Stimulus novelty uncovers coding diversity in visual cortical circuits

Marina Garrett*, Peter Groblewski*, Alex Piet*, Doug Ollerenshaw*, Iryna Yavorska*, Farzaneh Najafi*, Adam Amster, Corbett Bennett, Michael Buice, Shiella Caldejon, Linzy Casal, Florence D'Orazi, Scott Daniel, Saskia EJ de Vries, Daniel Kapner, Justin Kiggins, Jerome Lecoq, Peter Ledochowitsch, Sahar Manavi, Nicholas Mei, Christopher B. Morrison, Sarah Naylor, Natalia Orlova, Jed Perkins, Nick Ponvert, Clark Roll, Sam Seid, Derric Williams, Allison Williford, Ruweida Ahmed, Daniel Amine, Yazan Billeh, Chris Bowman, Nicholas Cain, Andrew Cho, Tim Dawe, Max Departee, Marie Desoto, David Feng, Sam Gale, Emily Gelfand, Nile Gradis, Conor Grasso, Nicole Hancock, Brian Hu, Ross Hytnen, Xiaoxuan Jia, Tye Johnson, India Kato, Sara Kivikas, Leonard Kuan, Quinn L'Heureux, Sophie Lambert, Arielle Leon, Elizabeth Liang, Fuhui Long, Kyla Mace, Ildefons Magrans de Abril, Chris Mochizuki, Chelsea Nayan, Katherine North, Lydia Ng, Gabriel Koch Ocker, Michael Oliver, Paul Rhoads, Kara Ronellenfitch, Kathryn Schelonka, Josh Sevigny , David Sullivan, Ben Sutton, Jackie Swapp, Thuyanh K Nguyen, Xana Waughman, Joshua Wilkes, Michael Wang, Colin Farrell, Wayne Wakeman, Hongkui Zeng, John Phillips, Stefan Mihalas, Anton Arkhipov, Christof Koch, Shawn R Olsen

Allen Institute, Seattle, WA, USA

The detection of novel stimuli is critical to learn and survive in a dynamic environment. Though novel stimuli powerfully affect brain activity, their impact on specific cell types and circuits is not well understood. Disinhibition is one candidate mechanism for novelty-induced enhancements in activity. Here we characterize the impact of stimulus novelty on disinhibitory circuit components using longitudinal 2-photon calcium imaging of Vip, Sst, and excitatory populations in the mouse visual cortex. Mice learn a behavioral task with stimuli that become highly familiar, then are tested on both familiar and novel stimuli. Mice consistently perform the task with novel stimuli, yet responses to stimulus presentations and stimulus omissions are dramatically altered. Further, we find that novelty modifies coding of visual as well as behavioral and task information. At the population level, the direction of these changes is consistent with engagement of the Vip-Sst disinhibitory circuit. At the single cell level, we identify separate clusters of Vip, Sst, and excitatory cells with unique patterns of novelty-induced coding changes. This study and the accompanying open-access dataset reveals the impact of novelty on sensory and behavioral representations in visual cortical circuits and establishes novelty as a key driver of cellular functional diversity.

The processing and prioritization of novel stimuli is critical for adaptive behavior in a constantly changing world (Rust and Cohen, 2022). Stimulus novelty directs attention, enhances perception, promotes exploration, and facilitates learning and memory (Jaegle et al., 2019; Ranganath and Rainer, 2003; Schomaker and Meeter, 2015). Novelty is a deviation from learned expectations; here we distinguish three forms of novelty depending on what has been learned, and over what timescale. *Contextual novelty* arises when a rare or unexpected stimulus appears among more common, repeated stimuli, thus deviating from local stimulus context. The *omission* of a stimulus from an expected or predictable sequence is a second form of novelty or surprise (Braga and Schönwiesner, 2022). Finally, *absolute stimulus novelty* is defined as the exposure to a stimulus that has not been experienced before.

The neural effects of novelty are widespread, affecting cortical, subcortical, and neuromodulatory systems (Braga and Schönwiesner, 2022; Kafkas and Montaldi, 2018; Ranganath and Rainer, 2003; Schomaker and Meeter, 2015; Tapper and Molas, 2020). A common observation across species and sensory modalities is that novel stimuli evoke larger sensory responses compared to familiar stimuli (Ranganath and Rainer, 2003; Schomaker and Meeter, 2015). This novelty effect (or it's inverse, the familiarity effect) is observed both at the macro-level using techniques such as EEG and fMRI in humans (Courchesne et al., 1975; Daffner et al., 2000; Hawco

and Lepage, 2014), and at the single cell level using single unit recordings in animals (Li et al., 1993; Meyer and Rust, 2018; Xiang and Brown, 1998; Zhang et al., 2022). Contextual novelty effects such as repetition suppression, mismatch negativity, and deviance detection are thought to arise through adaptation (Braga and Schönwiesner, 2022; Garrido et al., 2009; Grill-Spector et al., 2006; Hu et al., 2021; Nelken and Ulanovsky, 2007). However, the cell types and circuits that underlie the lifetime history-dependent effects of absolute stimulus novelty are unknown.

Disinhibition has been proposed as a mechanism of stimulus novelty detection (Schulz et al., 2021). An emerging body of literature has identified a cortical microcircuit with vasoactive intestinal peptide (Vip) expressing inhibitory interneurons enhancing the activity of excitatory neurons by selectively inhibiting somatostatin (Sst) expressing inhibitory interneurons (Fu et al., 2014; Karnani et al., 2016; Krabbe et al., 2019; Lee et al., 2013; Letzkus et al., 2015; Pfeffer et al., 2013; Pi et al., 2013). Supporting the hypothesis that the Vip-Sst disinhibitory circuit is at the heart of novelty processing, Vip neurons are activated by novel stimuli (Garrett et al., 2020; Melzer et al., 2021). Experiments in the primary visual cortex (Hayden et al., 2021), primary auditory cortex (Kato et al., 2015; Natan et al., 2017, 2015), and in the hippocampus (Arriaga and Han, 2019) show that Sst inhibitory neurons are more responsive to familiar compared to novel stimuli. Familiarity-enhanced responses in Sst cells could serve to suppress local excitatory cells in a stimulus-specific manner, thereby leading to reduced cortical responses to familiar stimuli. The absence of this Sst cell-mediated inhibition, through Vip-mediated disinhibition, may support enhanced excitatory responses to novel stimuli, thus gating plasticity and triggering the formation of new representations (Park et al., 2021; Schulz et al., 2021). However, experiments unifying these observations in a single paradigm are lacking.

In the visual cortex of mice, neurons not only code for visual features, but also for behavioral choices, reward expectation, locomotor activity, and sensorimotor mismatch (Fiser et al., 2016; Koay et al., 2022; Liu et al., 2015; Musall et al., 2019; Niell and Stryker, 2010; Pakan et al., 2018; Poort et al., 2015; Stringer et al., 2019). The extent to which novelty affects coding for behavior and task-related information is unknown.

Furthermore, excitatory neurons as well as Vip and Sst inhibitory subclasses are composed of multiple cell types with distinct gene expression patterns, morphologies, synaptic connectivity, and intrinsic electrical properties (Gouwens et al., 2020, 2019; Jiang et al., 2015; Kepecs and Fishell, 2014; Tasic et al., 2018; Tremblay et al., 2016; Yao et al., 2021; Zeng and Sanes, 2017). We sought to address whether different subpopulations of excitatory and inhibitory cells are differentially involved in novelty processing.

Here, we measure the effect of absolute stimulus novelty on excitatory, Vip, and Sst inhibitory populations in the mouse visual cortex during visually guided behavior. By recording in the context of a visual change detection task (Figure 1), we investigated how absolute stimulus novelty influenced activity following repeated, expected stimulus presentations, unexpected stimulus changes, and unexpected stimulus omissions. We find changes in population activity that are consistent with a role for disinhibition in stimulus novelty processing (Figure 2). By modeling the relationship of neural activity to behavioral responses, locomotion, pupil dynamics, licking, and reward consumption, we discover that novelty impacts coding for not only sensory but also behavioral and task-related information (Figure 3). Finally, we demonstrate diversity in novelty-induced changes in coding within populations of Vip, Sst, and excitatory cells through functional clustering of cellular coding across days (Figure 4).

These findings depended on the generation of a large-scale standardized dataset using a team science approach (de Vries et al., 2020; Siegle et al., 2021). This public dataset represents a rich substrate for mining neural activity associated with stimulus and behavioral processing, serving as an open resource for further discovery. The full dataset can be accessed at https://portal.brain-map.org/explore/circuits/visual-behavior-2p.

Measuring the impact of stimulus novelty on visual cortical circuits

To test the impact of stimulus novelty on neural coding by distinct cell types in the context of an active behavioral task, we used the Allen Brain Observatory 2-photon calcium imaging pipeline to collect an open-access dataset using standardized experimental procedures (de Vries et al., 2020) (Fig. 1a). We focused on neural populations involved in the cortical Vip-Sst disinhibitory circuit using transgenic mice that expressed the genetically encoded calcium indicator GCaMP6f in excitatory neurons (Slc17a7-IRES2-Cre;Camk2a-tTa;Ai93), Sst inhibitory neurons (Sst-IRES-Cre;Ai148), or Vip inhibitory neurons (Vip-IRES-Cre;Ai148) (Fig. 1b,c). Using both single and multi-plane 2-photon imaging, we measured activity from primary (VISp) and secondary (VISl) visual areas, across multiple cortical depths (Extended Data Fig 1a). In this study, we focused on 134 unique fields of view in 66 mice during 202 imaging sessions (14,495 total cells; 12,826 excitatory cells, 468 Sst cells, 1,201 Vip cells).

Prior to imaging, mice learned a visual change detection task through an automated training procedure (Groblewski et al., 2020)(Fig. 1d, Extended Data Fig. 1b). The final stage used a set of 8 natural scene images that became highly familiar to the mice over weeks of training (21+/-14 sessions with familiar image set prior to imaging) (Fig. 1e; Extended Data Fig. 1c-e). The task consisted of briefly presented images with an interleaved gray screen (250 ms stimulus, 500 ms gray screen); mice were rewarded for licking in response to a change in image identity within 750 ms (Fig 1d). This paradigm is a 'roving standard' change detection task in which the oddball stimulus on one trial becomes the standard for the next trial. Thus, each image can be either the standard or the oddball depending on the local context. Once mice were well-trained, they performed the task under the 2-photon microscope, first with the familiar image set, and subsequently with a novel natural scene image set they had not seen before (Fig. 1e; Extended Data Fig. 1b-e). Behavioral performance was consistent across familiar and novel image sessions (Fig. 1f, g; Extended Data Fig. 1f; Extended Data Fig. 2e,f), indicating mice had learned the task rules and generalized their change detection behavior to novel stimuli within a single session.

A key aspect of our experimental design was to include and measure a rich set of stimulus, behavioral, and task-related features to evaluate how the three forms of novelty impact not only sensory coding but also behavioral and task coding. In the sensory domain, we assessed neural activity associated with the repeated stimuli (standards), unexpected changes in stimulus identity (oddballs), and unexpected stimulus omissions (5% of non-change stimulus presentations were omitted during imaging sessions; Fig. 1d). For behavioral features, we monitored the animals' running speed, pupil diameter, and licking behavior. Task-related information included tracking hits and misses, and the delivery of water rewards (Fig. 1h). Finally, to assess how exposure to stimulus novelty impacts single cell coding dynamics across multiple days, we tracked individual neurons across multiple experimental sessions using familiar and novel images: one session with familiar images, the first day of exposure to novelty, and a subsequent session of re-exposure to the same novel image set, referred to as the novel+ session (Fig. 1e; Extended Data Fig. 2a-d, Extended Data Fig. 3).

While our study focuses on the impact of novelty on neural representations using these three session types, our publicly available dataset includes longitudinal tracking of neurons in up to ten 2-photon imaging sessions, comprised of both active behavior and passive viewing (Extended Data Fig. 2a, d), as well as the complete behavioral training and imaging history for all mice. The full imaging dataset includes 326 unique fields of view from 107 mice (50,482 total neurons imaged; 45,780 excitatory cells, 1,332 Sst cells, 3,370 Vip cells). The full behavior training dataset includes 4,787 sessions.

Stimulus novelty differentially impacts excitatory, Vip, and Sst cell populations

We first sought to determine whether stimulus novelty modulated the strength of sensory

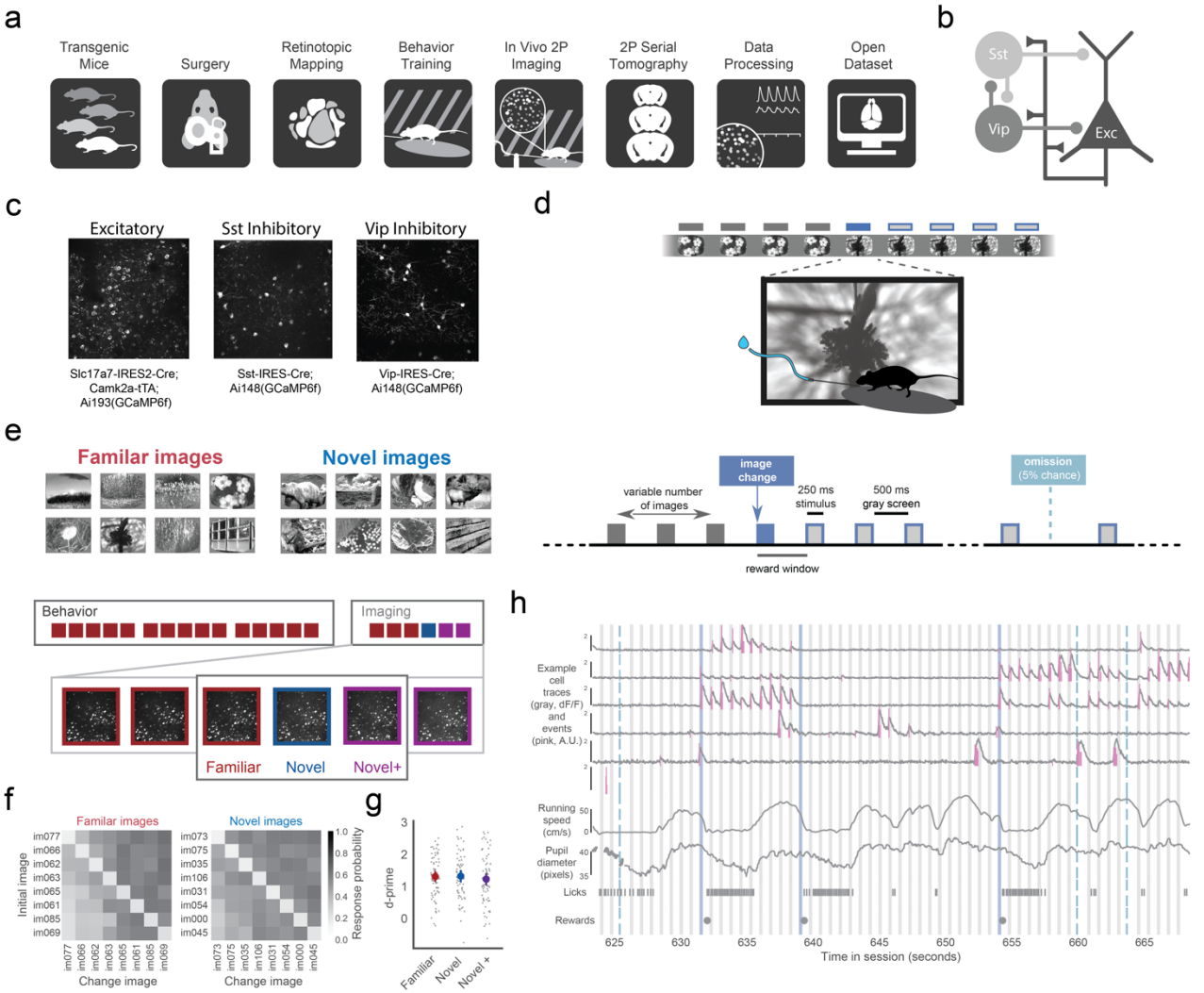


Figure 1 | Standardized measurement of behavior and neurophysiology in mice performing a change detection task with familiar and novel stimuli. a, Data was collected using the Allen Brain Observatory, a standardized high-throughput data collection platform where a series of experimental procedures are performed by specialized teams and subject to strict quality control procedures. **b,** Schematic of a disinhibitory circuit motif whereby Vip Inhibitory cells suppress Sst Inhibitory cells to enable disinhibition of Excitatory populations. **c,** in vivo 2-photon calcium imaging was performed in three transgenic mouse lines expressing GCamp6f in excitatory neurons (Slc17a7 expressing) or in distinct sub-classes of inhibitory neurons (Sst or Vip expressing). **d,** Mice were trained to perform a go/no-go change detection task with serially presented flashed stimuli. Stimuli were presented for 250ms, separated by a 500ms gray screen period. Mice learned to lick within 750ms after a change in image identity to earn a water reward. A variable number of repeated stimulus presentations (gray boxes with blue outlines) occurred after each change. 5% of non-change stimuli were randomly omitted during imaging sessions (but not during training). **e,** Mice were trained with a set of 8 natural scene images that became highly familiar over the course of training. Once mice were well trained, they performed the task under the 2-photon microscope with either the familiar image set or a novel set of 8 images that the mice had not seen before. Neural populations were tracked over multiple 2-photon recording sessions during task performance with the familiar and novel images. This study focuses the last familiar session, the first novel image session, and a subsequent novel image session where mice were re-exposed to the new image set (designated as novel+ throughout figures and text). **f,** Behavioral response probability across all image transitions for the familiar and novel image sets. Catch trials (non-changes) are on the diagonal. **g,** Behavior performance, measured as d-prime, for familiar, novel, and novel+ re-exposure sessions, demonstrating that mice can generalize task performance to novel stimuli. d-prime was calculated for the most detectable image in each image set as the most direct comparison of generalization. Behavior performance for all image transitions is shown in Extended Data Fig. 2. **h,** Simultaneous measurement of neural activity and behavior during the change detection task. First six rows show dF/F traces (gray) and detected calcium events (pink) for example Excitatory neurons. Running speed, pupil diameter, licks, and reward times are also measured. Gray bars indicate repeated stimulus presentations, blue bars indicate stimulus changes, and light blue dashed lines are stimulus omissions.

responses in each of the three cell classes. To do this we computed average stimulus-triggered population responses separately in sessions with familiar versus novel images. Responses were quantified as the averaged magnitude of events detected from changes in calcium fluorescence (see Methods). Excitatory cells had larger image-evoked responses during the novel session compared to the familiar session (Fig 2b,c; Extended Data Fig. 4a), for both expected, repeated images (Fig. 2e) and for

unexpected image changes (Fig. 2f). Consistent with a role for the Vip-Sst disinhibitory circuit in facilitating enhanced excitatory activity with novelty, Sst neurons displayed much reduced levels of stimulus-driven activity during the novel session compared to the familiar session, for both changes and repeated pre-change images, while Vip cells showed the opposite pattern and were strongly activated by change and pre-change images in the novel, but not the familiar, session (Fig. 2b,c,e,f; Extended Data Fig. 4).

With repeated experience, novel stimuli become familiar. We measured this familiarization process by recording activity in subsequent experimental sessions in which the novel image set was presented again to the mice during task performance. These sessions were acquired 1-3 days (2.1+/-0.9 sessions) after the initial novel session; we refer to this subsequent re-exposure with novel stimuli as the 'novel+' session (Extended Data Fig. 2a, b, c). Strikingly, in the Vip and Sst populations, the magnitude of image-evoked responses in the novel+ returned to the baseline level measured in the familiar session (Fig. 2b-d; Extended Data Fig. 4), despite having an order of magnitude less experience with the novel stimuli. Indeed, the response magnitude in the novel+ session was not statistically different from responses in the familiar session (Fig. 2e,f). Image responses in the excitatory population were substantially lower in the novel+ compared to novel session (Fig. 2e,f), but did not fully return to the familiar level. These results show that VIP mediated disinhibition is specific to the first session with novel images, and demonstrate a rapid timescale of stimulus familiarization in inhibitory populations, on the order of hours to days.

To test whether these changes in activity modified the sensory information that could be read-out from simultaneously measured population activity, we performed several decoding analyses. First, we found that population decoding of image identity was enhanced in novel sessions in excitatory and Vip cells (Extended Data Fig. 5a), indicating that novelty increases population selectivity for images in these two cell classes. Second, we found that we were able to decode changes from repeated stimulus presentations with higher accuracy from both the Vip and excitatory population during the novel compared to familiar sessions (Fig. 2g; Extended Data Fig. 5b), suggesting that novelty enhances the population response to unexpected image changes in these two cell classes. Interestingly, novelty-induced enhancements in decoding of image identity and image change were also reversed in the novel+ session in excitatory and Vip populations (Fig. 2g; Extended Data Fig. 5b). Novelty did not have a significant effect on image identity and image change decoding from Sst population activity.

We previously found that Vip neurons display ramping activity dynamics during the inter-stimulus interval in sessions with familiar stimuli, and continue to ramp up when images are unexpectedly omitted from a sequence of repeated images (Garrett et al., 2020). Consistent with prior results, this ramping activity in Vip neurons is present in familiar sessions but absent (or much reduced) in the first novel session (Fig. 2i,j). Remarkably, on subsequent exposure to the novel image set (novel+), ramping activity during stimulus omission re-emerged in the Vip population (Fig. 2k), representing a rapid experience-dependent change in neural coding and dynamics. Omission signals were more readily decoded from population activity during familiar and novel+ sessions than in the first novel session for Vip cells. There was no significant difference in omission decoding in excitatory or Sst cells (Fig. 2m; Extended Data Fig. 6a), despite the presence of some omission related activity in familiar sessions for the excitatory population (Fig 2i).

Given that Vip omission-related activity was stronger during sessions with familiar images, we wondered whether the omission signal predicts upcoming familiar image identity. However, it was not possible to decode image identity from population activity during omissions (Fig. 2n; Extended Data Fig 6b). We also compared omission-related activity during change detection behavior with familiar stimuli to passive viewing sessions with the same stimuli played in open loop and found that omission signals persisted, indicating that omission signals in Vip cells are related to stimulus familiarity independent of task engagement (Extended Data Fig. 7a,b).

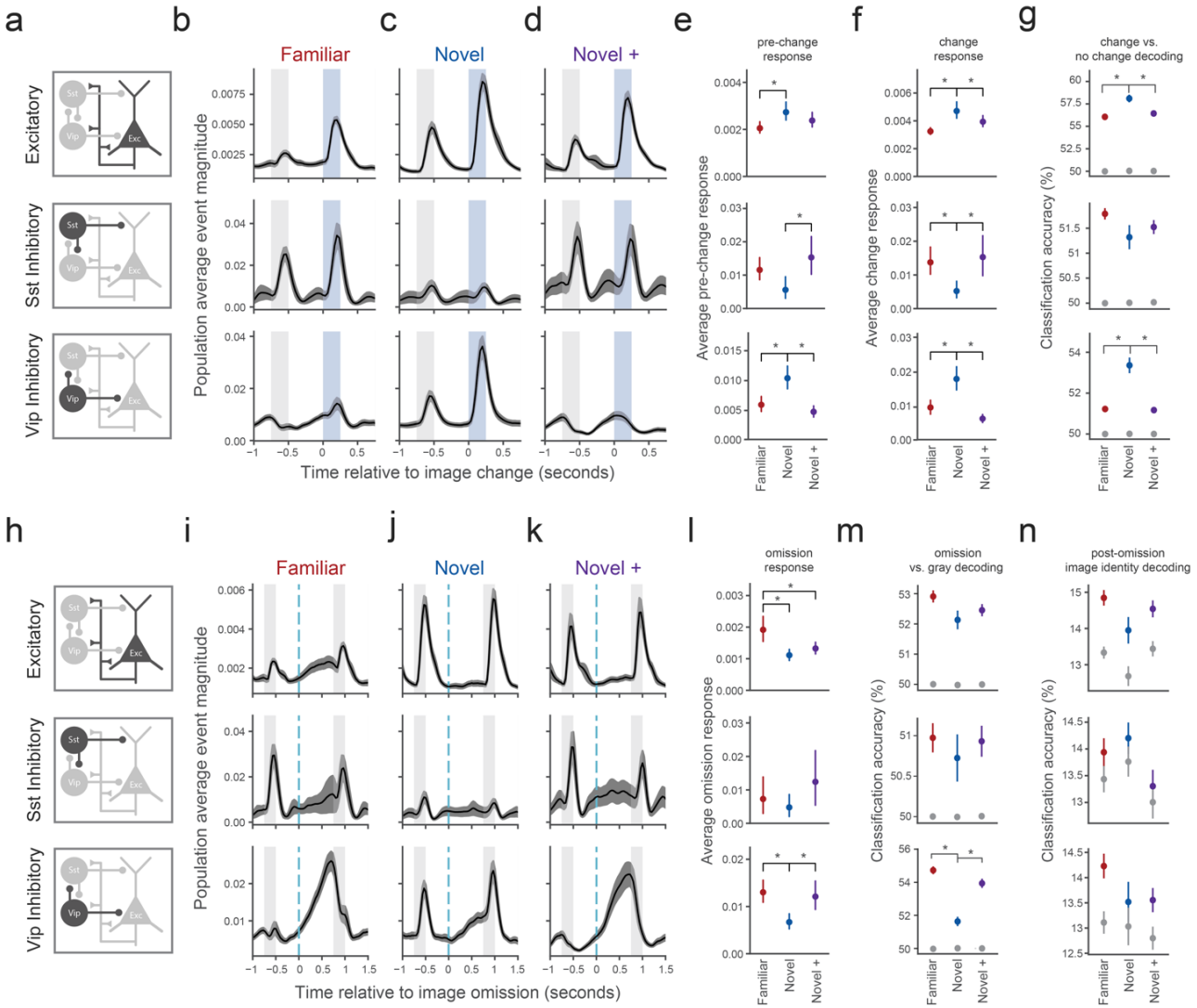


Figure 2 | Novelty differentially influences responses to expected and unexpected sensory stimuli across cell subclasses and is consistent with a disinhibitory effect. a, Disinhibitory circuit schematic highlighting the relevant cell subclasses represented in each row of b-g. Top row is Excitatory cells (n=34 mice, 12,826 cells), middle row is Sst cells (n=15 mice, 468 cells), bottom row is Vip cells (n=17 mice, 1,201 cells). **b,** Population activity in response to image changes during familiar image sessions. Activity is quantified as the magnitude of calcium events detected from fluorescence traces at each time point, averaged over all recorded neurons across mice. Error bars are +/- SEM. **c,** Population activity in response to image changes during the first novel image session. **d,** Population activity in response to image changes during re-exposure to the novel image set in the novel+ session. **e,** Average event magnitude for the image presentations prior to image changes (indicated by gray bars in b-d) across experience levels. Novelty enhances pre-change responses in Excitatory and Vip populations. Error bars in e-g are 95% confidence intervals. Differences between experience levels are marked as significant after a one-way ANOVA, followed by Tukey HSD ($p < 0.05$). **f,** Average event magnitude for image changes (indicated by blue bars in b-d). Novelty enhances change-evoked activity in Excitatory and Vip populations but suppresses activity of the Sst population. Response magnitude is not significantly different between familiar and novel+ sessions. **g,** Population decoding accuracy of change vs. pre-change stimuli shows enhanced representation of image change with novelty in Excitatory and Vip, but not Sst, populations. Classification accuracy is performed on neural event timeseries then averaged over 400ms window after image onset (Extended Data Fig. 5) for each session, then averaged across sessions. Gray points show classification accuracy of shuffled data. **h,** Disinhibitory circuit schematic highlighting the relevant cell class represented in each row of i-n. **i,** Population activity following image omissions during familiar image sessions. Activity is quantified as the magnitude of calcium events detected from fluorescence traces, averaged over all recorded neurons across mice. Error bars are +/- SEM. **j,** Population activity following image omissions during the first novel image session. **k,** Population activity following image omissions during re-exposure to the novel image set in the novel + session. **l,** Average response magnitude for omissions is reduced during novel sessions in Excitatory and Vip populations. Error bars are 95% confidence intervals. **m,** Population decoding accuracy of omission-evoked activity vs. activity during preceding gray screen is reduced by novelty in the Vip population. Classification accuracy is performed on neural timeseries then averaged over 750ms window after the time of omission vs. the 100ms window prior to the omission for each session, then averaged across sessions. Gray points show classification accuracy of shuffled data. **n,** Population decoding accuracy for the identity of the image shown before and after the omission, demonstrating that omission related activity does not convey image specific information. Gray points show classification accuracy of shuffled data. Differences between experience levels are marked as significant after one-way ANOVA followed by Tukey HSD ($p < 0.05$) for panels l, m, n.

Mice adjusted their running and licking behavior according to the stimulus presentation frequency, typically slowing just before a stimulus was presented, and licking shortly after. This pattern of behavior was consistent across familiar and novel sessions, demonstrating that behavioral differences do not account for the dramatic effects of novelty on neural activity (Extended Data Fig. 8b,c,e,f). The average pupil diameter was slightly larger during novel sessions, potentially related to an increased arousal with novelty. However the dynamics of pupil fluctuations following image changes and omissions were consistent across familiar and novel sessions (Extended Data Fig. 8a,d), indicating that the increase in Vip omission activity during familiar sessions was not tied to a similar increase in pupil diameter. Interestingly, pupil size was not different between novel and novel+ sessions, despite major differences in neural activity in these sessions, further dissociating behavioral and neural effects.

Because image contrast and sensory tuning can affect neural response magnitudes and dynamics similarly to stimulus novelty (Millman et al., 2020; Rust and Cohen, 2022), we confirmed that the results described above were truly due to the novelty of the stimulus and not the particular image set that was selected to be familiar or novel by training one group of mice with the opposite image set compared to the rest of the mice. In this cohort, image set B (the novel set for most mice in the study) was used for training and became highly familiar, while image set A (the familiar set for the majority of mice) was used as the novel image set. These results were consistent with novelty as the primary driver of neural activityl differences at the population level, rather than the image sets themselves (Extended Data Figure 7c-e).

Novelty alters multiplexed coding of sensory, behavioral, and task information

We next addressed the question of whether stimulus novelty impacts coding for behavioral and task information. We used a linear regression model with time-dependent kernels to describe each cell's activity trace as a linear combination of responses to sensory, behavioral, and task features (Engelhard et al., 2019; Musall et al., 2019; Pillow et al., 2008; Steinmetz et al., 2019). Features used in the model included image presentations, image omissions, hit and miss task events (i.e., image changes with and without licking responses and rewards), and behavioral measures including licking, running speed, and pupil diameter (Fig. 3a). The kernel weights for each cell describe the linear influence of each feature on neural activity, separated from other ongoing task events. Model performance was correlated with each cell's activity trace signal-to-noise ratio, indicating that cells with highly dynamic activity were well fit by the model (Extended Data Fig. 9c).

We computed the unique contribution of each feature to overall model performance by comparing the performance of the full model with reduced models fit without each feature or group of features (Fig. 3a). We refer to this fractional reduction in variance explained as the "coding score". As an illustrative example, we show the coding score and kernels for stimulus omissions for all cells in familiar sessions in Fig. 3b (left). Cells with non-zero coding scores have larger amplitude kernels. Averaging the omission kernels for all cells within each cell class (Vip, Sst, excitatory) revealed that our model captured the expected time course of ramping dynamics in Vip cells during familiar (Fig. 3b, right), novel (Fig 3c, left), and novel+ image sessions (Fig. 3c, right). Kernels and coding scores for all cells are shown in Extended Data Figs. 10-12.

We grouped features into four major model categories: images, omissions, behavioral, and task information (Fig. 3a) and examined how the different cell classes coded for these feature categories (Fig. 3d; see Extended Data Fig. 8d,e for all ungrouped features). We found that activity of excitatory and Sst cells was best explained by images, whereas Vip cell activity was best explained by a combination of images, omissions, and behavioral features (Fig. 3d). Task coding was relatively low on average across all cell classes but was higher for Sst and excitatory cells compared to Vip cells (Fig. 3d).

