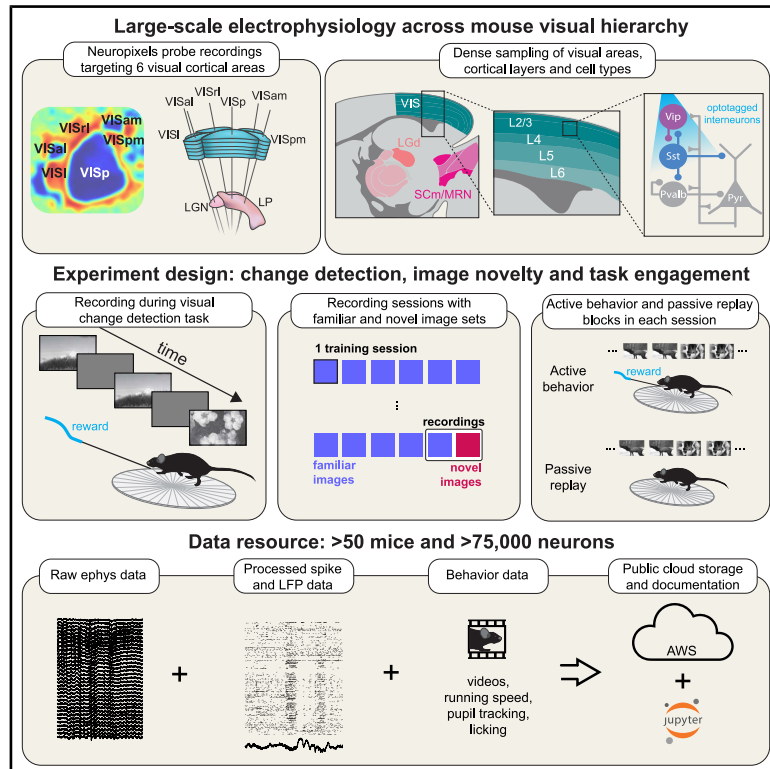


Map of spiking activity underlying change detection in the mouse visual system

Graphical abstract



Authors

Corbett Bennett, Samuel D. Gale, Gregory Heller, ..., Séverine Durand, Christof Koch, Shawn R. Olsen

Correspondence

shawno@alleninstitute.org

In brief

Large-scale Neuropixels recordings across the mouse visual system during an image-change detection task map how sensory, motor, and task engagement signals are distributed across the cortex, thalamus, and midbrain. Task and novelty signals differentially modulate activity along this hierarchy, and combined decoding and optogenetics suggest mice detect changes via an adaptation-based strategy within a critical time window.

Highlights

- Spike data from >75,000 neurons across mouse visual hierarchy
- Change detection behavior with stimulus novelty and passive replay manipulations
- Deep sampling of cortical layers and cell types reveals functional differences
- Neural data constrain where, when, and how change detection is implemented



Resource

Map of spiking activity underlying change detection in the mouse visual system

Corbett Bennett,¹ Samuel D. Gale,¹ Gregory Heller,² Tamina K. Ramirez,^{3,4} Hannah Belski,¹ Alex Piet,¹ Omid Zobeiri,⁵ Adam Amster,⁶ Anton Arkhipov,⁵ Alex Cahoon,⁷ Shiella Caldejon,¹ Mikayla Carlson,¹ Linzy Casal,⁷ Scott F. Daniel,⁶ Colin Farrell,⁷ Marina Garrett,¹ Ryan Gillis,¹ Conor Grasso,⁷ Ben J. Hardcastle,¹ Ross Hytnen,⁷ Tye Johnson,¹ Peter Ledochowitsch,⁸ Quinn L'Heureux,⁷ Dana Mastrovito,¹ Ethan G. McBride,¹ Stefan Mihalas,¹ Chris Mochizuki,⁷ Christopher B. Morrison,⁶ Chelsea Nayan,⁷ Nhan-Kiet Ngo,⁶ Kat North,⁷ Douglas R. Ollerenshaw,¹ Ben Ouellette,¹ Paul Rhoads,⁷ Kara Ronellenfitch,⁶ Martin Schroedter,⁷ Joshua H. Siegle,¹ Cliff Slaughterbeck,⁷ David Sullivan,⁹ Jackie Swapp,¹ Michael Taormina,⁶ Wayne Wakeman,⁷ Xana Waughman,¹⁰ Allison Williford,¹ John W. Phillips,⁷ Peter A. Groblewski,¹ Séverine Durand,¹ Christof Koch,⁷ and Shawn R. Olsen^{1,11,*}

¹Allen Institute/Neural Dynamics, Seattle, WA 98109, USA

²Department of Brain and Cognitive Sciences, Picower Institute for Learning and Memory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

³Department of Neuroscience, Columbia University, New York, NY 10032, USA

⁴Department of Neurology, Columbia University, New York, NY 10032, USA

⁵Allen Institute/Brain Health, Seattle, WA 98109, USA

⁶Allen Institute/Brain Science, Seattle, WA 98109, USA

⁷Allen Institute, Seattle, WA 98109, USA

⁸General Sense, Alachua, FL 32615, USA

⁹SIXTY D, Bend, OR 97702, USA

¹⁰University of Arizona College of Medicine, Tucson, AZ 85724, USA

¹¹Lead contact

*Correspondence: shawno@alleninstitute.org

<https://doi.org/10.1016/j.cell.2026.06.025>

SUMMARY

Visual behavior requires coordinated activity across hierarchically organized brain circuits. Understanding this complexity demands datasets that are both large-scale (sampling many areas) and dense (recording many neurons in each area). Here, we present a database of spiking activity across the mouse visual system—including the cortex, thalamus, and midbrain—while mice perform an image change detection task. Using Neuropixels probes, we record from >75,000 high-quality units in 54 mice, mapping area-, cortical-layer-, and cell-type-specific coding of sensory and motor information. Modulation by task engagement increased across the thalamocortical hierarchy but was strongest in the midbrain. Novel images recruited an expanded cortical population and modulated late cortical (but not thalamic) responses. Population decoding and optogenetics identified a critical time window for change detection and were consistent with mice using an adaptation-based rather than image-comparison strategy. This comprehensive resource provides a valuable substrate for understanding sensorimotor computations in neural networks.

INTRODUCTION

A central challenge in neuroscience is to understand how diverse neurons, connected through distributed circuits, transform sensory information into behavior. The mammalian visual system exemplifies this complexity—visual processing engages hierarchical interactions across many brain regions, including the cortex, thalamus, and midbrain, with each local circuit containing heterogeneous cell types that coordinate through intricate connectivity to guide behavior. These visual circuits must be both stable for learned associations and flexible for novel experiences. To understand how spiking activity in the visual system implements adaptable sensorimotor computations, we need comprehensive data-

sets that measure the activity of multiple brain regions and cell types simultaneously during visually guided behavior.

Technical advances, including Neuropixels probes, have revolutionized our ability to record from large populations of neurons.^{1–6} Several large-scale experimental efforts in mice have already begun to systematically map brain-wide neural activity during sensory processing and decision-making.^{7–10} However, these recordings are typically serial (one or two probes at a time) and shallow (sampling tens to hundreds of neurons from each area). Understanding how brain networks coordinate activity on the level of cell types and cortical layers requires a complementary approach: dense, simultaneous recordings from interconnected regions.



At the Allen Institute, we previously built standardized physiological pipelines for dense sampling of the thalamocortical visual system.^{4,11–13} Our prior Visual Coding Neuropixels dataset surveyed stimulus-evoked spiking activity in passively viewing mice,⁴ supporting numerous studies of cortical, thalamic, and hippocampal processing^{14–18} (see review by de Vries et al.¹⁹), but lacked a behavioral task.

Here, we describe a large-scale open dataset in which we used multi-regional Neuropixels recordings to map spiking activity across the mouse visual system while mice performed a visual change detection task. Change detection is a fundamental visual computation essential for survival, enabling rapid decisions in dynamic environments. Recent work in the mouse has begun to identify circuits in the visual system involved in this process,^{8,12,20–25} including a complementary study by the Allen Institute using two-photon calcium imaging to measure cortical cell populations in the same task.¹² However, a comprehensive map of spiking activity across multiple levels of the visual thalamocortical hierarchy is lacking.

Our experimental design included (1) simultaneous recordings from six visual cortical areas at retinotopically matched locations,¹ (2) dense sampling across cortical layers and identification of genetically defined interneuron subclasses through optotagging, (3) recordings from visual thalamic and midbrain regions involved in visual decision-making, and (4) a multi-epoch experiment structure capturing neural activity during behavior, passive stimulus replay, and processing of both novel and familiar stimuli.

This resource—the Visual Behavior Neuropixels dataset—is readily available on public cloud platforms (Figure S1; Table S1) with comprehensive tutorials and documentation (Table S2). In this study, we (1) characterize sensory and motor encoding across areas, layers, and cell types, (2) describe response modulation by task engagement and stimulus novelty, and (3) evaluate the timing and computational strategies for change detection.

RESULTS

Survey of spiking activity during visual change detection

We recorded spiking data simultaneously from six Neuropixels probes, each targeting a distinct visual cortical region (Figure 1A; targets: VISp, VISl, VISal, VISrl, VISam, and VISpm).^{1,4} Before the experiment, intrinsic signal imaging (ISI) was used to map the retinotopy of each cortical area. These maps then guided probe insertions to record from neurons with receptive fields near the center of the stimulus monitor (Figures 1A and 1B). In total, this dataset includes 76,091 high-quality units spanning cortical and subcortical regions from 54 mice over 103 behavioral sessions (Figures 1C and 1D).

Each session comprised five experimental epochs: (1) active behavior, during which mice performed the visual change detection task; (2) passive stimulation with gabor stimuli to map receptive fields; (3) spontaneous activity with no visual stimulation; (4) full-field luminance changes; and (5) “passive replay” of the identical image sequence during active behavior but with the lick spout retracted (Figure 1E). After these epochs, we optotagged cortical interneurons expressing ChR2 in SST-

cre;Ai32 or VIP-cre;Ai32 mice (Figure 1E; Figure S2; 434 SST+, 27 VIP+ units).

During the change detection task, mice viewed a series of briefly presented natural images and received a water reward for licking when the identity of the image changed (Figure 1F).^{20,26} During training, mice progressed through several stages (Figure 1G), reaching proficiency in 3–4 weeks (15.7 ± 6.9 sessions, mean $\pm$ SD). Once proficient, mice licked with high probability on go trials (when a change occurred) compared with catch trials (when a change did not occur) (Figure 1H; hit rate: 0.82 ± 0.08 , false alarm rate: 0.17 ± 0.09 , mean $\pm$ SD, $n = 53$ mice; $p = 2.4 \times 10^{-10}$, Wilcoxon signed-rank test; $d\text{-prime} = 2 \pm 0.4$, mean $\pm$ SD). Though previous work has shown that mice can use both visual cues and timing information to solve this task,²⁷ mice in this dataset overwhelmingly used visual cues (see STAR Methods).

In a typical experiment, we recorded spiking activity from ~ 700 high-quality units (739 ± 183 , mean $\pm$ SD) and collected several streams of behavioral data, including running speed, lick spout contacts, and videos of the eye, face, and body (Figure 1I).

Encoding of sensory and action variables

To characterize how units encoded visual stimuli, task events, and motor output during the change detection task, we fit a general linear model (GLM) to the spiking activity of each unit during the active behavior epoch. The GLM explains each unit's firing rate by a linear combination of 14 time-dependent kernels: eight kernels for the eight unique images shown during a given session (visual features); two kernels for hits and misses (task features); three kernels for licks, running speed, and pupil radius (action features); and a kernel for image omissions (Figure 2A). The model fit captured 6.5% of the total variance on average across all areas and 11% for visual cortical units (all: $6.5\% \pm 9.8\%$, mean $\pm$ SD; VIS: $11\% \pm 11.8\%$; cross-validated performance; Figure S3A), consistent with similar efforts.^{5,12}

To assess the degree to which units encoded sensory or action features, we compared the performance of the full model to one for which either the image kernels were removed (visual dropout) or the lick kernel was removed (lick dropout). Dropout scores measure the unique contribution of each feature to the variance explained for each unit. Visual dropout scores were highest in visual cortical and thalamic regions, with VISp and lateral visual areas VISl and VISal scoring higher than anterior/medial visual areas VISrl, VISpm, and VISam (Figure S3B). In the superior colliculus motor layers (SCm; intermediate and deep gray layers) and the midbrain reticular nucleus (MRN), $\sim 11\%$ of units had high visual dropout scores ($>1\%$ change in total variance explained) and 29% had high lick-dropout scores (Figure S3B), consistent with the premotor function of these midbrain structures.^{28,29} We noted a rough dorsal/ventral topography of visual/motor encoding in SCm and MRN (Figures S3C and S3D). A small fraction (1%) of visual cortical units exhibited high lick-dropout scores. These units were more common in higher visual areas and were biased to deep cortical layers (Figures S3E and S3F).

Two units may have similar dropout scores for a given feature but respond with different dynamics. To gain insight into response

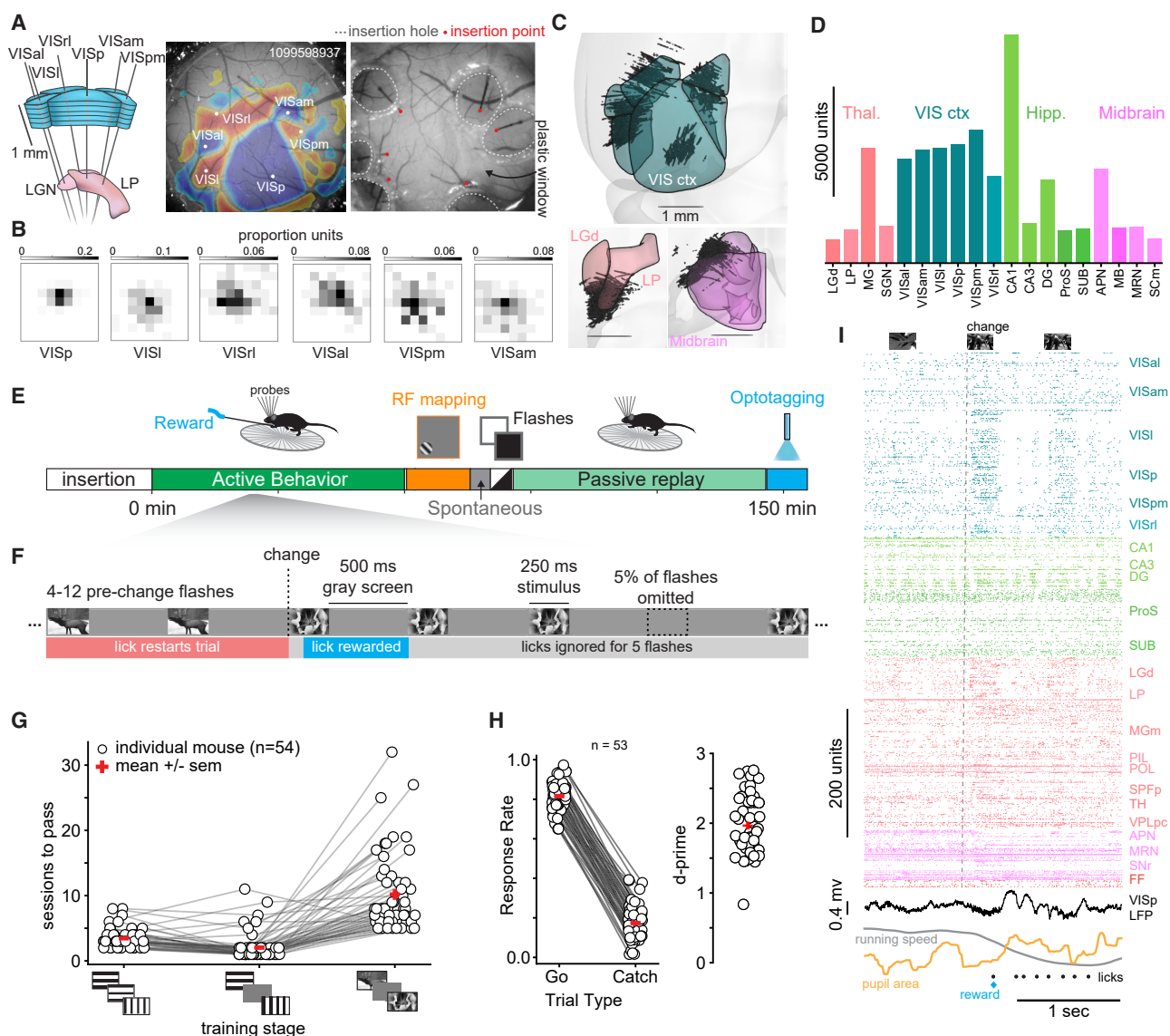


Figure 1. Survey of activity in the visual cortex, thalamus, and midbrain during visual change detection behavior

(A) Left: recording configuration with six Neuropixels probes. Middle: ISI sign map identifying cortical area boundaries to guide probe insertions. White circles indicate insertion sites. Right: image of the brain surface during recording.

(B) 2D histograms of population receptive field centers in each cortical area across experiments. Each bin spans 10° .

(C) Locations of recorded units relative to Allen Common Coordinate Framework (CCF) volumes. Top: visual cortical areas. Bottom: left, thalamus (LGd and LP); right, midbrain (SCm and MRN). Scale bars, 1 mm.

(D) Unit counts across all areas with at least 1,000 units.

(E) Recording session timeline.

(F) Timeline for example trial during the change detection task.

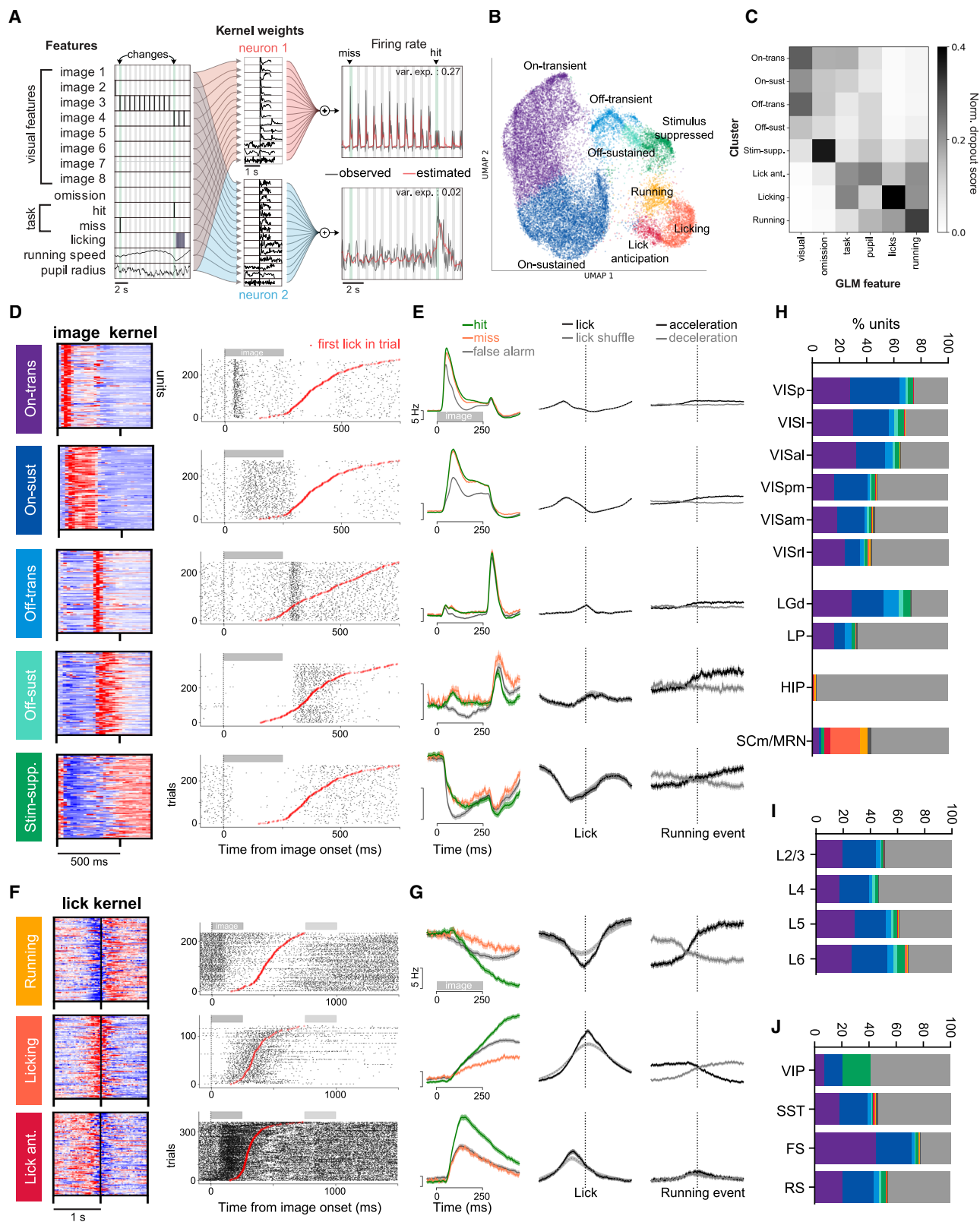
(G) Number of sessions to pass the three stages of behavioral training for each mouse (white circles). Red crosses show mean $\pm$ SEM across mice. One mouse from (G) failed to engage in the task during recording and was excluded from (H).

(H) Left: response rates for change trials (go) and non-change trials (catch) across mice (white circles). Rates are calculated after combining trials across both recording days. Right: d-prime values calculated from go and catch response rates. Red crosses show mean $\pm$ SEM across mice. One mouse from (G) failed to engage in the task during recording and was excluded from (H).

(I) Spike raster showing hit trial in example session. Dotted line denotes image change. VISp LFP (black), running speed (gray), pupil area (gold), reward (blue), and licks (black dots) are shown beneath raster.

dynamics, we clustered units based on their learned time-dependent kernels (STAR Methods). This yielded 11 clusters: for convenience, we label 5 of these clusters “sensory clusters” since their

activity was predominately (though not exclusively) explained by sensory features; similarly, 3 “action clusters” were predominately explained by action features; and 3 additional clusters



(legend on next page)

were either too small or too weakly associated with the GLM variables to be analyzed in detail (Figures 2B and 2C; kernels and distributions of explained variance are shown in Figure S4). The GLM accurately captured the mean firing rate patterns for hit, miss, false alarm, and correct reject trials, validating the kernels as a basis for functional clustering (Figures S5A and S5B). Additional features extracted from behavioral videos had only a marginal explanatory benefit (Figure S5C). Cross-validation with a bootstrapped random forest classification method confirmed cluster robustness (Figures S5D and S5E).

The 5 sensory clusters consisted of (1) an “on-transient” cluster with a sharp response at stimulus onset; (2) an “on-sustained” cluster with activity sustained throughout the stimulus presentation; (3) an “off-transient” cluster with a sharp response at stimulus offset; (4) an “off-sustained” cluster; and (5) a “stimulus-suppressed” cluster, which was inhibited by the visual stimulus and showed ramping activity during inter-stimulus intervals and stimulus omissions (Figures 2D and 2E; Figures S4 and S6A). Sensory encoding units were most common in the visual cortex and thalamus but were present in every area recorded, with “on” outnumbering “off” units (Figure 2H). Midbrain sensory units were mostly on-transient (Figure 2H). Within the cortex, stimulus-suppressed units were common in the deeper layers and were particularly enriched in VIP+ interneurons (Figures 2I and 2J; Figure S6B). Notably, units in this cluster showed stronger modulation by motor features than units in the other sensory clusters (Figure 2C; Figure S4B).

As a population, sensory cluster units did not show significant lick-aligned activity relative to a shuffled control but were weakly modulated by running (Figure 2E). Consistent with previous studies,^{30,31} VIP and SST cells in sensory clusters showed differential running modulation (Figure S3H). In the midbrain, neurons in sensory clusters had more mixed encoding of sensory and motor information compared with visual cortical neurons (Figure S3G).

The three action clusters consisted of (1) a “running” cluster with activity anticipating locomotion; (2) a “licking” cluster with activity during licking bouts; and (3) a “lick-anticipation” cluster whose activity preceded licking (Figures 2F and 2G; Figure S4). Action cluster units were most common in the midbrain (Figure 2H). In the visual cortex, these units were rare and biased toward anterior/medial higher visual areas VISam and VISrl (Figure 2H) and deep cortical layers (Figure 2I).

Lick-anticipation units in SCm/MRN appeared to reflect the decision to lick. These units were strongly active during stimuli

that elicited licking (hits and false alarms; Figure S7A) and responded to stimuli at short latency (~50 ms) during active behavior but not passive viewing (Figure S7A). Baseline activity was elevated before stimulus presentations that elicited false alarms (Figures S7A and S7C) and anticorrelated with response time (Figures S7G and S7I). Lick-anticipation units—but not lick units—also showed ramping activity that increased with time from the last lick bout, mirroring the propensity for mice to commit false alarms after long inter-lick-bout intervals (Figures S7A–S7F).

Critical time window for change detection decision

To investigate the chain of processing leading from image presentation to the lick response, we evaluated when sensory information (image identity and image change) and action information (licking) could be decoded from spiking activity in each area (Figure 3A). We used sensory cluster units to decode image identity and image change and action cluster units to decode licking (Figures S8A–S8C compare decoding performance in sensory and action clusters). Because sensory and/or action units were sparse in some areas, we pooled units across sessions, creating pseudopopulations of 100 units for decoding (STAR Methods). Baseline activity was subtracted to emphasize stimulus-driven changes in activity. For each decoder, we selected training data to isolate the variable of interest (Figure 3B).

Image identity information was present shortly after the visual latency and propagated along the visual hierarchy, with LGd leading (latency to half max: 38 ms), followed by VISp and the lateral areas (44–46 ms), then the medial areas VISpm and VISam (~53 ms), and finally SCm/MRN (64 ms) (Figures 3A and 3C).

Decoding image *change* was delayed relative to image identity (earliest change decoding: 53 ms; earliest identity decoding: 38 ms) and compressed across areas (change range, 12 ms; identity range, 25 ms; Figure 3C). Notably, higher-order visual thalamic area LP led all areas in change decoding (latency 53 ms), followed by LGd/VISp and lateral areas VISl/VISal (~60 ms), and lastly the medial areas VISpm/VISam together with SCm/MRN (~64 ms). LP led VISp both when aggregating units across sessions ($p = 0.001$, permutation test) and within sessions for which both areas were recorded (Figures S8D and S8E; $p = 5.25e-3$, Wilcoxon signed-rank test, $n = 14$ sessions). SCm/MRN was the only area lacking a delay between change and image decoding latencies.

Licks could be decoded reliably from SCm/MRN action neurons ~120 ms after image onset, placing an upper bound on the time

Figure 2. GLMs fit to single units reveal diverse sensory and action coding

- (A) Left: GLM schematic. Middle: kernel fits for two example neurons (red and blue shading). Right: observed (gray) and estimated (red) firing rates for example neurons. Kernels are convolved with features and summed to estimate firing rates.
- (B) Uniform manifold approximation and projection (UMAP) embedding of kernel fits for all recorded units colored by cluster label.
- (C) Normalized feature dropout scores across clusters. Matrix is normalized so columns add to 1.
- (D) Left: preferred-image kernels for 100 example neurons in each sensory cluster. Right: spike raster for one example neuron in each cluster aligned to the preferred image onset for non-change stimuli for which the mouse responded. Trials have been ordered by response time (red symbols).
- (E) Population-averaged firing rates for each cluster aligned to image onset for hit, miss, and false alarm trials (left); lick response (middle; response times were shuffled across trials, and response was recalculated to produce gray trace); and running events (right; acceleration from stillness in black; deceleration to stillness in gray). Running events were taken from passive replay epoch to avoid correlation with licking. Shaded regions indicate SEM across units.
- (F and G) Same as (D) and (E) but for action clusters with kernels aligned to lick response times. Time window is extended to show motor response.
- (H–J) For each region (H), cortical layer (I), and cortical cell class (J), the percentage of units belonging to each cluster. Light gray bar indicates units for which the GLM did not explain a significant fraction of variance (<5%). Dark gray indicates units in three clusters excluded from further analysis.

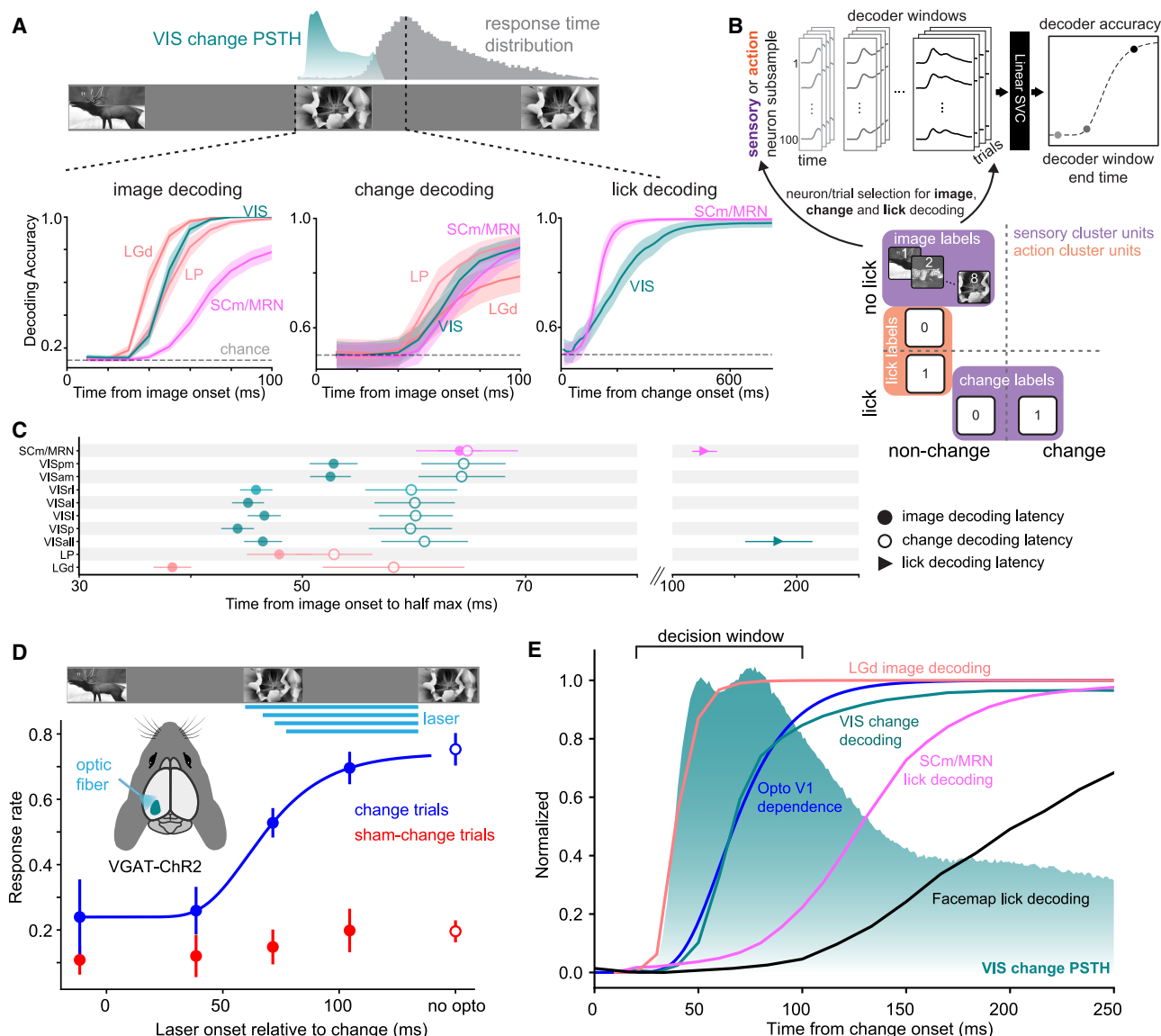


Figure 3. Decoding, behavior, and perturbations reveal a critical window for change detection

(A) Top: schematic depicting relative timing of image change onset, population response in visual cortex (teal), and lick response time distribution across mice (gray). Bottom: decoding time courses across areas for image, change, and lick decoding. Shaded regions indicate standard deviation across pseudopopulations.

