (2017).

83. J. Kamiński, A. Brzezicka, A. N. Mamelak, U. Rutishauser, Combined phase-rate coding by persistently active neurons as a mechanism for maintaining multiple items in working memory in humans. *Neuron* **106**, 256-264.e3 (2020).

84. E. Svoboda, M. C. McKinnon, B. Levine, The functional neuroanatomy of autobiographical memory: a meta-analysis. *Neuropsychologia* **44**, 2189–2208 (2006).

85. U. Rutishauser, I. B. Ross, A. N. Mamelak, E. M. Schuman, Human memory strength is predicted by theta-frequency phase-locking of single neurons. *Nature* **464**, 903–907 (2010).

86. W. B. Scoville, B. Milner, Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* **20**, 11–21 (1957).

87. H. Tomita, M. Ohbayashi, K. Nakahara, I. Hasegawa, Y. Miyashita, Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature* **401**, 699–703 (1999).

88. D. J. Felleman, D. C. Van Essen, Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47 (1991).

89. P. Grimaldi, K. S. Saleem, D. Tsao, Anatomical connections of the functionally defined "face patches" in the macaque monkey. *Neuron* **90**, 1325–1342 (2016).

90. M. Cerf, N. Thiruvengadam, F. Mormann, A. Kraskov, R. Q. Quiroga, C. Koch, I. Fried, On-line, voluntary control of human temporal lobe neurons. *Nature* **467**, 1104–1108 (2010).

91. B. Thirion, E. Duchesnay, E. Hubbard, J. Dubois, J.-B. Poline, D. Lebihan, S. Dehaene, Inverse retinotopy: inferring the visual content of images from brain activation patterns. *Neuroimage* **33**, 1104–1116 (2006).

92. T. Naselaris, C. A. Olman, D. E. Stansbury, K. Ugurbil, J. L. Gallant, A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *Neuroimage* **105**, 215–228 (2015).

93. T. S. Lee, D. Mumford, Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **20**, 1434–1448 (2003).

94. A. Yuille, D. Kersten, Vision as Bayesian inference: analysis by synthesis? *Trends Cogn. Sci.* **10**, 301–308 (2006).

95. B. Peters, J. J. DiCarlo, T. Gureckis, R. Haefner, L. Isik, J. Tenenbaum, T. Konkle, T. Naselaris, K. Stachenfeld, Z. Tavares, D. Tsao, I. Yildirim, N. Kriegeskorte, How does the primate brain combine generative and discriminative computations in vision?, *arXiv [q-bio.NC]* (2024). http://arxiv.org/abs/2401.06005.

96. Y. Ma, D. Tsao, H.-Y. Shum, On the principles of Parsimony and Self-consistency for the emergence of intelligence, *arXiv [cs.AI]* (2022). http://arxiv.org/abs/2207.04630.

97. J. Minxha, A. N. Mamelak, U. Rutishauser, Surgical and electrophysiological techniques for single-neuron recordings in human epilepsy patients. *Extracellular recording* (2018).

98. U. Rutishauser, E. M. Schuman, A. N. Mamelak, Online detection and sorting of extracellularly recorded action potentials in human medial temporal lobe recordings, in vivo. *J. Neurosci. Methods* **154**, 204–224 (2006).

99. M. Reuter, H. D. Rosas, B. Fischl, Highly accurate inverse consistent registration: a robust approach. *Neuroimage* **53**, 1181–1196 (2010).

100. W. M. Pauli, A. N. Nili, J. M. Tyszka, A high-resolution probabilistic in vivo atlas of human subcortical brain nuclei. *Sci Data* **5**, 180063 (2018).

101. B. Avants, J. T. Duda, J. Kim, H. Zhang, J. Pluta, J. C. Gee, J. Whyte, Multivariate analysis of structural and diffusion imaging in traumatic brain injury. *Acad. Radiol.* **15**, 1360–1375 (2008).

102. A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, I. Sutskever, Zero-shot text-to-image generation. *ICML* **abs/2102.12092**, 8821–8831 (2021).

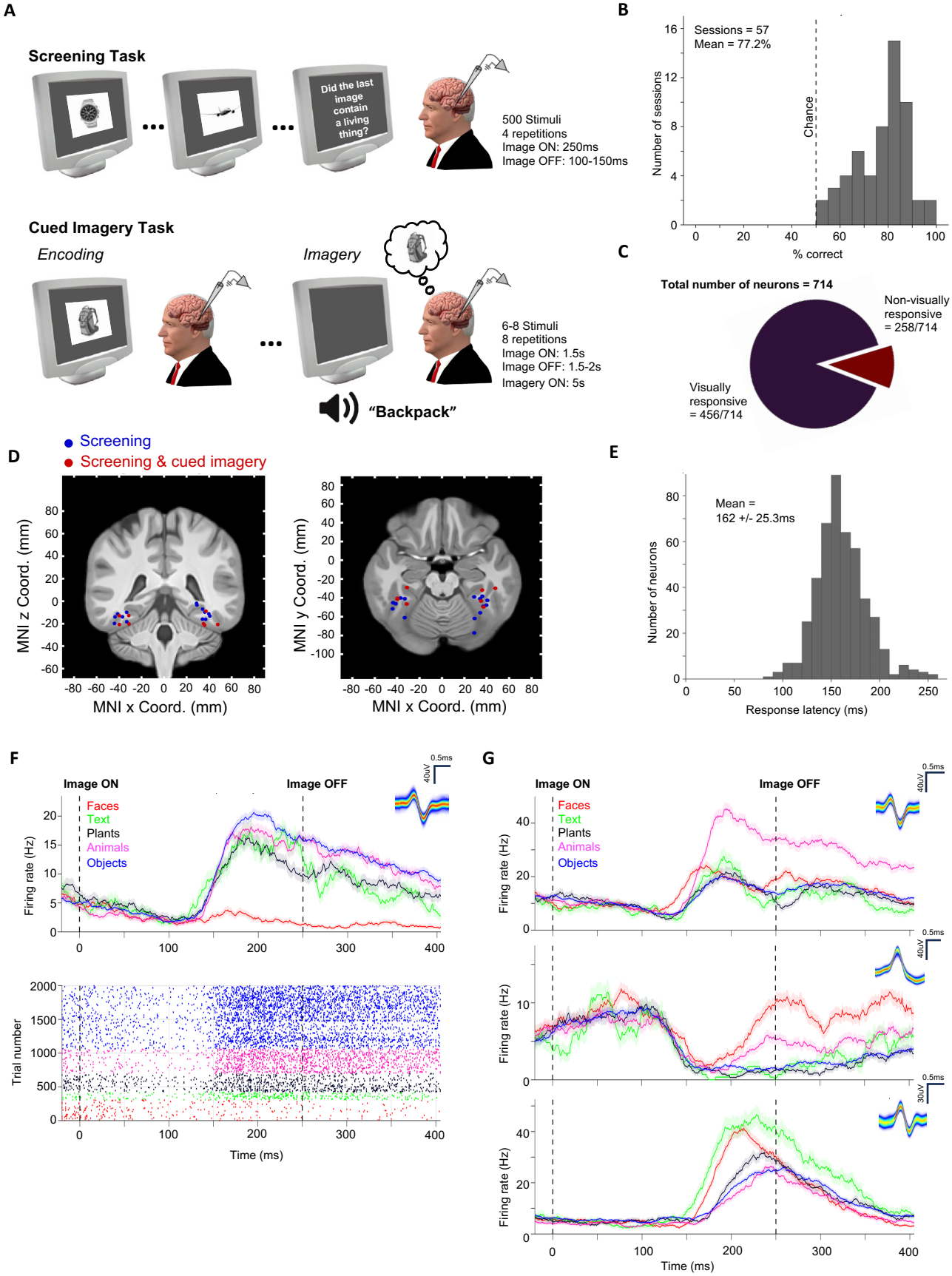
A
Screening Task
Did the last image contain a living thing?
500 Stimuli
4 repetitions
Image ON: 250ms
Image OFF: 100-150ms
Cued Imagery Task
Encoding
Imagery
6-8 Stimuli
8 repetitions
Image ON: 1.5s
Image OFF: 1.5-2s
Imagery ON: 5s
"Backpack"
B
Sessions = 57
Mean = 77.2%
Chance
Number of sessions
% correct
C
Total number of neurons = 714
Non-visually responsive = 258/714
Visually responsive = 456/714
D
Screening
Screening & cued imagery
MNI z Coord. (mm)
MNI x Coord. (mm)
MNI y Coord. (mm)
MNI x Coord. (mm)
E
Mean = 162 +/- 25.3ms
Number of neurons
Response latency (ms)
F
Image ON
Image OFF
Faces
Text
Plants
Animals
Objects
Firing rate (Hz)
Trial number
Time (ms)
G
Image ON
Image OFF
Faces
Text
Plants
Animals
Objects
Firing rate (Hz)
Time (ms)

Figure 1. Selective visual responses in human VTC.

(A) Task schematics. Patients performed two tasks: a screening task (top) and a cued imagery task (bottom). In the screening task, grayscale images with white backgrounds were displayed on a gray screen for 250 ms with the inter-trial interval jittered between 100-150 ms. Images subtended 6-7 visual degrees. At random intervals (min interval: 1 trial, max interval: 80 trials) a yes-no catch question would appear pertaining to the image that came just before it. **(B)** Accuracy of catch question responses for all sessions. On average patients answered catch questions correctly in 77 ± 11% of the trials (± s.d.) despite the rapid stimulus presentation, implying the stimuli were closely attended to. **(C)** 456 of the 714 recorded neurons were visually responsive. **(D)** Recording locations of the 27 microwire bundles (left and right) that contained at least one well isolated neuron in human VTC across all 16 patients in Coronal (left) and Axial (right) views. Montreal Neurological Institute coordinates can be read off the image or seen in Table S1. Each dot represents the location of one microwire bundle (8 channels). The locations marked in red were sessions used in a subsequent cued imagery task. **(E)** Distribution of response latencies for all 456 visually responsive VTC neurons. The mean response latency was 162 ± 25 ms (± s.d.). **(F)** An example neuron recorded during the screening task. This neuron was responsive to all categories except faces. The estimated response latency is 178 ms. Stimulus onset is at t = 0 and offset is at t = 250 ms. The inset shows the mean waveform of the neuron. **(G)** Further example neurons illustrating the diversity of response profiles. (Top) A strong category selective neuron, showing a lower response to any of the non-preferred categories. (Middle) A response profile characterized by an initial suppression of activity. (Bottom) A neuron that distinguishes its preferred from non-preferred category using a latency code along with a rate code.