We then split the data according to experience level (familiar, novel, novel+) and evaluated how the average coding score for each feature category was

influenced by stimulus novelty (Fig. 3e). These results were calculated for the full population of recorded cells (colored disconnected points in Fig. 3e) and for the subset of cells that were tracked across all 3 experience levels (gray lines in Fig. 3e). Consistent with changes to image-evoked activity with novelty (Fig. 2), excitatory and Vip populations increased the strength of image coding during the first novel image session, while image coding scores decreased in Sst cells. Omission coding by Vip cells was reduced by novelty and re-emerged in the subsequent novel+ session, consistent with model-free measures of activity (Fig. 2). The Sst population had a small but significant reduction in omission coding with novelty.

Vip cells showed the strongest encoding of behavioral information, which was selectively increased in the novel session (Fig. 3e). There was a small but significant increase in behavioral coding in the novel+ session for excitatory cells, suggesting that alterations in behavioral coding by excitatory neurons occurs over a slower timescale than in Vip cells. The Sst population showed a similar trend. Finally, task encoding also transiently increased during the first novel image session in Vip and excitatory populations, whereas Sst task coding scores transiently decreased. Task coding in all cell classes returned to the familiar level after the first session of exposure to novel images, indicating a relatively fast timescale mechanism for enhancing task-related signals with novelty. These results were consistently observed when using ungrouped model features (Extended Data Fig. 12a) or using stricter cell selection criteria (Extended Data Fig. 12b,c).

We further split our data by cortical area and cortical depth (Extended Data Fig. 14). In excitatory cells, we see stronger image coding in VISp and stronger coding for omissions, behavioral, and task variables in VISl in all sessions, demonstrating area specific coding independent of experience level. In contrast, area differences in coding across the Vip population were experience-dependent; omission and task coding were stronger in VISl during familiar sessions, while image coding was stronger in VISl in the novel+ session. This suggests a transition from expectation and task signaling to more sensory based coding with novelty in Vip cells. In the Sst population, behavioral coding was consistently higher in VISl across all experience levels, similar to excitatory cells. Sst omission coding was increased in VISp in the familiar session. Coding for all feature categories was enhanced in deeper layers across the excitatory population. No major differences by cortical depth were observed for Vip or Sst cells.

Together, these results demonstrate that stimulus novelty not only alters stimulus and omission encoding but also modifies coding for other behavioral and task features. In the case of image and task features, the average changes in the coding scores of Vip and excitatory cells were modified in the same direction, whereas the Sst population changed in the opposite direction—this is consistent with the Vip-Sst disinhibitory circuit. These effects were present across cortical depths, indicating that novelty shapes coding across the column of cortical circuitry. However, these results are based on population averages and may obscure meaningful heterogeneity in experience dependent coding changes in individual neurons, as can be observed in Fig. 3b and Extended Data Figs. 10-12. Therefore, we next sought to determine diversity in the pattern of changes in coding with novelty at the level of individual neurons.

Coding diversity within cell classes

To investigate single cell diversity in terms of the strength and type of information encoded across familiar and novel sessions, we focused on the subset of cells matched in all 3 sessions (Fig. 4a; Extended Data Fig. 3; 3306 excitatory cells, 200 Sst cells, 415 Vip cells). We characterized each cell's multi-session pattern of encoding as a 12-element vector composed of the coding score for each feature category (images, omissions, behavioral, and task) across the three experience levels (familiar, novel, novel+), normalized to the maximum explained variance across sessions for that cell (Figure 4a). The matrix of neurons by feature coding across sessions (Fig. 4b) was used as the input to an unsupervised clustering analysis (see Methods; Extended Data Fig. 15), performed separately on

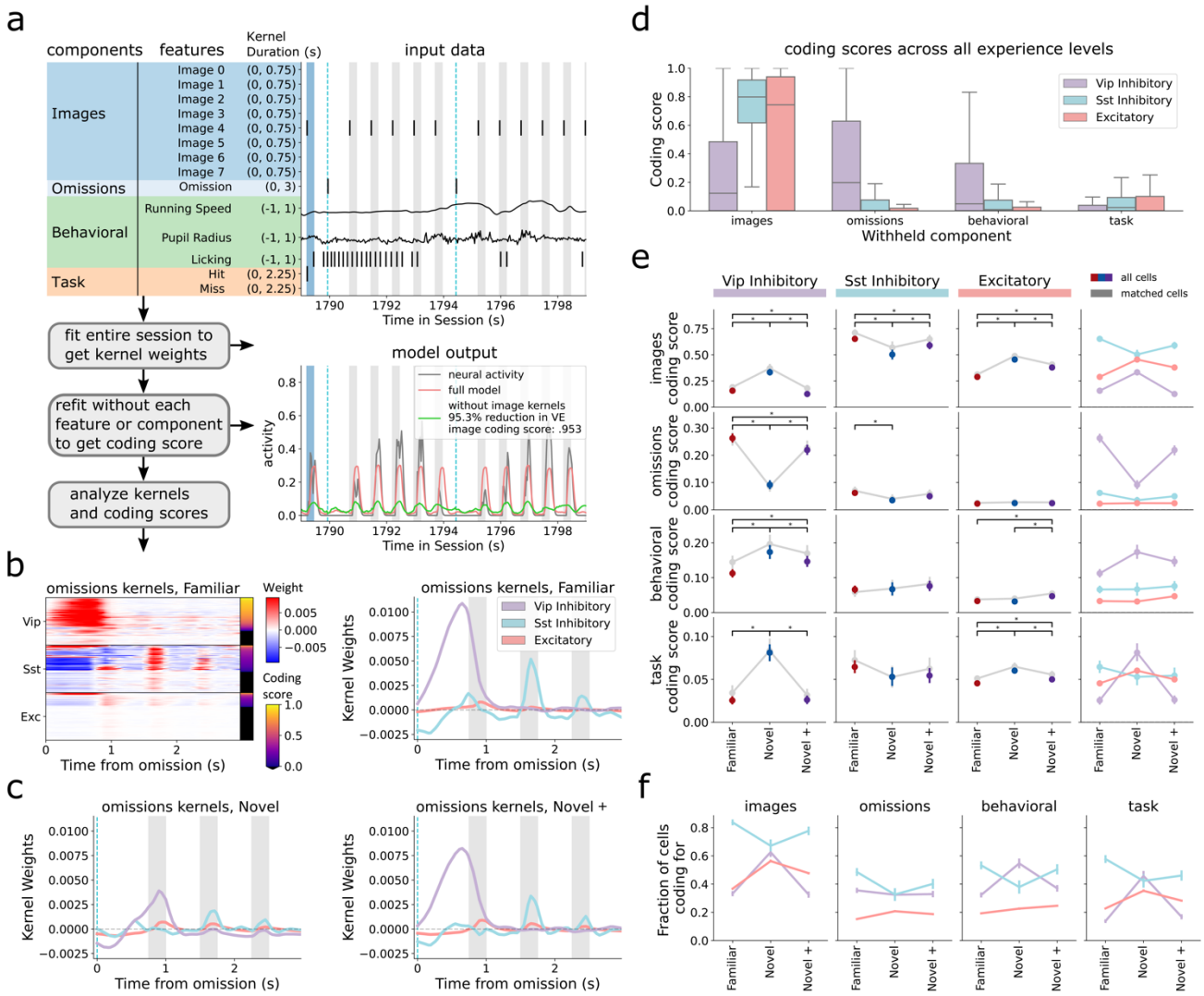


Figure 3 | Coding for stimulus, behavioral, and task information is cell class specific and is reversibly altered by novelty. a, Kernel regression model convolves input features with learned kernels and sums the output to predict neural activity. Input features can be thematically grouped into high level components: images, omissions, behavioral, and task features. The model was refit while withholding each feature or component to find the unique contribution of that feature, summarized as a coding score. **b,** Left panel, heatmap of learned omission kernels for all cells in familiar image set behavioral sessions. Cells are sorted by cell type, and then omission coding score. Right panel, average omission kernels for each cell type for familiar image sessions. **c,** Average omission kernels for the first novel session and novel+ re-exposure session. **d,** Distribution of coding scores for each model component, split by cell type. Boxplot shows the quartiles of each cell type and the whiskers mark +/- 1.5x interquartile Range (Q3-Q1). Outliers are not shown for clarity. **e,** Average coding score for each cell type (columns) by experience level (colors) and model component (rows). Colored points show all recorded neurons (mean +/- 95% CI). Connected gray lines shows the average for the subset of cells that were matched across all experience levels (mean +/- 95% CI). Differences between experience level are marked as significant after a one-way ANOVA, followed by Tukey HSD (p = 0.05). Statistics were computed on all cells. Far right column shows all cells of each cell type super-imposed. **f,** Fraction of cells with non-zero coding scores for each component split by experience and cell type (mean +/- 95% CI).

each cell class. We identified 10 excitatory clusters, 5 Sst clusters, and 10 Vip clusters (Fig 4b-d, f; Extended Data Fig. 15c). This workflow identified subpopulations of neurons with similar changes in feature coding with experience and demonstrated significant heterogeneity in response types within and across cell classes (Fig. 4b-d, average coding for each cluster in Extended Data Fig. 16). Importantly, neurons belonging to these clusters were identified across multiple mice (Extended Data Fig. 15f) and multiple clusters were identified in each mouse (Extended Data Fig. 15e), demonstrating that functional diversity is not due to differences in the behavior of individual animals or related factors.

The patterns of dynamic multi-day encoding shown by individual clusters were diverse and often different from the average patterns for a given cell subclass (mean coding for excitatory, Sst, and Vip shown in top rows of Fig. 4b, c & d).

The largest excitatory cluster (cluster 1, 22.1% of excitatory neurons) did not encode any features in any session. Most other clusters within the excitatory population primarily encoded images, and differed in the strength of encoding across familiar, novel, and novel+ experience levels (Fig. 4c). The second largest excitatory cluster (cluster 2, 18.6% of excitatory cells) showed strong image encoding that was specific to the first novel session. Excitatory cluster 3 (12% of excitatory cells) encoded images in both the first novel and subsequent novel+ sessions, indicating selectivity for the novel image set independent of experience level. Excitatory cluster 5 (9.6% of excitatory cells) encoded familiar images. The prevalence of novelty specific encoding (cluster 2) compared to image set specific encoding (clusters 3 & 5) highlights the substantial influence of stimulus novelty on sensory representations in the visual cortex. We also found several smaller clusters with multiplexed coding for images and behavioral (cluster 10, 3.2% of excitatory cells) or task features (cluster 6, 8.1% of excitatory cells, cluster 9, 5.3% of excitatory cells). These clusters had unique activity dynamics compared to image specific clusters (Extended Data Fig. 16a).

We identified fewer clusters in the Sst population compared to excitatory cells, which is consistent with lower variance of coding scores in the Sst population overall (Fig. 4e). Sst clusters also primarily encoded images, with differing strength of coding for familiar or novel stimuli. The largest Sst cluster (cluster 1, 33% of Sst cells) coded only for familiar images, while the second largest (cluster 2, 28.5% of Sst cells) showed image coding for both familiar and novel+ sessions. A third Sst cluster (cluster 3, 19.5% of Sst cells) coded for images specifically in the novel+ sessions. Together these three clusters can account for the average coding across the Sst population, but we also identified a cluster with the opposite pattern as the average, showing enhanced image coding during the first novel session (cluster 4, 16% of Sst cells).

Vip cells showed the most diversity (Fig. 4d). The largest Vip cluster (cluster 1, 22.4% of Vip cells) selectively coded for images during the first novel session. The second largest Vip cluster (cluster 2, 15.4%) was a non-coding cluster, like excitatory cluster 1. Other Vip cell clusters showed multiplexed coding of several features. Some clusters had weak image coding during the novel session, combined with omission coding during either the familiar or novel+ sessions (Vip cluster 3, 13.3% and cluster 4, 10.8%). Others showed joint coding of both images and omissions in the same session (Vip cluster 6, 8.7%, and cluster 8, 6%). One cluster had robust omission coding during familiar sessions which switched to robust image coding with novelty (Vip cluster 7, 6.5%), mirroring the population average. Finally, several Vip clusters were defined by their dynamic behavioral encoding across novelty conditions (Vip clusters 5, 9.4%, cluster 9, 4.8%, and cluster 10, 2.7%).

To evaluate whether specific functional patterns are biased to specific cortical areas or depths, we quantified for each cluster the distribution of cells across the visual areas VISp and VISl and across superficial (upper) versus deep (lower) cortical depths (upper defined as <250um from cortical surface, lower defined as >250um). None of the Sst and Vip clusters were significantly biased for depth or area (Extended Data Fig. 16), however these cell classes may still have layer specific effects due to their differential distribution across cortical depth (Tremblay et al., 2016). In contrast, nearly all the excitatory clusters were significantly biased either for cortical depth or area (Fig. 4g; Extended Data Fig. 16). Interestingly, the novelty encoding cluster (excitatory cluster 2) was biased to superficial depths of VISp, while the familiar image selective cluster was biased to lower depths, particularly in VISl (excitatory cluster 5).

It is important to note that these clusters are not of equal size (Fig. 4f). Some patterns of coding changes are overrepresented compared to others, suggesting they are not simply due to random drift in activity.

For instance, excitatory clusters 1 (22.1%) and 2 (18.6%) are ~4x larger than clusters 8 (5.4%) and 9 (5.3%), and ~2x larger than clusters 4 (9.6%) & 5 (9.6%) (Figure 4f). We further quantified the bias in cluster size with a shuffle analysis to compare the sizes of clusters in our data to a control condition where the identity of neurons was shuffled across experience levels, breaking any relationships in experience dependent coding in individual cells, while maintaining the overall distribution of coding scores within each experience level (Extended Data Fig. 17a,e). This demonstrated some clusters were significantly larger than expected by chance, whereas others were at chance level or even under-represented. Notably, novelty-enhanced excitatory and Vip clusters were significantly larger than expected by chance (excitatory clusters 2 & 3, Extended Data Fig. 17c; Vip cluster 1, Extended Data Fig. 17d). In contrast, novelty-suppressed clusters were larger than chance among Sst cells (Sst clusters 1 & 2, Extended Data Fig. 17b). These results indicate that specific patterns of changes in coding with novelty are selectively enriched or minimized within cell subclasses of the visual cortex.

As a further demonstration of the importance of novelty in defining functional clusters, we performed clustering on coding patterns measured in 3 repeated sessions with the same familiar stimulus (Extended Data Fig. 18). This procedure yielded fewer clusters for excitatory and Vip cells (6 excitatory clusters, 7 Vip clusters) and the largest clusters (i.e. the majority of neurons) showed more consistent patterns of coding across days (excitatory clusters 1 & 2, 69.6% of cells; Sst clusters 1& 2, 86.6% of cells; Vip clusters 1 & 2, 54.4% of cells), rather than dynamic patterns of changes as observed when clustering included novelty.

Ultimately, we were able to identify specific clusters of Vip, Sst, and excitatory cells that showed novelty-induced changes in image coding that were consistent with a potential disinhibitory relationship (Fig 4h). However, these clusters only represent a subset of the total population in each cell class. Thus, the heterogenous functional clusters we identified suggest a diverse set of distinct circuit motifs are engaged during novelty processing.

Discussion

Using the Allen Brain Observatory experimental pipeline, we performed longitudinal imaging of nearly 15,000 excitatory and inhibitory cortical cells in mice performing a visually guided behavioral task. Our study represents the first exploration of this rich open dataset that can be further modeled and mined for future discovery. We found that novelty altered responses to expected and unexpected sensory stimuli, stimulus omissions, and coding for behavioral and task information by Vip, Sst, and excitatory neurons in the mouse visual cortex. When we evaluated the pattern of changes in coding of individual neurons, we found that the average effects we observed were the product of multiple clusters of neurons with diverse functional properties, including differences in the direction and timescale of novelty effects. Vip neurons were the most diverse across feature encoding and novelty modulation, consistent with a highly plastic population of cells that signal salience in an experience-dependent manner.

A role for disinhibition in novelty processing

A recent modeling study (Schulz et al., 2021) demonstrated that suppression of a subpopulation of inhibitory cells can generate novelty responses in excitatory cells. We found that stimulus-evoked responses in Vip neurons are dramatically enhanced by stimulus novelty, while Sst responses are reduced, and excitatory responses increased, supporting the hypothesis that Vip neurons are the source of disinhibition as in Schulz et al.'s model.

Vip-mediated disinhibition has been shown to enhance perception, learning, and memory (Fu et al., 2014; Krabbe et al., 2019; Letzkus et al., 2015, 2011; Melzer et al., 2021; Park et al., 2021; Turi et al., 2019; Zhang et al., 2014) in similar ways to stimulus novelty (Park et al., 2021; Quent et al., 2021; Reichardt et al., 2020; Schomaker and Meeter, 2015). Vip-mediated disinhibition also plays a role in contextual modulation of visual responses and has been suggested to enhance sensitivity to weak but salient inputs (Garcia del Molino et al., 2017; Hertäg

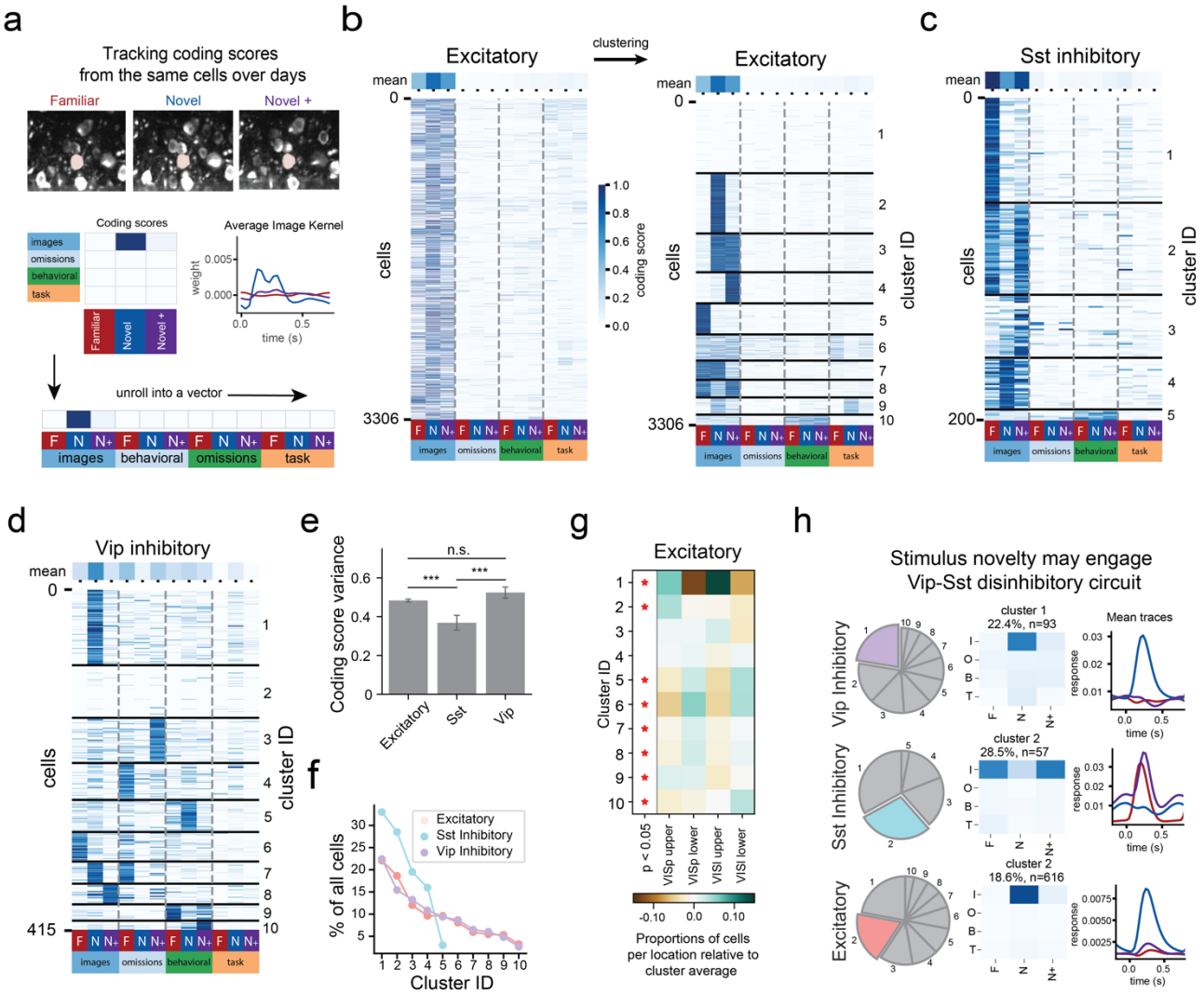


Figure 4 | Patterns of changes in coding with experience define functional subtypes within cell classes. a, Coding scores for the three experience levels (Familiar, Novel, Novel +) were computed for cells that were active, identified, and matched in all 3 sessions and formatted into a vector of four features by three experience levels for clustering. Inset shows average image kernels across experience level for the example cell shown. **b,** Left, all coding scores for Excitatory cells before clustering. Right, Excitatory coding scores after clustering procedure (see Methods & Extended Data Fig. 15) identifies 10 functional clusters with distinct patterns of changes in coding with experience. Clusters are separated by black lines and organized by cluster ID, in descending cluster size order. **c,** Clustered coding scores for Sst Inhibitory cells with five functional subtypes. **d,** Clustered coding scores for Vip Inhibitory cells with 10 functional subtypes, showing diverse patterns of changes in coding with experience. **e,** Sst Inhibitory cells show less variability in coding scores across the population than Excitatory and Vip Inhibitory cells. Difference in variance is measured as sum of squared error and marked as significant with ranksum test ($p<0.001$). **f,** The Sst inhibitory population has fewer clusters and exhibits less functional diversity than VIP inhibitory or Excitatory populations. **g,** Proportions of cells across visual area and imaging depth for each Excitatory cluster, relative to cluster average. Differences are marked as significant after Chi-squared test, followed by Benjamini-Hochberg correction. Comparable plots for Sst and Vip, which did not have significant differences are shown in Extended Data Fig. 16d. **h,** Functional clustering reveals evidence for disinhibitory Vip-Sst circuit. 22% of VIP neurons respond to images in the first Novel session only, while responses to images of 28% of Sst neurons are suppressed, and 18% of Excitatory cells are enhanced (left column). Mean coding scores of corresponding clusters are shown in feature by experience level space, with darker color indicating stronger coding score (middle column). Mean image-aligned responses of corresponding clusters are separated by experience level (familiar in red, novel in blue, and novel+ in purple), showing inverted dynamics of Sst relative to Vip and Excitatory.

and Sprekeler, 2019; Keller et al., 2020; Millman et al., 2020). Novelty induced disinhibition may function in a similar manner, by increasing the gain of excitatory responses for novel stimuli that, while salient, lack the strong recurrent connectivity that develops with learning. Novelty-induced disinhibition could then trigger the formation of new stimulus specific networks by gating plasticity (Miehl and Gjorgjieva, 2022; Wilmes and Clopath, 2019, 2019). In support of this idea, novelty can reorganize patterns of functional connectivity to enable plasticity and new learning (Park et al. 2021).

The mechanisms that enhance Vip neurons' responses to novelty remain unknown, but are likely to involve neuromodulatory (Bunzeck and Thiel, 2016; Ranganath and Rainer, 2003), top-down frontal (Zhang et al., 2014), and/or thalamic inputs (Williams and Holtmaat, 2019).

Importantly, we found that stimulus novelty effects on inhibitory responses were specific to the first behavior session with novel images, as Vip and Sst population activity dynamics returned to familiar levels upon re-exposure to the same image set in the novel+ session. Familiarization-induced changes in excitatory activity after re-exposure were less pronounced. These results are consistent with the idea that rapid inhibitory plasticity is required to re-establish stable network dynamics following disinhibition of excitatory populations, and to instruct longer term changes in excitatory circuits (Clopath et al., 2016, 2016; Miehl and Gjorgjieva, 2022; Sprekeler, 2017; Zenke et al., 2017). The striking differences in activity between the first novel session and subsequent novel+ re-exposure also suggests that a period of offline consolidation, such as during sleep, could be in involved in the expression of the familiarization effect (Aton et al., 2014; Montgomery et al., 2022).

Absolute and *contextual novelty* are distinct phenomena (Bunzeck and Düzel, 2006; Bunzeck and Thiel, 2016; Ranganath and Rainer, 2003) that are rarely studied concurrently. Our experimental design included both contextual novelty, in the form of unexpected changes to repeated stimuli, and stimuli that were novel in an absolute sense, allowing examination of potential interactions between them. Stimulus novelty increased the strength of activity for both changes and repeated stimulus presentations in excitatory and Vip populations, suggestive of a global effect (Fig. 2e,f). However, we also found an increase in the ability to decode changes from non-changes in excitatory and Vip population activity during novel image sessions (Figure 2g), indicating that stimulus novelty may enhance some aspects of contextual novelty signals.

Previous studies showed that Sst neurons are necessary for adaptation of excitatory responses to repeated stimuli in auditory cortex (Natan et al., 2015) and contribute to deviance detection in visual cortex (Hamm and Yuste, 2016). In our study, Sst population activity remained elevated during repeated stimulus presentations (Fig. 2b-d), consistent with a role in providing stimulus specific inhibition to support adaptation in excitatory populations, particularly when stimuli are familiar.

A third major subclass of inhibitory cells, parvalbumin-expressing interneurons (PV) which were not measured in this study, could also contribute to absolute and contextual novelty signals (Hayden et al., 2021; Kaplan et al., 2016).

Vip omission signals are reduced by novelty

Stimulus omissions are another form of contextual novelty in which expectation is violated, *i.e.,* a predicted input does not occur (Braga and Schönwiesner, 2022). In sessions with familiar stimuli, we found that Vip activity showed ramping behavior during the inter-stimulus interval and continued to ramp up during unexpected omissions. Although this omission-related ramping was dramatically reduced during the first novel session (Fig. 2j), it rapidly re-emerged during subsequent exposure to the novel images (Fig. 2k), indicating that Vip omission ramping activity is experience-dependent. Excitatory population activity also showed omission related signals in the familiar sessions, but not in the novel or novel+ sessions.

Other studies have documented experience-dependent omission responses in a population of excitatory neurons in the visual cortex during virtual navigation (Fiser et al., 2016). Ramping activity in the visual cortex has also been associated with reward anticipation (Chubykin et al., 2013; Poort et al., 2015; Shuler and Bear, 2006). Vip inter-stimulus ramping could relate to reward expectation during change detection behavior; however, we found that Vip omission ramping was not reward-dependent, as it also occurred in passive sessions with familiar stimuli (Extended Data Fig 7a,b). Contrast suppression is another possible explanation for omission related ramping activity, and Vip cells, as well as L2/3 excitatory cells, are known to be suppressed by high

contrast stimuli (Millman et al., 2020). Whether suppressed by contrast responses in visual cortex are also experience dependent is an open question.

In the frontal cortex, delay period activity of Vip neurons was found to support working memory and task performance (Kamigaki and Dan, 2017). Vip ramping in visual cortex could serve as a working memory signal to facilitate detection of image change for familiar stimuli. However, we found that Vip population activity during omissions did not carry information about stimulus identity (Fig. 2n), as would be expected for a working memory signal.

Instead, ramping activity in Vip cells could represent more general information about task timing (Rabinovich et al., 2022), temporal expectation (Nobre et al., 2007), prediction error (Orlova et al., 2021; Wacongne et al., 2011), salience (Itti and Koch, 2001), stimulus contrast (Millman et al., 2020) or a combination of these (Braga and Schönwiesner, 2022).

Novelty influences behavioral and task coding in visual cortex

Many studies have demonstrated coding for a variety of cognitive and behavioral signals in the mouse visual cortex (Dipoppa et al., 2018; Liu et al., 2015; Minderer et al., 2019; Musall et al., 2019; Poort et al., 2015; Rabinovich et al., 2022; Saleem et al., 2013; Steinmetz et al., 2019; Stringer et al., 2019), and these signals can be enhanced by learning and experience (Khan et al., 2018; Makino and Komiyama, 2015; Poort et al., 2015; Shuler and Bear, 2006). We found that stimulus novelty can modify encoding of behavioral (running, licking, pupil diameter) and task (hit, miss) variables in visual cortical neurons, in addition to stimulus and omission related information (Figure 3). The coordinated increase in sensory, behavioral, and task coding may serve to link novel stimuli to behavioral state and trial outcomes. Indeed, Vip-mediated disinhibition gates plasticity (Krabbe et al., 2019; Letzkus et al., 2015; Williams and Holtmaat, 2019; Zhang et al., 2014) and plays a role in behavioral modulation in visual cortex (Kuchibhotla et al., 2017; Pakan et al., 2018). Our results demonstrate that novelty can regulate this process.

Interestingly, the behavioral performance of our 66 animals was consistent across familiar and novel stimuli (Figure 1f, g; Extended Data Fig. 2). It is possible that the ability of mice to generalize their behavior performance to novel stimuli is enabled by enhanced coding for stimulus, behavioral and task-related information with novelty in Vip and Excitatory populations (Fig. 3e,f). One study found that novelty increased the cognitive flexibility of mice that were tasked with learning new behavioral contingencies (Park et al., 2021). Novelty can also increase attention and arousal (Schomaker and Meeter, 2015), driving a more robust encoding of novel stimuli into working memory (Mayer et al., 2011). Importantly, our data show that the influence of novelty on behavioral and cognitive representations is not only the provenance of higher-order association areas, but occurs as early as the primary visual cortex.

Functional diversity is revealed by novelty

Our clustering analysis (Figure 4) revealed diverse subpopulations of cells with distinct novelty-dependent patterns of coding. We identified 10 unique Vip clusters, 10 excitatory clusters, and 5 Sst clusters. These clusters differed in the direction and timescale of novelty-induced changes. For example, different excitatory clusters encoded novel images in either a transient (specific to the first novel session; excitatory cluster 2) or persistent (present in both novel and novel+ re-exposure; excitatory cluster 3) way. We also found Sst clusters that were transiently suppressed by novelty (Sst cluster 2) or remained suppressed across multiple sessions with novel images (Sst cluster 1). These separate populations may serve distinct computational roles. For instance, transient changes in coding could be involved in gating plasticity at specific time points, whereas those cells with more persistent changes may have a role in maintaining stimulus representations.

Vip inhibitory neurons were most diverse in their coding properties, both across features and with novelty. At the population level, Vip cells

dramatically altered their coding from omissions in familiar sessions to image coding in novel sessions. Clustering showed these changes largely resulted from the contributions separate clusters of Vip cells (Fig.4d, Extended Data Fig. 16). We also found that distinct Vip clusters differentially coded for behavioral features during familiar or novel sessions. A study in cingulate cortex also found a high degree of heterogeneity in Vip representations across behavioral contexts and time (Johnson et al., 2022), suggesting that functional flexibility may be a general property of Vip cells across the brain.

Novelty enhancement was most prevalent in excitatory cells in superficial layers (excitatory cluster 2), while familiarity enhancement was observed in excitatory clusters that were biased to deeper layers (excitatory cluster 5). Novelty processing has been linked to predictive coding (Quent et al., 2021; Reichardt et al., 2020). Prediction error signals have been identified predominantly in layer 2/3 of visual cortex (Homann et al., 2022; Keller and Mrsic-Flogel, 2018; Muzzu and Saleem, 2021) while deeper layers have been associated with memory signals (Doron et al., 2019; Murayama et al., 2009), consistent with our findings of novelty enhancement in superficial layers and familiarity coding in deeper layers.

Representational drift (Aitken et al., 2022; Deitch et al., 2021; Marks and Goard, 2021; Rule et al., 2019; Schoonover et al., 2021) could also contribute to differences in coding across days. However, several control analyses comparing cluster sizes to shuffled conditions (Extended Data Fig. 17), as well as clustering on multiple familiar sessions (Extended Data Fig 18) demonstrate that novelty is a key driver of the functional diversity we observe.