(B) Decoding strategy: each timestep on the decoding curve (top right) represents the performance of a linear decoder trained on activity in a window ranging from stimulus onset to that time point. Sensory cluster neurons were used for image and change decoding. Action clusters were used for lick decoding. Pseudotrials were created by sampling 100 neurons for each area/cluster combination across sessions. Bottom: trial types selected to decode image, change, and licking.

(C) Mean decoding latency (time to half max) across neuron subsamples for image (filled circles), change (open circles), and lick (filled triangles) decoding. Error bars represent standard deviation across pseudopopulations. "VISall" includes neurons from all visual cortical areas combined.

(D) Top: VISp silencing strategy; light was delivered at four intervals relative to stimulus onset. Bottom: mean lick response rates ($n = 4$ mice, 15 sessions) as a function of laser onset time for change trials (blue) and catch trials (red). Error bars indicate SEM across mice.

(E) Selected decoding curves are normalized to the maximum within 500 ms from change onset and plotted with the visual cortex change stimulus response to highlight the critical window during which mice compute change and decide to lick. Black curve represents the decoding of the lick response from the first 500 singular vectors of the face video.

window during which image information is used to inform the decision to lick (Figures 3A and 3C). Decoding licks from action neurons in the visual cortex lagged the midbrain by ~ 60 ms ($p = 0.001$; permutation test), possibly reflecting an efferent signal.

Action neurons were too sparse in individual visual cortical or thalamic regions to decode licking by this method.

This decoding analysis is correlational and estimates when, and to what degree, information present in spiking activity can

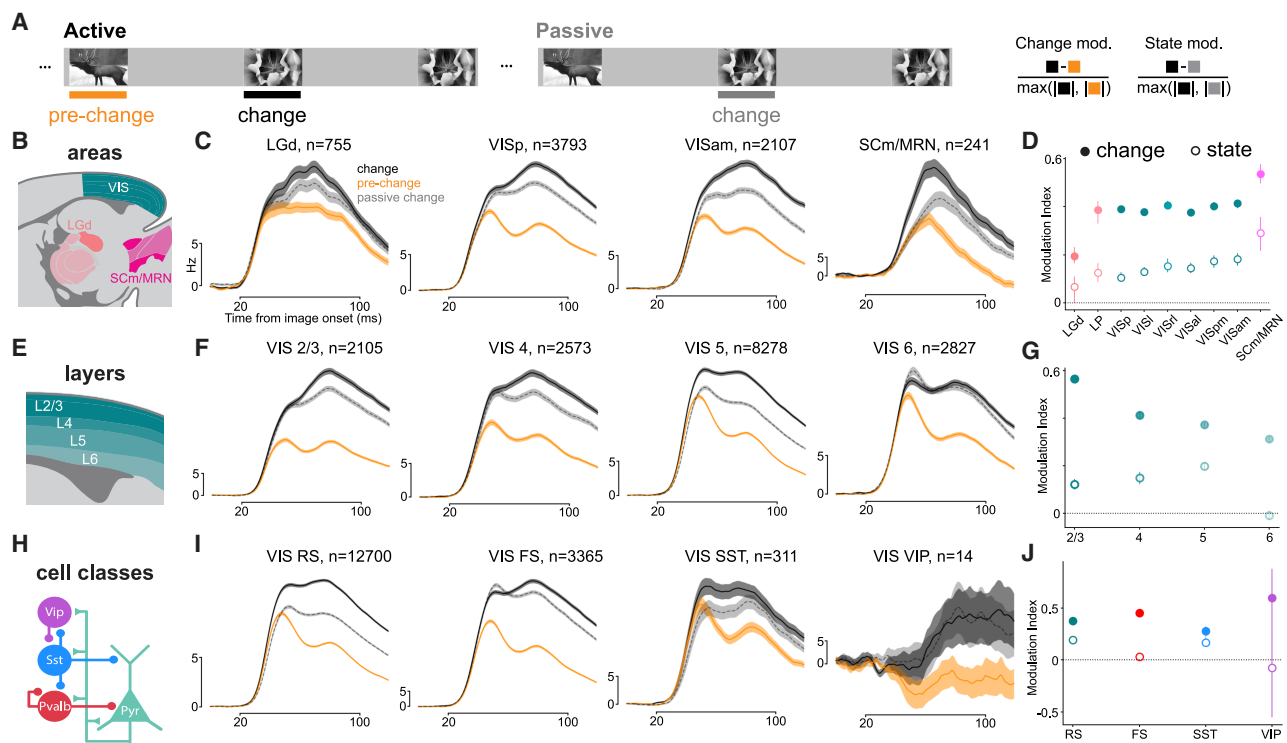


Figure 4. Modulation by image change and task engagement across the visual system

(A) Change and state modulation definitions. Modulation indices were computed on evoked activity for units in sensory clusters.

(B) Areas analyzed in (C) and (D).

(C) Population PSTHs for active change (black), active pre-change (orange), and passive change (gray). x axis scale bar indicates the decision window used to compute metrics in (D). Shading shows SEM across units.

(D) Change (solid symbols) and state (open symbols) modulation indices for nine brain regions (median with bootstrapped 95% confidence interval [CI]).

(E–J) Same as (B)–(D) but for cortical layers (E–G) and cell classes (H–J). Layer and cell-class data were pooled across visual cortical areas.

be read out from the neural population by an external observer. To causally test when activity in the visual cortex is used by mice to detect image changes, we next applied temporally precise optogenetics.

We optogenetically silenced VISp in VGAT-ChR2 mice,^{32,33} starting at various time points relative to the image change and continuing to the end of the response window ($n = 4$ mice, 15 sessions; Figure 3D). When VISp silencing started prior to the change in image presentation, the behavioral response rate was potentially reduced. However, behavior was not impacted when VISp silencing began 100 ms after the image change, indicating that mice required no more than 100 ms of VISp activity to reliably detect image changes. At an intermediate silencing time point, change detection was partially disrupted. This time course of VISp dependence closely matched the time course for change decoding in the visual cortex (Figure 3E).

The lick response is an upper bound on the action initiation time, since orofacial movements, including jaw opening, precede licking. To further constrain the time when overt motor output might influence neural activity through reafference, we trained a decoder to identify lick bouts from video data of the face. This provided a much faster estimate of the behavioral response time than lick latency (face decoder latency: 201 ms, median first lick time: 424 ms) but still lagged midbrain lick

decoding by ~ 75 ms (face decoder: 201 ms, midbrain decoder: 123 ms) (Figure 3E).

Given these bounds, we operationally defined a decision window extending from 20 to 100 ms after image presentation, bracketed by the first stimulus-driven spikes in the visual thalamus and the inflection point in midbrain lick decoding (Figure 3E).

Modulation by image changes and task engagement across areas, layers, and cell types

We next investigated which cell populations were most modulated by behaviorally relevant stimuli (image changes) and task engagement during the decision window. We focused on stimulus-evoked activity in sensory cluster units. We defined two modulation indices. First, a “change modulation index” compares spiking responses to image changes with responses to the stimulus immediately preceding changes (“pre-change” stimulus; Figure 4A). Positive values indicate that image changes evoke larger responses. Second, a “state modulation index” compares responses to change stimuli shown during the behavioral task to those during passive replay (Figure 4A). Positive values indicate larger evoked rates during active versus passive blocks.

Across areas, change modulation was stronger in the visual cortex than in LGd but strongest in SCm/MRN (Figures 4B–4D; Figures S9A and S9C; comparing black and gold curves). Within

the cortex, layer 2/3 units were more modulated than the lower layers (Figures 4E–4G; Figure S9C), while cortical SST+ units were less modulated than the other cell classes (Figures 4H–4J; Figure S9C).

State modulation increased across the thalamocortical visual hierarchy, and SCm/MRN units were most strongly modulated (Figures 4B–4D; Figures S9A and S9D). Across cortical layers, layer 6 was the only layer not significantly modulated by task engagement, while layer 5 showed the strongest modulation (Figures 4E–4G; Figure S9D). Across cortical cell types, regular spiking (RS) and SST units were more strongly modulated by task engagement than fast spiking (FS) units. VIP units were highly variable (Figures 4H–4J; Figure S9D). These findings could not be explained by differences in running speed across active and passive epochs (Figure S10) and were qualitatively similar when grouping units at the session level (Figures S12A–S12C).

In summary, both forms of modulation increased across the visual hierarchy and were strongest in SCm/MRN. In the cortex, modulation varied by layers and cell types, with layer 6 units showing the least modulation overall.

Stimulus novelty modulates late response in the visual cortex

The change modulation index above reveals how units respond to *contextual novelty* over seconds-long time scales,³⁴ i.e., modulation by the change image after several repetitions of the pre-change image. We also designed our study to investigate how long-term image familiarization shapes neural activity relative to stimuli the mouse has never seen (*absolute stimulus novelty*). Throughout training, each mouse encountered only one set of eight images (image set “G” or image set “H”; Figures 5A and 5B), after which we performed two recordings on consecutive days. One day used the familiar image set. The other used the novel image set, consisting of six new images and two “hold-over” images from the familiar set (purple in Figure 5B). For each cell population, we computed a novelty modulation index based on the population response to familiar and novel images (STAR Methods). Positive values indicate higher spiking rates on average to novel versus familiar images.

Visual thalamic regions, LGd and LP, showed minimal image novelty modulation, whereas all visual cortical areas showed strong novelty enhancement (Figures 5C–5E). Midbrain structures SCm/MRN had an intermediate preference for novel images (Figures 5C–5E). These results were qualitatively similar across the active and passive conditions as well as change and non-change stimuli (Figures S11A and S11B) and were consistent with mouse-level analysis (Figure S12D).

Within cortex, novelty modulation was strongest in the upper layers, with less modulation in layer 5 and little to no novelty enhancement in layer 6, particularly in the lateral/posterior visual areas (Figures 5F–5H; Figures S11E and S12E). This lack of modulation in layer 6, the main source of cortical feedback to the primary visual thalamus, is consistent with weak novelty modulation in LGd.

Across cortical cell types, RS, FS, and VIP population activity was strongly novelty-enhanced, whereas the SST population had similar activity for novel and familiar images (Figures 5I and 5J; Figure S12F). Prior calcium imaging experiments

showed that SST cells in VISp and VISl prefer familiar images.¹² To reconcile these findings, we split SST units by cortical area. We found that SST units in VISp as well as lateral areas VISl and VISal were indeed familiar-image preferring (Figure 5K; Figure S11D). By contrast, SST units in anterior/medial areas VISrl, VISpm, and VISam preferred novel images (Figure 5K; Figure S11D).

The lack of novelty modulation in the thalamus suggests the enhancement in the cortex could arise through intracortical interactions, predicting that novelty modulation should lag the early feedforward sweep of activity up the visual hierarchy. Indeed, we found that the first wave of image-evoked activity in LGd and cortex was similar for familiar and novel images (Figures 6A and 6B) and progressed along the anatomical hierarchy as expected (Figure 6C). Novelty modulation manifested as a clear second peak in the cortical peri-stimulus time histogram (PSTH) ~60 ms after stimulus onset (Figures 6B and 6D) and only appeared in the visual thalamus at the end of the decision window (~100 ms; Figures 6A and 6D). In RS cells, the disparity between early and late responses was most dramatic, with the initial peak slightly preferring the familiar image set (Figure 6B). The time course of novelty modulation in cortex was the same for mice trained on either image set G or H (Figure S11F), and analysis of holdover images showed similar results (Figure S11D).

The delayed novelty amplification we observed in the cortex could reflect higher firing rates in neurons that were already responsive during the early window and/or the recruitment of new neurons with longer latency responses. We quantified the fraction of cells with significantly elevated novel image responses in the first half (20–60 ms from stimulus onset; “early” epoch) and second half (60–100 ms; “late” epoch) of the decision window. In the early epoch, a similar fraction of RS cells responded to novel and familiar images, but in the late epoch 10%–15% more cells responded to the novel image set (Figure 6F). This pattern was observed in all cortical areas (Figure 6G). In the visual thalamus, neither epoch showed novelty-driven recruitment (Figures 6E and 6G).

Sensory cluster membership likewise shifted with novelty: in novel sessions, the fraction of on-sustained units increased, whereas on-transient and non-responsive units decreased (Figure 6H). Changes in the other sensory or action clusters were negligible (Figure 6H).

Together, these results suggest that many neurons not previously responsive to familiar images are recruited at long latency by novel images, expanding the active population. What are the consequences of these changes for image representation in the visual cortex? We found that the impact of novelty on image tuning was cell-type specific, consistent with recordings in primates.³⁵ Tuning of RS units in the visual cortex broadened due to amplification of less-preferred stimuli (Figure 6I), whereas FS modulation monotonically scaled with image preference (Figure 6J). Despite these changes, familiar and novel images were decoded with similar accuracy across visual areas (Figure S13A), although decoding of novel images lagged familiar images, particularly in higher visual areas (Figures 6K and 6L). This effect was consistent across cell types (Figures S13B and S13C). Thus, familiarization sharpens

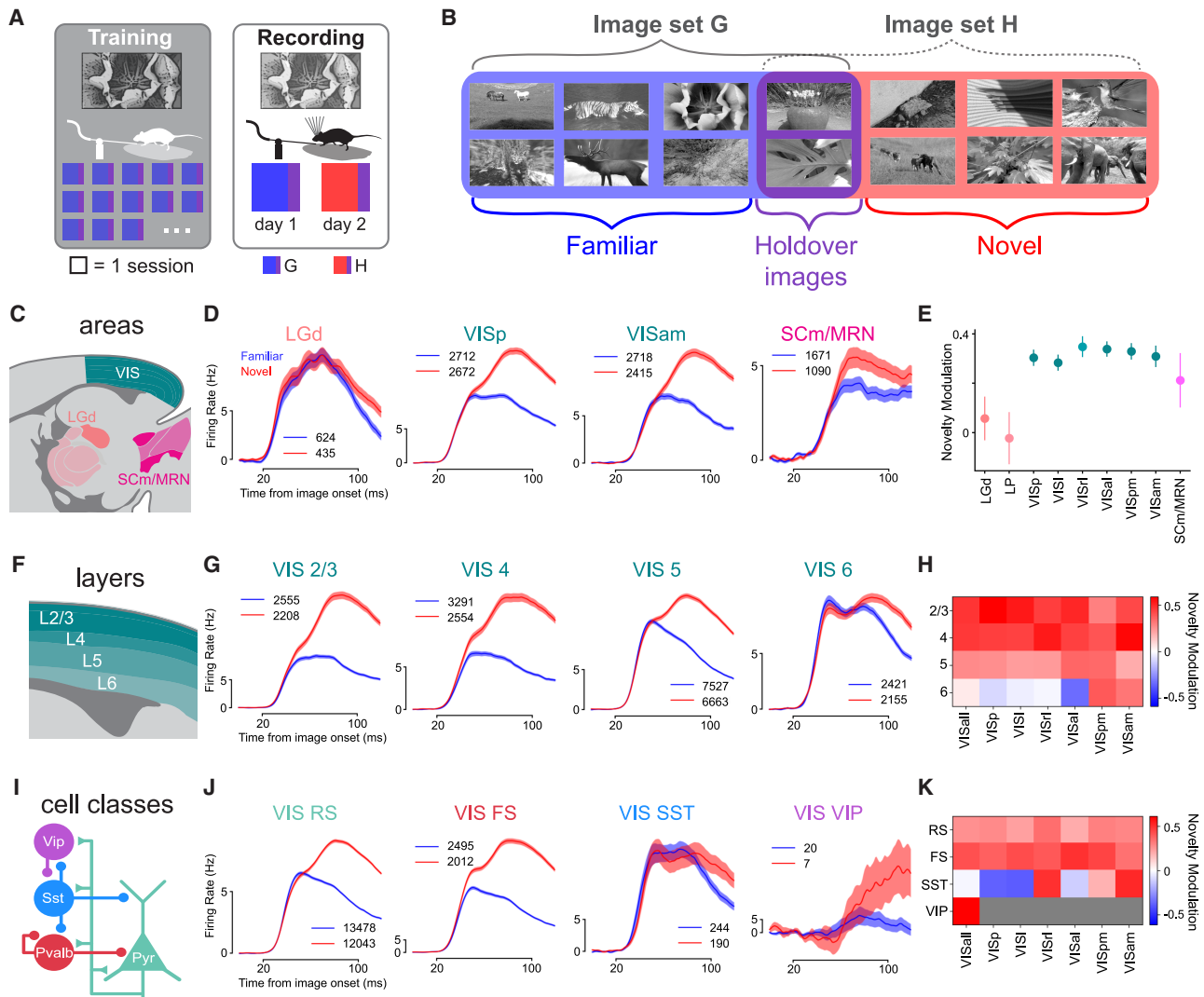


Figure 5. Modulation by stimulus novelty across the visual system

(A) Image-set exposure during training and recording days for an example mouse. Colored squares indicate behavioral sessions. During training (left), the mouse encounters one image set (G). During recording (right), the training image set is used for one day and a novel image set (H) for the other.

(B) G and H image sets contain eight images. Two images are shared (“holdover images,” purple background). The six images unique to each set were used to compare familiar and novel image responses.

(C) Area diagram.

(D) Population PSTHs for familiar (blue) and novel (red) images across four areas. Shading indicates SEM across units.

(E) Novelty modulation indices for nine areas. Novelty modulation was computed on activity during the decision window (20–100 ms from image onset). Error bars show standard deviation over bootstrap samples.

(F–K) Same as (C)–(E) but for cortical layers (F–H) and cell classes (I–K). Note that novelty modulation indices for VIP+ units were not calculated for individual areas due to their limited number.

temporal dynamics (all cell types) and stimulus tuning (RS cells), producing similar decoding with fewer spikes and faster latencies.

In four mice, we investigated the time course of novel image familiarization across days (Figure S13D). Consistent with previous results,¹² response magnitudes returned to familiar levels after one session of exposure to novel images (Figures S13E and S13F), though response dynamics remained slower (Figures S13E, S13G, and S13H).

Perturbing behavior with novel images suggests a change detection strategy

Given the dramatic changes in neural representation of novel versus familiar images, we next asked whether behavior with the novel image set could yield clues about the computational strategy mice used to solve the task. We considered two strategies, which make very different behavioral predictions for the two familiar holdover images (purple images in Figure 5B) during novel and familiar sessions.

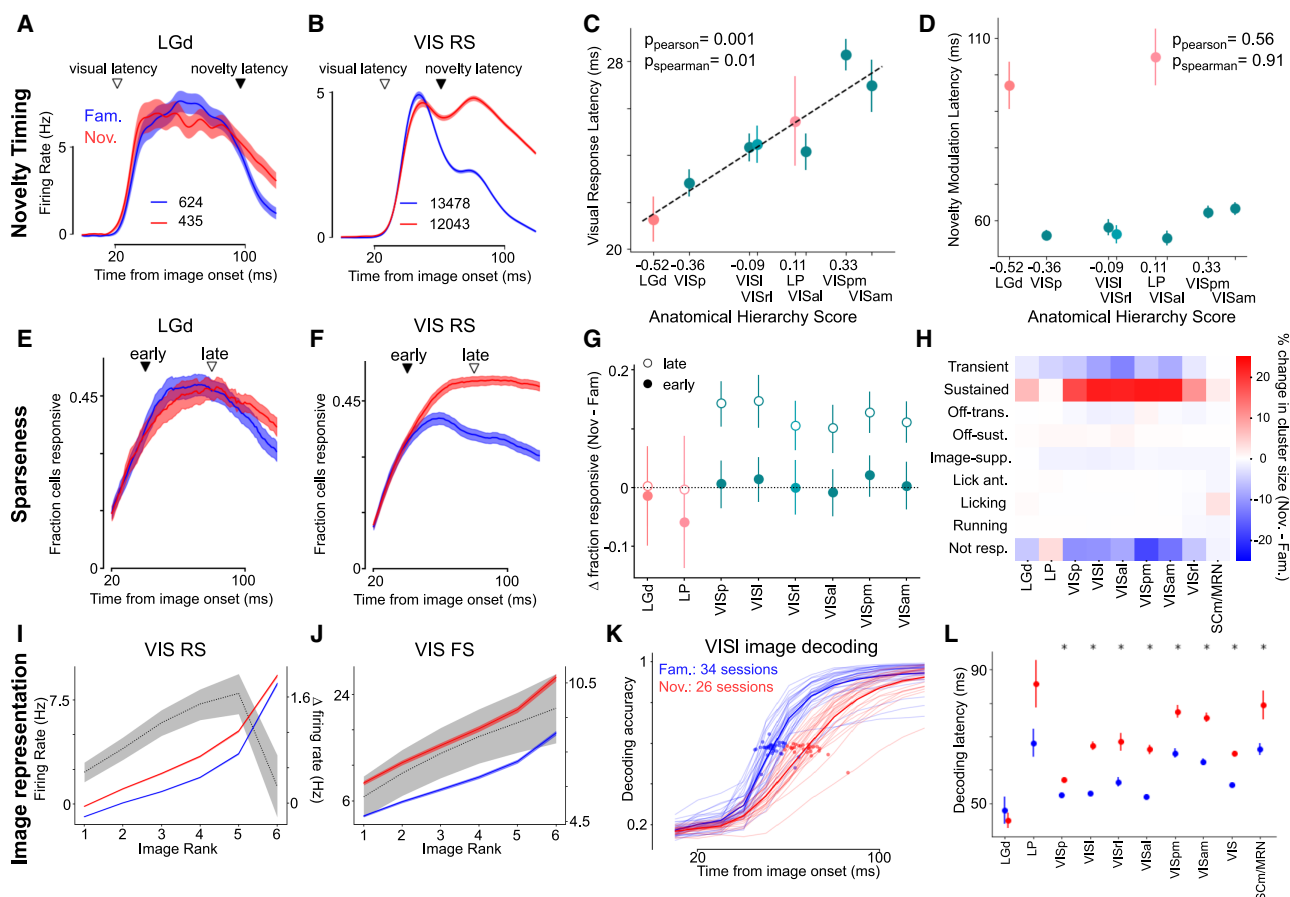


Figure 6. Novel images recruit slow, dense population responses in the cortex

(A) LGd population PSTH for familiar and novel non-change stimuli. Shading indicates SEM. Arrowheads indicate estimated visual latency (open symbol) and novelty latency (closed symbol).
 (B) Same as (A) for visual cortical RS units.
 (C) Visual response latency across the visual hierarchy. Error bars for (C) and (D) indicate the standard deviation of the bootstrap distribution of means (1,000 iterations).
 (D) Novelty modulation latency across the visual hierarchy. Error bars for (C) and (D) indicate the standard deviation of the bootstrap distribution of means (1,000 iterations).
 (E) Fraction of LGd units responsive over time for familiar and novel images. Arrowheads indicate the midpoint of the early (20–60 ms) and late (60–100 ms) response windows.
 (F) Same as (E) but for VISp units. Shading indicates SEM over sessions.
 (G) Difference in fraction of units responsive to novel (closed symbols) and familiar (open symbols) images for early and late response windows in (E) and (F). Error bars are 95% CI of the mean.
 (H) Change in cluster membership for units in novel versus familiar sessions.
 (I) Tuning curves of visual cortical RS units for familiar and novel stimuli (shading indicates SEM over units). Image responses were ranked before averaging. Dotted gray line shows the difference between novel and familiar curves (shading indicates the 95% bootstrapped CI).
 (J) Same as (I) for visual cortical FS units.
 (K) Image decoding time courses for VISl units during 34 familiar and 26 novel sessions. Light traces indicate individual sessions. Dark traces are means across sessions. Symbols indicate time to half max for each session.
 (L) Mean image decoding latency across areas. Error bars indicate SEM across sessions. * indicates $p < 0.05$ for the Wilcoxon rank-sum test after correction for multiple comparisons.

In the first strategy, mice remember the identity of the last image shown until the next image presentation. When the next image comes, they read out that image and compare it to the remembered image, licking for a mismatch. To model this “image-comparison” strategy, we trained a one-vs.-all image decoder on activity during the decision window for each of the 8 images. These decoders determine whether each image

presentation was the same as or different from the previous image. Decoder confidence was given by the distance of each image response from the decoder boundary (magenta curly bracket in Figure 7A).

In the second strategy, mice exploit the fact that image changes produce larger population responses than image repeats, irrespective of image identity (i.e., change modulation in

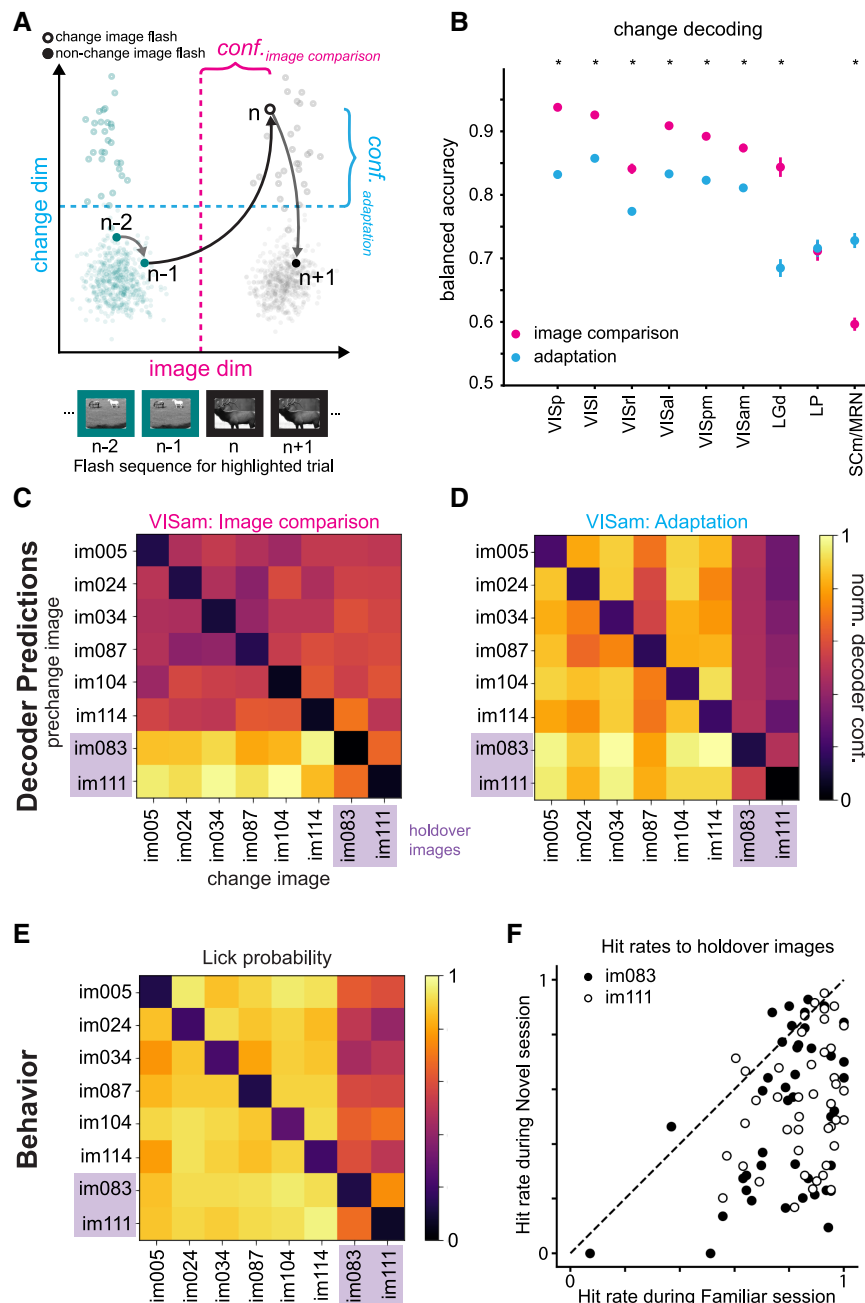


Figure 7. Behavior during the novel day suggests a change computation strategy

(A) Plot illustrating two possible strategies for detecting image change. Individual dots are population responses in the visual cortex to stimulus presentations plotted in two dimensions, maximizing separation between image identity (x axis) and between change and repeat stimuli (y axis). Responses to two example stimuli are shown in teal and black. Solid symbols are repeat presentations, and open symbols are change stimuli. Bold symbols show responses from one example trial when the image transitioned from the teal to the black image. Magenta line indicates the boundary for the “image-comparison” decoder. Cyan line indicates the boundary for the “adaptation” decoder. Distance from these boundaries measures the confidence of the respective decoder.

(B) Change decoding accuracy using confidence from either the image-comparison decoder (magenta) or the adaptation decoder (cyan). Error bars indicate SEM. * indicates $p < 0.05$ for the Wilcoxon signed-rank test after correction for multiple comparisons.

(C) Normalized mean decoder confidence across sessions for the image-comparison decoder for each image transition. Holdover images are the last two rows/columns.

(D) Same as (C) but for the adaptation decoder.

(E) Mouse behavior on the novel day, showing mean lick probability across sessions for each image transition.

(F) Response probability for the two holdover images during the familiar and novel sessions. Each symbol represents the response rate to one of the images for one session.

Figure 4). Mice could use this enhanced response to infer that the image changed without directly comparing the current image to a memory of the previous image. We refer to this as the “adaptation strategy,” though change modulation may not entirely arise from low-level adaptation mechanisms (see discussion). To model this strategy, we trained a decoder to classify individual image presentations as either change or non-change and quantified decoder confidence (cyan curly bracket in Figure 7A).

Both strategies identified image changes well above chance across all areas tested (Figure 7B). However, examining decoder confidence for each image transition revealed starkly different

predictions by the two strategies. The image-comparison decoder predicts a benefit for detecting changes from familiar to novel images (first six columns of bottom two rows in Figure 7C), leveraging the separation of novel and familiar responses in population space. By contrast, the adaptation decoder predicts a deficit in detecting transitions to familiar holdovers (rightmost columns in Figure 7D), confusing response attenuation due to adaptation with that due to

familiarization. Strikingly, we found that mice exhibit this same deficit (Figure 7E), responding to transitions to familiar holdover images at a much lower rate than novel images. This was not a general deficit in detecting changes for these stimuli, as mice responded at high rates when they appeared in the familiar session (Figure 7F). The same deficit was present in mice regardless of whether they trained on the G or H image sets, indicating that it reflected the novel context rather than the image set itself (Figures S14A and S14B).