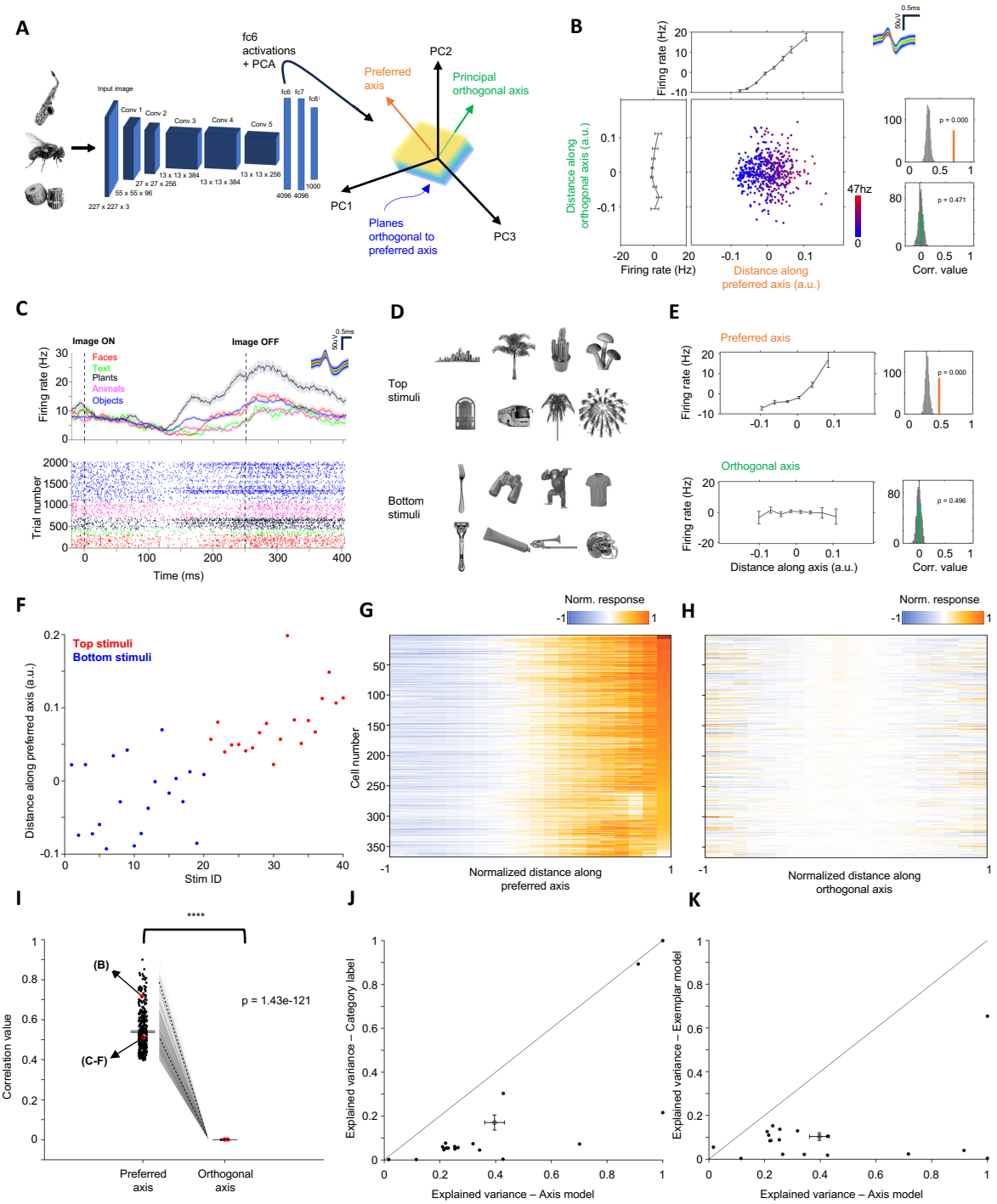
A
fc6 activations + PCA
Input image
Conv 1
Conv 2
Conv 3
Conv 4
Conv 5
fc6
fc7
fc8
227 x 227 x 3
55 x 55 x 96
27 x 27 x 256
13 x 13 x 384
13 x 13 x 384
13 x 13 x 256
4096
4096
1000
PC1
PC2
PC3
Preferred axis
Principal orthogonal axis
Planes orthogonal to preferred axis
B
Firing rate (Hz)
Distance along orthogonal axis (a.u.)
Distance along preferred axis (a.u.)
47hz
p = 0.000
p = 0.471
Corr. value
C
Image ON
Image OFF
Faces
Text
Plants
Animals
Objects
Firing rate (Hz)
Trial number
Time (ms)
D
Top stimuli
Bottom stimuli
E
Preferred axis
Orthogonal axis
Firing rate (Hz)
Distance along axis (a.u.)
p = 0.000
p = 0.496
Corr. value
F
Top stimuli
Bottom stimuli
Distance along preferred axis (a.u.)
Stim ID
G
Norm. response
Cell number
Normalized distance along preferred axis
H
Norm. response
Normalized distance along orthogonal axis
I
****
(B)
(C-F)
p = 1.43e-121
Correlation value
Preferred axis
Orthogonal axis
J
Explained variance – Category label
Explained variance – Axis model
K
Explained variance – Exemplar model
Explained variance – Axis model

Figure 2. Axis tuning in human VTC neurons.

(A) Schematic of the stimulus parametrization and axis computation procedure. Individual stimuli were parametrized as points in a 50-dimensional feature space that was built by performing PCA on the unit activations of AlexNet's fc6 layer and keeping the top 50 PCs. Neurons' preferred axes were computed in this space by projecting the features onto the mean subtracted responses leading to a large null space for each neuron. **(B)** Example axis-tuned neuron. Scatter shows the responses to the 500 stimulus images projected onto the neuron's preferred axis and principal (longest) orthogonal axis in the face feature space. Response magnitude is color-coded. (Top) Mean response as a function of distance along the preferred axis. (Left) Mean response as a function of distance along the orthogonal axis. (Top distribution) The correlation value between projection value and firing rate for the preferred axis (orange) showing a significant correlation ($p = 0.001$, bootstrap distribution 1000 repetitions). (Bottom distribution) Correlation between the projection value and firing along the orthogonal axis (green) is not significantly different from the null distribution for the orthogonal axis ($p = 0.496$, bootstrap distribution 1000 repetitions). **(C-F)** The axis model explains complex neural tuning that does not follow pre-defined categorical boundaries. **(C)** An example neuron is shown. Peri-stimulus time histogram (PSTH) and raster of the example neuron. A close look at the raster reveals robust responses to a few out of "preferred" category objects. **(D)** The top (most preferred) and bottom (least preferred) stimuli for the neuron shown in (C), no semantic category cleanly delineates between them. **(E)** Axis tuning for the neuron shown in (C-D). See (B) for notation. **(F)** Projection values of the stimuli shown in (D) for the neuron shown in (C-D) reveals a systematic relationship between the projection value and stimulus preference. **(G-J)** Population summary of the tuning to the preferred (G) and orthogonal (H) axes for all visually responsive neurons that were significantly axis tuned (367/456). Each row is a neuron and shows the increase in normalized response as a function of distance along the preferred axis. **(H)** Corresponding plot for the principal orthogonal axis (orthogonal axis capturing the most variation) showing no systematic change in response as a function of distance. **(I)** Pearson correlation between the projection value of the stimulus images onto the preferred and orthogonal axes and the firing rate response of the neuron ($n=367$). The neurons shown in (B) and (C-F) are marked in red. **(J)** Comparison of fit quality for axis model and category label for 18 neurons. These 18 neurons were chosen as they had significant variance explained in their responses across all 3 models. The axis model provides significantly better fits to actual responses than the category label. (39.62% axis, 17.04% category, $p = 9.42e-04$, Wilcoxon ranksum test). The error bars indicate standard error of the mean. **(K)** Comparison of fit quality for axis and exemplar models for the same 18 neurons. The axis model provides significantly better fits to actual responses than the exemplar model (39.62% axis, 10.37% exemplar, $p = 4.78e-05$, Wilcoxon ranksum test). The error bars indicate standard error of the mean.

