

A concentration of visual cortex-like neurons in prefrontal cortex

Received: 14 September 2023

Olivia Rose^{1,2} & Carlos R. Ponce¹ 

Accepted: 7 August 2024

Published online: 14 August 2024

 Check for updates

Visual recognition is largely realized through neurons in the ventral stream, though recently, studies have suggested that ventrolateral prefrontal cortex (vIPFC) is also important for visual processing. While it is hypothesized that sensory and cognitive processes are integrated in vIPFC neurons, it is not clear how this mechanism benefits vision, or even if vIPFC neurons have properties essential for computations in visual cortex implemented via recurrence. Here, we investigated if vIPFC neurons in two male monkeys had functions comparable to visual cortex, including receptive fields, image selectivity, and the capacity to synthesize highly activating stimuli using generative networks. We found a subset of vIPFC sites show all properties, suggesting subpopulations of vIPFC neurons encode statistics about the world. Further, these vIPFC sites may be anatomically clustered, consistent with fMRI-identified functional organization. Our findings suggest that stable visual encoding in vIPFC may be a necessary condition for local and brain-wide computations.

Areas V1, V2, V4, and inferotemporal cortex (IT) contain a high proportion of neurons that respond strongly to specific shapes at preferred retinal positions. These neurons appear to function as filters that fire maximally when their encoded pattern occurs within an input image¹, so they are frequently modeled as *kernels* in artificial neural networks (ANNs), such as convolutional networks^{2–4}. While there is excitement in using ANNs as models of the visual system, how far can we carry this analogy? One way to answer this is to investigate the state of visual information beyond sensory areas. IT and V4 project to ventral portions of lateral prefrontal cortex (vIPFC)^{5,6}. PFC contains heterogeneous neuronal populations, some showing visual tuning^{7,8}, with most representing behavioral task variables related to decision-making. Frequently, ANNs are trained for classification, where “decisions” are made by an output layer, usually a fully connected stage that combines the activity of units with receptive fields (RFs)⁹. These units have access to all regions of the visual input, and they encode complex conjunctions of lower-level features resembling objects and scenes. This can be demonstrated using feature visualization techniques, which create synthetic images that best describe the RF shape^{10,11} (Fig. 1a). So, is this decision stage in ANNs comparable to what happens in primate cortex – e.g., do visual vIPFC neurons encode full objects and scenes, like fully

connected units in an ANN, or do they encode simpler features like convolutional units?

Many PFC neurons show visual properties during passive viewing (without extensive task training^{7,8,12}), such as RFs^{13–16}, and selectivity for some image categories over others (simple geometric shapes⁸ or faces^{7,17}). However, these observations do not prove conclusively that all vIPFC neurons encode objects or scenes. Visual neurons can appear object-selective when they respond to lower-level features contained within the image^{18,19}, and discovering these features requires a systematic deconstruction of the image (via feature removal¹⁸ or visualization²⁰). Some studies worked to show that in PFC, intact faces were elemental units of representation⁷, but these relied on coarse image-scrambling techniques which can eradicate critical features. To explore if PFC neurons function are comparable to ANN fully connected units (Fig. 1a, bottom right), we set out to determine if vIPFC neurons could drive the synthesis of images using deep networks. This technique combines electrophysiology, search algorithms²¹ and generative adversarial networks pretrained to generate images based on natural statistics²².

We used the DeePSim¹¹ generative network, which encodes natural statistics via low-level features (e.g., short continuous lines, color patches), not pre-defined objects or scenes as in recent networks^{23,24}.

¹Department of Neurobiology, Harvard Medical School, Boston, MA, USA. ²Roy and Diana Vagelos Division of Biology & Biomedical Sciences, Washington University, St. Louis, MO, USA. ✉ e-mail: carlos_ponce@hms.harvard.edu

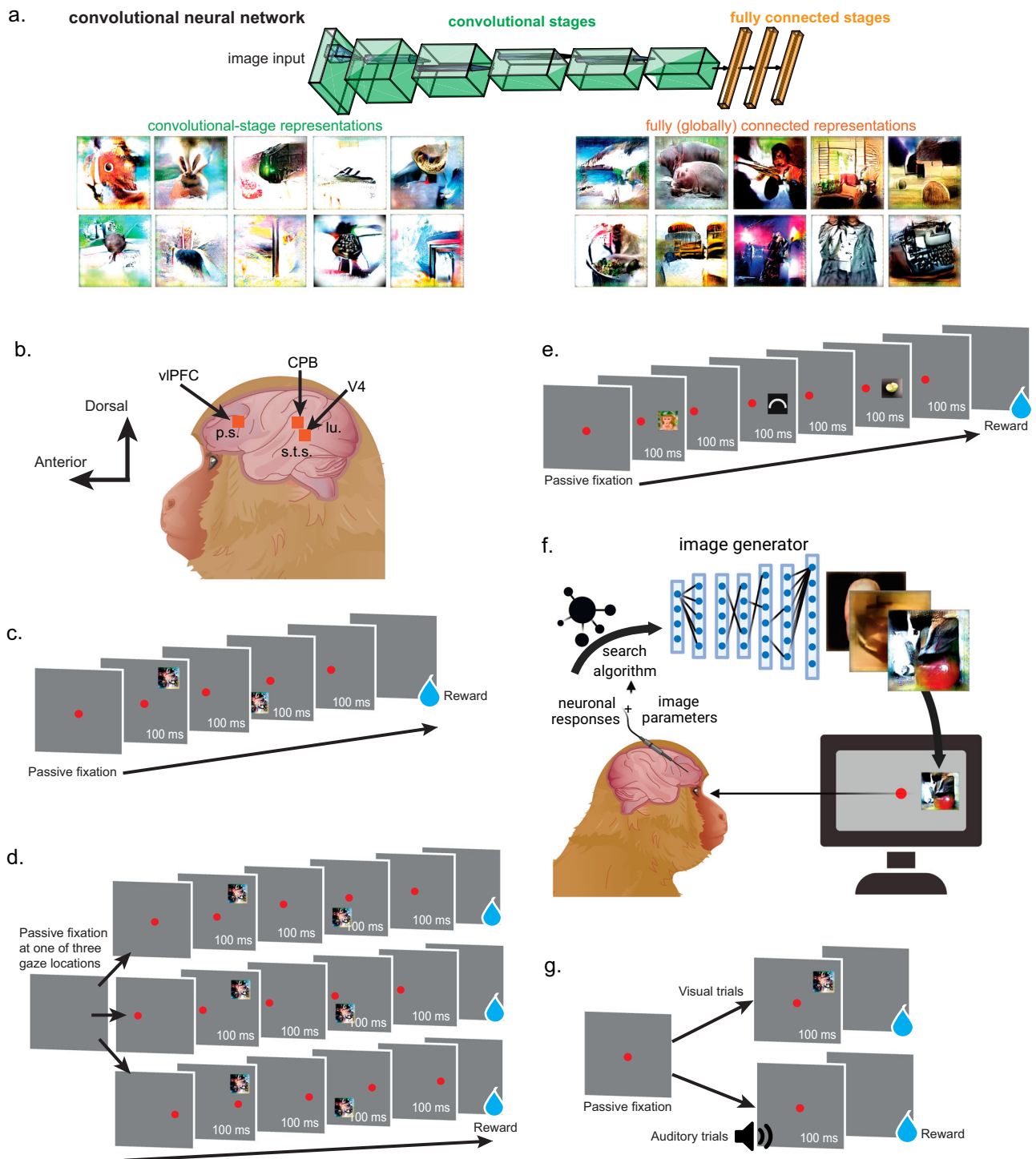


Fig. 1 | Experimental designs. **a** Common convolutional neural network architecture with convolutional layers and fully connected layers, with the last used for classification. Feature visualization shows low-level pattern encoding in convolutional layers and more object- and scene-like encoding in connected layers.

b Location of microelectrode arrays implanted in cortex of two macaques. VIPFC ventrolateral prefrontal cortex. CPB auditory caudal parabelt. P.S. principal sulcus, S.T.S. superior temporal sulcus, LU lunate. **c** Receptive field mapping task: stimulus is flashed in multiple locations on the screen, blue drop represents liquid reward delivery. **d** Retinotopic vs. allocentric coding experiments: designed as previous fixation task, only the horizontal position of the fixation point shifts pseudo-

randomly across trials. **e** Selectivity: different images are presented at the same position relative to the fixation point. **f** Image synthesis experiments: images are manipulated (“evolved”) to increase firing rate, using a closed-loop design comprising neuronal firing rate responses, an image generator (generative adversarial network, GAN), and a search algorithm (covariance matrix adaptation evolutionary strategy, CMAES). **g** Polysensory experiments: subject maintains fixation while an image or sound is played, followed by reward delivery. All tasks relied on passive fixation. $N = 32$ unique microwires per array, showing signals from single neurons and neuronal microclusters (multiunits), involving 16 and 21 experiments (monkey C and D).

We also characterized the basic visual properties of the same vIPFC neurons to determine if they also had receptive fields (RFs) and selectivity in the absence of task training. The answer has significant implications for both biological and computer vision research programs: given evidence that vIPFC is necessary for computations in IT²⁵, when designing deep models of the primate visual system, is it essential to include a vIPFC-like processing stage that is fully connected or convolutional?

Results

We trained two rhesus macaques (Monkeys C and D, both adult males) to perform a passive fixation task, keeping their gaze on a red spot at the center of the monitor as images were flashed. Each animal was implanted with two 32-channel floating microelectrode arrays (Microprobes for Life Sciences). Both animals had one array implanted in right vIPFC, ventral to the principal sulcus and anterior to the arcuate sulcus (Fig. 1b); the specific implant location was dictated by visual identification of these sulci and by avoiding local vasculature. The most likely location for the arrays was 8A, given their proximity to the ventral-dorsal border marked by the principal sulcus²⁶. Monkey C had a second 32-channel array implanted in area V4, between the lunate- and the inferior occipital sulcus. Monkey D had a second 32-channel array implanted in the auditory caudal parabelt (CPB, at the caudal tail of the gyrus dorsal to the superior temporal sulcus). Most array channels showed multiunit and hash activity, although there were also transient single units. Below, we report both single- and multiunit activity results as *sites*, although examples using single units will be specified.

