HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2024 October 15.

Published in final edited form as:

Cell Rep. 2024 August 27; 43(8): 114521. doi:10.1016/j.celrep.2024.114521.

Mixing novel and familiar cues modifies representations of familiar visual images and affects behavior

Noam Nitzan[1,2], **Corbett Bennett**[3], **J. Anthony Movshon**[2], **Shawn R. Olsen**[3], **György Buzsáki**[1,2,4,*]

[1]New York University Neuroscience Institute, New York University, New York, NY 10016, USA

[2]Center for Neural Science, New York University, New York, NY 10003, USA

[3]Allen Institute for Neural Dynamics, Seattle, WA 98109, USA

[4]Lead contact

SUMMARY

While visual responses to familiar and novel stimuli have been extensively studied, it is unknown how neuronal representations of familiar stimuli are affected when they are interleaved with novel images. We examined a large-scale dataset from mice performing a visual go/no-go change detection task. After training with eight images, six novel images were interleaved with two familiar ones. Unexpectedly, we found that the behavioral performance in response to familiar images was impaired when they were mixed with novel images. When familiar images were interleaved with novel ones, the dimensionality of their representation increased, indicating a perturbation of their neuronal responses. Furthermore, responses to familiar images in the primary visual cortex were less predictive of responses in higher-order areas, indicating less efficient communication. Spontaneous correlations between neurons were predictive of responses to novel images, but less so to familiar ones. Our study demonstrates the modification of representations of familiar images by novelty.

In brief

Based on a large-scale dataset from mice trained on a visual go/no-go change detection task, Nitzan et al. found that mice's behavior is impaired when familiar and novel stimuli are mixed. The decrease in performance was paralleled by a series of physiological correlates, which persisted during spontaneous activity.

Graphical Abstract

*Correspondence: gyorgy.buzsaki@nyulangone.org.

AUTHOR CONTRIBUTIONS

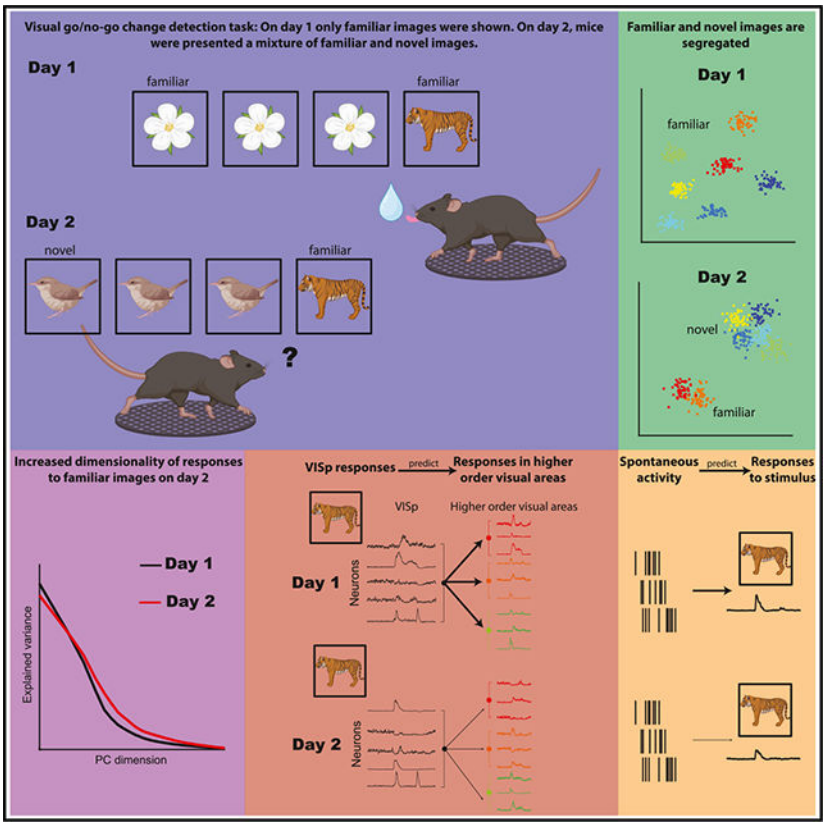
N.N. and G.B. designed the analytical process. C.B. and S.R.O. designed and supervised the data collection process. N.N. performed all analyses. J.A.M. contributed to formulating appropriate questions. N.N. and G.B. wrote the paper with inputs from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2024.114521.



INTRODUCTION

Reliable detection of change in the surrounding world is a survival advantage. The effectiveness of change detection depends on several factors, including brain state and the complexity of the environment, with its mixture of familiar and novel stimuli. Familiarity is not a feature of a physical cue but depends on the subject's experience with that cue. Extensive literature documents how neurons respond to specific familiar and novel stimuli and when searching for a target in the presence of distractors, but we have a limited understanding of the neuronal mechanisms that allow the detection of change in a visual scene. Responses to previously unexperienced stimuli are fundamental to cognitive operations, including curiosity, motivation, attention, and memory.[1] Novelty (unfamiliarity), surprise effect (unpredictability), stimulus history (recency), and salience induce behavioral orientation and attention tostimuli.[2,3] These behavioral changes are potentially driven by hypothetical "comparator" mechanisms distributed across many brain regions.[4–9] Novel visual stimuli trigger stronger evoked potentials, and cortical neurons respond more vigorously when an unexpected or unknown stimulus is presented compared to familiar stimuli.[10,11] As novel stimuli become familiar, neuronal responses attenuate over time.[10,12–14] This stimulus-specific reduction in neural activity has been referred to by various terms, including adaptation, mnemonic filtering, repetition suppression, and decremental responses.[15,16]

In real-life situations, novel stimuli are often mixed with familiar stimuli,[17] and the detection of even subtle changes in the animal's ecological milieu is often critical. Behavioral data from humans and other primates indicate that visual search for a familiar target pattern is impeded when it is embedded within novel distractors.[18–21] Yet, how the intrusion of novel stimuli affects the neuronal representation of familiar cues is not well understood. If novel representations were simply added to an ever-growing brain dictionary, adding new synapses and neurons would be required. This is an unlikely scenario, given that the total synaptic weights and firing rates in a given circuit do not change much with learning.[22,23] Another option is that novel stimuli bring about a reorganization among the old representations. Therefore, it is critical to learn whether and how novel inputs to the brain interact with existing representations and how such hypothetical reorganization affects the animal's ability to detect change.

To address how the intrusion of novel stimuli affects the stability of existing neuronal representations, we employed a paradigm that allowed large-scale recording by Neuropixels probe of neuronal activity in multiple cortical and subcortical structures simultaneously to study both single neuronal responses and their population dynamics in response to changing visual stimuli. In these experiments, head-fixed mice performed a go/no-go visual change detection task.[24] After several weeks of training with familiar images, subjects were shown an image set containing both familiar and novel images. Unexpectedly, we found that the behavioral performance in response to familiar images was substantially impaired when they were interleaved with novel images. The behavioral impairment was accompanied by a multitude of physiological changes. Novelty increased the tendency of visual cortical neurons for generalization, as indicated by larger fractions of cells modulated by a larger number of images. Unexpectedly, when familiar images were mixed with novel stimuli, they elicited weaker responses and recruited fewer cells. In addition, responses to familiar images became less correlated, and their dimensionality increased, indicating their perturbed stability. When familiar stimuli were mixed with novel stimuli, responses to the familiar stimuli in the primary visual cortex (VISp) became less predictive of activity in higher-order visual areas, indicating less efficient communication. The higher correlations between neuron pairs representing novel stimuli persisted during spontaneous activity in the absence of visual stimuli, compared to cell pairs representing both novel and familiar stimuli. Several physiological parameters of the modified neuronal responses were related to behavioral impairment. These findings suggest that novel information does not simply add to an existing scaffold, but novelty modifies the relationship between familiar stimuli and their previously formed neuronal representations.

RESULTS

Head-fixed mice were trained on a go/no-go visual change detection task (Figure 1A). Mice were shown a continuous series of natural images presented for 250 ms. On each trial, one image was presented for a variable number of times (5–11 times, according to a truncated geometric distribution). Mice were rewarded with a 3 μL water drop for licking within a 150–750 ms window whenever the identity of the image changed (HIT; Figure 1A). A lack of response was considered a MISS trial (omission error). If mice licked before the image change, the trial was aborted and restarted at the time of the next scheduled

image presentation (commission error). Mice underwent two recording sessions: in the first recording session, stimuli were drawn from a familiar set of eight images (set G) on which the mice were previously trained (30.5 ± 1.8 sessions, range 47). On the second recording day, they were shown a new image set (set H) containing six novel images and two familiar images from day 1 (henceforth “shared” images). Experimental sessions consisted of 60 min of active task performance, followed by 25 min of receptive field (RF) mapping using Gabor patches and full-field flashes. Following the RF mapping, the same sequences of images that were presented during the active part of the task were presented again, but this time the water spout was retracted (“passive replay”; Figure 1B).

Behavioral responses

Due to pre-training (>2 weeks; STAR Methods), mice performed at a high level on the first recording session (mean HIT rate per image 70%, range 64%–75% per image). On day 1, average HIT rates in response to the six day-1-only images (“familiar”) were strongly correlated with average HIT rates to the two shared images (im083 and im111; r = 0.90, Figure S1A), demonstrating that the familiar and shared images were perceived similarly. On day 2, mice maintained high HIT rates when stimulus change was signaled by previously unseen (novel) images (mean HIT rate per image 74%, range 71%–77%). Surprisingly, when the changed image was a shared (i.e., already familiar) image, the mice performed significantly worse (Figures 1C and S1C; mean HIT rate 47% per image), resulting in weaker correlations between average HIT rates for novel and shared images (r = 0.34, Figure S1A). The correlation between performances in response to shared stimuli on day 1 and day 2 was also lower (Figure S1B). To examine whether the decreased performance in response to a given image was associated with transitions from particular images, we computed the HIT probability conditioned on the previous image presented (Figure S1E). Conditional HIT probabilities did not depend on the previous image presented on either day, and lower HIT probabilities for shared images on day 2 were observed regardless of the preceding image. One possible explanation is an “extinction” effect, since after the active session on day 1, the images were also presented in the absence of reward for 60 min (passive replay; Figure 1B). We reasoned that if extinction was a proper explanation, the mice should relearn to associate shared image change and reward by improving their performance throughout the active session on day 2.[25] To examine this possibility, we computed the average MISS count in five-trial blocks (Figure 1D). Although mice had more misses (omission error) as the session progressed, the within-session dynamics of MISS trials were comparable between responses to familiar stimuli on day 1 and to novel stimuli on day 2. In contrast, MISS trials in response to shared images on day 2 were high throughout the session (Figure 1D). Further, while reaction times to the different images were not statistically different on day 1, mice responded slightly, but significantly, faster to shared images on day 2, which is inconsistent with the extinction hypothesis (Figure 1F). In further support against the extinction explanation, when, in a second cohort of mice ($n = 3$), the novel image set (H) was presented on day 1, the performance of the mice to shared images was reduced on day 1, but recovered on day 2, when they were presented together with other familiar images (G) (Figure S1F). Other behavioral variables, such as pupil diameter and running speed, did not show systematic variation and could not account for the decreased performance on day 2 (Figure S2).