The functional clusters that we identified within each broader class may correspond to different transcriptomic cell types (Tasic et al., 2018; Yao et al., 2021; Zeng, 2022; Zeng and Sanes, 2017). Future studies using methods such as in situ spatial transcriptomics performed on previously imaged cortical neurons can directly link the functional clusters we observe to more resolved transcriptomic cell types (Bugeon et al., 2022; Condylis et al., 2022; Moffitt et al., 2018). In addition, the functional clusters we identified may map onto diverse connectivity motifs beyond the Vip-Sst disinhibitory circuit (Campagnola et al., 2022; Cottam et al., 2013; Garcia-Junco-Clemente et al., 2019; Pfeffer et al., 2013; Pi et al., 2013; Zhou et al., 2017). Mapping the cell type specific inputs to distinct transcriptomic populations and determining the connectivity between different functional clusters will be essential to elucidate the source and mechanism of novelty processing and functional diversity.

Acknowledgments

We thank the Allen Institute founder, Paul G. Allen, for his vision, encouragement, and support. We thank the Allen Institute Laboratory Animal Services team for assistance with tissue collection, the Imaging team for assistance with imaging histology sections, the Transgenic Colony Management team for generating and maintaining the mouse lines, and the Animal Care Team for taking care of our mice.

References

Aitken, K., Garrett, M., Olsen, S., Mihalas, S., 2022. The geometry of representational drift in natural and artificial neural networks. PLOS Comput. Biol. 18, e1010716. https://doi.org/10.1371/journal.pcbi.1010716

Arriaga, M., Han, E.B., 2019. Structured inhibitory activity dynamics in new virtual environments. eLife 8, e47611. https://doi.org/10.7554/eLife.47611

Aton, S.J., Suresh, A., Broussard, C., Frank, M.G., 2014. Sleep Promotes Cortical Response Potentiation Following Visual Experience. Sleep 37, 1163–1170. https://doi.org/10.5665/sleep.3830

Braga, A., Schönwiesner, M., 2022. Neural Substrates and Models of Omission Responses and Predictive Processes. Front. Neural Circuits 16.

Bugeon, S., Duffield, J., Dipoppa, M., Ritoux, A., Prankerd, I., Nicoloutsopoulos, D., Orme, D., Shinn, M., Peng, H., Forrest, H., Viduolyte, A., Reddy, C.B., Isogai, Y., Carandini, M., Harris, K.D., 2022. A transcriptomic axis predicts state modulation of cortical interneurons. Nature 607, 330–338. https://doi.org/10.1038/s41586-022-04915-7

Bunzeck, N., Düzel, E., 2006. Absolute Coding of Stimulus Novelty in the Human Substantia Nigra/VTA. Neuron

51, 369–379. https://doi.org/10.1016/j.neuron.2006.06.021

Bunzeck, N., Thiel, C., 2016. Neurochemical modulation of repetition suppression and novelty signals in the human brain. Cortex 80, 161–173. https://doi.org/10.1016/j.cortex.2015.10.013

Campagnola, L., Seeman, S.C., Chartrand, T., Kim, L., Hoggarth, A., Gamlin, C., Ito, S., Trinh, J., Davoudian, P., Radaelli, C., Kim, M.-H., Hage, T., Braun, T., Alfiler, L., Andrade, J., Bohn, P., Dalley, R., Henry, A., Kebede, S., Mukora, A., Sandman, D., Williams, G., Larsen, R., Teeter, C., Daigle, T.L., Berry, K., Dotson, N., Enstrom, R., Gorham, M., Hupp, M., Dingman Lee, S., Ngo, K., Nicovich, P.R., Potekhina, L., Ransford, S., Gary, A., Goldy, J., McMillen, D., Pham, T., Tieu, M., Siverts, L., Walker, M., Farrell, C., Schroedter, M., Slaughterbeck, C., Cobb, C., Ellenbogen, R., Gwinn, R.P., Keene, C.D., Ko, A.L., Ojemann, J.G., Silbergeld, D.L., Carey, D., Casper, T., Crichton, K., Clark, M., Dee, N., Ellingwood, L., Gloe, J., Kroll, M., Sulc, J., Tung, H., Wadhwani, K., Brouner, K., Egdorf, T., Maxwell, M., McGraw, M., Pom, C.A., Ruiz, A., Bomben, J., Feng, D., Hejazinia, N., Shi, S., Szafer, A., Wakeman, W., Phillips, J., Bernard, A., Esposito, L., D'Orazi, F.D., Sunkin, S., Smith, K., Tasic, B., Arkhipov, A., Sorensen, S., Lein, E., Koch, C., Murphy, G., Zeng, H., Jarsky, T., 2022. Local connectivity and synaptic dynamics in mouse and human neocortex. Science 375, eabj5861. https://doi.org/10.1126/science.abj5861

Chubykin, A.A., Roach, E.B., Bear, M.F., Shuler, M.G.H., 2013. A Cholinergic Mechanism for Reward Timing within Primary Visual Cortex. Neuron 77, 723–735. https://doi.org/10.1016/j.neuron.2012.12.039

Clopath, C., Vogels, T.P., Froemke, R.C., Sprekeler, H., 2016. Receptive field formation by interacting excitatory and inhibitory synaptic plasticity. https://doi.org/10.1101/066589

Condylis, C., Ghanbari, A., Manjrekar, N., Bistrong, K., Yao, S., Yao, Z., Nguyen, T.N., Zeng, H., Tasic, B., Chen, J.L., 2022. Dense functional and molecular readout of a circuit hub in sensory cortex. Science 375, eabl5981. https://doi.org/10.1126/science.abl5981

Cottam, J.C.H., Smith, S.L., Häusser, M., 2013. Target-Specific Effects of Somatostatin-Expressing Interneurons on Neocortical Visual Processing. J. Neurosci. 33, 19567–19578. https://doi.org/10.1523/JNEUROSCI.2624-13.2013

Courchesne, E., Hillyard, S.A., Galambos, R., 1975. Stimulus novelty, task relevance and the visual evoked potential in man. Electroencephalogr. Clin. Neurophysiol. 39, 131–143. https://doi.org/10.1016/0013-4694(75)90003-6

Daffner, K.R., Mesulam, M.M., Scinto, L.F.M., Calvo, V., Faust, R., Holcomb, P.J., 2000. An electrophysiological index of stimulus unfamiliarity. Psychophysiology 37, 737–747. https://doi.org/10.1111/1469-8986.3760737

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., Roll, K., Garrett, M., Keenan, T., Kuan, L., Mihalas, S., Olsen, S., Thompson, C., Wakeman, W., Waters, J., Williams, D., Barber, C., Berbesque, N., Blanchard, B., Bowles, N., Caldejon, S.D., Casal, L., Cho, A., Cross, S., Dang, C., Dolbeare, T., Edwards, M., Galbraith, J., Gaudreault, N., Gilbert, T.L., Griffin, F., Hargrave, P., Howard, R., Huang, L., Jewell, S., Keller, N., Knoblich, U., Larkin, J.D., Larsen, R., Lau, C., Lee, E., Lee, F., Leon, A., Li, L., Long, F., Luviano, J., Mace, K., Nguyen, T., Perkins, J., Robertson, M., Seid, S., Shea-Brown, E., Shi, J., Sjoquist, N., Slaughterbeck, C., Sullivan, D., Valenza, R., White, C., Williford, A., Witten, D.M., Zhuang, J., Zeng, H., Farrell, C., Ng, L., Bernard, A., Phillips, J.W., Reid, R.C., Koch, C., 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

Deitch, D., Rubin, A., Ziv, Y., 2021. Representational drift in the mouse visual cortex. Curr. Biol. 31, 4327-4339.e6. https://doi.org/10.1016/j.cub.2021.07.062

Dipoppa, M., Ranson, A., Krumin, M., Pachitariu, M., Carandini, M., Harris, K.D., 2018. Vision and Locomotion Shape the Interactions between Neuron Types in Mouse Visual Cortex. Neuron 98, 602-615.e8. https://doi.org/10.1016/j.neuron.2018.03.037

Engelhard, B., Finkelstein, J., Cox, J., Fleming, W., Jang, H.J., Ornelas, S., Koay, S.A., Thiberge, S.Y., Daw, N.D., Tank, D.W., Witten, I.B., 2019. Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. Nature 570, 509–513. https://doi.org/10.1038/s41586-019-1261-9

Fiser, A., Mahringer, D., Oyibo, H.K., Petersen, A.V., Leinweber, M., Keller, G.B., 2016. Experience-dependent spatial expectations in mouse visual cortex. Nat. Neurosci. 19, 1658–1664. https://doi.org/10.1038/nn.4385

Fu, Y., Tucciarone, J.M., Espinosa, J.S., Sheng, N., Darcy, D.P., Nicoll, R.A., Huang, Z.J., 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

Garcia del Molino, L.C., Yang, G.R., Mejias, J.F., Wang, X.-J., 2017. Paradoxical response reversal of top-down modulation in cortical circuits with three interneuron types. eLife 6, e29742. https://doi.org/10.7554/eLife.29742

Garcia-Junco-Clemente, P., Tring, E., Ringach, D.L., Trachtenberg, J.T., 2019. State-Dependent Subnetworks of Parvalbumin-Expressing Interneurons

in Neocortex. Cell Rep. 26, 2282-2288.e3. https://doi.org/10.1016/j.celrep.2019.02.005

Garrett, M., Manavi, S., Roll, K., Ollerenshaw, D.R., Groblewski, P.A., Ponvert, N.D., Kiggins, J.T., Casal, L., Mace, K., Williford, A., Leon, A., Jia, X., Ledochowitsch, P., Buice, M.A., Wakeman, W., Mihalas, S., Olsen, S.R., 2020. Experience shapes activity dynamics and stimulus coding of VIP inhibitory cells. eLife 9, e50340. https://doi.org/10.7554/eLife.50340

Garrido, M.I., Kilner, J.M., Stephan, K.E., Friston, K.J., 2009. The mismatch negativity: A review of underlying mechanisms. Clin. Neurophysiol. 120, 453–463. https://doi.org/10.1016/j.clinph.2008.11.029

Gouwens, N.W., Sorensen, S.A., Baftizadeh, F., Budzillo, A., Lee, B.R., Jarsky, T., Alfiler, L., Baker, K., Barkan, E., Berry, K., Bertagnolli, D., Bickley, K., Bomben, J., Braun, T., Brouner, K., Casper, T., Crichton, K., Daigle, T.L., Dalley, R., de Frates, R.A., Dee, N., Desta, T., Lee, S.D., Dotson, N., Egdorf, T., Ellingwood, L., Enstrom, R., Esposito, L., Farrell, C., Feng, D., Fong, O., Gala, R., Gamlin, C., Gary, A., Glandon, A., Goldy, J., Gorham, M., Graybuck, L., Gu, H., Hadley, K., Hawrylycz, M.J., Henry, A.M., Hill, D., Hupp, M., Kebede, S., Kim, T.K., Kim, L., Kroll, M., Lee, C., Link, K.E., Mallory, M., Mann, R., Maxwell, M., McGraw, M., McMillen, D., Mukora, A., Ng, Lindsay, Ng, Lydia, Ngo, K., Nicovich, P.R., Oldre, A., Park, D., Peng, H., Penn, O., Pham, T., Pom, A., Popović, Z., Potekhina, L., Rajanbabu, R., Ransford, S., Reid, D., Rimorin, C., Robertson, M., Ronellenfitch, K., Ruiz, A., Sandman, D., Smith, K., Sulc, J., Sunkin, S.M., Szafer, A., Tieu, M., Torkelson, A., Trinh, J., Tung, H., Wakeman, W., Ward, K., Williams, G., Zhou, Z., Ting, J.T., Arkhipov, A., Sümbül, U., Lein, E.S., Koch, C., Yao, Z., Tasic, B., Berg, J., Murphy, G.J., Zeng, H., 2020. Integrated Morphoelectric and Transcriptomic Classification of Cortical GABAergic Cells. Cell 183, 935-953.e19. https://doi.org/10.1016/j.cell.2020.09.057

Gouwens, N.W., Sorensen, S.A., Berg, J., Lee, C., Jarsky, T., Ting, J., Sunkin, S.M., Feng, D., Anastassiou, C.A., Barkan, E., Bickley, K., Blesie, N., Braun, T., Brouner, K., Budzillo, A., Caldejon, S., Casper, T., Castelli, D., Chong, P., Crichton, K., Cuhaciyan, C., Daigle, T.L., Dalley, R., Dee, N., Desta, T., Ding, S.-L., Dingman, S., Doperalski, A., Dotson, N., Egdorf, T., Fisher, M., de Frates, R.A., Garren, E., Garwood, M., Gary, A., Gaudreault, N., Godfrey, K., Gorham, M., Gu, H., Habel, C., Hadley, K., Harrington, J., Harris, J.A., Henry, A., Hill, D., Josephsen, S., Kebede, S., Kim, L., Kroll, M., Lee, B., Lemon, T., Link, K.E., Liu, X., Long, B., Mann, R., McGraw, M., Mihalas, S., Mukora, A., Murphy, G.J., Ng, Lindsay, Ngo, K., Nguyen, T.N., Nicovich, P.R., Oldre, A., Park, D., Parry, S., Perkins, J., Potekhina, L., Reid, D., Robertson, M., Sandman, D., Schroedter, M., Slaughterbeck, C., Soler-Llavina, G., Sulc, J., Szafer, A., Tasic, B., Taskin, N., Teeter, C., Thatra, N., Tung, H., Wakeman, W., Williams, G., Young, R., Zhou, Z., Farrell, C., Peng, H., Hawrylycz, M.J., Lein, E., Ng, Lydia, Arkhipov, A., Bernard, A., Phillips, J.W., Zeng, H., Koch, C., 2019. Classification of electrophysiological and morphological neuron types in the mouse visual cortex. Nat. Neurosci. 22, 1182–1195. https://doi.org/10.1038/s41593-019-0417-0

Groblewski, P.A., Ollerenshaw, D.R., Kiggins, J.T., Garrett, M.E., Mochizuki, C., Casal, L., Cross, S., Mace, K., Swapp, J., Manavi, S., Williams, D., Mihalas, S., Olsen, S.R., 2020. Characterization of Learning, Motivation, and Visual Perception in Five Transgenic Mouse Lines Expressing GCaMP in Distinct Cell Populations. Front. Behav. Neurosci. 14.

Hamm, J.P., 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

Hawco, C., Lepage, M., 2014. Overlapping patterns of neural activity for different forms of novelty in fMRI. Front. Hum. Neurosci. 8.

Hayden, D.J., Montgomery, D.P., Cooke, S.F., Bear, M.F., 2021. Visual Recognition Is Heralded by Shifts in Local Field Potential Oscillations and Inhibitory Networks in Primary Visual Cortex. J. Neurosci. 41, 6257–6272. https://doi.org/10.1523/JNEUROSCI.0391-21.2021

Hertäg, L., Sprekeler, H., 2019. Amplifying the redistribution of somato-dendritic inhibition by the interplay of three interneuron types. PLOS Comput. Biol. 15, e1006999. https://doi.org/10.1371/journal.pcbi.1006999

Homann, J., Koay, S.A., Chen, K.S., Tank, D.W., Berry, M.J., 2022. Novel stimuli evoke excess activity in the mouse primary visual cortex. Proc. Natl. Acad. Sci. 119, e2108882119. https://doi.org/10.1073/pnas.2108882119

Hu, B., Garrett, M.E., Groblewski, P.A., Ollerenshaw, D.R., Shang, J., Roll, K., Manavi, S., Koch, C., Olsen, S.R., 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

Itti, L., Baldi, P., 2009. Bayesian surprise attracts human attention. Vision Res., Visual Attention: Psychophysics, electrophysiology and neuroimaging 49, 1295–1306. https://doi.org/10.1016/j.visres.2008.09.007

Itti, L., Koch, C., 2001. Computational modelling of visual attention. Nat. Rev. Neurosci. 2, 194–203. https://doi.org/10.1038/35058500

Jaegle, A., Mehrpour, V., Rust, N., 2019. Visual novelty, curiosity, and intrinsic reward in machine learning and the brain. Curr. Opin. Neurobiol. 58, 167–174. https://doi.org/10.1016/j.conb.2019.08.004

Johnson, C., Kretsge, L.N., Yen, W.W., Sriram, B., O'Connor, A., Liu, R.S., Jimenez, J.C., Phadke, R.A., Wingfield, K.K., Yeung, C., Jinadasa, T.J., Nguyen, T.P.H., Cho, E.S., Fuchs, E., Spevack, E.D., Velasco, B.E., Hausmann, F.S., Fournier, L.A., Brack, A., Melzer, S., Cruz-Martín, A., 2022. Highly unstable heterogeneous representations in VIP interneurons of the anterior cingulate cortex. Mol. Psychiatry 27, 2602–2618. https://doi.org/10.1038/s41380-022-01485-y

Kafkas, A., Montaldi, D., 2018. How do memory systems detect and respond to novelty? Neurosci. Lett. 680, 60–68. https://doi.org/10.1016/j.neulet.2018.01.053

Kamigaki, T., Dan, Y., 2017. Delay activity of specific prefrontal interneuron subtypes modulates memory-guided behavior. Nat. Neurosci. 20, 854–863. https://doi.org/10.1038/nn.4554

Kaplan, E.S., Cooke, S.F., Komorowski, R.W., Chubykin, A.A., Thomazeau, A., Khibnik, L.A., Gavornik, J.P., Bear, M.F., 2016. Contrasting roles for parvalbumin-expressing inhibitory neurons in two forms of adult visual cortical plasticity. eLife 5, e11450. https://doi.org/10.7554/eLife.11450

Karnani, M.M., Jackson, J., Ayzenshtat, I., Sichani, A.H., Manoocheri, K., Kim, S., Yuste, R., 2016. Opening Holes in the Blanket of Inhibition: Localized Lateral Disinhibition by VIP Interneurons. J. Neurosci. 36, 3471–3480. https://doi.org/10.1523/JNEUROSCI.3646-15.2016

Kato, H.K., Gillet, S.N., Isaacson, J.S., 2015. Flexible Sensory Representations in Auditory Cortex Driven by Behavioral Relevance. Neuron 88, 1027–1039. https://doi.org/10.1016/j.neuron.2015.10.024

Keller, A.J., Dipoppa, M., Roth, M.M., Caudill, M.S., Ingrosso, A., Miller, K.D., Scanziani, M., 2020. A Disinhibitory Circuit for Contextual Modulation in Primary Visual Cortex. Neuron 108, 1181-1193.e8. https://doi.org/10.1016/j.neuron.2020.11.013

Keller, G.B., Mrsic-Flogel, T.D., 2018. Predictive Processing: A Canonical Cortical Computation. Neuron 100, 424–435. https://doi.org/10.1016/j.neuron.2018.10.003

Khan, A.G., Poort, J., Chadwick, A., Blot, A., Sahani, M., Mrsic-Flogel, T.D., Hofer, S.B., 2018. Distinct learning-induced changes in stimulus selectivity and interactions of GABAergic interneuron classes in visual cortex. Nat. Neurosci. 21, 851–859. https://doi.org/10.1038/s41593-018-0143-z

Koay, S.A., Charles, A.S., Thiberge, S.Y., Brody, C.D., Tank, D.W., 2022. Sequential and efficient neural-population coding of complex task information. Neuron 110, 328-349.e11. https://doi.org/10.1016/j.neuron.2021.10.020

Krabbe, S., Paradiso, E., d'Aquin, S., Bitterman, Y., Courtin, J., Xu, C., Yonehara, K., Markovic, M., Müller, C., Eichlisberger, T., Gründemann, J., Ferraguti, F., Lüthi, A., 2019. Adaptive disinhibitory gating by VIP interneurons permits associative learning. Nat. Neurosci. 22, 1834–1843. https://doi.org/10.1038/s41593-019-0508-y

Kuchibhotla, K.V., Gill, J.V., Lindsay, G.W., Papadoyannis, E.S., Field, R.E., Sten, T.A.H., Miller, K.D., Froemke, R.C., 2017. Parallel processing by cortical inhibition enables context-dependent behavior. Nat. Neurosci. 20, 62–71. https://doi.org/10.1038/nn.4436

Lee, S., Kruglikov, I., Huang, Z.J., Fishell, G., Rudy, B., 2013. A disinhibitory circuit mediates motor integration in the somatosensory cortex. Nat. Neurosci. 16, 1662–1670. https://doi.org/10.1038/nn.3544

Letzkus, J.J., Wolff, S.B.E., Lüthi, A., 2015. Disinhibition, a Circuit Mechanism for Associative Learning and Memory. Neuron 88, 264–276. https://doi.org/10.1016/j.neuron.2015.09.024

Letzkus, J.J., Wolff, S.B.E., Meyer, E.M.M., Tovote, P., Courtin, J., Herry, C., Lüthi, A., 2011. A disinhibitory microcircuit for associative fear learning in the auditory cortex. Nature 480, 331–335. https://doi.org/10.1038/nature10674

Li, L., Miller, E.K., Desimone, R., 1993. The representation of stimulus familiarity in anterior inferior temporal cortex. J. Neurophysiol. 69, 1918–1929. https://doi.org/10.1152/jn.1993.69.6.1918

Liu, C.-H., Coleman, J.E., Davoudi, H., Zhang, K., Hussain Shuler, M.G., 2015. Selective Activation of a Putative Reinforcement Signal Conditions Cued Interval Timing in Primary Visual Cortex. Curr. Biol. 25, 1551–1561. https://doi.org/10.1016/j.cub.2015.04.028

Makino, H., Komiyama, T., 2015. Learning enhances the relative impact of top-down processing in the visual cortex. Nat. Neurosci. 18, 1116–1122. https://doi.org/10.1038/nn.4061

Marks, T.D., Goard, M.J., 2021. Stimulus-dependent representational drift in primary visual cortex. Nat. Commun. 12, 5169. https://doi.org/10.1038/s41467-021-25436-3

Mayer, J.S., Kim, J., Park, S., 2011. Enhancing visual working memory encoding: The role of target novelty. Vis. Cogn. 19, 863–885. https://doi.org/10.1080/13506285.2011.594459

Melzer, S., Newmark, E.R., Mizuno, G.O., Hyun, M., Philson, A.C., Quiroli, E., Righetti, B., Gregory, M.R., Huang, K.W., Levasseur, J., Tian, L., Sabatini, B.L., 2021. Bombesin-like peptide recruits disinhibitory cortical circuits and enhances fear memories. Cell 184, 5622-5634.e25. https://doi.org/10.1016/j.cell.2021.09.013

Meyer, T., Rust, N.C., 2018. Single-exposure visual memory judgments are reflected in inferotemporal cortex. eLife 7, e32259. https://doi.org/10.7554/eLife.32259

Miehl, C., Gjorgjieva, J., 2022. Stability and learning in excitatory synapses by nonlinear inhibitory plasticity. PLOS Comput. Biol. 18, e1010682. https://doi.org/10.1371/journal.pcbi.1010682

Millman, D.J., Ocker, G.K., Caldejon, S., Kato, I., Larkin, J.D., Lee, E.K., Luviano, J., Nayan, C., Nguyen, T.V., North, K., Seid, S., White, C., Lecoq, J., Reid, C., Buice, M.A., de Vries, S.E., 2020. VIP interneurons in mouse primary visual cortex selectively enhance responses to weak but specific stimuli. eLife 9, e55130. https://doi.org/10.7554/eLife.55130

Minderer, M., Brown, K.D., Harvey, C.D., 2019. The spatial structure of neural encoding in mouse posterior cortex during navigation. Neuron 1–17. https://doi.org/10.1016/j.neuron.2019.01.029

Moffitt, J.R., Bambah-Mukku, D., Eichhorn, S.W., Vaughn, E., Shekhar, K., Perez, J.D., Rubinstein, N.D., Hao, J., Regev, A., Dulac, C., Zhuang, X., 2018. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. Science 362, eaau5324. https://doi.org/10.1126/science.aau5324

Montgomery, D.P., Hayden, D.J., Chaloner, F.A., Cooke, S.F., Bear, M.F., 2022. Stimulus-Selective Response Plasticity in Primary Visual Cortex: Progress and Puzzles. Front. Neural Circuits 15, 815554. https://doi.org/10.3389/fncir.2021.815554

Musall, S., Kaufman, M.T., Juavinett, A.L., Gluf, S., 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

Muzzu, T., Saleem, A.B., 2021. Feature selectivity can explain mismatch signals in mouse visual cortex. Cell Rep. 37, 109772. https://doi.org/10.1016/j.celrep.2021.109772

Natan, R.G., Briguglio, J.J., Mwilambwe-Tshilobo, L., Jones, S.I., Aizenberg, M., Goldberg, E.M., 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

Natan, R.G., Rao, W., Geffen, M.N., 2017. Cortical Interneurons Differentially Shape Frequency Tuning following Adaptation. Cell Rep. 21, 878–890. https://doi.org/10.1016/j.celrep.2017.10.012

Nelken, I., Ulanovsky, N., 2007. Mismatch negativity and stimulus-specific adaptation in animal models. J. Psychophysiol. 21, 214–223. https://doi.org/10.1027/0269-8803.21.34.214

Niell, C.M., Stryker, M.P., 2010. Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex. Neuron 65, 472–479. https://doi.org/10.1016/j.neuron.2010.01.033

Nobre, A., Correa, A., Coull, J., 2007. The hazards of time. Curr. Opin. Neurobiol., Sensory systems 17, 465–470. https://doi.org/10.1016/j.conb.2007.07.006

Orlova, N., Najafi, F., Tsyboulski, D., Seid, S., Kivikas, S., Kato, I., Griffin, F., Leon, A., Hytnen, R.D., L'Heureux, Q., North, K., Swapp, J., Nayan, C., Hancock, N., Ahmed, R., Gelfand, E., Cho, A., Mace, K., Howard, R., Casal, L., Lambert, S., Lee, E.K., Caldejon, S., Waughman, X., Williford, A., Garrett, M., Ollerenshaw, D., Olsen, S.R., Groblewski, P.A., Saggau, P., Lecoq, J., 2021. Multiplane Mesoscope reveals distinct cortical interactions following expectation violations. https://doi.org/10.1101/2020.10.06.328294

Pakan, J.M.P., Currie, S.P., Fischer, L., Rochefort, N.L., 2018. The Impact of Visual Cues, Reward, and Motor Feedback on the Representation of Behaviorally Relevant Spatial Locations in Primary Visual Cortex. Cell Rep. 24, 2521–2528. https://doi.org/10.1016/j.celrep.2018.08.010

Park, A.J., Harris, A.Z., Martyniuk, K.M., Chang, C.-Y., Abbas, A.I., Lowes, D.C., Kellendonk, C., Gogos, J.A., Gordon, J.A., 2021. Reset of hippocampal–prefrontal circuitry facilitates learning. Nature 591, 615–619. https://doi.org/10.1038/s41586-021-03272-1

Pfeffer, C.K., Xue, M., He, M., Huang, Z.J., 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

Pi, H.-J., Hangya, B., Kvitsiani, D., Sanders, J.I., Huang, Z.J., Kepecs, A., 2013. Cortical interneurons that specialize in disinhibitory control. Nature 503, 521–524. https://doi.org/10.1038/nature12676

Pillow, J.W., Shlens, J., Paninski, L., Sher, A., Litke, A.M., Chichilnisky, E.J., Simoncelli, E.P., 2008. Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature 454, 995–999. https://doi.org/10.1038/nature07140

Poort, J., Khan, A.G., Pachitariu, M., Nemri, A., Orsolic, I., Krupic, J., Bauza, M., Sahani, M., Keller, G.B., Mrsic-Flogel, T.D., Hofer, S.B., 2015. Learning Enhances Sensory and Multiple Non-sensory Representations in Primary Visual Cortex. Neuron 86, 1478–1490. https://doi.org/10.1016/j.neuron.2015.05.037

Quent, J.A., Henson, R.N., Greve, A., 2021. A predictive account of how novelty influences declarative memory. Neurobiol. Learn. Mem. 179, 107382. https://doi.org/10.1016/j.nlm.2021.107382

Rabinovich, R.J., Kato, D.D., Bruno, R.M., 2022. Learning enhances encoding of time and temporal surprise in mouse primary sensory cortex. Nat. Commun. 13, 5504. https://doi.org/10.1038/s41467-022-33141-y

Ranganath, C., Rainer, G., 2003. Neural mechanisms for detecting and remembering novel events. Nat. Rev. Neurosci. 4, 193–202. https://doi.org/10.1038/nrn1052

Reichardt, R., Polner, B., Simor, P., 2020. Novelty Manipulations, Memory Performance, and Predictive Coding: the Role of Unexpectedness. Front. Hum. Neurosci. 14.

