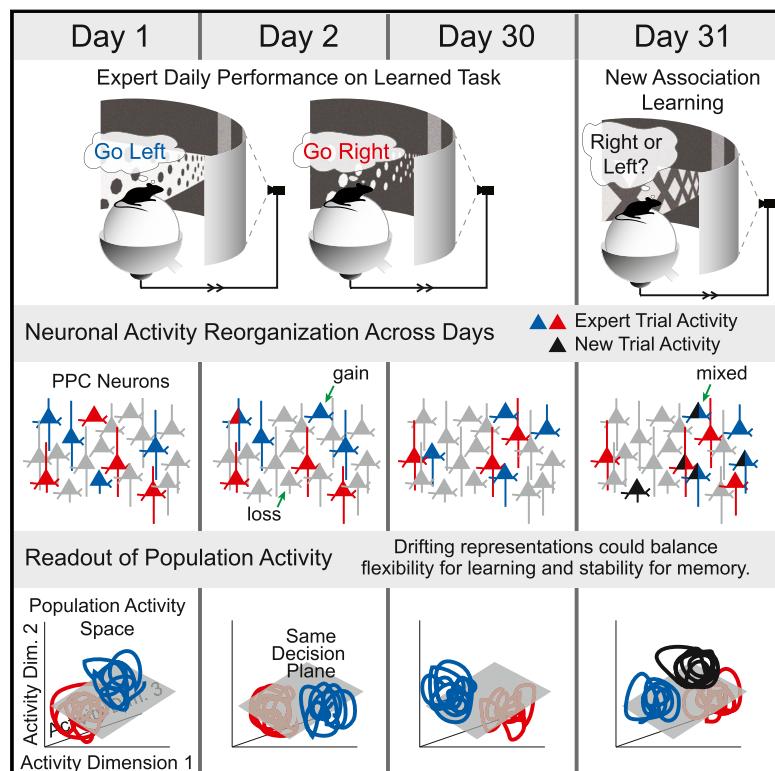


Dynamic Reorganization of Neuronal Activity Patterns in Parietal Cortex

Graphical Abstract



Authors

Laura N. Driscoll, Noah L. Pettit,
Matthias Minderer, Selmaan N. Chettih,
Christopher D. Harvey

Correspondence

harvey@hms.harvard.edu

In Brief

Contrary to the idea that representations of sensory stimuli or the activity patterns that accompany motor actions are stable, neuronal representations in the parietal cortex can change across days, possibly allowing for the tradeoff between stable encoding of information and flexibility for incorporating new information.

Highlights

- Parietal cortex neurons were tracked for a month during a navigation decision task
- Activity patterns changed relative to behavior in individual neurons over weeks
- Stable information can be read out from population activity for over a week
- Drifting representations could mediate a tradeoff between flexibility and stability

Dynamic Reorganization of Neuronal Activity Patterns in Parietal Cortex

Laura N. Driscoll,¹ Noah L. Pettit,¹ Matthias Minderer,¹ Selmaan N. Chettih,¹ and Christopher D. Harvey^{1,2,*}

¹Department of Neurobiology, Harvard Medical School, Boston, MA 02115, USA

²Lead Contact

*Correspondence: harvey@hms.harvard.edu

<http://dx.doi.org/10.1016/j.cell.2017.07.021>

SUMMARY

Neuronal representations change as associations are learned between sensory stimuli and behavioral actions. However, it is poorly understood whether representations for learned associations stabilize in cortical association areas or continue to change following learning. We tracked the activity of posterior parietal cortex neurons for a month as mice stably performed a virtual-navigation task. The relationship between cells' activity and task features was mostly stable on single days but underwent major reorganization over weeks. The neurons informative about task features (trial type and maze locations) changed across days. Despite changes in individual cells, the population activity had statistically similar properties each day and stable information for over a week. As mice learned additional associations, new activity patterns emerged in the neurons used for existing representations without greatly affecting the rate of change of these representations. We propose that dynamic neuronal activity patterns could balance plasticity for learning and stability for memory.