Many vIPFC sites showed position tuning consistent with V4 RFs

Each site was tested to determine if (a) it was visually responsive and (b) it had an RF. Images (5°-wide) were flashed along an invisible grid spanning 40° of visual angle (−20° to 20°) along the x- and y-axis while the animal kept fixation (Fig. 1c). A site was defined as visually responsive if it showed a statistically reliable change in firing rate after image onset. The same site was defined as having an RF if this firing rate change was dependent on position. To determine if neuronal sites had a significant change in firing rate in response to visual stimulation, we divided the first 200 ms of response after image onset into a *baseline* window (1–30 ms) and an *evoked* window (50–150 ms). We computed the mean firing rate in each time window per experiment, then performed a two-sided Student's *t*-test and corrected by false-discovery rate (FDR). We found that 32/32 (100%) of V4 sites showed a statistically reliable change in firing rate, whereas none of the 32 (0%) CPB sites did ($p < 0.05$, after FDR correction). Next, we measured visual responsiveness in the vIPFC sites across stimulation in all grid positions. We found that 16/32 (50%, Monkey C) and 12/32 (37.5%, Monkey D) of sites in Monkey D showed a significant change in firing rate after image onset ($N = 20$ experiments). To control for baseline window sizes, we also measured the responsiveness of PFC sites in separate experiments, using 5–10°-wide images at locations near the population RF. We focused on the change in activity at −150 to 0 ms before image onset vs. 50 to 200 ms after. We found that out of 434 recordings across days, 41.9% and 40.6% of sites showed responsiveness (using a shorter baseline window resulted in estimates of 46.6% and 43%). We conclude that many vIPFC neuronal sites signaled the onset of a picture on the screen in the absence of task training, similarly to V4 sites.

Next, we tested if this responsiveness depended on position. To determine the proportion of sites with RFs, we used each site's responses to perform a one-way ANOVA factor of stimulus position. We also computed each site's mean response per position and then visualized a mean RF across all trials ($N = 16$ experiments for Monkey C, 21 for Monkey D). We fit a 2-D Gaussian function²⁷ to each RF to obtain the horizontal and vertical positions and tuning widths. We found that all 32 V4 sites (100%, Monkey C) showed significant RFs across days

($p < 0.01$, one-way ANOVA with correction by false-discovery rate), while none of the 32 CPB sites (0%, Monkey D) did. In vIPFC, a reliable subset of sites demonstrated RFs consistently: out of 32 vIPFC channels per monkey, six sites in Monkey C (18.8%) and 12 in Monkey D (37.5%) showed robust RFs (18/64 total vIPFC channels, 28.1%). These RFs represented the contralateral visual hemifield to the implanted hemisphere, though a small fraction spanned the midline between contra- and ipsilateral hemifields (Fig. 2a–c, Supplemental Fig. 1a, b). Compared to V4 RFs, vIPFC RFs covered more of the visual field (i.e., had larger estimated widths; see Fig. 2d), though in our samples, vIPFC RFs were also more eccentric than the V4 RFs (Supplemental Table 1). V4 RFs were located around $(10.1, -3.2)^\circ \pm (0.003, 0.005)^\circ$ (horizontal and vertical coordinates, median \pm standard error via bootstrap), with respect to fixation $(0,0)^\circ$. The tuning widths of V4 RFs ranged from 4.5° to 5.0°, with a mean width of $4.7^\circ \pm 0.001^\circ$ (across X and Y). The sampled RFs in vIPFC were located around $(12.7, 11.5)^\circ \pm (0.1, 0.03)^\circ$ in Monkey C, and at $(10.2, -11.3)^\circ \pm (0.04, 0.05)^\circ$ in Monkey D (Fig. 2c). vIPFC RF tuning widths ranged from 4.6° to 12.2°, with a mean width of $7.7^\circ \pm 0.07^\circ$ in Monkey C and $7.2^\circ \pm 0.02^\circ$ in Monkey D (Fig. 2d). We performed a limited test to determine if these RFs were *retinocentric*—bound to the position of the fovea—or *allocentric*, influenced by absolute position in world coordinates; we observed that the RFs shifted as the fovea moved between fixation positions (Fig. 1d, Supplemental Fig. 2). We conclude that a substantial fraction of vIPFC sites showed robust RFs that were measurable even in the absence of task training. Below, we show these RFs are stable across months.

We found that sites with reliable RFs seemed to be close to one another within the layout of the array (Fig. 2b). To test if such clustering could arise from random sampling of an otherwise mixed (*salt-and-pepper*) distribution of sites having or lacking RFs, we calculated the mean distance (in μm) between channels having RFs, then performed a permutation test to estimate the probability of this clustering arising from chance. We randomly resampled channels across the array and computed the mean distances among them (each sample having the same number of RF channels, sampled with replacement, Monkey C: $N = 6$ channels; Monkey D: $N = 12$ channels, 1000 samples per monkey). For Monkey C, the observed mean distance between channels with RFs was $667.6 \pm 2.6 \mu\text{m}$, while the mean distance between randomly sampled channels was $1236.1 \pm 7.8 \mu\text{m}$ ($p = 0.01$). For Monkey D, the mean observed distance was $752.2 \pm 1.3 \mu\text{m}$, while the mean randomized distance was $1233.1 \pm 4.9 \mu\text{m}$ ($p = 0.002$). Thus, vIPFC sites with RFs were more likely to be close to one another, providing further evidence of the standing functional organization of RFs in previous studies^{13,28}.

Image selectivity in vIPFC without task training

Ventral stream neurons are commonly characterized in terms of their visual selectivity, which is observed even when animals lacked task training or were anesthetized^{29–31}. To determine if these vIPFC sites showed selectivity, we showed full-color photographs of categories such as *monkeys*, *humans* and other *animals*, *fruits*, *objects*, and *scenes* (Fig. 1e). In case some neurons were tuned for simpler features, we also included artificial stimuli such as Gabor wavelets, lines of different orientations, and curved objects previously used to test V4 neurons^{32,33}. For an independent description of the stimulus set, we used Google Cloud Vision (GCV) to assign labels to all presented images. GCV assigned 3403 labels across images (563 unique). The most common labels were *Terrestrial animal* (0.05%), *Primate* (0.04%), *Snout* (0.04%), *Macaque* (0.04%), *Rhesus macaque* (0.03%), *Plant* (0.03%), *Wildlife* (0.03%), and *Fur* (0.03%) (Fig. 3a, i–iii). The images were presented during a simple fixation task, 100 ms ON, 100 ms off, 3–6 images per trial before rewarding with a drop of juice. We measured the number of image-selective sites per area, and how their preferences changed this label distribution across areas. A visual neuron is typically defined as selective if it shows high activity to some

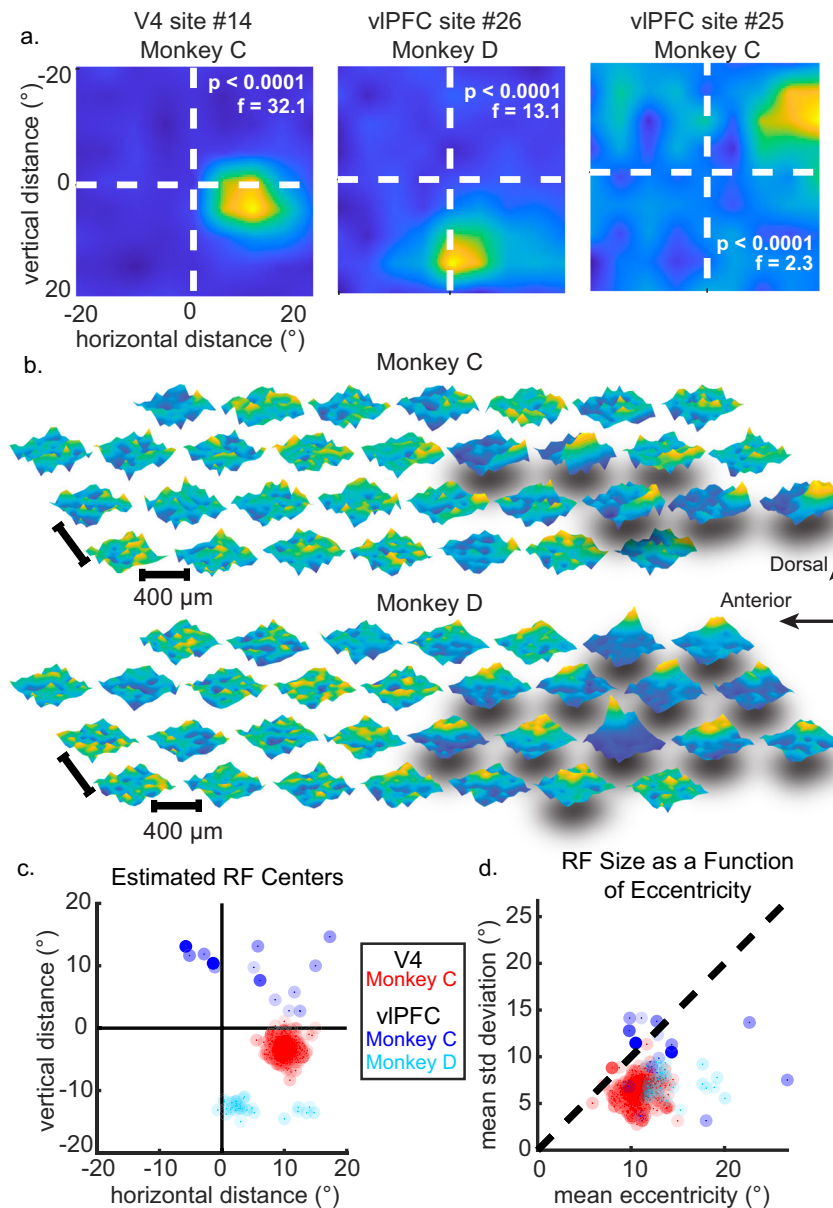


Fig. 2 | Receptive fields. **a** Interpolated heatmaps of RF examples from V4 and vIPFC multiunits, with yellow signaling higher firing rates. Dashed white lines indicate horizontal and vertical meridians of the experimental monitor. **b** Schematic of RF properties as a function of array geometry. Array sites with significant RFs denoted by shadows, and more peaked plots demonstrating stronger RFs. The arrangement of the RF heatmaps reflects the electrode geometry of the Microprobes Inc.'s floating microelectrode array (FMA). **c** Location of RF centers with

respect to the monitor. Solid black lines denote horizontal and vertical meridians. Red points correspond to individual V4 sites (all from Monkey C), while blue points indicate vIPFC sites (dark blue from Monkey C; light blue from Monkey D). Point opacity scales with effect size (F -statistic), with more solid points indicating stronger values. **d** RF size as a function of eccentricity from the fovea. Marker colors and transparency same as (c). Source data are provided as a Source Data file.

images more than others (Fig. 3b, i), as quantified by an ANOVA test (high F -ratio values indicates that the variability of responses across images was higher than within multiple presentations of the same image). We measured this F -ratio while controlling for the placement of the stimulus relative to each site's RF, since the stimulus could not be centered in every unique RF simultaneously. We expected that selective sites would show a larger F -ratio value, plus a negative relationship between F -ratio and increasing RF-stimulus distance. Sites without an RF were treated as having one centered at the point of fixation (0,0) $^{\circ}$.