Rule, M.E., O'Leary, T., Harvey, C.D., 2019. Causes and consequences of representational drift. Curr. Opin. Neurobiol., Computational Neuroscience 58, 141–147. https://doi.org/10.1016/j.conb.2019.08.005

Rust, N.C., Cohen, M.R., 2022. Priority coding in the visual system. Nat. Rev. Neurosci. 23, 376–388. https://doi.org/10.1038/s41583-022-00582-9

Saleem, A.B., Ayaz, A., Jeffery, K.J., Harris, K.D., Carandini, M., 2013. Integration of visual motion and locomotion in mouse visual cortex. Nat. Neurosci. 16, 1864–1869. https://doi.org/10.1038/nn.3567

Schomaker, J., 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

Schoonover, C.E., Ohashi, S.N., Axel, R., Fink, A.J.P., 2021. Representational drift in primary olfactory cortex. Nature 594, 541–546. https://doi.org/10.1038/s41586-021-03628-7

Schulz, A., Miehl, C., Berry, M.J., Gjorgjieva, J., 2021. The generation of cortical novelty responses through inhibitory plasticity. eLife 10, e65309. https://doi.org/10.7554/eLife.65309

Shuler, M.G., Bear, M.F., 2006. Reward Timing in the Primary Visual Cortex. Science 311, 1606–1609. https://doi.org/10.1126/science.1123513

Siegle, J.H., Jia, X., Durand, S., Gale, S., Bennett, C., Graddis, N., Heller, G., Ramirez, T.K., Choi, H., Luviano, J.A., Groblewski, P.A., Ahmed, R., Arkhipov, A., Bernard, A., Billeh, Y.N., Brown, D., Buice, M.A., Cain, N., Caldejon, S., Casal, L., Cho, A., Chvilicek, M., Cox, T.C., Dai, K., Denman, D.J., de Vries, S.E.J., Dietzman, R., Esposito, L., Farrell, C., Feng, D., Galbraith, J., Garrett, M., Gelfand, E.C., Hancock, N., Harris, J.A., Howard, R., Hu, B., Hytnen, R., Iyer, R., Jessett, E., Johnson, K., Kato, I., Kiggins, J., Lambert, S., Lecoq, J., Ledochowitsch, P., Lee, J.H., Leon, A., Li, Y., Liang, E., Long, F., Mace, K., Melchior, J., Millman, D., Mollenkopf, T., Nayan, C., Ng, L., Ngo, K., Nguyen, T., Nicovich, P.R., North, K., Ocker, G.K., Ollerenshaw, D., Oliver, M., Pachitariu, M., Perkins, J., Reding, M., Reid, D., Robertson, M., Ronellenfitch, K., Seid, S., Slaughterbeck, C., Stoecklin, M., Sullivan, D., Sutton, B., Swapp, J., Thompson, C., Turner, K., Wakeman, W., Whitesell, J.D., Williams, D., Williford, A., Young, R., Zeng, H., Naylor, S., Phillips, J.W., Reid, R.C., Mihalas, S., Olsen, S.R., Koch, C., 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

Sprekeler, H., 2017. Functional consequences of inhibitory plasticity: homeostasis, the excitation-inhibition balance and beyond. Curr. Opin. Neurobiol., Neurobiology of Learning and Plasticity 43, 198–203. https://doi.org/10.1016/j.conb.2017.03.014

Steinmetz, N.A., Zatka-Haas, P., Carandini, M., 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

Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C.B., Carandini, M., Harris, K.D., 2019. Spontaneous behaviors drive multidimensional, brainwide activity. Science 364, eaav7893. https://doi.org/10.1126/science.aav7893

Tasic, B., Yao, Z., Graybuck, L.T., Smith, K.A., Nguyen, T.N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M.N., Viswanathan, S., Penn, O., Bakken, T., Menon, V., Miller, J., Fong, O., Hirokawa, K.E., Lathia, K., Rimorin, C., Tieu, M., Larsen, R., Casper, T., Barkan, E., Kroll, M., Parry, S., Shapovalova, N.V., Hirschstein, D., Pendergraft, J., Sullivan, H.A., Kim, T.K., Szafer, A., Dee, N., Groblewski, P., Wickersham, I., Cetin, A., Harris, J.A., Levi, B.P., Sunkin, S.M., Madisen, L., Daigle, T.L., Looger, L., Bernard, A., Phillips, J., Lein, E., Hawrylycz, M., Svoboda, K., Jones, A.R., Koch, C., Zeng, H., 2018. Shared and distinct transcriptomic cell types across neocortical areas. Nature 563, 72–78. https://doi.org/10.1038/s41586-018-0654-5

Tremblay, R., Lee, S., Rudy, B., 2016. GABAergic Interneurons in the Neocortex: From Cellular Properties to Circuits. Neuron 91, 260–292. https://doi.org/10.1016/j.neuron.2016.06.033

Turi, G.F., Li, W.K., Chavlis, S., Pandi, I., O'Hare, J., Priestley, J.B., Grosmark, A.D., Liao, Z., Ladow, M., Zhang, J.F., Zemelman, B.V., Poirazi, P., Losonczy, A., 2019. Vasoactive Intestinal Polypeptide-Expressing Interneurons in the Hippocampus Support Goal-Oriented Spatial Learning. Neuron 101, 1150-1165.e8. https://doi.org/10.1016/j.neuron.2019.01.009

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., Dehaene, S., 2011. Evidence for a hierarchy of predictions and prediction errors in human cortex. Proc. Natl. Acad. Sci. 108, 20754–20759. https://doi.org/10.1073/pnas.1117807108

Williams, L.E., Holtmaat, A., 2019. Higher-Order Thalamocortical Inputs Gate Synaptic Long-Term Potentiation via Disinhibition. Neuron 101, 91-102.e4. https://doi.org/10.1016/j.neuron.2018.10.049

Wilmes, K.A., Clopath, C., 2019. Inhibitory microcircuits for top-down plasticity of sensory representations. Nat. Commun. 10, 5055. https://doi.org/10.1038/s41467-019-12972-2

Xiang, J.-Z., Brown, M.W., 1998. Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe. Neuropharmacology 37, 657–676. https://doi.org/10.1016/S0028-3908(98)00030-6

Yao, Z., van Velthoven, C.T.J., Nguyen, T.N., Goldy, J., Sedeno-Cortes, A.E., Baftizadeh, F., Bertagnolli, D., Casper, T., Chiang, M., Crichton, K., Ding, S.-L., Fong, O., Garren, E., Glandon, A., Gouwens, N.W., Gray, J., Graybuck, L.T., Hawrylycz, M.J., Hirschstein, D., Kroll, M., Lathia, K., Lee, C., Levi, B., McMillen, D., Mok, S., Pham, T., Ren, Q., Rimorin, C., Shapovalova, N., Sulc, J., Sunkin, S.M., Tieu, M., Torkelson, A., Tung, H., Ward, K., Dee, N., Smith, K.A., Tasic, B., Zeng, H., 2021. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. Cell 184, 3222-3241.e26. https://doi.org/10.1016/j.cell.2021.04.021

Zeng, H., 2022. What is a cell type and how to define it? Cell 185, 2739–2755. https://doi.org/10.1016/j.cell.2022.06.031

Zeng, H., Sanes, J.R., 2017. Neuronal cell-type classification: challenges, opportunities and the path forward. Nat. Rev. Neurosci. 18, 530–546. https://doi.org/10.1038/nrn.2017.85

Zenke, F., Gerstner, W., Ganguli, S., 2017. The temporal paradox of Hebbian learning and homeostatic plasticity. Curr. Opin. Neurobiol., Neurobiology of Learning and Plasticity 43, 166–176. https://doi.org/10.1016/j.conb.2017.03.015

Zhang, S., Xu, M., Kamigaki, T., Hoang Do, J.P., Chang, W.-C., Jenvay, S., Miyamichi, K., Luo, L., Dan, Y., 2014. Long-range and local circuits for top-down modulation of visual cortex processing. Science 345, 660–665. https://doi.org/10.1126/science.1254126

Zhou, X., Rickmann, M., Hafner, G., Staiger, J.F., 2017. Subcellular Targeting of VIP Boutons in Mouse Barrel Cortex is Layer-Dependent and not Restricted to Interneurons. Cereb. Cortex 27, 5353–5368. https://doi.org/10.1093/cercor/bhx220

Materials and Methods

Mice

All experiments and procedures were performed in accordance with protocols approved by the Allen Institute Animal Care and Use Committee. Male and female transgenic mice expressing GCaMP6 in various Cre-defined cell populations were used in these experiments[1]. The three genotypes used in this study were *Slc17a7*: Slc17a7-IRES2-Cre;Camk2a-tTA;Ai93(TITL-GCaMP6f), n=41; *Sst*: Sst-IRES-Cre;Ai148(TIT2L-GC6f-ICL-tTA2), n=19; *Vip*: Vip-IRES-Cre;Ai148(TIT2L-GC6f-ICL-tTA2), n=22. Prior to surgery mice were singly-housed and maintained on a reverse 12-hour light cycle (off at 9am, on at 9pm); all experiments were performed during the dark cycle.

Surgery

All mice received a headpost and cranial window surgery as previously described[2,3]. Briefly, surgery was performed on healthy mice that ranged in age from 5-12 weeks. Mice were deeply anesthetized with isoflurane prior to removing skin and exposing the skull. A custom titanium headframe was cemented to the skull and a circular piece of skull 5 mm in diameter was removed, durotomy performed, and a glass coverslip stack was cemented in place. Mice were given 2 weeks to recover from surgery before intrinsic signal imaging.

Intrinsic Signal Imaging

Intrinsic signal imaging (ISI) was used to measure the hemodynamic response of the cortex to visual stimulation across the entire field of view. As previously described, ISI was used to delineate functionally defined visual area boundaries for targeting of 2P experiments[3].

Data Acquisition

Mice were lightly anesthetized before imaging sessions began with a vasculature image acquired under green light. Next the imaging plane was defocused and the hemodynamic response to a visual stimulus was imaged under red light. The stimulus consisted of an alternating checkerboard pattern (20º wide bar, 25º square size) moving 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[4].

Data Processing

A minimum of three trials were averaged to produce altitude and azimuth phase maps, calculated from the discrete Fourier transform of each pixel. 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[5].

To provide a reliable map for subsequent targeting of 2-photon calcium imaging experiments, a consistent anatomical coordinate corresponding to the center of V1 (which maps to center of the retina) was used to realign the maps. A map of eccentricity from the V1 centroid was produced by shifting the origin of the map of visual eccentricity to the coordinates at the V1 centroid, thereby representing the retinotopic gradients relative to this point. A representation of the corresponding retinotopic location is present in nearly all higher visual areas (HVAs). Using these modified V1 and HVA targets for optical physiology experiments ensured that recorded neurons represent a consistent region on the retina, approximately at the center of the right visual hemifield.

Behavior Training

Water restriction and habituation

Throughout training mice were water-restricted to motivate learning and performance of the behavioral task [6]. Mice had access to water only during behavioral training sessions or when provided by a technician on non-training days. 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. The first day of behavior training began after 10 days of water restriction. Mice were trained 5 days per week (Monday-Friday) and were allowed to earn unlimited water during the daily 1-hour sessions; supplements were provided in a home cage water dish if the earned volume fell below 1.0mL

and/or body weight fell under 80-85% of initial baseline weight. On non-training days mice were weighed and received water provision to reach their target weight, but never less than 1.0 mL per day.

Apparatus

Mice were trained in custom-designed, sound-attenuating behavior enclosures equipped with a 24" gamma-corrected LCD monitor (ASUS, #PA248Q). Mice were head-fixed on a behavior stage with 6.5" running wheel tilted upwards by 10-15 degrees. The center of the visual monitor was placed 15 cm from the eye and visual stimuli were spherically warped to account for the variable distance from the eye toward the periphery of the monitor. Water rewards were delivered using a solenoid (NI Research, #161K011) to deliver a calibrated volume of fluid through a blunted, 17g hypodermic needle (Hamilton) positioned approximately 2-3 mm away from the animal's mouth using a custom-made 3-axis motorized stage.

Change detection task

Overview: The change detection task and automated procedure for training this task have previously been described in detail[7,8]. Briefly, mice were trained using a behavioral program implementing a go/no-go change detection task schematized in Figure 1. Mice were presented with a continuous stream of flashed visual stimuli (250ms stimuli interleaved with 500ms 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 (150-750ms after stimulus change) a water reward was delivered. A 'grace period' of 3 seconds occurred after each change, during which no additional image changes could occur, thereby providing time for reward consumption before the next trial was initiated. At the start of each trial, trial type was determined (87.5% "GO" or 12.5 % "CATCH") and a change time was drawn from a geometric distribution ranging from 4 to 12 flashes (3 to 9 seconds). If the mouse licked prior to the stimulus change the trial was reset ("aborted") up to 5 times before a new trial and change time was re-drawn. Thus, continuous licking would result in no opportunity to obtain reward. No other punishment was delivered for false alarms or aberrant licking.

Automated Training: Mice were trained using an automated training procedure that consisted of 4 stages of increasing complexity. On Day 1 of the automated training protocol mice received a short, 15-min "open loop" session during which non-contingent water rewards were delivered coincident with 90° changes in orientation of a full-field, static square-wave grating (Stage 0). This session was intended to 1) introduce the mouse to the fluid delivery system and, 2) provide the technician an opportunity to identify the optimal lick spout position for each mouse. Each session thereafter was run in "closed loop", and progressed through 3 phases of the operant task: Stage 1: static, full-field square wave gratings (oriented at 0° and 90°, with the black/white transition always centered on the screen and the phase chosen randomly on every trial), Stage 2: flashed, full-field square-wave gratings (0° and 90°, with phase as described in 1), and Stage 3: flashed full-field natural scenes (8 "Familiar" natural images shown in Figure 1). In a subset of mice the image sets labeled "Familiar" and "Novel" were switched in order to balance the experimental design.

Progression through training stages: Starting with Stage 1, the advancement criteria required mice to achieve a session maximum performance of at least d-prime=2 (calculated over a rolling 100 trial window without trial count correction) during two of the last 3 sessions. The fastest progression from Stage 1 to Stage 3 was 4 training days. Once mice exhibited consistent performance (d-prime >/= 1 over 3 consecutive sessions) they became eligible to transition to the 2-photon calcium imaging stage of the experiment.

2-photon Calcium Imaging

Calcium imaging data was acquired using two microscope platforms, each of which was built around our custom-designed Allen Brain Observatory behavior platform and common mouse-to-screen geometry as previously described[3,9,10].

Single Plane Imaging Apparatus

Single-plane calcium imaging was performed using a 2-photon microscope (Scientifica Vivoscope), as used by de Vries et al., 2020 and Garrett et al., 2019. Scientifica microscope design is based on 8 kHz resonant scanning mirror and employs conventional hardware (photomultiplier tubes, Hamamatsu; transimpedance amplifier, Femto; DAQ hardware, National Instruments; 16x imaging objective, Nikon) to collect emitted fluorescence and form an image on the

acquisition computer. The microscope is controlled by the company's proprietary LABView software SciScan. Laser excitation was provided by a Ti:Sapphire laser (Chameleon Vision, Coherent) at 910 nm. Pulse dispersion compensation was set at ~10,000 fs^2. Movies were recorded at 30Hz using resonant scanners over a 400 µm field of view.

Multi-Plane Imaging Apparatus

Multi-plane calcium imaging was performed using a Dual-Beam Mesoscope (Multiscope) that allowed us to double imaging throughput[11]. The second laser beam was introduced to the original 2P-RAM system[12], packaged into a compact opto-mechanical add-on unit, and optimized for ease of alignment. The dual-beam modification consisted of a 1) delay line used to split the original laser beam into 2 and delay one of the beams by half a period of the excitation laser; 2) a secondary z-scanner and 3) custom-built demultiplexing unit. The delay line allowed for temporal encoding of the excitation beam and further demultiplexing of the detected fluorescence based on the arrival time at the photodetector. A secondary z-scanner allowed us to send two beams to the two focal planes located along Z axis. Laser excitation was provided by a Ti:Sapphire ultrafast laser (Chameleon Ultra II, Coherent). Pulse dispersion compensation was optimized for GCaMP6 using a custom-built external pulse compensation module based on a single-prism four-path design (Sofroniew et al., 2016; Akturk, et al., 2006). The Multiscope was controlled with customized ScanImage software (VidrioTech) as well as an in-house developed Workflow Sequencing Engine. Like conventional 2-photon microscopes, the emitted fluorescence was detected using a single photomultiplier tube, and a custom analog demultiplexing circuit was used to separate fluorescence from two planes[11]. This was achieved by multiplying the PMT signal with two complimentary square waveforms, where each waveform corresponded to the temporal window during which fluorescence received by the PMT consisted of the signal from one of the focal planes. The duration of the integration window was 6.25 ns (half a period of the excitation laser's pulse train), which is not enough to fully capture the decay of fluorescence, which resulted in the tail of fluorescence leaking to the opposite integration window and causing inter-plane crosstalk (~10% remaining crosstalk on average). We used an ICA-based demixing algorithm to further clean up the data acquired in simultaneously imaged focal planes. National Instruments data acquisition hardware (PXI chassis, PXIe6363 DAQ boards) was used to control the microscope, form, and record the image.

Data Acquisition

Daily preparations for the 2-photon imaging experiments were conducted under ambient red light to maintain the reversed day-night cycle, and imaging itself was performed in the dark. Mice were head-fixed in a behavior stage identical to that used during behavior training. A water immersion objective was used for single-plane experiments on the Scientifica microscope whereas a water-based ultrasonic gel was used as immersion medium for the multi-plane experiments on the Multiscope. Two-photon movies (512x512 pixels, 31 Hz for single plane and 512x512 pixels, 11 Hz for each plane in multi-plane experiments), eye tracking (30 Hz), and behavior (30 Hz) were recorded and continuously monitored. Recording sessions were ~1 hour long, but could be interrupted if any of the following was observed: 1) mouse stress as shown by excessive secretion around the eye, nose bulge, and/or abnormal posture; 2) excessive pixel saturation (>1000 pixels) as reported in a continuously updated histogram; 3) loss of baseline intensity caused by bleaching and/or loss of immersion water in excess of 20%; 4) hardware failures causing a loss of data integrity. At the end of each experimental session, a z-stack of images (+/- 30 µm around imaging site, 0.75 µm step) was collected to evaluate cortical anatomy as well as z-motion during acquisition. In addition, a full-depth cortical z stack (~800 µm total depth, 5 µm step) was collected to document the imaging site location.

Quality Control

Each experimental session was analyzed for data integrity based on a broad range of operational parameters. A comprehensive report was automatically generated to track data trends, animal behavior, experimental failures and errors. To minimize bias, QC reports were reviewed by rig operators other than the operator who performed the session.

Assessment of the following quality metrics was performed after each imaging session. Failure to meet any of these criteria resulted in the session being retaken on a subsequent day.

1. Image Saturation: Initial movie frames were assessed to confirm that photonic saturation did not exceed 1000 pixels and the full dynamic range of the recording system was adequately covered.
2. Photobleaching: Epochs of fluorescence at the beginning and end of a session were compared to ensure that baseline fluorescence did not drop greater than 20%.

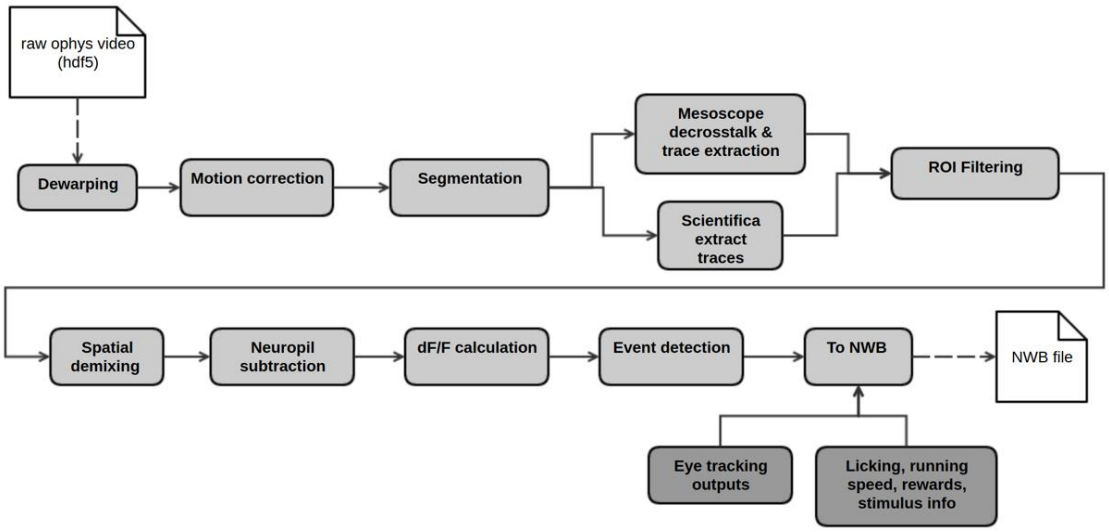
3. Field of View Targeting Validation: Targeted imaging locations were checked against the intrinsic signal imaging data, using a registered coordinate system, to confirm that data was collected from the correct visual area.
4. Z-Axis Stability: Stability of image recording was assessed by comparing a windowed average image from the first and last 5 minutes of the experiment to a z-stack of images (+/- 30 µm around imaging site, 0.75 µm step) collected at the end of each experimental session to calculate the amount of drift that occurred over the session. Experiments with z-drift above 10µm over the course of the entire session were excluded.
5. Animal Stress: Behavior videos were viewed to confirm that animals did not show excessive signs of stress. Any animal that showed eye secretion covering the pupil or excessive orbital tightening was returned to its home cage to recover. The presence of nose bulge, flailing and abnormal postures was also monitored.
6. Task Performance: Behavior performance in the change detection task was confirmed to have met (or exceed) a peak d-prime of 1.0.
7. Temporal Sync: Temporal alignment of data streams was confirmed.
8. Hardware/Software Failure: Multiple datastreams and metrics were assessed to ensure that incoming data integrity was not compromised by hardware and/or software related errors
9. Excessive Motion: Imaging frames were checked for residual motion (after motion correction algorithms had been applied)
10. Interictal Events: Presence of interictal events was assessed by measuring the full field fluorescence and calculating the intensity spike prominence and width for the first 10,000 frames of the 2P movie. Experiments with a non-zero probability of interictal events were then checked manually in order to exclude any potentially epileptic mice[13] (see Steinmetz et al., 2017).

A final, container-level QC assessment was performed once all data collection for a mouse was completed. This secondary assessment included assessing the following metrics:

1. Full Container Status: Containers (the set of imaging sessions for a given field of view) were confirmed to contain all required datasets.
2. Brain Health: 2-photon serial tomography sections were examined to assess general brain health. Health assessment includes checking for excessive bruising, brain abnormalities, deformities, necrotic tissue damage, and checking for any signs of laser damage.
3. Cell Matching: Imaging fields of view were checked across all experiments in a container to confirm successful targeting of the same field of view. This assessment was made visually, based on similarity in patterns of vasculature and the presence of clearly identifiable matched cells across sessions.

2-photon Data Processing

An overview of the data processing steps for both single- and multi-plane imaging is shown below.



Dewarping

To account for variation in scanning rate across each line in the 2p-imaging frame, due to the non-constant speed of the resonant scanner, data acquired with the Scientifica microscopes required a dewarping step (Multiscope instruments have built-in dewarping correction).

Correcting for the warping in the image involves taking a subset of the warped image's columns, towards the edges of the image where the scanning speed is lower, and combining them using parameters derived from calibration data acquired with a standard grid image. To determine correct dewarping parameters, the grid image is adjusted until the dewarped grid is uniform. Each side of the image is dewarped independently.

The following equations are used to determine which columns are chosen from the warped image and how they are combined. The jth column of the dewarped image will be given by the formula:

$$col[j] = (1 + \lfloor g[j-1] \rfloor - g[j-1]) * input_image[\lfloor g[j-1] \rfloor] + (g[j] - \lfloor g[j] \rfloor) * input_image[\lfloor g[j] \rfloor]),$$

where

$$f(x) = x - \text{int}\left[b\left(1 - \sin\left(\left(\frac{x}{3a} + \frac{1.0}{6.0}\right)\pi\right)\right) + 0.5\right], \quad x \in [0, a]$$

$$g(x) = (x + 0.5) - \left[b\left(1 - \sin\left(\left(\frac{x + 0.5}{3a} + \frac{1.0}{6.0}\right)\pi\right)\right)\right], \quad x \in [0, a]$$

Notice that this is taking a linear combination of the two (generally, adjacent) columns of the original image, using the decimal parts of the g function as coefficients. If these ever end up being non-adjacent columns from the original image, then we also add the value of the skipped column to col[j].
In the event that both f[j] and g[j - 1] are not nonnegative, then we fill that column with the average from the original image. And if f is nonnegative but g is not, then we simply replace that column with the exact same column from the original image.

For each experiment, there are four parameters that tell us how much the image has been warped. These are called *aL*, *aR*, *bL*, and *bR*. The L and R refer to the side of the image which is being dewarped. In the equations above, a and b are replaced with their corresponding parameter, depending on whether you are dewarping the left side or the right side. The parameters *aL* and *aR* specify how far (in pixels) into the image the warping occurs. The parameters *bL* and *bR* are more a measure of how warped the image is in those areas.

Motion Correction

We used Suite2P v0.9.3 rigid registration for motion correction of 2-photon movies. Suite2P performs an iterative phase-correlation-based registration on a small subset of frames to generate a reference image from the average projection of those frames. The registration of the entire movie then proceeds with registration by phase correlation of each frame to this reference image. We saved Suite2P's registered output tiff stacks and concatenated them into a single hdf5 format compatible with the rest of our processing pipeline.

Suite2P's parameter *maxregshift* has a default value of 0.1, which clips lateral motions of more than 10% of the FOV dimension. We found a few examples in our release data where our experiments had real, long-timescale lateral shifts of greater than 10%. We increased the *maxregshift* parameter to 0.2, to allow for registration of these experiments. A consequence of this change was that up to 20% lateral shifts were allowed by Suite2P over short timescales as well. We observed that these short timescale shifts, often single frames, were likely not physical shifts, but struggles of the registration algorithm to register low signal-to-noise frames. Across an entire experiment, we monitor the worst-case shifts and establish a motion border which invalidates any ROI which touches it. To prevent too much inflation of the motion border exclusion area by these unphysical short timescale shifts, we detrended the x and y lateral corrections with a median filter over a 3 second window. We clip frame displacements to a +/-5% window around any outlier more than 5% above or below the detrended corrections was truncated to 5%. This allowed us to handle

physical long timescale drifts and limit the impact of artifactual short timescale shifts. Any movie frames where the lateral shift was truncated were translated to the new truncated translation value.

For each movie, we produce the following outputs: the registered hdf5 movie for downstream processing, the motion correction shifts both from Suite2P directly and after our detrended clipping step, and Suite2P's output corrXY, the value of phase-correlation at the best alignment for each frame. For QC inspection, we generate plots showing the x and y lateral shifts and corrXY resulting from motion correction, as well as a video preview showing the movie before and after motion correction, side-by-side, averaged to 0.5 frames-per-second and a 10x playback speed.

Cell Segmentation & ROI Filtering

The Visual Behavior 2P project used the same segmentation procedure that was developed for the Visual Coding 2P dataset, published in de Vries et al., 2020. The active cell segmentation module was designed to locate active cells within a field-of-view (FOV) by isolating cellular objects using the spatial and temporal information from the entire movie. The goal of the active cell segmentation module is to achieve robust performance across experimental conditions with no or little adjustment, such as different mouse cell lines, fluorescent proteins (e.g., GCaMP6f or GCaMP6s), and FOV locations of visual areas and depths. The process begins with the full image sequence as input to apply both the spatial as well as temporal information to isolate an individual active cell of interest without data reduction, such as by PCA, and does not make assumptions about the number of independent components existing in the active cell movie. Also, in contrast to other methods, this approach separates the individual steps, including identifying and isolating each cellular object, computing confidence of each identified object (by object classification) and the step of resolving objects overlapping in x-y space (which lead to cross talk in traces), so that each can be improved upon if necessary.

Pre-segmentation

The motion corrected image sequence was spatially median filtered (using 3x3 pixel kernel) to reduce white noise. The sequence was then low pass filtered and downsampled by 1/8 temporally to enhance the signal-to-noise ratio (SNR). The processed image sequence was then divided into periods of fixed temporal length p, where p = 50 frames (~13.3 sec.). The maximum projection image from each period and the mean image (mu_image) of the whole sequence were computed. The maximum projection image from all temporal periods, called Periodical Projection frames PP(t), were further normalized to become Normalized Periodical Projection (NPP) frames:

$$NPP(t) = MF((PP(t) - mu_image), 3x3) * G(t)$$

Where G(t) is the frame intensity normalization gain computed based on the intensity histogram of each PP(t). This is to normalize any change in overall intensity across the experiment, and to reduce experiment-to-experiment variability. MF(3x3) is median filtering with a 3x3 pixel kernel.

Note in each NPP(t), a subset of cells can be found with changes in fluorescence during that time period. With sufficient experiment length, and with many sweeps of different stimuli, various repetitive cell firing patterns can be found. Cells with overlapping spatial positions in x and y can be observed as firing at different time frames, allowing the following detection process to identify them individually despite having spatial overlap.

ROI Detection

Adaptive and mathematic morphological image processing techniques were applied to process each NPP(t). After band-pass filtering, an initial binary object map was generated by thresholding the resulting image minus a low pass version of itself to capture spatially varying background intensity. Conditional dilation, erosion and connected component analysis were applied to filter the candidate binary objects and fill holes. The final set of regions of interest (ROIs) in each NPP were identified using another connected component labelling and a simple rule-based classifier. This classification was based on comparing measured morphometric attributes (object area, shape, intensity, uniformity, etc.) of the objects to the statistics derived from the targeted active cell components in the sample data sets. After each frame was processed, a set of candidate ROIs from each NPP were then grouped with candidate ROIs from all other NPPs. ROIs within a similar spatial location (defined by the distance between centroids < 5 µm) and with similar morphometric attributes (e.g., delta (area or shape) < 20%) across NPP frames were grouped as the same cell object. The ROI with the highest contrast and with shape and area within range statistically derived from sample data was selected to represent that cell in the mask image and in a composite image for visual QC. ROIs with different spatial locations (centroids > 5 µm apart) and or dissimilar morphometric attributes are recorded as different cells.

Occasionally two or more spatially overlapping ROIs could be found that were active in the same timeframe and were therefore detected as a single ROI for that NPP. Additional steps to classify them as multiple-cell objects were taken using their attributes of combined area and shape (eccentricity).

After ROI detection in each NPP frame and grouping of all ROIs across frames was completed, a set of "unique" active cell objects was identified. Cells near the FOV boundaries (3 µm) were eliminated from further consideration due to the fact that motion shifts can create boundary effects to the traces computed from these cells. To generate the segmentation mask image, all non-overlapping cells were placed in a single mask plane and overlapping cells were placed in subsequent planes, to ensure unique identification of all cells.

Crosstalk Removal in Multiscope Data

Crosstalk between focal planes is a fundamental limitation in multiplexed microscopy systems, such as the Multiscope. Details about system hardware and methods used to optimize crosstalk in the data are described above and can be found in Orlova et al., 2020 [11]. Crosstalk removal was performed on fluorescence traces using an ICA-based approach (implementation scikit-learn.FastICA), where independent components are estimated by minimizing Gaussianity of the data[14] (Hyvärinen, 1999). The assumption is that the two planes are a mixed observation of two clean sources and are mixed linearly using a mixing matrix. We assume a mixing matrix of the form [[1-a, a], [b, 1-b]], where a and b are in [0,1), and probably around 0.15. After FastICA, we transform the resulting mixing matrix to be of this form to recover the proper scaling of the mixed signals. Prior to FastICA, data undergoes whitening; we do not use the built-in whitening of the FastICA module (as it appears to contain a bug that affects scaling of the outputs).

Algorithm description

The plane whose traces we are correcting is referred to as the signal plane. The plane coupled to it (simultaneously acquired with temporal multiplexing) is referred to as the crosstalk plane. The algorithm is run on both permutations of signal plane and crosstalk plane, i.e., for each pair, it is run once with plane A as the signal plane and plane B as the crosstalk plane and once with plane B as the signal plane and plane A as the crosstalk plane.

Cell segmentation is performed on each plane independently to generate cell ROIs. We define the set of ROIs as those detected in the signal plane. We construct raw signal traces by measuring the average per-pixel flux in each ROI in the motion-corrected movie taken from the signal plane (trace extraction). We construct raw crosstalk traces by measuring the average per-pixel flux in the same ROIs in the motion-corrected movie taken from the crosstalk plane (these ROIs are those detected in the signal plane). For both the raw signal traces and the raw crosstalk traces, raw neuropil traces are constructed using the pixels bordering the ROIs. We flag any ROI whose footprint or neuropil intrudes on the motion correction border as invalid and remove it from further processing.

Signal and crosstalk planes are coupled and leak into to each other due the imperfections of the temporal multiplexing approach. We assume that, for each ROI, the measured, or raw signal and crosstalk traces are linear combinations of clean, or unmixed signal and crosstalk traces, i.e.

R = M U

where R is a 2xN matrix (N is the number of timesteps in the traces) such that R[0,:] is the raw signal trace and R[1,:] is the raw crosstalk trace. U is a 2xN matrix such that U[0,:] is the unmixed signal trace and U[1,:] is the unmixed crosstalk trace. M is the 2x2 mixing matrix relating the two. For each ROI, U[0,:] represents the trace coming from that ROI uncontaminated by crosstalk. That unmixed signal is the trace that we want to use in all future processing steps. We use Independent Component Analysis (ICA) [1] to solve for U and M. Specifically, we:

1) Subtract the mean from each trace and find the 2x2 matrix W which will transform (R-mean) into a "whitened" dataset whose correlation matrix is approximately the 2x2 identity matrix.
2) Use scikit learn's FastICA algorithm to solve for U and M
3) The scale of the traces is corrected by computing the scaling that will restore the mixing matrix to the assumed form of [[1-alpha, alpha], [beta, 1-beta]]. This scaling can be computed by multiplying the inverse of the mixing matrix by the vector [1,1]. This transformation is applied to the traces to restore the original scale.
4) If the off-diagonal elements of M are positive and less than 0.3, ICA has converged to a valid result. Identify the unmixed trace that is most closely correlated with the raw signal trace as the unmixed signal trace (ICA is agnostic regarding the ordering of unmixed signals).
5) If ICA failed to converge to a valid result, the ROI is marked as having failed ICA.

We use all the ROIs that pass the above algorithm to construct an average mixing matrix M for the signal plane. The inverse of this matrix is used to unmix any ROIs marked as "failed" in step (4). We find unmixed neuropil traces by

taking the mixing matrix used for the associated ROI and multiplying its inverse by the matrix of raw neuropil traces. If no ROIs in the plane converged to a valid result, it is impossible to construct an average mixing matrix for the plane and we mark the plane as having failed processing.