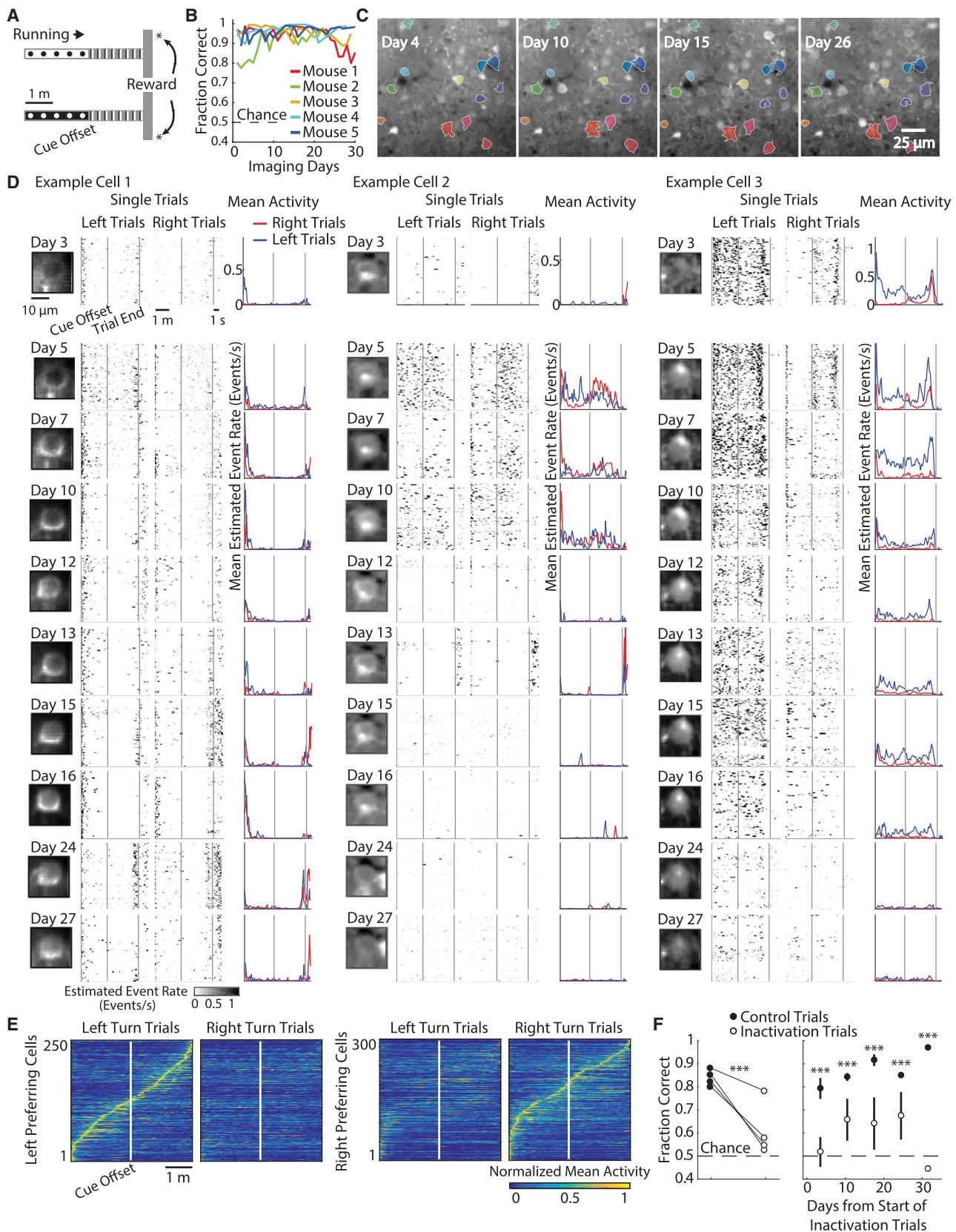
INTRODUCTION

Cortical neurons form associations between sensory stimuli and behavioral actions. For example, landmarks are associated with specific actions during navigation, and arbitrary categories of sensory cues can guide relevant behavioral responses (Freedman and Assad, 2016; Harvey et al., 2012). Past work studying the representations for sensorimotor associations has focused on brain areas at the interface of sensation and action, including posterior parietal cortex (PPC). Studies of PPC have commonly measured neuronal activity at single snapshots in time. For example, in typical experiments, one set of neurons is studied on one day and a separate population of neurons is examined on the next day. Therefore, temporal features of PPC representations across days and weeks have not been examined systematically. It is thus poorly understood whether PPC activity converges to a stable pattern following learning or whether neuronal activity patterns continue to change.