The number of image-selective sites varied by cortical region, with V4 having the highest percentage of selective sites, PBC having none, and vIPFC arrays an intermediate value (Supplemental Table 2). The V4

population showed an F -ratio range of 0.4–12.1 and a negative correlation between RF location and stimulus distance of -0.47 (Pearson coefficient, $p < 0.001$). The PBC population showed an F -ratio range of 0.6–2.1 with no significant correlation between RF location and stimulus distance (-0.01 , Pearson, $p = 0.61$). The vIPFC populations showed an F -ratio range of 0.6–3.3 (Monkeys C and D: 0.6–3.1 and 0.6–3.3), and both showed a statistically reliable correlation between RF location and stimulus distance (-0.24 , $p < 0.001$ and -0.07 , $p < 0.001$). In short, as a population, vIPFC neurons were conventionally selective for images, though not as numerous as V4.

For what visual attributes were these neurons selective? As a reminder, the stimulus set was chosen to include many monkeys, animals, and faces, so the most common GCV labels (Fig. 3a, i) across

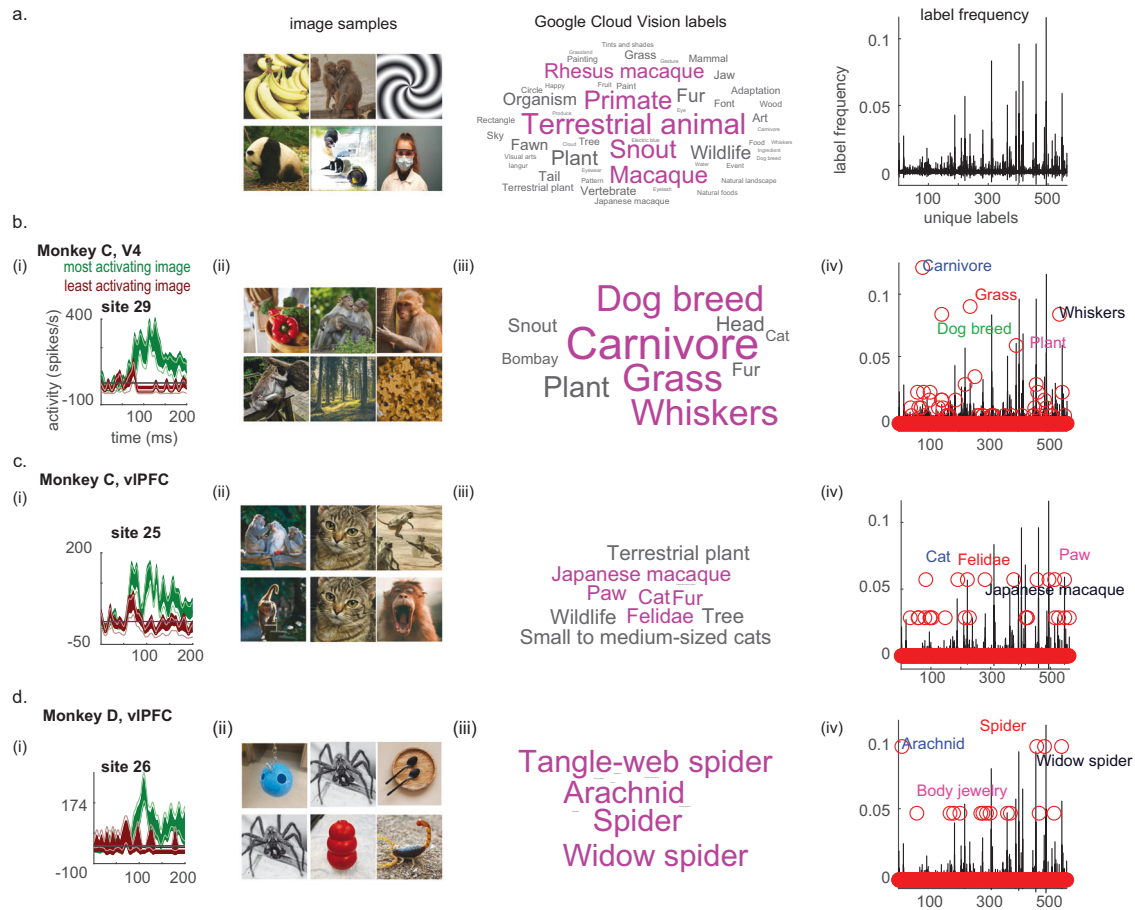


Fig. 3 | Image selectivity. **a** (i). Representative sample of images used. (ii). Labels assigned by Google Cloud Vision to the entire image data set. (iii) Baseline frequency of unique labels across the full image set. **b** (i). Peri-stimulus time histogram for one site in V4 (green trace marks response to top image, brown trace marks response to bottom image, mean \pm SEM), followed by (ii) sample of top preferred images, (iii) labels assigned to the top preferred images, and (iv) change in label distribution based on image preference. **c, d** Same, but for vIPFC sites. $N = 32$

unique microwires per array, involving $N = 1252$ – 1568 experiments involving signals from single units and neuronal microclusters (multiunit signals whose composition often vary across days, see Supplemental Table 2 for statistical details). Images shown are substitutions that resemble those used in the actual experiment, swapped due to copyright reasons. The original Figure with original images can be found in our Github page (see *Data Availability* section). Source data are provided as a Source Data file.

all images were *Terrestrial animal* (0.05%), *Primate* (0.04%), *Snout* (0.04%), *Macaque* (0.04%), *Rhesus macaque* (0.03%), *Plant* (0.03%), *Wildlife* (0.03%), and *Fur* (0.03%). If we had selected the images to perfectly align with neuronal selectivity, we would expect that the label distribution of each site’s preferred images should be statistically similar to this baseline distribution (Fig. 3a, ii-iii). In fact, V4 and vIPFC sites were not so predictable in their tuning (Fig. 3b-d, iii-iv).

We focused this analysis on sites that had RFs situated within 5° of stimulus center (where the image was typically 10° -wide). For each site per experiment, we identified its top preferred image and collected the Google Cloud labels given to that image, then examined the frequency of labels. We found that across the V4 sites, there were 47 unique labels, most prominently showing information related to *Carnivore* (i.e., cats and other animals, 0.12%), *Grass* (0.09%), *Whiskers* (0.09%), *Dog breed* (0.09%), and *Plant* (0.06%). For the vIPFC sites in monkey C, there were 26 unique labels, including *Cat* (0.06%), *Felidae* (0.06%), *Fur* (0.06%), *Japanese macaque* (0.06%), and *Paw* (0.06%) – briefly, many animal-related shape and texture information. For the vIPFC sites in monkey D, there were 16 unique labels: *Arachnid* (0.10%), *Spider* (0.10%), *Tangle-web spider* (0.10%), *Widow spider* (0.10%), and *Body jewelry* (0.05%). Example images are shown in Fig. 3b-d, Supplemental Table 2). In short, vIPFC and V4 sites showed a disproportionate

selectivity for animal features and textures, many of which are present in monkeys or primates, but other animal orders as well.

Common knowledge is that anterior IT and PFC neurons that are visually responsive tend to have large RFs. So, given that most PFC sites did not seem to have local RFs, we considered the possibility that their RF was “everywhere” (i.e., they had RFs as big as the monitor). This is also in line with the hypothesis of PFC neurons functioning more like output units in ANNs. In this scenario, neurons without a measurable RF would still show selectivity to images. To test this possibility, we looked for sites with *F*-ratio values that were large but for which we could not estimate classic RFs. We first analyzed array V4, which we knew had classical localized RFs. We found that V4 sites that were 10° away from the stimulus center showed a median *F*-ratio of 1.10 (0.96–1.32, 25–75%ile). Of these, only 0.20% of 1046 recordings had *p*-values below a significance threshold of 0.01. These are low values, so we considered them noise arising from the use of multiple comparisons and many recordings. We found the same results in both vIPFC populations: sites that were 10° away from the stimulus center showed a median *F*-ratio of 1.03 (Monkey C: 0.92–1.15, 25–75%ile; Monkey D: 0.91–1.11). Of these, 0.05% of 1192 recordings (Monkey C) and 0.04% of 681 (Monkey D) had *p*-values < 0.01 . So, we found no strong evidence that any vIPFC sites were selective over the entire screen.

In short, we found that vIPFC contained neurons that were consistent with those in the ventral stream, having RFs (Fig. 2) and image selectivity (Fig. 3). Yet, we also noted that as in V4, these vIPFC sites did not show selectivity that aligned strongly with one or two simple semantic categories. In the ventral stream, this indicates that neurons are not categorical, but instead tuned to lower-level visual attributes. Currently, the most effective way to determine if a neuron is tuned to lower-level features is to use image synthesis; the efficacy of this image generation technique stems from its use of images that initially lack form and cannot be assigned to any category, and thus can only be driven by a basic affinity to shapes, colors, and textures. So, could vIPFC neurons drive the synthesis of highly activating images?

Stable visual encoding in vIPFC without training, category, or semantics

To investigate if vIPFC selectivity reflected neuronal tuning for categorical or semantic information, or if this selectivity was driven by simpler visual attributes like shape, color, or texture, we implemented a closed-loop method for generating highly activating visual stimuli (Fig. 1f). In this approach, neurons guide generative networks to reconstruct images containing their preferred visual attributes. Each experiment starts with the sampling of abstract textures synthesized by the generator (each texture corresponding to a model *input vector*), which are then presented within a site's RF. Along with their corresponding input vectors, neuronal responses to each texture are propagated into an evolutionary search algorithm (covariance matrix adaptation evolutionary strategy or CMA-ES), which seeks to find new candidate input vectors whose images could maximize the site's firing rate. These new vectors are then converted into images and presented within the site's RF. This process continues with dozens of iterations, and with each iteration, more highly activating visual motifs are found and preserved, while less highly activating motifs are screened out. Over time, cortical neurons drive the real-time generation of their own visual stimuli, which typically elicit higher firing rates from a neuron than any pre-selected stimulus. We have called this algorithm *XDream* and the reconstructed image a *prototype*. We consider an experiment to be successful if the adaptive generator finds visual attributes that elicit significantly higher firing rates than initial images during the first ten iterations of the *XDream* cycle. *XDream* experiments were conducted on sites that both demonstrated an RF that session and exhibited an excitatory response after image presentation. In total, we conducted over 145 *XDream* experiments under guidance of vIPFC neurons (65 experiments for Monkey C; 80 experiments for Monkey D), targeting 19 unique vIPFC sites across both subjects (8 sites in Monkey C; 11 sites in Monkey D). *XDream* experiments could target distinct single units within a given site, or instead use input from the whole channel (multiunit/hash). We also included reference images, including photographs known to elicit strong responses from the sites, to control for changes in signal isolation.