We run the unmixed signal and unmixed crosstalk traces through an event detection algorithm described in Giovannucci et al. 2019[15]. We define events as a sequence of five timestamps whose sum of log probabilities exceeds a threshold (14 in our implementation). We calculate the probability relative to a Gaussian distribution centered on the mode trace value whose width is calculated considering the distribution of all trace values less than the mode. We define each timestamp that passes this test as an event in the trace.

For each ROI, we assess whether any events occur in the unmixed signal trace that are independent from the events in the unmixed crosstalk trace. We define independent events as those events in the unmixed signal trace which do not occur within two timestamps of an event in the unmixed crosstalk trace. If no such independent events occur in the unmixed signal trace, we mark the ROI as a ghost (i.e. an ROI that was only detected because of crosstalk contamination from the coupled plane) and discard it from future processing steps.

ROI Filtering

Not all ROIs generated by segmentation are complete individual cell bodies. To exclude ROIs that are not actually cell bodies from further analysis, the ROIs are labeled with a multi-label classifier that distinguishes ROIs that are considered well defined cell bodies from other ROIs. The set of reasons to exclude an ROI are: the ROI is a union of two or more cells; the ROI is a duplicate of another; the ROI is close to the edge of the FOV and is impacted by motion such that parts of the ROI are missing from the video; the ROI is likely an apical dendrite and not a cell body; or that the ROI is too small, too narrow, or too dim to confidently be considered a cell body.

The initial ROI filtering was generated by a set of heuristics based on depth, shape, area, intensity, signal-to-noise, and the ratio of mean to max intensity of the max projection. The initial filtering was used to generate a set of training labels, on which a multi-label classifier was trained. The multi-label classifier is implemented using a linear Support Vector Classifier trainer for each label (binary relevance) using metrics generated in segmentation combined with depth, driver, reporter, and targeted structure as features. The final ROI filtering workflow is to 1) label ROIs that fall within the motion cutoff regions at the border, 2) label ROIs using the binary relevance classifier, 3) label significantly overlapping ROIs as duplicates, and 4) label ROIs that significantly overlap two or more ROIs as unions.

Demixing Traces From Overlapping ROIS

The simplest way to extract fluorescence traces, given a set of ROI masks, is to average the fluorescence within each ROI. If two ROIs overlap, this procedure will artificially correlate their traces. Therefore, a model is used where every ROI has a trace which is distributed across its ROI in some spatially heterogeneous, time-dependent fashion:

$$F_{it} = \sum_{k} W_{kit} T_{kt}$$

where W is a tensor containing time-dependent weighted masks: W_{kit} measures how much of neuron k's fluorescence is contained in pixel i at time t. T_{kt} is the fluorescence trace of neuron k at time t - this is the desired value to estimate. F_{it} is the recorded fluorescence in pixel i at time t.

Importantly, this model applies to all ROIs, including those too small to be a neuron or otherwise filtered out.

Duplicate ROIs (defined as two ROIs with >70% overlap) and ROIs that are the union of two other ROIs (any ROI where the union of any other two ROIs accounts for 70% of its area) are filtered out before demixing, and the remaining filtering criteria are applied after demixing. Projecting the movie (F) onto the binary masks (A) reduces the dimensionality of the problem from 512x512 pixels to the number of ROIs:

$$\sum_{i} A_{ki} F_{it} = \sum_{k,i} A_{ki} W_{kit} T_{kt}$$

where A_{ki} is one if pixel i is in ROI k and zero otherwise–these are the ROI masks from segmentation, after filtering out duplicate and union ROIs. At a particular time point t, this yields the simple linear regression:
$AF(t) = (AW^T(t))T(t)$

where the weighted masks W are estimated by the projection of the recorded fluorescence F onto the binary ROI masks A. On every imaging frame t, the linear least squares solution T are computed in to extract each ROI's trace at that time point.

It was possible for ROIs to have negative or zero demixed traces T. This occurred if there were union ROIs (one ROI composed of two neurons) or duplicate ROIs (two ROIs in the same location with approximately the same shape) that the initial detection missed. If this occurred, those ROIs and any that overlapped with them were removed from the experiment. This led to the loss of ~1% of ROIs.

Neuropil Subtraction

The recorded fluorescence from an ROI was contaminated by the fluorescence of the neuropil immediately above and below the cell due to the point-spread function of the microscope. In order to correct for this contamination, the amount of contamination was estimated for each ROI. The estimated F_N was done by taking an annulus of 10 mm around the cellular ROI, excluding pixels from any other ROIs. In order to remove this contamination, the extent to which ROI was affected by its local neuropil signal was evaluated.

The recorded traces were modeled as F_M as $F_M = F_C + rF_N$, where F_C is the unknown true ROI fluorescence trace and F_N is the fluorescence of the surrounding neuropil. In order to estimate the contamination ratio r for each ROI, the error was minimized $E = \langle (F_C - (F_M - rF_N))^2 + \lambda\Lambda(F_C)\rangle_t$ by jointly optimizing for r and F_C. $\Lambda(F_C)$ is the first temporal derivate of the cellular trace, weighted by $\lambda = 0.05$; this smoothness constraint on the cellular trace allows per-ROI optimization for r. $\langle\cdot\rangle_t$ denotes an average over time. Gradient descent was used on r. At each step of the gradient descent, F_C was solved at the zero gradient of E. Gradient descent was performed on the first half of the traces and computed E on the second, so that it is a cross-validation error. After computing r and F_C for an ROI, the neuropil-subtracted trace $F_M - rF_N$ was used as the basis for all subsequent analysis in order to avoid any residual effects of the smoothness constraint.

To standardize the learning rate and initial conditions of the gradient descent, each ROI's neuropil trace was normalized to (0,1). The measured ROI trace was normalized by the same amount, used a learning rate of 10 and initial condition of $r = 0.001$. The gradient descent was stopped at the first local minimum of E. If the resulting r was greater than 1 or less than 0, or final cross-validation error E greater than $2\ |\langle F_M\rangle_t|$, the gradient descent was attempted again with a 10x slower learning rate. If those convergence criteria were still not met, an initial condition of $r = 0.5$ was used. If those convergence criteria were still not met, that ROI was flagged and after computing r for all other ROIs in the experiment, set r for un-converged ROIs to the mean.

To validate the performance of our algorithm, it was tested on a publicly-available benchmark dataset (Chen et al., 2013). A distribution of contamination ratios was obtained, centered nearly on the author's choice of 0.7 (mean r of 0.68 vs their choice of 0.7), but with significant heterogeneity. For this benchmark dataset and using the same optimization parameters as for the Allen Brain Observatory – Visual Coding data, 6 of 36 cells failed the initial neuropil subtraction with r>1, and would have gone through the additional steps outlined above.

dF/F Calculation

We generate normalized, detrended traces of neuronal activity by performing the following algorithm on each trace:

First, we estimate the standard deviation of the noise in the trace. We do this by centering the trace on the curve resulting from applying a median filter with kernel size 3.33 seconds to the trace. To avoid including signal events in our estimate of noise, we discard any values that exceed 1.5 times the absolute value of the minimum of this centered trace. We make a first estimate of the standard deviation of noise in the trace as 1.4826 times the median absolute deviation of this truncated, centered trace. We further discard any values that exceed 2.5 times this first estimate of the standard deviation and finally return 1.4826 times the median absolute deviation of the remaining centered trace values as the standard deviation of noise in the trace.

Next, we calculate the dF/F trace. We define the baseline activity in the trace as the result of applying a median filter with kernel size 600s to the trace. We subtract this baseline from the raw trace and normalize the difference by the baseline. At timesteps where the baseline trace is less than the standard deviation of noise calculated above, we normalize by the standard deviation of noise, instead. This gives the dF/F trace.

Finally, we detrend the dF/F trace. We estimate the trend in the dF/F trace by applying a median filter with kernel length 3.33 seconds. To prevent anomalously large trends from arising, we use the same algorithm we used to estimate the standard deviation of noise in the raw trace to estimate the standard deviation of noise in

the dF/F trace. We constrain the trend to never exceed 2.5 times this estimated standard deviation of noise in the dF/F trace. We subtract the trend from the dF/F trace and report this as the detrended dF/F trace.

Session to Session Cell Matching

Multiple 2-photon calcium imaging movies were acquired for each imaging plane across multiple imaging sessions. To map cells between sessions, we used an automated matching algorithm. The module has 4 steps:

1. Determining the spatial transform between each pair of sessions using image registration techniques on the average projection images of the sessions.
2. Applying the derived spatial transform to segmented cell masks.
3. For each pair of sessions, solving the linear assignment problem (i.e. bipartite graph matching) to determine which pairs of masks are most likely to be related, based on the inter-centroid distance of the segmented masks and the intersection over union (IOU) of the masks. A 10 pixel distance threshold is placed on the inter-centroid distance to ensure only closely overlapping masks were matched.
4. A graph combination method was used to join all the bipartite graphs to determine the most likely label sets that match across all sessions.

Code for pairwise session cell matching is available here: https://github.com/AllenInstitute/ophys_nway_matching

Determining the spatial transformation

The module first used an intensity-based method to register the average intensity projection images of each pair of sessions, producing a Euclidean transformation that registered each image pair. We found that a single image registration strategy did not work for all cases. In some cases, the true lateral shifts between two experiments were too large for the ECC registration algorithm. In other cases, the border in the projection images biased the registration towards matching borders and not cell soma. In still other experiments, a histogram equalization step that helped most experiments resulted in a degraded registration.

We implemented a "meta-registration" where 4 different registration sequences were attempted. Each sequence employed an initial cropping step to eliminate the motion border problem. The sequences employed varying combinations of ECC, PhaseCorrelation, and contrast adjustment. After each of the 4 sequences were attempted, we chose the best candidate based on the structural similarity metric (SSIM) between the two registered images. If none of the candidate sequences improved SSIM beyond that achieved with unregistered images, we flagged the registration as a failure for QC inspection.

In a few cases, there were still some unexplained failures to register. In these cases, we found that the average projection intensity images had low contrast, but the maximum projection intensity images had more visible features. By substituting in the maximum projections for the registration step, we achieved good registration and good cell matching for these containers.

Applying the spatial transformation to cell masks

The cell masks were materialized into a set of images, such that no ROIs overlapped in any one image. We then applied the relative transformation found from the intensity projections to these materialized images, to get the cell masks in an aligned space.

Solving the linear assignment problem between 2 sessions

To map cells, a bipartite graph matching algorithm (the Blossom method) was used to find correspondence of cells between sessions. The algorithm used cell labels in the pair-wise experiments as nodes, and the edges of the graph were weighted with weights, w :

$$w = \frac{d}{d_{max}} + \left(1.0 - \frac{I}{U}\right)$$

where d is the distance between the 2 mask centroids (in pixels), d_{max} = 10 pixels, I is the number of pixels that are shared between the 2 masks (the intersection), and U is the total number of pixels covered by the two masks combined (the union). The graph edge for any pair of masks with d > 10 pixels was assigned a very large weight, eliminating that pair as a candidate match.

The original version of this module used the "Hungarian" method to minimize the total weight. SciPy has an equivalent implementation of that method. We found that the solution from the "Blossom" method was more robust to

order permutations of the inputs. We used the networkx implementation of the "Blossom" method. As this implementation seeks to maximize total weight, we inverted the weights, taking care to prevent division-by-zero:

$$w = \frac{1}{1 + w}$$

By maximizing the summed weights of edges, the bipartite matching algorithm found the "best match" between cells in each pair of experiments.

Joining the pairwise bipartite graphs

All the pairwise bipartite graphs were naively combined. This combined graph is not constrained to avoid labeling conflicts. Matching was performed on the full set of ROIs identified by the segmentation algorithm, but ROIs that were deemed invalid by the ROI filtering step were not included in the final set of matched cells.

For quality control checks, we produced standard plots of pairwise registrations across sessions, as well as quantification of the fraction of ROIs matched across sessions. Session pairs with visible failures to register (clear features in the image were not aligned), or with very low fraction of matched cells (<10% matching between session pairs), were manually evaluated and an attempt was made to optimize the registration process to produce accurate registration. Session pairs that were not possible to register were excluded from the dataset.

The code for performing field of view registration and cell matching is tested in a CI/CD system and openly available at: https://github.com/AllenInstitute/ophys_nway_matching

Calcium Event Detection

We used the FastLZeroSpikeInference ("FastLZero") to identify events in traces derived from the 2P movies[16]. FastLZero fits a trace to a sum of exponentially decaying spikes, with the timescale of decay determined by a single parameter (gamma), while imposing an L0 regularization to minimize the number of events used to fit. Gamma has been empirically determined for each genetically defined cell type by running the MLSpike autocalibration routine on all available recordings and determining the maximum likelihood decay constant across the distribution of fitted decay constants for available neurons of any given cell type[17].

The relative weight in the optimization penalty between number of events used and best residuals is controlled by a regularization factor (lambda) that controls the tradeoff between missing events (false negatives) and overfitting of noise with extraneous events (false positives). Lambda needs to be determined empirically as a function of the noise level present in each recording. We iteratively searched the regularization factor space to find the regularization factor where the smallest event provided by FastLZero was at least two times the estimated noise level of the fluorescence trace. The noise estimator was initially optimized for experiments recorded at 31Hz (single-plane imaging with Scientifica 2-photon microscopes) and validated against ground truth data manually annotated by a human expert.

Multiscope experiments were recorded at 11Hz, and we observed that the same multiplicative factor (2.0) resulted in many more small amplitude events than in the 31Hz data when probed with synthetic calcium data. One can rationalize this difference intuitively by noting that the characteristic timescales of fluorescence decay do not change based on sampling rate, but the noise estimation does. Thus, our noise estimator, which was only validated for 31Hz data, systematically underestimated the noise for 11Hz data. To correct for the effects of sampling rate on the noise estimation, for Multiscope, we empirically determined the equivalent multiplicative factor to be 2.6 by identifying the value that minimized the discrepancy between event magnitude traces extracted (using factor of 2.0) from Scientifica data sampled at 31 Hz, and event magnitude traces extracted from the same data but after downsampling by a factor of 3.

Eye Tracking Data Processing

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). 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 and the ellipse fit parameters were saved to disk. 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.

All code for processing of the eye tracking data is visible in the brain_observatory.behavior.eye_tracking_processing module of the AllenSDK. The eye tracking data is available as the 'eye_tracking' dataframe in the session object. The dataframe has the following columns:

- Timestamps: the timestamp of every frame
- {pupil, eye or cr}_center_x: the x-position of the center of the pupil, eye or corneal reflection ellipse fit on that frame, in pixel space of the eye tracking movie frame.
- {pupil, eye or cr}_center_y: the y-position of the center of the pupil, eye or corneal reflection ellipse fit on that frame, in pixel space of the eye tracking movie frame.
- {pupil, eye or cr}_width: the (more) horizontal half-axis of the ellipse fit of the pupil, eye or corneal reflection
- {pupil, eye or cr}_height: the (more) vertical half-axis of the ellipse fit of the pupil, eye or corneal reflection
- {pupil, eye or cr}_phi: the angle to the (more)horizontal axis of the ellipse fit of the pupil, eye or corneal reflection, measured CCW from the x-axis of the image frame.
- likely_blink: a Boolean defining frames that have been identified as likely outlier fits, which is often caused by blinking/squinting of the eye.
- pupil_area: the area of the pupil, assuming that the pupil is a circle with a diameter defined by the major axis of the ellipse fit. Set to NaN where likely_blink == True.
- pupil_area_raw: the area of the pupil, assuming that the pupil is a circle with a diameter defined by the major axis of the ellipse fit. No outliers/likely blinks are removed.
- {eye or cr}_area: the area of the eye or corneal reflection ellipse fit. Set to NaN where likely_blink == True.
- {eye or cr}_area_raw: the area of the eye or corneal reflection ellipse fit. No outliers/likely blinks are removed.

Briefly, the ellipse fit parameters produced by the processing pipeline were loaded for each session. It was assumed that the pupil is actually round, but when viewed obliquely, it 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 on 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 can be low, and ellipse fits either fail completely or may be fit erroneously. To avoid including these erroneous fits in analysis, an algorithm attempts to identify these blink frames, adding a column to the eye_tracking dataframe called 'likely_blink'. 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 areas 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.

It is important to note that the outlier frames identified by the likely_blink algorithm are not always caused by blinks. Instead, they may be frames where the DeepLabCut algorithm simply failed to identify a reasonable fit, triggering the outlier detection portion of the algorithm.

Data analysis

Behavior performance

Reward rate was computed as the average rewards per second in a rolling 320 second window, over each 60-minute behavior session. Engaged periods of the session were defined as having >2/3 rewards per minute.

Response probability for image changes (i.e. hit rate, off-diagonals in Figure 1f and Extended Data Fig. 2e) was computed as the fraction of stimulus changes during engaged periods with a correct licking response. Response probability for non-changes (i.e. false alarm rate, diagonals in Figure 1f and Extended Data Fig. 2e) was computed as the fraction of stimulus presentations that could have been a change (i.e. not during reward window, reward consumption "grace period", or 2 second period prior to drawing a change time) when an incorrect response was emitted. Response probability was computed for each session then averaged across sessions with a given image set (Familiar or Novel) to produce Figure 1f and Extended Data Fig. 2e.

The value of d-prime, also known as sensitivity index, was calculated as the relative difference between the inverse cumulative normal distribution function of the hit rates and false alarm rates. Hit and false alarm rates were computed as described above (hit rate as response probability for image changes and false alarm rate as response probability for stimulus presentations that could have been a change but did not change). d-prime was computed for all image transitions during engaged periods for Extended Data Figs. 1f and 2d. For Figure 1g, d-prime was calculated using the hit rate for the 2 most detectable image transition pairs, compared to the corresponding non-change stimulus presentations for the images in the pairs, in order to facilitate a direct comparison of the impact of novelty on behavior performance, independent of image detectability.

Normalized pupil width

Pupil width, defined as the horizontal half-axis of the ellipse fit of the pupil (described above in "Eye Tracking Data Processing"), was pre-processed by filtering out "likely blinks" and linearly interpolating values. Pupil width was then normalized to the 5-minute gray screen period at the beginning of the session to account for any differences in pupil area resulting from slightly different placement of the dichroic mirror reflecting the eye to the eye tracking camera. This allowed a direct comparison of normalized pupil width across Familiar, Novel, and Novel+ sessions in Extended Data Fig. 8a, d. Normalized pupil width was aligned to the time of image changes or image omissions within each session and averaged, then averaged across all sessions of a given type.

Temporal alignment of 2P and behavior

Temporal alignment is needed to link and analyze data streams collected at different frame rates. In our experiments, 2-photon movie timeseries collected on Scientifica microscopes for single-plane imaging were collected at 30Hz. 2-photon movie timeseries collected on the Multiplane Mesoscope were collected at 11Hz. Running speed, licking, and reward delivery were collected at 30Hz, at the frequency of the stimulus display. Eye tracking movies were collected at 60Hz.

To generate stimulus locked traces of neural activity, we extracted a subset of each cell's dF/F trace (or detected events) in a window around each stimulus onset time. To prevent shifting of timestamps to align 2-photon data to stimulus times, we interpolated neural activity traces to a 30Hz sampling rate relative to the stimulus onset time for each trial. This allowed more accurate averaging of neural signals across trials, as each trial's timestamps were interpolated relative to the stimulus start time. A demonstration of this procedure can be found on GitHub, along with the functions to perform the alignment.

Response characterization

All response analyses were performed using detected calcium events smoothed with a causal half-gaussian filter with scale 65ms. To remove aliasing between the neural activity timestamps and stimulus timestamps all continuous signals were interpolated onto stimulus-aligned timestamps, as described above.

Trial level metrics

For each 2-photon experiment, a pandas dataframe was created containing stimulus aligned traces (aligned and extracted as described above) and associated metadata for every stimulus presentation and stimulus omission during the behavior session. In addition to a segment of each neuron's trace in a window around the stimulus onset time, we computed the mean response in a 500ms window after stimulus onset (hereafter referred to as mean response for a

given stimulus presentation). For stimulus omissions, we used a 750ms window to capture the full window prior to the next stimulus, based on empirical observation of omission ramping in Vip neurons. The significance of the response to each stimulus presentation was computed by comparing the mean response in the 500ms (or 750ms) stimulus window with the mean response in a 500ms (or 750ms) window taken from a shuffled distribution of that neuron's activity during the gray screen spontaneous activity period where no stimulus was present. A p-value was calculated by comparing these values across 10,000 shuffles and computing the fraction of those shuffles where the mean stimulus response was larger than the mean response in the shuffled data.

Population average timeseries

To produce a population average timeseries across all neurons within a given cell class, stimulus aligned traces were aggregated over a given condition (ex: change aligned or omission aligned) and averaged for each neuron, then averaged across all sessions of a given type (ex: Familiar, Novel, Novel+) and displayed +/-SEM in Figures 2b-d, i-k. The average timeseries for all neurons are shown as heatmaps for each condition in Extended Data Figure 4.

For behavioral timeseries, the stimulus aligned traces for all trials of a given condition were averaged for each session, then averaged across sessions, and displayed +/- SEM in Extended Data Fig. 8.

Population average metrics

To compute the trial averaged response to image changes, we selected all stimulus presentations corresponding to image changes (regardless of image identity) in a given session, then took the average of the mean response in the 500ms window following change onset across all change trials for each neuron. The same procedure was used to compute each neuron's average response across all pre-change stimuli. For omissions, the same procedure was applied using a 750ms window after the time the stimulus would have been displayed if it were not omitted. See 'get_mean_df' function here:

For each of these conditions (change, pre-change, omitted), we created a distribution of mean response values across neurons for each experience level (Familiar, Novel, Novel+) and report the mean +/-95% CI in Figures 2e,f&l.

To test whether response distributions were significantly different across experience levels, areas, or depths, we used one-way ANOVA, followed by Tukey HSD to compare population means across the conditions being compared. A p-value of 0.05 was used as the significance threshold.

Fraction responsive neurons

A neuron was considered "responsive" for a specific condition if at least 25% of trials for that condition (ex: changes, omissions) had a p-value > 0.05 (p-value determined by comparing each trial's mean response to a shuffled distribution as described above in "trial level metrics" section). The fraction of responsive neurons for each session was then calculated as the number of "responsive" neurons divided by the total number of neurons. The average fraction of responsive neurons across sessions for each experience level is reported in Extended Data Figure 4.

Population decoding

General methodology for decoding population activity

A linear SVM (Python scikit-learn package) was trained on each bin of the population activity in each session. Population activity was aligned on either the image changes or omissions (Figures 2g, m, n, Extended Data Figs. 5, 6). Time bins within the window [-0.5 0.75] sec surrounding the image change (or omission) were used for classification. One classifier was trained per time bin (non-overlapping 93ms time bins for Mesoscope experiments, and 33ms time bins for Scientifica experiments). Decoding analysis was performed on the full population of neurons recorded in each session (Figure 2g, m, n), also on the "cell matched" population (Extended Data Fig. 4). To obtain the "cell matched" population, neurons were matched across 3 sessions: the last Familiar session, the 1st Novel session, and the 2nd Novel session. Similar results were found from both populations (full and cell matched).

Details of the decoding method are explained in Orlova et al. 2019[11]. In brief, to break any dependencies on the sequence of trials, we shuffled the order of trials for the entire population. L2 regularization was used to avoid over-fitting. 5-fold cross validation was performed by leaving out a random 1/5 subset of trials to test the classifier performance, and using the remaining trials for training the classifier. This procedure was repeated 50 times. A range of regularization values was tested, and the one that gave the smallest error on the validation dataset was chosen as the optimal regularization parameter. Classifier accuracy was computed as the percentage of testing trials in which the class was accurately predicted by the classifier and summarized as the average across the 50 repetitions of trial subsampling.

A minimum of 10 trials and 3 neurons was required to run the SVM on a session. The inferred spiking activity of each neuron was z-scored before running the SVM.

Decoding image change occurrence

To address whether population activity in single trials carries a signal about image changes, population activity following an image change was classified against the activity following a non-change image (Figure 2g; Extended Data Fig. 5b). For the non-change image, the image immediately preceding the image change was used. One SVM was trained per time bin in the window [-0.5 0.75] sec surrounding the pre-change image and image change. The input to each SVM was formed in the following way: population activity at time t relative to each image change (class 1; size: n x m, where n: number of image changes; m: number of neurons) was vertically concatenated with the population activity at time t relative to each pre-change image (class 0). This was done for all image changes and pre-change images in the session (SVM input size: 2*n x m, where n: number of image changes; m: number of neurons), allowing to study how the population activity at time t after an image change was distinct from the activity at time t after a non-change image. The output of the SVM constituted of two classes, representing image changes and no changes.

Decoding image identity

To address whether population activity in single trials carries information about image identity, population activity following an image change was classified according to the image identity (Extended Data Figure 5a). One SVM was trained per time bin in the window [-0.5 0.75] sec surrounding the image change. The input to each SVM was formed in the following way: population activity at time t relative to the image change was concatenated for all image changes (SVM input size: n x m, where n: number of image changes; m: number of neurons). The output of the SVM constituted of 8 classes, representing the 8 different images presented in each session.

Decoding omission occurrence

To address whether population activity in single trials carries a signal about omissions, population activity following an omission was classified against the activity during the gray screen (Extended Data Fig. 6a). Gray screen activity was taken from the time bin immediately preceding the omission. One SVM was trained per time bin in the window [-0.5 0.75] sec surrounding the omission. The input to each SVM was formed in the following way: population activity at time t relative to each omission (class 1) was concatenated with the population activity at time t-1 relative to each omission (class 0, gray-screen activity). This was done for all omissions in the session (SVM input size: 2*n x m, where n: number of omissions; m: number of neurons). The output of the SVM constituted of two classes, representing omissions and no omissions, i.e., gray screens.

Decoding image identity from omission-evoked activity

To address whether population activity following omissions carries information about the identity of the subsequent image, population activity following an omission was classified according to the post-omission image identity (Extended Data Fig. 6b). One SVM was trained per time bin in the window [-0.5 0.75] sec surrounding the omission. The input to each SVM was formed in the following way: population activity at time t relative to omission was concatenated for all omissions (SVM input size: n x m, where n: number of omissions; m: number of neurons). The output of the SVM constituted of 8 classes, representing the 8 different images that followed the omission in each session.

Summary quantification of decoding results

To quantify the magnitude of population decoding accuracy, the decoding accuracy trace of each experiment was averaged over 400 ms after image (or omission) onset for decoding image changes, image identity, and post-omission image identity, and over 750ms after omission onset for omission decoding analysis. This quantity was averaged across all experiments collected using both single and multi-plane two-photon imaging (Figures 2g, m, n)

Statistical tests

We used two-way ANOVA, followed by Tukey HSD to compare population decoding accuracies across experience levels (Familiar, Novel 1, Novel >1). Two-sided t-test was used to compare correlations between the real and shuffled data, for each experience level. A p-value of 0.05 was used as the significance threshold.

Regression model

Our regression model is a linear model with time-dependent kernels with a gaussian noise model. Each model feature is a vector $f_i(t)$, and was convolved with a learned kernel $k_i(t)$ to produce the predicted model component of that feature $f_i(t) * k_i(t)$. Model components were summed together to produce the full model response $r(t) = \sum_i f_i(t) * k_i(t)$. The learned kernels were instantiated as vectors of weights. The length of each feature kernel was determined using domain knowledge and trial and error. For continuous features (running speed and pupil diameter) $f_i(t)$ was a time-series containing the data series at each time point. For discrete features $f_i(t)$ was 1 on timesteps where that feature occurred, and 0 elsewhere.

Stimulus pre-processing

For each cell, we fit the model to detected calcium events smoothed with a causal half-gaussian filter with scale 65ms. This smoothing step was performed in order to fit the model to a continuous signal. To remove aliasing between the neural activity timestamps and feature timestamps all continuous signals were interpolated onto stimulus-aligned timestamps. Discrete features were binned onto the nearest stimulus-aligned timestamp. For a given experiment, if a feature had less than 5 discrete events, that feature was not included in the model. Continuous features (pupil diameter and running speed) were standardized to have mean zero, and unit variance.

Toeplitz Matrix Implementation

Convolutions are time-invariant linear shift operators, which allowed us to implement the convolution of kernels with feature timeseries as a banded toeplitz matrix operation $y = Wx$. Here y is the 1d time-series for a cell, 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.

Closed form solution and ridge regularization

The toeplitz matrix implementation with a gaussian noise model yields the standard solution for ordinary least squares regression: $x = (W^T W)^{-1} W^T y$. We added an L2 ridge regression penalty to the cost function, resulting in the following solution: $x = (W^T W + \lambda I)^{-1} W^T y$ where λ is the L2 penalty. We fit the model using 5-fold cross validation. We split each session into 50 intervals, and randomly assigned 10 intervals to each cross-validation fold. Therefore, each of the 5 folds were intermingled in time. To determine the hyper-parameter λ we evaluated the model on a grid of potential λ values from 0 to 500, and for each experiment we selected the λ that resulted in the best test-set performance across cells in that experiment. The training/test splits were different for hyper-parameter selection than fitting the model for analysis.

Model evaluation

We evaluate our model by computing the explained variance: $\frac{(\sigma^2_{total} - \sigma^2_{residual})}{\sigma^2_{total}}$. If the smoothed calcium event trace had no activity (no events) on a cross-validation fold, then then cell has no variance on that fold. Therefore, for numerical stability, when a cell had no activity on a cross-validation fold, the model explained variance was set to 0.

Unique contribution of model features

To evaluate the contribution of each kernel in explaining each cell'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 compute a coding score which measures the fraction of the full model's explained variance that is lost when using the reduced model. This coding score therefore captures the unique explained variance from that kernel or set of kernels. The raw coding score for feature *i* is given by: $\frac{(1 - VE_{reduced,i})}{VE_{full}}$. Some model features are only present during a portion of the session, we therefore adjust the raw coding score to only evaluate the explained variance on the time points when that feature was present. The adjusted coding score for feature *i* is given by: $\frac{(1 - VE(t_i)_{reduced,i})}{VE(t_i)_{full,i}}$. Here t_i indicates timepoints where feature *i* was in the model, and $VE(t_i)$ is the explained variance only on the timepoints where feature *i* was in the model. For numerical stability, the coding scores were bounded between (0,1). Further, if the full model explained less than 0.5% of the variance on the relevant time points, the coding score was set to 0. In Extended Data Fig.

9e we also compute coding scores for reduced models where all features were removed except for the feature of interest. In this case the coding score is computed by: $\frac{VE(t_i)_{reduced,\,i}}{VE(t_i)_{full,\,i}}$.

Statistical tests on coding scores

We used one-way ANOVA, followed by Tukey HSD to compare population coding scores across experience levels (Familiar, Novel 1, Novel >1). A p-value of 0.05 was used as the significance threshold.

Functional clustering

Data Selection

For the functional clustering analysis in Figure 4 we selected neurons that were successfully matched across sessions with each of the three experience levels (Familiar, Novel, Novel+; Extended Data Fig. 2a-c, Extended Data Fig. 3), regardless of their overall activity level or encoding strength. Feature coding of each of these neurons was represented by a 12-element vector of coding scores (see 'unique contribution of model features') describing the four main model components (all images, behavioral, omissions, and task) across three experience levels.

Across Session Normalization of Coding Scores

Our functional clustering analysis is restricted to cells recorded across all three experience levels and specifically seeks to assess changes across experience levels. We therefore normalized the coding scores across sessions to account for changes in the overall full model explained variance for a single cell across sessions.

$$VE(t_i)VE(t_i)maxfull,i\frac{(1-VE(t_i)_{reduced,\,i})}{VE(t_i)_{\max full,i}}$$

Spectral Clustering

We performed a clustering workflow that used spectral clustering in combination with a consensus clustering method[18] to isolate robust clusters that are not influenced by random initial conditions. In spectral clustering, the clusters are determined by their 'connectivity matrix', where nearby points (nodes in the graph) are assigned to the same cluster. We computed the connectivity matrix using a radial basis function. Next, the data was projected into a low-dimensional space using Laplacian embedding, and then each datapoint was assigned a label using the K-means strategy.