In addition to HIT rate, we also examined the incidence of premature licks, which resulted in aborted trials (commission errors). On day 1, the fraction of aborted trials did not significantly differ across images (Figure 1E). However, on day 2, premature licks were significantly less frequent during the presentation of shared images, compared to novel stimuli, resulting in significantly fewer aborted trials (Figures 1E, S1D, and S1G). The increased omission, yet decreased commission, errors were manifested in a lower sensitivity index (d') on day 2 and higher decision criteria during shared image trials (Figures S1H and S1I), indicating that mice adopted a more conservative decision criterion for familiar images.

In a third cohort, mice ($n = 10$) were trained and tested on day 1 with image set H and on day 2 with image set G, and we replicated the main behavioral results (Figures S1J and S1K). This control experiment indicated that stimulus novelty per se, rather than the specific images in each image set, was driving the changes in performance.

Novelty modulates neuronal responses to familiar stimuli

We first examined the responses of individual neurons to the presented images. Six Neuropixels probes each targeted separate visual cortical areas, including VISp, as well as lateral and medial higher-order visual areas (Figure 2A). In addition, the probes also recorded from subcortical structures, including the visual thalamus (lateral geniculate nucleus [LGd] and lateral posterior nucleus [LP]), hippocampus (HPC), and midbrain (MB; Figure S3A). Peri-stimulus time histograms (PSTHs) of the population responses (computed for each image separately, based on repeated presentations, and averaged across stimulus categories; STAR Methods) were similar between images on day 1 (Figures 2B and S3B). As expected,[10,11,26] novel stimuli presented on day 2 elicited stronger responses in visual cortical areas, hippocampus, and midbrain but less so in the thalamus (Figure 2B). Novelty also increased the fractions of modulated cells in most areas (Figure S3C) and resulted in more sustained responses,[27] but response onset to novel stimuli was delayed compared to shared stimuli (Figures 2B, 2C, S3D, and S3E). Importantly, neuronal excitability measures in response to shared stimuli were altered on day 2 when they were mixed with novel stimuli. Response magnitude to shared stimuli was lower on day 2 compared to that of day 1 (Figure 2D), and the fraction of positively modulated cells decreased (Figure 2E), while the fraction of negatively modulated cells increased (Figure 2F; see also Figures S4A–S4C). These changes were similar in different visual cortical areas, while other areas showed mixed effects (Figures S4E–S4G). A cross-validated linear regression model trained to predict HIT rate/image from these metrics performed significantly worse on shared images on day 2 compared to day 1 (Figure S4D).

To visualize spiking activity patterns evoked by the presentation of natural images, we used Rastermap for sorting neural responses along a one-dimensional manifold.[28] This analysis uncovered multiple activity patterns expressed by clusters of neurons that alternated during the active behavior (Figure 3A). Some of the clusters were preferentially tuned to a single image, while other neuronal responses were non-specific to images or were related to the animal's motor behavior (Figure 3B). Both the size and the fraction of clusters that were preferentially tuned to shared images were significantly reduced on day 2 compared to day 1, indicating that responses to those images were sparser (Figures S4J and S4K).

PSTHs of individual neurons across the dataset exhibited a large degree of variability with respect to their selectivity, temporal response profile, and response magnitude (Figure 3C). A selectivity index was defined as the fraction of images that yielded significant spike responses. In VISp, the largest fraction of responding neurons (~15%) exhibited a low degree of selectivity by responding to all eight images, while only a small fraction of neurons (~8%) was selectively tuned to only one image. Selectivity decreased from day 1 to day 2, with fewer neurons responding to a single image and more neurons to all eight images, due to more generalized responses to novel stimuli. Lateral visual areas (VISl, VISal) followed the same trend, whereas medial visual areas showed intermediate degrees of selectivity (Figure 3D). In a striking contrast, responding hippocampal neurons (DG, CA3, CA1, and SUB) showed the opposite trend, with the majority of significantly modulated neurons responding to a single image and only a minor fraction responded to all eight images (Figure 3D).

Image decoding from population responses

In addition to single-neuron responses, we examined how well the identity of the images could be decoded by neuronal populations.[29,30] Due to the diverse image tuning profiles, the image identity could be readily decoded from the normalized spike counts of simultaneously recorded neurons using a linear decoder (Figure S5; STAR Methods). Decoding accuracy from withheld data was highest in the visual cortex and visual thalamus, but was also significant in the hippocampus and midbrain (Figures S5A–S5D). Image decodability was present already in the first 50 ms bin and reached a maximum by 100 ms in the visual cortex and remained high throughout the stimulus duration and even 200 ms after the stimulus offset (Figure S5E). The same trends were observed when decoding image identity was predicted from responses in individual visual cortical areas separately (Figure S5F). The temporal response profile was similar in the thalamus. In the hippocampus, image decoding was delayed and reached a maximum in the 100–150 ms bin and rapidly diminished thereafter (Figure S5E). The decoding accuracy of single images was high in the visual cortical areas, intermediate in the thalamus and midbrain, and lowest in the hippocampus (Figures S5G–S5H). The accuracy of image decoding from hippocampal spiking activity was reduced from day 1 to day 2 (Figure S5G), likely due to the higher confusion between novel stimuli (Figure S5B). Thus, while image specificity was highest in the hippocampus at the single-neuron level (Figure 3D), when a larger population of neurons was available, image decoding was more effective in the visual cortex, due mainly to the large numbers of responding neurons with different rates and patterns to individual images (Figure 3C).

Novel stimuli bring about modification of network dynamics

To explore potential mechanisms responsible for the differential responses to shared stimuli across days, we compared their population responses using multiple complementary methods. We first inspected the intrinsic dimensionality of visual cortical activity in response to shared stimuli on both days by computing their eigenspectra. Leading dimensions on day 1 explained more variance compared to day 2, and this was reversed for dimensions explaining less variance, resulting in a flattening of eigenspectra on day 2 (Figure 4A). We quantified this change using the participation ratio (PR), which measures the spread of explained variance across dimensions.[31] PR values of shared images on day

2 were significantly higher than those of the same images on the previous day, indicating an increase in the intrinsic dimensionality of responses to familiar images when interspersed with novel ones (Figure 4B). Dimensionality in higher-order visual areas was higher compared to that of the VISp (Figure S6A). These results were corroborated by fitting the eigenspectra of cross-validated principal components and estimating their power-law exponent.[29] This analysis indicated a significantly flatter eigenspectrum of shared images on day 2 than on day 1 (Figure S6B). Other areas included in the dataset also displayed a significant increase in the dimensionality of responses to familiar images on day 2 compared to day 1, but unlike the visual cortex, the dimensionality of novel images was also significantly higher (Figures S6C–S6E).

Next, we asked whether the hypothesized destabilization of shared image representations affected the interactions between the different visual areas. To address this question, we used a cross-validated ridge regression model, which takes the normalized single-trial residual activity (i.e., after subtracting the appropriate PSTH) of VISp units in response to each of the stimuli and uses this input to predict the activity in higher-order visual areas, as well as in a held-out VISp subpopulation (Figures 4C and 4D; STAR Methods). As expected, on day 1, predictive performances were similar for familiar and shared images (Figure 4E). They were highest in VISp and decreased in areas with a higher anatomical hierarchy score assigned by a previous anatomical study.[32] In contrast, the predictive performance of activity in response to shared images on day 2 was significantly lower than that of novel images (Figure 4F). Importantly, the prediction of spiking to shared images in higher-order visual areas from VISp was also significantly lower on day 2 compared with day 1 (Figure 4G). These differences were significant in higher-order visual areas, but not in the VISp (against withheld data), indicating less effective communication between VISp and higher-order visual areas in day 2, but not within VISp. These differences were maintained during passive viewing, indicating that they were not the result of action, motivation, or attention signals during the active part of the session (Figure S6F).

Novelty reduces change-evoked responses to familiar stimuli

It has been shown that viewing simple visual patterns such as gratings induces stimulus-specific adaptation,[33] but it is not clear whether adaptation is modulated by stimulus familiarity. We examined the extent of adaptation by computing the ratio of responses to changed and repeated presentations of the same image, averaged across trials. Confirming previous observations,[24,34] higher-order visual cortical areas were increasingly sensitive to changes in image identity, resulting in a strong correlation between the ratio of responses to changed/repeated shared stimuli and anatomical hierarchy (Figures 5A, S7A, and S7B). In stark contrast, when interleaved with novel stimuli, excess activity to changed familiar stimuli was markedly reduced, and the positive correlation with anatomical hierarchy was abolished (Figure 5A). These results were maintained during passive viewing, suggesting that they cannot be fully accounted for by action signals (Figure S7C). We validated these results by training a cross-validated linear decoder to classify image change from repeated presentation of the same image. Prediction accuracy was comparable across familiar and novel images but significantly decreased for shared images on day 2, indicating diminished differences between firing responses to changed vs. repeated shared stimuli (Figure 5B).

These differences were observed across other areas included in the dataset and were maintained during passive replay (Figures S7E–S7G). In sum, in the presence of novel stimuli, responses to already familiar stimuli undergo a weaker adaptation.

Novelty reduced the within-session stability of familiar representations

Although responses to novel stimuli were stronger and recruited more cells (Figure 2), their dimensionality was comparable to that of familiar stimuli on the previous day (Figure 4). We asked whether this is because neuronal representations of novel stimuli change less over time within the course of the session.[35,36] To compare the drift of neuronal responses to familiar and novel neural images, we calculated the population vector (PV) correlations between spike counts of visual cortex neurons across same-image trials for the first appearance of the image (Figure S8A). On day 1, we observed similar PV drifts across different images (Figure S8B), quantified by the decay rates and intercepts (estimated by exponential fit; Figures S8B and S8C). The decay rates of the shared images were similar on day 1 and day 2, whereas the PVs of novel images exhibited a slower decay (Figure S8C). Those differences were observed in individual subjects, suggesting that they are not driven by a small subset of mice (Figure S8D). The same decay was also maintained during passive viewing, suggesting that it is not due to action signals (Figures S8E and S8F), and the decay was similar in different visual cortical areas (Figure S8G). Subcortical areas, including the hippocampus, visual thalamus, and midbrain, exhibited similar drift, but with no significant differences between novel and familiar images (Figures S8H–S8J).

To test the hypothesis that a failure to respond to image change involves a momentary destabilization of the neuronal representation of that image, we calculated the difference between the PV correlation on a MISS trial and the immediately preceding and following HIT trials of the same image (Figure S8L). PV correlations on the surrounding HIT trials were, on average, higher than those on MISS trials on both recording days (Figure S8M). However, the difference in PV correlation on MISS trials was significantly larger for shared images on day 2 compared to both novel images and the same images on day 1 (Figure S8M, $p < 0.001$, Kruskal-Wallis with Tukey-Kramer *post hoc* tests). The average decrease in PV correlation in a session was positively correlated with MISS probability for both familiar and novel images (Figure S8N). These results suggest that mixing familiar and novel stimuli perturb the stability or strength of neuronal representations of familiar images.