V4 prototypes. First, we confirmed that the image-synthesis technique worked in V4, even with neurons having more eccentrically located receptive fields than we had ever previously tested^{34,35}. Thirty-one of the 32 V4 sites showed reliable RFs and visually evoked activity; these were subsequently tested for the ability to generate prototypes. Of these sites, 29 V4 sites (93.5% of tested sites; 90.6% of all sites) successfully guided prototype generation. Out of 54 *XDream* experiments led by V4 sites, 43 reliably converged on visual attributes that drove a site's firing rate significantly above the initial (non-optimized) textures, for an overall success rate of 79.6% across 29 unique V4 sites (Fig. 4c). Single units, multiunits, and whole-site hash in V4 could all drive successful prototype synthesis (see Supplemental Tables 3-4). Additionally, 35 of the 43 successful experiments also demonstrated a significantly higher firing rate to the synthetic images than the natural

reference photographs (81.4% of successful experiments; 65% of all experiments).

vIPFC prototypes. We found vIPFC sites in both animals that led to prototype generation. As noted in previous sections, there were vIPFC sites with RFs (Monkey C, 8/32 sites, Monkey D, 11/32 sites), which were subsequently investigated with synthesis experiments. Between 25–55% of vIPFC sites with RFs led to prototype generation, converging on visual motifs that drove the site's firing rate to a value significantly higher than the initial 10 blocks of synthetic images (Monkey C, 2/8 RF sites, 6% of all vIPFC sites in the array; Monkey D, 6/11 RF sites, 18.8% of all vIPFC sites). All evolving sites were within the previously identified RF clusters. Parsing these results by experiment, between 11–34% of individual sessions led to successful prototypes (Monkey C, $N=65$ experiments, Monkey D, $N=80$, Fig. 4a-c). There was no systematic dependency of evolution success on signal type: in Monkey C, most successful *XDream* experiments were driven by single units (85.7% of all evolutions, all of which originated from the same site, and located within the RF channel cluster) and the rest by multiunits, distinctly separable from the hash (Fig. 4g; Supplemental Table 4). In Monkey D, all successful *XDream* experiments were driven by either distinct multi-units (85.2% led by multi-unit activity) or whole-site hash (14.8%). The fact that multi-unit and hash signals could yield prototypes hints at a functional organization in vIPFC – face- and color-tuned fMRI patches have been observed in this area^{17,28}. Often, we found that sites showed higher firing rates to synthetic images over real-world photographs: in Monkey C, this was present in seven successful evolution experiments (42.9%, all cases driven by the vIPFC single unit), while for Monkey D, this occurred in 20 successful evolution experiments (74.1%) across four unique sites. Overall, these results suggest that the generator was well-suited to capture visual representations encoded by given vIPFC sites, despite the use of much simpler visual motifs than typically used to probe vIPFC neurons outside of a task context.

Prototypes produced under vIPFC guidance were simple, not unlike those found in V4. Upon visual inspection, most prototypes generated by vIPFC sites appeared to feature dark bounded forms against cleared, lighter backgrounds (Fig. 4d, e); this clearing is present in the ventral stream and can be measured using information-theoretic metrics³⁴. These vIPFC prototypes appeared remarkably stable across days. To measure the stability of repeated prototypes from the same unit, we used a series of morphological operations to remove small artifacts from the image while preserving the global form of the prototype – more precisely, a segmentation mask per prototype. We computed segmentation masks for all final prototypes from a given vIPFC site isolating the bounded features from around the cleared background. After vectorizing each segmentation mask, we computed the distance in correlation space (one minus the Pearson correlation), in which shorter distances indicate higher correlation between masks (Fig. 4e). We found that prototype masks derived from a given vIPFC site were generally more similar to each other (distance of 0.73 ± 0.01 , mean \pm SEM) than to 1) prototype masks from other vIPFC sites (distance of 0.97 ± 0.01 , mean \pm SEM), 2) prototype masks from V4 sites (distance of 0.95 ± 0.01), and 3) masks from shuffled versions of the prototypes under investigation (i.e., randomly permuting the prototype on the pixel level, then generating a segmentation mask of the shuffled prototype; distance of 0.96 ± 0.01 , mean \pm SEM, Fig. 4f). The first two comparisons control for any fundamental biases of the adaptive generator, while the third controls for pixelwise responses independently of global form. To determine whether the observed vIPFC mask similarity was likely to be obtained through chance, we conducted a randomization test and found that the vIPFC mask similarity was highly unlikely to be due to chance alone ($p=0.001$). In summary, we found that vIPFC single- and multiunits could drive synthetic image generation stably across months, providing a

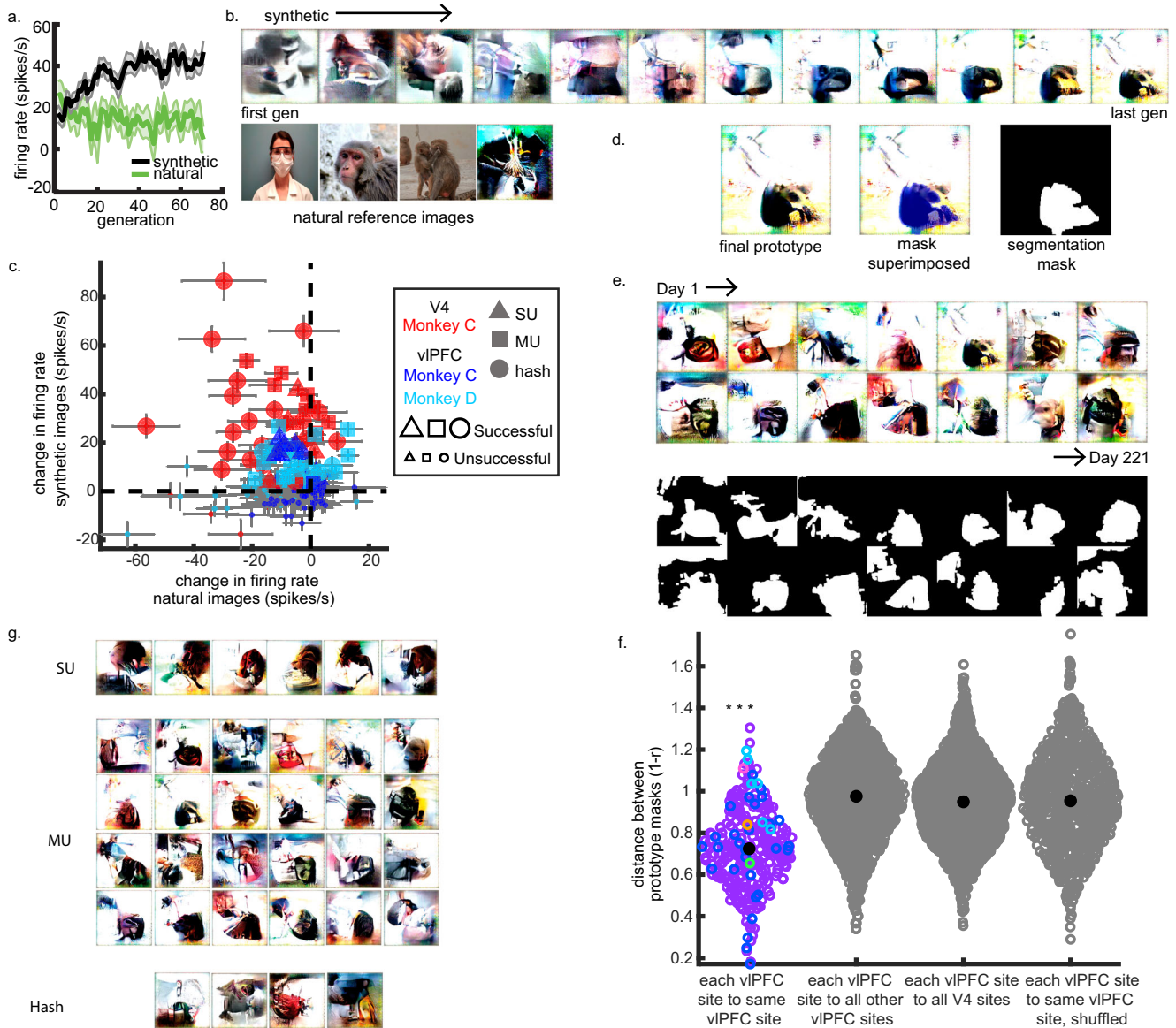


Fig. 4 | Prototype synthesis experiments. **a** Firing rate trajectory during a successful vIPFC prototype experiment. The black trace shows the mean firing rate for synthetic images, while the green trace shows the mean firing rate for natural images (mean \pm SEM). **b** Stimuli presented during the experiment in (a). Representative firing rate trajectory toward synthetic images produced by the closed-loop generator (top row) against fixed reference photographs (bottom row). **c** Population figure of the mean change in firing rate to natural images (x-axis) against synthetic images (y-axis), per experiment. Red markers mark prototype experiments targeting V4 sites, while blue markers denote vIPFC sites (dark blue=Monkey C; light blue=Monkey D). Marker shape represents unit type: triangles for single-units (SU), squares for distinct multiunits (MU), and circles for whole-site hash. Large markers indicating successful experiments. Each point shows a standard error of the mean change, computed via bootstrap ($N = 77$ total experiments, see Supplemental Tables 3 and 4). **d** Segmentation masking of a prototype. **e** Final prototypes from a vIPFC multiunit isolated for over seven months (top) with

their respective segmentation masks (bottom). **f** Within-site prototype similarity ("stability") across the population of successful vIPFC prototypes. For a given site, distances between all prototype masks were computed in correlation space, with shorter distances signaling higher correlation between masks. Colors differentiate individual vIPFC sites across both subjects, with the multiunit from (e) in purple. Distances between prototype masks from each vIPFC site (e.g., between site 1 prototypes from sessions 1 through 3) were compared against, 1) prototypes from all other vIPFC sites (e.g., distances between site 1/session 1 and all other non-site 1 prototypes generated by vIPFC sites), 2) prototypes from V4 sites (e.g., site 1/session 1 and all prototypes generated by V4 sites), and 3) shuffled versions of successful prototypes from the targeted vIPFC site (e.g., site 1/session 1 prototype and the same prototype randomly permuted pixelwise). Median distances denoted by large black markers. Triple asterisks denote statistical significance at $p < 0.001$ by way of randomization test. **g** Prototypes plotted as a function of signal type. Source data are provided as a Source Data file.

functional organization scale for previous fMRI-level observations of visual feature encoding in vIPFC.