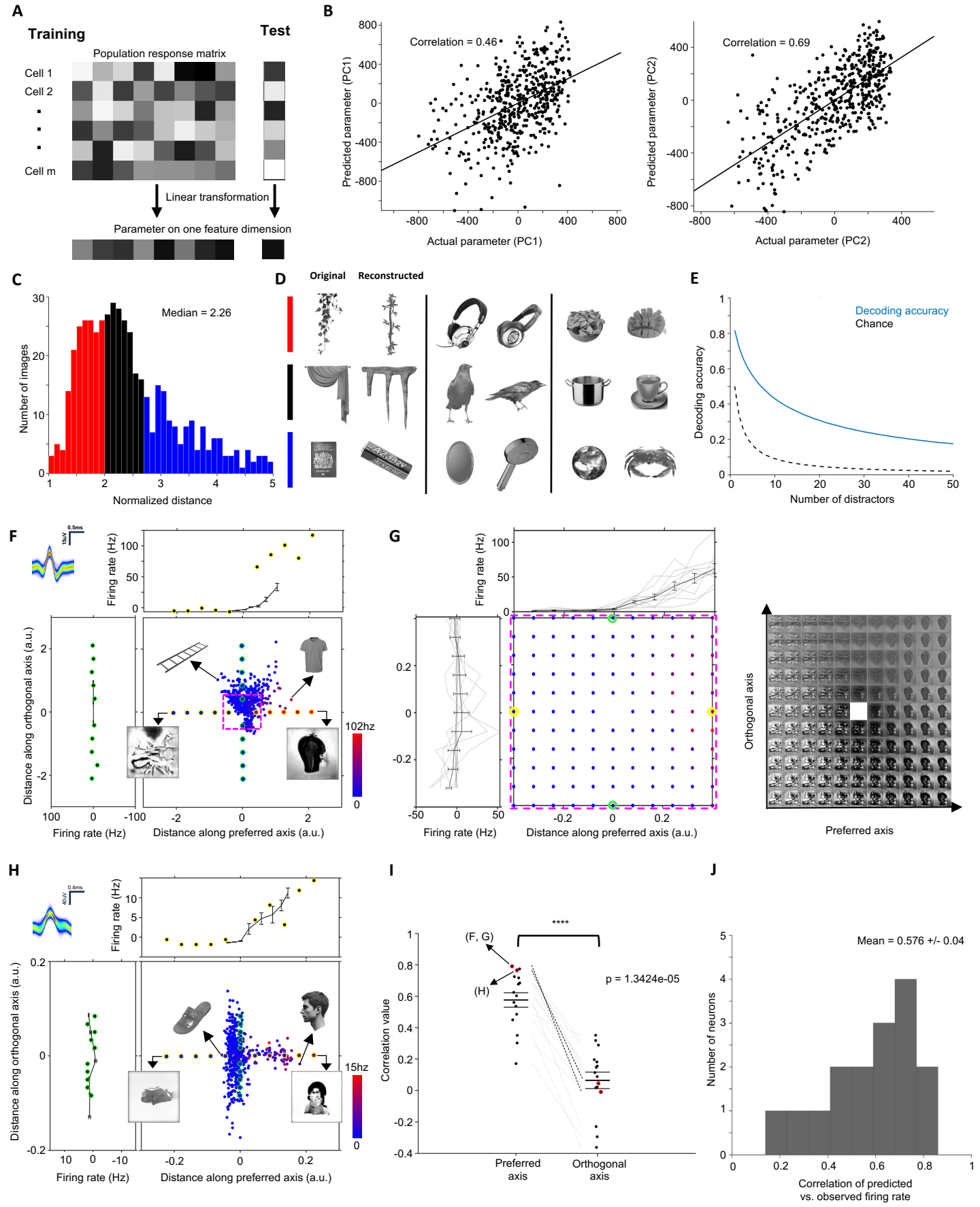
A
Training
Test
Population response matrix
Cell 1
Cell 2
Cell m
Linear transformation
Parameter on one feature dimension
B
Correlation = 0.46
Predicted parameter (PC1)
Actual parameter (PC1)
Correlation = 0.69
Predicted parameter (PC2)
Actual parameter (PC2)
C
Median = 2.26
Number of images
Normalized distance
D
Original
Reconstructed
E
Decoding accuracy
Chance
Decoding accuracy
Number of distractors
F
Firing rate (Hz)
Distance along orthogonal axis (a.u.)
Distance along preferred axis (a.u.)
102hz
G
Firing rate (Hz)
Distance along preferred axis (a.u.)
Orthogonal axis
Preferred axis
H
Firing rate (Hz)
Distance along orthogonal axis (a.u.)
Distance along preferred axis (a.u.)
15hz
I
(F, G)
(H)
****
p = 1.3424e-05
Correlation value
Preferred axis
Orthogonal axis
J
Mean = 0.576 +/- 0.04
Number of neurons
Correlation of predicted vs. observed firing rate

Figure 3. Decoding object features via linear regression and generating super stimuli for VTC neurons using the axis model and a GAN.

[Note: The human face in panel H of this figure has been replaced with a synthetic face generated by a diffusion model (*102*), in accordance with the bioarxiv policy on displaying human faces.]

(A-E) Reconstruction of objects from neural responses. **(A)** Decoding model. We used responses to all but one object (500-1 = 499) to determine the transformation between responses and feature values by linear regression, and then used that transformation to predict the feature values of the held-out object. **(B)** Decoding accuracy. Predicted vs. actual feature values for the first (left, p = 1.04e-25, paired t-test) and second (right, p = 0, paired t-test) dimensions of object space (see methods). **(C)** Distribution of normalized distances between reconstructed feature vectors and the best-possible reconstructed feature vectors for 482/500 images (see methods) to quantify decoding accuracy across the population. The normalized distance takes into account the fact that the object images used for reconstruction did not include any of the object images shown to the patients. A normalized distance of 1 means the reconstruction found the best solution possible. **(D)** Images were split into tertiles based on normalized distance. Examples of the reconstructions in the first tertile (top row), second (middle row) and third (bottom row) as compared to the original stimulus images being reconstructed. **(E)** Decoding accuracy as a function of the number of distractor objects drawn randomly from the stimulus set (see methods). The black dashed line represents the decoding accuracy one would expect by chance. **(F)** Axis plot of an example neuron showing the positions of generated stimuli sampled along the preferred axis (yellow) and orthogonal axis (green) relative to the other stimulus images. The maximum and minimum projection valued images are displayed along the preferred axis. The vertical and horizontal line plots are the binned firing rate of the stimulus images as one moves along each axis with the generated images overlaid on top. The systematic increase along the preferred axis with an almost identical response to all images along the orthogonal axis is clearly visible. **(G)** (left) The responses of the neuron in (F) to a grid of images sampled in the space spanned by the preferred and orthogonal axes. The extent of the grid is indicated by the purple bounding box in (F). (right) The stimuli that comprised the grid. The neuron showed increases in firing rate to stimuli that varied in directions parallel to the preferred axis. **(H)** Another example neuron with responses to both screening and synthetic stimuli (see (F) for notation). **(I)** Population summary. For each of the 16 neurons across 4 patients, the Pearson Correlation between the projection value of the generated images and the firing rate response of the neuron is visualized for the preferred and orthogonal axes respectively. The two example neurons shown in (F-G) and (H) are marked in red. **(J)** Distribution of the correlation values between the predicted and observed firing rate responses to the synthetic images. For each neuron, the preferred axis was computed using the 500 stimulus images and used to learn the transformation between projection values and firing rates (as in A, see methods). This transformation was used to compute predicted responses of the neuron to the synthetic images. The distribution shown is the correlation of those predicted values (using the axis of the neuron in the first session) to the responses of the matched neuron in the second session observed during the experiment (see methods for matching procedure).

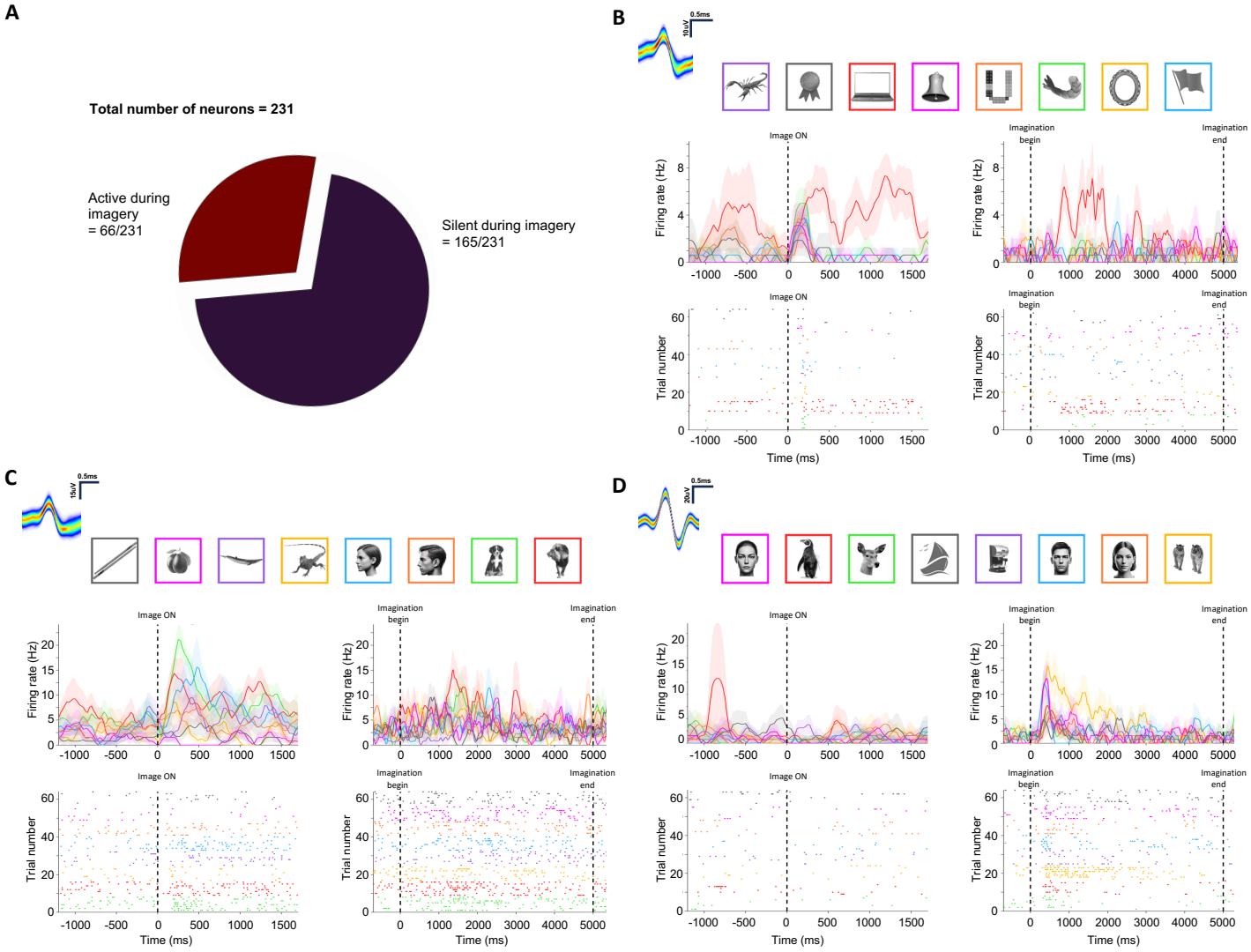
A
Total number of neurons = 231
Active during imagery = 66/231
Silent during imagery = 165/231
B
0.5ms
10uV
Image ON
Imagination begin
Imagination end
Firing rate (Hz)
Trial number
Time (ms)
C
0.5ms
15uV
D
0.5ms
20uV

Figure 4. Cued Imagery task reveals robust reactivation of VTC neurons.

[Note: The human faces in panels B-D of this figure have been replaced with synthetic faces generated by a diffusion model (*102*), in accordance with the bioarxiv policy on displaying human faces.]

(A) The proportion of total recorded VTC neurons that were active (or reactive) during imagery. **(B-D)** Example neurons. Left PSTH and raster shows the response of the neuron during the encoding while right PSTH and raster shows the response during imagery. The stimulus images used in the task are shown above arranged in ascending order of projection onto the neurons preferred axis (computed during screening). **(B)** This neuron's preferred stimulus was the laptop during encoding and it reactivated robustly during imagery of the laptop. Note that this was not an axis tuned neuron and as such the laptop is not the right most image. **(C)** Example of an axis tuned unit that reactivated to multiple stimuli (red & green) in a graded manner during imagery. **(D)** A small number of neurons recorded (15/231) were quiet during encoding but strongly active during imagery. Once such example is shown.