Novel and familiar representations are segregated

While the response properties of single neurons to familiar and novel stimuli have been extensively compared,[10,11,37,38] it is unclear how visual cortex population activity is affected by novelty. To explore how changes in single-neuron firing are manifested on the population level across days, we compared population responses to familiar and novel stimuli, using two complementary methods. First, we computed a distance metric for each pair of images (expressed as the Frobenius norm between the normalized spike counts matrices) separately for each visual cortical area (Figure S9A). While in all pairs the two images were distinct from each other in both sessions using this metric (Figure S9B), the average distances on day 1 were relatively small across all visual areas (Figures S9C and S9D). In contrast, the distances between familiar (shared) and novel images on day 2 were several-fold larger

(Figures S9C and S9E). Similarly, we also found that edge angles between trial-averaged population response vectors to novel and familiar stimuli were significantly greater than those between pairs of novel-novel or familiar-familiar images, indicating a larger separation along the signal axis (Figures S9F–S9J). In a second approach, we asked whether different images are represented by overlapping sets or different combinations of neurons. To this end, we identified the set of neurons that were significantly modulated by each image (Figure 3; STAR Methods), and, for each pair of images, we computed the Jaccard similarity between these binary indicator vectors. The similarity between familiar and novel images was lower in all areas on day 2 ($p < 0.001$; Figure S9K), indicating a large separation between neurons representing familiar and novel stimuli. In sum, these results indicate that visual cortex activity patterns undergo orthogonalization in response to novelty.

Neuronal pattern changes predict behavioral performance

To further explore the changes in population activity across days and link them to task performance, we applied a non-linear dimensionality reduction method[39] to the normalized spike counts of visual cortex neurons during the active part of the task. This analysis revealed separate clusters corresponding to the different images as well as an additional cluster corresponding to the intermittent gray screen and omission trials (Figure 6A). A closer comparison of the embeddings in different areas revealed two unexpected trends. First, while clusters corresponding to the different images were well separated from one another and the gray-screen cluster in VISp, higher-order visual cortex areas showed a less robust separation (Figures 6A and S10A). Second, in higher-order visual cortex areas, clusters corresponding to shared images on day 2 were embedded substantially closer to the gray-screen cluster or even within it (Figure 6A). Because UMAP is a non-linear approach best suited to data visualization, we formally quantified this effect by applying principal-component analysis (PCA) to the data. We used the first three dimensions of the data, which explained a substantial (32% ± 1%) portion of the variance, and computed the average difference in distance from the gray-screen cluster between familiar (day 1) or novel (day 2) images and shared images and compared this measure with each region's anatomical hierarchy score (Figure 6B). We found that this difference linearly increased with anatomical hierarchy on day 2, resulting in a significant correlation, whereas day 1 differences were scattered around zero and were not correlated with anatomical hierarchy (Figure 6B). We hypothesized that, commensurate with the diminishing difference in response magnitude to shared images from baseline activity (Figure 2), the decreased performance in response to shared images on day 2 was the result of diminishing distinctiveness between the underlying neural representations. To test this idea, we computed the correlation between the distance of each image from the gray-screen cluster and the MISS probability of that image (Figure 6C). We found significant negative correlations between the distance of a cluster corresponding to a given image from the gray-screen cluster and the MISS probability of that image on day 2 in all visual areas. The magnitude of this correlation across visual areas was significantly correlated with their hierarchy score (Figure 6C).

To further link neuronal activity to behavior, we applied the same analysis to the firing responses of visual cortex neurons (combining all visual cortex areas), restricted to the presentation of changed images on HIT and MISS trials on both recording days. Clusters

corresponding to the different images were again well separated on both days (Figures 6D and 6E, left). We computed the average distance in PCA space between each cluster's centroid and all other cluster centroids. On day 1, normalized distances varied across images but were consistent across sessions (e.g., im078, Figure S10B) and were comparable on HIT and MISS trials (Figure S10F). In contrast, on day 2, centroid distances of novel and shared images showed opposite trends: the distances between clusters corresponding to novel images were significantly smaller on MISS trials compared both to the same images during HIT trials and to familiar images, consistent with the idea that MISS errors resulted from diminished distinctiveness in neural representations. On HIT trials, and even more so on MISS trials, the shared images were embedded significantly farther away from the rest of the clusters (Figure 6F; see Figures S10B–S10E for individual images). This differential embedding was not the result of different motor patterns (i.e., licking), because it was also observed during passive viewing without licking (Figure S10G). Similar results were obtained when applying this analysis for each visual area separately (Figure S10H). The normalized centroid distances of the shared images were significantly positively correlated with the probability of MISS trials for familiar and shared images, but not for novel images (Figure 6G). Thus, while the tendency of mice to miss novel images was associated with reduced distances between the representations of those images, representations of the interleaved familiar (shared) images were most distinct in MISS trials.

Mixing novel and familiar images affects the correlations of visual cortical neurons during spontaneous activity

If the segregation of familiar and novel representations and the subsequent destabilization of the representation of shared stimuli imparts enduring changes in the correlational structure of visual cortex neurons, we hypothesized that this altered relationship should persist in the absence of external stimuli. To test this hypothesis, we trained a cross-validated generalized linear model to predict the firing rates of individual withheld visual cortex neurons during spontaneous activity at the end of the active behavior from the weighted sum of firing rates of the rest of the visual cortex population ("peer prediction").[40] After optimizing the weights to predict spontaneous spiking data, we used the stimulus-evoked firing rates of neurons from the active task to predict the response of the left-out neuron (Figure 7A). This analysis allowed us to reconstruct the peer-predicted tuning curves of visual cortex neurons (Figure 7B). We then evaluated the reconstruction accuracy by correlating the actual and the predicted tuning curves (Figure 7C). On day 1, correlations were comparable across familiar and shared images and consistent with previous results in the auditory cortex.[41] The reconstruction accuracy of responses to novel images was significantly higher compared to familiar ones, indicating that the correlational structure of cells responding to novel images was less variable between stimulus-evoked and spontaneous activity. In contrast, the reconstruction of responses to shared images on day 2 was significantly lower compared to that of day 1 (Figure 7C). These results indicate that the population cofiring patterns that preferentially engaged neurons during stimulus presentations endured during the following spontaneous activity, but this effect was diminished for familiar stimuli when they were interleaved with novel ones.

DISCUSSION

Using a change detection task, we examined how intruding novel stimuli affect the neuronal representation of old cues. When familiar images were embedded in a stream of novel pictures, their neuronal correlates were altered, including the predictions of neuronal patterns in higher-order areas from spiking activity in VISp. The altered neuronal changes were predictive of behavioral outcomes. The findings suggest that novel cues can affect the neuronal representation of familiar images and their behavioral responses.

Novel stimuli affect the representation of familiar images

The only condition for obtaining a reward in the present paradigm was to detect a change across various images, irrespective of whether the images were familiar or novel. Behavioral performances in response to familiar and novel stimuli were comparable in terms of HIT rates, an indication that the schema[42] learned during the weeks of training was generalized to the novel stimuli on day 2, despite the overall increased firing rates evoked by novel images. Yet, the animals' overall behavior was quite different on day 1 and day 2, mainly due to the decreased HIT rates (omission errors) in response to the familiar (shared) stimuli. At the same time, the mice licked less discriminately and at a more constant rate during the presentation of novel images (commission errors), indicating both stimulus-specific and global changes in their behavior.

Presumably, the main factor in the change detection task is the perceived magnitude of change between the current and the previous image. On day 2, novel stimuli may have been regarded by the mice as a larger departure from the norm (i.e., perceived change), explaining why HIT rates were comparable to those of familiar (day 1) and novel (day 2) images. At the same time, this larger contrast to novel images may have reduced the perceived magnitude of chance (or "surprise"[43]) in response to shared stimuli, resulting in an increased fraction of MISS trials. In turn, we hypothesize that the reduced behavioral performance can be explained by the altered neuronal representation of the familiar (shared) images in the presence of novel cues. This interpretation is supported by several physiological measures. The latency to the first evoked spike to familiar images was shorter compared to that for novel stimuli[44–46] in visual cortical areas, commensurate with the reduced latency of behavioral responses. Yet, the mixture of old and novel images perturbed the representation of the old images. The effects of this perturbation were manifested by multiple differences between the representations of shared images on both days. Responses to shared images on day 2 were closer to baseline activity, as indicated by a lower modulation index on day 2 (Figure 2), and increasingly similar to responses to the gray screen, particularly in higher-order visual cortex areas (Figure 6). Unexpectedly, the dimensionality of population responses to shared stimuli increased from day 1 to day 2 (Figure 4B), despite the responses becoming sparser on day 2. In addition, neuronal responses to shared images on MISS trials were particularly distinct from the responses to novel images, and their intercluster distance positively correlated with MISS probability. The reduction in PV correlation on MISS trials was larger for shared images on day 2 compared to both day 1 and novel images (Figure S8). Further, the instantaneous decrease in PV correlation was positively correlated with MISS probability for both the familiar and the novel images. When interleaved with novel stimuli,

familiar stimuli underwent weaker adaptation, resulting in a diminished sensitivity to image change across different visual areas. Finally, the readout of spiking activity in the VISp by higher-order areas for shared images was reduced from day 1 to day 2 (Figure 4), indicating reduced efficacy of representation of the shared images in higher-order visual areas in the presence of distracting novel stimuli. The reorganization of visual circuits introduced by the novel images is further supported by the altered correlational structure introduced by the novel stimuli that was evident by the low overlap between cell pairs responding to familiar and novel stimuli (Figure S9) and persisted during the following spontaneous activity (Figure 7). Overall, these findings demonstrate that mixing familiar stimuli with novel images temporally alters the relationship between the same physical stimuli and their neuronal representations.

The hypothesis of destabilization of the representation of familiar images is seemingly at odds with the slower decay of PV correlation as a function of trial numbers for novel stimuli. The reduced representational drift may be explained by the enhanced attention to novel stimuli.[47] The persistence of the PV of novel stimuli throughout the session may explain the persistent destabilization of the neuronal representation of familiar images.

In our experiments, the familiar stimuli embedded in a sequence of novel images were associated with altered neuronal firing patterns, yet the extent of this destabilization effect remains unknown, as mice were tested on the novel image set only once. Insights into this question may come from a related, 3 day experiment, in which mice were shown familiar images on day 1, novel images on day 2, and then the same novel images on day 3. The neuronal responses were similar on day 1 and day 3, suggesting that novelty can become familiar overnight.[26] Thus, one would expect that if the six novel images shown on day 2 were repeatedly presented over several days, they would have blended with the familiar eight stimuli presented on day 1, and the physiological differences we described here between old and novel images would disappear.