The adaptation model further predicted a higher false alarm rate for novel images (diagonal of Figure 7D), a pattern also

found in mouse behavior (Figure 7E; Figure S14C). In addition, in all areas except LGd, the adaptation decoder better predicted trial-by-trial licking compared with image comparison (Figure S14D), particularly in higher visual areas and SCm/MRN, where it assigned the greatest feature weights to units from the on-transient, on-sustained, and lick-anticipation clusters (Figure S14F). We found little evidence for a persistent representation of image identity during the inter-stimulus gray period (Figure S14E), as the image-comparison strategy would predict, though this representation could be maintained in unsampled regions.

DISCUSSION

This study presents a large-scale database of spiking activity from mice performing a visual change detection task. Dense sampling across areas, layers, and cell types reveals diverse functional encoding, a narrow time window linking cortical activity to choice, and pronounced modulation by novel stimuli. Together, these results—and the open dataset that enabled them—map how sensory and behavioral signals are distributed across the mouse visual system and are shaped by experience.

Coding across areas, layers, and cell types

By recording across the visual system from early sensory regions (LGd and VISP) to midbrain premotor areas (SCm and MRN), we mapped sensory and motor encoding during visually guided behavior. Our clustering analysis identified a set of canonical visual response types, including on-transient, off-transient, on-sustained, off-sustained, and contrast-suppressed cells, all of which have correlates in retinal^{36,37} and thalamic^{38,39} processing pathways. In addition, we found a small number of action-encoding units in early visual cortical areas, supporting the emerging view that movement information is widely distributed.^{5,40–42} Conversely, we found visual sensory neurons in SCm and MRN—classically motor structures^{28,29}—that responded to images at short latency and supported rapid, high-accuracy image decoding. Such distributed sensorimotor representations may support active sensation, where stimulus and movement information must be integrated to predict action consequences.⁴³

Dense recordings also revealed layer- and cell-type-specific findings, two of which we highlight here: (1) layer 6 maintains stable sensory responses, and (2) SST+ interneurons show area-specific novelty modulation.

Layer 6 was less modulated by both task engagement and stimulus novelty than other layers. Since layer 6 provides the primary feedback projection to LGd, this is consistent with the similarly low modulation in LGd. This compartmentalization may allow layer 6 and LGd to maintain a stable sensory representation across behavioral states and experiences. Layer 6 is known to influence gain in LGd and other cortical layers in a stimulus-dependent manner,^{44–48} and the lack of strong task engagement and novelty modulation could allow layer 6 to implement this control independent of cognitive and behavioral factors. We note that a previous study⁴⁹ reported layer 6 modulation across major state changes (sleep to high arousal), a broader axis than the task-engaged versus unengaged comparison here.

Using optotagging, we identified SST+ interneurons across all 6 visual cortical areas. SST+ neurons in VISP and lateral areas (VISl and VISa) preferred familiar images (consistent with previous calcium imaging results in VISP¹³), whereas SST+ neurons in anterior/medial areas (VISr, VISam, and VISpm) preferred novel images, revealing region-specific functional heterogeneity. Stronger recurrent input to SST+ cells in medial versus lateral areas⁵⁰ could explain this difference, given our evidence that novelty modulation arises through recurrent cortical interactions.

Time window for change detection behavior

A key aspect of our analysis was to define the critical time window (20–100 ms post-stimulus) when stimulus change information in the visual cortex informs behavioral choices. We found tight temporal correspondence between image change decoding from cortical activity and behavioral effects of optogenetic silencing of the primary visual cortex. This short window is consistent with previous studies that evaluated when the visual cortex is necessary for stimulus detection and discrimination.^{51,52} Intriguingly, stimulus changes could be decoded earliest in the higher-order visual thalamic area LP (mouse pulvinar)—even earlier than VISP—suggesting that LP may be a critical node in the change-detection computation. Fast change detection signals could arise in LP through a disynaptic pathway coursing from retinal ganglion cells to collicular wide field vertical (WVF) cells⁵³ to posterior LP cells.^{54–56} Both WVF⁵⁷ and posterior LP cells⁵⁴ have large spatial receptive fields, which could confer enhanced sensitivity to full-field image changes.

The decoding lag between LP and VISP does not by itself indicate a processing lag, but this result is consistent with LP's proposed role in amplifying prediction errors in cortical layer 2/3 for unexpected stimuli.⁵⁸ Our finding that layer 2/3 is particularly sensitive to image changes provides additional support for its role in prediction error signaling.⁵⁹ The rapid propagation of image change signals through this thalamocortical network, culminating in midbrain motor preparation within 120 ms, reveals the remarkable efficiency of this sensorimotor transformation.

Within this critical time window, our results provide insight into the computational strategy mice use to solve the task. Mouse behavior showed specific deficits in detecting changes to familiar holdover images in the novel context, consistent with deficits predicted by an adaptation but not an image-comparison strategy. Further studies will be required to definitively establish the role of response adaptation by designing behavioral tasks that decouple changes in visual gain and image identity.

The adaptation strategy may be computationally advantageous because it requires only one decision boundary, whereas distinguishing each image from the others would require up to eight. It also avoids the energetically costly persistent spiking required for image-template maintenance, instead relying on short-term plasticity mechanisms, including synaptic depression.^{60,61}

Novelty expands the active cortical population and enhances sustained visual response

During the short 80-ms decision window following stimulus presentation, we observed distinct waves of activity in the visual

cortex. The first wave (20–60 ms after stimulus onset) is a feed-forward sweep minimally affected by stimulus novelty. A second wave between 60–100 ms is strongly modulated by stimulus novelty. During this window, novel images engage an expanded set of on-sustained cortical neurons, recruited largely from a reservoir of non-responsive cells (Figure 6). VIP+ interneurons are also activated by novel images during this late window (Figure 5J), suggesting that VIP+ cells—perhaps through their inhibition of SST+ cells and consequent disinhibition of excitatory cells^{12,30,62–64}—may gate recruitment of this expanded population. This second wave of activity could reflect recurrent cortical processing, including local circuits within VISp,⁶⁵ feedback from higher-order visual cortical areas,⁶⁶ or top-down feedback from frontal cortical regions.⁶² Other mechanisms, such as fast novelty-evoked neuromodulator release, could also contribute.⁶⁷

The expansion of neurons responding to novel images may be computationally beneficial, allowing the cortex to adapt to new stimulus statistics without disrupting existing representations.^{68,69} The delayed novelty preference in the cortex is not simply relayed to midbrain premotor regions (SCm/MRN), which show less novelty modulation. This may reflect the source of cortical input to these regions (primarily layer 5, which shows weaker novelty modulation than the upper layers) and/or integration with other non-novelty-modulated inputs. Together, these results suggest that the cortex specializes in adaptive sensory representation, while premotor midbrain regions are tuned for behavioral relevance, evident in their strong stimulus change and state modulation.

Limitations of the study

Although we measured activity across many brain regions spanning the thalamus, cortex, and midbrain, additional structures will be important to sample in future work. These include higher-order visual cortical areas VISli and VISpor, prefrontal cortical structures implicated in novelty processing,⁷⁰ and several additional sub-compartments of the superior colliculus and LP (superficial SC and anterior/medial LP), which likely influence visual processing in the areas we characterize in this paper.

Our results are also primarily correlational. Future perturbation experiments should test causal mechanisms by which specific areas, layers, and cell types contribute to the computations and behavioral responses we describe. For instance, the mechanisms supporting the adaptation strategy remain to be determined and could involve short-term synaptic depression of excitatory inputs and/or the recruitment of specific inhibitory cell types, as implicated by studies on unexpected stimulus processing in visual^{22,71} and auditory^{72,73} cortex. These experiments should be guided by circuit modeling efforts constrained by the dataset we provide.

A database for mining and modeling

Our open dataset provides a comprehensive characterization of spiking activity in the mouse visual system during a visually guided behavior and is a rich resource for further data mining and modeling. We packaged all spiking, LFP, and behavior data (including full training histories for each mouse) into Neurodata Without Borders (NWB) files and deposited them on public cloud storage platforms (Figure S1; Table S1), with tutorials and

technical resources to support access and analysis (Table S2). Several studies have already used this database to explore various topics, including omission responses across brain regions,⁷⁴ the impact of novel images on the representation of familiar images,⁷⁵ and how dynamics across the visual hierarchy approach criticality.⁷⁶

This dataset is an ideal substrate for examining multi-regional interactions¹⁴ and spike propagation through hierarchical networks.⁴ Its comprehensive coverage of multiple areas, cortical layers, and cell types during behavior makes it a powerful resource for data-intensive models, including detailed biophysical models,^{77,78} cell-type-specific population models,^{79,80} and neural network foundation models.^{77,81,82} The full impact of this database will emerge through community engagement in mining, analysis, and modeling efforts.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to the lead contact, Shawn R. Olsen (shawno@alleninstitute.org).

Materials availability

This study did not generate any new, unique reagents.

Data and code availability

- Data: all behavior, spike, and LFP data were packaged into NWB files and deposited in a public S3 bucket (<s3://visual-behavior-neuropixels-data/visual-behavior-neuropixels/>), as well as the DANDI archive (DANDI: <https://doi.org/10.48324/dandi.000713/0.240702.1725>). Raw electrophysiology and behavior video data have been deposited in a public S3 bucket (<s3://allen-brain-observatory/visual-behavior-neuropixels/raw-data/>).
- Code: code to access and analyze these data can be found in our public tutorial resource (https://allensdk.readthedocs.io/en/latest/visual_behavior_neuropixels.html). Original code has been deposited at <https://github.com/AllenInstitute/VBN-Manuscript> and is publicly available as of the date of publication.
- Additional information: any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was supported by the Allen Institute. We thank the Transgenic Colony Management, Animal Care, and Laboratory Animal Services teams at the Allen Institute for caring for the mice in this study. A. Arkhipov and O.Z. were further supported by the National Institute of Mental Health of the National Institutes of Health under award no. U01MH130907. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

AUTHOR CONTRIBUTIONS

Conceptualization: C.B., S.D.G., M.G., D.R.O., P.A.G., and S.R.O.; data curation: C.B., H.B., G.H., T.K.R., A. Amster, S.F.D., B.J.H., E.G.M., C.B.M., and S.D.; formal analysis: C.B., S.D.G., A.P., O.Z., and D.M.; funding acquisition: A. Arkhipov and C.K.; investigation: C.B., H.B., S.D.G., G.H., T.K.R., S.C., C.G., T.J., C.N., N.-K.N., K.N., B.O., P.R., J.H.S., X.W., M.C., and R.G.; methodology: C.B., S.D.G., G.H., A.P., T.K.R., M.G., J.H.S., and S.D.; project administration: L.C., J.H.S., A.W., P.A.G., J.W.P., and S.R.O.; resources: A.C., P.L., Q.L., M.S., C.S., D.S., and K.R.; software: C.B., S.D.G., G.H., A.P., O.Z., A. Amster, S.F.D., B.J.H., R.H., P.L., C.M., C.B.M., D.R.O.,

J.H.S., and W.W.; supervision: C.B., A. Arkhipov, C.F., S.M., D.S., M.T., A.W., S.D., P.A.G., C.K., and S.R.O.; validation: C.B., H.B., G.H., T.K.R., E.G.M., S.D., and S.R.O.; visualization: C.B., S.D.G., O.Z., and S.R.O.; writing – original draft: C.B. and S.R.O.; writing – review & editing: C.B., S.D.G., G.H., A.P., O.Z., A. Arkhipov, M.G., B.J.H., P.L., J.H.S., S.D., P.A.G., C.K., and S.R.O.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used Copilot and Claude to review and generate code and edit text. After using these tools, the authors reviewed, validated, and edited the content as needed and take full responsibility for the content of this publication.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
 - Animals
- **METHOD DETAILS**
 - Surgery
 - Intrinsic Signal Imaging
 - Behavior Training
 - Neuropixels Recordings
 - Experimental design and visual stimuli
 - Optotagging
 - Ex Vivo Imaging
 - Spike Sorting
 - VISp silencing experiment
 - Data processing
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Receptive Fields
 - Hit rate, false alarm rate and d-prime
 - Correlation of VISp interneuron activity with locomotion
 - Regression Model
 - Functional Clustering
 - Peri-stimulus/event histograms
 - Image, change and lick decoding
 - Facemap lick decoding
 - Image decoding for familiar and novel sessions
 - Change, state and novelty modulation indices
 - Visual response and novelty modulation latencies
 - Change detection strategy analysis
 - Statistical Comparisons

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cell.2026.06.025>.

Received: October 22, 2025

Revised: March 28, 2026

Accepted: June 16, 2026

Published: July 8, 2026

REFERENCES

1. Durand, S., Heller, G.R., Ramirez, T.K., Luviano, J.A., Williford, A., Sullivan, D.T., Cahoon, A.J., Farrell, C., Groblewski, P.A., Bennett, C., et al. (2023). Acute head-fixed recordings in awake mice with multiple Neuropixels

- probes. *Nat. Protoc.* 18, 424–457. <https://doi.org/10.1038/s41596-022-00768-6>.
2. Jun, J.J., Steinmetz, N.A., Siegle, J.H., Denman, D.J., Bauza, M., Barbarits, B., Lee, A.K., Anastassiou, C.A., Andrei, A., Aydin, Ç., et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature* 551, 232–236. <https://doi.org/10.1038/nature24636>.
3. Machado, T.A., Kauvar, I.V., and Deisseroth, K. (2022). Multiregion neuronal activity: the forest and the trees. *Nat. Rev. Neurosci.* 23, 683–704. <https://doi.org/10.1038/s41583-022-00634-0>.
4. Siegle, J.H., Jia, X., Durand, S., Gale, S., Bennett, C., Graddis, N., Heller, G., Ramirez, T.K., Choi, H., Luviano, J.A., et al. (2021). Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature* 592, 86–92. <https://doi.org/10.1038/s41586-020-03171-x>.
5. Steinmetz, N.A., Zatka-Haas, P., Carandini, M., and Harris, K.D. (2019). Distributed coding of choice, action and engagement across the mouse brain. *Nature* 576, 266–273. <https://doi.org/10.1038/s41586-019-1787-x>.
6. Bennett, C., Ouellette, B., Ramirez, T.K., Cahoon, A., Cabasco, H., Brown-ing, Y., Lakunina, A., Lynch, G.F., McBride, E.G., Belski, H., et al. (2024). SHIELD: Skull-shaped hemispheric implants enabling large-scale electrophysiology datasets in the mouse brain. *Neuron* 112, 2869–2885.e8. <https://doi.org/10.1016/j.neuron.2024.06.015>.
7. Chen, S., Liu, Y., Wang, Z.A., Colonell, J., Liu, L.D., Hou, H., Tien, N.-W., Wang, T., Harris, T., Druckmann, S., et al. (2024). Brain-wide neural activity underlying memory-guided movement. *Cell* 187, 676–691.e16. <https://doi.org/10.1016/j.cell.2023.12.035>.
8. Khilkevich, A., Lohse, M., Low, R., Orsolic, I., Bozic, T., Windmill, P., and Mrsic-Flogel, T.D. (2024). Brain-wide dynamics linking sensation to action during decision-making. *Nature* 634, 890–900. <https://doi.org/10.1038/s41586-024-07908-w>.
9. Bondy, A.G., Charlton, J.A., Luo, T.Z., Kopec, C.D., Stagnaro, W.M., Venditto, S.J.C., Lynch, L., Janarthanan, S., Oline, S.N., Harris, T.D., et al. (2025). Brain-wide coordination of internal signals during decision-making. Preprint at bioRxiv. <https://doi.org/10.1101/2024.08.21.609044>.
10. Angelaki, D., Benson, B., Benson, J., Birman, D., Bonacchi, N., Bougrova, K., Bruijns, S.A., Carandini, M., Catarino, J.A., et al.; International Brain Laboratory (2025). A brain-wide map of neural activity during complex behaviour. *Nature* 645, 177–191. <https://doi.org/10.1038/s41586-025-09235-0>.
11. de Vries, S.E.J., Lecoq, J.A., Buice, M.A., Groblewski, P.A., Ocker, G.K., Oliver, M., Feng, D., Cain, N., Ledochowitsch, P., Millman, D., et al. (2020). A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. *Nat. Neurosci.* 23, 138–151. <https://doi.org/10.1038/s41593-019-0550-9>.
12. Garrett, M., Groblewski, P., Piet, A., Ollerenshaw, D., Najafi, F., Yavorska, I., Amster, A., Bennett, C., Buice, M., Caldejon, S., et al. (2023). Stimulus novelty uncovers coding diversity in visual cortical circuits. Preprint at bioRxiv. <https://doi.org/10.1101/2023.02.14.528085>.
13. Arkhipov, A., da Costa, N., de Vries, S., Bakken, T., Bennett, C., Bernard, A., Berg, J., Buice, M., Collman, F., Daigle, T., et al. (2025). Integrating multimodal data to understand cortical circuit architecture and function. *Nat. Neurosci.* 28, 717–730. <https://doi.org/10.1038/s41593-025-01904-7>.
14. Jia, X., Siegle, J.H., Durand, S., Heller, G., Ramirez, T.K., Koch, C., and Olsen, S.R. (2022). Multi-regional module-based signal transmission in mouse visual cortex. *Neuron* 110, 1585–1598.e9. <https://doi.org/10.1016/j.neuron.2022.01.027>.
15. Akella, S., Ledochowitsch, P., Siegle, J.H., Belski, H., Denman, D.D., Buice, M.A., Durand, S., Koch, C., Olsen, S.R., and Jia, X. (2025). Deciphering neuronal variability across states reveals dynamic sensory encoding. *Nat. Commun.* 16, 1768. <https://doi.org/10.1038/s41467-025-56733-w>.

16. Fritsche, M., Solomon, S.G., and de Lange, F.P. de (2022). Brief Stimuli Cast a Persistent Long-Term Trace in Visual Cortex. *J. Neurosci.* 42, 1999–2010. <https://doi.org/10.1523/JNEUROSCI.1350-21.2021>.
17. Purandare, C., and Mehta, M. (2023). Mega-scale movie-fields in the mouse visuo-hippocampal network. *eLife* 12, RP85069. <https://doi.org/10.7554/eLife.85069>.
18. Nitzan, N., Swanson, R., Schmitz, D., and Buzsáki, G. (2022). Brain-wide interactions during hippocampal sharp wave ripples. *Proc. Natl. Acad. Sci. USA* 119, e2200931119. <https://doi.org/10.1073/pnas.2200931119>.
19. de Vries, S.E.J., Siegle, J.H., and Koch, C. (2023). Sharing neurophysiology data from the Allen Brain Observatory. *eLife* 12, e85550. <https://doi.org/10.7554/eLife.85550>.
20. Garrett, M., Manavi, S., Roll, K., Ollerenshaw, D.R., Groblewski, P.A., Ponvert, N.D., Kiggins, J.T., Casal, L., Mace, K., Williford, A., et al. (2020). Experience shapes activity dynamics and stimulus coding of VIP inhibitory cells. *eLife* 9, e50340. <https://doi.org/10.7554/eLife.50340>.
21. Glickfeld, L.L., Histed, M.H., and Maunsell, J.H.R. (2013). Mouse Primary Visual Cortex Is Used to Detect Both Orientation and Contrast Changes. *J. Neurosci.* 33, 19416–19422. <https://doi.org/10.1523/JNEUROSCI.3560-13.2013>.
22. Hamm, J.P., and Yuste, R. (2016). Somatostatin Interneurons Control a Key Component of Mismatch Negativity in Mouse Visual Cortex. *Cell Rep.* 16, 597–604. <https://doi.org/10.1016/j.celrep.2016.06.037>.
23. Jin, M., and Glickfeld, L.L. (2020). Mouse Higher Visual Areas Provide Both Distributed and Specialized Contributions to Visually Guided Behaviors. *Curr. Biol.* 30, 4682–4692.e7. <https://doi.org/10.1016/j.cub.2020.09.015>.
24. Wang, L., McAlonan, K., Goldstein, S., Gerfen, C.R., and Krauzlis, R.J. (2020). A Causal Role for Mouse Superior Colliculus in Visual Perceptual Decision-Making. *J. Neurosci.* 40, 3768–3782. <https://doi.org/10.1523/JNEUROSCI.2642-19.2020>.
25. Cone, J.J., Bade, M.L., Masse, N.Y., Page, E.A., Freedman, D.J., and Maunsell, J.H.R. (2020). Mice Preferentially Use Increases in Cerebral Cortex Spiking to Detect Changes in Visual Stimuli. *J. Neurosci.* 40, 7902–7920. <https://doi.org/10.1523/JNEUROSCI.1124-20.2020>.
26. Groblewski, P.A., Ollerenshaw, D.R., Kiggins, J.T., Garrett, M.E., Mochizuki, C., Casal, L., Cross, S., Mace, K., Swapp, J., Manavi, S., et al. (2020). Characterization of learning, motivation, and visual perception in five transgenic mouse lines expressing GCaMP in distinct cell populations. *Front. Behav. Neurosci.* 14, 104. <https://doi.org/10.3389/fnbeh.2020.00104>.
27. Piet, A., Ponvert, N., Ollerenshaw, D., Garrett, M., Groblewski, P.A., Olsen, S., Koch, C., and Arkhipov, A. (2024). Behavioral strategy shapes activation of the Vip-Sst disinhibitory circuit in visual cortex. *Neuron* 112, 1876–1890.e4. <https://doi.org/10.1016/j.neuron.2024.02.008>.
28. Inagaki, H.K., Chen, S., Ridder, M.C., Sah, P., Li, N., Yang, Z., Hasanbegovic, H., Gao, Z., Gerfen, C.R., and Svoboda, K. (2022). A midbrain-thalamus-cortex circuit reorganizes cortical dynamics to initiate movement. *Cell* 185, 1065–1081.e23. <https://doi.org/10.1016/j.cell.2022.02.006>.
29. Gandhi, N.J., and Katnani, H.A. (2011). Motor Functions of the Superior Colliculus. *Annu. Rev. Neurosci.* 34, 205–231. <https://doi.org/10.1146/annurev-neuro-061010-113728>.
30. Fu, Y., Tucciarone, J.M., Espinosa, J.S., Sheng, N., Darcy, D.P., Nicoll, R.A., Huang, Z.J., and Stryker, M.P. (2014). A Cortical Circuit for Gain Control by Behavioral State. *Cell* 156, 1139–1152. <https://doi.org/10.1016/j.cell.2014.01.050>.
31. Reimer, J., Froudarakis, E., Cadwell, C.R., Yatsenko, D., Denfield, G.H., and Tlilas, A.S. (2014). Pupil fluctuations track fast switching of cortical states during quiet wakefulness. *Neuron* 84, 355–362. <https://doi.org/10.1016/j.neuron.2014.09.033>.
32. Li, N., Chen, S., Guo, Z.V., Chen, H., Huo, Y., Inagaki, H.K., Chen, G., Davis, C., Hansel, D., Guo, C., et al. (2019). Spatiotemporal constraints on optogenetic inactivation in cortical circuits. *eLife* 8, e48622. <https://doi.org/10.7554/eLife.48622>.
33. Zhao, S., Ting, J.T., Atallah, H.E., Qiu, L., Tan, J., Gloss, B., Augustine, G.J., Deisseroth, K., Luo, M., Graybiel, A.M., et al. (2011). Cell type-specific channelrhodopsin-2 transgenic mice for optogenetic dissection of neural circuitry function. *Nat. Methods* 8, 745–752. <https://doi.org/10.1038/nmeth.1668>.
34. Schomaker, J., and Meeter, M. (2015). Short- and long-lasting consequences of novelty, deviance and surprise on brain and cognition. *Neurosci. Biobehav. Rev.* 55, 268–279. <https://doi.org/10.1016/j.neubiorev.2015.05.002>.
35. Woloszyn, L., and Sheinberg, D.L. (2012). Effects of Long-Term Visual Experience on Responses of Distinct Classes of Single Units in Inferior Temporal Cortex. *Neuron* 74, 193–205. <https://doi.org/10.1016/j.neuron.2012.01.032>.
36. Masland, R.H. (2012). The Neuronal Organization of the Retina. *Neuron* 76, 266–280. <https://doi.org/10.1016/j.neuron.2012.10.002>.
37. Baden, T., Berens, P., Franke, K., Román Rosón, M., Bethge, M., and Euler, T. (2016). The functional diversity of retinal ganglion cells in the mouse. *Nature* 529, 345–350. <https://doi.org/10.1038/nature16468>.
38. Piscopo, D.M., El-Danaf, R.N., Huberman, A.D., and Niell, C.M. (2013). Diverse Visual Features Encoded in Mouse Lateral Geniculate Nucleus. *J. Neurosci.* 33, 4642–4656. <https://doi.org/10.1523/JNEUROSCI.5187-12.2013>.
39. Román Rosón, M.R., Bauer, Y., Kotkat, A.H., Berens, P., Euler, T., and Busse, L. (2019). Mouse dLGN Receives Functional Input from a Diverse Population of Retinal Ganglion Cells with Limited Convergence. *Neuron* 102, 462–476.e8. <https://doi.org/10.1016/j.neuron.2019.01.040>.
40. Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C.B., Carandini, M., and Harris, K.D. (2019). Spontaneous behaviors drive multidimensional, brain-wide activity. *Science* 364, 255. <https://doi.org/10.1126/science.aav7893>.
41. Musall, S., Kaufman, M.T., Juavinett, A.L., Gluf, S., and Churchland, A.K. (2019). Single-trial neural dynamics are dominated by richly varied movements. *Nat. Neurosci.* 22, 1677–1686. <https://doi.org/10.1038/s41593-019-0502-4>.
42. Salkoff, D.B., Zagha, E., McCarthy, E., and McCormick, D.A. (2020). Movement and Performance Explain Widespread Cortical Activity in a Visual Detection Task. *Cereb. Cortex* 30, 421–437. <https://doi.org/10.1093/cercor/bhz206>.
43. Rao, R.P.N. (2024). A sensory-motor theory of the neocortex. *Nat. Neurosci.* 27, 1221–1235. <https://doi.org/10.1038/s41593-024-01673-9>.
44. Spacek, M.A., Crombie, D., Bauer, Y., Born, G., Liu, X., Katzner, S., and Busse, L. (2022). Robust effects of corticothalamic feedback and behavioral state on movie responses in mouse dLGN. *eLife* 11, e70469. <https://doi.org/10.7554/eLife.70469>.
45. Olsen, S.R., Bortone, D.S., Adesnik, H., and Scanziani, M. (2012). Gain control by layer six in cortical circuits of vision. *Nature* 483, 47–52. <https://doi.org/10.1038/nature10835>.
46. Bortone, D.S., Olsen, S.R., and Scanziani, M. (2014). Translaminar Inhibitory Cells Recruited by Layer 6 Corticothalamic Neurons Suppress Visual Cortex. *Neuron* 82, 474–485. <https://doi.org/10.1016/j.neuron.2014.02.021>.
47. Denman, D.J., and Contreras, D. (2015). Complex Effects on In Vivo Visual Responses by Specific Projections from Mouse Cortical Layer 6 to Dorsal Lateral Geniculate Nucleus. *J. Neurosci.* 35, 9265–9280. <https://doi.org/10.1523/JNEUROSCI.0027-15.2015>.
48. Born, G., Schneider-Soupiadis, F.A., Eriskens, S., Vaiceliunaite, A., Lao, C.L., Mobarhan, M.H., Spacek, M.A., Einevoll, G.T., and Busse, L. (2021). Corticothalamic feedback sculpts visual spatial integration in mouse thalamus. *Nat. Neurosci.* 24, 1711–1720. <https://doi.org/10.1038/s41593-021-00943-0>.
49. Augustinaite, S., and Kuhn, B. (2020). Complementary Ca²⁺ Activity of Sensory Activated and Suppressed Layer 6 Corticothalamic Neurons Reflects Behavioral State. *Curr. Biol.* 30, 3945–3960.e5. <https://doi.org/10.1016/j.cub.2020.07.069>.

