Cell Reports, Volume *43*

Supplemental information

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

Noam Nitzan, Corbett Bennett, J. Anthony Movshon, Shawn R. Olsen, and György Buzsáki

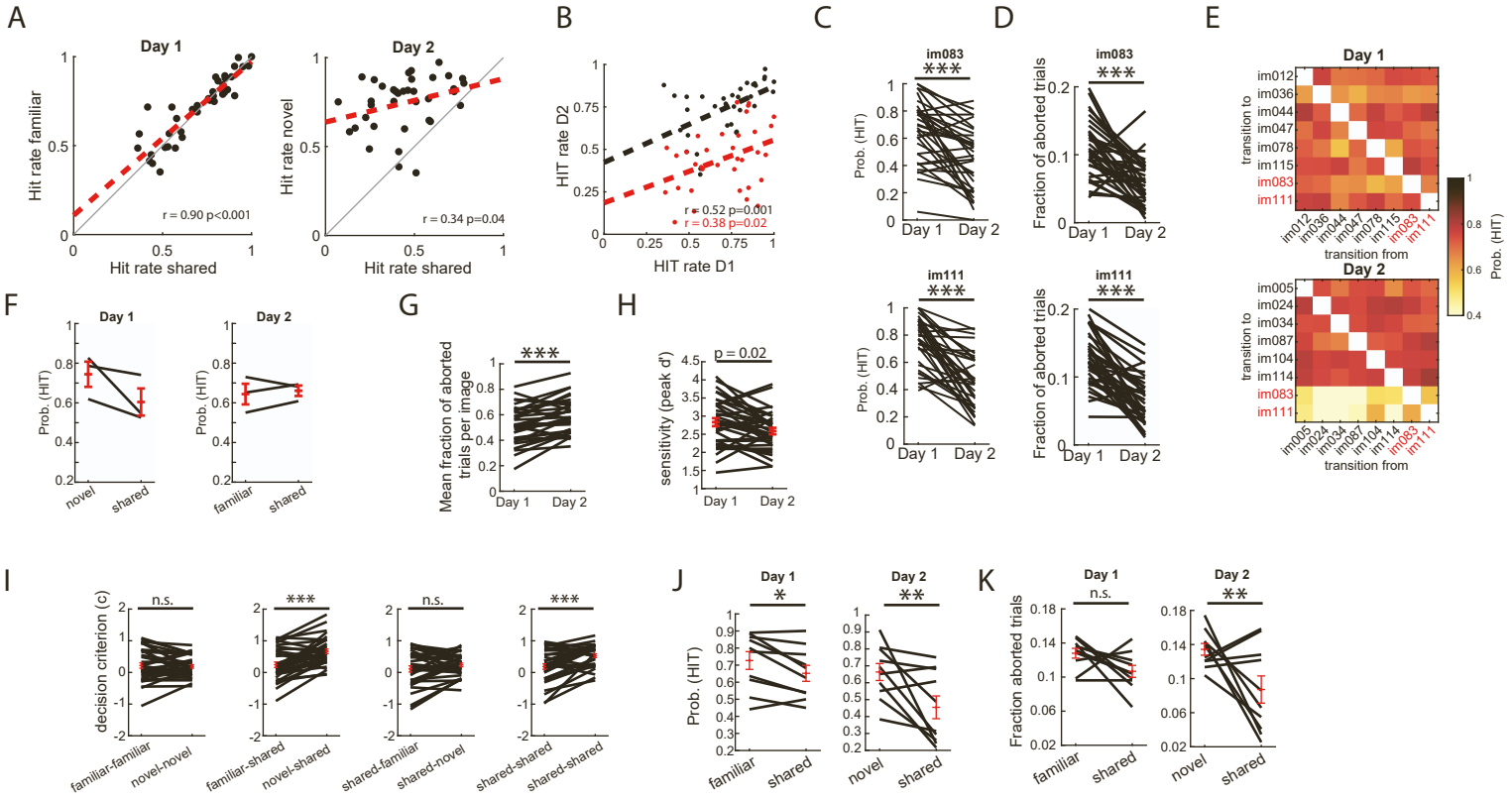


Figure S1. Task performance. A) Average hit rates of individual mice in response to shared images, plotted against hit rates to familiar images on day 1 (left) or novel images on day 2 (right). On day 1, hit rates to shared images were strongly correlated with hit rates to familiar images (Pearson's rho = 0.90, p<0.001) as indicated by the proximity of the regression line (red) and the diagonal (gray). In contrast, day 2 hit rates to shared images were only weakly correlated with hit rates to novel images (Pearson's rho = 0.34, p = 0.04). **B)** Average HIT rates for familiar images on day 1, plotted against HIT rates on day 2 for novel images (black) and for shared images on day 1, plotted against the same images on day 2 (red; n = 35 sessions from mice measured across both days). **C)** Comparison of hit rates on day 1 and day 2 for the shared images ($p < 0.001$ for both shared images on day 2, Wilcoxon signed-rank test). **D)** Comparison of aborted trials on day 1 and day 2 for the shared images ($p < 0.001$ for both shared images on day 2, Wilcoxon signed-rank test). **E)** Average transition matrix showing image-wise HIT probability conditioned on the previous image on day 1 (top) and day 2 (bottom). **F)** Hit rates in a complementary dataset where mice were trained on image-set G, and then shown image-set H on day 1 and image-set G on day 2 (differences not significant due to small sample size, Wilcoxon signed-rank test; n = 3 mice). **G)** Mean fraction of aborted trials (calculated over all trials of a given image) per mouse, for familiar images on day 1 and novel images on day 2 (p<0.001, Wilcoxon signed rank test, n = 35 sessions). **H)** Sensitivity (d') for each subject on day 1 and day 2 (n = 35 sessions; Wilcoxon signed-rank test). **I)** Decision criterion (c) of individual subjects, computed separately for transitions from familiar to familiar (novel to novel) images (left), familiar to shared (novel to shared) images (middle left), shared to familiar (shared to novel) images (middle right) or shared to shared images (right). Only transitions to shared images on day 2 showed a significant increase compared to the previous day (p<0.001, Wilcoxon signed-rank test; n = 35 sessions). **J)** Hit rates (averaged across all images of the same type) for familiar and shared images on day 1 (left) and novel and shared images on day 2 (right) in a complimentary dataset where mice (n=10) were trained on image set H (familiar image set, day 1) and tested on image set G (novel) on day 2. On day 1, differences are marginally significant (p = 0.04), but the effect size is smaller compared to day 2 (difference of means day 1 = 0.07, day 2 = 0.21; *p<0.05, **p<0.01; Wilcoxon signed-rank test) . **K)** Same as J but for the fraction of aborted trials in the same dataset (n.s., not significant; Wilcoxon signed-rank test).