An analogy may be drawn between the current observations and human psychophysical experiments employing a visual search paradigm. Search for a novel target among familiar distractors was markedly faster than search for a familiar target among novel distractors, and the reaction time increased with increasing numbers of novel distractors.[18,19] Likewise, the behavioral performance of expert subjects on a bisection task deteriorated when different stimulus types were interleaved,[48] indicating that mixing different stimulus categories may impair task sensitivity. Similarly, parallels to the altered representations of shared images on day 2 may be drawn with fMRI experiments with human subjects[17,49] in which the participants had to compare match-mismatch expectations of object sequences (A > B > C > D). Following repeated presentations, the participants viewed the same quartet of objects in the same sequence, an entirely new sequence, or a sequence in which only the last two objects in the sequence were reversed. Similar to our experiments, hippocampal activation to the entirely new sequence did not differ much from activation produced by the familiar sequence. In contrast, hippocampal activation was altered when the objects corresponded to a mixture of old and novel sequences.

Overall, our findings suggest that the perturbations to the representation of familiar images may be explained by the aberrant interactions along the visual hierarchy. However, top-down inputs to visual cortex may also play a critical role in visual discrimination.[50–52] It remains to be tested how top-down inputs modulate local computations in the visual cortex during the change detection task to guide behavior.

Limitations of the study

The current paradigm tested the effects of mixing familiar and novel stimuli under fixed conditions (i.e., a 1/3 ratio). It is unclear whether the observed destabilization effect would persist under different conditions (e.g., a 3/1 ratio). It is tempting to speculate that the magnitude of destabilization would vary with the ratio of novel to familiar images. Insights to this question may come from human studies showing that visual search for familiar targets embedded within novel distractors is incrementally impeded with increasing numbers of novel distractors, supporting the hypothesis that the destabilization effect would decrease with a decreasing ratio of novel to familiar images. To test this hypothesis, new experiments with varying ratios of novel to familiar images will be needed. Although the only requirement in the present task was to lick after detecting a change in the presented images, mice produced other types of behaviors as well. The most prominent of these was the extremely high-speed running that was maintained during the entire session (Figure S2). Calorie loss associated with such intense exercise is substantial and may reflect maladaptive "collateral" behavior,[53,54] likely induced by the head-fixed condition. How constraining the mouse's behavior by head fixing affected behavioral performance and neurophysiological patterns is not clear at the moment. Eye and head movements are essential features of vision and likely critical elements of image transformation. Future experiments in freely moving mice will be needed to address these caveats.

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, György Buzsáki (Gyorgy.Buzsaki@nyulangone.org).

Materials availability—This study did not generate new unique reagents.

Data and code availability

- Data: The data reported in this paper is publicly available on http://portal.brain-map.org/explore/circuits/visual-behavior-neuropixels. Any additional information required to reanalyze the data reported in this work paper is available from the Lead Contact upon request.
- Code: The code used to analyze the data is available at https://github.com/buzsakilab/buzcode

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Mice—For all analyses presented in this paper, we use data from the Allen Brain Institute Visual Behavior Neuropixels dataset. The main dataset included both male and female mice of the following genotypes: 10 WT mice, 19 Sst-IRES-Cre/wt;Ai32(RCL-ChR2(H134R)_EYFP)/wt and 11 Vip-IRES-Cre/wt;Ai32(RCL-ChR2(H134R)-_EYFP)/wt mice. 35 of these mice were measured across both days. For the supplementary datasets shown in Figure S1 data included 3 WT mice and 10 Sst-IRES-Cre/wt;Ai32(RCL-ChR2(H134R)_EYFP)/wt mice.

METHOD DETAILS

Behavior—Mice were trained for 1 hour/day and progressed through multiple training phases by meeting specific progression criteria defined by d-prime and the number of contingent (i.e., not aborted) trials per training session. Mice were deemed ready to be transitioned to the recording stage if the following criteria were met during the final training phase: 1) A peak d-prime (calculated over a 100 trials rolling window) of >1 for three consecutive sessions. 2) At least 100 contingent trials on three consecutive sessions. 3) Mean reward number of >120 over at least three sessions. The active task lasted for 1 h and consisted of 31.3 ± 0.05 go trials (where the identity of images changed) on day 1 and 29.3 ± 0.07 go trials on day 2 (mean ± s.e.m.).

QUANTIFICATION AND STATISTICAL ANALYSIS

Signal detection measures—d-prime and decision criteria were computed based on hit rates and false-alarm rates, which were based on catch trials, where the identity of the image did not change but the animal licked in the response window. Hit and false alarm rates were corrected to account for low trial counts by clipping their values using the following formula:

$$\frac{1}{2N_H} \leq R_H \leq 1 - \frac{1}{2N_H}$$

$$\frac{1}{2N_F} \leq R_F \leq 1 - \frac{1}{2N_F}$$

Where R_H and R_F are hit and false-alarm rates, respectively, and N_H and N_F are the numbers of go and catch trials, respectively. d-prime is defined as:

$$d' = \mathrm{Z}(R_H) - \mathrm{Z}(R_F)$$

The decision criterion is defined as:

$$\mathrm{c} = -(\mathrm{Z}(R_H) + \mathrm{Z}(R_F))/2$$

where Z is the inverse cumulative normal distribution function.

Unit quality criteria—For all analyses, we only included units that with less than 0.5% ISI violations, presence ratio higher than 0.9, amplitude cutoff below 0.1 and firing rate higher than 0.1 Hz.

Estimation of pupil area—Eye tracking data was acquired at 30 Hz and pre-processed by the Allen Institute. The pupil diameter, defined as the mean of the pupil height and width was normalized by the median of each session.

Rastermap embedding—One-dimensional embedding of neural activity was performed using Rastermap (https://github.com/MouseLand/RasterMap) with the following parameters: n_clusters = 20, n_PCs = 40, locality = 0.75, time_lag_window=15.

Event triggered histograms and significance of modulation—Peri-stimulus time histograms (PSTHs) were computed by counting spiking activity around stimulus time into 1 ms bins. The mean firing rate was then calculated by dividing by the bin size and number of stimuli. To deem units significantly modulated by a given stimulus, spike times were jittered in a ± 0.5 s window to generate 1000 surrogate PSTHs for each unit. We then calculated the sum of squared differences between the actual PSTH and the mean of surrogate PSTHs and compared it to the sum of squared differences of each surrogate PSTH with the mean of surrogate PSTHs. Units were deemed significantly modulated if this difference exceeded the 97.5 percentile, corresponding to $p<0.05$. The modulation index was computed by calculating the percent of firing rate change during the stimulation window compared to a 100 ms preceding baseline window and divided by the mean baseline firing rate and the modulation sign (i.e., up or down-modulation) was defined as the direction of this index.

Response lags—Response lags to the different stimuli were calculated as the median time to first spike across all trials. We only considered response lags of significantly modulated units.

Decoding of image identity—We used a linear multiclass SVM trained using 10-fold cross validation to predict natural image identity. To exclude the potential influence of expectation signals we only included the first image in a trial (i.e., the changed image). Input features were min-max normalized spike counts from all units in a given session and brain region. Decoding accuracy is reported as the mean of the cross-validated accuracy.

Measurements of population response geometry—To measure the linear difference between neural representations of each pair of images we used the measure of the Frobenius norm computed as:

$$\| A^{p,q} \|_F = \| A^p - A^q \|_F = \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} |a^p_{i',j} - a^q_{ij}|^2}$$

Where A^p and A^q are min-max normalized spike count by image presentation matrices for image p and q.

The edge angle between the representations of different images was computed as:

$$\Theta_{p,q} = cos^{-1}\left[\frac{\boldsymbol{u}_p \cdot \boldsymbol{u}_q}{\| \boldsymbol{u}_p \| \| \boldsymbol{u}_q \|}\right]$$

where $\Theta_{p,q}$ is the angle between vectors $\boldsymbol{u}_p$ and $\boldsymbol{u}_q$ denoting average responses to images p and q.

Jaccard similarity was calculated as:

$$J_{p,q} = |\boldsymbol{u}_p \cap \boldsymbol{u}_q| / |\boldsymbol{u}_p \cup \boldsymbol{u}_q|$$

Where $\boldsymbol{u}_p(i)$ and $\boldsymbol{u}_q(i)$ are indicator variables denoting whether the i^{th} neuron is significantly modulated by images p and q, respectively and $|\mathbf{u}|$ is the cardinal set of $\mathbf{u}$.

Regression models—To predict per-session HIT rates, we used a 10-fold cross-validated linear regression model using the fractions of up- and down-modulated cells, as well as the session-averaged visual cortex modulation index as input features. Model performance is reported as the distributions of mean-squared error values across folds.

To predict target population activity from VISp activity, spiking activity from each visual area during the 250 ms image presentation window was counted in 50 ms bins, the appropriate PSTH was subtracted from each single-trial response and the resulting responses were z-score normalized. To exclude the contribution of motor patterns, we excluded the first (i.e., changed) image in a trial. Target population activity was predicted using:

$$\widehat{Y}_{\text{Ridge}} = X\boldsymbol{B}_{\text{Ridge}}$$

Where $\boldsymbol{B}_{\text{Ridge}} = \left(X^T X + \lambda I\right)^{-1} X^T Y$ is the least-squares solution and λ is a constant that determines the strength of regularization chosen using a 10-fold cross validation.

Population vector correlation—The population vector (PV) correlation was calculated as Pearson's correlation between the spike counts of neurons over repeated presentations of the same stimulus. To account for potential expectation signals, we only used the first image in a trial (i.e., the changed image). Average PV correlations are referenced to the first presentation of an image.

UMAP visualization—Uniform Manifold Approximation and Projection (UMAP) was performed on spiking data binned at 50 ms resolution using parameters n_neighbors = 20, metric = Euclidean, min_dist = 0.1, spread = 0.1, components=3. We used the MATLAB implementation available under (https://www.mathworks.com/matlabcentral/fileexchange/71902).

Quantifying the dimensionality of responses to natural images—We used the participation ratio (PR)[31] to quantify the dimensionality of neural responses. The

eigenspectrum of responses to each image was obtained by applying principal component analysis to spike count matrices, and the PR was defined as:

$$PR = \frac{\left(\sum_{i=1}^{N} \lambda_i\right)^2}{\sum_{i=1}^{N} (\lambda_i)^2}$$

To allow for comparison across different sessions, the PR was divided by the number of neurons.

Cross-validated principal components were computed using a k-fold approach (k=5): singular vectors V_n were first calculated from the training data X_{train} and the test data X_{test} was then projected onto those singular vectors to yield a cross-validated PC score $U_n = X_{test}V_n$. The variance was then calculated from each cvPC score.

Peer prediction analysis—We used a generalized linear model (GLM) to reconstruct tuning curves to the different images from correlations during spontaneous activity. First, spike counts during the 5-minutes spontaneous activity were binned at 250 ms resolution and a 10-fold cross-validated GLM was trained to predict the spike counts of a left-out visual cortex neuron from the spike counts of all other visual cortex neurons using the Poisson distribution. The model weights were optimized by minimizing the mean squared error between the observed and predicted spike count. We then used the weights obtained from spontaneous activity in combination with the firing rates of peer neurons in response to image presentation to predict the response of the left-out neuron to that stimulus.