50. Li, J.Y., Hass, C.A., Matthews, I., Kristl, A.C., and Glickfeld, L.L. (2021). Distinct recruitment of feedforward and recurrent pathways across higher-order areas of mouse visual cortex. *Curr. Biol.* 31, 5024–5036.e5. <https://doi.org/10.1016/j.cub.2021.09.042>.
51. Gale, S.D., Strawder, C., Bennett, C., Mihalas, S., Koch, C., and Olsen, S.R. (2024). Backward masking in mice requires visual cortex. *Nat. Neurosci.* 27, 129–136. <https://doi.org/10.1038/s41593-023-01488-0>.
52. Resulaj, A., Ruediger, S., Olsen, S.R., and Scanziani, M. (2018). First spikes in visual cortex enable perceptual discrimination. *eLife* 7, e34044. <https://doi.org/10.7554/eLife.34044>.
53. Kühn, N.K., Li, C., Baimacheva, N., Zimmer, J., Reinhard, K., Bonin, V., and Farrow, K. (2025). Dendritic architecture enables de novo computation of salient motion in the superior colliculus. *Curr. Biol.* 35, 3799–3811.e8. <https://doi.org/10.1016/j.cub.2025.06.060>.
54. Bennett, C., Gale, S.D., Garrett, M.E., Newton, M.L., Callaway, E.M., Murphy, G.J., and Olsen, S.R. (2019). Higher-Order Thalamic Circuits Channel Parallel Streams of Visual Information in Mice. *Neuron* 102, 477–492.e5. <https://doi.org/10.1016/j.neuron.2019.02.010>.
55. Fang, Q., Chou, X.L., Peng, B., Zhong, W., Zhang, L.I., and Tao, H.W. (2020). A Differential Circuit via Retino-Colliculo-Pulvinar Pathway Enhances Feature Selectivity in Visual Cortex through Surround Suppression. *Neuron* 105, 355–369.e6. <https://doi.org/10.1016/j.neuron.2019.10.027>.
56. Chou, X.L., Fang, Q., Yan, L., Zhong, W., Peng, B., Li, H., Wei, J., Tao, H.W., and Zhang, L.I. (2020). Contextual and cross-modality modulation of auditory cortical processing through pulvinar mediated suppression. *eLife* 9, e54157. <https://doi.org/10.7554/eLife.54157>.
57. Gale, S.D., and Murphy, G.J. (2014). Distinct Representation and Distribution of Visual Information by Specific Cell Types in Mouse Superficial Superior Colliculus. *J. Neurosci.* 34, 13458–13471. <https://doi.org/10.1523/JNEUROSCI.2768-14.2014>.
58. Furutachi, S., Franklin, A.D., Aldea, A.M., Mrcsic-Flogel, T.D., and Hofer, S.B. (2024). Cooperative thalamocortical circuit mechanism for sensory prediction errors. *Nature* 633, 398–406. <https://doi.org/10.1038/s41586-024-07851-w>.
59. Jordan, R., and Keller, G.B. (2020). Opposing Influence of Top-down and Bottom-up Input on Excitatory Layer 2/3 Neurons in Mouse Primary Visual Cortex. *Neuron* 108, 1194–1206.e5. <https://doi.org/10.1016/j.neuron.2020.09.024>.
60. Hu, B., Garrett, M.E., Groblewski, P.A., Ollerenshaw, D.R., Shang, J., Roll, K., Manavi, S., Koch, C., Olsen, S.R., and Mihalas, S. (2021). Adaptation supports short-term memory in a visual change detection task. *PLOS Comput. Biol.* 17, e1009246. <https://doi.org/10.1371/journal.pcbi.1009246>.
61. Li, J.Y., and Glickfeld, L.L. (2023). Input-specific synaptic depression shapes temporal integration in mouse visual cortex. *Neuron* 111, 3255–3269.e6. <https://doi.org/10.1016/j.neuron.2023.07.003>.
62. Zhang, S., Xu, M., Kamigaki, T., Hoang Do, J.P., Chang, W.-C., Jenvay, S., Miyamichi, K., Luo, L., and Dan, Y. (2014). Selective attention. Long-range and local circuits for top-down modulation of visual cortex processing. *Science* 345, 660–665. <https://doi.org/10.1126/science.1254126>.
63. Pi, H.-J., Hangya, B., Kvitsiani, D., Sanders, J.L., Huang, Z.J., and Kepecs, A. (2013). Cortical interneurons that specialize in disinhibitory control. *Nature* 503, 521–524. <https://doi.org/10.1038/nature12676>.
64. Pfeffer, C.K., Xue, M., He, M., Huang, Z.J., and Scanziani, M. (2013). Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nat. Neurosci.* 16, 1068–1076. <https://doi.org/10.1038/nn.3446>.
65. Niell, C.M., and Scanziani, M. (2021). How Cortical Circuits Implement Cortical Computations: Mouse Visual Cortex as a Model. *Annu. Rev. Neurosci.* 44, 517–546. <https://doi.org/10.1146/annurev-neuro-102320-085825>.
66. Glickfeld, L.L., and Olsen, S.R. (2017). Higher-Order Areas of the Mouse Visual Cortex. *Annu. Rev. Vis. Sci.* 3, 251–273. <https://doi.org/10.1146/annurev-vision-102016-061331>.
67. Rangel-Gomez, M., and Meeter, M. (2016). Neurotransmitters and Novelty: A Systematic Review. *J. Psychopharmacol.* 30, 3–12. <https://doi.org/10.1177/0269881115612238>.
68. Fusi, S., Miller, E.K., and Rigotti, M. (2016). Why neurons mix: high dimensionality for higher cognition. *Curr. Opin. Neurobiol.* 37, 66–74. <https://doi.org/10.1016/j.conb.2016.01.010>.
69. Steinberg, J., Advani, M., and Sompolsky, H. (2021). New role for circuit expansion for learning in neural networks. *Phys. Rev. E* 103, 022404. <https://doi.org/10.1103/PhysRevE.103.022404>.
70. Ranganath, C., and Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nat. Rev. Neurosci.* 4, 193–202. <https://doi.org/10.1038/nrn1052>.
71. Bastos, G., Holmes, J.T., Ross, J.M., Rader, A.M., Gallimore, C.G., Wargo, J.A., Peterka, D.S., and Hamm, J.P. (2023). Top-down input modulates visual context processing through an interneuron-specific circuit. *Cell Rep.* 42, 113133. <https://doi.org/10.1016/j.celrep.2023.113133>.
72. Natan, R.G., Briguglio, J.J., Mwila-bwe-Tshilobo, L., Jones, S.I., Aizenberg, M., Goldberg, E.M., and Geffen, M.N. (2015). Complementary control of sensory adaptation by two types of cortical interneurons. *eLife* 4, e09868. <https://doi.org/10.7554/eLife.09868>.
73. Yarden, T.S., Mizrahi, A., and Nelken, I. (2022). Context-Dependent Inhibitory Control of Stimulus-Specific Adaptation. *J. Neurosci.* 42, 4629–4651. <https://doi.org/10.1523/JNEUROSCI.0988-21.2022>.
74. Nitzan, N., and Buzsáki, G. (2025). Diversity of omission responses to visual images across brain-wide regions. *Sci. Adv.* 11, eadv5651. <https://doi.org/10.1126/sciadv.adv5651>.
75. Nitzan, N., Bennett, C., Movshon, J.A., Olsen, S.R., and Buzsáki, G. (2024). Mixing novel and familiar cues modifies representations of familiar visual images and affects behavior. *Cell Rep.* 43, 114521. <https://doi.org/10.1016/j.celrep.2024.114521>.
76. Harris, B., Gollo, L.L., and Fulcher, B.D. (2024). Tracking the Distance to Criticality in Systems with Unknown Noise. *Phys. Rev. X* 14, 031021. <https://doi.org/10.1103/PhysRevX.14.031021>.
77. Billeh, Y.N., Cai, B., Gratiy, S.L., Dai, K., Iyer, R., Gouwens, N.W., Abasi-Asl, R., Jia, X., Siegle, J.H., Olsen, S.R., et al. (2020). Systematic Integration of Structural and Functional Data into Multi-scale Models of Mouse Primary Visual Cortex. *Neuron* 106, 388–403.e18. <https://doi.org/10.1016/j.neuron.2020.01.040>.
78. Chen, G., Scherr, F., and Maass, W. (2022). A data-based large-scale model for primary visual cortex enables brain-like robust and versatile visual processing. *Sci. Adv.* 8, eabq7592. <https://doi.org/10.1126/sciadv.abq7592>.
79. Aitken, K., Campagnola, L., Garrett, M.E., Olsen, S.R., and Mihalas, S. (2024). Simple synaptic modulations implement diverse novelty computations. *Cell Rep.* 43, 114188. <https://doi.org/10.1016/j.celrep.2024.114188>.
80. Ito, S., Piet, A., Bennett, C., Durand, S., Belski, H., Garrett, M., Olsen, S.R., and Arkhipov, A. (2024). Coordinated changes in a cortical circuit sculpt effects of novelty on neural dynamics. *Cell Rep.* 43, 114763. <https://doi.org/10.1016/j.celrep.2024.114763>.
81. Azabou, M., Arora, V., Ganesh, V., Mao, X., Nachimuthu, S., Mendelson, M., Richards, B., Perich, M., Lajoie, G., and Dyer, E. (2023). A Unified, Scalable Framework for Neural Population Decoding. *Adv. Neural Inf. Process. Syst.* 36, 44937–44956.
82. Zhang, Y., Wang, Y., Benetó, D.J., Wang, Z., Azabou, M., Richards, B., Winter, O., Dyer, E., Paninski, L., et al.; International Brain Laboratory (2024). Towards a “universal translator” for neural dynamics at single-cell, single-spike resolution. Preprint at ArXiv, arXiv:2407.14668v2.
83. Groblewski, P.A., Sullivan, D., Lecoq, J., de Vries, S.E.J., Caldejon, S., L'Heureux, Q., Keenan, T., Roll, K., Slaughterback, C., Williford, A., et al. (2020). A standardized head-fixation system for performing large-scale,

- in vivo physiological recordings in mice. *J. Neurosci. Methods* 346, 108922. <https://doi.org/10.1016/j.jneumeth.2020.108922>.
84. Siegle, J.H., López, A.C., Patel, Y.A., Abramov, K., Ohayon, S., and Voigts, J. (2017). Open Ephys: an open-source, plugin-based platform for multi-channel electrophysiology. *J. Neural Eng.* 14, 045003. <https://doi.org/10.1088/1741-2552/aa5eea>.
85. Peirce, J.W. (2009). Generating stimuli for neuroscience using PsychoPy. *Front. Neuroinform.* 2, 10. <https://doi.org/10.3389/neuro.11.010.2008>.
86. Rudy, B., Fishell, G., Lee, S., and Hjerling-Leffler, J. (2011). Three Groups of Interneurons Account for Nearly 100% of Neocortical GABAergic Neurons. *Dev. Neurobiol.* 71, 45–61. <https://doi.org/10.1002/dneu.20853>.
87. Jia, X., Siegle, J.H., Bennett, C., Gale, S.D., Denman, D.J., Koch, C., and Olsen, S.R. (2019). High-density extracellular probes reveal dendritic backpropagation and facilitate neuron classification. *J. Neurophysiol.* 121, 1831–1847. <https://doi.org/10.1152/jn.00680.2018>.
88. Renier, N., Wu, Z., Simon, D.J., Yang, J., Ariel, P., and Tessier-Lavigne, M. (2014). iDISCO: A Simple, Rapid Method to Immunolabel Large Tissue Samples for Volume Imaging. *Cell* 159, 896–910. <https://doi.org/10.1016/j.cell.2014.10.010>.
89. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>.
90. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
91. Syeda, A., Zhong, L., Tung, R., Long, W., Pachitariu, M., and Stringer, C. (2024). Facemap: a framework for modeling neural activity based on orofacial tracking. *Nat. Neurosci.* 27, 187–195. <https://doi.org/10.1038/s41593-023-01490-6>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
High electroendosmosis agarose	Sigma-Aldrich	Cat #: A9793; CAS: 9012-36-6
Certified Low-Melt Agarose	BIO-RAD	Cat #: 1613112; CAS: 9012-36-6
Tergazyme 1%	Alconox	N/A
Chloromethyl di-alkyl indocarbocyanine (CM-Dil dye, 1 mM in ethanol)	ThermoFisher	Cat #: V22888
Deposited data		
Behavior + spike data	This paper	URI: s3://visual-behavior-neuropixels-data/visual-behavior-neuropixels/ ; DANDI: https://doi.org/10.48324/dandi.000713/0.240702.1725
Behavior training data	This paper	URI: s3://visual-behavior-neuropixels-data/visual-behavior-neuropixels/ ; DANDI: https://doi.org/10.48324/dandi.000713/0.240702.1725
Local Field Potential (LFP) data	This paper	URI: s3://visual-behavior-neuropixels-data/visual-behavior-neuropixels/ ; DANDI: https://doi.org/10.48324/dandi.000713/0.240702.1725
Raw AP band data	This paper	URI: s3://allen-brain-observatory/visual-behavior-neuropixels/raw-data/
Raw LFP band data	This paper	URI: s3://allen-brain-observatory/visual-behavior-neuropixels/raw-data/
Eye, face, body videos	This paper	URI: s3://allen-brain-observatory/visual-behavior-neuropixels/raw-data/
Experimental models: Organisms/strains		
Mouse C57BL/6J	The Jackson Laboratory	JAX: 000664; RRID: IMSR_JAX:000664
Mouse Vip-IRES-Cre	The Jackson Laboratory	JAX: 031628; RRID: IMSR_JAX:031628
Mouse Sst-IRES-Cre	The Jackson Laboratory	JAX: 028864; RRID: IMSR_JAX:028864
Mouse VGAT-ChR2-YFP	The Jackson Laboratory	JAX: 014548; RRID: IMSR_JAX:014548
Mouse Ai32	The Jackson Laboratory	JAX: 024109; RRID: IMSR_JAX:024109
Software and algorithms		
OpenEphys GUI	Siegle et al. ⁸⁴ : https://doi.org/10.1088/1741-2552/aa5eea	Version: v0.4.6
Kilosort 2	Stringer et al. ⁴⁰ : https://doi.org/10.1126/science.aav7893	www.github.com/MouseLand/Kilosort2
Facemap	Syeda et al. ⁹¹ : https://doi.org/10.1038/s41593-023-01490-6	https://github.com/MouseLand/facemap
Analysis Code	This paper	https://github.com/AllenInstitute/VBN-Manuscript

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Animals

Mice were maintained in the Allen Institute animal facility and used in accordance with protocols approved by the Allen Institute's Institutional Animal Care and Use Committee. The main dataset consisted of electrophysiology experiments from 27 C57BL/6J mice (15 males, 12 females), 31 Sst-IRES-Cre; Ai32 mice (16 males, 15 females), and 23 Vip-IRES-Cre; Ai32 mice (16 males, 7 females). All mice were adults at the time of recording (160 ± 29 days; mean $\pm$ standard deviation). Optogenetic perturbation

experiments were performed in four VGAT-ChR2 mice (1 male, 3 females). Four-day electrophysiology experiments were performed in five Vip-IRES-Cre; Ai32 mice (3 male, 2 females). Following surgery, all mice were single-housed and maintained on a reverse 12-hour light cycle in a shared facility with room temperatures between 68° and 72°F and humidity between 30 and 70%. All experiments were performed during the dark cycle. During behavior training, mice were water restricted to 85% of their initial body weight with ad libitum access to food.

METHOD DETAILS

Surgery

Mice received a headpost and cranial window as previously described.^{4,83} Briefly, mice (aged 54 ± 6 days; mean $\pm$ standard deviation) were deeply anesthetized with isoflurane prior to exposing the skull. An incision was made to expose the skull and the head was levelled with respect to pitch, roll and yaw. The headframe was positioned to center the well at 3.1 mm lateral and 1.3 mm anterior to lambda. The headframe was then affixed to the skull using C&B Metabond (Parkell). Once the Metabond was dry, the skull was rotated by 20° to make a circular craniotomy (5 mm diameter) over visual cortex. After removal of the skull, a durotomy was performed and a circular glass coverslip (5 mm diameter with additional 1 mm lip to affix to skull) was lowered into the craniotomy and affixed to the skull with Vetbond. The bottom of the coverslip was coated with a layer of polydimethylsiloxane (SYLGARD 184, Sigma-Aldrich) to reduce adhesion to the brain surface. A ring of Kwik-Cast was added along the perimeter of the coverslip to hold it in place. Following surgery, mice were monitored for overall health and cranial window clarity over a 7-10 day recovery period.

Four mice were implanted with SHIELD implants⁶ instead of glass coverslips to permit recordings over four consecutive days.

Intrinsic Signal Imaging

Intrinsic signal imaging (ISI) was used to measure the hemodynamic response of the cortex to visual stimulation across the entire cranial window.⁴ ISI results were then used to delineate visual cortical area boundaries and guide probe insertions.

Mice were lightly anesthetized with 1-1.4% isoflurane and eye drops were applied (Lacri-Lube Lubricant Eye Ointment). Mice were headfixed for imaging normal to the cranial window. Before ISI, an image of the vasculature was taken under green light to register ISI data to the brain surface. Next the imaging plane was defocused and the hemodynamic response to a visual stimulus was imaged under red light. The visual stimulus was an alternating checkboard pattern (25° square size) superimposed on a 20° wide bar drifting across a mean luminance gray background. On each trial, the stimulus bar was swept across the four cardinal axes 10 times in each direction at a rate of 0.1 Hz.

A minimum of three trials were averaged to produce altitude and azimuth phase maps, calculated from the discrete Fourier transform of each pixel at the stimulus frequency. A sign map was produced from the phase maps by taking the sine of the angle between the altitude and azimuth map gradients. In the sign maps, each cortical visual area appears as a contiguous red or blue region (Figure 1A).

Behavior Training

Water restriction and habituation

Throughout training mice were water-restricted to motivate learning and performance of the behavioral task. Mice had access to water only during behavioral training sessions or when provided by a technician on non-training days. Before water restriction, mice were weighed for three consecutive days to establish a baseline weight. During the first week of water restriction mice were habituated to daily handling and increasing durations of head fixation in the behavior enclosure over a five-day period. Mice were trained 5 days per week (Monday-Friday) and were allowed to earn unlimited water during the daily 1 hour sessions. Supplemental water was supplied in the home cage if the volume earned during the task fell below 1 mL or the body weight fell below 85% of the baseline weight. On non-training days mice were weighed and received enough water to reach 85% of their baseline weight (but never less than 1 mL). For some mice, the target weight was adjusted to 90% of the baseline weight to maintain health.

Apparatus

Mice were trained in custom built behavior enclosures equipped with a 24" gamma-corrected LCD monitor (ASUS, #PA248Q). Mice were head-fixed on a behavior stage and allowed to run on a 6.5" running wheel. The center of the visual monitor was placed 15 cm from the right eye and visual stimuli were spherically warped to correct for distortion at the periphery of the monitor. Water rewards were delivered using a gravity-fed solenoid (NI Research, #161K011) to deliver a calibrated volume of water through a blunted 17g hypodermic needle (Hamilton) positioned approximately 2 mm away from the animal's mouth using a custom 3-axis motorized stage.

Change Detection Task

This task has been previously described in detail.^{12,26} Briefly, mice were trained in a go/no-go change detection task (schematized in Figure 1F). Mice were presented with a continuous stream of natural images (250 ms duration followed by 500 ms of isoluminant gray screen) and were trained to lick a reward spout when the identity of the stimulus changed. If mice responded correctly within a short, post-change response window (150-750 ms after stimulus change) a water reward was delivered. A grace period of 3 seconds occurred after each change to allow for water consumption. During this period no additional changes could occur. At the start of each trial, the change stimulus was selected. Given that there were eight total images, $\sim 7/8$ of trials were GO trials (the change stimulus selected differed from the current stimulus) and $1/8$ were CATCH trials (the change stimulus selected was the same as the

current stimulus). A change time was selected from a geometric distribution ranging from 5 to 12 stimulus presentations. If the mouse licked prior to the stimulus change the trial restarted (aborted). If a given trial was aborted 5 times, a new trial and change time were drawn. No other punishment was delivered for aberrant licking. Licks during the response window were scored as hits on GO trials and false alarms on CATCH trials. During electrophysiology experiments but not during prior training, approximately 5% of image stimuli were omitted, extending the gray screen between stimuli from 500 ms to 1.25 s. The stimulus immediately preceding a change could not be omitted, and two consecutive stimuli were never omitted.

Automated Training

Mice were trained using an automated procedure that comprised six stages. On the first training day (Stage 0), mice were introduced to the rig and the lickspout during a 15 minute open-loop session. Non-contingent 10 μ L water rewards were delivered coincident with 90° changes in the orientation of a full-field static square-wave grating. All subsequent training stages were run in closed-loop and required the mouse to lick for water reward as follows:

- Stage 1: Mice were presented with static, full-field square wave gratings (oriented at 0° or 90°) and rewarded with 10 μ L of water for licking when the orientation of the grating changed. Progression to the next stage required a d-prime of at least 2 for 2 of the last 3 sessions.
- Stage 2: A 500 ms inter-stimulus gray screen was introduced between grating stimuli (250 ms duration; oriented at 0° or 90°). Progression criteria were identical to Stage 1.
- Stage 3: Flashed gratings were replaced with natural images. Progression was triggered after 3 sessions.
- Stage 4: Reward volume was reduced to 7 μ L. Progression required a peak d-prime (calculated over a rolling window of 100 trials) greater than 1 for 3 consecutive sessions and at least 100 contingent (non-aborted) trials.
- Stage 5: A 5 minute epilogue was appended at the end of the session, introducing mice to the gabor stimuli and full field flashes used to map receptive field properties. Reward volume was reduced to 5 μ L. A five minute waiting period was added to the beginning of the session to acclimate mice to experiment preparation time during electrophysiology experiments. Mice were eligible to be handed off for electrophysiology when they achieved a mean reward number of 120 over the last 3 sessions in addition to the Stage 4 criteria.

After training, mice were habituated to the electrophysiology rigs for 5-10 days before experiments during which the water reward was further reduced (3-5 μ L) to prevent satiation.

In a previous dataset from the Allen Institute using the same change detection task,¹² many mice relied at least partially on a timing strategy, licking more than expected by chance about 5 stimuli from the end of the last trial (i.e. the first eligible change image presentation).²⁷ To encourage mice to use a visual strategy in this study, we added a strict reward threshold to the passing criterion for the final training stage. Applying the same dynamic logistic regression model used in Piet et al.,²⁷ we quantified the impact of the added criterion with a strategy index score, indicating whether the mouse relied more heavily on visual (positive score) or timing (negative score) cues. Indeed, we found that the mice in this dataset were significantly more biased towards a visual strategy than mice in the previous dataset (strategy index mean $\pm$ SEM: 9.5 ± 1.3 for this dataset, -5.1 ± 0.59 for Garrett et al.¹²; $p=1.2\text{e-}23$, Wilcoxon rank-sum test).

Neuropixels Recordings

Preparation of the mouse for recording

After the completion of a successful ISI map, a custom insertion window was generated for each mouse.¹ First, six insertion targets were manually drawn on the VISP-aligned eccentricity map using a web-based annotation tool. Targets were positioned at the center of retinotopy for each area. The coordinates of each target were used to automatically generate the outlines of the insertion window, which was subsequently laser-cut out of 0.5 mm clear PETG plastic (Ponoko). When seated in the headframe well, the window facilitates access to the brain via holes over each of the six visual areas.

On the morning of the first recording day, the cranial coverslip was removed and replaced with an insertion window containing holes aligned to six cortical visual areas as previously described.^{1,4} Briefly, on the morning of the first recording day, the mouse was anesthetized with isoflurane (3-5% induction and 1.5% maintenance) and the eyes were protected with ocular lubricant. Body temperature was maintained at 37.5°C. The Kwik-Cast surrounding the cranial window was removed and a silicone suction cup was attached to the window. If necessary, a bent syringe needle was used to free the coverslip from Vetbond, and the cranial window was lifted from the brain surface. The insertion window was then placed above the brain and affixed to the headframe well with Metabond. An agarose mixture was injected underneath the window and allowed to solidify. The mixture consisted of 0.4 g high EEO Agarose (Sigma-Aldrich), 0.42 g Certified Low-Melt Agarose (BioRad), and 20.5 ml ACSF (135.0 mM NaCl, 5.4 mM KCl, 1.0 mM MgCl_2 , 1.8 mM CaCl_2 , 5.0 mM HEPES). A layer of silicone oil was added over the holes in the insertion window to prevent the agarose from drying out. A plastic cap was screwed into the headframe well to protect the preparation from cage debris. After this procedure (~30 minutes), mice were returned to their home cage and allowed to recover for at least 2 hours (typically 3-4) before the recording experiment.

After the first recording session, the probes were retracted, silicone oil was removed from the window surface, and the well was sealed with additional agarose added on top of the implant window. The protective cap was re-attached and the mouse was returned

to its home cage. On the following day, the mouse was returned to the rig, the excess agarose was removed from the implant window and a fresh layer of silicone oil was applied before reinsertion of the probes for the second recording. Receptive fields generated from the first recording session for each visual area were used to adjust the probe insertion targets if necessary to maximize RF overlap with the center of the screen.

For mice with SHIELD implants, the mouse was briefly anesthetized on the Friday before the experimental week, and the SORTA-Clear plug protecting the implant was removed and replaced with Kwik-Cast. On the morning of each recording day, the Kwik-Cast plug was removed and the implant was covered with agarose and a thin layer of silicone oil. At the end of the recording, the agarose was removed and replaced with a fresh layer of Kwik-Cast.

Probes

All neural recordings were performed with Neuropixels 1.0 probes.² The electrodes closest to the tip were always used, providing a maximum of 3.84 mm of tissue coverage. All recordings were referenced to the large pad at the tip of the probe. Our goal was to insert six probes per experiment.

Each probe was coated in CM-Dil (1 mM in ethanol; Thermo Fisher, V22888) to visualize probe tracks after tissue clearing and ex vivo imaging. The probes were dipped in a well filled with the dye for approximately 1 minute.

Head fixation

The mouse was placed on the running wheel and fixed to the headframe clamp. The plastic cap was removed from the headframe well and an aluminum cone was lowered to prevent the mouse's tail from striking the probes. An infrared dichroic mirror was placed in front of the right eye to allow the eye-tracking camera to operate without interference from the visual stimulus. A black curtain was then lowered over the front of the rig, placing the mouse in complete darkness except for the visual stimulus monitor.

Grounding

A 32 AWG silver wire (A-M Systems) was epoxied to the headframe before the initial headframe/cranial window surgery. This wire becomes electrically conductive with the brain surface after the application of the ACSF/agarose mixture beneath the insertion window. The wire was pre-soldered to a gold pin embedded in the headframe well, which mates with a second gold pin on the protective cone. The cone pin was soldered to 22 AWG hook-up wire (SparkFun Electronics), which was connected to both the behavior stage and the probe ground. Before the experiment, the brain-to-probe ground path was checked using a multimeter.

The reference connection on the Neuropixels probes was permanently soldered to ground using a silver wire, and the headstage grounds (which are contiguous with the Neuropixels probe grounds) were connected in parallel to the animal ground.

Probe Insertion

After the mouse was secured in the headframe, the cartridge holding the probes was lowered so the probe tips were approximately 2.5-mm above the brain surface. The probes were then manually lowered one by one to the brain surface until spikes were visible on the electrodes closest to the tip. After the probes penetrated the brain to a depth of around 100 μm , they were inserted automatically at a rate of 200 $\mu\text{m min}^{-1}$ (total of 3.5 mm or less into the brain) to avoid damage caused by rapid insertion. After the probes reached their targets, they were allowed to settle for at least 10 minutes.

Data acquisition and synchronization

Neuropixels data was acquired at 30 kHz (spike band) and 2.5 kHz (LFP band) using the Open Ephys GUI.⁸⁴ Gain settings of 500 $\times$ and 250 $\times$ were used for the spike band and LFP band, respectively. Probes were connected to a PXIe card inside a National Instruments chassis (Neuropixels 1.0). Raw neural data was streamed to a compressed format for archiving, which was extracted before analysis.

Videos of the eye, body and face were acquired at 60 Hz. The angular velocity of the running wheel was recorded at the time of each stimulus frame, at approximately 60 Hz. Synchronization signals for each frame were acquired by a dedicated computer with a National Instruments card acquiring digital inputs at 100 kHz, which was considered the master clock. A 32-bit digital 'barcode' was sent with an Arduino Uno (SparkFun DEV-11021) every 30 s to synchronize all devices with the neural data. Each Neuropixels probe has an independent sample rate between 29,999.90 Hz and 30,000.31 Hz, making it necessary to align the samples offline to achieve precise synchronization. The synchronization procedure used the first matching barcode between each probe and the master clock to determine the clock offset, and the last matching barcode to determine the clock scaling factor.

Lick Detection

Licks were detected with a capacitive lick sensor during training. However, this sensor was incompatible with Neuropixels recordings due to the electrical artifact generated during licking. To address this issue, a custom piezo lick sensor was designed for electrophysiology rigs consisting of a non-conductive PEEK needle (Hamilton) resting on a piezo microphone (TE Connectivity, 1007079-1). Licks were detected by the physical displacement of the spout during contact with the tongue. To add rigidity, the PEEK needle was reinforced either with a wooden applicator or by a stainless-steel sleeve (Hamilton) which stopped approximately a centimeter from the end of the needle. The analog microphone signal was digitized with a SAMD21 microprocessor and band-pass filtered between 3 and 35 Hz. The resulting signal was then passed through a software comparator to generate the final TTL lick signal.

Stimulus Monitor

Visual stimuli were generated using custom scripts based on PsychoPy⁸⁵ and were displayed using an ASUS PA248Q LCD monitor, with 1,920 $\times$ 1,200 pixels (55.7 cm wide, 60 Hz refresh rate). Stimuli were presented monocularly, and the monitor was positioned 15 cm from the right eye of the mouse and spanned 120 $^\circ \times$ 95 $^\circ$ of visual space before stimulus warping. Each monitor was gamma corrected and had a mean luminance of 50 cd m^{-2} . To account for the close viewing angle of the mouse, a spherical warping was

applied to all stimuli to ensure that the apparent size and spatial frequency were constant across the monitor as seen from the mouse's perspective.

Probe removal and cleaning

After each experiment, the probes were retracted from the brain at a rate of ~ 1 mm/s and the mouse was removed from head fixation and returned to its home cage. The probes were then immersed in a well of 1% Tergazyme for ~ 12 hours.

Quality Control for Neuropixels recording sessions

Out of a total of 180 sessions, 27 were excluded from the release due to the following QC failures: white foam buildup on the edge of the eye partially covering the pupil (1), software failures compromising critical data streams (2), visual stimulus synchronizing failure (2), cortical bleeding or compromised brain health resulting in low unit activity and/or atypical visual responses (18), gap in data acquisition (2), discovery of purulent material over right hemisphere during ex-vivo imaging (2).

Of the remaining 153 sessions, 50 were released with a quality control flag and omitted from the analysis in this manuscript. We flagged two potential abnormalities as follows:

Abnormal histology: Ex-vivo imaging revealed potential tissue damage in 27 mice. Individual brains were flagged for 1) discoloration indicating large bleeds or hypoxic tissue, 2) swelling of the hippocampus, 3) tissue bruising, or 4) skull deformity compressing the right cortex (contralateral to window). All recording sessions were flagged for these mice, as acute events like large subcortical bleeds or diffuse bruising could not be confidently ascribed to the first or second recording day (49 sessions total).

Abnormal activity: We flagged 12 sessions as having potential epileptiform activity, characterized by a sharp increase in the firing rate across the probe (lasting 1–5 seconds) followed by a sustained suppression of activity (lasting several minutes) for at least one probe during the session. Of these 12 sessions, 11 were also flagged for abnormal histology.

The remaining 103 sessions (with no QC failures or flags) were used for the analyses presented in this paper.

Experimental design and visual stimuli

Every experiment followed the timeline diagrammed in [Figure 1E](#). Briefly, mice encountered the following stimulus blocks (in order of appearance):

Active behavior block: the mouse performed the detection of change task described above. The images used during this block came from one of two image sets of 8 natural images (G or H). On one recording day, mice were exposed to the same image set they had seen during behavior training ('Familiar' sessions). On the other recording day, mice were shown a novel image set ('Novel' sessions). Two images were shared across these image sets ('holdover' images) and were thus familiar on both sessions. In total, 41 mice were trained on image set G, and shown G on the first recording day and H on the second; 10 mice were trained on H, and shown H on the first recording day and G on the second; 3 mice were trained on G and shown H on the first recording day and G on the second.

Images were presented for 250 ms, with a 500 ms gray screen between images. During recording sessions, images were omitted (resulting in a gray screen lasting 1.25 seconds) with a probability of 5%. However, the image immediately preceding the change image and the image immediately following an omission were never omitted, producing a true omission probability of $\sim 3\%$. At the end of this block, the lick spout was retracted for the remainder of the session.

Gabor stimuli: Gabor stimuli windowed to have a 20° diameter were presented in a 9×9 grid to map receptive fields. Grid positions were spaced by 10 degrees. Gabors had a temporal frequency of 4 Hz, a spatial frequency of 0.08 cycles/degree and were shown at 3 orientations (0, 45, 90 degrees). They were 250 ms in duration without gaps between stimuli. There were 15 trials for each condition (81 positions, 3 orientations).

Gray screen: Following the Gabor stimuli, there were 5 minutes of gray screen without stimuli to measure spontaneous activity.

Full-field flashes: Flashes were black or white at 80% contrast. They were 250 ms in duration with 2 seconds between flash starts. There were 75 trials for each condition (light and dark flashes).

Passive replay: The stimuli encountered by the mouse during the active behavior block were re-shown frame-for-frame. Note however that the lick spout was retracted for this block and the mouse was therefore unable to earn rewards.