A common framework for memory proposes a direct and fixed mapping of neuronal activity with sensory stimuli and behavioral actions. In this framework, learned associations develop by linking these fixed representations (Messinger et al., 2001; Sakai and Miyashita, 1991). During learning, synaptic and other biophysical changes are hypothesized to minimize errors in the link between sensory stimuli and behavioral outputs, eventually converging to a stable solution (Ganguly and Carmena, 2009; Peters et al., 2014). This view proposes that a memory engram is a collection of the same neurons that are activated every time the learned association is recalled (Tonegawa et al., 2015). Alternatively, experimental data indicate that synaptic connections between neurons continually change over time. In some cases, only a small fraction of synaptic connections persist over weeks (Attardo et al., 2015; Stettler et al., 2006; Trachtenberg et al., 2002). Theoretical models propose that continuous change in neuronal circuits can optimize a tradeoff between stability and flexibility by sampling from multiple solutions of activity patterns and connectivity that similarly convey relevant information (Ajemian et al., 2013; Kappel et al., 2015; Rokni et al., 2007). Such changes could present computational benefits, such as limiting the likelihood of overtraining and convergence to local optima.

Recent methods to track the activity of the same neurons over days have led to studies of neuronal activity patterns over time. Studies in sensory cortex have revealed that representations of stimulus features are generally stable over the examined periods (Andermann et al., 2010; Mank et al., 2008; Margolis et al., 2012; Peron et al., 2015; Poort et al., 2015; Rose et al., 2016; Tolias et al., 2007). In motor cortex, the stability of activity patterns for generating actions is a controversial topic. Many studies have noted stability in motor cortex activity, but some have identified subtle shifts in tuning over days (Chestek et al., 2007; Ganguly and Carmena, 2009; Huber et al., 2012; Padoa-Schioppa et al., 2004; Peters et al., 2014; Rokni et al., 2007; Stevenson et al., 2011). Also, a recent study reported a mix of stable and changing features in the activity of HVC during birdsong (Liberti et al., 2016). The largest changes in neuronal activity patterns have been noted in the hippocampus. Upon repeated exposure to the same environment, place cell activity was gained and lost in individual cells (Kentros et al., 2004; Ziv et al., 2013).

These studies suggest a possible range of stability in neuronal representations, perhaps with greater stability in areas more closely related to sensation or action and less stability in the hippocampus. However, these emerging ideas are based on a relatively small number of studies that collectively have sampled



(legend on next page)

limited numbers of brain regions. In addition, some previous work focused on changes in the absence of learned behaviors, during tasks that did not require activity in the region being studied, and without systematic quantification of the relationship between activity and behavior over time (Attardo et al., 2015; Ziv et al., 2013). These limitations have arisen in part due to difficulties in tracking the activity of the same neurons over time because, with methods such as electrophysiology or supracellular resolution imaging, it can be difficult to obtain reliable metrics of cell identities across days.

Here, we tested whether arbitrary stimulus-action pairings develop stable representations in mouse PPC after learning. We tracked the activity of populations of neurons and behavioral patterns across weeks as mice performed a navigation-based task in virtual reality at near-perfect levels. We focused on activity in PPC because it is essential for performing this task and in rodents contributes to learned sensorimotor associations (Harvey et al., 2012; McNaughton et al., 1994; Nitz, 2006; Whitlock et al., 2012). Activity patterns in individual neurons changed greatly over days and weeks, such that the population of neurons with the most task-relevant information drifted over time. Despite changes in single cells, the PPC population maintained a steady state with the same statistics of population activity. Information about the task could be decoded from population activity above chance levels across days using a stable readout despite changes in individual neurons. Also, representations of newly learned cue-response relationships developed without greatly perturbing existing representations. We propose that drift in neuronal activity patterns could be important for mediating a tradeoff between stable encoding of information and flexibility for incorporating new information in PPC.