We performed additional simulations to determine if we could recover simpler prototypes by combining the selectivity of single units with different complex preferences (Supplemental Fig. 3), but this did not seem to be the case. Further, to examine the differences in prototypes across areas, we focused on variability. We began by

measuring the variability of the vIPFC and V4 sites during the evolution experiments using the Fano factor—the variance divided by mean of every site's spike-rate response distribution. Because every synthetic image was presented only once, we conducted this analysis using the reference images, which were shown once per generation. The median Fano factor values per area were 11.1 ± 0.3 (V4, Monkey C, \pm SE, interquartile range 7.38, sample size $N = 558$), 12.0 ± 0.4 (vIPFC, Monkey C,

±SE, interquartile range 10.56, sample size $N=838$), and 11.0 ± 0.2 (Monkey D, ±SE, interquartile range 8.89, sample size $N=960$). These values varied reliably across areas: a Kruskal-Wallis test revealed a difference in the median dispersion among the groups, $H(2)=11.89$, $p=2.6 \times 10^{-3}$. Yet, performing a two-sample test between V4 and vIPFC (Monkey D) shows a p -value of 0.24, showing no statistical difference between these groups. This suggests that the difference in Fano factor across areas is due to Monkey C vIPFC sites; as a group, they were more variable in their visual responses, and notably, this is a region that also showed more mixed auditory-visual selectivity than the Monkey D vIPFC sites.

Next, we asked if there were differences in the variability of prototypes themselves, as a function of area. Based on one of our previous publications, we know that the input codes to the generator (the latent codes or image “genes”) map closely to image space, such that nearby points in the latent space correspond to similar images in pixel space. We leveraged this information to ask: how variable were the prototypes produced by V4 vs. those produced by vIPFC? We measured the dispersion of latent vectors as a function of closed-loop iteration (generation). Dispersion was computed by obtaining the covariance matrix of all genes per generation ($N=40$) and measuring the trace of the covariance matrix (the sum of the individual dimension variances). We saw that genetic dispersion started high at generation 2, then decreased and settled around the 30th generation. To quantify this relationship, we fit an exponential decay function to these curves of the form $y(x) = A \cdot e^{bx} + c$ to estimate the slope value b . We found that mean slope values were different: for V4, -0.135 ± 0.001 , and for vIPFC, -0.130 ± 0.001 for both monkeys. This means that V4 sites were faster about reducing the variability in their prototypes, which we interpret to be a property of their stronger visual selectivity. We performed a randomization test to estimate the probability of that the observed difference could arise from the same distribution and found this was $p=3.2 \times 10^{-2}$ (see Methods). So, we conclude that V4 reduced the stochastic variability in prototype creation faster than vIPFC did. However, this did not mean that the final prototypes in vIPFC were more reliably more stochastic than those in V4: the median prototype dispersion after responses converged was 22279 ± 22 for V4, 22468 ± 18 for vIPFC (Monkey C), and 22172 ± 17 for vIPFC (Monkey D). A Kruskal-Wallis test revealed a statistical difference among the groups, $H(2)=133.29$, $p=1.1 \times 10^{-29}$. So, the V4 and vIPFC Monkey D prototypes were both less variable than the vIPFC monkey C sites, and this was likely related to the Fano factor result from above. So, we conclude that the prototype variability was stable, but it depended on the relative variability of individual sites, even to natural images.

Polysensory tuning of vIPFC sites

Neurons in vIPFC are often polysensory, responding to images and sounds^{36,37}. Here, we asked if our sampled vIPFC sites showed these properties as well. Visual cortex neurons show poor tuning to stimuli that are not images, therefore we considered the hypothesis that vIPFC neurons with strong visual encoding (*i.e.*, capable of driving image synthesis or encoding prototypes) might also show poor responses to non-visual stimuli. Alternatively, we considered the hypothesis that vIPFC neurons might combine prototype encoding as well as responsiveness to sounds, at least in multiunit signals. To test these hypotheses, we used a simple experimental design: in a given trial, the monkey held fixation for a fraction of a second, then a stimulus—either an image or a sound—was presented for 100 ms (Fig. 1g). If the animal kept fixation during the stimulus, they got a liquid reward. In each trial, each image could be a photograph, line shape, Gabor function, or a sine-wave with a given frequency. We then measured the change in activity for a given neuronal site from the first 50 ms after stimulus onset to the window of 60–200 ms (change in activity was defined as *responsiveness*). We found that our choice of images and sounds were effective at driving activity in sites in either V4, CPB, or vIPFC: in V4, 75% of 32 sites

responded to images with a median response was 17.3 ± 2.3 events/s ($p < 0.02$ per Wilcoxon sign rank paired test, after false discovery rate correction using the Benjamini and Hochberg procedure). While none of the V4 sites showed increased activity to sounds, as a population they were suppressed weakly by sounds (median response -1.3 ± 0.2 , see Supplemental Fig. 4a, b). No individual V4 site was responsive to both modalities. In CPB, 62.5% of 32 sites showed a median increase in response of 5.0 ± 1.6 events/s and 0% of 32 sites showed reliable changes in median firing rate to images (2.0 ± 0.4 events/s); as in V4, no site was responsive to both modalities. Thus, while visual cortex sites were driven primarily by images and auditory cortex sites by sounds, intermingled sites in both cortices also showed a bit of responsiveness to the other stimulus type. This type of polysensory crosstalk has been described in sensory cortex before³⁸.

Compared to sites from visual and auditory cortex, vIPFC sites were more likely to respond to both images and sounds. vIPFC sites reliably showed image-related modulation of at least a few events per second (Monkey C, 71.9% of 32 sites were responsive to images at a P -threshold of 0.02, showing a median response of 5.0 ± 1.0 events/s; Monkey D, 18.8%, median response 2.2 ± 2.1 events/s). Many of these sites also showed sound-related activity of at least a few events per second (Monkey C, 75.0% of 32 sites, median response 3.5 ± 0.8 events/s; Monkey D, 6.2% of 32 sites, though the median response was -0.5 ± 0.4 , so many sites were also suppressed). In both animals, there were sites that were statistically responsive to both images and sounds (18/32 channels in Monkey C, 2/32 channels in Monkey D). When noting the relative activity of sites in response to the images and sounds, we found that the sites in vIPFC that could reliably drive image synthesis fell within the distribution of V4 sites (Supplemental Fig. 4c). Thus, we conclude that vIPFC multiunit sites can drive image synthesis even if they show some tuning to other stimulus modalities.

Discussion

Neurons in vIPFC are strongly interconnected with neurons in the ventral stream, and lesion studies have shown that this is important for visual processing²⁵. Yet what do vIPFC neurons do exactly to improve information processing in the ventral stream? Here, we defined the properties of vIPFC neurons to explore the importance of their feedback to the ventral stream, and to determine if computational models of the visual system should implement PFC-like modules as convolutional layers or more complicated modules. While vIPFC neurons are known to show stimulus-position tuning suggestive of classic RFs¹³ and image selectivity^{7,8,17,28,39} in the absence of task training¹², not all of these properties have been previously and thoroughly conducted in the same neuronal sites and across sensory cortices using targeted image deconstruction. We found a population of vIPFC sites with visual encoding in line with ventral stream neurons: subsets of vIPFC neurons demonstrate stable position-based tuning, which also encode specific low-level visual motifs.

RFs from vIPFC sites were a bit larger and more eccentric than those from V4 sites. RFs throughout the brain tend to be larger in size in more anterior, later stages along the ventral visual stream, while RF size also scaled as a function of eccentricity from the fovea; as such, the vIPFC RF sizes may reflect either or both of these organizational properties. These RFs, like V4 RFs, could best be aligned on retinal coordinates, rather than allocentric positions on the monitor. Additionally, these RFs were associated with *selectivity*: sites responded differentially across diverse images presented at the center of their RFs. While photographs were effective stimuli, we found evidence that these responses could not be accounted for entirely by affective, motivational, or extraretinal processes, as vIPFC neurons could successfully guide adaptive generators – models that create textures lacking objects or other familiar patterns. Though firing rates to synthesized images significantly rose throughout successful experiments, the responses to real-world photographs decreased, suggesting both

V4 and vIPFC sites otherwise show adaptation to repeated stimuli. Visual attributes in the synthesized images were consistent with attributes contained in preferred real-world images. Arguably, the synthetic images were simpler than others we have observed in the ventral stream³⁴, although this is probably due, at least in part, to the eccentric RFs of these sites: the visual system has lower acuity at the periphery than foveal vision does. Finally, as in visual cortex, these representations were stable across months.

Together, these results suggest visual processing beyond the traditional ventral visual stream, with vIPFC potentially functioning as a concurrent or additional stage of *convolutional* processing for visual recognition. Because visual RFs from V1 to IT are implemented as convolutional operations, these results would suggest that a PFC-like layer should also be partially convolutional. At this point, it would not be convincing to argue that vIPFC neurons show visual properties that are uniquely sophisticated or beyond the capabilities of the ventral stream itself. However, as this line of investigation is extended to incorporate other tasks, properties might emerge, justifying a more heterogeneous computational model of the ventral visual stream. Future studies will need to be conducted to determine whether the described visual cortex-like vIPFC neurons are recruited for visual recognition tasks, or are necessary for multisensory integration, the performance of visually-guided behaviors, and/or social communication, allowing the merging of theories of adaptive coding in PFC⁴⁰ with stable, naturalistic functions. One interesting possibility is that the purpose of visual encoding in PFC facilitates visually driven-action selection: studies such as Rushworth et al.⁴¹ have shown that ventrolateral and orbital PFC are important for linking spatially localized visual information (e.g., object identity) with motor response selection⁴². In PFC, it might be more efficient to select actions when visual features are spatially associated with the response. We found that subsets of neurons in vIPFC encoded for coarse shapes independently of task training, suggesting that these neurons reflect a natural association between this specific visual shape and a motor movement – most likely a saccade—since the vIPFC neurons' receptive fields were eccentric.

Limitations of the study

One major limitation in our conclusions is the interpretation of the cluster of sites with reliable RFs. It is a possibility that this cluster reflects the border between two areas. While the literature shows no clear consensus on the areal borders within the region anterior to the arcuate sulcus and posterior/ventral to the principal sulcus, Petrides and Pandya (2001) illustrated that this area holds the confluence of area 8A, 45A, and 45B. The more dorsal aspect of this region encroaches onto the frontal eye field (FEF), although we doubt that our recordings pertain this area, as FEF lies more along the anterior wall of the arcuate sulcus^{43–45}. We also conducted a brief (three-day long) pilot experiment with microstimulation following the parameters and design of Schwedhelm et al.⁴⁶, but did not observe any eye movements. Because our primary goal was to understand the visual encoding of vIPFC neurons, not map out cortical areas, this is a limitation that was built into our design, although the finding invites further investigation.