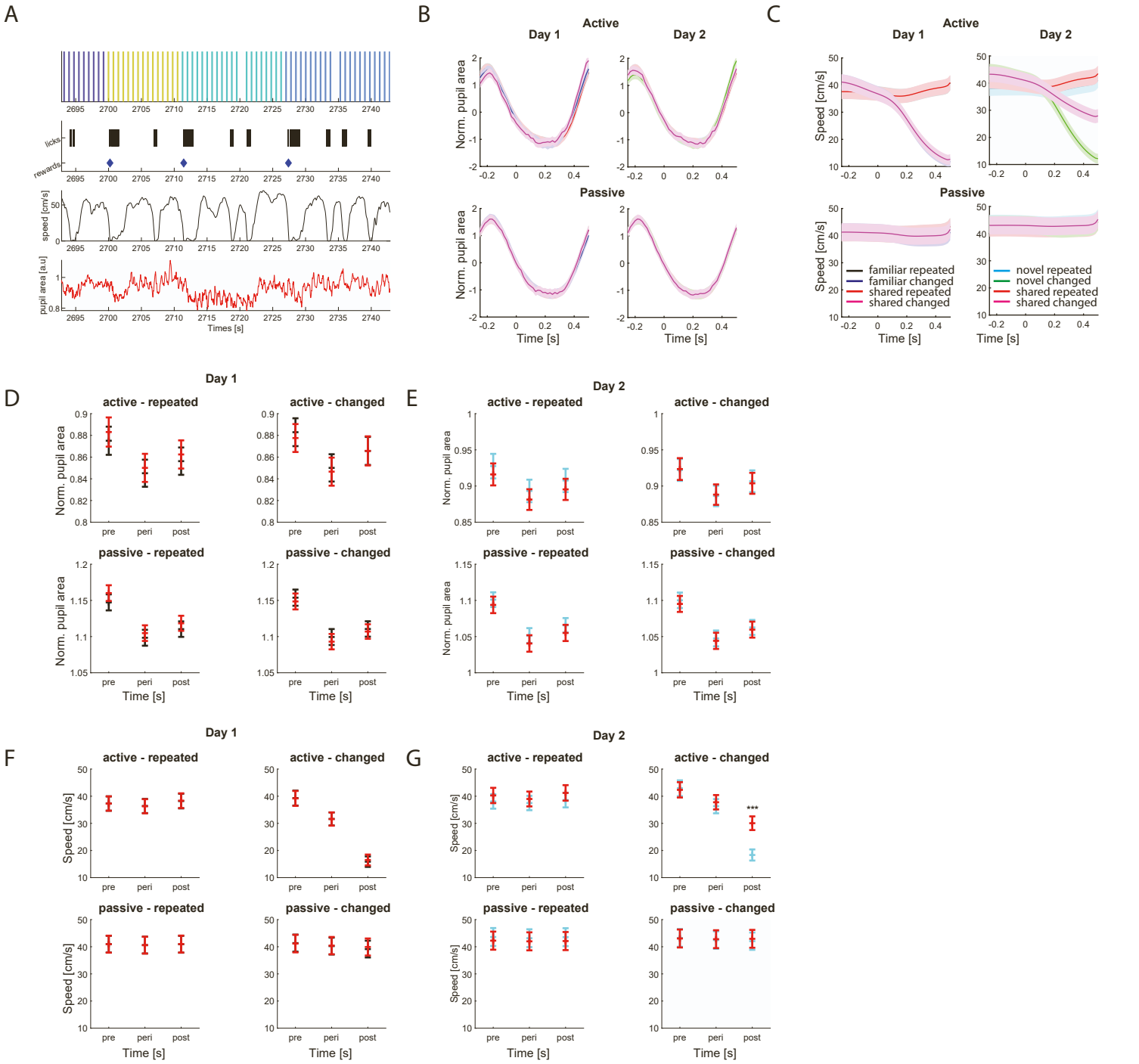


Figure S2. Behavioral correlates of the visual change detection task. A) Behavioral responses from 50 seconds of an example session. Top panel: Image presentation times, color-coded by image identity. Top middle panel: Lick (top, black ticks) and reward (bottom, blue diamonds) timing. Bottom middle panel: Running speed. Bottom panel: Pupil area. **B)** Top left: Normalized pupil area (mean ± s.e.m.) surrounding the onset of repeated familiar stimuli (black) or repeated shared stimuli (red) during the active task (n = 38 sessions). Top right: Same, for changed familiar (blue) and shared (magenta) stimuli. Bottom: Same, for passive viewing. Pupil data was z-scored normalized for visualization. **C)** Same as B but for running speed. Note the decrease in running speed after the presentation of a changed image, as mice slow down to consume the water reward. Also note that this decrease is less prominent for shared images on day 2 (top right), reflecting the higher miss probability on those trials. **D)** Top left: Average (mean ± s.e.m.) pupil area either before (-250-0 ms), during (0-250 ms) or post stimulation (250-500 ms) of repeated familiar (black) or shared (red) stimuli on day 1. No statistical differences between the groups. Top right: Same, for changed stimuli. Bottom: Same as top panels, for passive viewing. **E)** Same as D, but for novel (cyan) and shared (red) stimuli on day 2. **F-G)** Same as D-E but for running speed. Only novel and shared stimuli during the active task showed significant differences ($p<0.001$, Wilcoxon rank-sum test).