RESULTS

Tracking Behavior and Neuronal Activity over Weeks

We trained mice to perform a two-alternative forced-choice task based on navigation through a T-maze in visual virtual reality (Harvey et al., 2012; Figure 1A). At the beginning of the T-stem, mice saw one of two visual cues (white or black walls). Mice then ran through a delay portion of the T-stem in which the walls were identical between trial types. Upon reaching the T-intersection, mice reported the cue identity by making a left or right turn to receive a reward. Mice achieved expert behavioral performance that was mostly stable over weeks (Figure 1B).

We imaged the activity of layer 2/3 PPC neurons. Our field of view corresponded to the area the Allen Brain Institute has named the V1Sa area of PTLp, as revealed by separate experiments using retinotopic mapping (Mouse and Coordinate, 2016; Figure S1). For consistency with earlier work, we called this region PPC and note that it is anterior to visual area AM and overlapping with, but shifted medially from, what previous work called area A (Wang and Burkhalter, 2007). Imaging was performed every day with occasional 1-day gaps. We developed approaches to identify the same neurons on each day (Figures 1C, 1D, and S2; STAR Methods). First, we identified fluorescence signal sources (putative cells) on each day independently, on the basis of temporally correlated fluctuations between pixels, rather than manual, anatomical methods that can fail to separate nearby cells, dendrites, and axons. Cells had to have calcium transients to be identified by this method. Second, we aligned putative cells across days by matching regions-of-interest (ROIs) with similar locations and neighborhoods from one day to the next. Finally, we visually compared each identified cell across all days to ensure that cells appeared consistent in the anatomical images on each day (Figure 1D). We considered cells only on days in which they were identified with high confidence. Not all cells were identified on every day.

Necessity of PPC Activity for Post-learning Performance of the Task

The activity patterns of PPC neurons on a single day were consistent with those reported previously (Harvey et al., 2012; Morcos and Harvey, 2016). Individual neurons were transiently active, such that PPC activity tiled the duration of a trial (Figure 1E). Many of these responses were reliable and selective for a particular trial type. For example, some cells were more active on black cue-right turn trials than on white cue-left turn trials or vice versa (Figure 1E).

PPC activity was necessary for the mouse to perform the behavioral task. In a separate cohort of mice, at a location centered at PPC, we activated channelrhodopsin-2 in parvalbumin-expressing interneurons to inhibit excitatory activity on a subset of trials. Inactivation decreased the mouse's behavioral performance from ~85% correct to just above chance levels (Figure 1F). Similar results were obtained weeks after the mouse achieved plateau performance, suggesting that PPC activity was necessary for performing the task, even in the post-learning phase. These results were consistent with our earlier pharmacological inactivation experiments and other studies showing a role for rodent PPC in visual decision tasks (Goad et al., 2016;

Figure 1. Chronic Imaging during Stable Performance of a Virtual-Navigation Decision Task

- (A) Schematic of the task.
 - (B) Behavioral performance for five mice.
 - (C) Example imaging plane with a subset of cells identified across days in color. See Figure S1 for PPC coordinates.
 - (D) Example cells over weeks. Left columns: mean fluorescence image is shown. Middle columns: deconvolved fluorescence signal on correct white cue-left turn and black cue-right turn trials is shown. Right columns: mean activity on correct white cue-left turn (blue) and black cue-right turn (red) trials is shown. See Figure S2 for cell-identification protocol.
 - (E) Sorted peak-normalized mean activity of neurons with a significant peak of activity, combined from 1 day each for 5 mice.
 - (F) Left: task performance on optogenetic inactivation and control trials for 4 mice, combined across days. Right: task performance for inactivation and control trials within 7-day time bins is shown. Error bars: mean \pm SEM across mice. n = 3, 3, 4, 2, and 1 mice for the bins, respectively. ***p < 0.001; control versus inactivation trials; permutation test. Control mouse not expressing ChR2: p = 0.18.
- See also Figures S1 and S2.