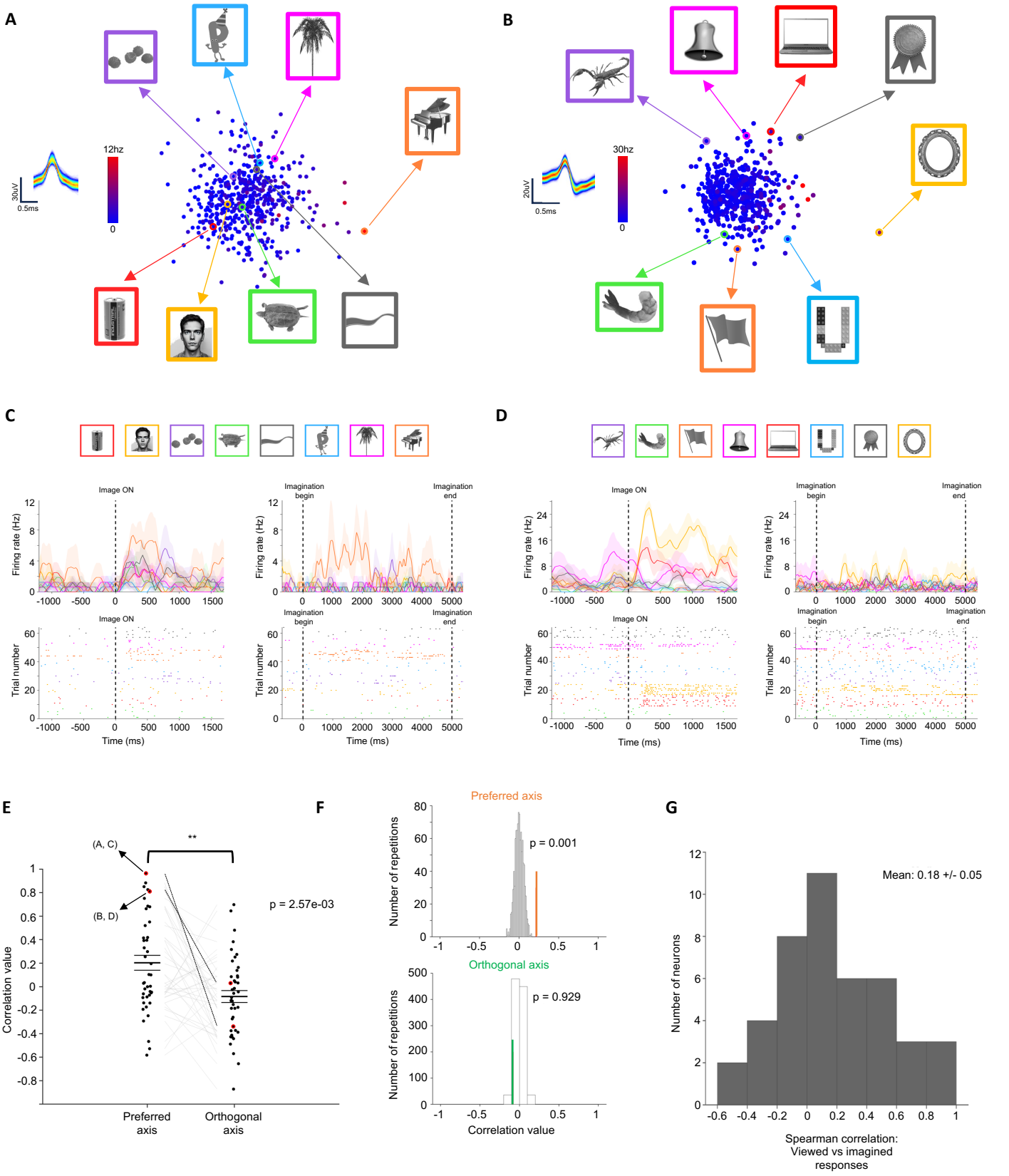
A
12hz
0
30uV
0.5ms
B
30hz
0
20uV
0.5ms
C
Image ON
Imagination begin
Imagination end
Firing rate (Hz)
Trial number
Time (ms)
D
Image ON
Imagination begin
Imagination end
Firing rate (Hz)
Trial number
Time (ms)
E
(A, C)
(B, D)
**
p = 2.57e-03
Correlation value
Preferred axis
Orthogonal axis
F
Preferred axis
p = 0.001
Number of repetitions
Orthogonal axis
p = 0.929
Number of repetitions
Correlation value
G
Mean: 0.18 +/- 0.05
Number of neurons
Spearman correlation:
Viewed vs imagined
responses

Figure 5. Neurons in VTC reactivate in a manner that respects the visual code.

[Note: The human face in panels A&C of this figure have been replaced with a synthetic face generated by a diffusion model (*102*), in accordance with the bioarxiv policy on displaying human faces.]

(A-D) Two example neurons. (A, B) Axis plots with the subset of 8 stimuli used for imagination indicated. Inset shows the waveform. **(C, D)** Response during encoding/viewing (left) and imagery (right). The top panel shows the stimuli, arranged in order of increasing projection value along the preferred axis. **(E-F)** Population summary. **(E)** Pearson correlation between the projection value onto the axis computed during screening and firing rate during imagery is shown for the preferred and orthogonal axes (n=43 neurons). Across neurons the preferred axes showed a significantly higher positive value (r_{pref} = 0.20, r_{ortho} = -0.084, p = 2.57e-03 Wilcoxon ranksum test). The neurons discussed in (A-B) and (C-D) are marked in red. **(F)** Comparison of the mean correlation across all reactivated neurons to the null distribution. Responses were significantly correlated to the projection value onto the preferred axis (p = 0.001, shuffled distribution with 1000 repetitions) but not to the projection value onto the orthogonal axis (p = 0.929, shuffled distribution with 1000 repetitions). **(G)** Correlation between the firing rates during viewing and imagination of the same stimuli. The responses to each stimulus in encoding and imagery were averaged across trials and the Pearson correlation coefficient was computed between those two vectors for each neuron. The mean value is 0.18, significantly larger than 0 (p = 2.40e-3, one sample ttest).

Supplementary materials for

A shared code for perceiving and imagining objects in human ventral temporal cortex

V. S. Wadia, C. M. Reed, J. M. Chung, L. M. Bateman, A. M. Mamelak,
U. Rutishauser[ψ], D. Y. Tsao[ψ]

This PDF file includes:

Supplementary Results
Supplementary Materials and Methods
Figs. S1 – S6
Table S1, S2

Supplementary results

Axis code is independent of the specific convolutional network used to parametrize stimuli (relevant for Figures S3 & S4)

All analyses reported in the main text use a feature space built from layer fc6 of AlexNet. However, there now exist a plethora of deep convolutional neural network models that achieve high performance on object recognition (*1*). We therefore set out to compare the ability of several such models to explain the responses of VTC neurons to general objects.

The models tested include AlexNet (*2*), both VGG-16 and 19 (*3*), VGG-Face (*4*), the eigen object model wherein the space is built by performing PCA on the pixel level representation directly (*5*, *6*), and four CORNet models (*7*, *8*). AlexNet, VGG-16/19, and the CORNet family of networks are trained to classify images into 1000 object categories with varying architectures. AlexNet has 8 layers, with 5 convolutional and 3 fully connected layers. VGG-16 and 19 have 16 and 19 layers respectively, with 3 fully connected layers and the rest being convolutional layers. VGG models are also known for leveraging the smallest possible receptive field in their convolutional layers (3x3). The CORNet family of networks consists of 4 networks: CORNet-Z is purely feedforward; CORNet-R includes some recurrence which has been shown to be essential for object recognition in the primate visual system (*9*); CORNet-RT has the same structure as R but includes 'biological unrolling' wherein the input at time $t+1$ in layer n is the same as input to layer $n-1$at time t so that information flows through the layers sequentially (*7*); and finally CORNet-S has the most complicated of architecture, including recurrent and skip connections between the layers (*8*). Despite the individual differences all four networks have architectures inspired by the primate visual system with layers corresponding to V1, V2, V4, and IT. VGG-Face has the same architecture as VGG-16 but is trained to identify 2622 celebrities (*4*).

To quantify the ability of each network to explain human VTC responses we learned a linear mapping between the features of each model and the neural responses (*10*). As we did in our earlier axis tuning computations and to avoid overfitting, we reduced dimensionality of the feature representations via principal-component analysis (PCA) yielding 50 features for each object and model. In our main analyses, we used leave-one-out cross validation and for each neuron fit the responses of 500-1 images to the 50 features via linear regression before predicting the response of the neuron to the left-out image using the same linear transform. The explained variance by the linear transform was used as an initial measure of goodness-of-fit. Beyond this we also computed the encoding and decoding error for each neuron with every model. Encoding error was computed as follows: the observed population vector to each object was compared to the predicted population response vector and the observed population response vector to a random other object in the set. If the angle between the observed and predicted response to the chosen object was smaller than the angle between the predicted response and the distractor, the prediction was considered correct (Figure S3B). Decoding error was computed via the same method except the feature vector of the object was predicted from

the neural responses. In other words, the roles of the neural responses and the model features were reversed (see methods). A model was considered to explain neural responses well if it has high explained variance and low encoding/decoding error. We found that with the exception of VGG-Face and the eigen object model that performed significantly worse, there was no significant difference in explained variance between the models (Figure S3A; p = 8.72e-10, AlexNet vs VGG-Face, Wilcoxon ranksum test; p = 4.49e-25, AlexNet vs eigen model, Wilcoxon ranksum test). The most complicated CORNet, CORNet-S outperformed the purely feedforward CORNet-Z (p = 1.69e-3, CORNet-S vs CORNet-Z, Wilcoxon ranksum test) but not its recurrent counterparts (p = 0.59, CORNet-S vs CORNet-R; p = 0.601, CORNet-S vs CORNet-RT, Wilcoxon ranksum test).

A note on the heterogeneity of mental representations across people (relevant for Figure S6)

There is a large body of evidence to support the notion that the subjective vividness of visual imagery varies greatly between individuals (*11–13*), with some individuals demonstrating a complete inability to generate a mental image (aphantasia) (*14*) while others have near-photorealistic mental images (hyperphantasia). Moreover, various neuroimaging studies have shown differences in fMRI bold signals — both intensity in early visual areas (*15*) and functional connectivity between areas (*12*, *16*) — between subjects that reported different amounts of vividness in imagery. Growing evidence of these differences has led to the conclusion that examining mental imagery at the group level with current tools (fMRI and psychophysics) is not appropriate — leading to the end of the “imagery debate” (*17*, *18*).