Optotagging

Optotagging Protocol

At the end of every experiment (regardless of genotype), an optotagging protocol was run during which the cortical surface was stimulated with blue light. In *Sst-IRES-Cre;Ai32* and *Vip-IRES-Cre;Ai32* mice, this protocol allowed us to identify putative Sst+ and Vip+ cortical interneurons by a dramatic increase in spiking activity time-locked to photostimulation. We chose to focus on these two cell classes because they make up two of the three major non-overlapping populations of cortical interneurons.⁸⁶ The third major class (parvalbumin-expressing cells), can largely be distinguished by their narrow spike waveforms.⁸⁷ Blue light was delivered by a 473 nm laser (Laser Quantum, model Ciel or Cobolt model 06-MLD). The light source was coupled to a 400 μ m diameter fiber optic cable (Thorlabs) or bifurcated fiber bundle (Thorlabs, BFYL4LF01), with the tip(s) positioned such that blue light illuminated the entire cranial window. Two types of stimuli at 3 different light levels were randomly interleaved: a 10 ms pulse, and a 1 s raised cosine ramp. For the pulse stimulus, a 0.5 ms ramp was applied at the beginning and end of the pulse. Stimuli were presented at intervals of 1.5 s plus a uniformly distributed delay between 0 and 0.5 s.

Cell type label assignment

To assign putative cell type labels to units recorded in *Vip-IRES-Cre;Ai32* or *SST-IRES-Cre;Ai32* mice, we defined criteria based on the following metrics characterizing each unit's response to the laser stimuli. Note that, for all analysis, we ignored the first 1.5 ms after laser onset and the last 1.5 ms before laser offset to avoid potential light artifacts in the recording, yielding 7 ms for analysis of the 10 ms laser pulse and 997 ms for the 1 second cosine ramp stimulus. We defined four criteria, requiring that optotagged neurons exhibit fast, reliable spiking to the laser pulse onset and sustained firing for the long cosine-ramp stimulus.

Z-scored response to the 10 ms pulse: The baseline spike rate was determined by randomly drawing 7 ms windows from the inter-trial intervals (ignoring activity within 5 ms of laser onset and offset) and counting spikes for these bins. We required that the mean spike count during the high power 10 ms pulse stimulus be greater than the mean rate over these baseline windows plus two standard deviations.

Response latency for the 10 ms pulse: For each trial of the high power 10 ms pulse condition, we measured the time of the first spike after laser onset. We required that the median latency be less than 8 ms.

Response jitter for 10 ms pulse: We required that the median absolute deviation of first spike times as defined above be less than 2 ms.

Fraction of time with significantly elevated spike rate for the cosine ramp stimulus: Baseline windows (997 ms duration) were randomly drawn from intertrial intervals as described above and spikes were counted in 10 ms bins for each window. The response to the high power cosine stimulus was also binned in 10 ms bins. For each bin, a Mann-Whitney U test was performed to compare the spike counts from the baseline and stimulus windows. We required that at least 30% of the bins be significantly ($p < 0.01$) elevated from baseline. Note that this is a conservative criterion and may exclude Chr2-expressing neurons with only a transient response to the laser.

In addition to these cell type labels, we defined regular spiking (RS) units as those that did not pass the above criteria and had a trough-to-peak duration of greater than 0.4 ms for the mean spike waveform at the peak channel (the channel with the greatest difference between the peak and trough). Fast spiking (FS) units were required to have a peak-to-trough duration of less than 0.4 ms.

Ex Vivo Imaging

Tissue Clearing

Mice were anesthetized (5% isoflurane for induction), then perfused with 4% paraformaldehyde (PFA). The brains were preserved in 4% PFA, rinsed with PBS the next morning, and stored at 4°C in PBS. A modified iDISCO method was then used to clear the brains.⁸⁸ On the first day, the brains were immersed in increasing concentrations of methanol (20, 40 and 60%) for an hour each, then overnight in 80% methanol. On the second day, they were dipped in 100% methanol (twice for one hour) and then into a mixture of 1/3 methanol and 2/3 dichloromethane overnight. On the third day, the brains were moved from pure dichloromethane (2 × 20 min) to pure dibenzyl ether, where they remained for 2-3 days until clearing was complete.

Optical projection tomography

Whole-brain 3D imaging was accomplished with optical projection tomography (OPT), as described previously.⁴ The OPT instrument consisted of collimated light sources for transmitted illumination (on-axis white LED, Thorlabs MNWHL4 with Thorlabs SM2F32-A lens and Thorlabs DG20-600 diffuser) or fluorescence excitation (off-axis Thorlabs M530L3, with Thorlabs ACL2520U-DG6-A lens and Chroma ET535/70m-2P diffuser), a 0.5× telecentric lens (Edmund Optics 62-932) with emission filter (575 nm LP, Edmund Optics 64-635), and a camera (IDS UI-3280CP). The specimen was mounted on a rotating magnetic chuck attached to a stepper motor, which positioned it on the optical axis within a glass cuvette filled with dibenzyl ether. The stepper motor and illumination triggering were controlled with an Arduino Uno (SparkFun DEV-11021) and custom shield including a Big Easy Driver (SparkFun ROB-12859). Instrument communication and image capture was accomplished with MicroManager.

A series of 400 images were captured with transmitted LED illumination. The specimen was rotated 0.9° between images (thus making one full rotation through the image series). The same procedure was repeated with the fluorescence excitation LED. Each channel was stored as a separate OME-TIFF dataset before extracting individual planes and metadata required for reconstruction using a custom Python script.

Isotropic 3D volumes were reconstructed from these projection images using NRecon (Bruker). The rotation axis offset and region-of-interest bounds were set for each image series pair using the transmitted channel dataset, then the same values applied to the fluorescence channel dataset. A smoothing level of 3 using a Gaussian kernel was applied to all images. Reconstructions were exported as single-plane 16-bit TIFF images taken along the rotation axis with final voxel size of 7.9 μm per side.

Registering probes to the Common Coordinate Framework

Reconstructed brains were registered to the Allen Institute Common Coordinate Framework (CCFv3) as previously described.⁴ Reconstructed volumes were downsampled to 10 μm per voxel and roughly aligned to the CCFv3 template brain using an affine transform. The volume was then cropped to a size of 1,023 × 1,024 × 1,024 and converted to Drishti format (<https://github.com/nci/drishti>). Next, 6–54 registration points were marked in up to 14 coronal slices of the individual brain by comparing to the CCFv3 template brain. Fluorescent probe tracks were manually labelled in coronal slices of the individual brain, and the best-fit line was found using singular value decomposition. The registration points were used to define a 3D nonlinear transform (VTK thinPlateSplineTransform), which was used to translate each point along the probe track into the CCFv3 coordinate space. Each CCFv3 coordinate corresponds to a unique brain region, identified by its structure acronym (for example, CA3, LP, VISp, etc.). A list of CCFv3

structure acronyms along each track was compared to the physiological features measured by each probe (for example, unit density, LFP theta power, visual responsiveness). The locations of major structural boundaries were manually identified to align the CCFv3 labels with the physiology data. After the manual alignment procedure, each recording channel (and its associated units) was assigned to a unique CCFv3 structure. White matter structures were not included; any units mapped to a white matter structure inherited the grey matter structure label that was immediately ventral along the probe axis.

Because two recordings were made in most animals (maximum of 12 insertions in each brain), information about the relative position of each probe on the first and second insertions was required to assign an annotated probe track to its appropriate recording session. Two data sources were consulted to make this judgement: 1) the operator's record of where each probe was inserted relative to surface vasculature landmarks for each recording day (annotated as described in *Identification of cortical visual area targets*), and 2) the record of the movement of each probe's manipulator. Together, these records were used to estimate the position of a given probe on the first recording day relative to the second and thus assign annotated probe tracks to their corresponding session.

Identification of cortical visual area targets

To confirm the identity of the cortical visual areas, images of the probes taken during the experiment were compared to images of the brain surface vasculature taken during the ISI session. Key points were selected along the vasculature on both images, and a perspective transform (OpenCV) was performed to warp the insertion image to the retinotopic map. The probe entry points were then manually annotated, and an area was assigned to each probe based on the ISI-derived visual area boundaries. Penetration points that could not be unambiguously associated with a particular visual area were classified as 'VIS'. If the cortical area label obtained via CCFv3 registration did not match the area identified in the insertion image overlay, the insertion image overlay took precedence.

Cortical layer labels

CCFv3 coordinates were used as indices into the template volume to extract layer labels for each cortical unit (L1, L2/3, L4, L5, or L6). The relative thickness of each layer, which can vary both within and across areas, is based on the average of the 1,675 individual brains used to create the template volume.

Probe annotation quality control

For 41 out of 905 insertions, probe tracks could not be confidently annotated in the imaging volume. For these tracks, the beginning and end of cortex were identified using electrophysiological hallmarks (e.g. unit density, LFP power, visual responsiveness), and cortical units were assigned to regions based on ISI as described above. All subcortical units for these probes were assigned to 'grey'.

Spike Sorting

To preserve consistency with the Allen Institute Visual Coding Neuropixels data release, all data processing and spike-sorting was performed according to the pipeline described previously (https://github.com/AllenInstitute/ecephys_spike_sorting).⁴

Data pre-processing and spike-sorting

Before spike-sorting, the spike-band data passed through four steps: offset removal, median subtraction, filtering and whitening. First, the median value of each channel was subtracted to center the signals around zero. Next, the median across simultaneously sampled channels was subtracted to remove common-mode noise. For each sample, the median value of channels $n:24:384$, where $n = [1, 2, 3, \dots, 24]$, was calculated, and this value was subtracted from the same set of channels. The original data were overwritten with the median-subtracted version, with the median value of each block of 16 channels saved separately, to enable reconstruction of the original signal if necessary. The median-subtracted data file is sent to the Kilosort2 MATLAB package (<https://github.com/mouseland/Kilosort2>, commit 2fba667359dbd0e52e67fa848f197e44cf5ef),⁴⁰ which applies a 150-Hz high-pass filter, followed by whitening in blocks of 32 channels. The filtered, whitened data are saved to a separate file for the spike-sorting step.

Spike-sorting quality control

Units were automatically classified as noise based on three criteria: waveform spread either restricted to a single channel or extending over more than 25 channels, waveform shape (no peak and trough, based on wavelet decomposition), or multiple spatial peaks (waveforms are non-localized along the probe axis). A suite of waveform metrics was then calculated to further refine non-noise units. The following metrics were used to filter units for analysis in this study:

Presence ratio: The session was divided into 100 equal-sized blocks; the presence ratio is defined as the fraction of blocks that include one or more spikes from a particular unit. Units with a low presence ratio are likely to have drifted out of the recording or could not be tracked by Kilosort2 for the duration of the experiment. Units were required to have a presence ratio greater than 0.9.

ISI (interspike intervals) violation ratio: This metric searches for refractory period violations that indicate a unit contains spikes from multiple neurons. It is defined as the ratio of the rate of contaminating spikes to the overall spike rate for a given unit and is calculated by counting the number of violations of less than 1.5 ms, dividing by the amount of time for potential violations surrounding each spike, and normalizing by the overall spike rate. It is always positive (or 0), but has no upper bound. Units were required to have an ISI violation ratio less than 0.5.

Amplitude cutoff: This metric provides an approximation of a unit's false negative rate. First, a histogram of spike amplitudes is created, and the height of the histogram at the minimum amplitude is extracted. The percentage of spikes above the equivalent amplitude on the opposite side of the histogram peak is then calculated. If the minimum amplitude is equivalent to the histogram

peak, the amplitude cutoff is set to 0.5 (indicating a high likelihood that more than 50% of spikes are missing). This metric assumes a symmetrical distribution of amplitudes and no drift, so it will not necessarily reflect the true false negative rate. Units were required to have an amplitude cutoff of less than 0.1.

VISp silencing experiment

To silence VISp activity, a multimode optic fiber (250 μm diameter, ThorLabs) was placed above the ISI-defined retinotopic center of V1 in VGAT-ChR2 mice expressing channelrhodopsin in cortical inhibitory neurons. The fiber was connected to a 488 nm blue laser calibrated to deliver ~ 2.5 mW at the fiber tip. Light was delivered on 30% of trials, turning on at one of four intervals from stimulus change onset: -12 ms (range: -13, -11), 39 ms (37, 39), 72 ms (70, 72) and 105 (104, 106). One mouse licked reliably for the laser onset and was excluded from analysis.

Data processing

Running wheel data

Running speed was monitored using an analog encoder mounted at the center of a 6.5" diameter running wheel. To correct for noise in the raw voltage signal, outlier values (more than 10 standard deviations from the mean) were removed and the final speed trace was lowpass filtered with a 10 Hz zero-phase Butterworth filter (`scipy.signal.filtfilt`⁸⁹).

Video data

A standardized pipeline was built for fitting ellipses to the pupil, eye (visible perimeter of the eyeball), and corneal reflection of the right eye, based on points tracked using the open source software DeepLabCut (<https://github.com/DeepLabCut/DeepLabCut>) as previously described.¹² We used DeepLabCut, initialized with a pre-trained ResNet 50 deep residual network, to track (up to) 12 points along the perimeters of the eye, pupil, and corneal reflection. Ellipses were then fit to the tracking points. Validation against hand-annotated 'ground truth' frames confirmed that a single 'universal' model, trained on a broad selection of data samples, robustly generalized on held-out data across different physiology rigs and individual animals.

It was assumed that the pupil is round but when viewed obliquely appears as an ellipse, the major axis of which reflects the pupil diameter. Thus, the area of the pupil is calculated on every frame as the area of a circle with a diameter defined by the longest axis of the ellipse fit for that frame. The area of the ellipses for both the corneal reflection and eye ellipse fits are calculated using the standard formula for the area of an ellipse. On frames where the animal is blinking, tracked points may be missing or the confidence of tracked points may be low, and ellipse fits either fail completely or fit erroneously. To avoid including these erroneous fits in analysis, an algorithm attempts to identify these 'likely blink' frames. Likely blinks are identified as frames where either the eye or pupil fit is missing, or where the z-scored value of the eye or pupil area exceeds 3. In addition, two frames before and after every likely blink identified by the above methods are also labeled as likely blinks. This is to avoid the possibility of analyzing erroneous fits caused by a partially opened eye. Note that this algorithm may flag not only genuine blink frames but also frames where the DeepLabCut algorithm simply failed to identify a reasonable fit.

QUANTIFICATION AND STATISTICAL ANALYSIS

For all analyses, only non-noise units from sessions without QC flags and which passed the presence ratio, ISI violation ratio and amplitude cutoff criteria were included.

Receptive Fields

Receptive fields (Figure 1) were calculated for each unit by taking the mean spike rate in the 250 ms following the presentation of a gabor patch in each of 81 locations (arranged in a 9×9 grid; see *Experimental design and visual stimuli* above). For each session, receptive fields for units in each visual area were averaged together to create a population receptive field. A 2D gaussian was then fit to this population receptive field and the peak position was taken as the receptive field center.

Hit rate, false alarm rate and d-prime

To calculate behavioral response rates, trials were first filtered for engagement. Engaged trials were defined as those for which the reward rate was at least 2 rewards per minute over the preceding 25 trials. Hit rates were then calculated as the fraction of engaged 'Go' trials for which the mouse licked within the response window after the change stimulus. False alarm rates were defined as the fraction of engaged 'Catch' trials for which the mouse responded after a 'sham change' stimulus. To properly compare hit and false alarm rates, sham changes were drawn from the same geometric distribution as change stimuli. For the analysis in Figure 1H, trials were aggregated across the two recording days before analysis. For one mouse, response rates were too low to reliably estimate the false alarm rate and the data from this animal were excluded. D-prime was calculated as the difference between the inverse of the cumulative distribution function (`scipy.stats.norm.ppf`) at the hit and false alarm rates.

Correlation of VISp interneuron activity with locomotion

To quantify how VISp FS, VIP and SST units were modulated by locomotion, we correlated neural activity during the spontaneous epoch with running speed. Spiking was binned in 100 ms bins and running was aligned to these bins by resampling through linear

interpolation. To create a cross-correlogram, we calculated the Pearson's correlation between each unit's activity and the running speed at each of 101 lags spanning ± 5 seconds.

Regression Model

The regression model was adapted from that used in Garrett et al.¹² to be compatible with electrophysiology data. Briefly, we fit neural activity with a linear regression model featuring time-dependent kernels. Each model feature was a vector $f_i(t)$, which was convolved with a learned kernel $k_i(t)$ to produce the predicted model component for that feature. Model components were then summed together to produce the full model response. For continuous features (running speed and pupil diameter) the feature vectors were timeseries. For discrete features, $f_i(t)$ was 1 on timesteps where that feature occurred and 0 elsewhere. The complete list of features and their associated kernel widths is as follows:

Hits

1 at stimulus onset for change stimuli during hit trials and 0 elsewhere; 1.5 second kernel width with 0 second offset.

Misses

1 at stimulus onset for change stimuli during miss trials and 0 elsewhere; 1.5 second kernel width with 0 second offset.

Omissions

1 at onset of omitted stimuli (i.e. when the stimulus would have otherwise occurred) and 0 elsewhere; 1.5 second kernel width with 0 second offset.

Running

continuous time series of running wheel data; 2 second kernel width with a -1 second offset such that the kernel spans $(-1, 1)$ seconds around the current time point.

Pupil

continuous time series of pupil diameter; 2 second kernel width with a -1 second offset.

Licks

1 for time bins during which a lick was registered and 0 elsewhere; 2 second kernel width with a -1 second offset.

Images

(one for each unique image, totaling 8): 1 at stimulus onset for all presentations of the respective image and 0 elsewhere; 0.75 second kernel width with 0 second offset.

The model thus took the form $y = Wx$, where y is the 1d time series for a given unit (spiking activity binned in 25 ms bins), x is a vector of all kernel weights concatenated together, and W is a Toeplitz matrix with diagonal bands that map kernel weights onto time-shifted features. To solve for the kernel weights, we added an L2 ridge regression penalty to the standard solution for ordinary least squares regression, resulting in the following solution:

$$x = (W^T W + \lambda I)^{-1} W^T y,$$

where λ is the L2 penalty. The model was fit with 5-fold cross-validation: we split each session into 50 intervals and randomly assigned 10 intervals to each cross-validation fold (each fold was intermingled in time). The hyper-parameter λ was selected for each session as the value (ranging from 0 to 500) that resulted in the best test-set performance across all units in that session. The training/test sets for hyper-parameter selection were different than those used for fitting the model for analysis.

To evaluate the unique contribution of each kernel to a given unit's activity, we fit a series of reduced models where individual kernels or groups of kernels (components) were removed from the model and the model was refit. We then computed a dropout score, which measures the absolute contribution of a given feature i to the variance explained for each unit: $VE_{full} - VE_{reduced,i}$.

To measure how well the GLM captured trial-averaged activity, unit spiking was binned in 25 ms bins to match the GLM prediction time course. Peri-stimulus time histograms (PSTH) were then made aligning both the observed neural activity and the GLM predicted activity for each unit to stimuli in four trial types: hit, miss, correct reject and false alarm. Observed and predicted activity traces were resampled by linear interpolation to align to stimulus start times. We then concatenated the four resulting trial-type PSTHs for observed and predicted activity separately and calculated what fraction of variance in the observed trial-averaged responses could be explained by the predicted trial-averaged responses.

Functional Clustering

Neurons were grouped into functional clusters based on their linear regression kernel weights (see *Regression Model* section). The neurons with total explained variance less than 0.05 were automatically assigned to a "non-coding" cluster. We clustered the remaining units based on the following kernels: the preferred image for each unit, omissions, hits, misses, running, pupil diameter, and licks. The feature matrix for clustering was generated by concatenating the kernel weights for all responsive units ($N_{neuron} \times 443_{weights}$). An iterative algorithm consisting of an outer loop and an inner loop was then used for the clustering process. Briefly, for each step of the outer loop, the inner loop iterated through all existing clusters, attempting to split each one. These candidate splits were then evaluated with a silhouette score and the best was chosen, producing $N+1$ clusters for the next iteration of the outer loop.

In more detail, the inner loop attempted to split the i th cluster as follows:

- (1) Uniform Manifold Approximation and Projection (UMAP) was initialized with randomly selected parameters to reduce the feature matrix to 3 dimensions, resulting in an $N_i \times 3$ matrix, where N_i is the number of units in the i th cluster.
- (2) A random sample of 5000 units was selected (if N_i was less than 5000, all units were used)
- (3) Spectral clustering (scikit-learn⁹⁰) with random seed initialization was used to divide the sample units into two clusters.
- (4) A random forest classifier (scikit-learn) was trained on the output labels of the sample units from step (4) to label the remaining units.
- (5) Steps 1–4 were repeated 21 times. The cluster assignments for all 21 iterations were saved for each unit. The pattern of cluster assignments for a given unit can be considered a 21-digit binary code. The two most common codes across neurons defined the two most consistent clusters over the 21 clustering iterations. Each neuron was assigned to the cluster ('code') with which it shared the most matching bits (Hamming similarity).

The outer loop was terminated based on a combination of the silhouette score and visual inspection of the clusters.

To assess the robustness of the clusters, we ran a bootstrapped classifier on the same feature matrix used for clustering. For each classification run (100 total iterations), we used UMAP to reduce the dimensionality of the feature matrix to $N_{units} \times 3$. A Random Forest classifier (scikit-learn RandomForestClassifier) was then trained on the cluster labels of the training set (20% of the units, randomly selected for each iteration) and subsequently used to predict labels for the test set (remaining 80% of units). Finally, a confusion matrix was constructed from the test set predictions by pooling across all 100 iterations (Figure S5).

To assess how inclusion of additional behavior regressors in the GLM might influence our sensory/action cluster assignments, videos of the face were analyzed with Facemap^{40,91} to extract 500 motion singular vectors (SVs). We then constructed a second GLM using these SVs as features, time-shifting them from -100 ms to 100 ms in 25 ms steps to account for potential time shifts in their relation to neural activity. This model was fit to the residuals from the previously described full model. Ridge regression, implemented with 5-fold cross-validation (using a 60% training, 20% testing, and 20% evaluation split per fold), was used to prevent overfitting. We then tested what fraction of the residual variance could be explained by this second model.

Peri-stimulus/event histograms

Unit spiking activity was first collected in a response tensor for each session as follows: for every stimulus in the behavioral task, spikes during the interval from stimulus onset to the next stimulus onset (750 ms) were binned in 1 ms bins and binarized, resulting in a Boolean tensor of shape [number of units $\times$ number of stimuli $\times$ 750]. Trials of the desired type were then selected from this tensor and averaged for each unit. Unit responses were further filtered by area, cell type, layer and/or cluster as appropriate and aggregated across sessions before averaging. Unless otherwise noted, unit responses were smoothed with a symmetric exponential filter ($\tau = 3$ ms) before averaging across units.

For the lick-triggered PETHs in Figure 2, unit activity was aligned to lick times during non-change stimuli. Shuffled lick responses were computed to identify what fraction of lick-triggered activity might be attributed to task events (e.g. visual stimuli) that correlated with licking. To create shuffled lick responses, lick times relative to stimulus onset were shuffled across trials and the lick triggered average was recomputed.

For run-triggered peri-event histograms (PETHs) in Figure 2, the above tensor was not used. Rather, acceleration and deceleration events were identified from the running wheel data, and the spiking activity for each unit was binned in 1 ms bins and aligned to these events and averaged. Acceleration events were identified as follows: 1 second windows of running data were extracted for which the mean speed in the first half was less than 1 cm/s and the mean speed for the second half was greater than 5 cm/s. For each window, the acceleration event time was then defined as the time of maximum difference between adjacent time points (if the speed at this point was less than 1 cm/s) or the last point for which the running speed was less than 1 cm/s between the time of the minimum speed and the time of maximum difference. Deceleration events were similarly defined but now for windows for which the speed reduced from greater than 5 cm/s in the first half to less than 1 cm/s in the second half. Within these windows, the deceleration event time was defined as the point of maximum deceleration (minimum speed difference at adjacent time points) after the peak speed was reached. Unit responses were smoothed with a symmetrical exponential filter ($\tau = 3$ ms) before averaging. These PETHs were calculated during the passive viewing epoch to avoid the correlation between running and licking present during active behavior.

Unless otherwise noted, PSTHs were baseline subtracted: the mean activity during the 50 ms before stimulus onset was subtracted from the stimulus response for each unit before averaging across units. PSTHs were always computed using engaged behavior trials.

The PSTH for visual cortex units in Figure 3 was computed for change stimuli during hit trials. Units in all six visual cortical areas were combined.

PSTHs in Figure 4 were computed for change, pre-change and passive change stimuli as follows: change stimuli were taken from hit trials; pre-change stimuli were defined as the stimuli immediately preceding change stimuli for hit trials. Change and pre-change stimulus responses were computed for each image separately and then averaged across images for each unit. Passive change stimuli were calculated for the same change stimuli but during passive replay.

PSTHs to familiar and novel stimuli in Figure 5 were computed for change stimuli from hit trials. Only non-holdover images (i.e. the 6 images unique to each image set) were used to compute both familiar and novel image responses. The PSTHs for familiar and novel stimuli in Figure 6 were computed using non-change stimuli, defined as stimuli during engaged behavior that 1) occurred at least 5

presentations since the last change (and were therefore outside the reward consumption window), 2) were not omitted, 3) did not immediately follow an omitted stimulus, and 4) occurred at least two stimulus presentations from the last lick response.

Image, change and lick decoding

Decoding from sensory and action clusters in [Figure 3](#) was performed on pseudopopulations of neurons aggregated across sessions as follows:

For each session, neurons of the relevant cluster and region were identified, and stimulus presentations meeting the decoding category criteria were selected as illustrated in [Figure 3B](#). Briefly, for image decoding, non-change presentations of each of the eight images were selected for which the mouse did not lick; for change decoding, non-change stimuli for which the mouse licked were compared to change stimuli for which the mouse licked; and for lick decoding, non-change stimuli for which the mouse did not lick were compared with non-change stimuli for which the mouse did lick. These stimuli were selected to isolate the axis of interest for each decoding analysis (image identity, change vs non-change, and lick vs no lick). All stimuli were taken from periods of engaged behavior (defined as trials during epochs for which the reward rate was ≥ 2 rewards over the preceding minute).

These unit and stimulus presentation ids were then used to index the session units $\times$ stimuli $\times$ 750 ms tensor described above in *Peri-stimulus/event histograms*. Activity was binned in 10 ms bins and aggregated in a list of tensors across sessions. The qualifying unit ids for each session were concatenated across sessions, and a pseudopopulation of 100 units was selected randomly from this list with replacement. For each unit in the pseudopopulation, the corresponding session tensor was loaded and that unit's activity was extracted for all trials in each of the relevant decoding categories (for example, change and non-change trials for change decoding). For each trial category, trial activity for that unit was randomly assigned to train and test splits (50/50 split). Units from the same session were constrained to have the same train and test splits. Pseudotrials were then constructed across all units in the pseudopopulation by randomly selecting, for each unit, 100 trials for each trial category with replacement. Training data was constructed from each unit's training split and testing data from each unit's test split.

Once train and test pseudotrials were constructed, individual decoders were trained on progressively longer intervals of activity from stimulus onset (for example, the first decoder saw only the first 10 ms bin after stimulus onset, but the fifth decoder saw 5 bins capturing the first 50 ms of activity). Decoders were linear support vector classification models (scikit-learn's LinearSVC with $C = 1.0$ and `class_weight = 'balanced'`). This entire process was repeated for 1000 randomly selected pseudopopulations, and the balanced accuracy for test data was averaged across runs to quantify decoding performance.

To test whether the time lag between LP and VISp change decoding might reflect different sampling of units from the two regions across sessions, we performed a session-level analysis ([Figures S8D and S8E](#)) as follows: 1) sessions for which at least 10 neurons were recorded from both LP and VISp were identified; 2) for each of these sessions, we performed change decoding for random samples of 10 neurons (with replacement) from each area. The number of samples was selected per area to ensure that each neuron had a greater than 99% probability of being selected at least once. 3) Decoding curves were averaged across samples, and decoding latencies (time to half maximum) were extracted, producing one latency per session per area. Sessions for which the mean decoding curve for either area failed to reach 60% accuracy (10% above chance) were excluded from further analysis (2 of 16 sessions). A Wilcoxon signed-rank test was then performed on these session-level decoding latencies.

Facemap lick decoding

To decode licking activity from video data ([Figure 3](#)), the side view video for each session was cropped to capture the face. The resulting cropped video was then processed with Facemap^{40,91} to extract the top 500 motion SVs. Lick and non-lick trials were selected and a linear SVC decoder was trained on progressively longer video intervals from stimulus onset as described above. Decoders were trained with five-fold cross validation and performance on test data was averaged across sessions.

Image decoding for familiar and novel sessions

To compare the time course of image decoding during familiar and novel sessions, populations of 40 units were subsampled for each session and region without replacement. The number of sampling iterations was selected such that there was a 99% chance of sampling each unit at least once (sessions with less than 40 units for a given region were excluded). For each subsample, unit activity was binned in 10 ms bins and linear SVC decoders were trained with five-fold cross-validation to predict image identity from progressively longer time windows as described above. Recall was calculated for each image separately (number of times an image was correctly predicted divided by the total number of times it was presented), and the median was taken across subsamples. Recall was averaged across images (excluding hold-over images) to produce decoding time courses for each session.

Change, state and novelty modulation indices

Change and state modulation indices for each unit were calculated on the mean activity during the decision window (20–100 ms after stimulus onset) after baseline subtraction (baseline was defined as the 50 ms preceding stimulus onset). These indices compare activity across two conditions and were computed as follows:

$$\text{Modulation Index} = \frac{\text{condition}_1 - \text{condition}_2}{\max(|\text{condition}_1|, |\text{condition}_2|)}$$

For change modulation, condition_1 represents the mean firing rate during the decision window during change stimuli, and condition_2 represents the mean firing rate during the decision window during pre-change stimuli. Both change and pre-change stimuli were taken from hit trials.

For state modulation, condition_1 represents the response to change stimuli during active behavior, and condition_2 represents the response to change stimuli during passive replay.

For each unit, the mean response was calculated for each image separately before averaging across images.

For novelty modulation, the index could not be calculated for individual units as different populations were recorded on the two recording days. Instead, units were filtered to meet the relevant criteria (e.g. region, cortical layer or cell-type) and separated according to whether they were recorded during the familiar session or the novel session. For each group, the mean response during the decision window was calculated for every unit as above and aggregated into a list. The familiar and novel responses were then re-sampled 1000 times with replacement. For each iteration, condition_1 was defined as the mean response across the sampled novel units, and condition_2 was defined as the mean response across the sampled familiar units. The final modulation index was then taken as the median value across iterations.

To examine the effect of running speed on state modulation (Figure S10), trials were subsampled to match the running speed distribution across active and passive epochs for each session before calculating unit-wise PSTHs and modulation indices as follows: 1) the mean running speed during a 200 ms window before each stimulus was binned in 5 cm/s bins separately for active and passive epochs; 2) for each bin, the minimum trial count was determined across the two epochs; 3) this minimum trial count was then subsampled without replacement from the trials in the corresponding bin for each epoch. Repeating this procedure for each bin produced identical running distributions for active and passive trials (Figure S10A).

Visual response and novelty modulation latencies

To determine the visual and novelty modulation latencies for a given region, each unit's mean response to non-change stimuli (excluding holdover images) and omitted stimuli were smoothed with a symmetrical exponential kernel ($\tau = 3$ ms) and baseline subtracted. Units recorded during familiar and novel sessions were separated and a bootstrapped estimation of latency was calculated as follows: for each iteration, familiar image responses, novel image responses, familiar-session omission responses and novel-session omission responses were resampled across units with replacement. D-prime values were then calculated at each time point as follows: for visual latency, corresponding image and omission response distributions were compared; for novelty modulation latency, familiar and novel image response distributions were compared. Each comparison thus yielded a 250 element vector of d-prime values for the 250 ms stimulus window. The time of the maximum d-prime for this vector was identified, and the latency was defined as the first point working backwards from the maximum at which d-prime was ≤ 0.1 , a threshold found empirically to identify the foot of the population visual response. If the maximum d-prime was less than 0.1, the latency was set to NaN for that iteration. For the final visual latency, familiar and novel visual latencies were averaged.