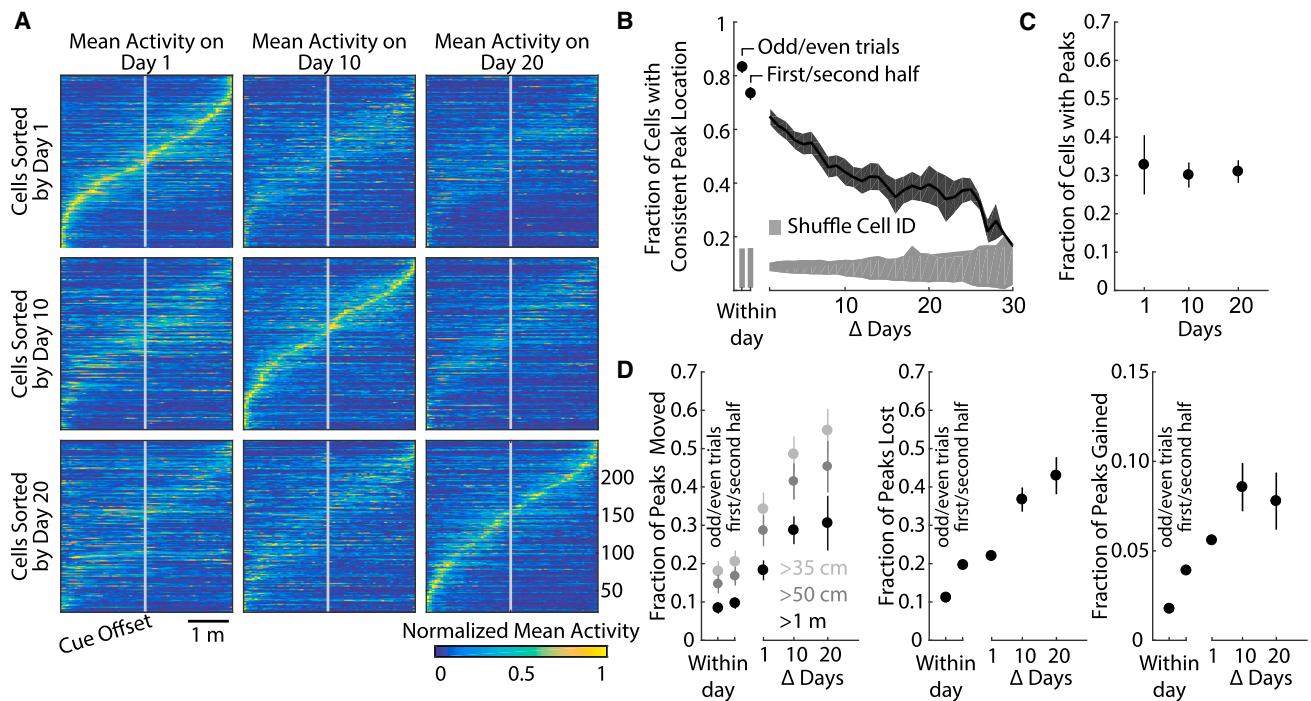


Figure 2. Reorganization of Activity within a Trial across Days

(A) Normalized mean activity of neurons identified on all 3 imaging days with a statistically significant peak in the sorted day. Sorting was the same for each day within a row and was different across rows.

(B) For cells with a highly significant (99% confidence) peak of activity on a given day, the fraction of cells that had a significant (95% confidence) peak of activity at a similar location (<70 cm shift) on a subsequent day ($p < 10^{-8}$ versus time; ANOVA). Shading: mean \pm SEM. $n = 5$ mice, except for large intervals (Figure S1A). The gray area indicates 95% confidence intervals of chance levels based on shuffling the cell IDs separately on each day.

(C) Fraction of cells with a significant peak. $p = 0.85$ versus time; ANOVA. Error bars: mean \pm SEM. $n = 5, 5$, and 4 mice, respectively.

(D) Left: for cells with a significant peak on day n and day $n + x$, the fraction of peaks that shifted by greater than 0.35 m, 0.5 m, and 1 m. Fraction moved 0.35 m versus time: $p = 0.019$; ANOVA. Center: for cells with a highly significant peak on day n , the fraction of cells that did not have a significant peak on day $n + x$ is shown. Fraction lost versus time: $p < 10^{-9}$; ANOVA. Right: for cells without a significant peak of activity on day n , the fraction of cells with a highly significant peak on day $n + x$ is shown. Fraction gained versus time: $p = 0.96$; ANOVA.