In order to understand whether there is a correlation between the data discussed here and subjective vividness, patients also completed the Vividness of Visual Imagery Questionnaire (VVIQ) (*19*). These responses were recorded in-person starting with the 3rd patient in this study and retroactively via video call for the previous ones. Remarkably, all the patients discussed here had very high scores in the vividness scale, with every single one of them falling into the ‘hyperphantasic’ category (Figure S6D). It therefore remains unclear whether the recapitulation of sensory context demonstrated in this study extends to individuals with weak visual imagery capabilities.

Supplementary Methods

Model comparisons

Extraction of features from stimulus images

Each stimulus image was fed into one of the following models to extract the corresponding features:

Eigen object model:

PCA was performed on the original images of the 500 stimulus objects and the top 50 PCs were extracted to compare with other models.

AlexNet:

We used a pre-trained MATLAB implementation of AlexNet. This is an 8 layer deep convolutional neural network with 5 convolutional layers and 3 fully connected layers, trained to classify images into 1000 object categories.

VGG Family:

We used pre-trained MATLAB implementations of VGG-16, a 16 layer deep convolutional neural network that contains 16 layers with 13 convolutional layers and 3 fully connected layers trained to classify images into 1000 object categories (*3*), VGG-Face which has the same structure as VGG-16 but is trained to recognize the faces of 2622 celebrities (*4*), and VGG-19 which has 19 layers (16 convolutional and 3 fully connected) trained on the same task as VGG-16 (*3*).

CORNet Family:

We used a pre-trained PyTorch implementation of CORNet. The CORNet family contains 3 architectures: CORNet-Z, CORNet-R, and CORNet-S. Each architecture includes 4 main layers that correspond to V1, V2, V4, and VTC. CORNet-Z is the simplest model and is purely feedforward. CORNet-R takes the otherwise feedforward network and introduces recurrent dynamics within each area. CORNet-S is the most complex containing within area recurrent connections, skip connections and the most convolutional layers. Our plots include a CORNet-RT plot which refers to a version of CORNet-R that does biological temporal unrolling (*7*) (see fig 2 in reference).

The parameters of AlexNet and VGG were obtained from MATLAB's Deep Learning Toolbox. The CORNets were downloaded from (https://github.com/dicarlolab/CORnet).

Explained variance computation

See main methods ('Axis model' subheading). The reported results are the ratio of explained to explainable variance in the 321/367 axis tuned neurons that had a > 10% explainable variance.

Encoding/Decoding error computation

For the encoding analysis, the response of each neuron was first zscored, then the same procedure as the explained variance computation was followed to obtain a predicted response to every single object. To quantify prediction accuracy, we examined the angle between the predicted population response to each object and its actual population response (target) or the population response to a different object (distractor). If the angle between the predicted response vector and the distractor was smaller than the angle between the predicted response vector and the target this was counted as an error. Overall encoding error was quantified as the average errors across 1000 pairs of target and distractor objects.

For the decoding analysis we used exactly the same procedure, however the roles of the neural responses and the object features were reversed. We first normalized each dimension of object features to have zero mean and unit variance, then for each image we used a leave-one-out procedure to fit a linear transform using the responses to 499 images and then predict the features of the left-out image. Decoding error was computed as the average decoding error across all target and distractor pairs in feature space.

Quantification of axis consistency

The consistency of the preferred axis of a neuron (Figure S2E) was determined as follows: The image set was randomly split into two subsets of 796 and 797 images, a preferred axis was computed for each set and the Pearson correlation was computed between the two. This procedure was repeated 100 times and the axis consistency was defined as the to get an aggregate correlation value.

Supplementary References

1. Y. Taigman, M. Yang, M. Ranzato, L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification" in *2014 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2014), pp. 1701–1708.

2. A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25** (2012).

3. K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, *arXiv [cs.CV]* (2014). http://arxiv.org/abs/1409.1556.

4. O. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition. *BMVC 2015 - Proceedings of the British Machine Vision Conference 2015* (2015).

5. L. Sirovich, M. Kirby, Low-dimensional procedure for the characterization of human faces. *J. Opt. Soc. Am. A* **4**, 519–524 (1987).

6. M. A. Turk, A. P. Pentland, Face Recognition Using Eigenfaces (1991). https://www.cin.ufpe.br/~rps/Artigos/Face%20Recognition%20Using%20Eigenfaces.pdf.

7. J. Kubilius, M. Schrimpf, A. Nayebi, D. Bear, D. L. K. Yamins, J. J. DiCarlo, CORnet: Modeling the Neural Mechanisms of Core Object Recognition, *bioRxiv* (2018)p. 408385.

8. J. Kubilius, M. Schrimpf, K. Kar, R. Rajalingham, H. Hong, N. Majaj, E. Issa, P. Bashivan, J. Prescott-Roy, K. Schmidt, Others, Brain-like object recognition with high-performing shallow recurrent ANNs. *Adv. Neural Inf. Process. Syst.* **32** (2019).

9. K. Kar, J. Kubilius, K. Schmidt, E. B. Issa, J. J. DiCarlo, Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nat. Neurosci.* **22**, 974–983 (2019).

10. L. Chang, B. Egger, T. Vetter, D. Y. Tsao, Explaining face representation in the primate brain using different computational models. *Curr. Biol.* **31**, 2785-2795.e4 (2021).

11. J. Pearson, T. Naselaris, E. A. Holmes, S. M. Kosslyn, Mental Imagery: Functional Mechanisms and Clinical Applications. *Trends Cogn. Sci.* **19**, 590–602 (2015).

12. N. Dijkstra, S. E. Bosch, M. A. J. van Gerven, Vividness of Visual Imagery Depends on the Neural Overlap with Perception in Visual Areas. *J. Neurosci.* **37**, 1367–1373 (2017).

13. F. Galton, Visualised Numerals, *Nature Publishing Group UK* (1880). https://doi.org/10.1038/021252a0.

14. A. Zeman, M. Dewar, S. Della Sala, Lives without imagery - Congenital aphantasia. *Cortex* **73**, 378–380 (2015).

15. X. Cui, C. B. Jeter, D. Yang, P. R. Montague, D. M. Eagleman, Vividness of mental imagery: individual variability can be measured objectively. *Vision Res.* **47**, 474–478 (2007).

16. J. Fulford, F. Milton, D. Salas, A. Smith, A. Simler, C. Winlove, A. Zeman, The neural correlates of visual imagery vividness – An fMRI study and literature review. *Cortex* **105**, 26–40 (2018).

17. S. M. Kosslyn, Image and brain: the resolution of the imagery debate. *J. Cogn. Neurosci.* **7**, 415–420 (1995).

18. J. Pearson, S. M. Kosslyn, The heterogeneity of mental representation: Ending the imagery debate. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 10089–10092 (2015).

19. D. F. Marks, Vividness of Visual Imagery Questionnaire. *British Journal of PsychologyJournal of Mental Imagery*, doi: 10.1037/t05959-000 (1973).

20. A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, I. Sutskever, Zero-shot text-to-image generation. *ICML* **abs/2102.12092**, 8821–8831 (2021).

21. P. Bao, L. She, M. McGill, D. Y. Tsao, A map of object space in primate inferotemporal cortex. *Nature* **583**, 103–108 (2020).

22. D. L. K. Yamins, H. Hong, C. F. Cadieu, E. A. Solomon, D. Seibert, J. J. DiCarlo, Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8619–8624 (2014).

23. C. F. Cadieu, H. Hong, D. L. K. Yamins, N. Pinto, D. Ardila, E. A. Solomon, N. J. Majaj, J. J. DiCarlo, Deep neural networks rival the representation of primate IT cortex for core visual object recognition. *PLoS Comput. Biol.* **10**, e1003963 (2014).

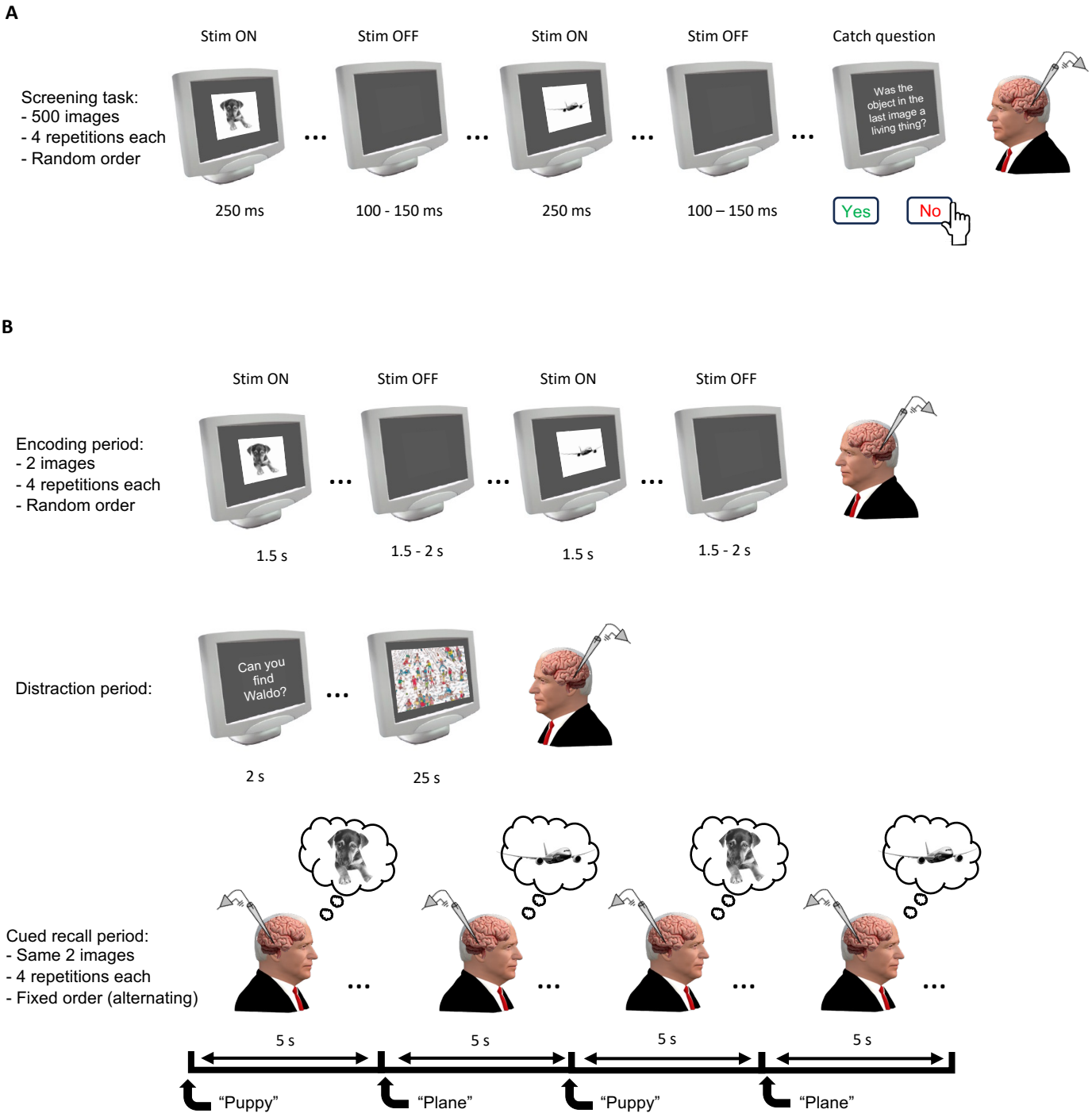
A
Stim ON
Stim OFF
Stim ON
Stim OFF
Catch question
Screening task:
- 500 images
- 4 repetitions each
- Random order
Was the object in the last image a living thing?
250 ms
100 - 150 ms
250 ms
100 – 150 ms
Yes
No
B
Stim ON
Stim OFF
Stim ON
Stim OFF
Encoding period:
- 2 images
- 4 repetitions each
- Random order
1.5 s
1.5 - 2 s
1.5 s
1.5 - 2 s
Distraction period:
Can you find Waldo?
2 s
25 s
Cued recall period:
- Same 2 images
- 4 repetitions each
- Fixed order (alternating)
5 s
5 s
5 s
5 s
"Puppy"
"Plane"
"Puppy"
"Plane"