This spectral clustering algorithm was run 150 times with optimal k clusters determined as described below (see 'Selecting Optimal Number of Clusters'). To perform consensus clustering on these 150 iterations we computed a co-clustering association matrix and then calculated the probability of any given neuron being assigned to the same cluster with every other neuron. Lastly, we applied hierarchical agglomerative clustering to the association matrix to obtain final cluster labels. We used scikit-learn for clustering analysis.

To answer whether clustering resulted in reduction of variability in the dataset, we computed sum of squared values for clustered and unclustered data: $\sum(\mu - x_i)^2$ Where, for clustered data, μ is the mean of coding scores of cells in that cluster, while for unclustered data μ is the mean of encoding scores for all cells. x_i is an array of encoding scores for each cell. For all three cell classes, clustering significantly reduced the variance in coding scores. However, the reduction in variance was smaller for Vip inhibitory neurons than Sst or excitatory cells (Mean SSE values post-clustering: Excitatory = 0.17, Sst = 0.17, Vip = 0.23).

Selecting Optimal Number of Clusters

For each cell class we estimated the optimal number of clusters using two methods: gap statistic and eigengap values (i.e. spectral gap). To compute the gap statistic, we shuffled the coding features to create a null distribution. In this shuffle, the dropout scores for each regressor and experience level are shuffled within their own category across cells, which preserves the original distributions but breaks structure across days and regressors. Gap statistics were computed by clustering both shuffled and original data 20 times for k clusters range from k = 2 to k = 25. For each k number of clusters, we computed the mean of within cluster variability as measured by pairwise Euclidean distance for cells in each cluster: $\overline{D} = \frac{\sum_n \bar{d}_{(i,j)}}{k}$, where $\overline{D}$ is the grand mean of means of pairwise distances, $\bar{d}$, between points *i* and *j* in each *n* cluster with total k clusters. The peak difference between reference (null hypothesis based on shuffled data) and measured (alternative hypothesis based on original data) values indicates the largest reduction in data variability due to clustering (Extended Data Fig. 15c). To obtain eigengap values, we computed the eigenvalues of the Laplacian graph[19]

(Extended Data Fig. 15c). We used the last non-zero peak of the sorted eigenvalue differences as an optimal number of clusters. Based on inspection of both the gap statistic and eigengap results we determined an optimal number of clusters for each of the three cell classes (excitatory, k = 10; Sst, k = 5; Vip, k = 10).

Cell ID Shuffle

To investigate whether the functional clusters that we obtained from our dataset are cell specific or are a property of coding score distributions across days, we performed a cell ID shuffle analysis (Expended Data Figure 17a). By shuffling cell ID matching across days while preserving coding scores within experience level, we perturbed within cell coding changes across days while maintaining overall effects of novelty at a population level. We then followed the same clustering steps as described above using predetermined number of clusters (Exc k = 10, Sst k = 5 Vip = 10, Extended Data Fig. 17b-d), after each shuffle iteration. On each iteration, post-shuffle clusters and original clusters were matched using lowest sum of squared value (SSE, Extended Data Fig. 17e). If all SSE values exceeded maximum threshold for a match (Exc = 0.10, Sst = 0.15, Vip = 0.15), we concluded that a given original cluster was not found among post-shuffled clusters, and the matched cluster size was set to zero. We then compared mean size of matched post-shuffled clusters with the size of the original clusters to determine the pattern of coding across days in each cluster is equally likely to obtain by chance. Cluster size differences were marked as significant post chi-squared-test with Benjamini-Hochberg correction.

Distribution of Clusters Across Cortical Depth and Areas

To determine whether each neural cluster was evenly distributed across the four test locations (V1/LM at two cortical depths, upper < 250 µm, lower => 250 µm, Figure 4g, Extended Data Fig. 16), we performed a chi-squared test to compare the difference between observed and expected proportions of cells in each cluster. We used the Benjamini-Hochberg procedure to correct for multiple comparisons. To compute proportions, we first normalized number of cells in each cluster and location to the total number of cells in a given location across clusters. Expected proportion of cells per cluster was computed as an average of normalized observed proportions for each cluster.

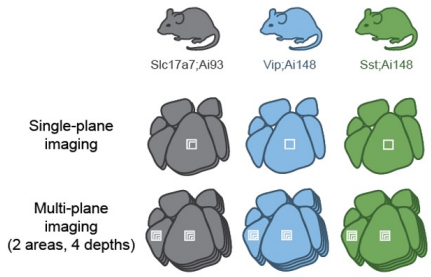
References

1. Madisen, L. *et al.* Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. *Neuron* **85**, 942–958 (2015).
2. Groblewski, P. *et al.* A standardized head-fixation system for performing large-scale, in-vivo physiological recordings in mice. *bioRxiv* (2020).
3. de Vries, S. E. J. *et al.* A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. *Nat. Neurosci.* **23**, 138–151 (2020).
4. Kalatsky, V. a & Stryker, M. P. New paradigm for optical imaging: temporally encoded maps of intrinsic signal. *Neuron* **38**, 529–45 (2003).
5. Garrett, M. E., Nauhaus, I., Marshel, J. H. & Callaway, E. M. Topography and Areal Organization of Mouse Visual Cortex. *J. Neurosci.* **34**, 12587–12600 (2014).
6. Guo, Z. V. *et al.* Procedures for Behavioral Experiments in Head-Fixed Mice. *PLoS ONE* **9**, e88678 (2014).
7. Groblewski, P. A. *et al.* Characterization of Learning, Motivation, and Visual Perception in Five Transgenic Mouse Lines Expressing GCaMP in Distinct Cell Populations. *Front. Behav. Neurosci.* **14**, (2020).
8. Garrett, M. E. *et al.* Experience shapes activity dynamics and stimulus coding of VIP inhibitory and excitatory cells in visual cortex. *eLife* 686063 (2019) doi:10.1101/686063.
9. Groblewski, P. A. *et al.* A standardized head-fixation system for performing large-scale, in vivo physiological recordings in mice. *J. Neurosci. Methods* **346**, 108922 (2020).
10. Siegle, J. H. *et al.* Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature* (2021) doi:10.1038/s41586-020-03171-x.
11. Orlova, N. *et al.* Multiplane Mesoscope reveals distinct cortical interactions following expectation violations. *bioRxiv* 2020.10.06.328294 (2020).
12. Sofroniew, N. J., Flickinger, D., King, J. & Svoboda, K. A large field of view two-photon mesoscope with subcellular resolution for in vivo imaging. *Nat. Methods* **10**, 413–420 (2013).

13. Steinmetz, N. A. *et al.* Aberrant Cortical Activity in Multiple GCaMP6-Expressing Transgenic Mouse Lines. *Eneuro* **4**, ENEURO.0207-17.2017 (2017).
14. Hyvarinen, A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* **10**, 626–634 (1999).
15. Giovannucci, A. *et al.* OnACID : Online Analysis of Calcium Imaging Data in Real Time. 1–28 (2017) doi:10.1101/193383.
16. Jewell, S. & Witten, D. Exact spike train inference via ℓ0 optimization. *Ann. Appl. Stat.* (2018) doi:10.1214/18-AOAS1162.
17. Deneux, T. *et al.* Accurate spike estimation from noisy calcium signals for ultrafast three-dimensional imaging of large neuronal populations in vivo. *Nat. Commun.* **7**, 12190 (2016).
18. Monti, S. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data.
19. von Luxburg, U. A tutorial on spectral clustering. *Stat. Comput.* **17**, 395–416 (2007).

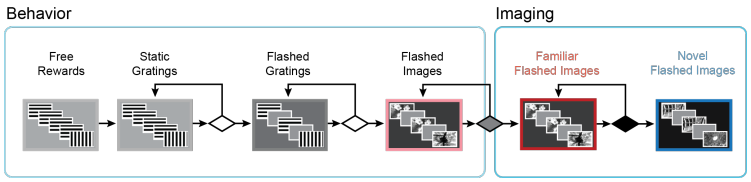
Extended Data Figure 1. Dataset description & behavior training procedure

a

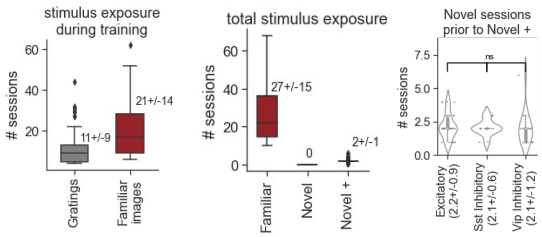


Imaging configuration	Cell class	Areas	Cortical Depths	# Mice	# Imaging planes	# Sessions	# Unique neurons
Single-plane	Excitatory	VISp	175, 375	26	26	78	6,329
	Sst Inhibitory	VISp	275	10	10	30	167
	Vip Inhibitory	VISp	175	11	11	33	350
Multi-plane	Excitatory	VISp, VISl	75, 175, 275, 375	8	36	25	6,497
	Sst Inhibitory	VISp, VISl	75, 175, 275, 375	5	20	16	301
	Vip Inhibitory	VISp, VISl	75, 175, 275, 375	6	31	20	851
			Total	66	134	202	14,495

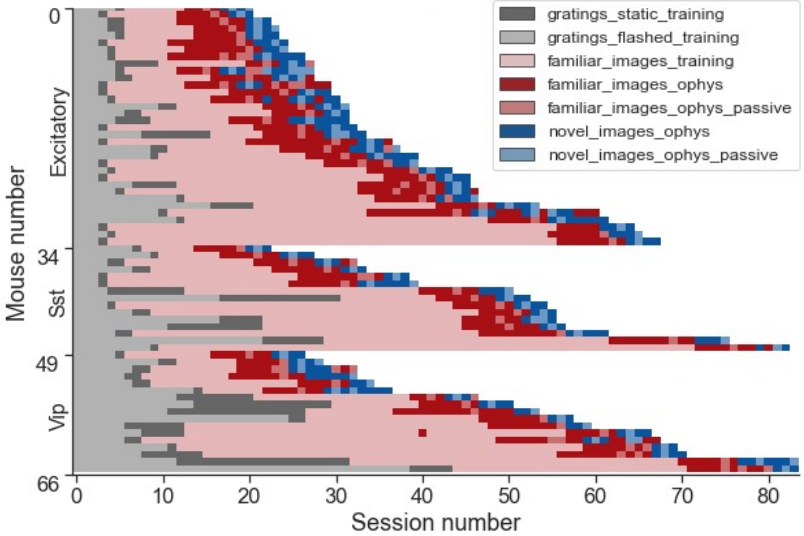
b



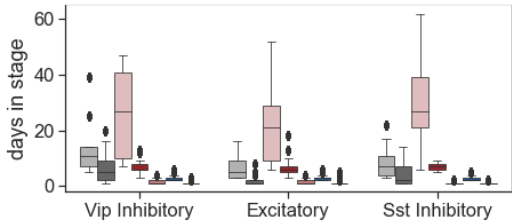
c



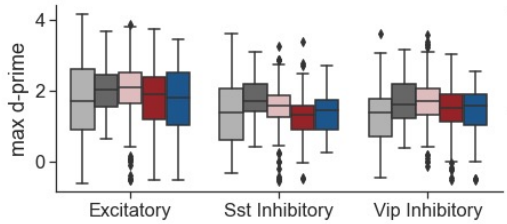
d



e

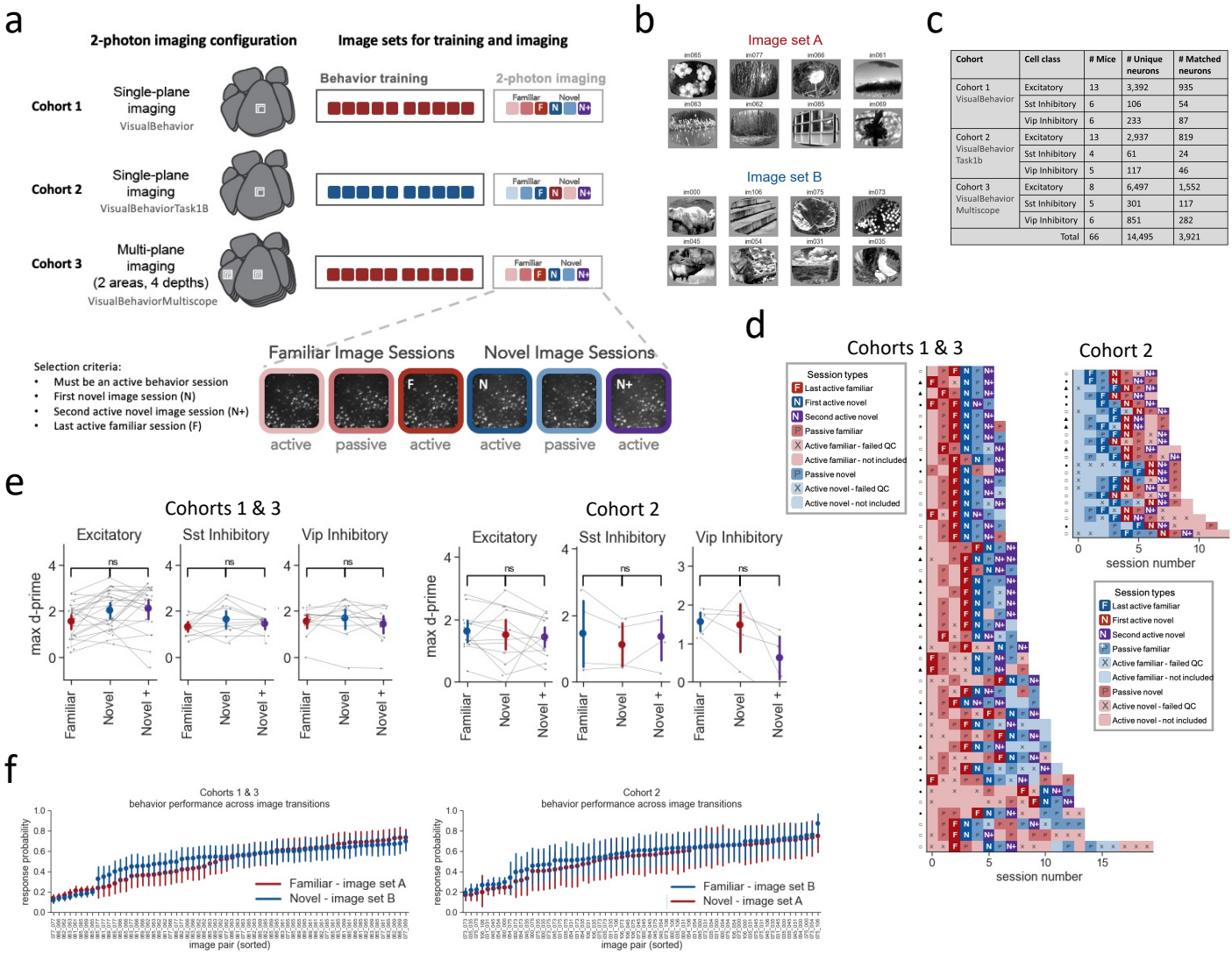


f



Extended Data Fig. 1 | Dataset description and behavior training procedure. a, Schematic and table describing single-plane and multi-plane imaging configurations for each transgenic mouse line. Single plane imaging sessions always occurred in the primary visual cortex, VISp at a single cortical depth per session. Multi-plane imaging sessions consisted of up to 8 imaging planes in two cortical areas, VISp and VISl, across 4 cortical depths per area in each session. **b,** Schematic of behavioral training procedure. Mice are automatically progressed through a series of training stages, beginning with static gratings and ending with flashed natural scene images, before being transitioned to 2-photon imaging. During the 2-photon imaging stage of the experiment, mice perform the task with either familiar or novel images. In some sessions, mice passively view the same stimuli in open loop mode with the lick spout retracted after being given their daily water. **c,** Left, number of sessions with image stimuli and gratings during behavioral training. Middle, number of sessions with familiar or novel images prior to 2-photon imaging sessions of each of the 3 types included in this study (Familiar, first exposure to Novel images, and a subsequent re-exposure to the novel image set, Novel+). Right, number of sessions with exposure to novel images prior to the Novel+ session included in physiology analysis. **d,** Training history for all mice included in the study, sorted by transgenic mouse line and then by total number of sessions per mouse. Colors indicate training stages, as in b. **e,** Number of sessions spent in each behavior stage across transgenic mouse lines (legend shared with panel d). **f,** Behavioral performance, quantified as the maximum d-prime value over the course of a session, across training stages per mouse line (legend shared with panel d).

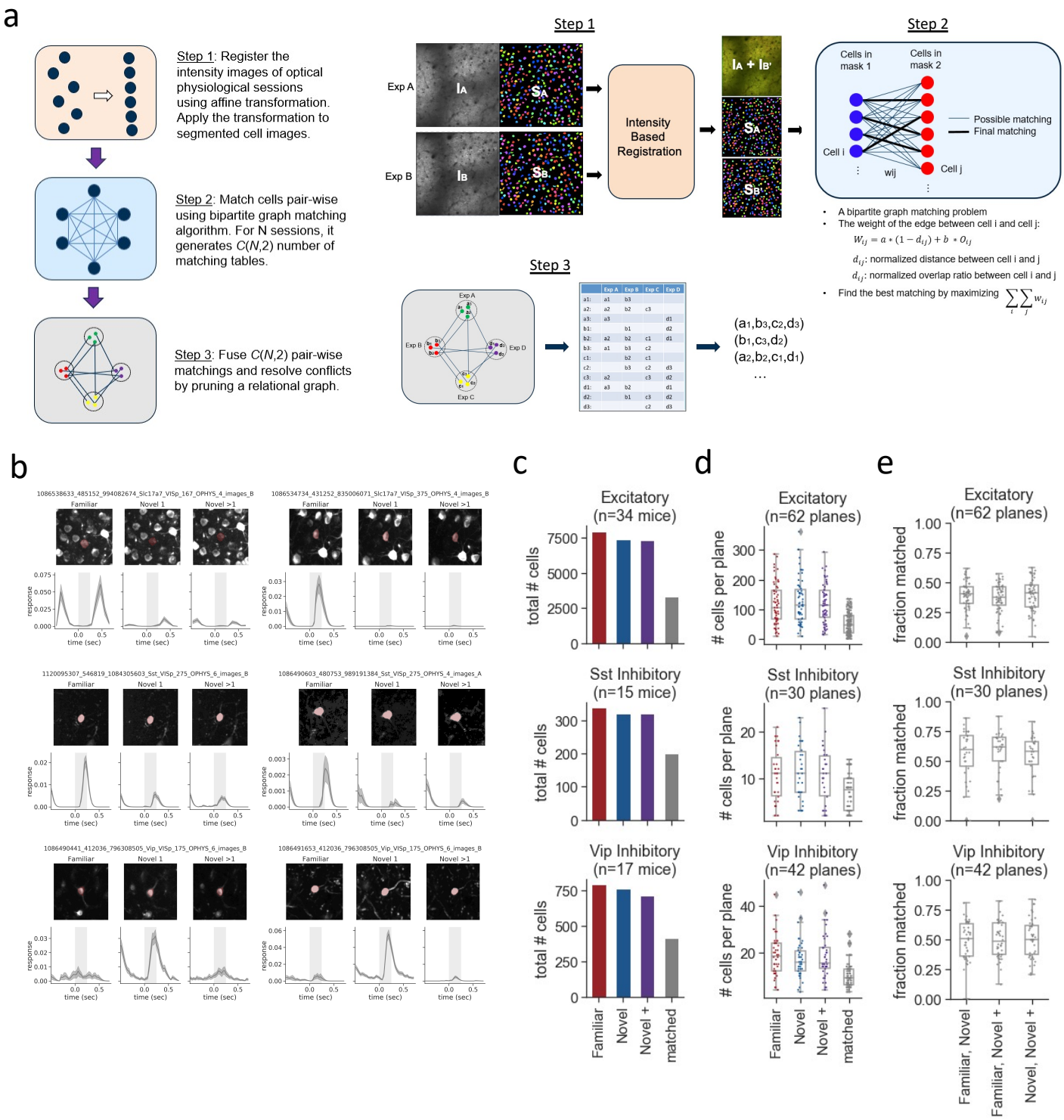
Extended Data Figure. 2. Selection criteria for physiology dataset and behavior during 2P



Cohort	Cell class	# Mice	# Unique neurons	# Matched neurons
Cohort 1 VisualBehavior	Excitatory	13	3,392	935
	Sst Inhibitory	6	106	54
	Vip Inhibitory	6	233	87
Cohort 2 VisualBehavior Task1b	Excitatory	13	2,937	819
	Sst Inhibitory	4	61	24
	Vip Inhibitory	5	117	46
Cohort 3 VisualBehavior Multiscope	Excitatory	8	6,497	1,552
	Sst Inhibitory	5	301	117
	Vip Inhibitory	6	851	282
	Total	66	14,495	3,921

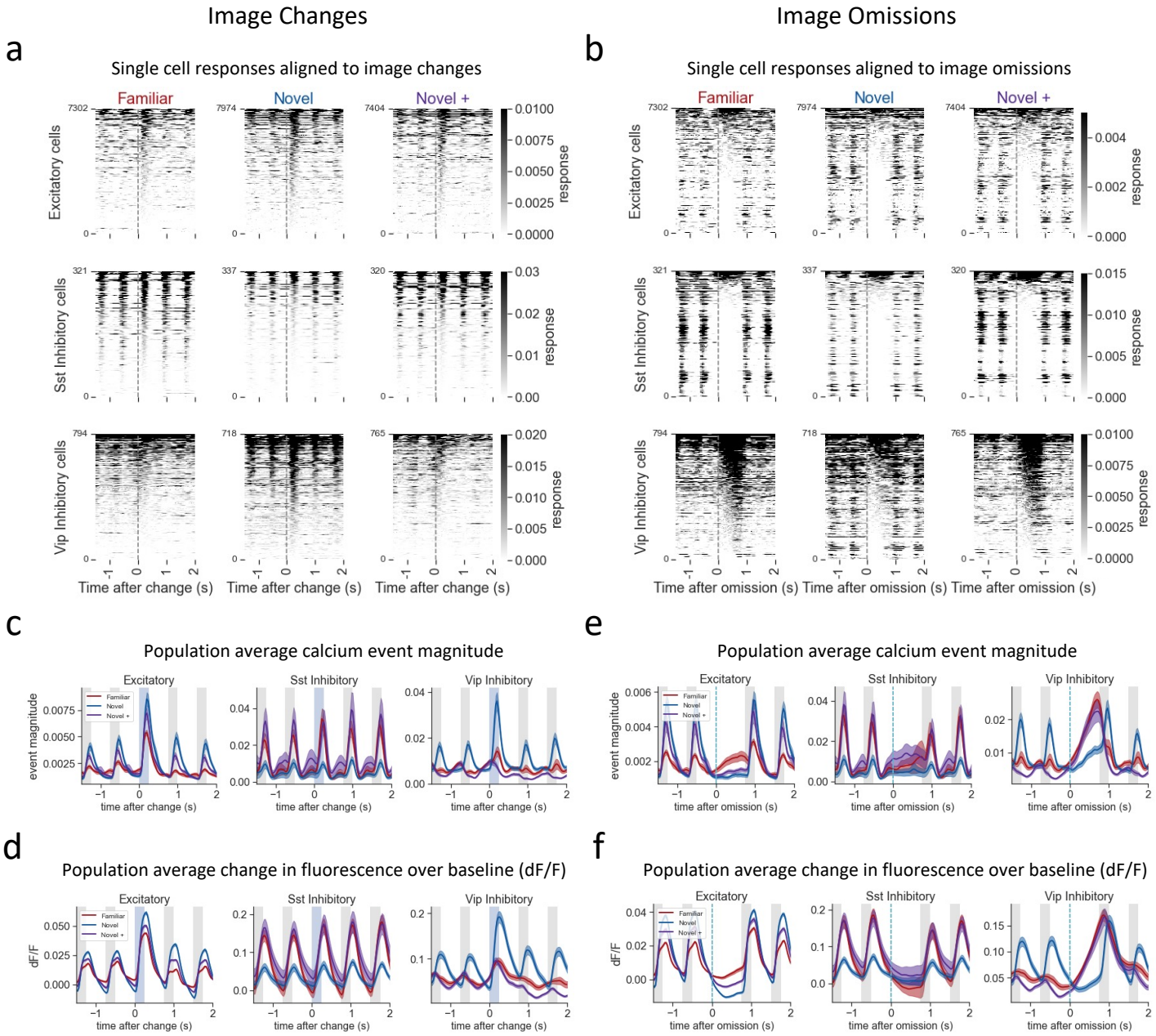
Extended Data Fig. 2 | Selection criteria for physiology dataset and behavior during 2P imaging. a, Schematic depicting the 3 cohorts of mice that were trained and imaged as part of the study, for both single-plane and multi-plane 2-photon imaging configurations. Cohorts 1 and 3 were trained using image set A (red), and cohort 2 was trained using image set B (blue). During imaging, active behavior sessions were interleaved with passive viewing sessions where mice viewed the stimulus in open loop mode with the lick spout removed after mice had received their daily water. Only active behavior sessions meeting specific criteria were included in analysis (with the exception of Extended Data Fig. 7a,b). **b**, Each image set consisted of 8 natural scene images. Image set A (red) was the familiar set for cohorts 1 & 3, image set B (blue) was the familiar set for cohort 2. **c,** For each transgenic line in each cohort (cohorts described in panel a), table describing numbers of mice, unique neurons (total segmented cells across all 3 sessions), and matched neurons (cells detected in all 3 sessions) for the session types included in the study (Familiar, Novel, Novel+). **d,** 2-photon imaging session sequence for all mice (rows), highlighting the Familiar, Novel, and Novel+ sessions that were used for analysis (darker colors, see legend). Top, cohorts 1 & 3, mice trained with image set A (red), image set B (blue) is novel. Bottom, cohort 2, mice trained with image set B (blue), image set A (red) is novel. Symbols on y-axis indicate cell population labeled with GCaMP6f in each mouse; square: Excitatory, triangle: Sst Inhibitory, circle: Vip Inhibitory. Only active behavior sessions were included. Passive sessions (indicated with letter P) were interleaved among active behavior sessions but were not used in the primary analysis for this study (but see Extended Data Fig. 7a). Sessions selected for analysis include the last active behavior session with Familiar images (F), the first novel image session (N), and a subsequent novel image session (N+). 2-photon imaging sessions that did not pass quality control (QC) are indicated with an X. **e,** Behavior performance, quantified as the maximum d-prime value for each session, across Familiar, Novel, and Novel+ experience levels for each cohort, demonstrating that performance is consistent across familiar and novel images and does not depend on which set was used for training. Gray lines show individual mice. Colors indicate image sets and quantify the average +/- 95% confidence intervals across mice. **f,** Response probability, during engaged periods, for each of the 64 possible image transitions for image set A (red) and image set B (blue) in the different cohorts of mice. Cohorts 1 & 3 were trained with image set A and cohort 2 was trained with image set B. Image pairs on x-axis are sorted by mean response probability. The image pairs with the highest detectability were selected for quantification of d-prime in Fig. 1d to facilitate comparison of behavior for similarly detectable images.

Extended Data Figure 3 – Across session cell matching



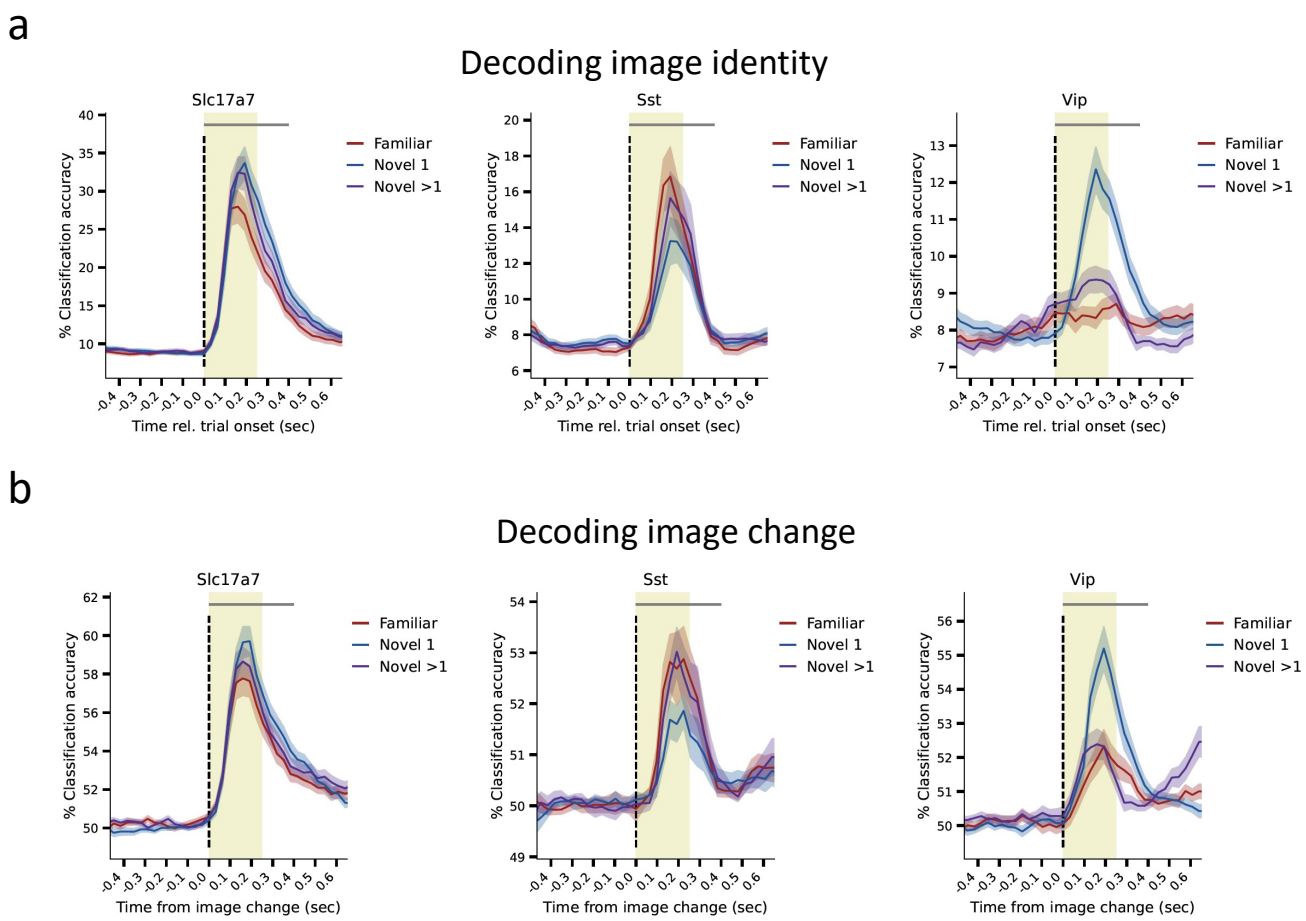
Extended Data Fig. 3 | Across session cell matching. **a,** Schematic of cell matching procedure. Average projection images from a pair of 2-photon imaging sessions are aligned to each other and the registration transform is applied to the segmented ROI masks (step 1). A bi-partite graph matching algorithm is used to identify the best ROI match between the two sessions using intersection over union (IOU) of the ROI masks, and the inter-centroid distance with a maximum allowable distance of 10 pixels (step 2). Each pair of sessions was registered and matched, and cell specimen IDs were unified across all pairwise matches (step 3). **b,** Example images of matched cells and average image responses in each of the 3 sessions. two example cells (columns) shown for each cell class (rows: Excitatory, Sst, Vip). **c,** Total number of cells detected in each session (Familiar, Novel, Novel +) and matched across all 3 sessions. **d,** Number of cells detected in each session and matched across all sessions for each imaging plane (each point is one imaging plane). **e,** Fraction of cells matched for each pair of sessions for each cell class. Individual points are unique imaging planes.