Statistical analyses—Data were analyzed in Matlab (2021b). Throughout the paper, data are presented as mean ± SEM or, when indicated, median ± 95% confidence intervals. Data are displayed as box plots representing median, lower and upper quartiles and whiskers representing most extreme data points or as median ± 95% confidence intervals computed from 5,000 resamples. Statistical tests for two groups were performed using Wilcoxon rank-sum test or signed-rank test when applicable. Statistical tests for multiple groups were performed using Kruskal-Wallis test followed by Tukey-Kramer posthoc tests.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank the Allen Institute for making the dataset available. This work was supported by NIH grants R01MH122391, R01EY022428, and U19NS107616 and by the Simons Foundation (534019). N.N. is supported by a German Research Foundation Walter Benjamin fellowship (NI 2057/1-1) and a Swiss National Science Foundation Mobility fellowship (P500PB_21440).

REFERENCES

1. Van Kesteren MTR, Ruiter DJ, Fernández G, and Henson RN (2012). How schema and novelty augment memory formation. Trends Neurosci. 35, 211–219. 10.1016/J.TINS.2012.02.001. [PubMed: 22398180]
2. Sokolov EN (1963). Higher nervous functions; the orienting reflex. Annu. Rev. Physiol. 25, 545–580. 10.1146/annurev.ph.25.030163.002553. [PubMed: 13977960]
3. Zhang K, Bromberg-Martin ES, Sogukpinar F, Kocher K, and Monosov IE (2022). Surprise and recency in novelty detection in the primate brain. Curr. Biol. 32, 2160–2173. 10.1016/j.cub.2022.03.064. [PubMed: 35439433]
4. Kafkas A, and Montaldi D (2018). Expectation affects learning and modulates memory experience at retrieval. Cognition 180, 123–134. 10.1016/J.COGNITION.2018.07.010. [PubMed: 30053569]
5. Knight RT (1996). Contribution of human hippocampal region to novelty detection. Nature 383, 256–259. 10.1038/383256a0. [PubMed: 8805701]
6. Stern CE, Corkin S, González RG, Guimaraes AR, Baker JR, Jennings PJ, Carr CA, Sugiura RM, Vedantham V, and Rosen BR (1996). The hippocampal formation participates in novel picture encoding: Evidence from functional magnetic resonance imaging. Proc. Natl. Acad. Sci. USA 93, 8660–8665. 10.1073/PNAS.93.16.8660. [PubMed: 8710927]
7. Tulving E, and Kroll N (1995). Novelty assessment in the brain and long-term memory encoding. Psychon. Bull. Rev. 2, 387–390. 10.3758/BF03210977/METRICS. [PubMed: 24203720]
8. Ranganath C, and Rainer G (2003). Cognitive neuroscience: Neural mechanisms for detecting and remembering novel events. Nat. Rev. Neurosci. 4, 193–202. 10.1038/nrn1052. [PubMed: 12612632]
9. Schomaker J, and Meeter M (2015). Short- and long-lasting consequences of novelty, deviance and surprise on brain and cognition. Neurosci. Biobehav. Rev. 55, 268–279. 10.1016/j.neubiorev.2015.05.002. [PubMed: 25976634]
10. Homann J, Koay SA, Chen KS, Tank DW, and Berry MJ (2022). Novel stimuli evoke excess activity in the mouse primary visual cortex. Proc. Natl. Acad. Sci. USA 119, e2108882119. 10.1073/pnas.2108882119. [PubMed: 35101916]
11. Garrett M, Manavi S, Roll K, Ollerenshaw DR, Groblewski PA, Ponvert ND, Kiggins JT, Casal L, Mace K, Williford A, et al. (2020). Experience shapes activity dynamics and stimulus coding of VIP inhibitory cells. Elife 9, e50340. 10.7554/eLife.50340. [PubMed: 32101169]
12. Natan RG, Briguglio JJ, Mwilambwe-Tshilobo L, Jones SI, Aizenberg M, Goldberg EM, and Geffen MN (2015). Complementary control of sensory adaptation by two types of cortical interneurons. Elife 4, e09868. 10.7554/ELIFE.09868. [PubMed: 26460542]
13. Nejad NG, English G, Apostolelli A, Kopp N, Yanik MF, and von der Behrens W (2023). Deviance Distraction and Stimulus-Specific Adaptation in the Somatosensory Cortex Reduce with Experience. J. Neurosci. 43, 4418–4433. 10.1523/JNEUROSCI.1714-22.2023. [PubMed: 37169591]
14. Kato HK, Chu MW, Isaacson JS, and Komiyama T (2012). Dynamic Sensory Representations in the Olfactory Bulb: Modulation by Wakefulness and Experience. Neuron 76, 962–975. 10.1016/J.NEURON.2012.09.037. [PubMed: 23217744]
15. Clark A (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci. 36, 181–204. 10.1017/S0140525X12000477. [PubMed: 23663408]
16. Spratling MW (2017). A review of predictive coding algorithms. Brain Cognit. 112, 92–97. 10.1016/J.BANDC.2015.11.003. [PubMed: 26809759]
17. Kumaran D, and Maguire EA (2009). Novelty signals: a window into hippocampal information processing. Trends Cognit. Sci. 13, 47–54. 10.1016/j.tics.2008.11.004. [PubMed: 19135404]
18. Mruczek REB, and Sheinberg DL (2005). Distractor familiarity leads to more efficient visual search for complex stimuli. Perception and Psychophysics 67, 1016–1031. 10.3758/BF03193628. [PubMed: 16396010]
19. Wang Q, Cavanagh P, and Green M (1994). Familiarity and pop-out invisual search. Percept. Psychophys. 56, 495–500. 10.3758/BF03206946/METRICS. [PubMed: 7991347]
20. Gauthier I, and Tarr MJ (1997). Becoming a “Greeble” Expert: Exploring Mechanisms for Face Recognition. Vis. Res. 37, 1673–1682. 10.1016/S0042-6989(96)00286-6. [PubMed: 9231232]

21. Logothetis NK, Pauls J, and Poggio T (1995). Shape representation in the inferior temporal cortex of monkeys. Curr. Biol. 5, 552–563. 10.1016/S0960-9822(95)00108-4. [PubMed: 7583105]
22. Dragoi G, Harris KD, and Buzsáki G (2003). Place representation within hippocampal networks is modified by long-term potentiation. Neuron 39, 843–853. 10.1016/S0896-6273(03)00465-3. [PubMed: 12948450]
23. Royer S, and Paré D (2003). Conservation of total synaptic weight through balanced synaptic depression and potentiation. Nature 422, 518–522. 10.1038/nature01530. [PubMed: 12673250]
24. Siegle JH, Jia X, Durand S, Gale S, Bennett C, Graddis N, Heller G, Ramirez TK, Choi H, Luviano JA, et al. (2021). Survey of spiking in the mouse visual system reveals functional hierarchy. Nature 592, 86–92. 10.1038/s41586-020-03171-x. [PubMed: 33473216]
25. Dunsmoor JE, Niv Y, Daw N, and Phelps EA (2015). Rethinking Extinction at Cell Press. Neuron 88, 47–63. 10.1016/j.neuron.2015.09.028. [PubMed: 26447572]
26. Garrett M, Groblewski P, Piet A, Ollerenshaw D, Najafi F, Yavorska I, Amster A, Bennett C, Buice M, Caldejon S, et al. (2023). Stimulus novelty uncovers coding diversity in visual cortical circuits. Preprint at bioRxiv, 1–40. 10.1101/2023.02.14.528085.
27. Meyer T, Walker C, Cho RY, and Olson CR (2014). Image familiarization sharpens response dynamics of neurons in inferotemporal cortex. Nat. Neurosci. 17, 1388–1394. 10.1038/nn.3794. [PubMed: 25151263]
28. Stringer C, Zhong L, Syeda A, Du F, Pachitariu M, and Pachitariu M (2023). Rastermap: a discovery method for neural population recordings. Preprint at bioRxiv, 1–24. 10.1101/2023.07.25.550571.
29. Stringer C, Pachitariu M, Steinmetz N, Carandini M, and Harris KD (2019). High-dimensional geometry of population responses in visual cortex. Nature 571, 361–365. 10.1038/s41586-019-1346-5. [PubMed: 31243367]
30. Hung CP, Kreiman G, Poggio T, and DiCarlo JJ (2005). Fast readout of object identity from macaque inferior temporal cortex. Science 310, 863–866. 10.1126/science.1117593. [PubMed: 16272124]
31. Altan E, Solla SA, Miller LE, and Perreault EJ (2021). Estimating the dimensionality of the manifold underlying multi-electrode neural recordings. PLoS Comput. Biol. 17, e1008591. 10.1371/journal.pcbi.1008591. [PubMed: 34843461]
32. Harris JA, Mihalas S, Hirokawa KE, Whitesell JD, Choi H, Bernard A, Bohn P, Caldejon S, Casal L, Cho A, et al. (2019). Hierarchical organization of cortical and thalamic connectivity. Nature 575, 195–202. 10.1038/s41586-019-1716-z. [PubMed: 31666704]
33. Movshon JA, and Lennis P (1979). Pattern-selective Adaptation in Visual Cortical Neurones. Nature 278, 850–852. 10.1038/278850a0. [PubMed: 440411]
34. Vinken K, Vogels R, and Op de Beeck H (2017). Recent Visual Experience Shapes Visual Processing in Rats through Stimulus-Specific Adaptation and Response Enhancement. Curr. Biol. 27, 914–919. 10.1016/j.cub.2017.02.024. [PubMed: 28262485]
35. Deitch D, Rubin A, and Ziv Y (2021). Representational drift in the mouse visual cortex. Curr. Biol. 31, 4327–4339. 10.1016/j.cub.2021.07.062. [PubMed: 34433077]
36. Aitken K, Garrett M, Olsen S, and Mihalas S (2022). The geometry of representational drift in natural and artificial neural networks. PLoS Comput. Biol. 18, e1010716. 10.1371/JOURNAL.PCBI.1010716. [PubMed: 36441762]
37. Woloszyn L, and Sheinberg DL (2012). Effects of Long-Term Visual Experience on Responses of Distinct Classes of Single Units in Inferior Temporal Cortex. Neuron 74, 193–205. 10.1016/j.neuron.2012.01.032. [PubMed: 22500640]
38. Huang G, Ramachandran S, Lee TS, and Olson CR (2018). Neural correlate of visual familiarity in macaque area V2. J. Neurosci. 38, 8967–8975. 10.1523/JNEUROSCI.0664-18.2018. [PubMed: 30181138]
39. McInnes L, Healy J, Saul N, and Großberger L (2018). UMAP: UniformManifold Approximation and Projection. J. Open Source Softw. 3, 861. 10.21105/JOSS.00861.
40. Harris KD, Csicsvari J, Hirase H, Dragoi G, and Buzsáki G (2003). Organization of cell assemblies in the hippocampus. Nature 424, 552–556. 10.1038/nature01834. [PubMed: 12891358]

41. Luczak A, Barthó P, and Harris KD (2009). Spontaneous Events Outline the Realm of Possible Sensory Responses in Neocortical Populations. Neuron 62, 413–425. 10.1016/j.neuron.2009.03.014. [PubMed: 19447096]