Change detection strategy analysis

To determine whether mouse behavior was better approximated by an 'adaptation' or 'image-comparison' strategy, the stimulus-wise responses of all units in a region of interest for a given session were averaged across the decision window (20-100 ms after stimulus onset), resulting in a response vector of length N_{units} for each stimulus presentation. The resulting response vectors were then used to train two decoding strategies as follows:

Adaptation decoder: A linear decoder (scikit-learn LinearSVC, with `class_weight = 'balanced'`) was trained to classify stimuli as change or non-change. 5-fold cross-validation was used to assess the balanced accuracy of the decoder on held-out data.

Image comparison decoders: For each image, a one-vs-all decoder was trained with 5-fold cross-validation (scikit-learn LinearSVC, with `class_weight = 'balanced'`) to differentiate stimulus responses to the selected 'template' image from responses to any of the other seven images (8 total decoders corresponding to 8 unique template images). For each stimulus presentation, the test-fold predictions of all 8 decoders were saved. The image-identity of the previous stimulus was then used to choose the appropriate decoder's prediction and thus predict whether the current stimulus was the same as or different from the previous stimulus.

Applying these decoding strategies produced two vectors of length $N_{\text{stim presentations}}$ for each session and region (depicted in Figure 7A). One vector stored the adaptation decoder confidence (scikit-learn's decision function) that any given stimulus was a change stimulus. The other stored the image comparison decoder's confidence that any given stimulus was the same as the previous stimulus.

These decoder confidence vectors were then used to predict two binary labels for each stimulus using logistic regression: *is_change* (whether a particular stimulus was a change stimulus, Figure 7B), and *is_lick* (whether the mouse licked for a particular stimulus, Figure S10D). For lick predictions, only non-change stimuli were considered, as model performance on these error trials was not confounded by there also being a stimulus change.

Statistical Comparisons

Comparisons for change and state modulation indices across regions, cortical layers, cell-types or clusters were made using the Wilcoxon rank-sum statistic with each unit considered an independent sample (`scipy.stats.ranksum`). For session-level analysis, comparisons were made using a permutation test as follows: for each session, the relevant metric for units in the group of interest (i.e. region, layer or cell-type) was averaged, producing one value per session per group. Sessions with less than five units in a given group were not used for that group. To compare two groups, values for sessions in which both groups were recorded were swapped randomly. For sessions in which only one group was recorded, group labels were randomly permuted across sessions. For each permutation iteration (10000 total), the mean difference between groups was taken to build a null-distribution. Finally, a two-tailed empirical p-value was calculated comparing the observed group difference to this null distribution.

Comparisons for novelty modulation were made with a permutation test. Units from a given group were aggregated across familiar and novel sessions separately. When comparing groups, group labels were permuted within familiar and novel conditions to preserve experience labels. For each permutation iteration (10000 total), a new novelty modulation index was calculated on these permuted groups to build a null distribution. Finally a two-tailed empirical p-value was calculated comparing the observed modulation to the null distribution. For mouse-level analysis, only mice for which 5 units in a given group were recorded for both the familiar and novel sessions were considered. A novelty modulation index was calculated for each mouse, and the resulting labels were permuted to build a null distribution from which an empirical p-value could be calculated as described for the session-level analysis above.

Comparisons of latencies for decoders trained on pseudopopulations (i.e. LP vs VISp change decoding; SC/MRN vs VIS lick decoding) were made with a permutation test. Units from both areas were pooled across sessions, and unit area labels were permuted. Decoding was then run on these permuted populations as described in *Image, change and lick decoding*. This process was repeated for 1000 permutations. To create a null distribution, a decoding latency difference between the two areas was calculated for each permutation. Finally, an empirical p-value comparing the observed decoding latency difference to this null distribution was determined.

For all analyses involving multiple comparisons, the Benjamini–Hochberg false discovery rate correction was applied (`statsmodels.stats.multitest.multipletests` with method ‘fdr_bh’).

Supplemental information

Map of spiking activity underlying change detection in the mouse visual system

Corbett Bennett, Samuel D. Gale, Gregory Heller, Tamina K. Ramirez, Hannah Belski, Alex Piet, Omid Zobeiri, Adam Amster, Anton Arkhipov, Alex Cahoon, Shiella Caldejon, Mikayla Carlson, Linzy Casal, Scott F. Daniel, Colin Farrell, Marina Garrett, Ryan Gillis, Conor Grasso, Ben J. Hardcastle, Ross Hytnen, Tye Johnson, Peter Ledochowitsch, Quinn L'Heureux, Dana Mastrovito, Ethan G. McBride, Stefan Mihalas, Chris Mochizuki, Christopher B. Morrison, Chelsea Nayan, Nhan-Kiet Ngo, Kat North, Douglas R. Ollerenshaw, Ben Ouellette, Paul Rhoads, Kara Ronellenfitch, Martin Schroedter, Joshua H. Siegle, Cliff Slaughterbeck, David Sullivan, Jackie Swapp, Michael Taormina, Wayne Wakeman, Xana Waughman, Allison Williford, John W. Phillips, Peter A. Groblewski, Séverine Durand, Christof Koch, and Shawn R. Olsen

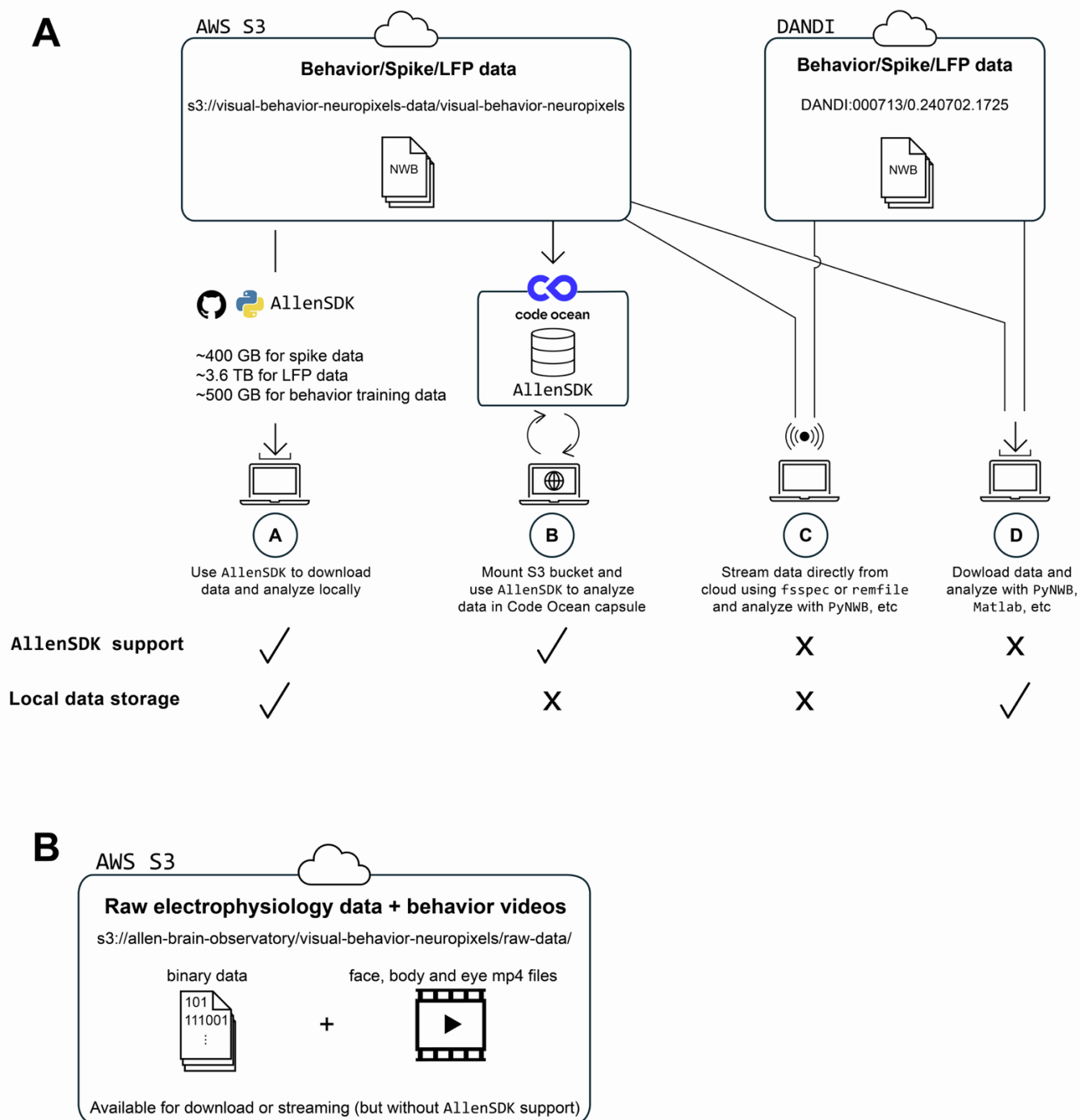


Figure S1: Accessing the Allen Institute’s Visual Behavior Neuropixels dataset. Related to STAR Methods. **A)** Diagram depicting four different ways users can access this dataset. NWB files containing behavior, spike and LFP data are all stored in a public AWS S3 bucket as well as in the DANDI archive. Users can access this dataset in a number of ways, including: (A) Using the AllenSDK to download NWB files to local storage; (B) mounting the S3 bucket in Code Ocean to analyze on cloud compute resources, (C) using tools like fsspec or remfile to stream data directly from the cloud to be analyzed with PyNWB (or other packages capable of reading HDF5 files); or (D) downloading data to local storage for analysis without the AllenSDK. **B)** In addition to the NWB files above, raw electrophysiology data as well as behavior videos of the eye, face and body are available in a separate S3 bucket. The AllenSDK does not manage download or analysis of these data.

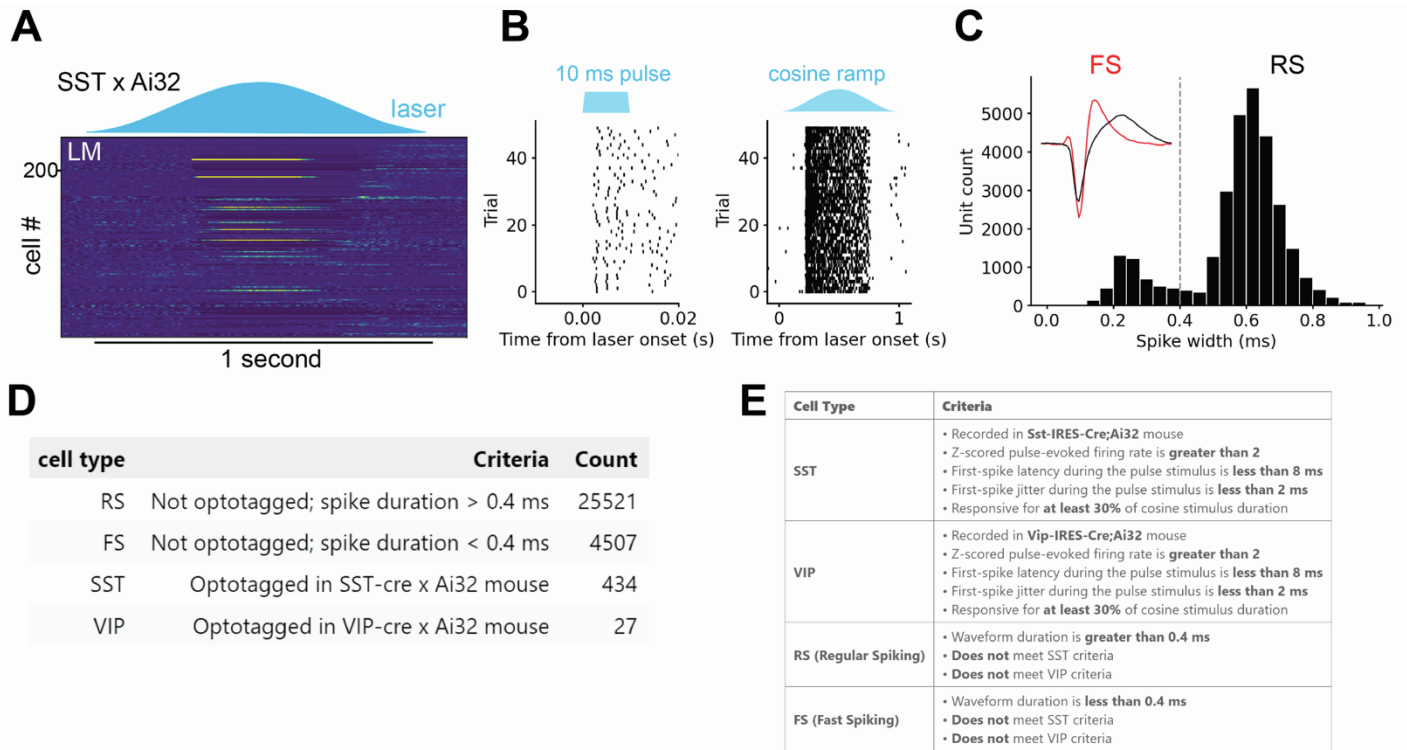


Figure S2: Identifying cortical cell types. Related to Figure 1. A) Heatmap showing mean response to cosine-ramp laser stimulus in SST-Cre x Ai32 mouse. Units (rows) are ordered by depth. Note several units with sustained response to laser (yellow). **B)** Raster for one example optotagged SST unit aligned to short pulse (left) and cosine ramp (right) laser stimuli. **C)** Histogram of peak-to-trough waveform duration for cortical units. Units with a duration of less than 0.4 ms were identified as FS. **D)** Table describing criteria and unit counts for each cortical cell type. **E)** Criteria for cell-type classification. Units were classified as optotagged if 1) their mean laser evoked firing rate was greater than two standard deviations above baseline for the short pulse at highest power, 2) the latency of their response to the pulse stimulus was less than 8 ms, 3) the jitter in their first spike during the pulse stimulus was less than 2 ms and 4) their firing rate was significantly elevated above baseline for at least 30% of the cosine ramp stimulus.

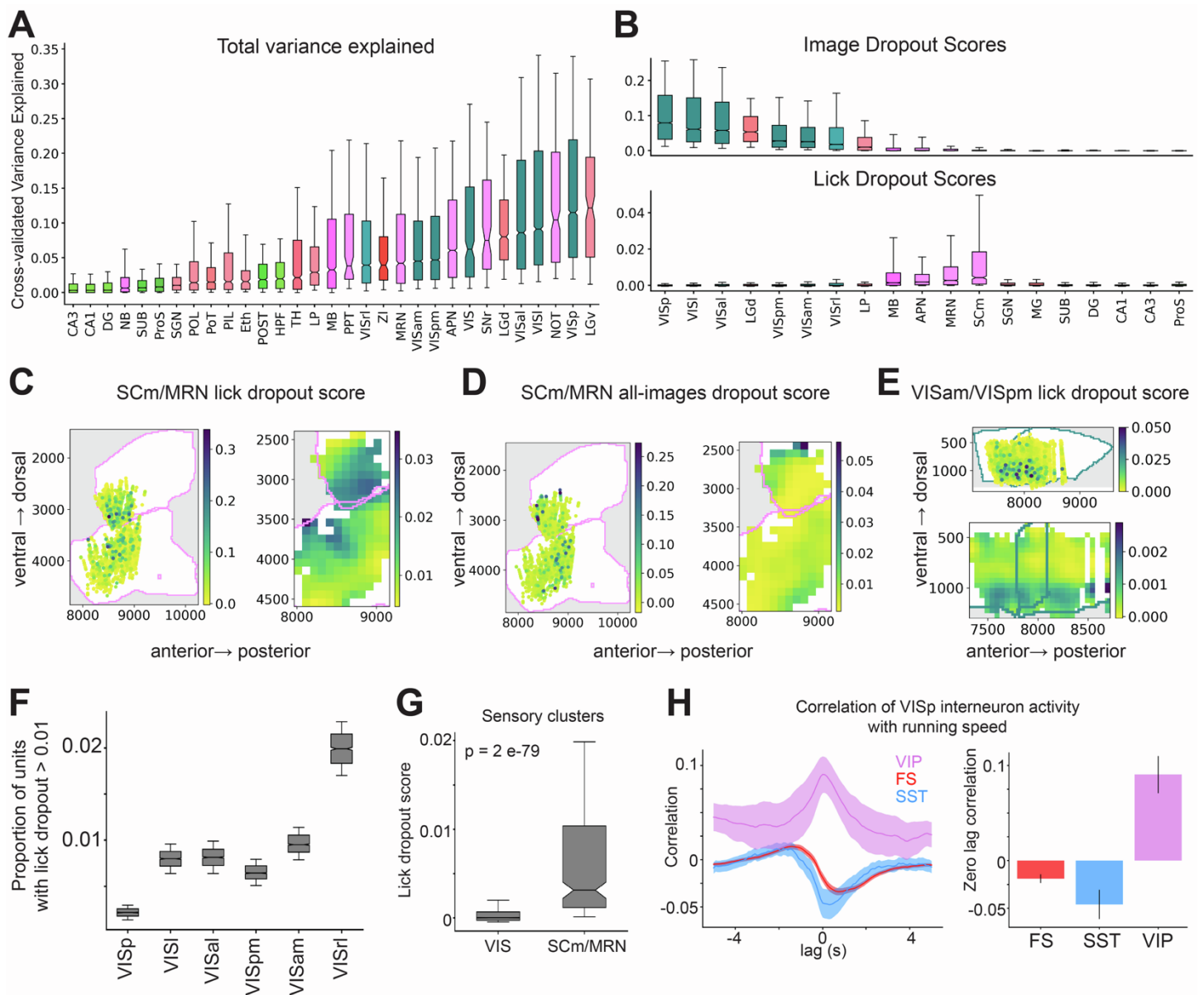


Figure S3: Further analysis of GLM results. Related to Figure 2. **A)** Cross-validated variance explained by GLM fit for units in all recorded areas. **B)** Visual (top) and lick (bottom) dropout scores across recorded areas. Box plots are colored according to area-color mapping used by Allen Institute Reference Atlas ontology (Wang et al. 2020). **C)** Distribution of lick dropout scores in SCm/MRN. Left: Lick dropout scores for each unit in SCm/MRN collapsed to horizontal plane. Right: same data as left, but spatially binned and smoothed. Pink contours outline SCm (dorsal) and MRN (ventral). **D)** Same as (C) but for visual dropout scores. Visual dropout scores were calculated by dropping all image kernels. **E)** Same as (C) but for lick dropout scores in VISam/VISpm. Teal contours outline VISam (anterior) and VISpm (posterior). Darker colors indicate larger dropout scores. **F)** Proportion units with high lick dropout scores across six visual cortical areas. **G)** Lick dropout scores for visual cortical units and SCm/MRN units assigned to sensory clusters. P-value calculated using Wilcoxon rank-sum test. Box plot notches indicate median; box edges indicate 25th and 75th percentiles; whiskers span 10-90th percentiles. **H)** Left: Mean cross-correlogram between running speed and unit activity for three cortical interneuron cell-types in VISp. Shading indicates SEM over units. Right: Mean correlation at zero lag for each cell-type. Error bar indicates SEM over units.

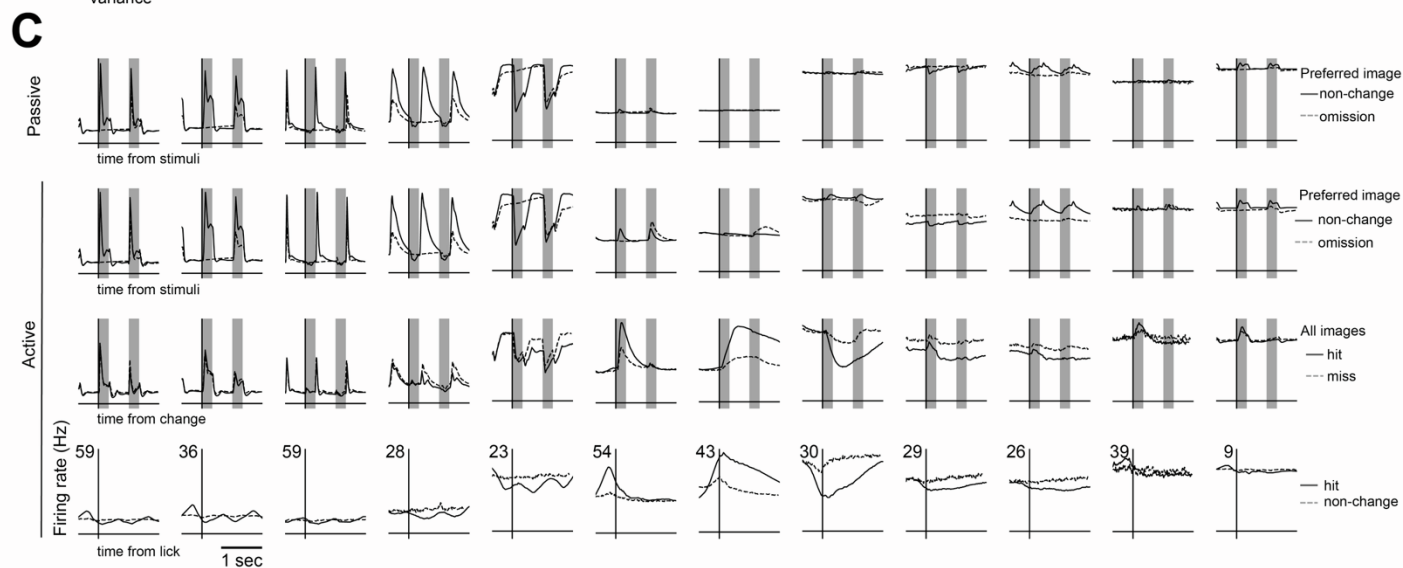
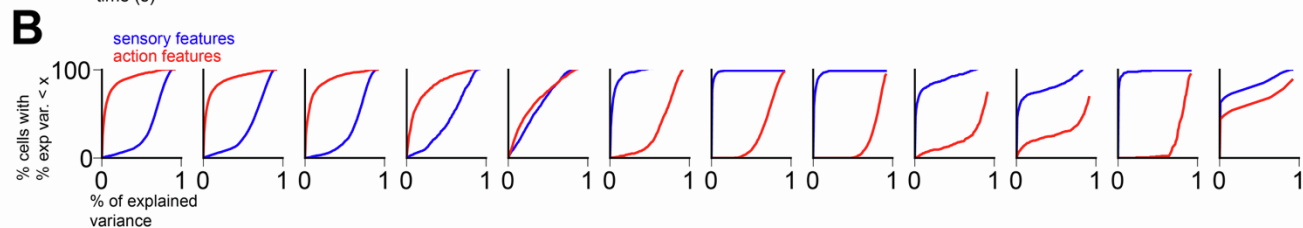
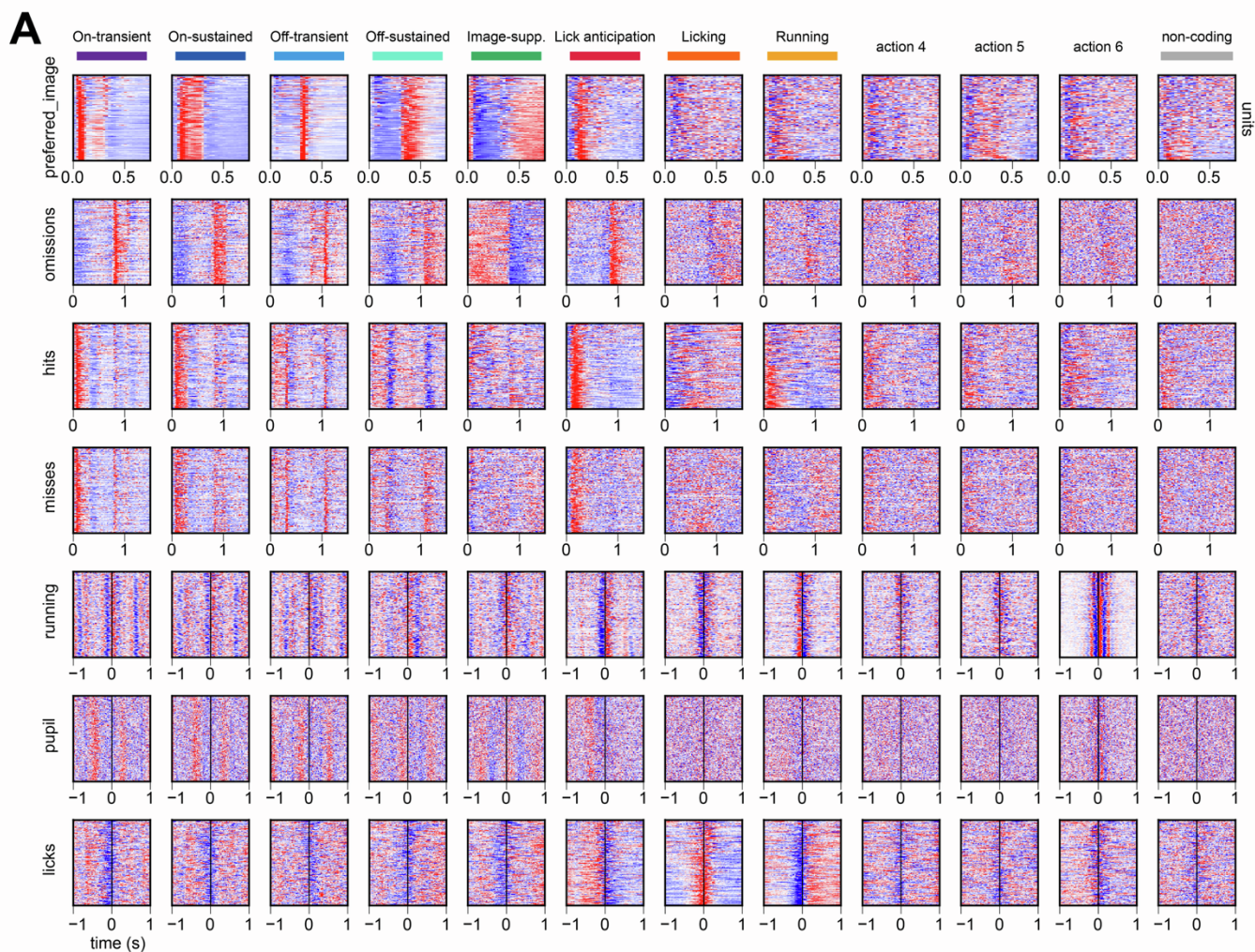


Figure S4: Kernel fits and population responses for each cluster. Related to Figure 2. **A)** Temporal kernels for all units in all clusters. **B)** Cumulative histograms showing what percentage of the explained variance is attributable to sensory (blue) or action (red) features. **C)** Population PSTHs for each cluster aligned to different task events in either the active or passive experimental epoch.