Another limitation of this study is that we had access to only 32 sites in vIPFC, and consequently we found a small number of vIPFC sites demonstrating visual responses, particularly for understanding systems-wide implications. However, we are encouraged that we found similar results in two monkeys (even though all two-monkey experimental designs are underpowered if one attempts to claim a species-level finding)⁴⁷ and across single- and multiunits.

Methods

All procedures received ethical approval by the Washington University School of Medicine and the Harvard Medical School Institutional Animal Care and Use Committees and conformed to NIH guidelines

provided in the Guide for the Care and Use of Laboratory Animals. All relevant ethical regulations for animal and non-human primate testing and research were followed. Because only male monkeys were available, sex could not be considered in the study design. Animals were purchased through a federally approved vendor.

Animal handling

Before experiments, the monkeys were trained using long periods of operant conditioning, as we have found that this pays off in years of calm and enthusiastic cooperation with by the animals. First, investigators spent two weeks teaching them to track/follow/touch a short stick for fruit and treats. Then the animals were encouraged to permit short stick contact with their collars for rewards (-1-2 weeks) and then to allow a leash to be placed on the collar (-2 weeks). The monkey chairs were present throughout training for familiarization. Once comfortable with the leash, the chair was then secured against the cage entrance and the leash is loosely guided into the chair: the animals were taught that the cage door may open while they have a leash. Rewards were provided when they showed relaxed behavior. After 3-4 days, they were gently but firmly guided into the chair using the leash, and the door closed behind them. After 1-2 days of repeating this in the vivarium, they were moved to the laboratory where they spend time watching TV shows with investigators. For collar placement and imaging, chemical restraint was used (ketamine and dexmedetomidine).

Experimental setup

Experiments were run in two identical experimental rigs, each controlled by a computer running MonkeyLogic (NIMH)⁴⁸. Stimuli were presented on ViewPixx monitors (ViewPixx Technologies, QC, Canada) at a resolution of 1920 × 1080 pixels (120 Hz, 61 cm diagonal); subjects were positioned 58 cm from the monitor during all experiments. Images were presented at the screens' maximum resolution after rescaling to match the size of the image in degrees of the visual field. Eye position was tracked using ISCAN infrared gaze tracking devices (ISCAN Inc., Woburn MA). In all experiments, the animals performed simple fixation tasks (i.e., holding their gaze on a 0.25°-diameter circle, within a -1.2–1.5°-wide window on the center of the monitor for 2–3 s to obtain a drop of liquid reward (water or diluted fruit juice, depending on the subject's preference). Rewards were delivered using a DARIS Control Module System (Crist Instruments, Hagerstown, MD).

Neuronal profiles

Two male rhesus monkeys (C and D, *Macaca mulatta*, 14–15 kg, 9-10 years of age) were each implanted with two 32-channel chronic floating microelectrode arrays in the left cortical hemisphere. Both monkeys were implanted with one 32-channel array in the prefrontal gyrus inferior to the arcuate sulcus and anterior to the principal sulcus (vIPFC); Monkey C had a second array implanted on the prelunate gyrus (visual area V4), while Monkey D had a second array implanted on the upper banks of the superior temporal sulcus, along the caudal parabelt, a region recognized as part of auditory association cortex. Based on standard anatomical descriptions, we refer to these sites as ventrolateral prefrontal cortex (vIPFC), V4, and caudal parabelt (CPB), respectively. Array locations were chosen solely based on sulcal landmarks and on local vasculature, resulting in an unbiased sample of neurons. We use the term *site* to refer to both single- and multi-units; in all experiments, sites were mostly multi-units and hash, with some transient single units across days.

Microelectrode arrays

Arrays were manufactured by Microprobes for Life Sciences (Gaithersburg, MD). Each array consisted of a ceramic base with 32 working electrodes (plus four reference/ground electrodes) made of platinum/iridium, with impedance values of 0.7–1.2 MΩ and lengths of 1.6–2.4 mm (4 mm for grounds and reference electrodes). For array

implantation, the animals were initially anesthetized using ketamine and dexmedetomidine, then maintained using isoflurane. All procedures were conducted under aseptic conditions in accordance to university policy and USDA regulations. Intraoperatively, a craniotomy was performed in the occipito-temporal skull region, followed by a durotomy that allowed visualization of sulcal and vasculature patterns; arrays were inserted into the cortical parenchyma at a rate of ~0.5 mm every three minutes, regulated by visual inspection of tissue dimpling. Arrays were fixed in place using titanium mesh (Bioplate, Los Angeles, CA) over collagen-based dural graft (DuraGen, Integra LifeSciences, Princeton NJ), and cables protected using Flexacryl (Lang Dental, Wheeling, Illinois). Omnetics connectors were housed in custom-made titanium pedestals (Crist Instrument Co., Hagerstown, MD). We used microelectrode arrays because they are very stable in terms of electrode drift, both within individual experimental sessions and across weeks and months; stability is important for the closed-loop experiments because any given synthetic stimulus is presented only once during the evolution experiment. Because the single-unit data yield is lower with chronically implanted arrays, most of our results are relevant to cortical sites, not individual neurons, although every major conclusion was replicated with both single units, multi-unit signals, and visual hash.

Data preprocessing

Data were recorded using Omnplex data acquisition systems (Plexon Inc., Dallas, Texas). At the start of every session, the arrays were connected to Omnplex digital head stages; all channels underwent spike auto-thresholding (incoming signals were recorded if they crossed a threshold amplitude value, determined within each channel, such as 2.75 standard deviations relative to zero-volt baseline). Manual sorting of units took place if the 2D principal component features of detected waveforms showed clearly separable clusters; clusters not overlapping with the main “hash” cluster were described as “single units” if they showed a refractory period; all other signals were described as “multi-units.” All subsequent analyses were performed using Matlab (2020a-2023b).

Visual responsiveness. To determine whether each vIPFC site had a significant change in firing rate in response to stimulation in different regions of the visual field, we divided the first 200 ms following image onset into time windows: the early window comprised the first 1–30 ms after image onset, while the late window period comprised 50–150 ms. We computed the mean firing rate in each time window per experiment, then performed a two-sided Student’s *t*-test between the mean firing rates during the baseline windows and the evoked windows for each site. Finally, we corrected all resulting *p*-values for multiple comparisons using false-discovery rate. We also repeated this analysis with a longer time window for the baseline activity, collected over the fixation period.

Position tuning experiments. To test if vIPFC sites showed position-dependent responses to images, the animals performed a passive fixation task while a 5°-width image was flashed for 100 ms at randomly sampled locations within an invisible grid across the monitor, ranging from (–20, 20)°. Initially, we tested stimuli ranging from 1° in width to 10°, observing that a 5° width typically elicited reliably robust responses from vIPFC neurons and from then used it most frequently (Fig. 1c). We included all experiments using a 5° stimulus to map any potential receptive fields (RFs), and normalized response variability by z-scoring the firing rate across sessions. **Analyses.** For each site, stimulus-evoked responses were defined as the mean spike rate during the first 50–200 ms after image onset (late window) minus the mean spike rate during the first 1–30 ms after image onset (early window). The mean responses to each position were interpolated into a (100x100)-pixel image (using *griddata.m*) and then fit using a 2-D

Gaussian function (*fimgaussfit.m*²⁷). This fitted Gaussian allowed us to estimate the center location (in Cartesian coordinates, where the origin corresponds to the locus of fixation) and approximate widths (in degrees of visual angle) of any potential RFs. A given recording site was defined as positionally tuned if the firing rate were modulated by image position per one-way ANOVA, with *p* < 0.05 after correction for false discovery rate⁴⁹.

Cluster analysis. Sites with reliable RFs appeared close to one another within the layout of the array (Fig. 2b). To quantify this observation, we calculated the mean distance (in μm) between channels with RFs, then performed a permutation test to determine the probability of spatial clustering arising from chance. Sites within the array were equidistantly spaced 400 μm apart in a honeycomb-like formation⁵⁰, with the entire array width spanning 4 mm; therefore, more tightly concentrated sites would have shorter distances than a scattered- or salt-and-pepper arrangement. For each array, we geometrically computed the distances between all possible pairings of sites with significant RFs, then calculated the mean distance across these sites. As there is an absolute distance of 400 μm between any two adjacent sites, we would expect to see mean distances of at least 400 μm. To test whether the observed distances could result due to chance alone, we implemented a permutation test. We randomly resampled with replacement for the same number of significant channels (Monkey C: *n* = 6 channels; Monkey D: *n* = 12 channels) and computed the mean distances between the randomly sampled channels, then replicated this process for 1000 iterations. We then computed the probability of observing the measured true distance by chance: to obtain a *p*-value, we divided the number of instances where the mean distance across randomly sampled sites measured less than or equal to the observed true mean distance by the total number of distances (number of observed + randomized).

Coordinate-transformation experiments. To determine the coordinate system encoded by position-tuned vIPFC neurons, subjects performed trials of active fixation while a 5°-width image was flashed for 100 ms on randomly sampled locations within a proportional invisible cartesian grid, as previously described. However, between trials, the fixation point randomly shifted between three locations: monitor center (as in previous experiments; cartesian coordinates (0°, 0°), now referred to as “center gaze”), as well as 5° (in visual angle) to the right (“rightward gaze,” cartesian coordinates (5°, 0°) with respect to center) and left (“leftward gaze,” cartesian coordinates (–5°, 0°) with respect to center) of monitor center. **Analyses. Inclusion criterion.** We carried out five experiments per monkey, estimating a site’s RF at each of the three gaze directions, yielding three RF maps per site. These experiments were done months after the arrays were implanted, so to control for signal degradation, we had an additional inclusion criterion. We used the independently collected RF-mapping data (from previous months) to make sure the site still had a stable RF in the first place. Specifically, we used the same data in the RF-mapping section to first generate a reference RF per site (*N* = 27 experiments, Monkey C; *N* = 38, Monkey D, sampled across the array’s lifespan). We then compared this reference RF with the center-gaze RF obtained in this new batch of experiments: for each gain-field experiment, we generated RFs for every site from only trials in the center-gaze condition, resulting in five RF maps per site. After interpolation into a 100 x 100 grid (as described in the RF section), no further fitting was performed on the RF maps. We then calculated the Pearson’s correlation between the site’s mean RF and the RF map for each of the five gain field experiments, generating a correlation value for each gain field experiment. An experiment was considered for analyses when the RF map during center gaze trials had at least a 0.6 correlation to the mean RF for that site; this filtered out sessions where the signal was completely obfuscated by noise. Center gaze trials were only used to determine the inclusion criteria, and from

then on were excluded from final analyses. For further analyses, we averaged across the vertical dimension (y-axis) of the visual field, as the experiment only varied the gaze direction across the x-axis.