42. Tse D, Langston RF, Kakeyama M, Bethus I, Spooner PA, Wood ER, Witter MP, and Morris RGM (2007). Schemas and memory consolidation. Science 316, 76–82. 10.1126/science.1135935. [PubMed: 17412951]

43. Walsh V (2003). A theory of magnitude: common cortical metrics of time, space and quantity. Trends Cognit. Sci. 7, 483–488. 10.1016/J.TICS.2003.09.002. [PubMed: 14585444]

44. Grill-Spector K, Henson R, and Martin A (2006). Repetition and the Brain: Neural Models of Stimulus-specific Effects. Trends Cognit. Sci. 10, 14–23. 10.1016/j.tics.2005.11.006. [PubMed: 16321563]

45. Manahova ME, Spaak E, and de Lange FP (2020). Familiarity Increases Processing Speed in the Visual System. J. Cognit. Neurosci. 32, 722–733. 10.1162/JOCN_A_01507. [PubMed: 31765601]

46. Donohue SE, Bartsch MV, Heinze HJ, Schoenfeld MA, and Hopf JM (2018). Cortical Mechanisms of Prioritizing Selection for Rejection in Visual Search. J. Neurosci. 38, 4738–4748. 10.1523/JNEUROSCI.2407-17.2018. [PubMed: 29691330]

47. Zemla R, Moore JJ, Hopkins MD, and Basu J (2022). Task-selective place cells show behaviorally driven dynamics during learning and stability during memory recall. Cell Rep. 41, 111700. 10.1016/J.CELREP.2022.111700. [PubMed: 36417882]

48. Clarke AM, Grzeczkowski L, Mast FW, Gauthier I, and Herzog MH (2014). Deleterious effects of roving on learned tasks. Vis. Res. 99, 88–92. 10.1016/J.VISRES.2013.12.010. [PubMed: 24384405]

49. Kumaran D, and Maguire EA (2007). Match-mismatch processes underlie human hippocampal responses to associative novelty. J. Neurosci. 27, 8517–8524. 10.1523/JNEUROSCI.1677-07.2007. [PubMed: 17687029]

50. Gilbert CD, and Li W (2013). Top-down Influences on Visual Processing. Nat. Rev. Neurosci. 14, 350–363. 10.1038/nrn3476. [PubMed: 23595013]

51. Zhang S, Xu M, Kamigaki T, Do JPH, Chang WC, Jenvay S, Miyamichi K, Luo L, and Dan Y (2014). Selective attention. Long-range and local circuits for top-down modulation of visual cortex processing. Science 345, 660–665. 10.1126/science.1254126. [PubMed: 25104383]

52. Makino H, and Komiyama T (2015). Learning enhances the relative impact of top-down processing in the visual cortex. Nat. Neurosci. 18, 1116–1122. 10.1038/nn.4061. [PubMed: 26167904]

53. Bruner S, and Revusky SH (1961). Collateral Behavior in Humans. J. Exp. Anal. Behav. 4, 349–350. 10.1901/jeab.1961.4-349. [PubMed: 13874003]

54. Wilson MP, and Keller FS (1953). On the selective reinforcement of spaced responses. J. of Comp. Physiol. Psychol. 56, 495–500. 10.1037/h0057705.

Highlights

- Mice perform poorly on a change detection task when familiar and novel images are mixed
- Neuronal responses to familiar images are perturbed when they are mixed with novel stimuli
- Communication between visual areas during familiar stimuli is perturbed by novel stimuli
- Mixing novel and familiar stimuli alters spontaneous correlations in the visual cortex

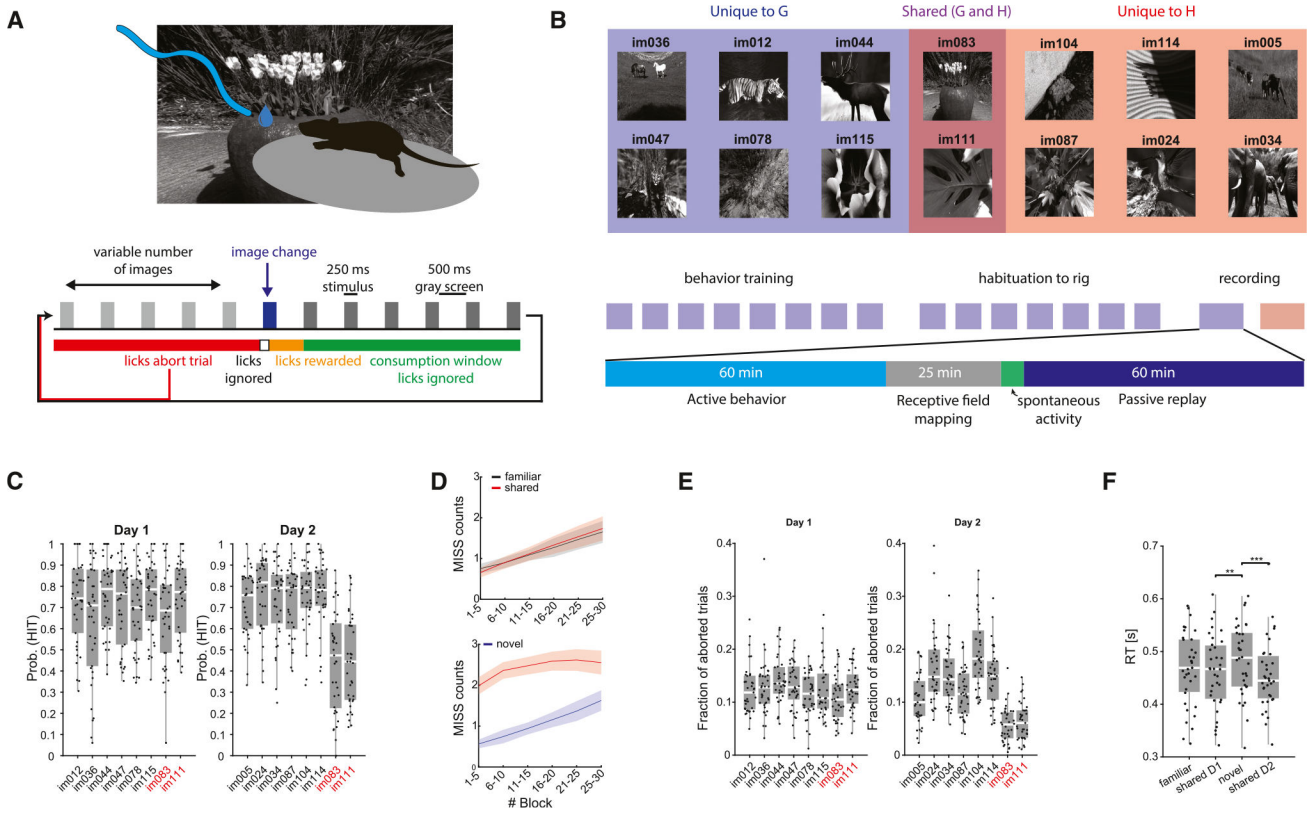


Figure 1. Visual change detection task

(A) Mice performed a visual go/no-go change detection task in which subjects are shown a continuous series of natural images and are rewarded with water dropsfor correctly reporting a change in the identity of the image. Premature licking resulted in a 300 ms time-out and the trial was restarted.

(B) Each mouse underwent two recording sessions. On the first recording day, subjects were shown a familiar image set (G) to which they had been exposed during the preceding training. On the second recording day, mice were shown a different image set (H) comprising six novel images and two familiar images from the previous set.

(C) Hit rate distributions for the different mice for each image on day 1 (left) and day 2 (right). On day 2, hit rates of shared images were significantly lower comparedto the novel images ($p < 0.001$, Kruskal-Wallis with Tukey-Kramer *post hoc* tests; $n = 38$ sessions on day 1 and 37 sessions on day 2).

(D) Top: average (mean ± SEM) number of MISS trials in five-trial blocks for familiar (black) and shared (red) images on day 1. Bottom: same, for novel (blue) and shared (red) images on day 2.

(E) Fraction of aborted trials per image (calculated over the overall aborted trials) for the various images on day 1 (left) and day 2 (right). Note that mice showed lower rates of premature licking on day 2 when presented with familiar images ($p < 0.001$ for both shared images on day 2, Kruskal-Wallis with Tukey-Kramer post *hoc* tests).

(F) Distribution of average reaction time (RT) per image type across the two recording days. RTs were significantly shorter on day 2 for shared images compared tonovel images

($p < 0.001$, Friedman repeated measure analysis with Tukey-Kramer *post hoc* tests; $n = 35$ sessions from subjects measured across both days).

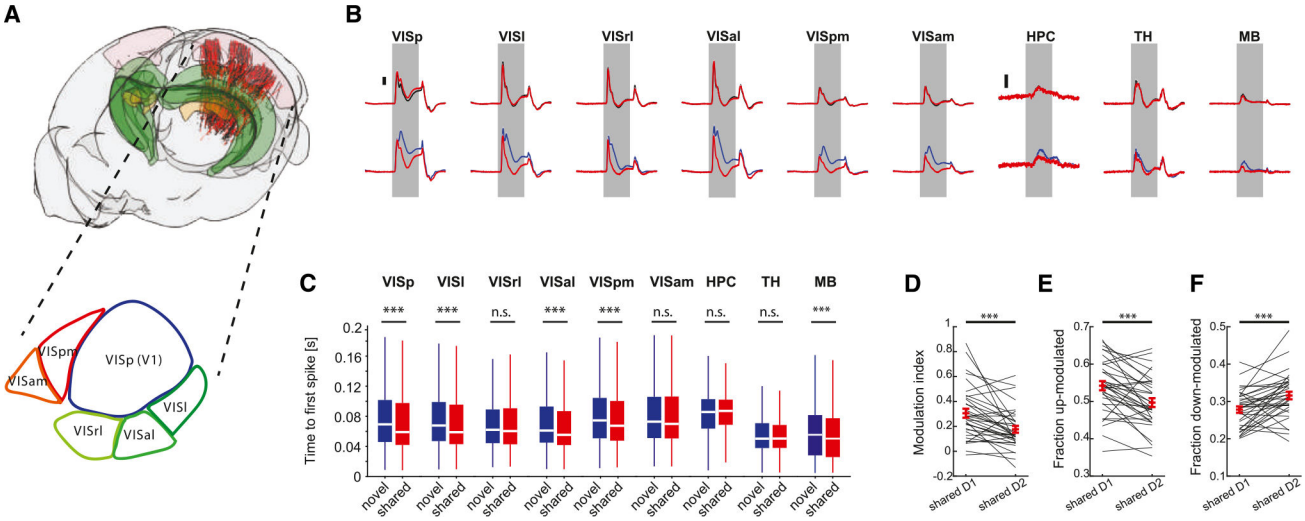


Figure 2. Novelty induced the redistribution of unit activity

(A) Top: cartoon of the mouse brain showing the location of neurons recorded across all sessions on day 1 (n = 38 sessions, black dots) and day 2 (n = 37 sessions, red dots). Visual cortex, hippocampus, and visual thalamus are depicted in pink, green, and yellow, respectively. Bottom: schematic illustration of the mouse visual cortex.