Extended Data Figure 4 – Average cell traces aligned to changes and omissions



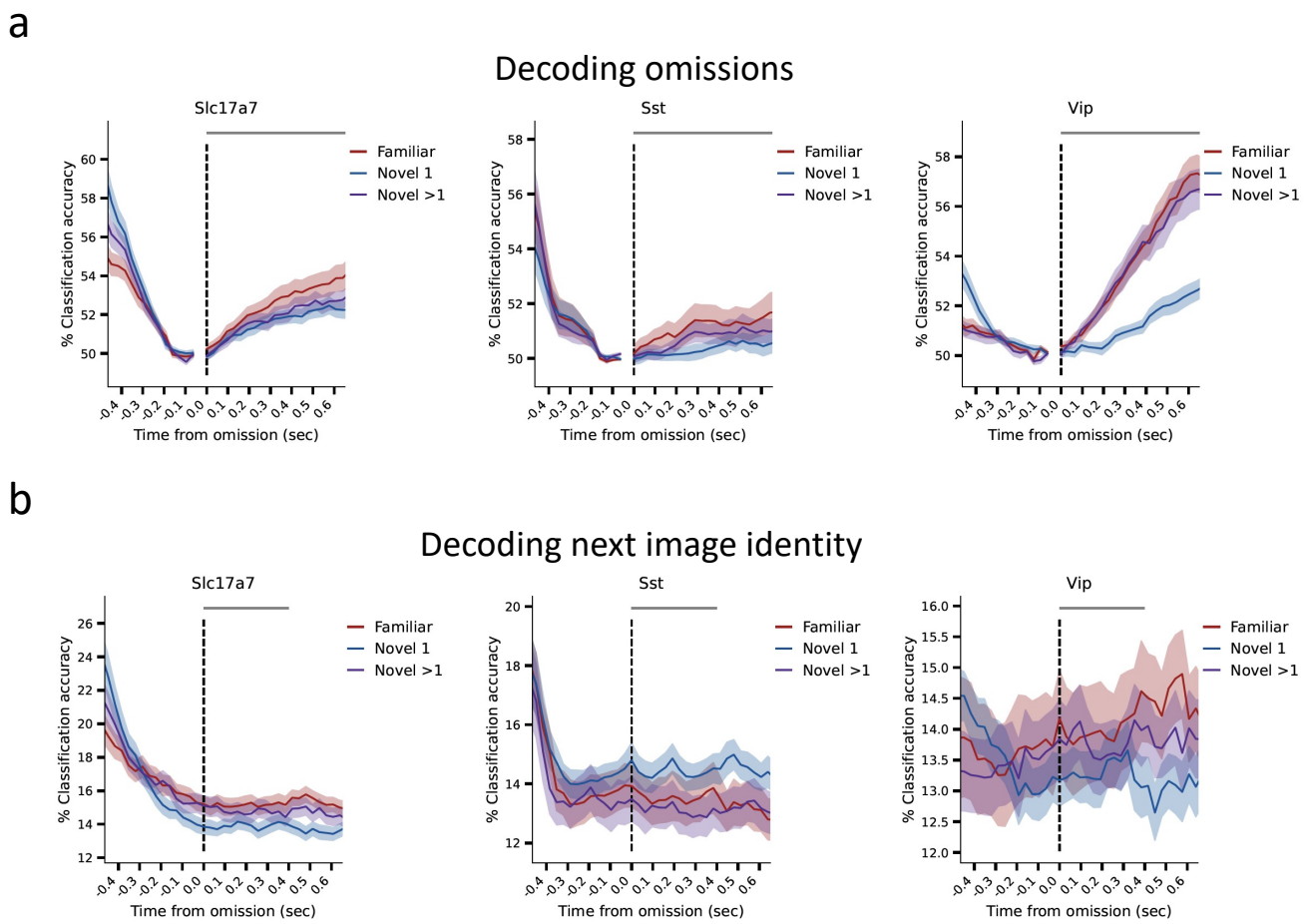
Extended Data Fig. 4 | Average cell traces aligned to changes and omissions. a, Average response aligned to image changes for each cell (y-axis) in a [-1.5, 2] second window (x-axis) relative to the time of the change. Rows correspond to each cell subclass (Excitatory, Sst, Vip), and columns correspond to experience levels (familiar, novel, novel+). Responses are quantified as event magnitudes, in arbitrary units. **b,** Average response aligned to image omissions for each cell (y-axis) in a [-1.5, 2] second window (x-axis) relative to the time where the omitted stimulus would have been presented. Responses are quantified as event magnitudes, in arbitrary units. **c,** Population average change aligned activity, quantified using the timeseries of event magnitudes detected from calcium fluorescence signals. Population averages correspond to the average of all cells in panel a, split by experience level and cell class. **d,** Population average change aligned activity, quantified using normalized fluorescence values (dF/F, change in fluorescence over baseline). **e,** Population average omission aligned activity, quantified using the timeseries of event magnitudes detected from calcium fluorescence signals. Population averages correspond to the average of all cells in panel a, split by experience level and cell class. **f,** Population average omission aligned activity, quantified using normalized fluorescence values (dF/F, change in fluorescence over baseline). Differences across experience levels are consistent across methods of signal quantification.

Extended Data Figure 5. Decoding images and changes



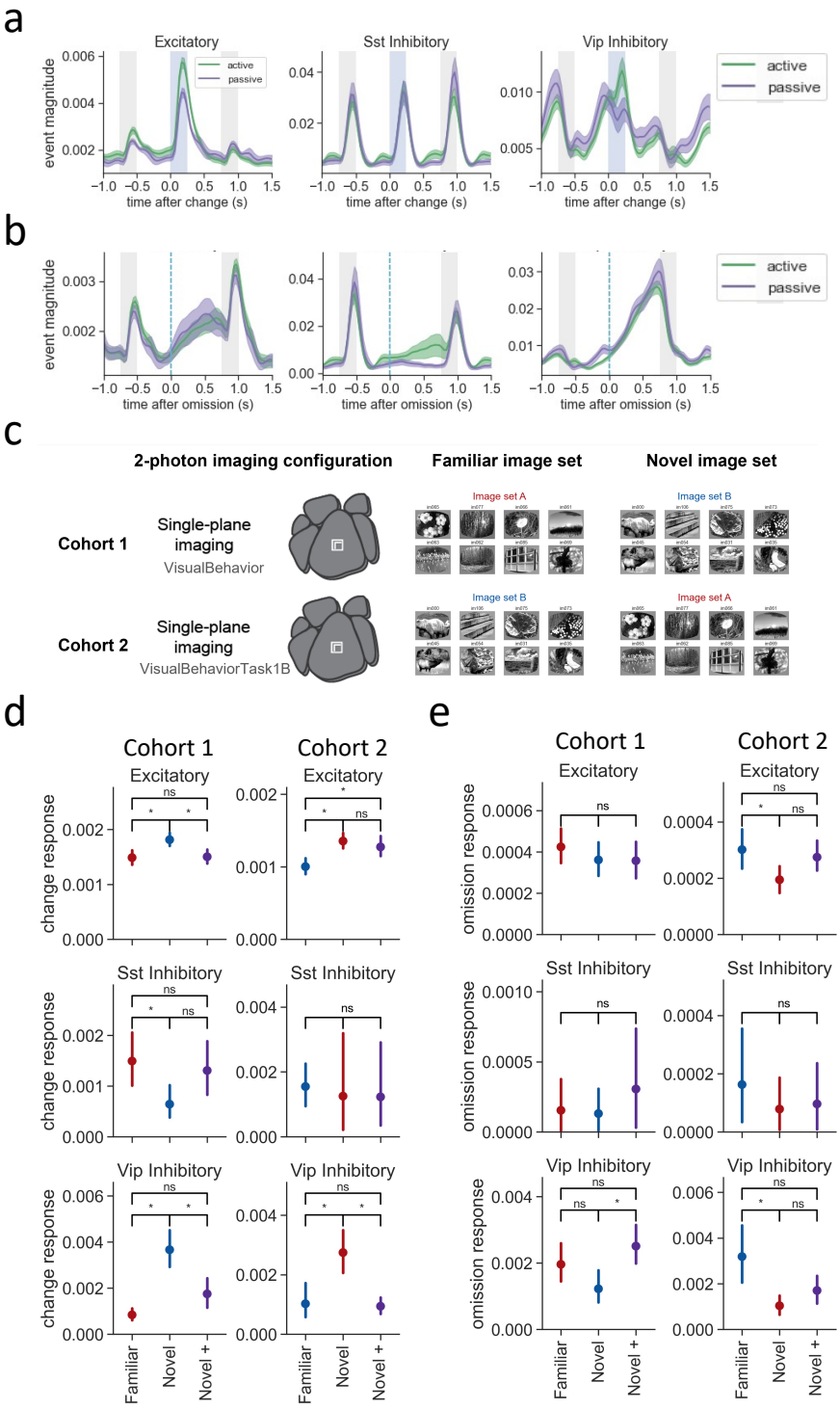
Extended Data Fig. 5 | Decoding image identities and image changes from image-aligned traces. A separate decoder was trained at each time point of the image-aligned traces to acquire a time course for the decoding analysis. Data shown for each experience level (red: familiar; blue: 1st novel session; purple: subsequent novel session), and each cell class (left: Excitatory; middle: Sst right: Vip). **a,** Decoding image identity using an 8-class classifier. Image-change-aligned traces for the 8 image types in each session were used in the analysis. A separate decoder was trained at each time point to classify the population activity as one of the 8 images. Population activity contains information about image identity within a 100 ms after the image is presented. This is the case in all classes, except for VIP neurons in the familiar and novel+ sessions. Note, the baseline is not exactly at the expected chance value (i.e. 1/8), because of the different number of trials per image. **b,** Decoding images changes from non-changes using a binary classifier. For the non-change image, the image immediately before the image change was used. A separate decoder was trained at each time point. Decoder was trained to classify the population activity at each time point as image change or non-change. Image change aligned traces and non-change aligned traces represented the two classes in the decoder.

Extended Data Figure 6. Decoding omissions and images after omissions



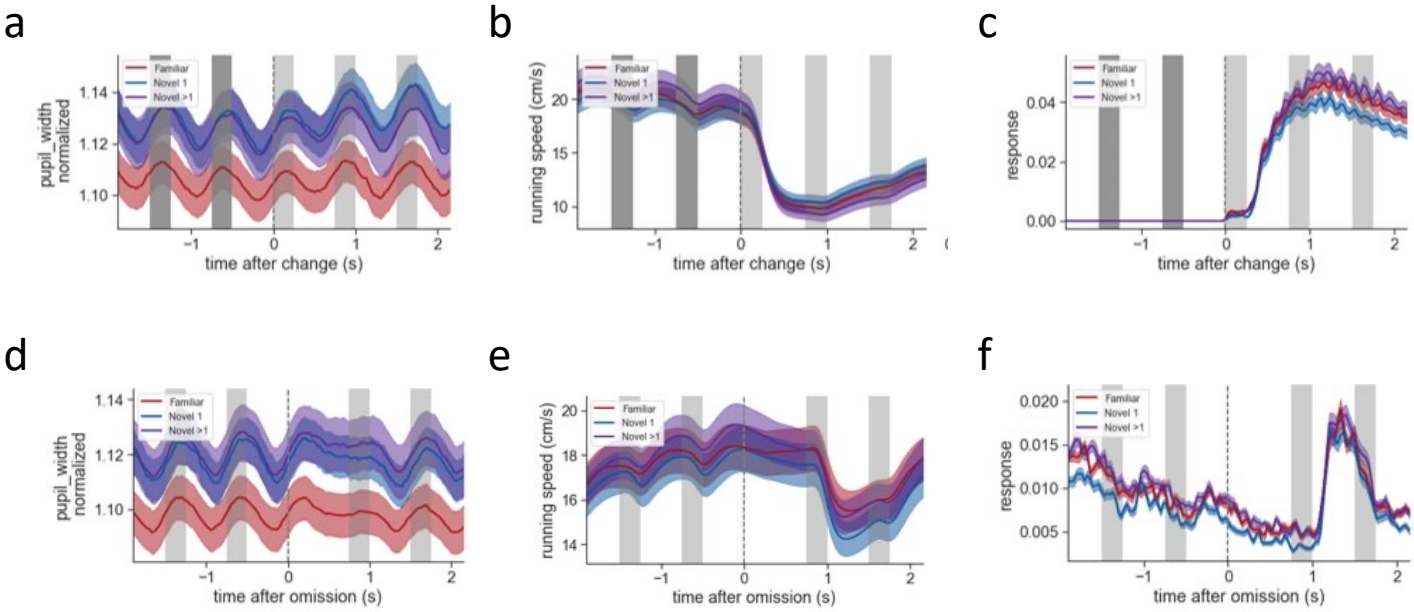
Extended Data Fig. 6 | Decoding omissions and image identities from omission-aligned traces. a, Decoding analysis to classify activity from omissions vs. gray screen epochs. To get a time course for the decoding analysis, a separate decoder was trained on each time point of the omission-aligned traces. Decoders were trained to perform a binary classification of the population activity at each time point relative to the population activity at baseline. Baseline activity was determined from the time point immediately preceding the omission onset. VIP population activity showed was most informative about omissions in familiar sessions. Excitatory population activity contained weaker information about omissions. **b,** 8-class classification was performed to assess whether population activity at each time point following omissions contained information about the image identity following omissions. It was not possible to decode image identity from omissions in any of the cell populations.

Extended Data Figure 7 – Passive vs. active comparison and reverse image set controls



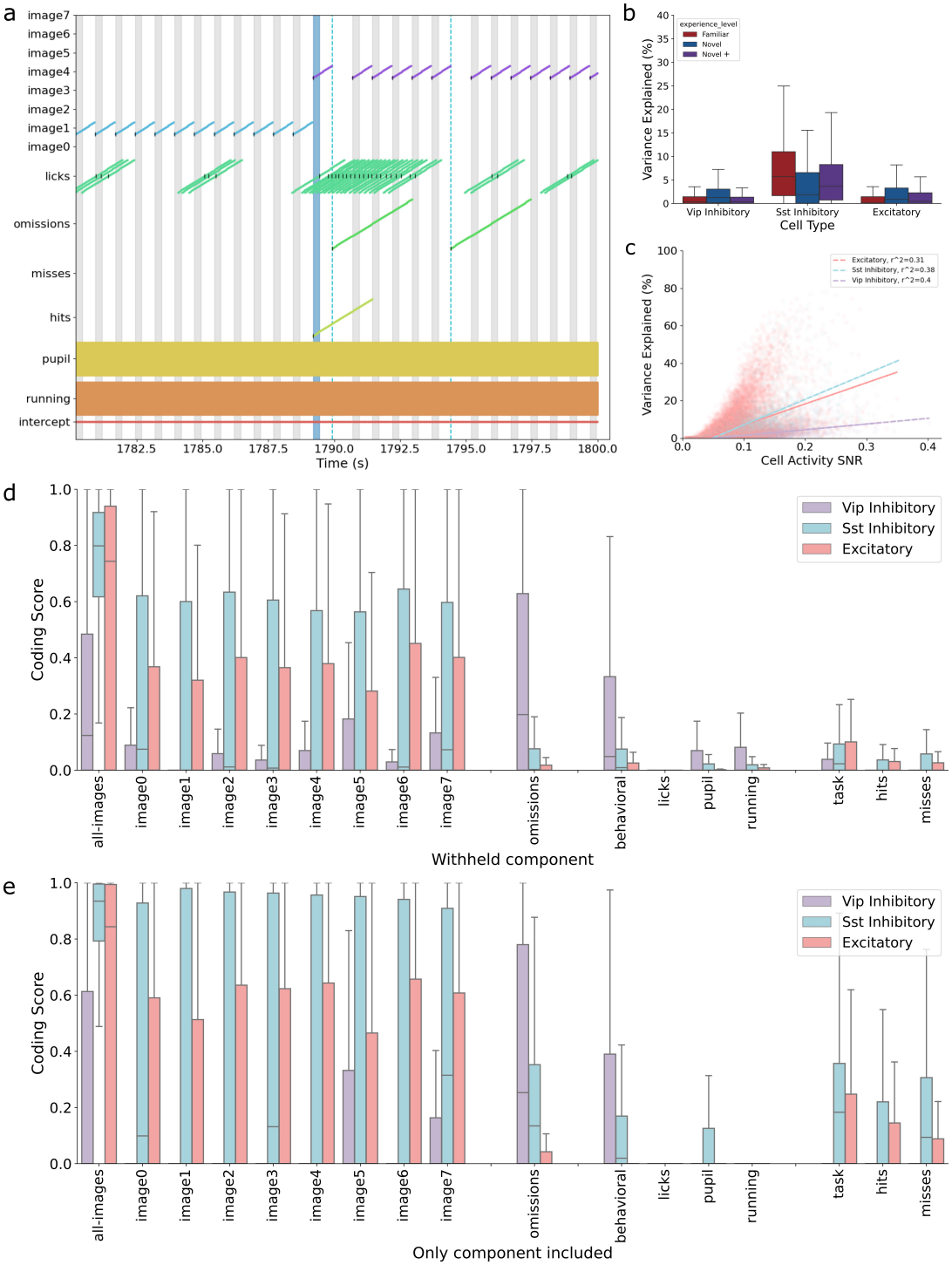
Extended Data Fig. 7 | Passive vs. active comparison and reverse image set controls. a, Image change-aligned population average responses for each cell class, comparing active behavior (green) and passive viewing (purple) sessions with familiar images. In passive viewing sessions, the change detection task was played in open loop mode, with the lick spout retracted such that no reward was available. Mice were given their daily water allocation prior to the session so they were not thirsty and motivated. **b,** Omission aligned population average responses for each cell class, comparing active behavior (green) and passive viewing (purple) sessions with familiar images. VIP omission response does not depend on task performance. **c,** Schematic of cohorts of mice trained with different image sets. Cohort 1 was trained with image set A which became highly familiar, then tested with image set B as the novel set. Cohort 2 was trained with image set B which became highlight familiar, then tested with image set A as the novel set. All recordings for both cohorts were in VISp. **d,** Average response to image changes for each cell class (rows) and cohort (columns). Activity is measured as average event magnitude in a 500ms window after image changes, averaged over all neurons across mice in each cohort. Colors correspond to image sets as in panel c. Differences in activity between familiar and novel image sets are consistent between cohorts and do not depend on which image set was used for training. **e,** Average response to image omissions for each cell class (rows) and cohort (columns). Activity is measured as average event magnitude in a 750ms window after the time the omitted stimulus would have been presented, averaged over all neurons across mice in each cohort. Colors correspond to image sets as in panel c.

Extended Data Figure 8. Average pupil, running and licking across experience



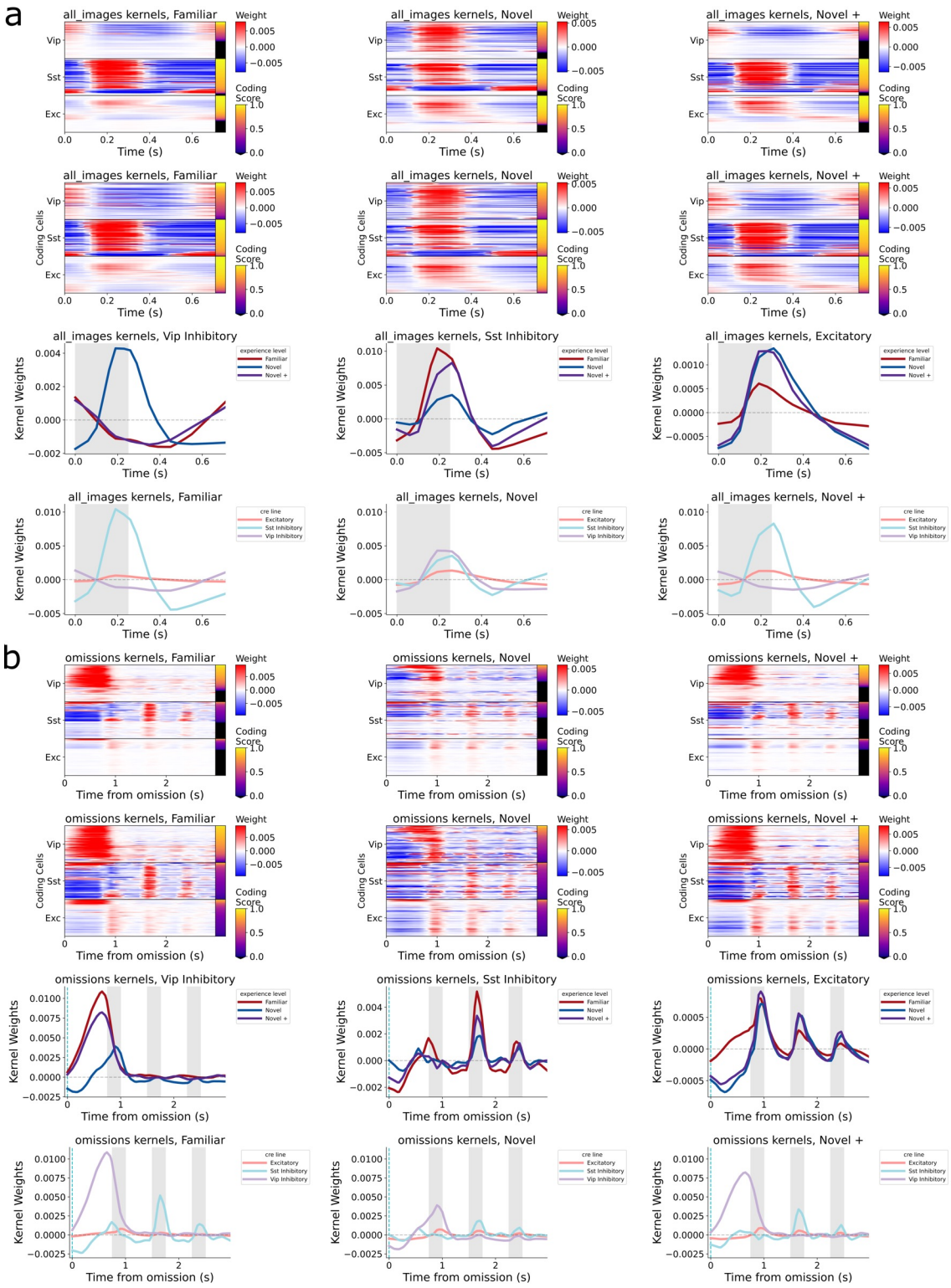
Extended Data Fig. 8 | Average pupil, running and licking across experience levels. **a,** Image change aligned pupil width averaged across sessions for each experience level (Familiar, Novel, Novel +). Pupil width during behavior performance is normalized to the average pupil diameter during the 5-minute gray screen period prior to the start of the stimulus. Pupil width is larger in Novel and Novel + sessions. **b,** Average change aligned running speed averaged across sessions for each experience level. **c,** Change aligned lick rate across experience levels. **d,** Omission aligned pupil width averaged across sessions for each experience level. Overall pupil width is larger during Novel and Novel+ sessions but the dynamics of the pupil response are consistent. **e,** Average omission aligned running speed across experience levels. **f,** Lick rate aligned to omissions across experience levels. Mice are more likely to lick in response to the image presentation directly after a stimulus omission.

Extended Data Figure 9. GLM Validation



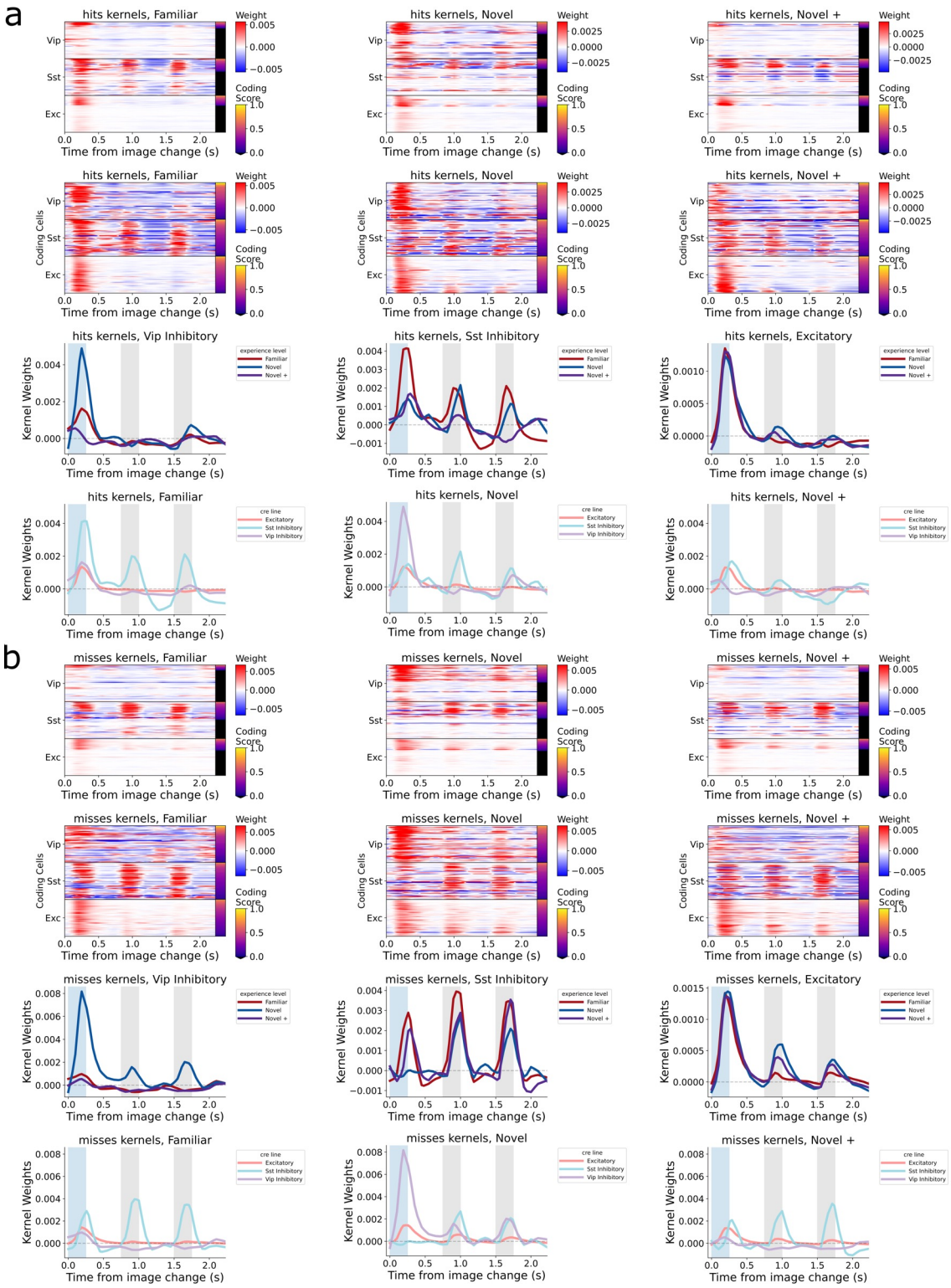
Extended Data Fig. 9 | GLM validation. **a,** Illustration of kernel length and timing for each feature. Kernels for each of the 8 images in each image set start at image onset (gray bands mark image presentations) and extend for 750ms. Kernels for licking are aligned to each detected lick (black tick marks) and extend forward and backward in time 1s. Kernels for omissions are aligned to the time point when each image would have been presented (dashed blue line) and extend for 3s. Kernels for hits and misses are aligned to the onset of the change image (blue band marks change image presentation) and extend for 2.25s. Kernels for running speed and pupil diameter are continuously convolved with the data timeseries and extend forward and backward in time 1s. **b,** Distribution of explained variance for each cell type split by experience level. The box shows the quartiles of each cell class and the whiskers mark +/- 1.5x Interquartile Range (Q3-Q1). Outliers are not shown for clarity. **c,** The explained variance of the GLM for each cell against the signal-to-noise ratio (SNR) of the neural activity trace. Here SNR is the mean of each cell's df/f signal divided by its standard deviation. Lines show the best linear fit. **d,** Distribution of coding scores for each model feature, split by cell type. **e,** Distribution of coding scores computed by refitting reduced models with only a single feature. **d,e,** The box shows the quartiles of each cell class and the whiskers mark +/- 1.5x Interquartile Range (Q3-Q1). Outliers not shown for clarity.

Extended Data Figure 10. Image and Omission kernels



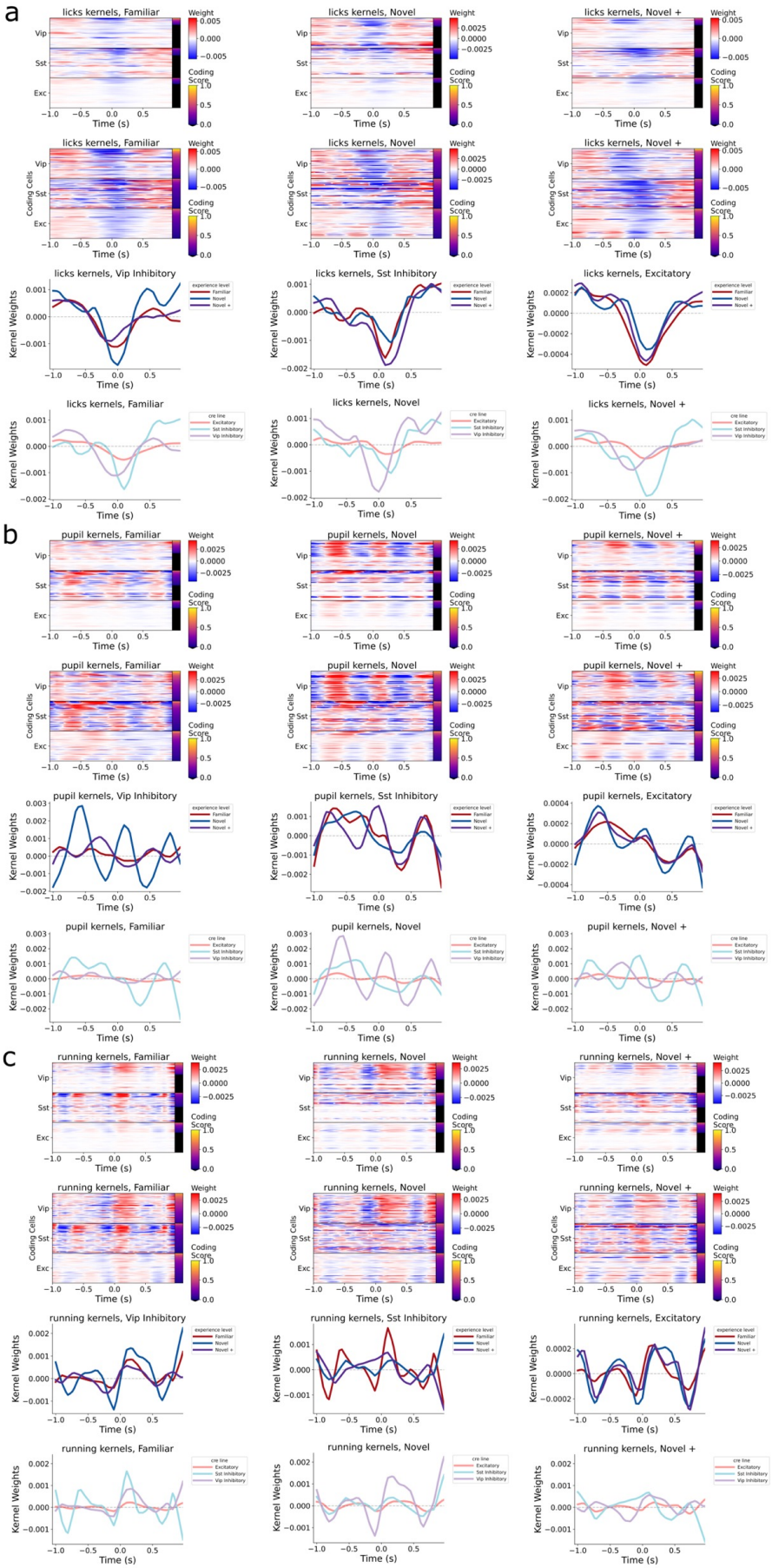
Extended Data Fig. 10 | Image and omission kernels for all cells. a, Image kernels across experience. Top row, heatmap of learned image kernels for all cells. Cells are sorted by cell class, and then by image coding score. Second row, same as top row, but only cells with non-zero image coding scores. Third row, average image kernels split by experience level (colors) for each cell class (columns). Bottom row, average image kernels split by cell class (colors), for each experience level (columns). **b,** Omission kernels across experience. Top row, heatmap of learned omission kernels for all cells. Cells are sorted by cell class, and then by omission coding score. Second row, same as top row, but only cells with non-zero omission coding. Third row, average omission kernels split by and experience level (colors), for each cell class (columns). Bottom row, average omission kernels split by cell class (colors), for each experience level (columns).