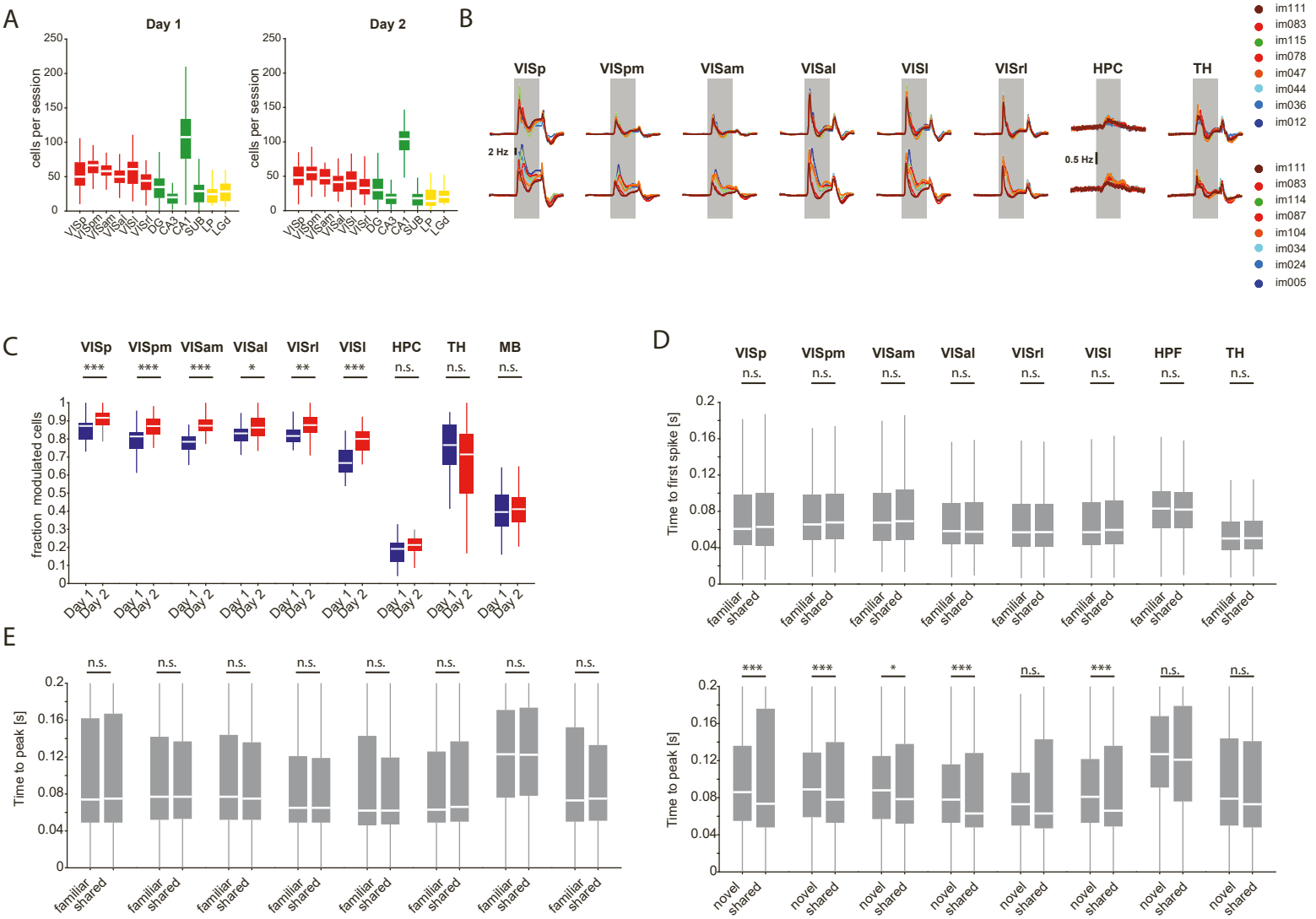


Figure S3. Novelty induced changes in unit firing. A) Distributions of cell counts per session for the main areas in the dataset on day 1 (left) and day 2 (right). **B)** Baseline subtracted mean responses of units in the different areas included in the dataset to the various images on day 1 (top) and day 2 (bottom). Note the differences in response magnitude between novel and shared images on day 2. Scale bar, 0.5 Hz for hippocampus, 2 Hz for all other areas (n = 961- 7,800 neurons per area). **C)** Distributions of fraction of modulated cells per area on day 1 (blue) and day 2 (red) (*p<0.05, **p<0.01, ***p<0.001, Wilcoxon rank sum test). **D)** Distributions of median lags to first spike after image presentation for familiar vs. shared images on day 1 (n.s., not significant, Wilcoxon rank-sum test). **E)** Left: Distributions of average peak response time after image presentation for familiar vs. shared images on day 1. Right: Same, for day 2 (*p<0.05, ***p<0.001, Wilcoxon rank-sum test).

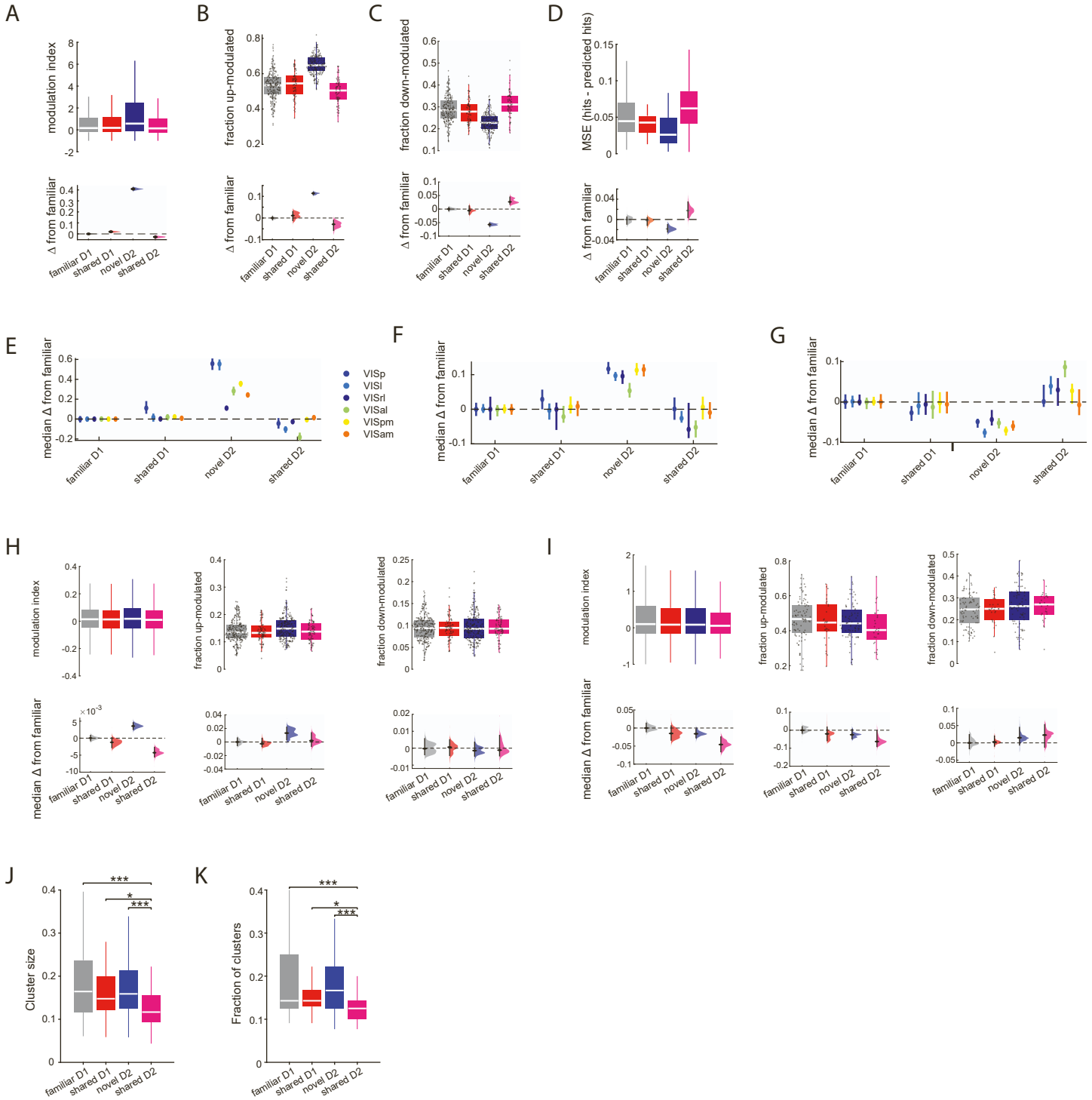


Figure S4. Comparison of unit modulation across days. A) Top: Distributions of modulation index of visual cortex neurons in response to familiar images on day 1 (grey), shared images on day 1 (red), novel images on day 2 (blue) and shared images on day 2 (magenta). Bottom: Effect size estimate depicted as the distribution of differences between the medians of each group computed from 5,000 bootstrapped resamples and the median of familiar images on day 1. Black bars depict 95% CIs. **B-C)** Same as A, but for the fraction of positively (B) and negatively (C) modulated visual cortex units. **D)** Distributions of mean squared errors between actual hit rates and hit rates predicted from withheld data by a linear regression model using the metrics in A-C as features. **E-G)** Same as A-C, but for individual visual cortical areas. Data are presented as median ± 95% confidence intervals. **H)** Same as A-C, for the hippocampus. **I)** Same as A-C, for the visual thalamus. **J)** Distributions of cluster size (fraction of neurons in a given cluster relative to the overall number of neurons in all clusters), for clusters corresponding to familiar images on day 1 (gray), shared images on day 1 (red), novel images on day 2 (blue) and shared images on day 2 (magenta). The size of clusters preferentially tuned to shared images on day 2 is significantly reduced compared to other images (***p<0.001, *p<0.05, Kruskal-Wallis with Tukey-Kramer posthoc tests; n = 38 day 1 sessions and 37 day 1 sessions). **K)** Same J, but for the overall fraction of clusters in a session.