(B) Baseline subtracted mean responses of units in the different areas included in the dataset to familiar (black) and shared (red) images on day 1 (top) and novel(blue) and shared (red) images on day 2 (bottom). Note the differences in response magnitude between novel and shared images on day 2. Scale bars, 0.5 Hz for hippocampus and 2 Hz for all other areas (n = 961–7,800 neurons per area).

(C) Distribution of median lags to first spike after image presentation for novel vs. shared images on day 2. Note the significantly shorter lags in response to sharedimages for the majority of the visual cortical areas.

(D) Comparison of modulation index to shared images on day 1 and day 2 (***$p < 0.001$, Wilcoxon signed-rank test, n = 35 mice recorded on both days).

(E and F) Same as (D), for the fraction of significantly up-modulated cells and down-modulated cells, respectively.

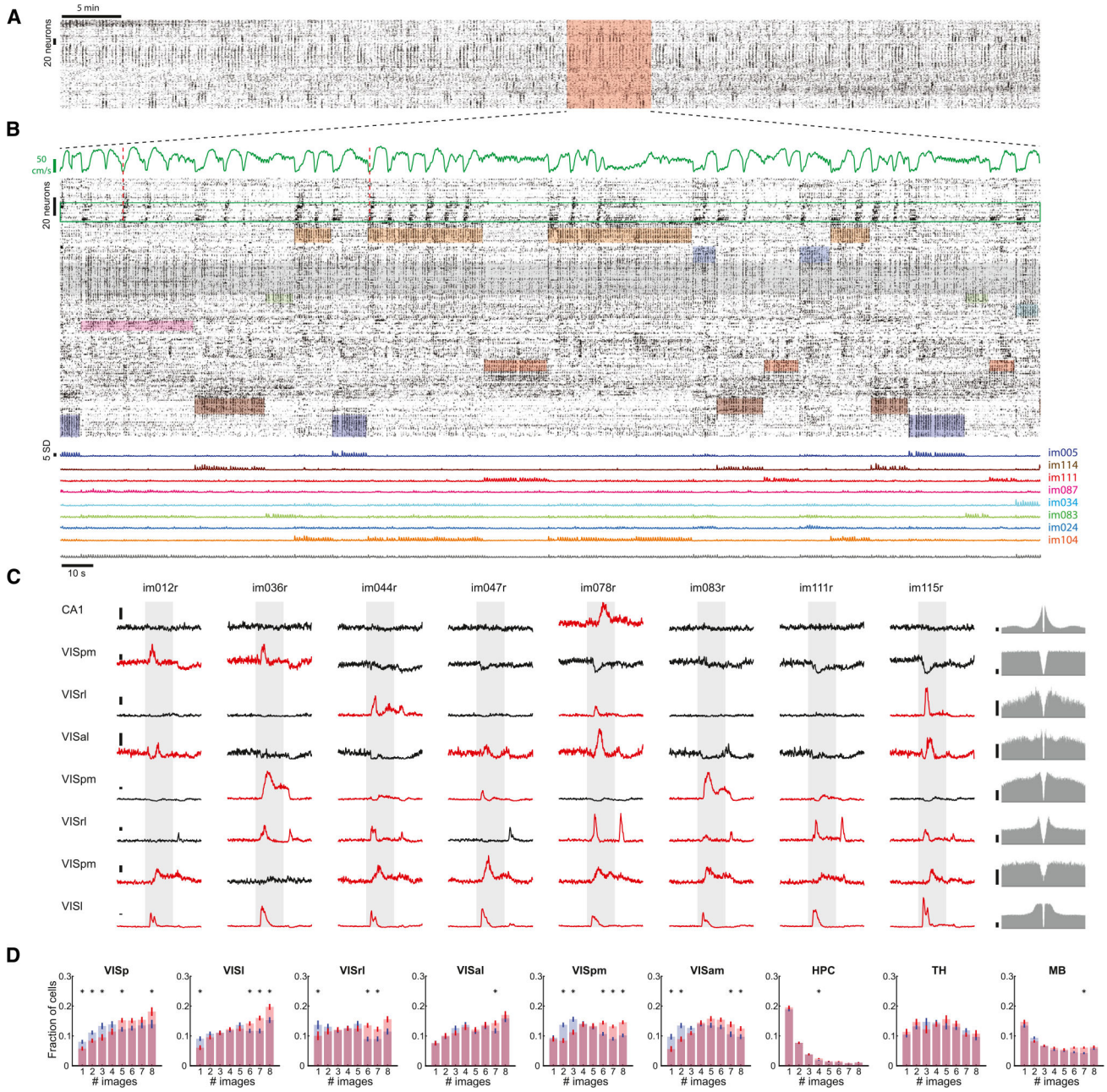


Figure 3. Heterogeneous unit activity in response to natural images

(A) Spiking activity from the active part of one example session sorted based on correlated activity between neurons using Rastermap.

(B) Magnified view of a 5 min epoch in (A) highlighted in red. Top: running speed. Middle: stimulus-specific unit activity in response to natural images. Shaded colored areas mark windows of different image presentation and are vertically restricted to clusters of neurons that preferentially respond to that stimulus. Shaded gray area marks a cluster of non-specific neurons that fire in response to all eight stimuli. Green box highlights neurons that preferentially fire at the initiation of running (two examples are marked by dashed red lines). Bottom: average Z-scored activity of the clusters highlighted above.

(C) Peri-stimulus time histograms (PSTHs) from one example session showing the responses of eight units to the different natural images presented on day 1 sorted by their degree of selectivity. Significantly up-modulated PSTHs are shown in red. Gray shaded area, 250 ms stimulus presentation window. The units' autocorrelograms are shown on the right. Scale bars, 5 Hz.

(D) Average (mean ± SEM) fraction of cells modulated by different numbers of images across the different major brain areas included in the datasets for day 1 (blue) and day 2 (red). Significant differences between day 1 and day 2 are marked with an asterisk ($p < 0.05$, Wilcoxon rank-sum test; $n = 38$ sessions on day 1 and 37 sessions on day 2).

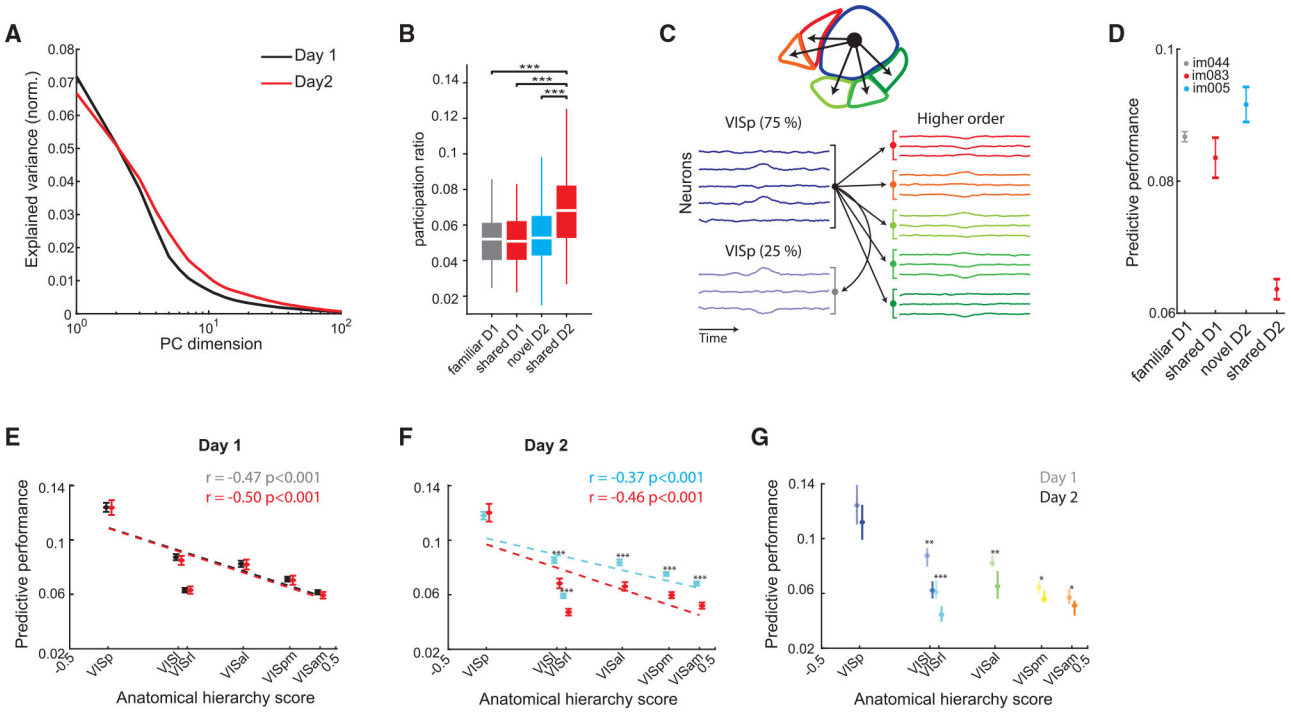


Figure 4. Novelty perturbs the representations of familiar stimuli and their propagation along the visual cortical hierarchy

(A) Example eigenspectra of responses to shared stimuli on day 1 (black) and day 2 (red).

(B) Distribution of participation ratio quantifying the intrinsic dimensionality of visual cortex population activity in response to the different image types(***$p < 0.001$, Kruskal-Wallis with Tukey-Kramer *post hoc* tests).

(C) Illustration of the cross-validated ridge regression model used VISp activity to predict the activity in higher-order visual areas or a withheld VISp subpopulation.

(D) Performance (mean ± SEM across folds) of a cross-validated ridge regression model trained to predict VISam activity from VISp activity in an example mouse for a familiar image and a novel image on day 1 and 2, respectively, as well as the same shared image on both days. Note the decreased performance for the shared image on day 2.</p/>(E) Predictive performance (mean ± SEM) for all higher-order visual areas, as well as a held-out VISp population for familiar (gray) and shared (red) images on day 1 (familiar/shared differences are not significant; $n = 26–30$ sessions per area). Dotted lines, linear regression lines.

(F) Same as (E), for novel (cyan) and shared (red) images on day 2. Predictive performance of activity during the presentation of shared images was significantly lower than that of novel images in all higher-order visual areas, but not in VISp ($p < 0.001$, Wilcoxon rank-sum test; $n = 31–33$ sessions per area).

(G) Comparison of predictive performance (median ± 95% confidence intervals) for shared images on day 1 (opaque colors) and day 2 (bright colors) for the different visual cortical areas, sorted by their anatomical hierarchy score (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, Wilcoxon rank-sum test with Bonferroni-Holm corrections).