Figure S1. Detailed schematics of both screening and cued imagery tasks.

Outlines showing the task structure for both screening and cued imagery tasks including number of images used, different task stages, and stimulus order.

(A) Screening task schematic. In the screening task, grayscale images with white backgrounds were displayed on a gray screen for 250 ms with the inter-trial interval jittered between 100-150 ms. Images subtended 6-7 visual degrees. At random intervals (min interval: 1 trial, max interval: 80 trials) a yes-no catch question would appear pertaining to the image that came just before it. Each image was repeated 4 times for a total of 2000 trials **(B)** Cued imagery task schematic. In the cued imagery task, a subset (6-8) of the 500 images used for screening were used, chosen to have spread across both the preferred and principal orthogonal axes. Each trial consisted of 2 images, and an initial encoding period wherein images were displayed on a gray screen for 1.5 s with the inter trial interval jittered between 1.5 – 2 s. After each image was viewed 4 times, a distraction period occurred wherein patients were required to spend 30 s on a visual search puzzle. Finally, after the distraction period they were cued verbally by the experimenter to imagine both images in the trial one by one in alternating order until both had been visualized 4 times. Each image appeared in 2 trials for a total of 8 repetitions of viewing and imagery per image.

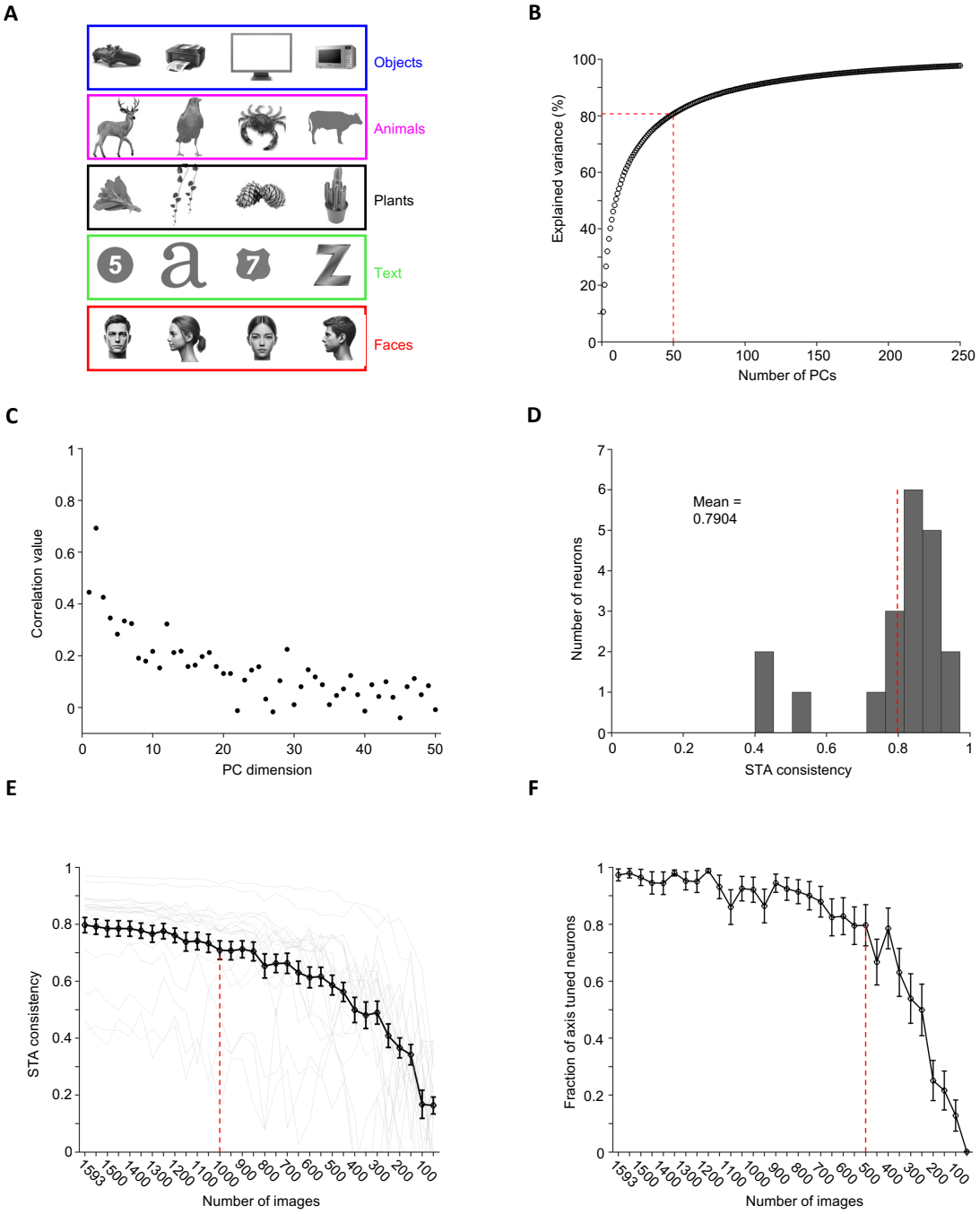
A
Objects
Animals
Plants
Text
Faces
B
Explained variance (%)
Number of PCs
C
Correlation value
PC dimension
D
Mean =
0.7904
Number of neurons
STA consistency
E
STA consistency
Number of images
F
Fraction of axis tuned neurons
Number of images

Figure S2. Stimuli and parameters chosen for screening task.

[Note: The human faces in panel A of this figure have been replaced with synthetic faces generated by a diffusion model (*20*), in accordance with the bioarxiv policy on displaying human faces.]

(A) 500 stimuli from face, text, plant, animal, and object categories were shown to the patients. Example images from each of the 5 categories (taken from www.freepngs.com). **(B)** The cumulative explained variance of fc6 unit responses over 250 PCs. 50 dimensions explained 80.78% of the variance. **(C)** Correlation between predicted feature values using the axes computed via the 500 stimulus screening and actual feature values for all 50 dimensions (related to Figure 3B) **(D-F)** In one early session a patient performed a more comprehensive version of the screening task, with 1593 stimuli (*21*). The 500 used in this project were subsampled from this larger set. The responses of the 22 axis tuned neurons in this session were used to determine appropriate parameters going forward. **(D)** Distribution of axis consistency (computed using half-splits, see methods) for axis tuned neurons in this session. Red line indicates the mean value. **(E)** Axis consistency as a function of the number of stimuli used to compute axes. The red line at 1000 indicates how consistent axes would be if computed using 500 stimuli. Axis consistency remains stable until the stimulus number drops below 400 in each half split. This informed the choice of 500 stimuli. **(F)** Proportion of axis tuned neurons detected as a function of the number of stimuli used. 78% of the axis tuned neurons detected using 1593 stimuli were still detected using 500 stimuli. At lower stimulus numbers the neuron count drops precipitously.