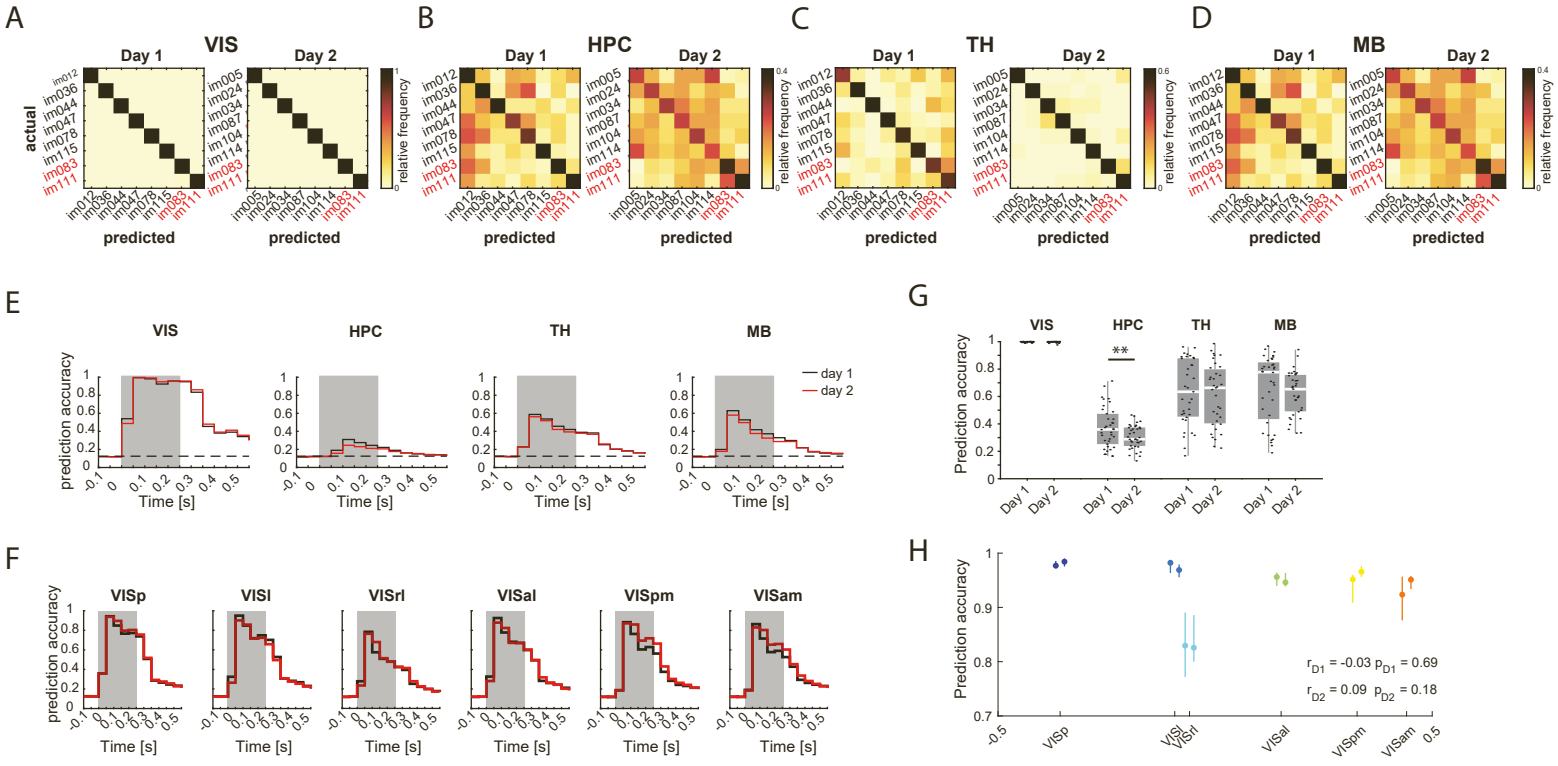


Figure S5. Image decodability. A) Average confusion matrices showing decodability of natural images from spike counts of visual cortex units over the entire 250 ms stimulus period for day 1 (right) and day 2 (left, n = 38 sessions on day 1 and 37 sessions on day 2). **B-D)** Same, for hippocampus, visual thalamus and midbrain areas. **E)** Average time-resolved prediction accuracy normalized spike counts in 50-ms windows around the stimulus period (shaded gray area) on day 1 (black) and day 2 (red). **F)** Average time-resolved prediction accuracy using normalized spike counts in 50-ms windows around the stimulus period (shaded gray area) on day 1 (black) and day 2 (red) for the various visual cortical areas. **G)** Distributions of decoding accuracy from the different areas on day 1 and day 2. Only the hippocampus showed a significant decline in prediction accuracy on day 2 ($p<0.01$, sign-rank test, n = 33-37 sessions on each day). **H)** Prediction accuracy (median ± 95% confidence intervals) of image identity for the different visual cortical areas on day 1 (left) and day 2 (right). Visual areas are ordered according to their anatomical hierarchy score, although no significant correlation was detected between prediction accuracy and anatomical hierarchy (Pearson's rho and p-values are displayed at the bottom right corner).