To measure potential gain effects of gaze direction for each cortical site, we set out to investigate if a given channel's RF changed in retinotopic position or in peak magnitude as a function of gaze condition. To do this, we compared each channel's RF at one gaze condition (-5,0) vs. another gaze condition (+5,0). However, this comparison would only be valid if the channel had an RF in the first place. To determine if a given channel had an RF, we used a third set of trials where the gaze condition was at (0,0). For each gaze condition, we performed a one-way ANOVA test (with stimulus position as the sole factor), correcting the observed *p*-values using false discovery rate tests (*mafdr.m*). We limited our first pass of analysis to channels that showed a significant RF ($p \leq 0.05$ after FDR correction) during the center gaze condition (24 V4 sites in Monkey C, four vIPFC sites in Monkey C, and 13 vIPFC sites in Monkey D, for a total of 17 active vIPFC sites), and then analyzed independent trials from the two remaining gaze conditions.

For some sub-analyses, we estimated putative RF centers using trials presented during center gaze direction. For sites exhibiting significant position tuning during center gaze (one-way ANOVA, stimulus position as sole factor, $p < 0.05$ after correction for false discovery rate), we then selected all trials during which images were presented within 2° vertically, across all horizontal positions, of the estimated RF center. These trials were further indexed by each of the three gaze directions. All stimulus positions were converted from absolute grid coordinates to retinal coordinates by adding the x-coordinate of the fixation location to the x-coordinates of each grid position. For example, a trial presenting an image at (10°,0°) stimulates different retinal positions depending on gaze direction; under center fixation, the image appears at (10°,0°), while during the leftward gaze condition (-5° from center), the image would appear at (5°,0°) with respect to the retina. Following conversion to retinal coordinates, we fit 2-D Gaussian functions to the evoked responses for each sampled stimulus position for each of the three gaze directions. We then performed permutation tests to determine the estimated RF center.

Stimulus selectivity. Subjects performed a fixation task while images were flashed at the center of a given site's RF. Image size was scaled with estimated RF width, with most experiments conducted using stimuli 5- to 10°-wide. Each image was presented for 100 ms, followed by a 100 ms off period during which the subject had to maintain fixation; trials typically consisted of 4–6 images presented sequentially, and images were presented for a minimum of four repetitions. Image sets consisted of visually diverse photographs and artificial stimuli, including images of conspecifics (e.g. macaque faces and full-body macaques in nature), monkey social encounters (e.g., fear grins, monkeys fighting), familiar and unfamiliar humans, foods/beverages (e.g. fruit, water, juice), other animals (e.g. dogs, cats), familiar objects (e.g. toys, caging), artificial stimuli (e.g. lines of different orientations, Gabor patches), and randomly sampled images from the image repository ImageNet⁵¹ (e.g. chairs, appliances). **Analyses.** We performed a one-way ANOVA on vIPFC site responses at the mapped RF, with factor of stimulus identity, correcting for multiple comparisons using false discovery rate.

Neuron-guided image synthesis experiments (XDream). vIPFC and V4 sites exhibiting retinotopic tuning became candidates for closed-loop neuron-guided image synthesis. In these experiments, images were formed from random textures devoid of semantic interpretation, guided by neuronal responses into acquiring shapes that best match the neuron's preferred stimulus features. We refer to stimuli synthesized by a deep generative network under neuronal guidance as *prototypes*³⁴. To assess if vIPFC sites encoded prototypes, subjects

performed a fixation task while stimuli were presented for 100 ms and re-shaped using an optimization algorithm (covariance matrix adaptation evolutionary strategy, CMA-ES). Synthesized images and real-world photographs (particularly of macaque faces, bodies, and social scenes) were interleaved during trials. Experiments ran for dozens of blocks (*generations*). Image synthesis experiments were considered successful when they resulted in images that caused the firing rate to significantly increase from the rate evoked by the first 10 blocks (right-tailed Wilcoxon rank sum between responses to synthetic versus natural images, $p < 0.01$). Additionally, responses to natural images served as a control to ensure responses did not change due to artifactual reasons, such as changes in signal isolation. **Segmentation masks.** To relate prototypes to one another, we used segmentation masks. Masks were generated by calculating the local regional luminance minima of each final prototype, using the morphological operations *opening-by-reconstruction* and *closing-by-reconstruction*⁵². Prototypes were converted from full-color images to grayscale, followed by morphological opening, eroding, reconstructing, closing, dilating, and further reconstructing with a disk-shaped structuring element of a 20-pixel radius. These morphological transformations allowed us to determine global form in prototypes with limited intrusion by any uninformative stochasticity introduced by the generator. After determining the lowest pixel luminance (*i.e.*, the local regional minima) present in the reconstructed grayscale prototype, we generated a segmentation mask by finding all pixels within the prototype having luminance values of at least 70% of the local minima. This simplistic pixelwise approach to segmentation masking allowed for effective “foreground” (darker pixels, frequently bounded forms) and “background” (lighter pixels, like a cleared background that contrasted strongly with the bounded forms) segmentation of visual motifs without clear semantic labels. The distances between vectorized prototype masks (*pdist2.m*) were measured in Pearson correlation space, yielding a correlation matrix where shorter distances between prototype masks indicate a higher correlation in their overlap. **Randomization test.** To determine whether observed mask distances could have been obtained due to chance, we conducted a randomization test that randomly shuffled the mask correlation matrix indices to generate a distribution of randomized mask distances from successful vIPFC prototypes. For each of 1,000 iterations, we shuffled the correlation matrix, saved the median randomized distance of the shuffled correlation matrix per iteration, and then calculated the probability of obtaining the observed vIPFC mask distances through chance alone (number of instances where randomized median distances were shorter than or equal to the observed median mask distance, divided by the total number of iterations and observations).

Variability in prototypes. We estimated differences in the variability of prototypes as a function of area. For each evolution, we measured the dispersion of latent vectors within each generation. Dispersion was computed by obtaining the covariance matrix of all genes per generation ($N - 40$) and measuring the trace of the covariance matrix (the sum of the individual dimension variances). We saw that genetic dispersion started high and then decreased and settled around the 30th generation. To quantify this relationship, we fit an exponential decay function to these curves of the form $y(x) = A * e^{b * x} + c$ to estimate the slope value *b*. To determine if the differences in slopes across areas could arise from the same distribution, we performed a randomization test. Specifically, to estimate the probability that the difference in slope between V4 and vIPFC could arise from the same underlying distribution, we combined the dispersion values from the V4 and vIPFC distributions, and then randomly sampled two groups from the mixed distributions, fitting each group with the exponential decay function, and measuring the difference between both slope values. We repeated this random re-sampling for 500 iterations and compared the actual observed difference to this mixed distribution.

Polysensory experiments. Subjects performed a fixation task. In every experiment, the animal began each trial by fixating on a 0.15° red circle placed at the center of the monitor. After holding fixation within a 1°-radius window for 500 ms, a stimulus was presented for 100 ms, followed by a blank screen for 150 ms. If the animal held fixation for that period, it received a liquid reward; breaking fixation led to a time out. Trials for this task presented a single stimulus before dispensing a drop of liquid reward. The stimulus could be an image or a sound. The image set comprised 10–20 photographs and artificial stimuli including Gabor functions, uniformly colored curved shapes, monkeys, or photographs randomly sampled from the image repository ImageNet. Image size scaled with estimated RF width, with most experiments conducted using stimuli 5–10°-wide. Across experiments, the image position was either the monitor center (0,0)° or placed in locations where we had observed neuronal receptive fields, such as those in V4, or vIPFC (see *RF* section). The sounds were simple tones, the number of unique tones was matched to the number of unique combinations of image/positions in the experiment (N). Per experiment, the set of frequency values was a sample of N between 250 to 2000 Hz, evenly spaced within that range. All stimuli were presented for a minimum of four repetitions each. The reward was water or juice, and its delivery was always associated with an audible click of the dispensing valve opening and closing. We conducted 14 experiments in Monkey C (66 ± 4 unique stimuli per experiment, 20 unique images, 446 unique sounds) and Monkey D (58 ± 6 unique stimuli, 18 unique images, 40 unique sounds). **Analyses.** The goal was to determine if a given array site responded to images, sounds, or both. We had access to 32 channels in V4 (Monkey C), 32 in CPB (Monkey D), and 32 in vIPFC (Monkeys C, D). For each experiment, we measured each site's firing rate activity after stimulus onset, transformed the peristimulus time histogram from events/s into a z-score, and concatenated the z-scored PSTHs across days; the z-score transformation was done to control for day-to-day variability in thresholding). Responsiveness was defined as the change in firing rate activity from the first 50 ms after stimulus onset, to the window of 60–200 ms. Each responsiveness value was concatenated across days. To determine if the mean site responsiveness increased reliably above zero, we performed a Wilcoxon sign-rank test (paired, one-tailed, $N = 11$, 14 for Monkeys C and D) to define the probability that the values arose from a distribution with mean 0. All p -values were corrected for false-discovery rate.

Statistics & reproducibility

The overall study design addressed reproducibility as follows: first, to make a generalizable statement of about neurons in V4 and PFC, neurophysiological recording sites were randomly sampled, Sampling was achieved by placing chronic microelectrode arrays in regions marked solely by sulcal patterns, which are practically universal markers for the macaque monkey, without any previous pre-mapping via imaging. Next, we used two animals to confirm that all major results could be replicated. If results were based on a single animal, they were explicitly marked as such. Every site in an array was tested and used for experiments if they showed physiological responses. Each site was tested multiple times over months to make sure some observations could be replicated over time. The robustness of the statistical tests was informed by having at least 5–6 repetitions of each position in receptive-field mapping and in the selectivity experiments. We used machine learning modeling to confirm that our closed-loop image search algorithm had enough statistical power to achieve high activations from hidden units and then by neurons, for example, by having at least $N = 40$ images per optimization iteration (“generation”). Sites were excluded from analysis if they did not show visual responsiveness (determined by measuring activity in a late time window after image onset). Every animal and cortical site was used as both their own control and

treatment group (pre-optimization, and response-optimized). Blinding could not be used.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All correspondence and material requests should be addressed to the corresponding author (C.R.P). Data generated in this study have been deposited in the Zenodo database with <https://doi.org/10.5281/zenodo.12565507>. The data are available under full access, access can be obtained by downloading the.mat files. The raw data is under ongoing analysis and available by request. The processed data is available at Zenodo. The data used in this study is available in the Zenodo database with <https://doi.org/10.5281/zenodo.12565507>. Source data are provided with this paper.

Code availability

Code and links to useful repositories that can be used to reproduce most of the results and figures (see Data Availability statement) are deposited in a Github repository (<https://github.com/Poncelab/visual-PFC>), 10.5281/zenodo.12566772.