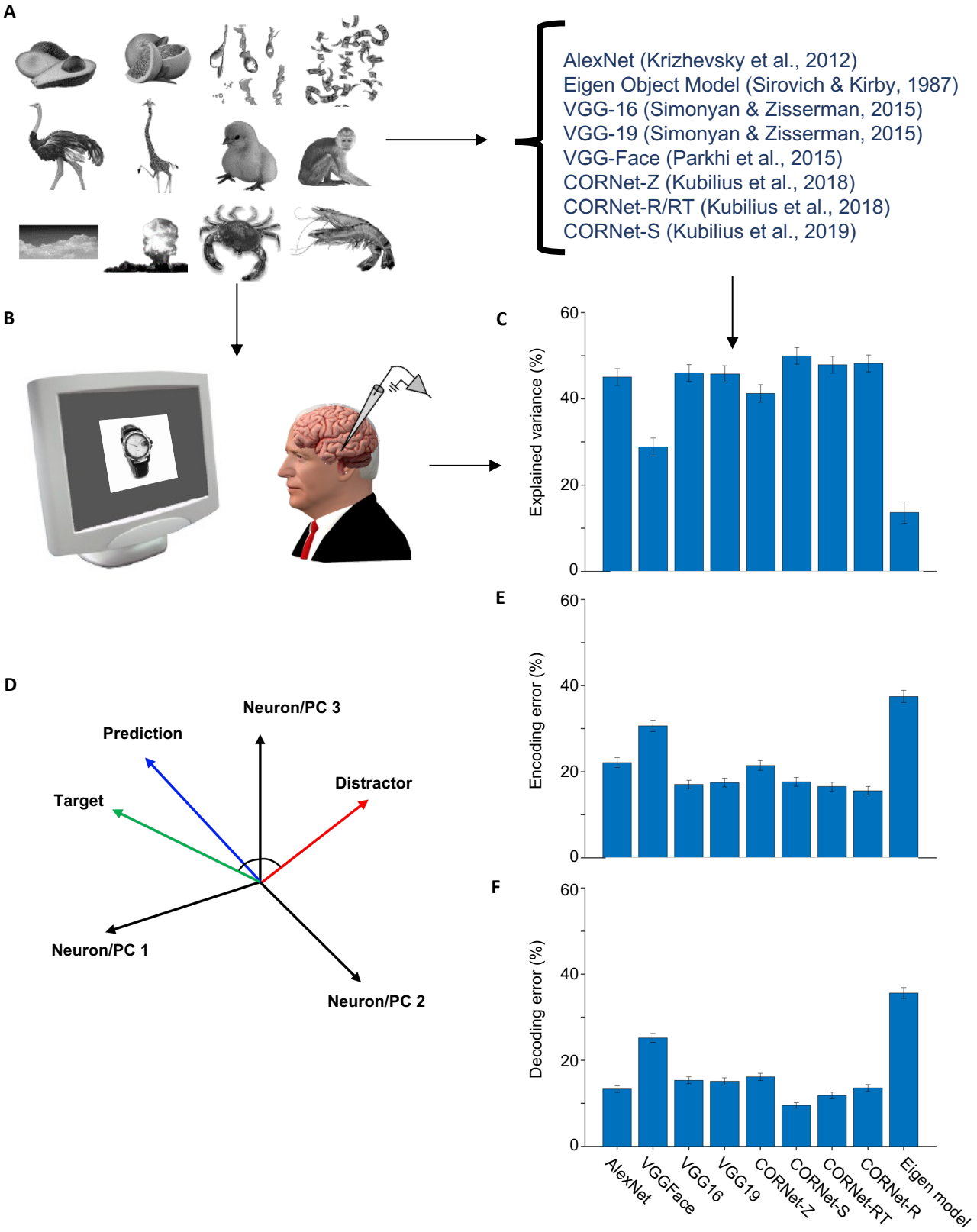
A
AlexNet (Krizhevsky et al., 2012)
Eigen Object Model (Sirovich & Kirby, 1987)
VGG-16 (Simonyan & Zisserman, 2015)
VGG-19 (Simonyan & Zisserman, 2015)
VGG-Face (Parkhi et al., 2015)
CORNet-Z (Kubilius et al., 2018)
CORNet-R/RT (Kubilius et al., 2018)
CORNet-S (Kubilius et al., 2019)
B
C
Explained variance (%)
0
20
40
60
E
Encoding error (%)
0
20
40
60
D
Neuron/PC 3
Prediction
Distractor
Target
Neuron/PC 1
Neuron/PC 2
F
Decoding error (%)
0
20
40
60
AlexNet
VGGFace
VGG16
VGG19
CORNet-Z
CORNet-S
CORNet-RT
CORNet-R
Eigen model

Figure S3. Comparing how well various models explained human VTC responses to object images.

(A) The 500 grayscale stimulus images used in the screening task were parametrized using 8 different models from 4 different model families (AlexNet, the eigen object model, VGG, and CORNet). The same number of features were extracted from units of the different models using principal components analysis (PCA) for comparison. **(B)** Responses from VTC neurons were recorded as patients viewed these objects 4 times each. For recording locations and task schematic see Figure 1D and Figure S1A respectively. **(C)** The explained variances for each model after 50 features were extracted using PCA. For each neuron, explained variance was normalized by the explainable variance (see methods). Error bars represent SEM for the recorded neurons. The eigen model, VGG-Face, and CORNet-Z performed worse than the other models with no significant differences between the rest ($p = 8.72e-10$, AlexNet vs VGG-Face, Wilcoxon ranksum test; $p = 4.49e-25$, AlexNet vs eigen model, Wilcoxon ranksum test). **(D-F)** The various models were compared with respect to how well they could predict the neuronal responses or the object features. In both cases a leave-one-out procedure was used to learn and test the transformation between responses and features. To quantify encoding error for example, for each object we compared predicted responses to individual objects in the neural state space to the actual responses to that object and a distractor object. **(D)** If the angle between the predicted response and the actual response was smaller than the angle between the predicted response and the distractor the encoding was considered correct. To quantify decoding error, we reversed the roles of the neural responses and the object features and decoded object features before comparing the decoded features to the actual features for a given object and a distractor object. **(E)** Encoding error across all models. **(F)** Decoding error across all models. The eigen model, VGG-Face, and the purely feedforward CORNet-Z had larger encoding/decoding errors than the other models, consistent with them explaining much less variance as well.

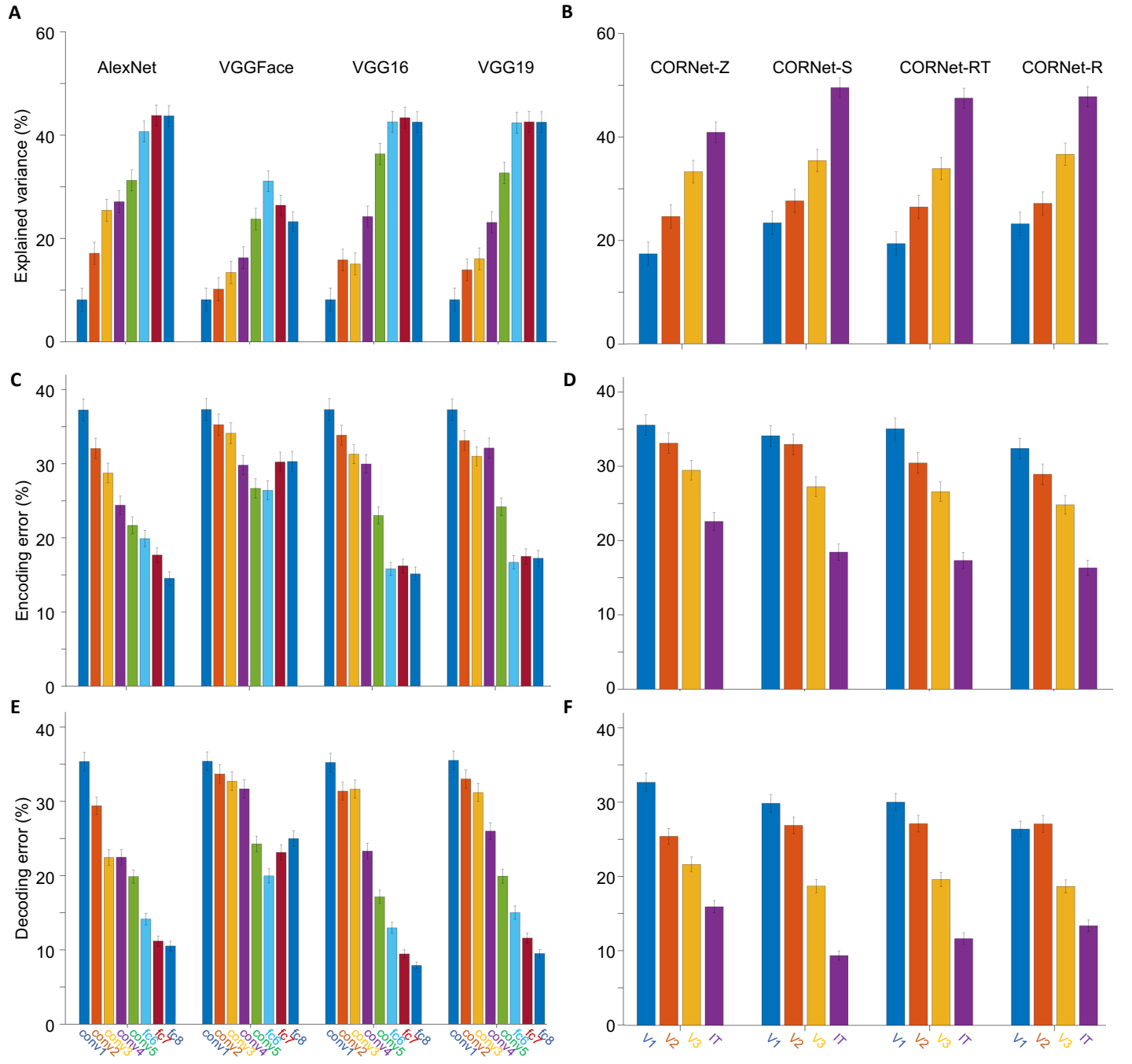
A
B
C
D
E
F
AlexNet
VGGFace
VGG16
VGG19
CORNet-Z
CORNet-S
CORNet-RT
CORNet-R
Explained variance (%)
Encoding error (%)
Decoding error (%)
0
20
40
60
10
30
conv1
conv2
conv3
conv4
conv5
fc6
fc7
fc8
V1
V2
V3
IT

Figure S4. Comparing explained variance, encoding, and decoding error across layers of various models.

Inspired by previous work (*21–23*) units in the penultimate layer were used to build the object space used in all analyses. Here we chose to compare performance across all fully-connected layers of AlexNet and the VGG networks, and all output layers of the CORNet networks used in Figure S3.

(A) Explained variance for all the layers of Alexnet and VGG models (before relu, dropout, or pooling). The performance between the full connected layers is similar. **(B)** Explained variance across the layers of the various CORNet models (V1, V2, V4, IT). The 'IT' or VTC layer performs the best in all CORNet versions, with little difference across the various forms with the exception of the purely feedforward version (CORNet-Z) performing worse than its recurrent counterparts. **(C)** Encoding error across layers of Alexnet and VGG models. As expected, the encoding error is lowest for the fully connected layers with the highest explained variance. **(D)** Encoding error across layers of the CORNets. **(E)** Decoding error across layers of Alexnet and VGG networks. **(F)** Decoding error across layers of CORNets.