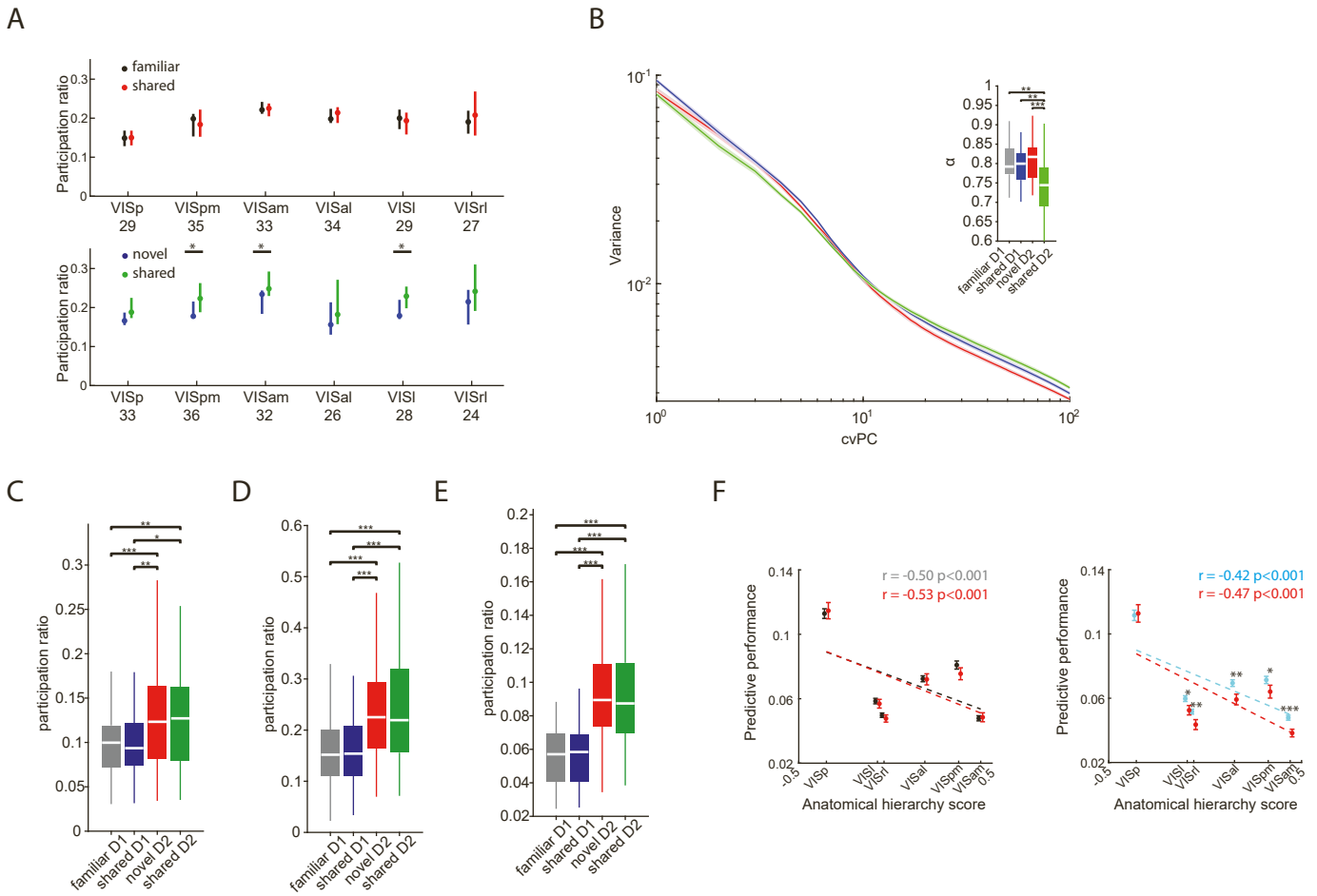
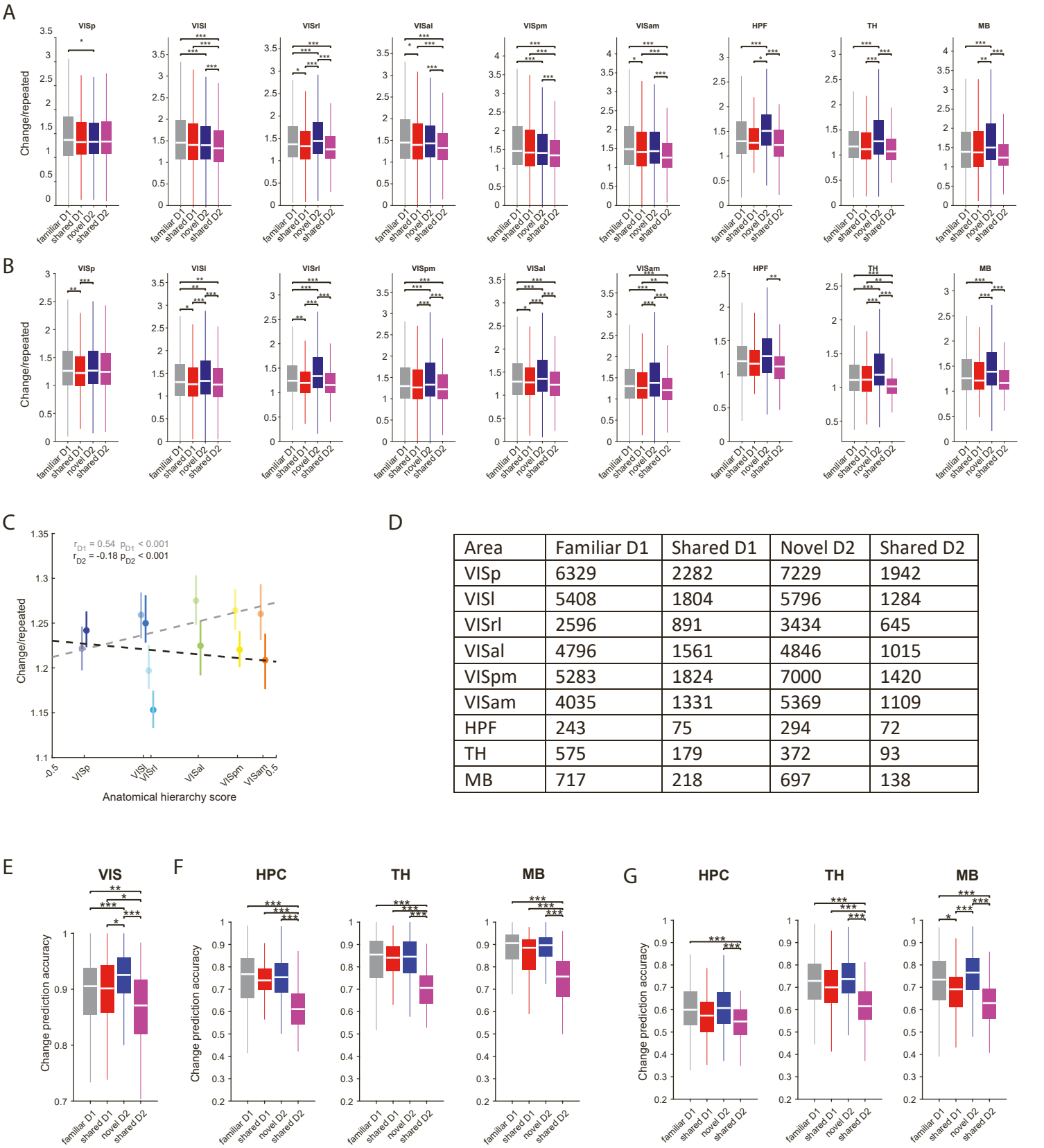


Figure S6. Population geometry measures. A) Participation ratio for the various visual areas on day 1 (top) and day 2 (bottom). The number of sessions for each region is indicated below the region's name (*p<0.05, Wilcoxon rank-sum test). **B)** Average (mean ± s.e.m.) eigenspaectra of first 100 components obtained from cross-validated principal component analysis (cvPCA) of spike-count by trial matrices of the responses to the different image groups. Inset: alpha coefficients obtained from the linear fit of $1/n^{\alpha}$ (**p<0.01, ***p<0.001, Kruskal-Wallis with Tukey-Kramer posthoc tests; day 1: n = 38 sessions, day 2: n = 37 sessions). **C)** Distributions of participation ratio quantifying the intrinsic dimensionality of hippocampal population activity in response to the different images (*p<0.05, **p<0.01, ***p<0.001, Kruskal-Wallis with Tukey-Kramer posthoc tests; n = 38 sessions on day 1 and 37 sessions on day 2). **D)** Same as C, but for visual thalamus (n = 31 sessions on day 1 and 24 sessions on day 2). **E)** Same as C, but for midbrain (n = 28 sessions on day 1 and 20 sessions on day 2). **F)** Left: Predictive performance (mean ± s.e.m) for all higher-order visual areas, as well as a held-out VISp population for familiar (gray) and shared (red) images on day 1 during passive viewing (familiar/ shared differences are not significant; n = 26-30 sessions per area). Right: Same, 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<0.05, **, p<0.01, ***p<0.001, Wilcoxon rank-sum test; n = 31-33 sessions per area).