References

1. Van Essen, D. C., Anderson, C. H. & Felleman, D. J. Information processing in the primate visual system: an integrated systems perspective. *Science* **255**, 419–423 (1992).
2. Riesenhuber, M. & Poggio, T. Hierarchical models of object recognition in cortex. *Nat. Neurosci.* **2**, 1019–1025 (1999).
3. Yamins, D. L. K. et al. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl Acad. Sci. USA* **111**, 8619–8624 (2014).
4. Serre, T. Deep Learning: The Good, the Bad, and the Ugly. *Annu. Rev. Vis. Sci.* **5**, 399–426 (2019).
5. Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G. & Mishkin, M. The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends Cogn. Sci.* **17**, 26–49 (2013).
6. Ungerleider, L. G., Gaffan, D. & Pelak, V. S. Projections from inferior temporal cortex to prefrontal cortex via the uncinate fascicle in rhesus monkeys. *Exp. Brain Res.* **76**, 473–484 (1989).
7. Scialajda, S., Wilson, F. & Goldman-Rakic, P. Face-selective neurons during passive viewing and working memory performance of rhesus monkeys: evidence for intrinsic specialization of neuronal coding. *Cereb. Cortex N. Y. N.* **1991** **9**, 459–475 (1999).
8. Riley, M. R., Qi, X.-L. & Constantinidis, C. Functional specialization of areas along the anterior–posterior axis of the primate prefrontal cortex. *Cereb. Cortex* **27**, 3683–3697 (2017).
9. Cong, S. & Zhou, Y. A review of convolutional neural network architectures and their optimizations. *Artif. Intell. Rev.* **56**, 1905–1969 (2023).
10. Erhan, D., Bengio, Y., Courville, A. & Vincent, P. Visualizing Higher-Layer Features of a Deep Network. *Tech. Rep. Univeristé Montr.* (2009).
11. Dosovitskiy, A. & Brox, T. Generating Images with Perceptual Similarity Metrics based on Deep Networks. *Adv. Neural Inf. Process. Syst. NIPS* (2016).
12. Pu, S., Dang, W., Qi, X.-L. & Constantinidis, C. Prefrontal neuronal dynamics in the absence of task execution. <https://doi.org/10.1101/2022.09.16.508324> (2022).
13. Suzuki, H. & Azuma, M. Topographic studies on visual neurons in the dorsolateral prefrontal cortex of the monkey. *Exp. Brain Res.* **53**, 47–58 (1983).

14. Mikami, A., Ito, S. & Kubota, K. Visual response properties of dorsolateral prefrontal neurons during visual fixation task. *J. Neurophysiol.* **47**, 593–605 (1982).
15. Viswanathan, P. & Nieder, A. Comparison of visual receptive fields in the dorsolateral prefrontal cortex and ventral intraparietal area in macaques. *Eur. J. Neurosci.* **46**, 2702–2712 (2017).
16. Viswanathan, P. & Nieder, A. Visual Receptive Field Heterogeneity and Functional Connectivity of Adjacent Neurons in Primate Frontoparietal Association Cortices. *J. Neurosci.* **37**, 8919–8928 (2017).
17. Tsao, D. Y., Schweers, N., Moeller, S. & Freiwald, W. A. Patches of face-selective cortex in the macaque frontal lobe. *Nat. Neurosci.* **2008** *118* **11**, 877–879 (2008).
18. Kobatake, E. & Tanaka, K. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J. Neurophysiol.* **71**, 856–867 (1994).
19. Bardon, A., Xiao, W., Ponce, C. R., Livingstone, M. S. & Kreiman, G. Face neurons encode nonsemantic features. *Proc. Natl. Acad. Sci. USA* **119**, e2118705119 (2022).
20. Olah, C., Mordvintsev, A. & Schubert, L. Feature Visualization. *Distill* **2**, e7 (2017).
21. Yamane, Y., Carlson, E. T., Bowman, K. C., Wang, Z. & Connor, C. E. A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nat. Neurosci.* **11**, 1352–1360 (2008).
22. Goodfellow, I. et al. Generative Adversarial Nets. in *Advances in Neural Information Processing Systems* (eds. Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. & Weinberger, K. Q.) 27 (Curran Associates, Inc, 2014).
23. Brock, A., Donahue, J. & Simonyan, K. Large Scale GAN Training for High Fidelity Natural Image Synthesis. *7th Int. Conf. Learn. Represent. ICLR 2019* 1–35 (2018).
24. Karras, T., Laine, S. & Aila, T. A style-based generator architecture for generative adversarial networks. in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* vols 2019-June 4396–4405 (IEEE Computer Society, 2019).
25. Kar, K., Kubilius, J., Schmidt, K., Issa, E. B. & DiCarlo, J. J. Evidence that recurrent circuits are critical to the ventral stream’s execution of core object recognition behavior. *Nat. Neurosci.* **22**, 974–983 (2019).
26. Petrides, M. & Pandya, D. N. Dorsolateral prefrontal cortex: comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection patterns. *Eur. J. Neurosci.* **11**, 1011–1036 (1999).
27. Orloff, N. Fit 2D Gaussian with Optimization Toolbox. *MATLAB Central File Exchange* <https://www.mathworks.com/matlabcentral/fileexchange/41938-fit-2d-gaussian-with-optimization-toolbox> (2020).
28. Haile, T., Bohon, K., Romero, M. & Conway, B. Visual stimulus-driven functional organization of macaque prefrontal cortex. *NeuroImage* **188**, 427–444 (2019).
29. Desimone, R., Albright, T. D., Gross, C. G. & Bruce, C. Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci.* **4**, 2051–2062 (1984).
30. Campbell, F. W., Cleland, B. G., Cooper, G. F. & Enroth-Cugell, C. The angular selectivity of visual cortical cells to moving gratings. *J. Physiol.* **198**, 237–250 (1968).
31. Gross, C. G., Bender, D. B. & Rocha-Miranda, C. E. Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science* **166**, 1303–1306 (1969).
32. Pasupathy, A. & Connor, C. E. Responses to contour features in macaque area V4. *J. Neurophysiol.* **82**, 2490–2502 (1999).
33. Hegd , J. & Van Essen, D. C. A comparative study of shape representation in macaque visual areas v2 and v4. *Cereb. Cortex N. Y. N.* **1991** **17**, 1100–1116 (2007).
34. Rose, O., Johnson, J., Wang, B. & Ponce, C. R. Visual prototypes in the ventral stream are attuned to complexity and gaze behavior. *Nat. Commun.* **2021** *121* **12**, 1–16 (2021).
35. Wang, B. & Ponce, C. R. Tuning landscapes of the ventral stream. *Cell Rep.* **41**, 111595 (2022).
36. Sugihara, T., Diltz, M. D., Averbeck, B. B. & Romanski, L. M. Integration of Auditory and Visual Communication Information in the Primate Ventrolateral Prefrontal Cortex. *J. Neurosci.* **26**, 11138–11147 (2006).
37. Romanski, L. M. & Hwang, J. Timing of audiovisual inputs to the prefrontal cortex and multisensory integration. *Neuroscience* **214**, 36–48 (2012).
38. Maunsell, J. H. R., Sclar, G., Nealey, T. A. & DePriest, D. D. Extraretinal representations in area V4 in the macaque monkey. *Vis. Neurosci.* **7**, 561–573 (1991).
39. Romanski, L. M. & Diehl, M. M. Neurons responsive to face-view in the primate ventrolateral prefrontal cortex. *Neuroscience* **189**, 223–235 (2011).
40. Duncan, J. An adaptive coding model of neural function in prefrontal cortex. *Nat. Rev. Neurosci.* **2**, 820–829 (2001).
41. Rushworth, M. F. et al. Attentional selection and action selection in the ventral and orbital prefrontal cortex. *J. Neurosci.* **25**, 11628–11636 (2005).
42. Lee, D., Rushworth, M. F. S., Walton, M. E., Watanabe, M. & Saka-gami, M. Functional Specialization of the Primate Frontal Cortex during Decision Making. *J. Neurosci.* **27**, 8170–8173 (2007).
43. Robinson, D. A. & Fuchs, A. F. Eye movements evoked by stimulation of frontal eye fields. *J. Neurophysiol.* **32**, 637–648 (1969).
44. Bruce, C. J., Goldberg, M. E., Bushnell, M. C. & Stanton, G. B. Primate frontal eye fields. II. Physiological and anatomical correlates of electrically evoked eye movements. *J. Neurophysiol.* **54**, 714–734 (1985).
45. Schall, J. D. Neuronal activity related to visually guided saccades in the frontal eye fields of rhesus monkeys: comparison with supplementary eye fields. *J. Neurophysiol.* **66**, 559–579 (1991).
46. Schwedhelm, P., Baldauf, D. & Treue, S. Electrical stimulation of macaque lateral prefrontal cortex modulates oculomotor behavior indicative of a disruption of top-down attention. *Sci. Rep.* **7**, 17715 (2017).
47. Fries, P. & Maris, E. What to do if N is two? *J. Cogn. Neurosci.* **34**, 1114–1118 (2021).
48. Hwang, J., Mitz, A. R. & Murray, E. A. NIMH MonkeyLogic: Behavioral control and data acquisition in MATLAB. *J. Neurosci. Methods* **323**, 13–21 (2019).
49. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
50. Musallam, S., Bak, M. J., Troyk, P. R. & Andersen, R. A. A floating metal microelectrode array for chronic implantation. *J. Neurosci. Methods* **160**, 122–127 (2007).
51. Russakovsky, O. et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **115**, 211–252 (2015).
52. Eddins, S. Opening by reconstruction. *Steve on Image Processing with MATLAB* <https://blogs.mathworks.com/steve/2008/07/14/opening-by-reconstruction/> (2008).

Acknowledgements

We thank Mary Carter and Elizabeth Cleaveland for their technical help with data acquisition and surgical procedures. We also thank Drs. Rick Born, Margaret Livingstone, and John Assad for their comments during experimental design and manuscript preparation. Research reported in this publication was supported by the Common Fund of the National Institutes of Health under award 1DP2EY035176-01 (C.R.P.), the David

and Lucile Packard Foundation (2020-71377, C.R.P.) by the National Eye Institute under awards 1F31EY034784-01A1 (O.R.) and Core Grant 2P30EY012196-26 (Margaret S. Livingstone).

Author contributions

Conceptualization, Methodology, Experimentation, Visualization, Investigation: O.R. and C.R.P. Writing - Original draft preparation: O.R. Supervision: C.R.P. Writing - Reviewing and Editing: O.R. and C.R.P.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-51441-3>.

Correspondence and requests for materials should be addressed to Carlos R. Ponce.

Peer review information *Nature Communications* thanks Peichao Li, and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024