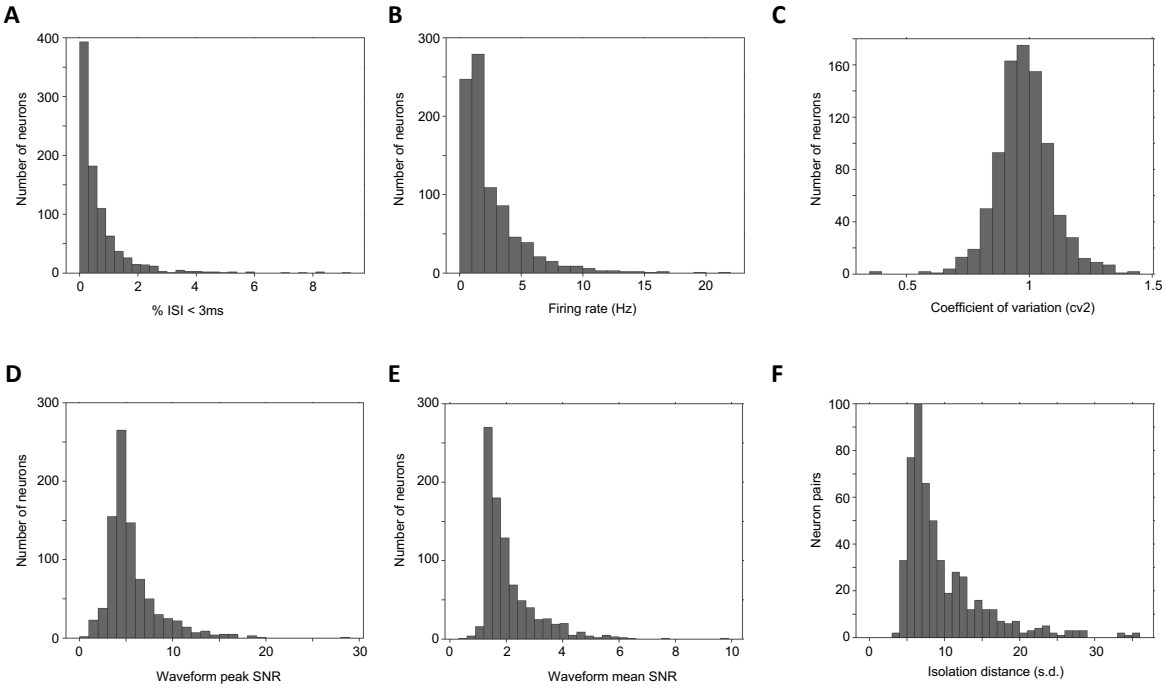
A
Number of neurons
% ISI < 3ms
B
Number of neurons
Firing rate (Hz)
C
Number of neurons
Coefficient of variation (cv2)
D
Number of neurons
Waveform peak SNR
E
Number of neurons
Waveform mean SNR
F
Neuron pairs
Isolation distance (s.d.)

Figure S5. Spike quality metrics for all identified putative single units.

(A) Proportion of inter-spike intervals (ISI) below 3ms. **(B)** Average firing rate. **(C)** Coefficient-of-variation. **(D)** Signal-to-noise ratio (SNR) for the peak of the mean waveform across all spikes as compared to the standard deviation of the background noise. **(E)** Mean SNR of the waveform. **(F)** Pairwise distance between all pairs of neurons on channels where more than one neuron was isolated.

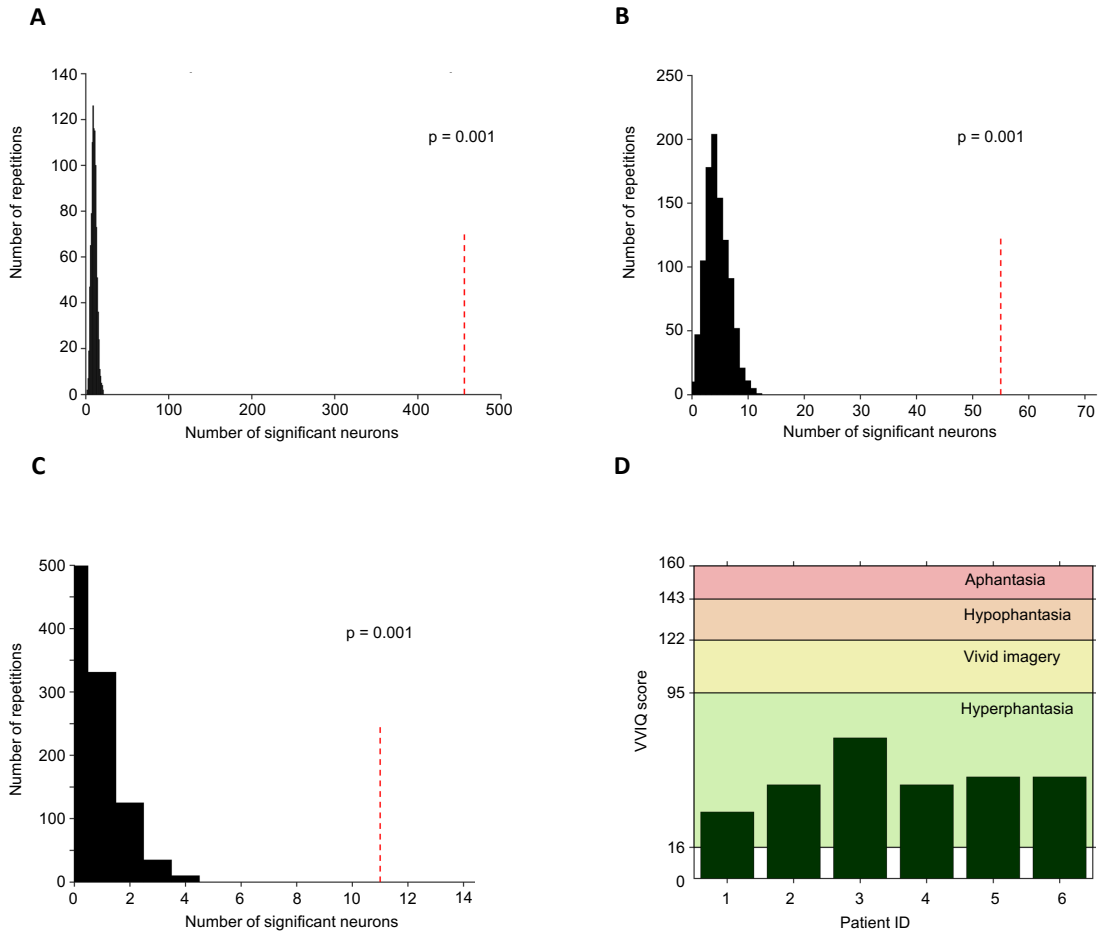
A
Number of repetitions
Number of significant neurons
p = 0.001
B
Number of repetitions
Number of significant neurons
p = 0.001
C
Number of repetitions
Number of significant neurons
p = 0.001
D
VVIQ score
Patient ID
Aphantasia
Hypophantasia
Vivid imagery
Hyperphantasia

Figure S6. Bootstrap statistics for number of neurons selected as visually responsive and active during imagery.

(A) Significance of the number of visually responsive VTC neurons. The dashed red line indicates the number of neurons selected as visually responsive (456/714). The null distribution (black) was estimated by re-running the identical selection procedure (sliding window ANOVA, see methods) after randomly shuffling the trial labels. Shuffling the trial labels destroys the association between the spiking response and trial identity, but keeps everything else (trial number, stim ON time etc.) intact. The shuffling procedure was carried out for 1000 repetitions. The mean of the null distribution was 16, implying that the chance level for selecting a neuron as visually responsive is 16/714 or ~2%. The p-value reported is the percentage of null distribution values that are greater than the chosen number of neurons. In this case, $p = 0$ and is reported as 1/number of repetitions $p = 0.001$. **(B, C)** Significance of the number of neurons in VTC active during imagery. Given that the selection criteria for activation during imagery is either a number of consecutive significant bins of a sliding window ANOVA or sliding window ttest we computed a null distribution for each. The null distribution for each is computed by re-running the identical selection procedure for 1000 repetitions after randomly shuffling the spike times for the ttest and the trial labels for the ANOVA in the visual responsivity test (see A). The mean of the null distribution for the ttest (B) was ~5, implying that the chance level for labeling a neuron as active during imagery via ttest is 5/231 or ~2% and $p = 0.001$, while for the ANOVA the mean is ~1 so the chance level is 1/231 or ~0.5% and $p = 0.001$. **(D)** All patients completed the 'Vividness of visual imagery' questionnaire which is a self-assessment of one's visualization capabilities. All patients recorded from in this study were 'hyperphantasic' or having very vivid visualizations.

Patient IDs	LVTC (x, y, z)	RVTC (x, y, z)
P73CS	N/A	36.06,-38.29,-13.78
P75CS	N/A	34.12, -55.84, -11.08
P76CS	N/A	34.96, -43.33, -24.04
P77CS	-41.26, -46.26, -15.39	34.03, -40.86, -20.00
P78CS	N/A	28.78, -38.95, -5.94
P79CS	-31.19, -28.97, -24.28	47.59, -29.77, -24.81
P80CS	-39.73, -40.46, -16.97	37.15, -49.70, -12.93
P81CS	-33.27, -61.10, -13.95	29.14, -61.91, -7.52
P82CS	-36.95, -40.17, -17.92	40.58, -42.64, -17.30
P84CS	-31.36, -46.56, -16.80	35.91, -31.69, -25.32
P85CS	-39.76, -41.34, -24.80	N/A
P86CS	-33.16, -40.89, -22.33	28.86, -77.30, -10.03
P87CS	-44.08, -51.67, -23.83	36.87, -48.95, -20.19
P88CS	-43.60, -45.15, -18.00	39.66, -47.79, -15.25
P92CS	-34.38, -41.74, -23.60	38.30, -34.42, -20.74
P93CS	-41.03, -34.84, -27.86	39.40, -38.25, -25.13

Table S1. MNI coordinates of microwire bundles in which at least 1 VTC neuron was recorded. Related to Figure 1.

Patient IDs	Sex	Age	Screening		Cued Imagery	
			# Sessions	# resp neurons	# Sessions	# resp neurons
P73CS	F	61	1	16	N/A	N/A
P75CS	F	29	1	6	N/A	N/A
P76CS	F	26	6	22	3	12
P77CS	F	46	1	6	N/A	N/A
P78CS	F	54	2	1	N/A	N/A
P79CS	F	44	7	80	3	44
P80CS	F	25	4	13	2	6
P81CS	F	28	4	38	N/A	N/A
P82CS	M	43	3	3	N/A	N/A
P84CS	M	60	4	47	2	27
P85CS	F	66	6	27	2	9
P86CS	F	39	2	18	N/A	N/A
P87CS	F	27	1	11	N/A	N/A
P88CS	F	29	1	15	N/A	N/A
P92CS	F	30	6	52	2	33
P93CS	F	51	8	101	N/A	N/A

Table S2. Number of sessions and neurons recorded.

Summary of the number of neurons recorded in each subject. In some subjects both Screening and Cued Imagery tasks were performed.