Extended Data Figure 11. Hit and Miss kernels



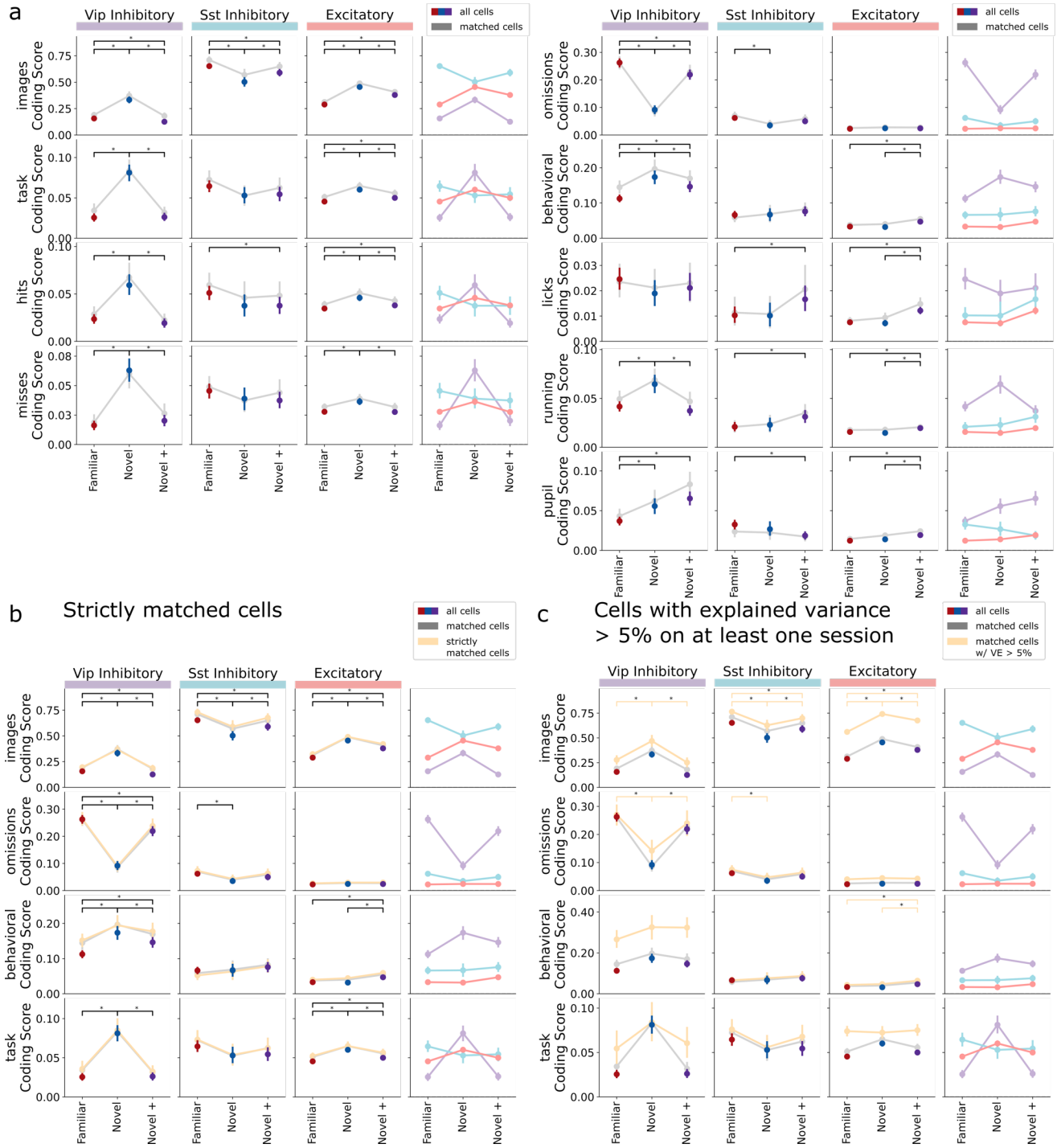
Extended Data Fig. 11 | Task kernels for all cells. a, Hit kernels across experience. Top row, heatmap of learned hits kernels for all cells. Cells are sorted by cell type, and then hits coding score. Second row, same as top row, but only cells with non-zero hit coding scores. Third row, Average hits kernels split by cell class and experience, for each cell class. Bottom row, average hits kernels split by cell class and experience, for each experience level. **b,** Miss kernels across experience. Top row, heatmap of learned miss kernels for all cells. Cells are sorted by cell class, then by miss coding score. Second row, same as top row, but only cells with non-zero miss coding scores. Third row, average miss kernels split by cell class and experience, for each cell class. Bottom row, average miss kernels split by cell type and experience, for each experience level.

Extended Data Figure 12. Licking, pupil, and running kernels



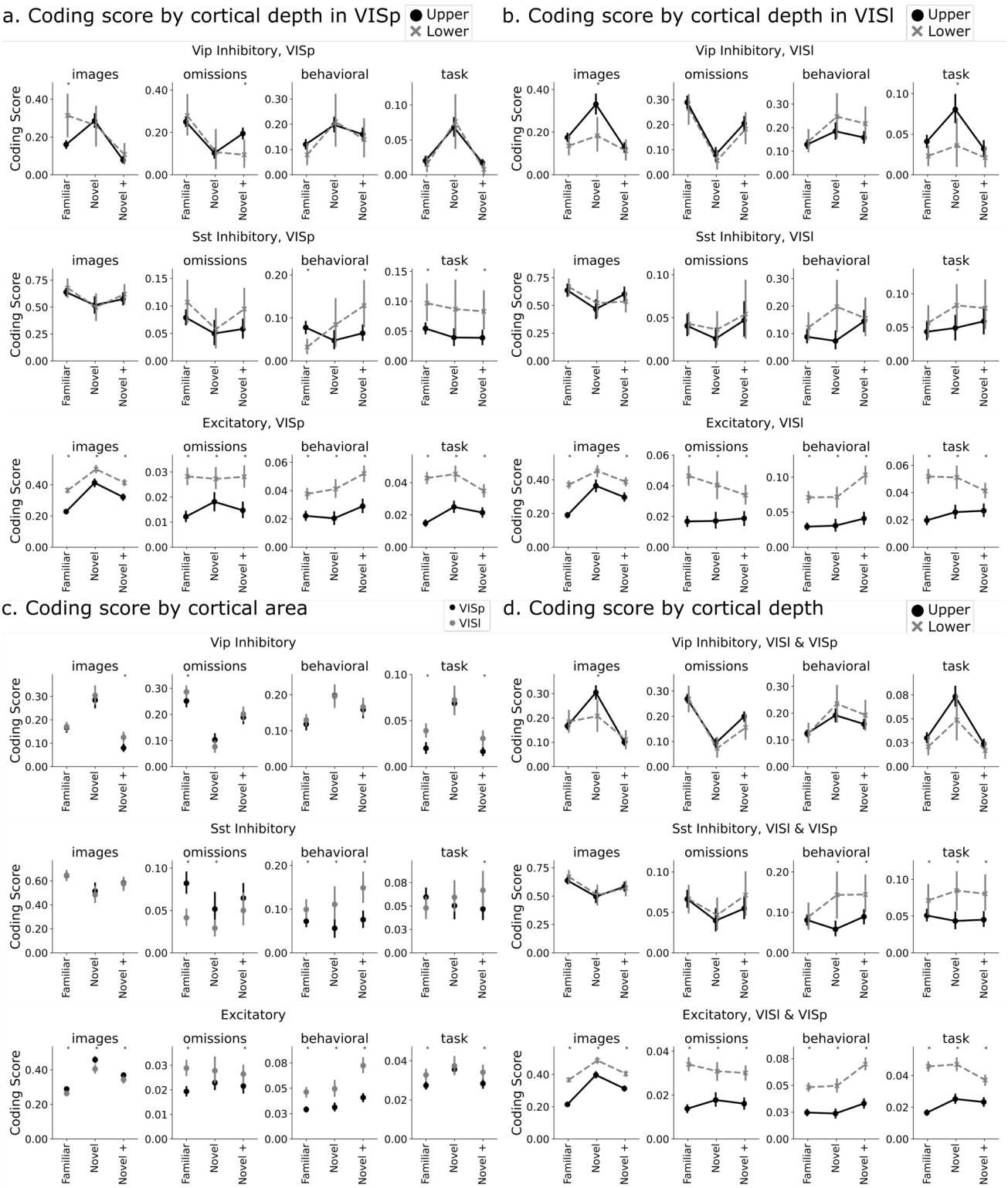
Extended Data Fig. 12 | Behavior kernels for all cells. a, Lick kernels across experience. Top row, heatmap of learned lick kernels for all cells. Cells are sorted by cell type, and then by lick coding score. Second row, same as top row, but only cells with non-zero lick coding scores. Third row, average lick kernels split by experience level (colors), grouped by cell class (columns). Bottom row, average lick kernels split by cell class (colors), grouped by experience level (columns). **b,** Pupil diameter kernels across experience. Top row, heatmap of learned pupil diameter kernels for all cells. Cells are sorted by cell class, and then pupil diameter coding score. Second row, same as top row, but only cells with non-zero pupil diameter coding scores. Third row, average pupil diameter kernels split by experience level, (colors) grouped by cell class (columns). Bottom row, average pupil diameter kernels split by cell type (colors), grouped by experience level (columns). **c,** Running speed kernels across experience. Top row, heatmap of learned running speed kernels for all cells. Cells are sorted by cell type, and then by running speed coding score. Second row, same as top row, but only cells with non-zero running speed coding scores. Third row, average running speed kernels split by experience level (colors), grouped by cell class (columns). Bottom row, average running speed kernels split by cell type (colors), grouped by experience level (columns).

Extended Data Figure 13. Additional details on coding scores



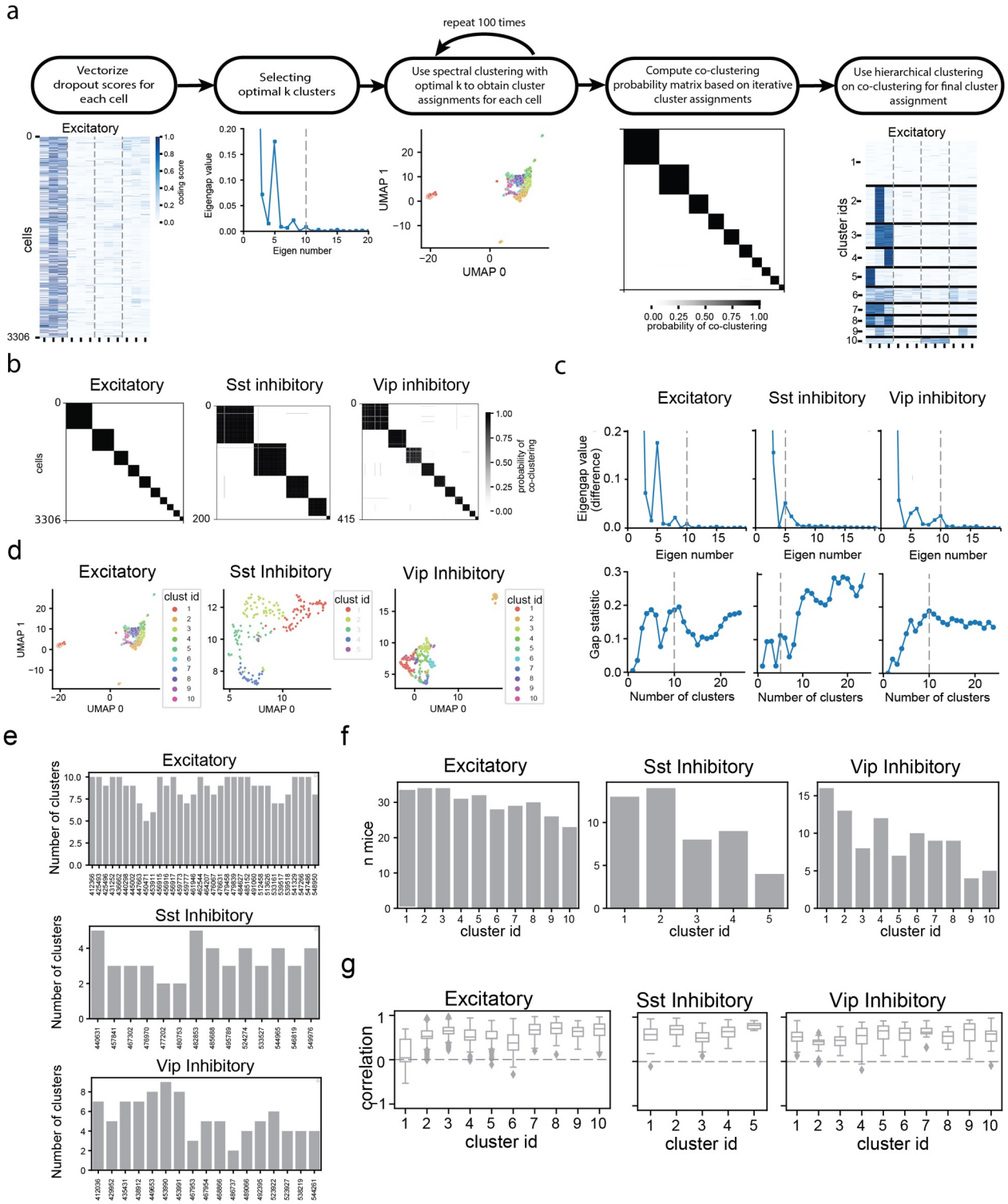
Extended Data Fig. 13 | Coding scores for all model features by experience. a, Same as Fig. 3e for all model features. Average coding score for each cell class (columns) by experience level and model component (rows). Colored points show all cells (mean +/- 95% CI). Connected gray lines show only cells matched across all experience levels. Differences between experience levels are marked as significant after a one-way ANOVA, followed by Tukey HSD (p = 0.05). Statistics were computed on all cells. **b,** Same as Fig. 3e with the addition of strictly matched cells that were matched across all experience levels including the last familiar session, and the first novel repeat. **c,** Same as Fig. 3e with the addition of cells that were matched across all experience levels and the GLM explained at least 5% of the cell's variance on at least one session.

Extended Data Figure 14. Coding scores by cortical depth and area



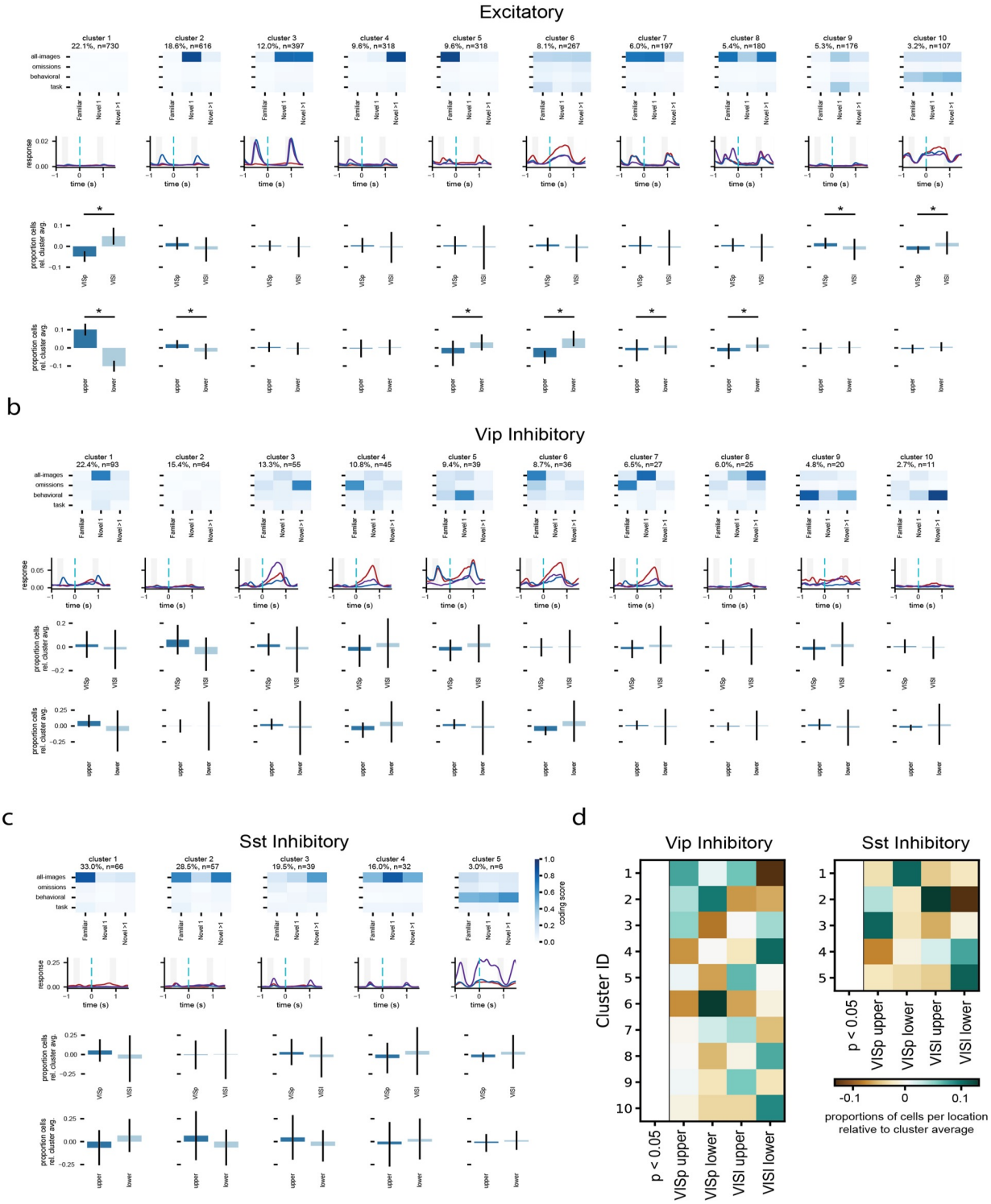
Extended Data Fig. 14 | Coding scores by cortical area and depth. a, Same as Fig. 3e, split by cortical depth within V1, **a,** Same as Fig. 3e, split by cortical depth within LM**. a & b**, Average coding score for each cell class (rows) by experience level and model component (rows) (mean +/- 95% CI, all cells). Differences between cortical depth are marked as significant with the result of a two-sided independent sample t-test ($p < 0.05$). Cells were divided into upper and lower layers of the cortex using a dividing line of 250um below the cortical surface. **c,** Same as Fig. 3e split by cortical area. Average coding score for each cell class (rows) by experience level and model component (rows) (mean +/- 95% CI, all cells). Differences between cortical areas are marked as significant with the result of a two-sided independent sample t-test ($p < 0.05$).

Extended Data Figure 15 – Clustering methods



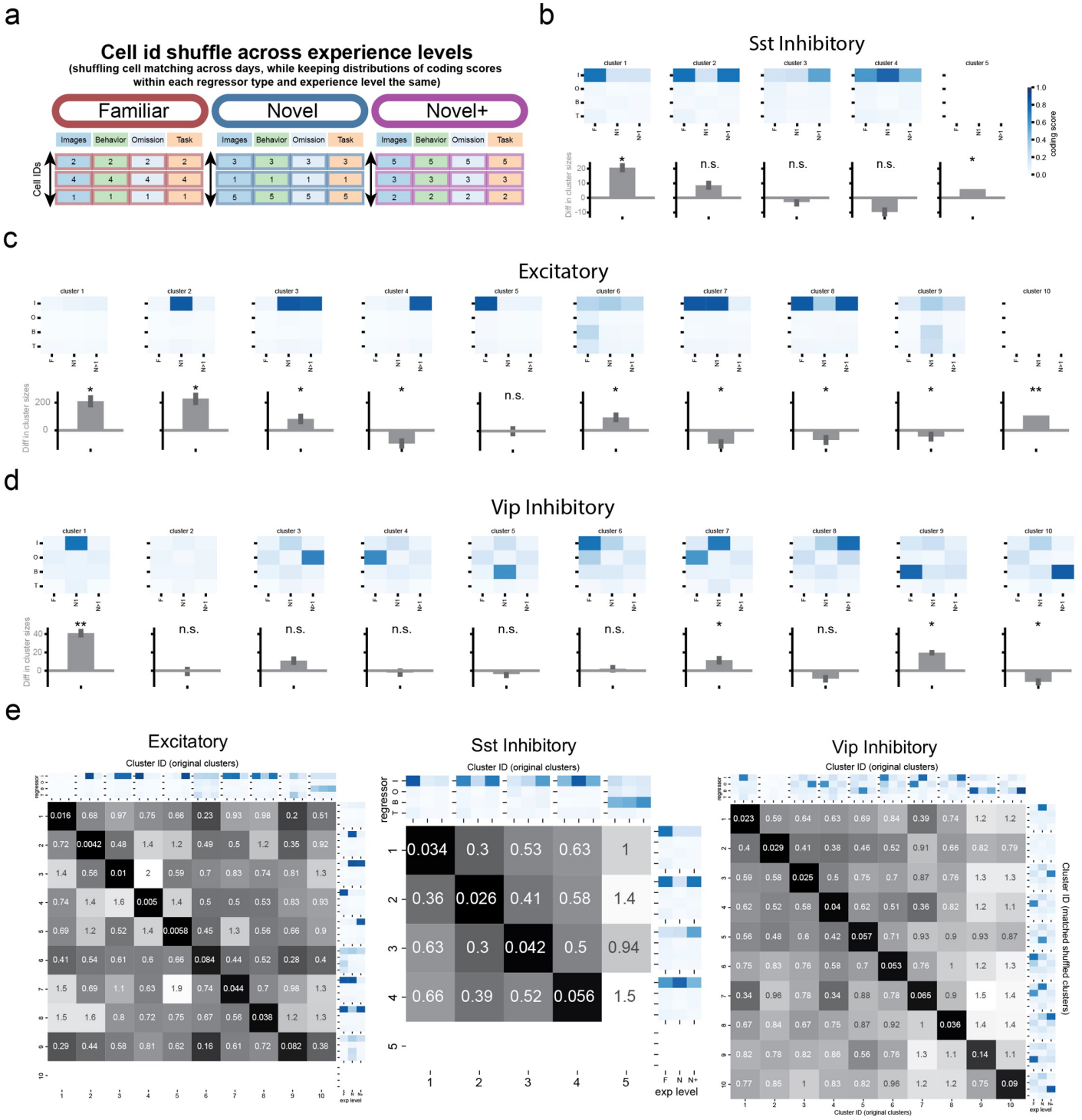
Extended Data Fig. 15 | Clustering methods. a, Schematic of clustering steps. Spectral clustering is used to cluster vectorized coding scores for each cell class separately. **b,** Co-clustering probability matrices for three cell classes used for final assignment of cluster IDs. Darker color indicates higher probability of cells belonging to the same cluster after 150 iterations. **c,** Top row, eigengap values for three cell classes, excitatory, Sst, Vip, used to choose optimal number of clusters. Gray dashed lines indicate number of clusters identified for each cell class. Bottom row, gap statistic value for all cell classes. **d,** UMAP representation of functional clusters for three cell classes. Colors represent distinct clusters. **e,** Bar plot representing number of clusters that were identified across mice. Multiple clusters were identified in mouse animal, and no one mouse contributed to only one cluster, indicating that functional clusters are not mouse specific. **f,** Bar plot showing number of mice that contributed to each functional cluster demonstrating that each cluster was comprised of cells from multiple mice. **g,** Within cluster correlation of individual cell coding scores with the mean coding score for each cluster.

Extended Data Figure 16 – Average coding scores for all clusters and layer distributions.



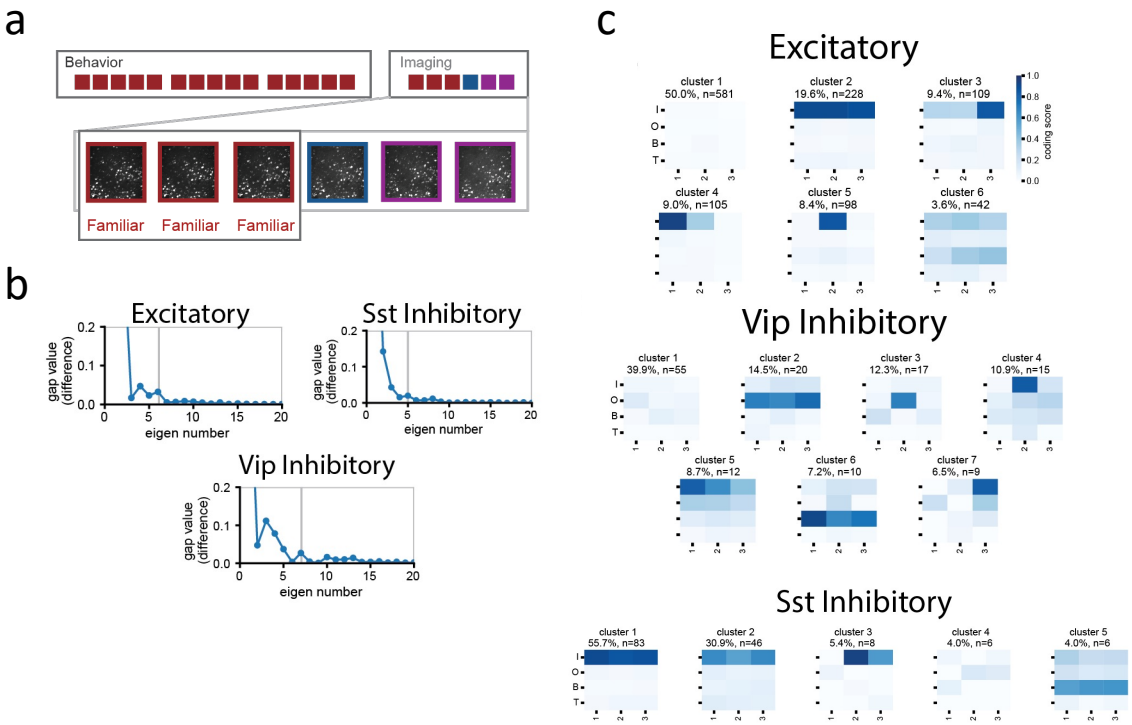
Extended Data Fig. 16 | Average coding scores for all clusters and layer distributions. a, Functional clusters for Excitatory cells. Top row, mean coding scores for each cluster, organized as feature category (rows) by experience level (columns), darker colors indicate higher coding scores. Second row, omission aligned population average across neurons belonging to each cluster plotted by experience level (familiar in red, novel in blue, novel+ re-exposure in purple). Gray bars indicate image presentations, blue dotted line indicates time of stimulus omission. Third row, proportion of cells in each visual area (VISp – primary visual cortex, VISl – lateromedial area), relative to the proportion of cells in that cluster. Bottom row, proportion of cells in each cluster cortical depth (upper <= 250 μm, lower > 250 μm), relative to the proportion of cells in that cluster. Differences in proportions are marked as significant with the result of a chi-squared test with Benjamini-Hochberg multiple comparison correction ($p < 0.05$). **b,** Same as a, for Vip Inhibitory clusters. **c,** Same as A for Sst Inhibitory clusters. **d,** Same as Fig 4g, for Vip Inhibitory and Sst Inhibitory subclasses. Proportion of cells in each combination of visual area and cortical depth for each cluster, relative to the proportion of cells in that cluster (Chi-squared test, followed by Benjamini-Hochberg correction).

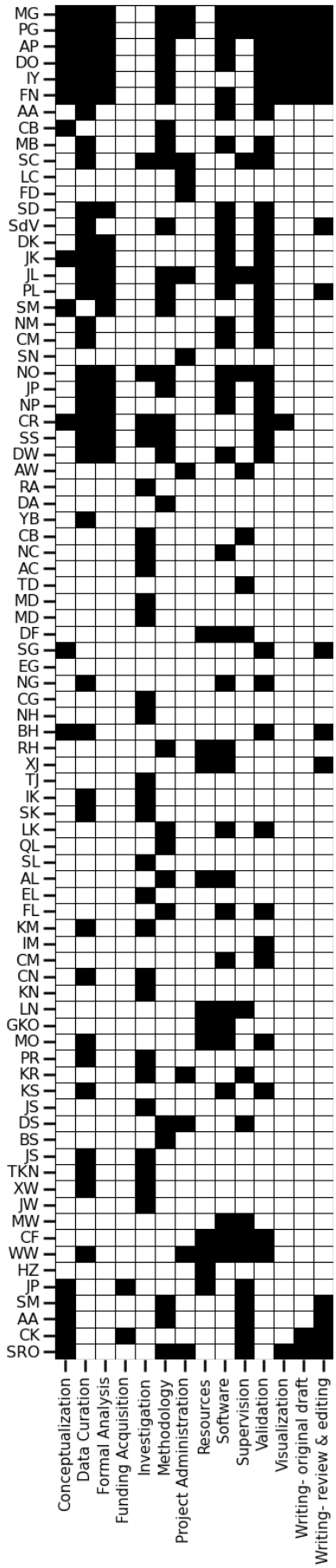
Extended Data Figure 17 – Functional clusters are not random.



Extended Data Fig. 17 | Functional clusters are not random. Shuffling cell ids across experience levels alters the size of novelty coding clusters. **a,** Schematic of shuffle procedure. Cell IDs were shuffled across days (experience levels) without disruption of coding scores for different regressors within each cell, resulting in data distributions identical to the original dataset. **b,** Top row, average coding scores of cells after shuffling experience levels and matching to the original clusters for Sst Inhibitory cells. No heatmap is shown if no cells were matched across all shuffle iterations. See panel e for matching procedure. Bottom row, difference in size (# of neurons) of original cluster relative to matched shuffled cluster, demonstrating that some cluster patterns are more prevalent in the real data compared to shuffled data. **c,** Same as panel b but for Excitatory cells. **d,** Same as B but for Vip Inhibitory cells. e**,** Mean sum-of-squared-error (SSE) values comparing shuffled clusters (rows) to original clusters (columns). Rows with no values indicate that no shuffled clusters were found to match the original clusters.

Extended figure 18 – Clustering of familiar sessions.



Extended Data Fig. 18 | Clustering results using three sequential sessions with familiar images. Clustering on familiar sessions results in fewer clusters with more consistent coding across days compared to novelty clusters. **a,** Schematic showing which sessions were used for clustering. **b,** The optimal number of clusters identified using eigengap values (see Methods). **b,** Average coding scores of clusters for familiar sessions, organized by feature category (Images, Omissions, Behavioral, Task) and experience level (Familiar, Novel, Novel+) for each cell class. The largest excitatory and Vip clusters were non-coding (excitatory familiar cluster 1, 50% of all excitatory cells; Vip familiar cluster 1, 40% of all Vip cells). The second largest excitatory cluster encoded images across all 3 familiar sessions (excitatory familiar cluster 2, 19.6% of all excitatory cells) and the second largest Vip cluster encoded omissions across all 3 familiar sessions (Vip familiar cluster 2, 14.5% of all Vip cells). Sst clusters derived from clustering on familiar sessions alone also showed a high degree of consistency across sessions, with both cluster 1 (55.7% of all Sst cells) and cluster 2 (30.9% of Sst cells) encoding images in all 3 sessions.

Extended Data Table 1 – Author contributions table using CRediT taxonomy