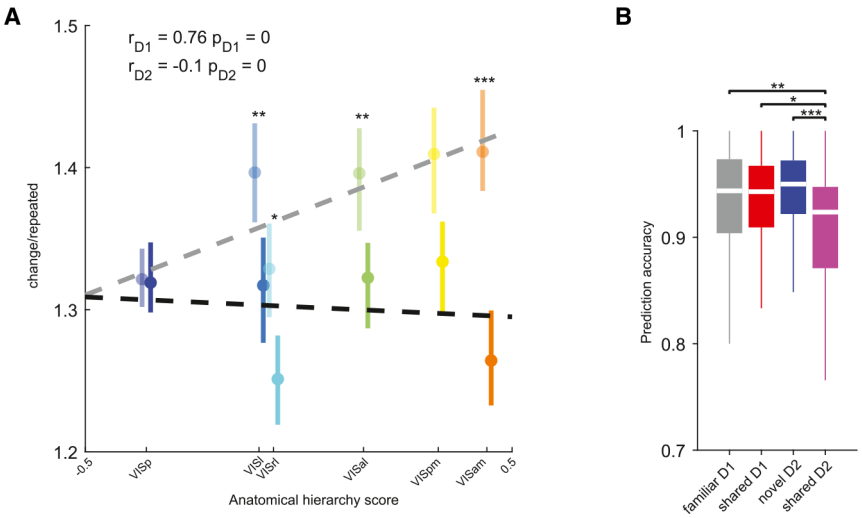


Figure 5. Novelty affects adaptation to familiar stimuli

(A) Ratio of responses (median ± 95% confidence intervals) to changed/repeated presentations of shared images on day 1 (opaque) and day 2 (bright colors) during the active task, plotted against each area's anatomical hierarchy score. Ratios were significantly lower on day 2 in higher-order visual areas, but not in VISp (**$p < 0.01$, ***$p < 0.001$, Bonferroni-Holm corrected Wilcoxon rank-sum test; see Figure S7 for sample size summary). Pearson's correlation coefficients and p values are indicated in the top left corner. Regression lines are plotted as dashed lines in the respective color.

(B) Prediction accuracy of a cross-validated linear classifier trained to distinguish changed and repeated presentations of the same image from normalized spike counts of all visual cortex neurons during the active task (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, Kruskal-Wallis with Tukey-Kramer *post hoc* tests; $n = 38$ sessions on day 1 and 37 sessions on day 2).

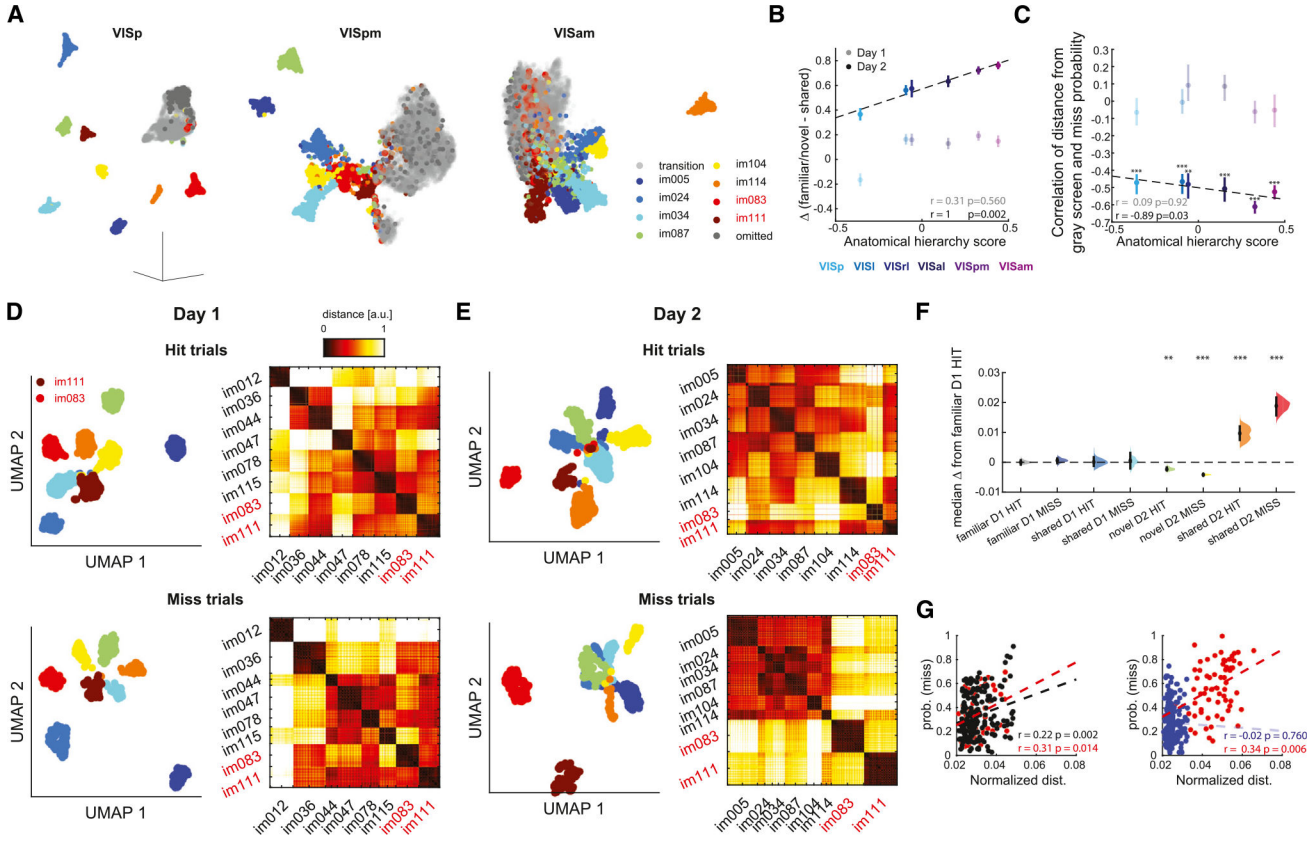


Figure 6. Differential embedding of novel and familiar neural representations

(A) Left: example UMAP embedding of responses to the different natural images presented on day 2 from primary visual cortex (VISp). Middle and right: same, but for two higher-order visual cortex areas (VISpm and VISam, respectively). Note that, while VISp responses to different images form distinct clusters, higher-order visual areas show a less clear separation. Also note that in higher-order visual areas responses to shared images (im083 and im111, red and brown, respectively) are embedded near or within the responses to the gray screen (gray).

(B) Anatomical hierarchy score of the different visual areas plotted against the average distance from the gray cluster (mean ± SEM) of shared images subtracted from the distance from the gray cluster of familiar (day 1) or novel (day 2) images. The color code for the different areas is shown below (n = 38 day 1 sessions and 37 day 2 sessions).

(C) Anatomical hierarchy score, plotted against the correlation between the distance of an image from the gray screen and the miss probability for that image (averaged across all images in a session). Significant correlations were observed only on day 2. Same color code as in (B) (***$p < 0.001$; **$p < 0.01$; Spearman's correlation; n = 38 and 37 sessions on day 1 and 2, respectively).

(D) Top left: two-dimensional UMAP embedding of neural responses (normalized spike counts within 50 ms windows during image presentation) of visual cortex units for HIT trials on day 1 from one example mouse. Responses are color coded by image identity. Top right: dissimilarity matrix showing the Euclidean distance between the embedding coordinates of the responses. Bottom: same, for MISS trials.

(E) Same as (D), but for day 2. Note that on day 2, neural responses to familiar images (red and brown clusters) are embedded farther away from novel images, particularly on MISS trials.

(F) Effect size estimate of average centroid distance depicted as the distribution of differences between the medians of each group computed from 5,000 bootstrapped resamples and the median of familiar image HIT trials on day 1. Black bars depict 95% CIs (**$p < 0.01$, ***$p < 0.001$ to familiar day 1 HIT, Kruskal-Wallis with Tukey-Kramer post hoc tests).

(G) Left: miss probability on day 1 plotted against average centroid distance for familiar images shown on day 1 only (black) or images shared across both days (red). Right: same, for novel images (blue) and familiar images shown on day 2. Pearson's correlation and p values are indicated in the bottom right.

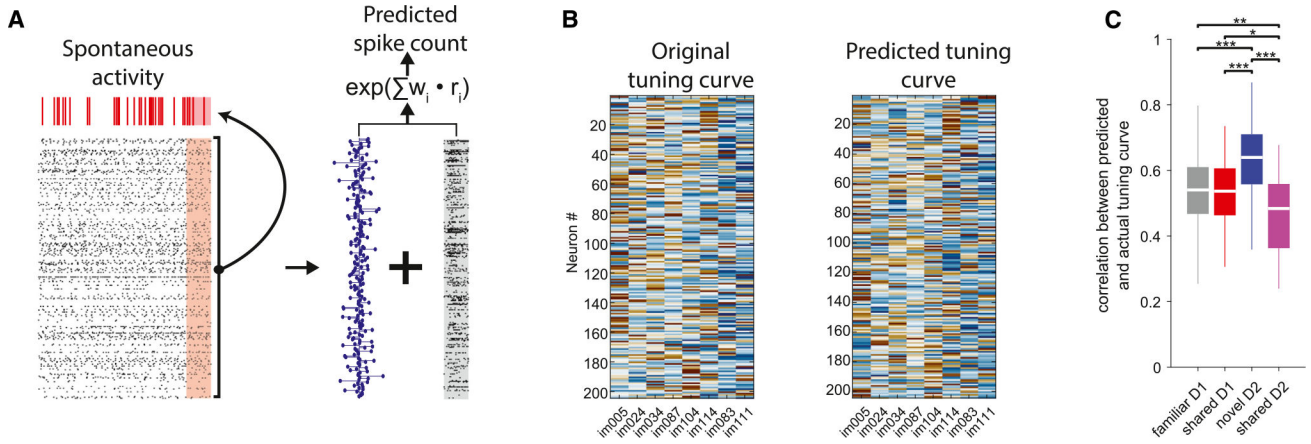


Figure 7. Differential participation of neurons modulated by familiar and novel images in spontaneous activity assemblies

(A) Left: a cross-validated generalized linear model was trained to predict the firing rate of a withheld visual cortex neuron (top raster plot, red) in 250 ms bins of spontaneous activity (red shaded area) from the firing rates of the remaining visual cortex neurons (bottom raster plot, black). Right: the optimized model weights (blue stem plot) were used in combination with the stimulus-evoked firing rates of visual cortex neurons (gray shaded raster) to predict the firing rate of the withheld neuron to that stimulus.

(B) Example actual (left) and predicted (right) tuning curves from one session.

(C) Distribution of correlation values between actual and predicted tuning curves, used to assess prediction accuracy (***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, Kruskal-Wallis with Tukey-Kramer post hoc tests; n = 38 sessions on day 1 and 37 sessions on day 2). Note decreased prediction of the same (shared) stimuli on day 2 compared to day 1.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
MATLAB 2023a	MathWorks	https://www.mathworks.com/
Analysis tools	Buzsaki Lab	https://github.com/buzsakilab/buzcode
Uniform Manifold Approximation and Projection (UMAP)	Stephen Meehan	https://www.mathworks.com/matlabcentral/fileexchange/71902
RasterMap	Stringer et al., 2023[28]	https://github.com/MouseLand/RasterMap