Area	Familiar D1	Shared D1	Novel D2	Shared D2
VISp	6329	2282	7229	1942
VISl	5408	1804	5796	1284
VISrl	2596	891	3434	645
VISal	4796	1561	4846	1015
VISpm	5283	1824	7000	1420
VISam	4035	1331	5369	1109
HPF	243	75	294	72
TH	575	179	372	93
MB	717	218	697	138

Figure S7. Adaptation to familiar and novel natural images. A) Adaptation of familiar (gray), shared day 1 (red), novel (blue) and shared day 2 (magenta) stimuli during the active task, quantified as the ratio of PSTHs in response to changed and repeated stimuli, shown separately for the different areas included in the dataset (***p<0.001, **p<0.01, *p<0.05, Kruskal-Wallis with Tukey-Kramer posthoc tests; see panel D for a summary of cell numbers included in each area). **B)** Same as A, but for passive viewing. **C)** Ratio of responses (median ± 95% confidence intervals) to changed/repeated presentations of shared images on day 1 (opaque) and day 2 (bright colors) during passive viewing, plotted against each area's anatomical hierarchy score. Pearson's correlation coefficients and p-values are indicated in the top left corner. Regression lines are plotted as dashed lines in the respective colors. **D)** Summary of the numbers of units included in the analysis. We considered units with a significant PSTHs (p<0.05) and a modulation index > 0.15. **E)** Prediction accuracy of a cross-validated linear classifier trained to distinguish changed and repeated presentations of the same image from normalized spike counts of visual cortex neurons during passive viewing (***p<0.001, **p<0.01, *p<0.05 Kruskal-Wallis with Tukey Kramer posthoc tests; n = 38 sessions on day 1 and 37 sessions on day 2). **F)** Left: Prediction accuracy of a cross-validated linear classifier trained to distinguish changed and repeated presentations of the same image from normalized spike counts of hippocampal neurons during active task. Middle and right: Same, for visual thalamus and midbrain (***p<0.001, Kruskal-Wallis with Tukey Kramer posthoc tests; n = 38 sessions on day 1 and 37 sessions on day 2). **G)** Same as F, but for passive viewing.

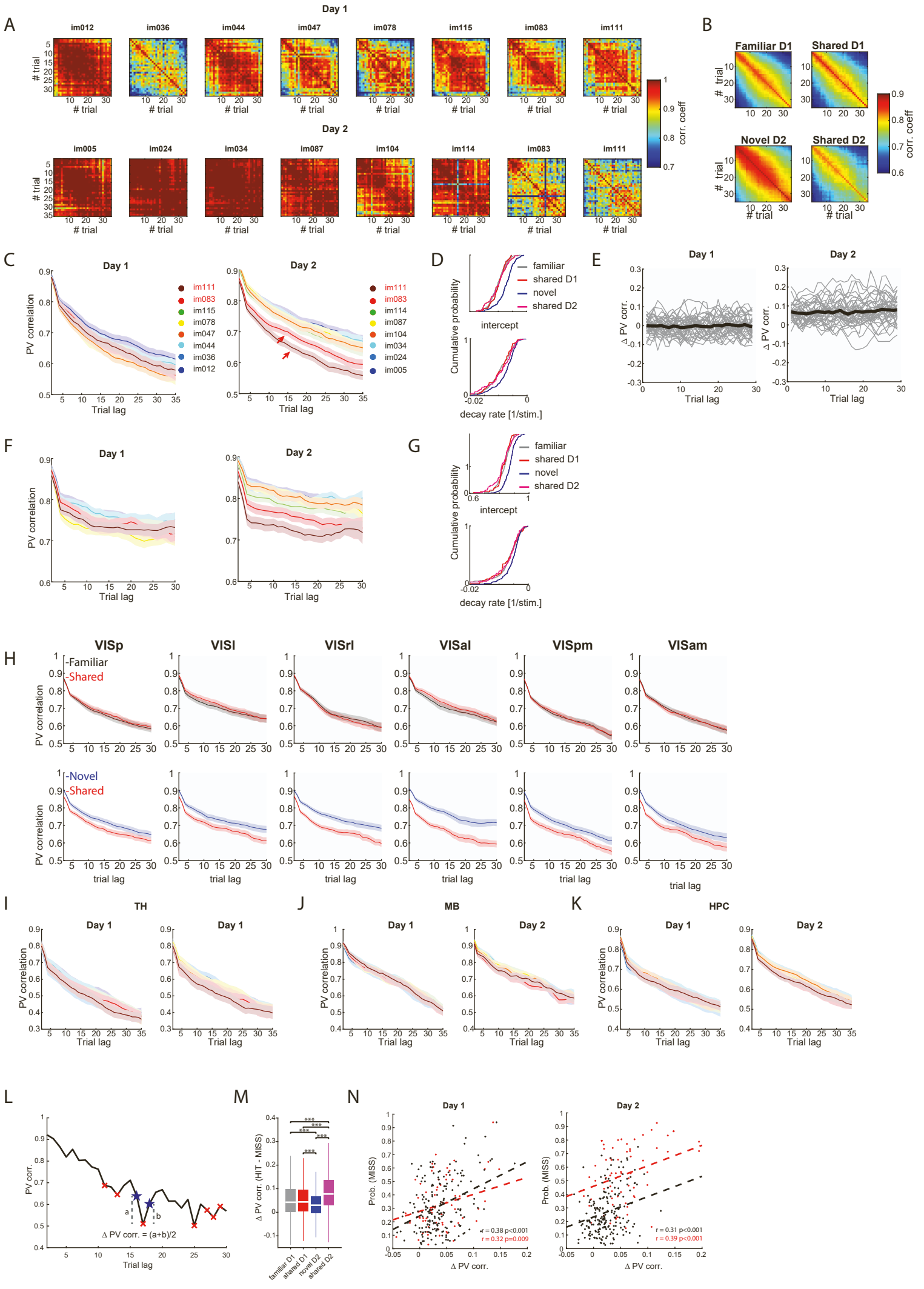

A
Day 1
im012
im036
im044
im047
im078
im115
im083
im111
Day 2
im005
im024
im034
im087
im104
im114
im083
im111
trial
corr. coeff
B
Familiar D1
Shared D1
Novel D2
Shared D2
C
Day 1
Day 2
PV correlation
Trial lag
D
familiar
shared D1
novel
shared D2
intercept
Cumulative probability
decay rate [1/stim.]
E
Day 1
Day 2
Δ PV corr.
Trial lag
F
Day 1
Day 2
PV correlation
Trial lag
G
familiar
shared D1
novel
shared D2
intercept
Cumulative probability
decay rate [1/stim.]
H
VISp
VISl
VISrl
VISal
VISpm
VISam
-Familiar
-Shared
-Novel
-Shared
PV correlation
trial lag
I
TH
Day 1
Day 1
PV correlation
Trial lag
J
MB
Day 1
Day 2
PV correlation
Trial lag
K
HPC
Day 1
Day 2
PV correlation
Trial lag
L
PV corr.
a
b
Δ PV corr. = (a+b)/2
Trial lag
M
Δ PV corr. (HIT - MISS)
familiar D1
shared D1
novel D2
shared D2
N
Day 1
Day 2
Prob. (MISS)
Δ PV corr.
r = 0.38 p<0.001
r = 0.32 p=0.009
r = 0.31 p<0.001
r = 0.39 p<0.001