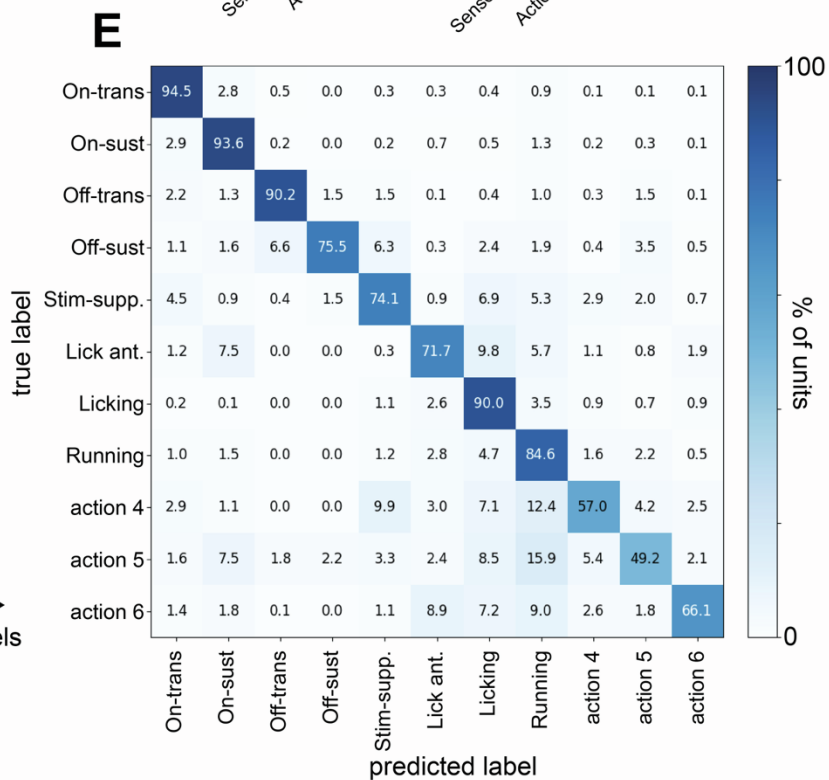
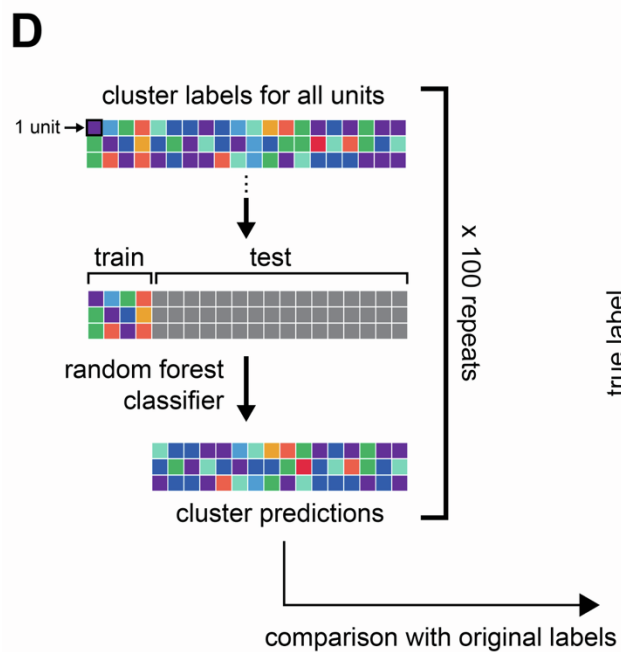
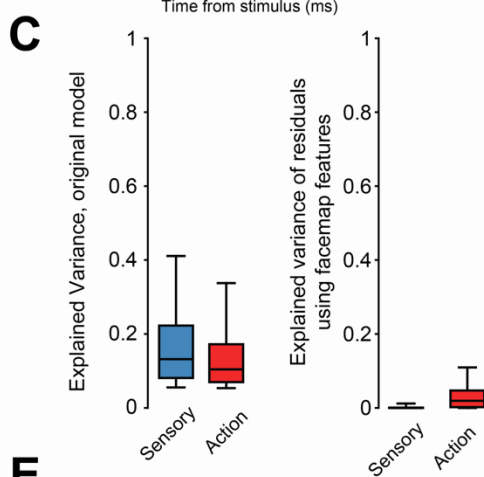
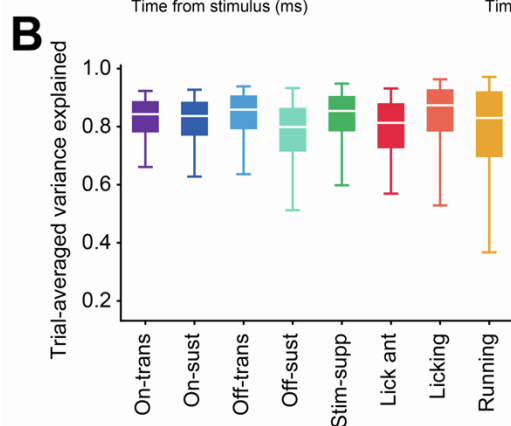
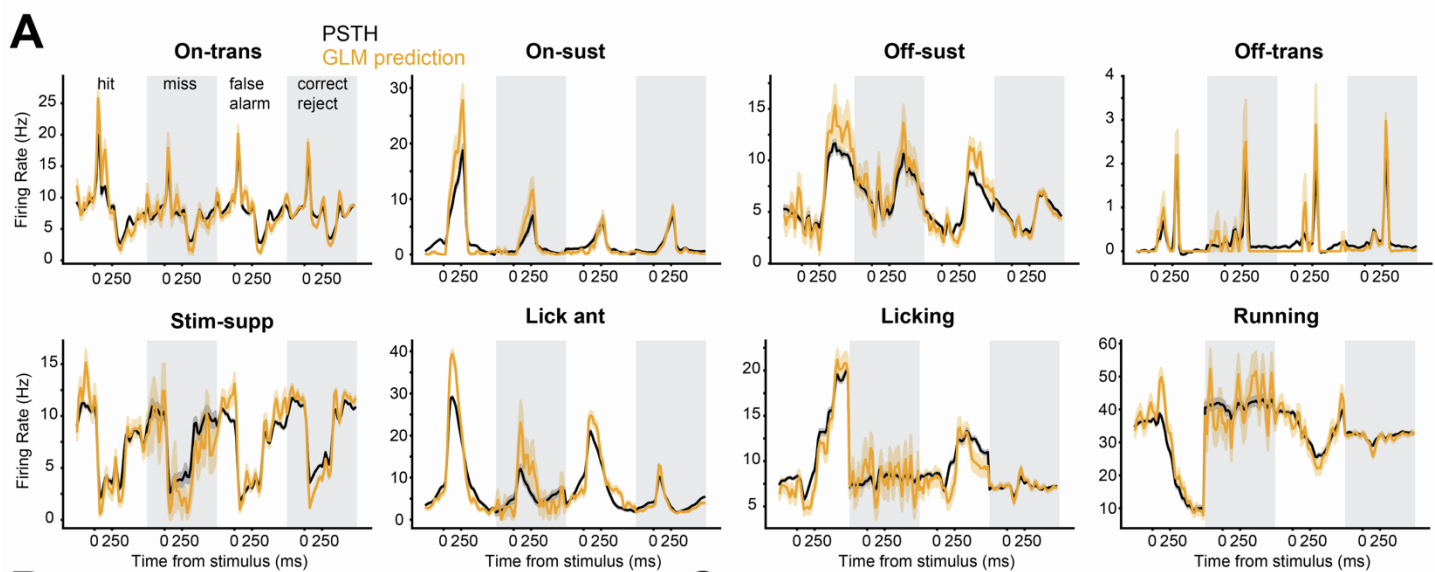


Figure S5: GLM and clustering validation. Related to Figure 2. **A)** PSTH (black) and mean GLM prediction (gold) across four trial types (hit, miss, false alarm, and correct reject) for example units in each cluster. **B)** Boxplot showing variance explained by GLM for trial-averaged responses across all units in each cluster. White line indicates median. Box spans 25-75 percentiles. Whiskers span 5-95 percentiles. **C)** Left: cross-validated total variance explained by the GLM for units in the sensory and action clusters. Right: cross-validated variance explained of the residuals from the GLM fit, now using 500 Facemap SVDs as regressors (see Methods). Facemap features explained some residual variance for action but not sensory units. Boxes span 25th and 75th percentiles. Whiskers span 5th and 95th percentiles. Medians indicated by black lines. **D)** Diagram depicting bootstrapped classifier method. For each iteration (100 total), units were randomly divided into 20/80 train/test splits and a random forest classifier was trained on the cluster labels from the training data to predict the cluster labels of the test set. The input to the classifier was the same feature matrix used in clustering after UMAP dimensionality reduction to three feature dimensions per unit. Predictions were pooled across bootstrap iterations to construct the confusion matrix. **E)** Confusion matrix indicating the frequency with which units in each cluster were assigned to each cluster by the classifier.

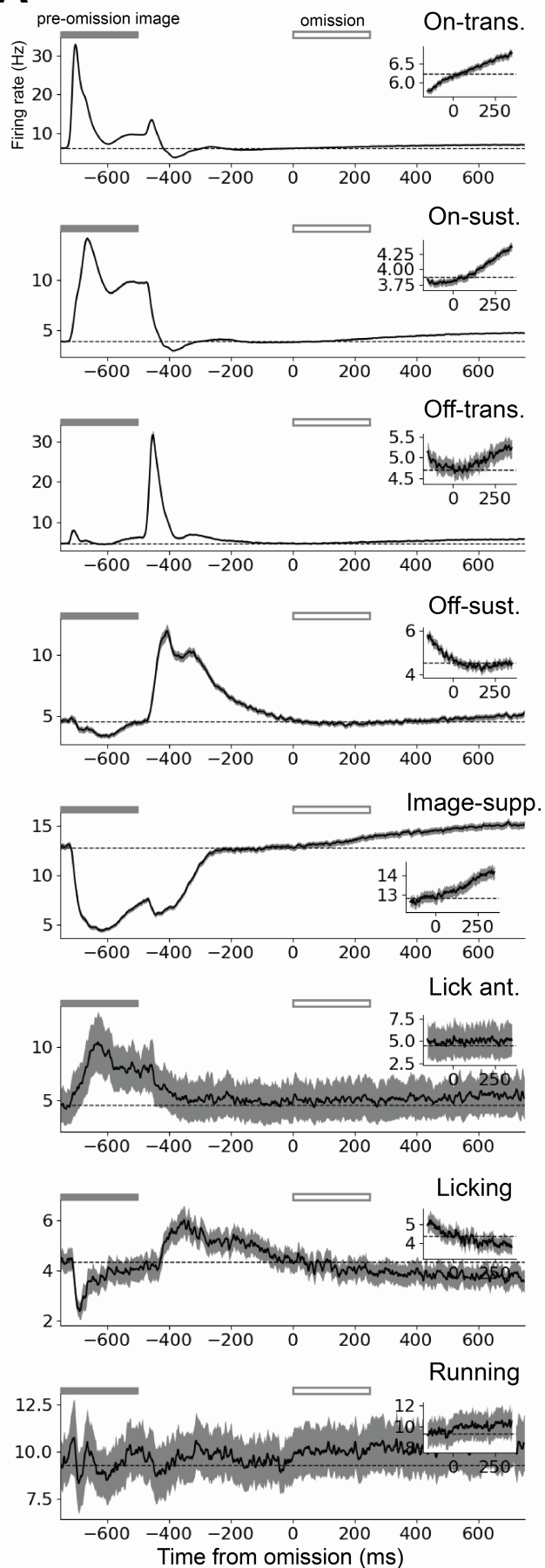
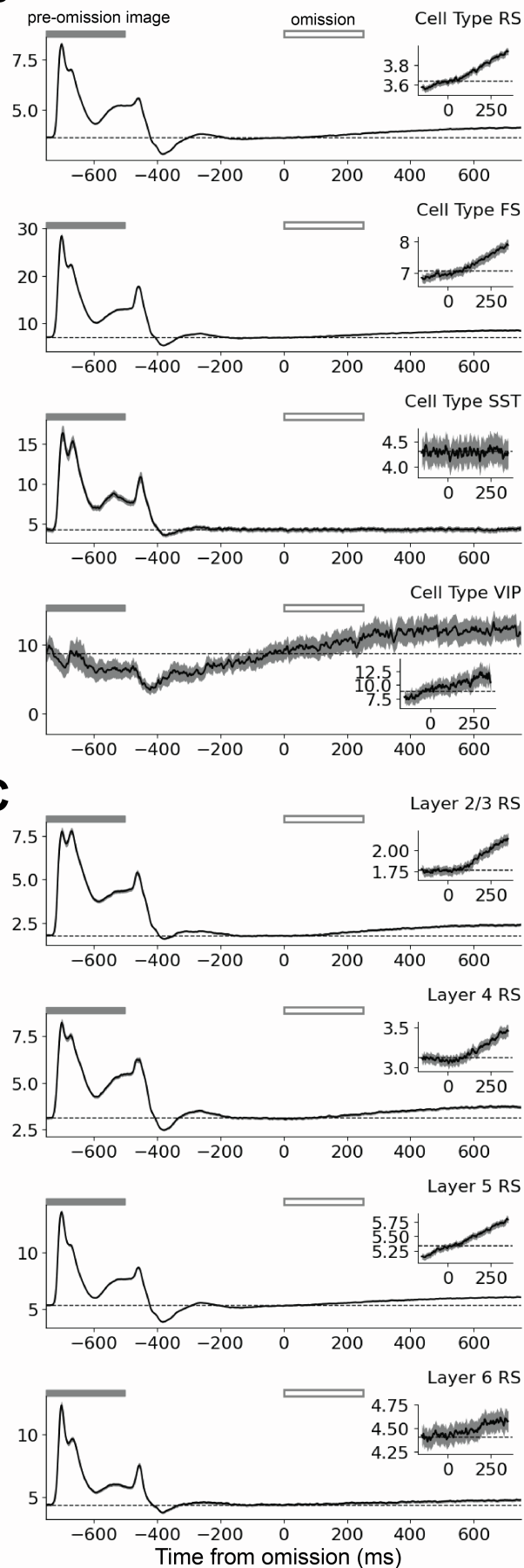
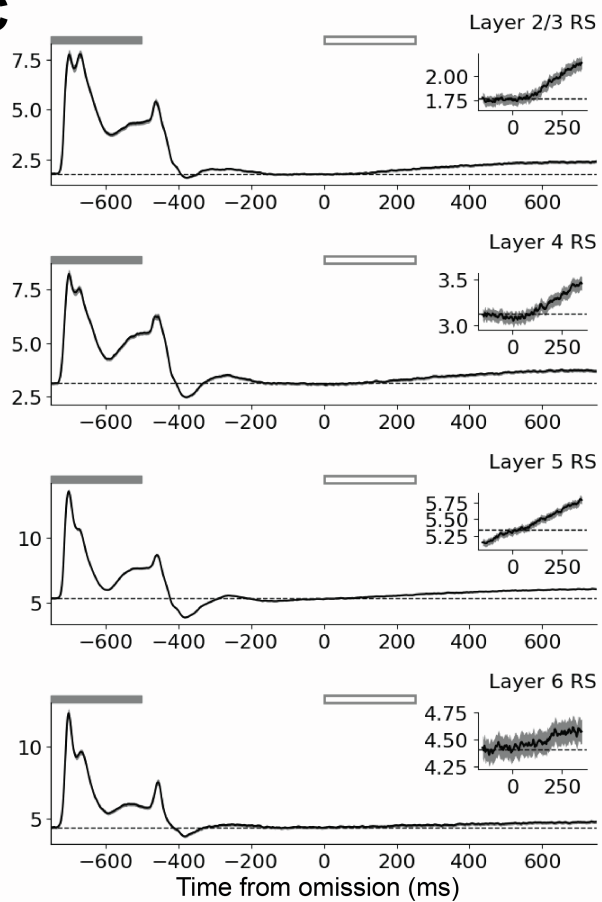
A**B****C**

Figure S6: Response to stimulus omissions in visual cortex. Related to Figure 2. A) Omission responses for visual cortical units in each of the eight GLM clusters. Filled grey rectangle at top of plot indicates the pre-omission image presentation. Open rectangle indicates when the omitted stimulus would have appeared. Plot inset expands on time of omission. **B)** Same as (A) but partitioning visual cortical units by cell type. **C)** Same as (A) but for RS units partitioned by cortical layer. Shaded regions give SEM.

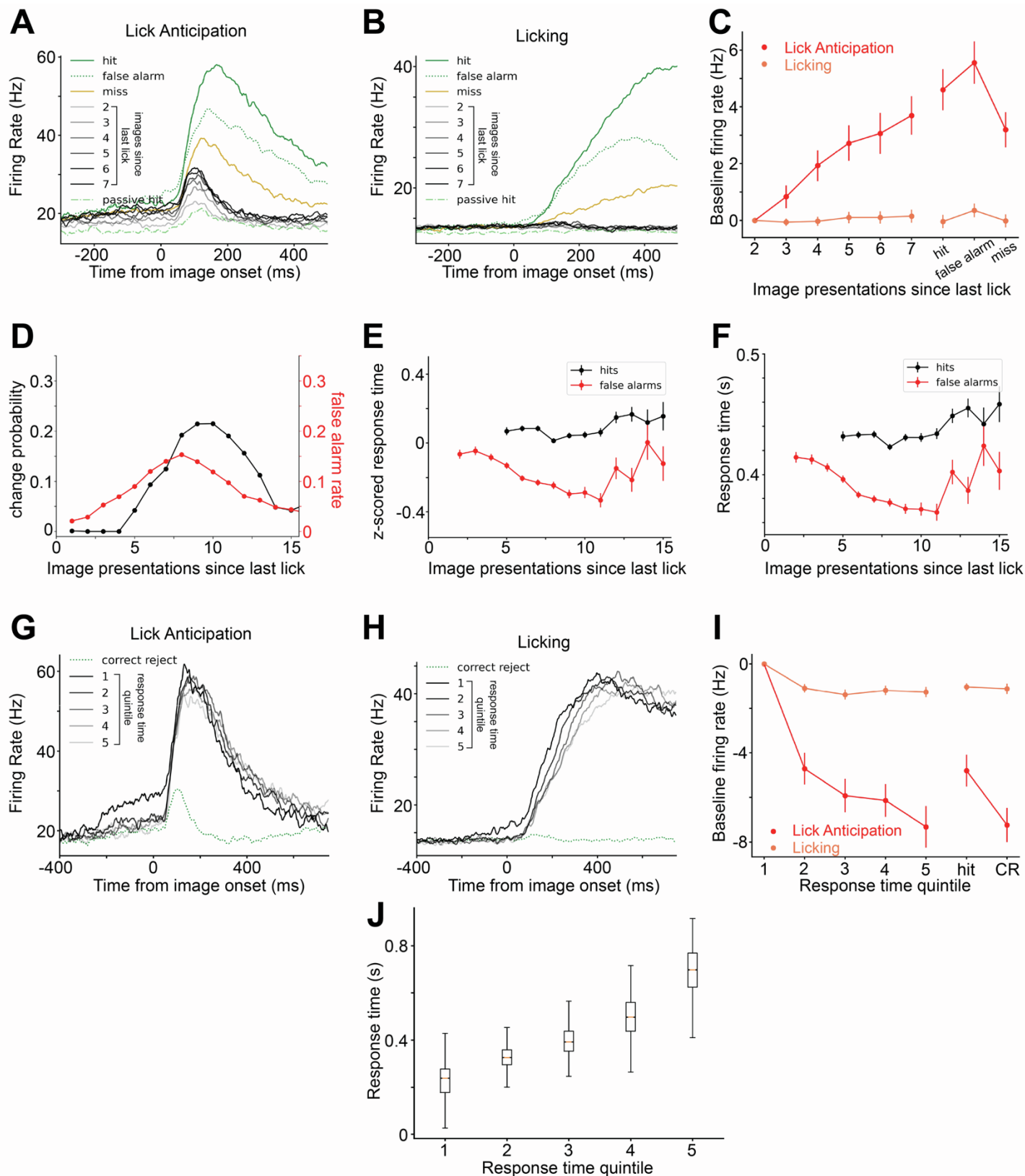


Figure S7: Baseline modulation of SCm/MRN lick anticipation units. Related to Figure 2. **A)** Population response of SCm/MRN units assigned to lick anticipation cluster broken out by the number of stimulus presentations since the last lick response (2-7 in legend) as well as hit, miss and false alarm trials. ‘Passive hit’ (green, dot-dash) responses refer to change stimuli shown during the passive viewing epoch which correspond to the hit stimuli (green, solid) during the active epoch. **B)** same as (A) but for SCm/MRN units assigned to the licking cluster. **C)** Mean of baseline activity (200 ms before stimulus onset) against presentations since last lick. **D)** Probability of image change (black) and false alarm rate (red)

against presentations since last lick. **E)** Response time as function of presentations since last lick for hit and false alarm trials. **F)** Same as (E) but the response times for each session have been z-scored before averaging across sessions. Error bars are SEM. **G,H,I)** Same as (A,B,C), but now activity is broken out by response time quintile and correct reject trials. Response time quintiles were calculated for each session separately before combining across sessions. Quintile 1 corresponds to trials with fastest response times. **J)** Response times by quintile. Boxes span 25-75th percentile. Orange line indicates median. All active conditions in (A-C) and (G-I) have been matched for mean running speed during the 200 ms prior to stimulus onset.

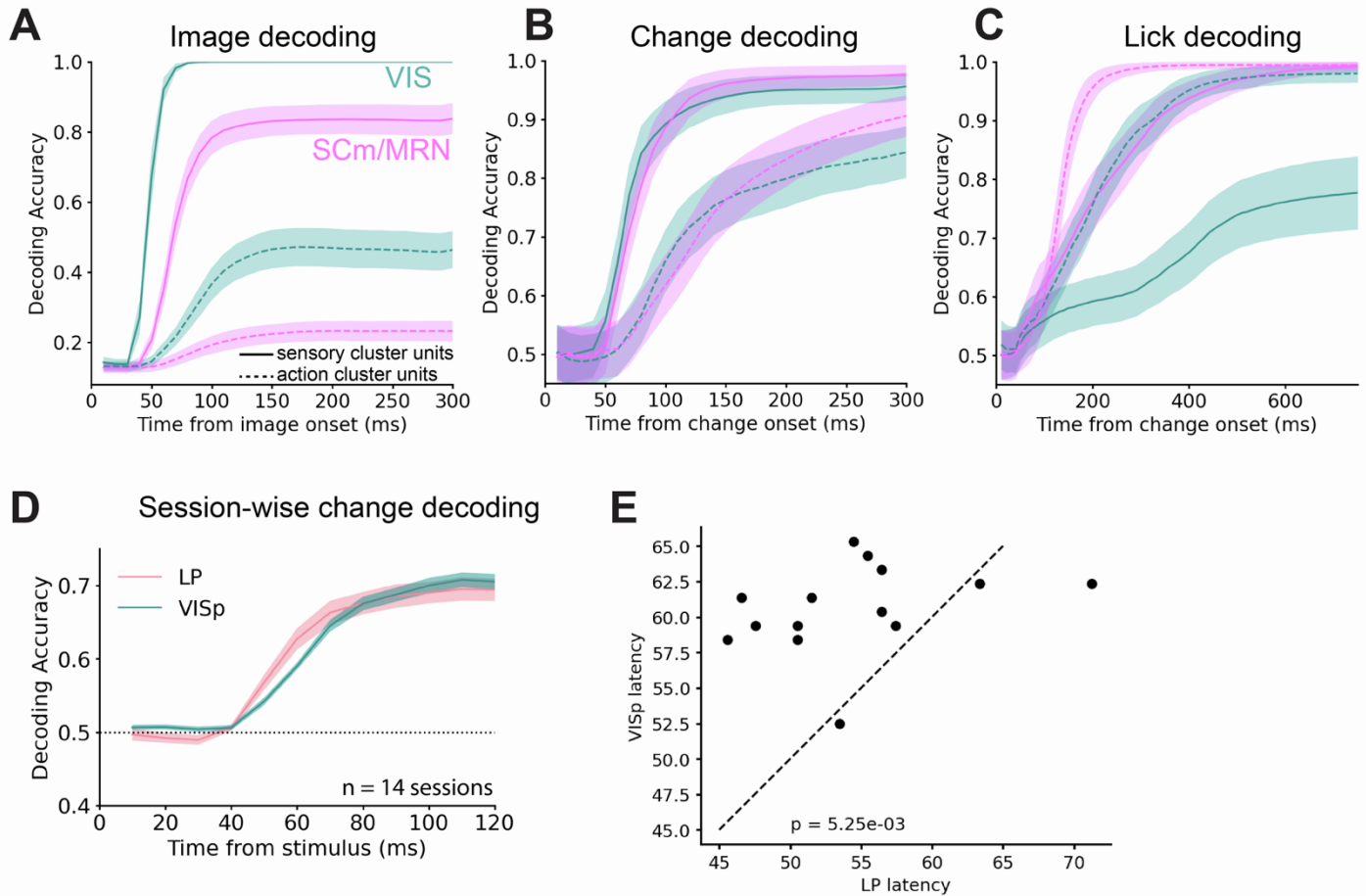


Figure S8: Comparing decoding time courses across clusters (sensory vs action) and areas (VISp vs LP). Related to Figure 3. A) Image decoding time courses for VIS (teal) and SCm/MRN (magenta) units from sensory (solid lines) or action (dashed lines) clusters. **B, C)** Same as (C) but for change and lick decoding. Note that action cluster units were too sparse in individual thalamic and cortical visual areas to perform this analysis. Shaded regions indicate standard deviation over 1000 pseudopopulations. **D)** Session-wise change decoding from sensory units in VISp and LP. Shaded regions give SEM over sessions. For each session, image change was decoded for samples of 10 sensory units from each area (see Methods). **E)** Scatter plot of LP and VISp decoding latencies from 14 sessions. P-value calculated with Wilcoxon signed-rank test.

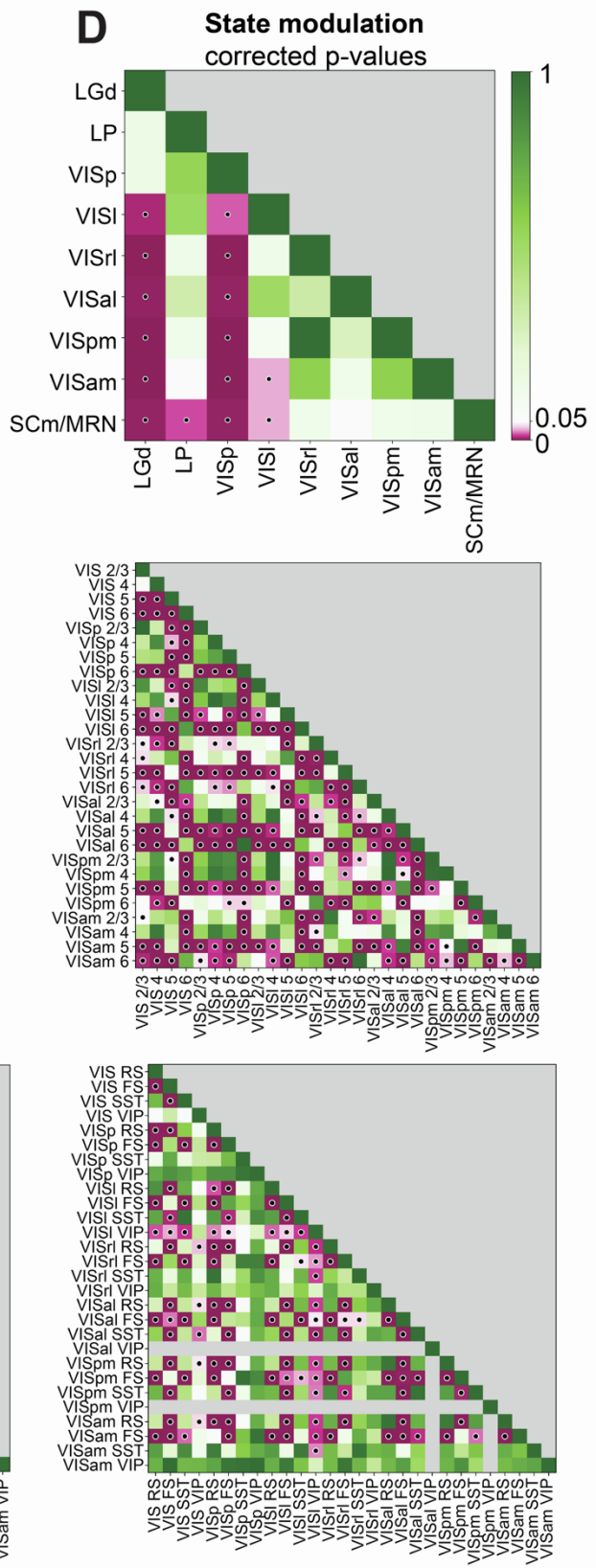
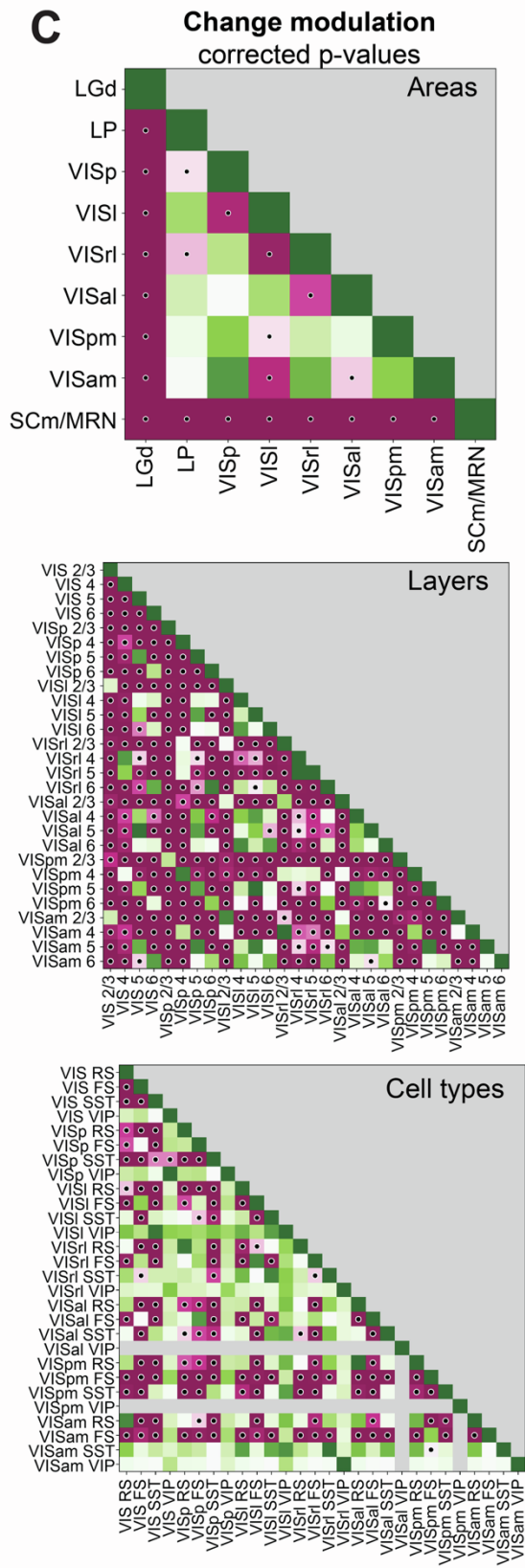
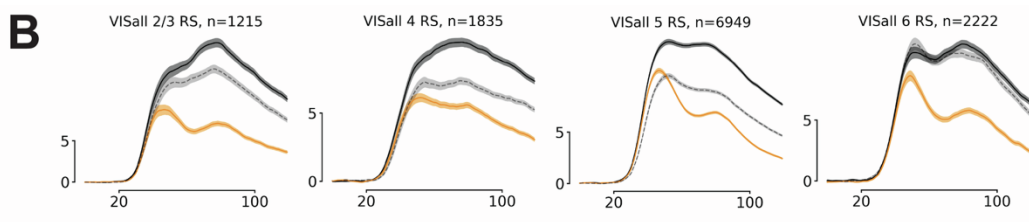
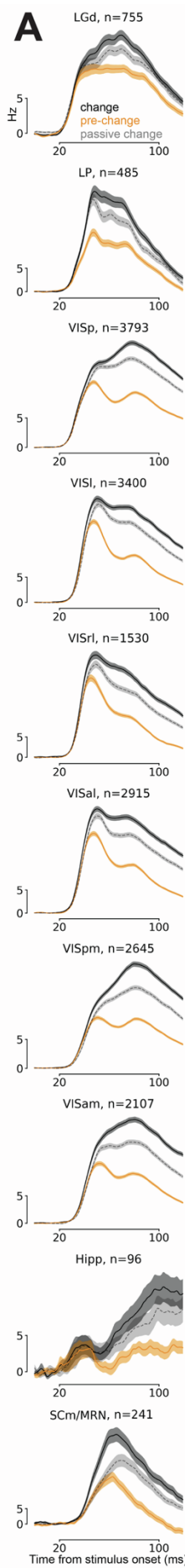


Figure S9: Change and state modulation across areas. Related to Figure 4. A) Population responses to change (black) and non-change (gold) stimuli during active behavior. Gray shows change stimuli during passive viewing. Time scale bar also indicates window over which change and state modulation were calculated (20-100 ms from stimulus onset). **B)** Same as (A), but plotted for RS cells in each cortical layer. **C)** P-value matrices for change modulation values across areas (top), cortical layers (middle) and cortical cell-types (bottom). Symbols (black circles with white border) indicate significant comparisons. All p-values were calculated with Wilcoxon rank-sum test and corrected for multiple comparisons by the Benjamini-Hochberg method. **D)** Same as (C) but for state modulation.