Harvey et al., 2012; Licata et al., 2017; Raposo et al., 2014). We note that, although inactivation was centered on PPC, such activity manipulations may have effects that spread beyond PPC (Otchy et al., 2015).

Reorganization of Sequential Activity across Maze Locations

To compare the activity patterns of neurons across days, we first focused on sequential activity throughout a trial. On each day, a sequence of neuronal activity was present (Figure 2A). To determine whether this sequence was the same from day to day, we sorted neurons based on where in the maze they had a reliable peak of activity. We then used the same sorting on earlier or later days (Figure 2A). Cells that had a significant peak of activity on a given day were unlikely to have a significant peak of activity at the same or nearby position over weeks (Figure 2B). Over time, the likelihood of a consistent peak position approached levels expected from a random reorganization of neuronal identities (Figure 2B). Changes in activity peaks from the first half to the second half of the session were small but greater than from measurement noise (changes quantified between odd and even trials within a session; $p = 0.017$). The changes across time resulted

from cells with a peak of activity on one day either losing that peak or having a shift in the peak's location on subsequent days, both of which increased in likelihood with time from when a peak was identified (Figure 2D). The loss of peaks of activity was offset by an approximately constant rate at which cells initially lacking a peak of activity gained an activity peak (Figure 2D), resulting in a consistent fraction of active cells with activity peaks on each day (Figure 2C). Together, these results indicate that activity patterns changed over time, with major reorganization occurring over several weeks.

Different Populations of Neurons with Trial-Type-Specific Activity Patterns across Days

We also investigated changes in activity patterns related to information about the trial type. Specifically, we asked whether the neurons that had different activity patterns on trials with different cues and choices were the same across time. For each neuron on each day, we used a decoder to quantify how well that neuron's activity predicted trial type across the duration of a trial (white cue-left turn versus black cue-right turn). On a given day, a significant fraction of active neurons had a decoding accuracy above chance ($29.1\% \pm 1.1\%$ of neurons; $p < 0.05$