Figure S8. Representational drift of familiar and novel images. A) Example population vector (PV) correlation matrices from one mouse for each stimulus, showing the correlations between spike counts of visual cortex neurons to repeated presentations of the same stimulus on day 1 (top row) and day 2 (bottom row). **B)** Mean trial-to-trial PV correlation averaged across animals and image type for familiar images shown on day 1 only (top left), shared images on day 1 (top right), novel images (bottom left) and shared images on day 2 (bottom right). **C)** Left: Average PV correlation as a function of stimulus presentation (change images only) on day 1. Right: Same, but for day 2. Note that representations of novel images decay at a slower rate compared to familiar images (red arrows). **D)** Top: Cumulative distributions of intercepts obtained from exponential fits to PV decorrelations of familiar, shared and novel images. Novel images have significantly higher intercept compared to all other groups (***p<0.001, Kruskal Wallis with Tukey Kramer posthoc tests). Bottom: Same for beta coefficients. Novel images decay significantly slower compared to all other groups (novel-familiar, p<0.001; novel-shared D1, p<0.01; novel-shared D2, p<0.05, Kruskal Wallis with Tukey Kramer posthoc tests). **E)** Left: Difference between PV correlation of responses to familiar and shared images on day 1, shown for individual mice (gray lines; n=38). Mean difference is plotted in black. Note that for all subjects, the differences are centered around zero. Right: Same, but for PV correlations to familiar images subtracted from those of novel images on day 2. Note that for the vast majority of subjects, those differences are centered above zero (n=37 sessions). **F-G)** Same as B-C, but for passive viewing. Novel images have significantly higher intercept compared to all other groups (***p<0.001, Kruskal Wallis with Tukey Kramer posthoc tests). and the PV correlation decay of novel images is significantly slower compared to all other groups (novel-familiar, p<0.01; novel-shared D1 and D2, p<0.05; Kruskal Wallis with Tukey Kramer posthoc tests). **H)** Top: Average PV correlation of familiar (black) or shared (red) stimuli as a function of stimulus presentation on day 1 shown separately for the different visual cortical areas. Bottom: Same, for novel (blue) and shared (red) stimuli on day 2. **I)** Left: Average PV correlation as a function of stimulus presentation for visual thalamus neurons on day 1. Right: Same, for day 2 . **J-K)** Same as H, for midbrain and hippocampus, respectively. **L)** Example PV correlation from one session. MISS trials are marked with red crosses. Δ PV correlation was defined as the average distance between PV correlation on a MISS trial to the immediately preceding and following HIT trials (blue stars). **M)** Distributions of Δ PV correlation values for the different image types on both days. On day 2, the amplitude of drop in PV correlation of shared images was significantly larger compared to the same images on the previous day (***p<0.001, Kruskal-Wallis with Tukey Kramer posthoc tests; n = 38 sessions on day 1 and 37 sessions on day 2). **N)** Left: Average decrease of PV correlation per session, plotted against MISS probability on that session for familiar (black) and shared (red) images. Pearson's correlation and the associated p-values are indicated in the bottom right corner. Right: Same, for day 2