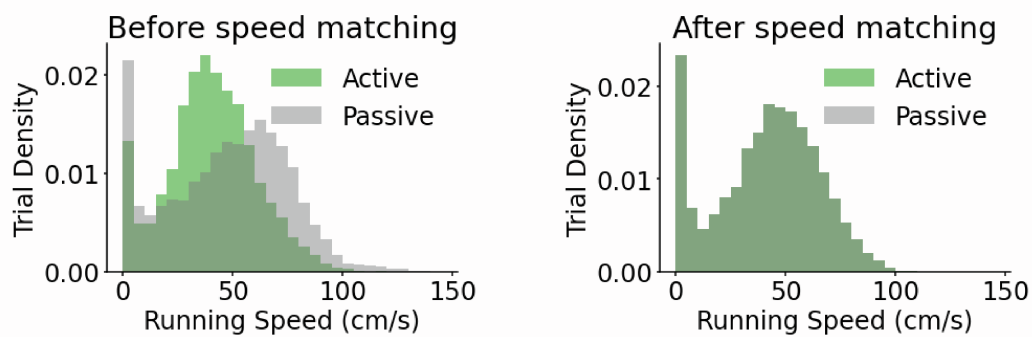
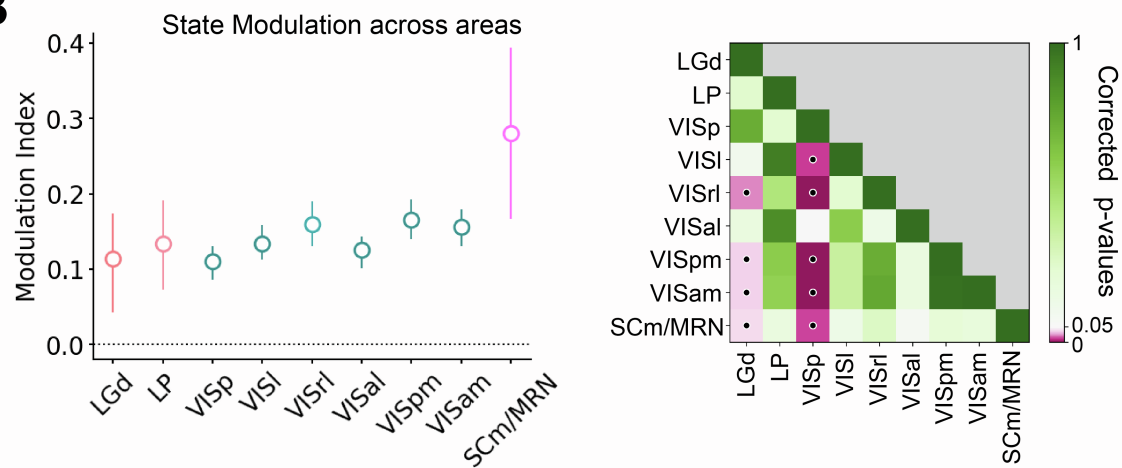
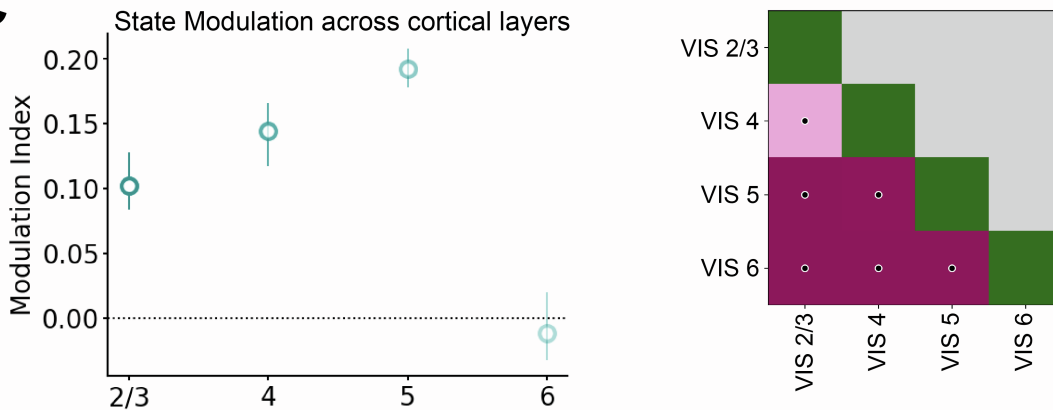
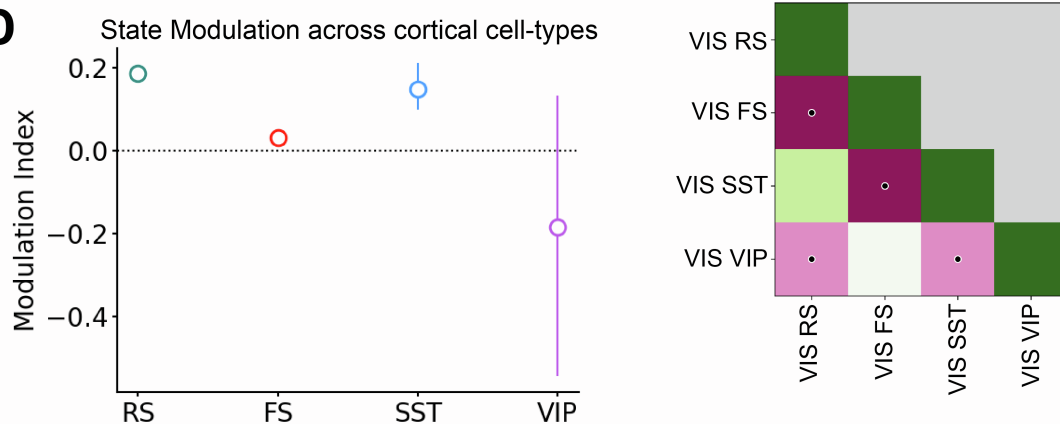
A**B****C****D**

Figure S10: State modulation does not reflect differences in running behavior across active and passive blocks. Related to Figure 4. A) Histogram of running speed during active (green) and passive (gray) epochs before (left) and after (right) speed matching across trials. For each trial, speed was averaged over the 200 ms before change onset. Note that after speed matching, the histograms are perfectly overlapping (dark green color). **B)** Left: State modulation across areas using only speed matched trials. Symbols indicate median and error bars indicate bootstrapped 95% CI of the median. Right: p-value comparison matrix for data on right. **C,D)** Same as (B) but for cortical layers (C) and cell types (D). All p-values were calculated with Wilcoxon rank-sum test and corrected for multiple comparisons by the Benjamini-Hochberg method.

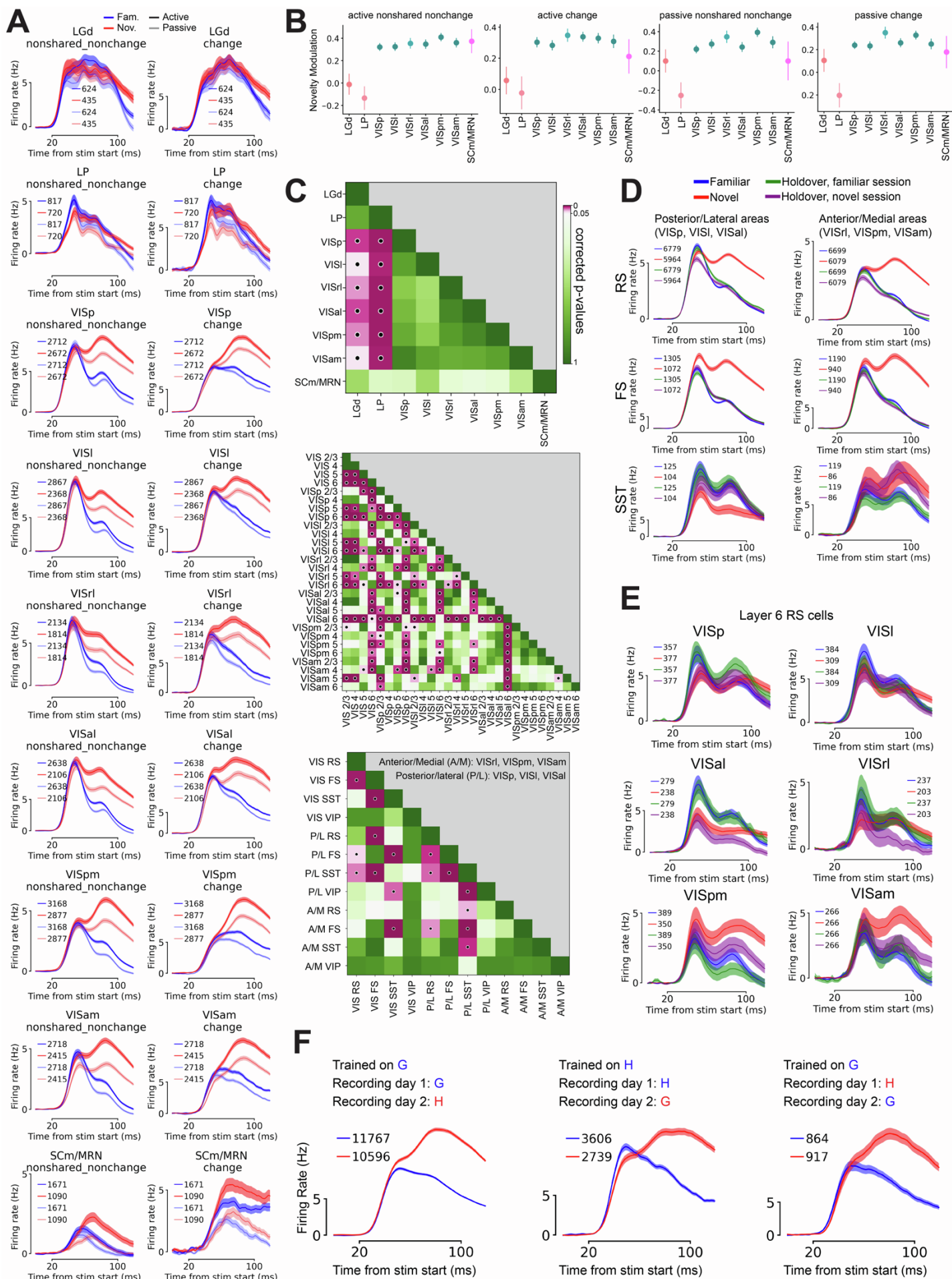


Figure S11: Further characterization of novelty modulation. Related to Figure 5. A) Population responses to non-change stimuli during familiar (blue) and novel (red) sessions for active behavior (dark) and passive viewing (light). For cortical areas, RS population is shown. **B)** Novelty modulation index calculated for change and non-change stimuli during active behavior and passive viewing. Symbols indicate mean $\pm$ standard deviation over 1000 bootstrap samples. **C)** P-values for novelty modulation indices across areas (top), cortical areas and layers (middle) and cortical areas and cell types (bottom). In bottom plot, P/L indicates posterior-lateral areas VISp, VISl, and VISal. A/M indicates anterior-medial areas VISrl, VISpm, and VISam. All p-values were calculated with a permutation test and corrected for multiple comparisons by the Benjamini-Hochberg method. **D)** Population responses for RS (top), FS (middle) and SST (bottom) units to familiar images (blue), novel images (red), holdover images during the familiar session (green) and holdover images during the novel session (purple). Units are either pooled across VISp, VISl and VISal (top row) or VISrl, VISpm and VISam (bottom row). **E)** Population responses of layer 6 RS units to same stimuli as (D) plotted for each visual cortical region. **F)** Population responses to change stimuli pooled across visual areas for three cohorts experiencing different training and recording trajectories. Note that for all three training/recording trajectories the response to novel stimuli is amplified and delayed relative to familiar stimuli.

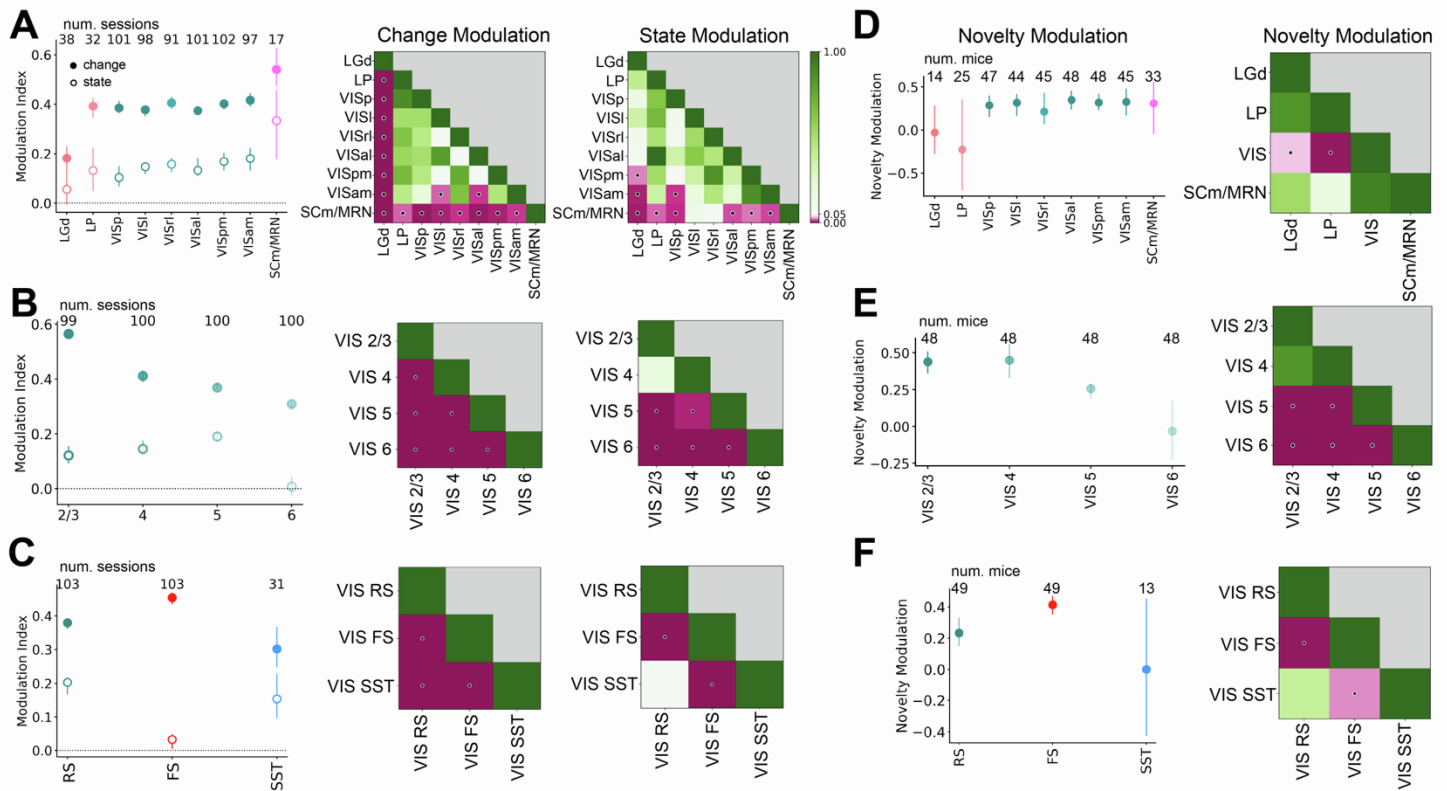


Figure S12: Change, state and novelty modulation over sessions and mice. Related to Figure 4. **A)** Left: Change (closed symbols) and state (open symbols) modulation calculated for each area on a session-wise basis. Symbols give median across sessions. Error bars indicate bootstrapped 95% CI of median across sessions. Right: P-value matrices for session-wise comparisons. P-values were calculated by a permutation test and corrected for multiple comparisons by the Benjamini-Hochberg method. **(B,C)** Same as (A) but for layers and cell-types respectively. **D)** Left: Novelty modulation for each area calculated on a mouse-wise basis. Symbols give median across mice. Error bars indicate bootstrapped 95% CI of median across mice. Right: P-value matrix for mouse-wise comparisons. **(E,F)** same as (D) but for layers and cell-types respectively.

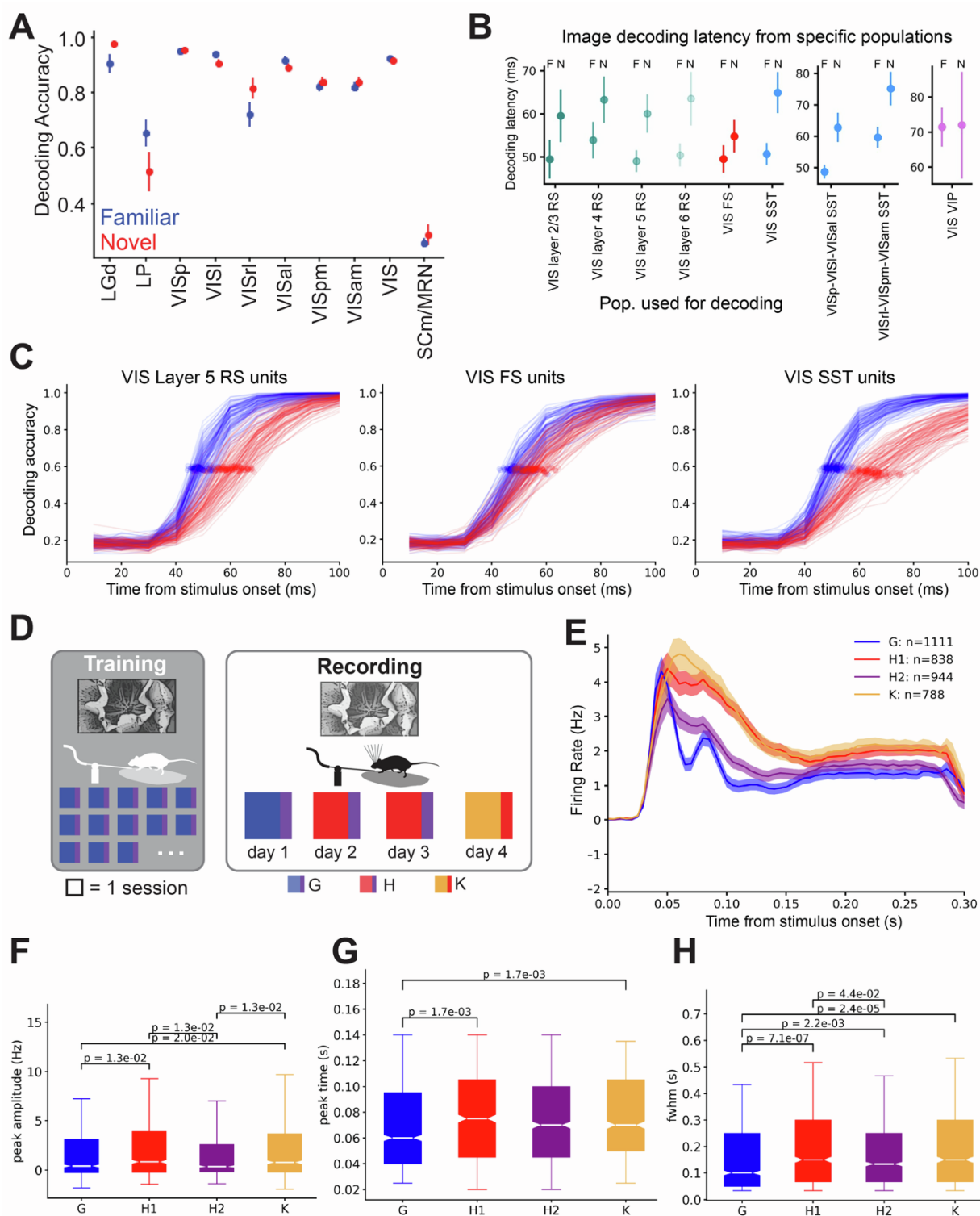


Figure S13: Cell-type specific decoding latencies and time course of image familiarization across sessions. Related to Figure 6. **A)** Image decoding accuracy for familiar (blue) and novel (red) images. Error bars indicate SEM across sessions. **B)** Image decoding latency using only units from particular layers and/or cell-types as indicated on x-axis. For each unit group, decoding latency for familiar images (first point) and novel images (second point) are shown separately. Symbols indicate mean $\pm$ standard deviation across 40-unit pseudopopulations. For VIP cells, 7-unit pseudopopulations were used. **C)** Example decoding time courses for familiar (blue) and novel (red) images for three cortical cell-types. Each curve represents one 40-unit pseudopopulation. Symbols indicate latencies (time to half-max) for each curve. **D)** Experimental design for four-day experiment: mice were trained on image set G and then recorded for four consecutive days with image sets G, H, H and K respectively. Image set K consisted of six novel images and two holdovers from image

set H. **E)** Population responses from visual cortical units to six non-holdover images from G (blue), H on recording day 2 (novel, red), H on recording day 3 (repeat, purple) and K (novel, orange). Shaded regions indicate SEM across units. **F)** Peak response amplitude for visual cortical units to images in (E). Notch indicates median. Box indicates 25th and 75th percentiles. Whiskers indicate 10th and 90th percentiles. **G)** Peak response time. **H)** Full width at half max. P-values calculated with Wilcoxon rank-sum test with correction for multiple comparisons.

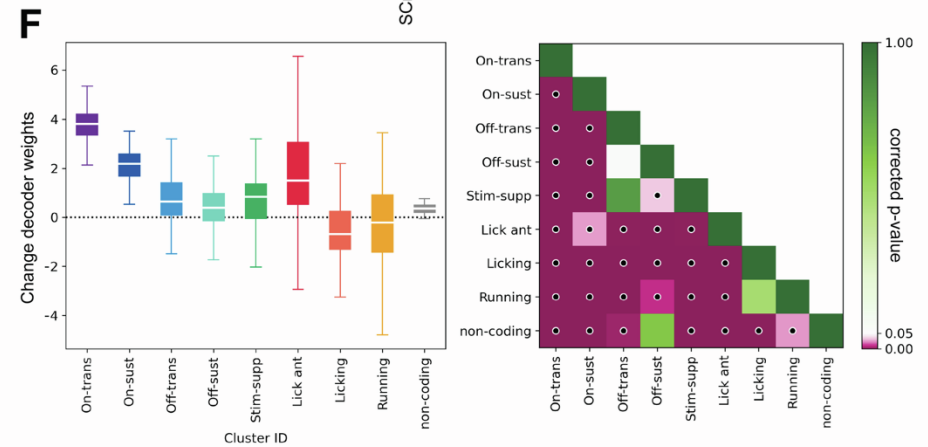
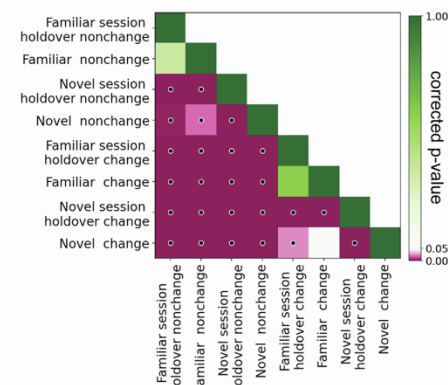
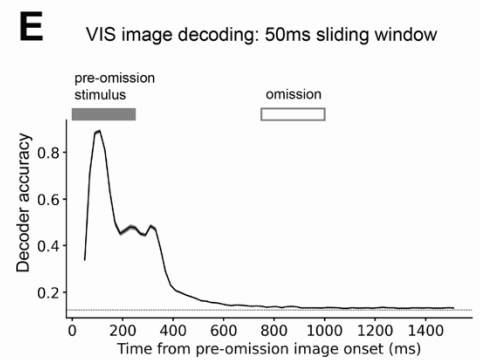
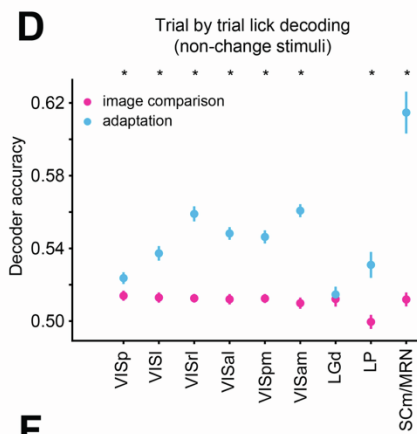
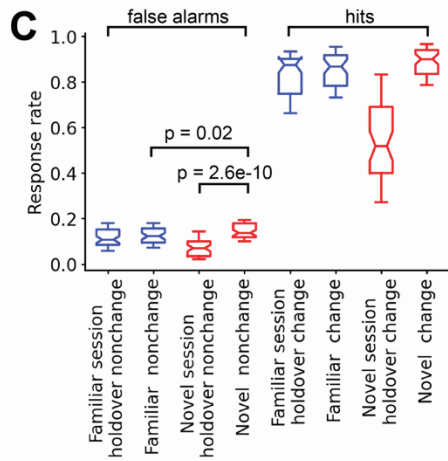
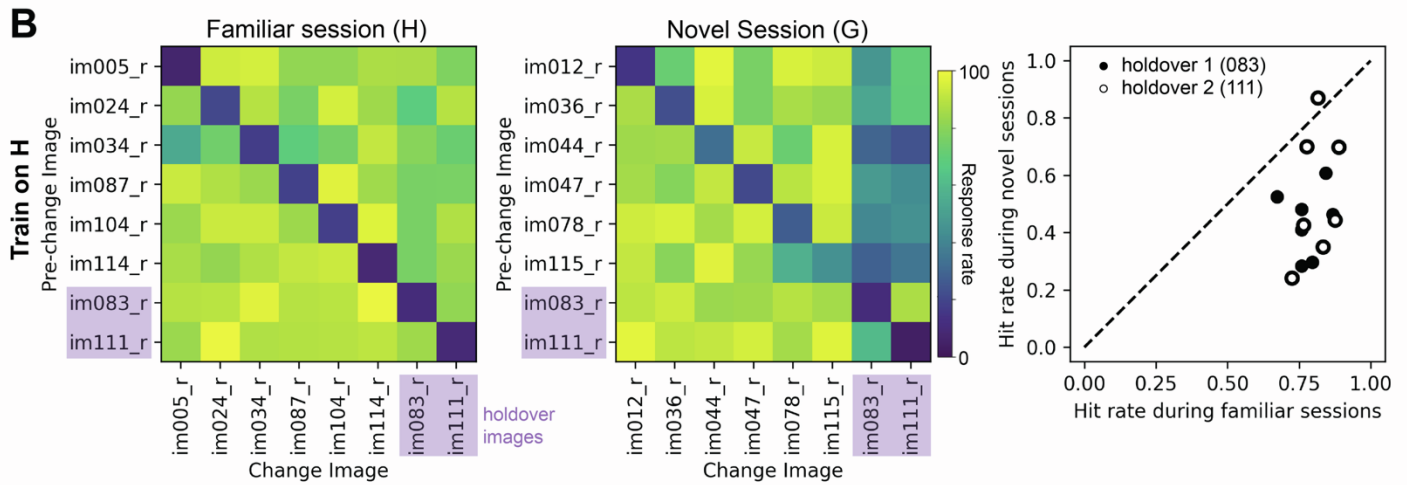
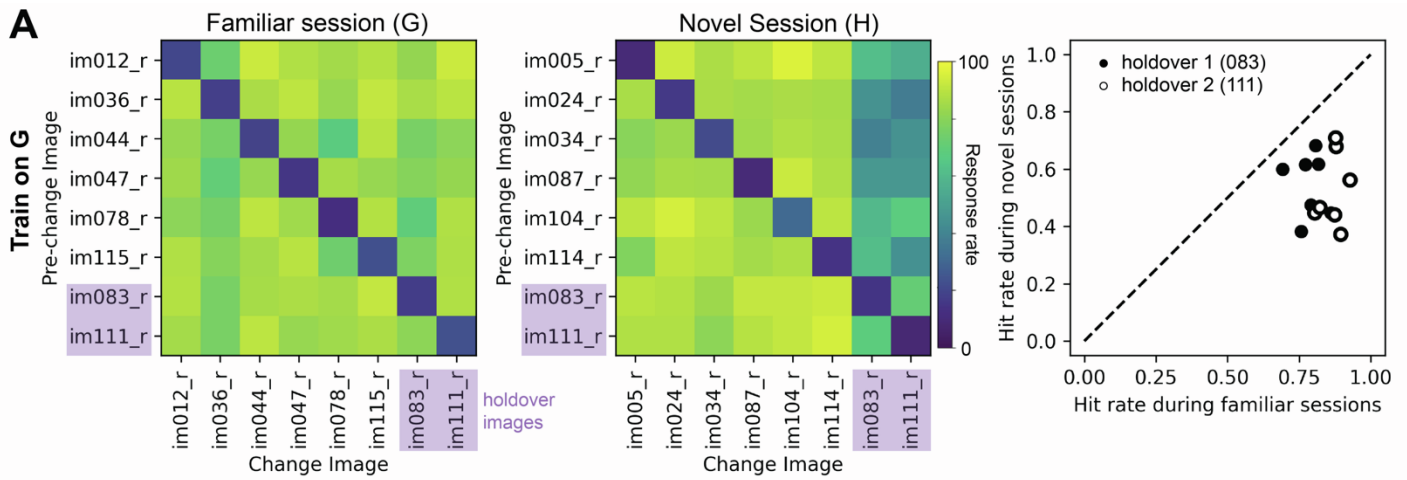


Figure S14: Visual change detection strategy. Related to Figure 7. A) Left: Mean response probability for each image transition in image set G for mice trained on G. Holdover images appear in final two rows and columns (purple shading). Middle: Mean response probability for image set H for mice trained on G. Right: Mean response probability for each image transition to holdover images (non-diagonal elements of final two columns) during familiar vs novel session. Image 111 shown in open symbols; image 83 in closed symbols. **B)** Same as (A) but for mice trained on image set H. **C)** Top: Response probabilities during familiar and novel sessions to various stimulus conditions. Bottom: P-value comparisons calculated with Wilcoxon rank-sum test corrected for multiple comparisons. **D)** Lick decoding accuracy during repeat stimuli using the image-comparison decoding strategy or the change decoding strategy. * indicates $p < 0.05$ after correction for multiple comparisons. **E)** Image decoding from visual cortical neurons over time. Decoding was run on activity in a 50 ms sliding window with a new decoder trained for each time step. Pre-omission stimuli were chosen to illustrate how image information decays over the longer inter-stimulus interval. Dotted line indicates chance performance. **F)** Left: Boxplot showing change-decoder weights given to units of each functional cluster, indicating their relative contributions to the “adaptation” strategy. Right: P-value matrix for cluster feature weight distributions calculated by Wilcoxon rank-sum test and corrected for multiple comparisons.

Table S1: Data Resources associated with this dataset. Related to STAR Methods.

Data Resource	Format	Uniform Resource ID (URI)
Behavior + spike data	NWB	s3://visual-behavior-neuropixels-data/visual-behavior-neuropixels/, https://doi.org/10.48324/dandi.000713/0.240702.1725
Behavior training data	NWB	s3://visual-behavior-neuropixels-data/visual-behavior-neuropixels/, https://doi.org/10.48324/dandi.000713/0.240702.1725
Local Field Potential (LFP) data	NWB	s3://visual-behavior-neuropixels-data/visual-behavior-neuropixels/, https://doi.org/10.48324/dandi.000713/0.240702.1725
Raw AP band data	Binary	s3://allen-brain-observatory/visual-behavior-neuropixels/raw-data/
Raw LFP band data	Binary	s3://allen-brain-observatory/visual-behavior-neuropixels/raw-data/
Eye, face, body videos	MP4	s3://allen-brain-observatory/visual-behavior-neuropixels/raw-data/

Table S2: Technical Resources associated with this dataset. Related to STAR Methods.

Technical Resource	Format	Uniform Resource ID (URI)
AllenSDK tutorials	Jupyter Notebooks	https://allensdk.readthedocs.io/en/latest/visual_behavior_neuropixels.html
Visual Behavior Neuropixels Databook	Jupyter Book	https://allenswdb.github.io/physiology/ephys/visual-behavior/VB-Neuropixels.html
Visual Behavior Neuropixels Technical Whitepaper	PDF	https://portal.brain-map.org/circuits-behavior/visual-behavior-neuropixels

AllenSDK tutorials: Jupyter notebooks demonstrating how to access data with the AllenSDK and perform basic analysis on behavior, spike and LFP data.

Visual Behavior Neuropixels Databook: Jupyter Book providing background information about the experimental design and change detection task, as well as in-depth descriptions of all metadata tables in this dataset.

Visual Behavior Neuropixels Technical Whitepaper: Document providing detailed information about the Allen Brain Observatory data collection pipeline and rig hardware.