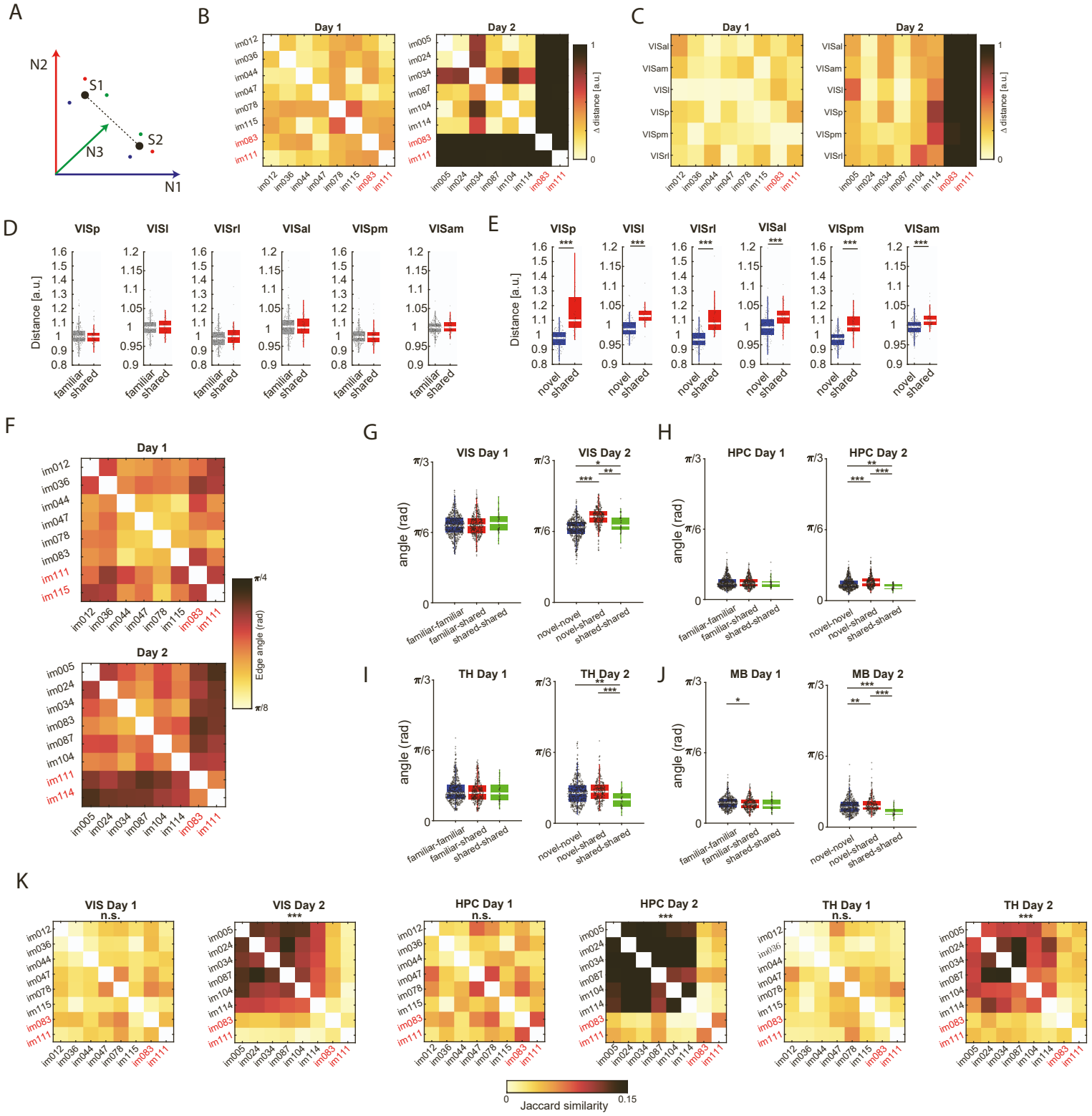


Figure S9. Geometry of population activity for novel and familiar stimuli. A) Schematic illustration of dissimilarity measurement. Each colored dot represents the mean firing rate of neurons 1-3 (N1-N3) to stimulus 1 (S1) or 2 (S2). The response centroid is marked by a black circle and the Euclidean distance between the two centroids is shows as a dotted line. **B)** Dissimilarity matrices showing the neuronal distance (Frobenius norm) between the normalized spike counts of visual cortex neurons (VISal) in response to the different images from one example sessions on day 1 (left) and day 2 (right). Note that on day 2, familiar images are more remote than novel images, as compared to the distances within novel image representations. **C)** Average distance between each image and all other images, for the 6 visual areas in the dataset on day 1 (left) and day 2 (right). Note the increase in average distance for familiar images on day 2 (n=37 mice). **D)** Distributions of average dissimilarity between each image and all other images for familiar and shared images on day 1 (differences not significant). **E)** Same, for novel and shared images on day 2 (***p<0.001, Wilcoxon rank sum test; n = 38 sessions on day 1 and 37 sessions on day 2). **F)** Edge angles between pairs of different images shown on day 1 (left) and day 2 (right) from an example mouse. **G)** Distributions of visual cortex edge angles between familiar and shared images on day 1 (left) and between novel and shared images on day 2 (right). **H-J)** Same as H, for HPC, TH and MB, respectively (*p<0.05, **p<0.01, ***p<0.001, Kruskal-Wallis with Tukey-Kramer post-hoc tests). **K)** Jaccard similarity of neurons significantly modulated by the different images and the different areas on day 1 (left) and day 2 (right). Note the low similarity between novel and familiar images in all areas on day 2, indicating a reduced overlap between the neurons representing those images (n.s., not significant; ***p<0.001, Wilcoxon rank-sum test between familiar/novel and shared).

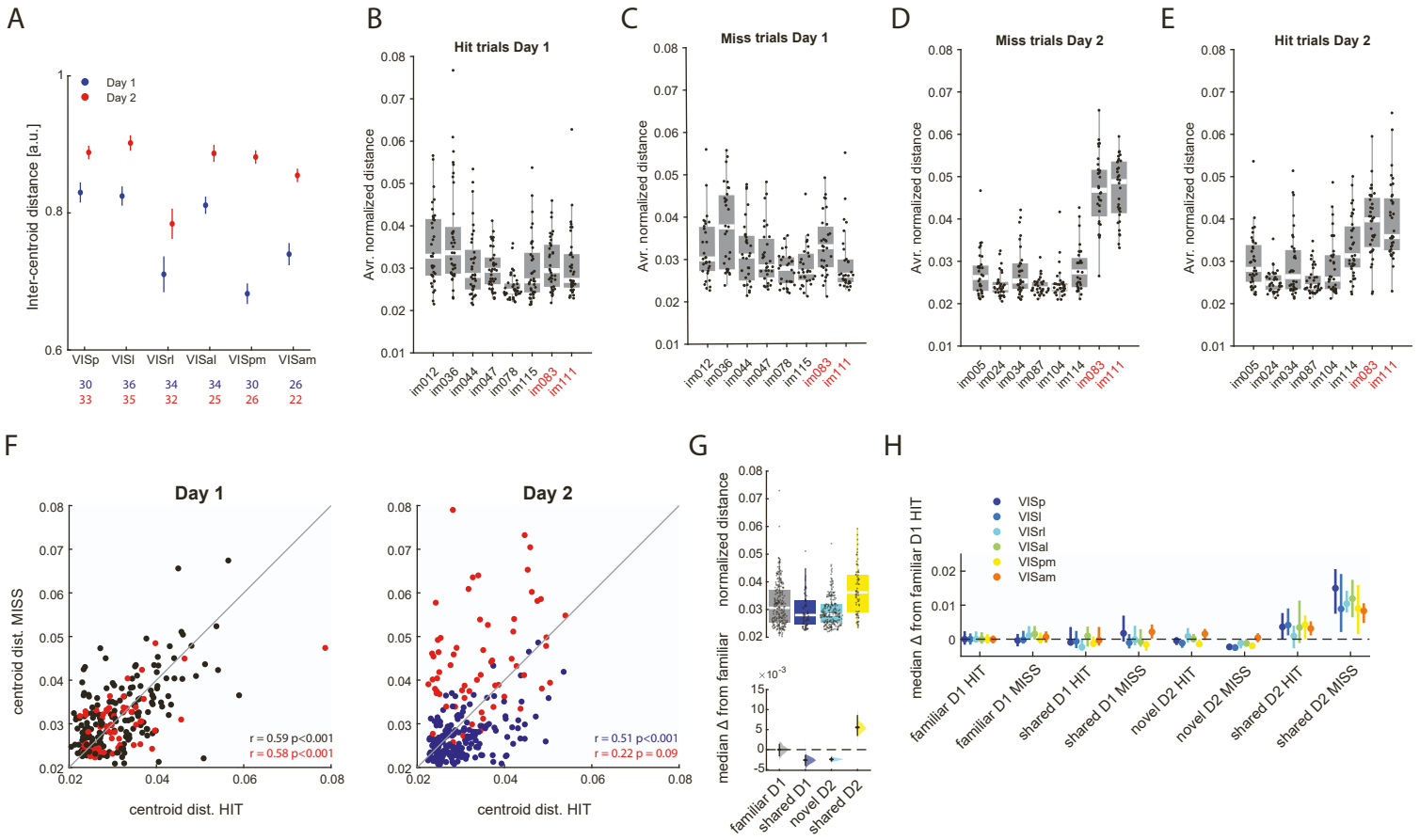


Figure S10. Embedding statistics of familiar and novel clusters. A) Average between-cluster distances (mean ± s.e.m.) between the centroid of one image cluster to all other centroids across the different visual areas on day 1 (blue) and day 2 (red). The number of sessions per region is indicated below in the corresponding color. **B)** Distributions of between-cluster distances during HIT trials on day 1 shown for each image separately. Shared images are highlighted in red. **C)** Same as B but for MISS trials on day 1. **D-E)** Same as B-C but for day 2. **F)** Average cluster distances on HIT trials plotted against average cluster distances on MISS trials on day 1 (left) and day 2 (right). On each day, shared images are plotted in red while all other images are shown in black. Note that on day 2, shared images show greater separation on MISS trials. **G)** Top: Distributions of between-cluster distances during passive image presentation on day 1 and day 2, shown separately for familiar, shared and novel images. Bottom: Effect size estimate depicted as the distribution of differences between the medians of each group computed from 5,000 bootstrapped resamples and the median of familiar images on day 1. Black bars depict 95% CIs. **H)** Between-cluster distances (median ± 95% confidence intervals) during HIT and MISS trials on day 1 and day 2, shown separately for familiar, shared and novel images and separately for the different visual cortex areas (legend).