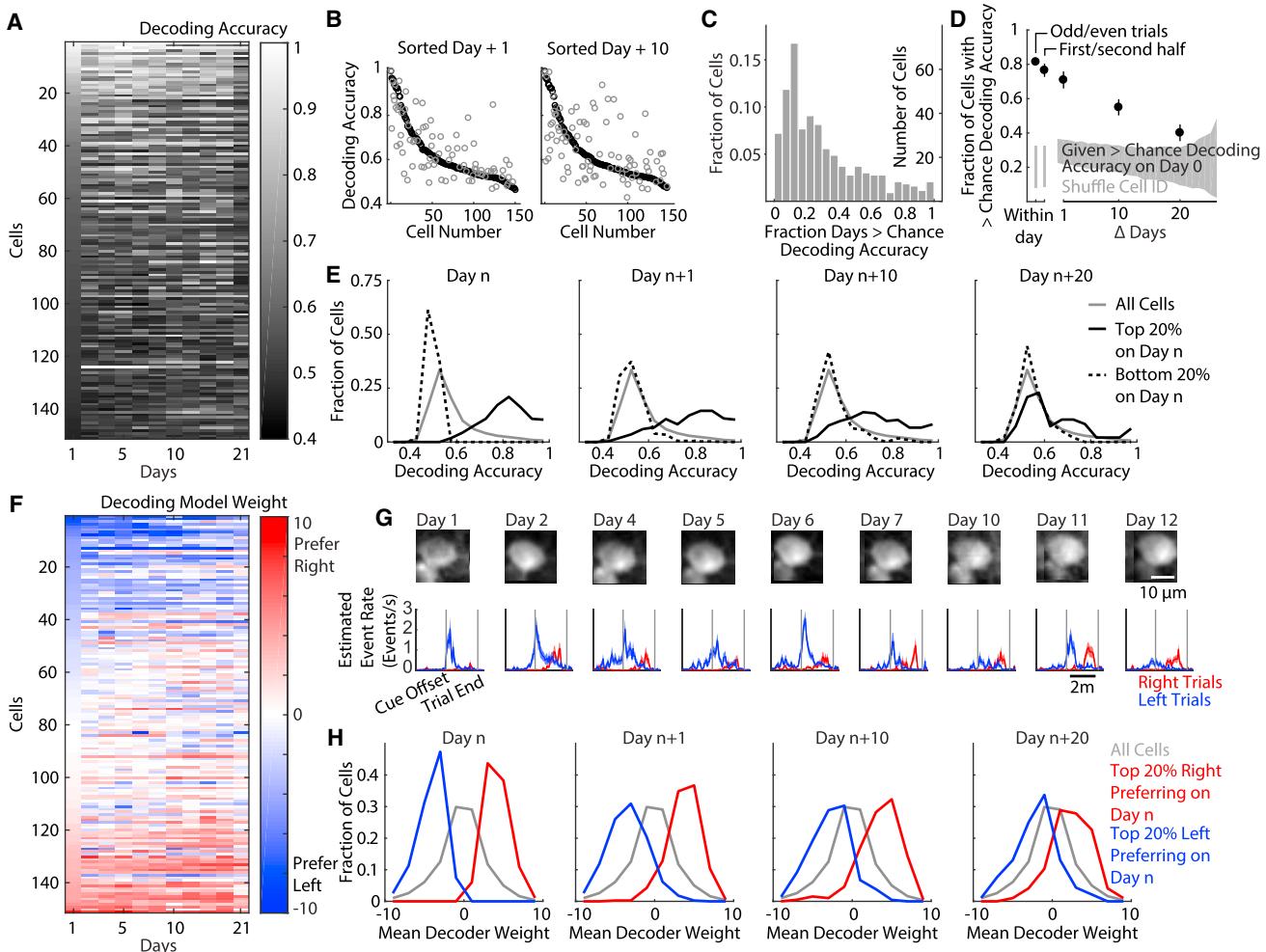


Figure 3. Reorganization of Information about Trial Type across Days

- (A) Decoding accuracy for trial type based on the activity of individual neurons, sorted by day 1.
- (B) For an example day, cells were sorted by their trial-type decoding accuracy. Decoding accuracy is shown on the day used for sorting (black) and on subsequent days (gray).
- (C) For the cells that were identified on ≥ 15 days, the fraction of days in which a cell's activity had above chance decoding accuracy, relative to the number of days in which the cell was identified. Chance: $p < 0.05$; permutation test.
- (D) Given high confidence for significant decoding accuracy on day n (99% confidence), the fraction of cells with greater than chance decoding accuracy (95% confidence) on day $n + x$. $p < 10^{-11}$ versus time; ANOVA. Error bars: mean \pm SEM. $n = 5$ mice. Gray area, 95% confidence intervals of chance levels based on shuffling cell IDs separately on each day.
- (E) Decoding accuracies for cells with the top and bottom 20% decoding on day n over time.
- (F) Trial type preference for the cells in (A) sorted by trial type preference on day 1, based on decoding model weights averaged over spatial bins in the maze.
- (G) Example cell with dynamic trial-type information. Top: mean fluorescence image is shown. Bottom: mean activity on correct white cue-left turn (blue) and black cue-right turn (red) trials is shown.
- (H) Decoder weights for the cells with the 20% largest and 20% smallest decoding weights on day n over time.

compared to decoding with shuffled trial labels). These neurons generally had consistent decoding accuracies within a session ($p = 0.28$ versus changes in odd/even trials; Figure 3D). However, neurons that had high decoding accuracies on a given day did not necessarily have significant decoding accuracies on subsequent days (Figures 3A and 3B). A large majority of neurons had significant decoding accuracy on fewer than half of the days in which they were identified (Figure 3C). Moreover, only $\sim 2\%$ of these neurons had significant decoding accuracy on all days in

which they were identified. The likelihood that a cell with greater than chance decoding accuracy on a given day had significant decoding accuracy on a subsequent day decreased with the interval between compared days and approached levels consistent with a random reorganization of cell identities (Figure 3D). In addition, we tracked the subpopulation of neurons that was the most highly informative about trial type on a given day. Over time, the distribution of decoding accuracy within this subpopulation approached and largely overlapped with the

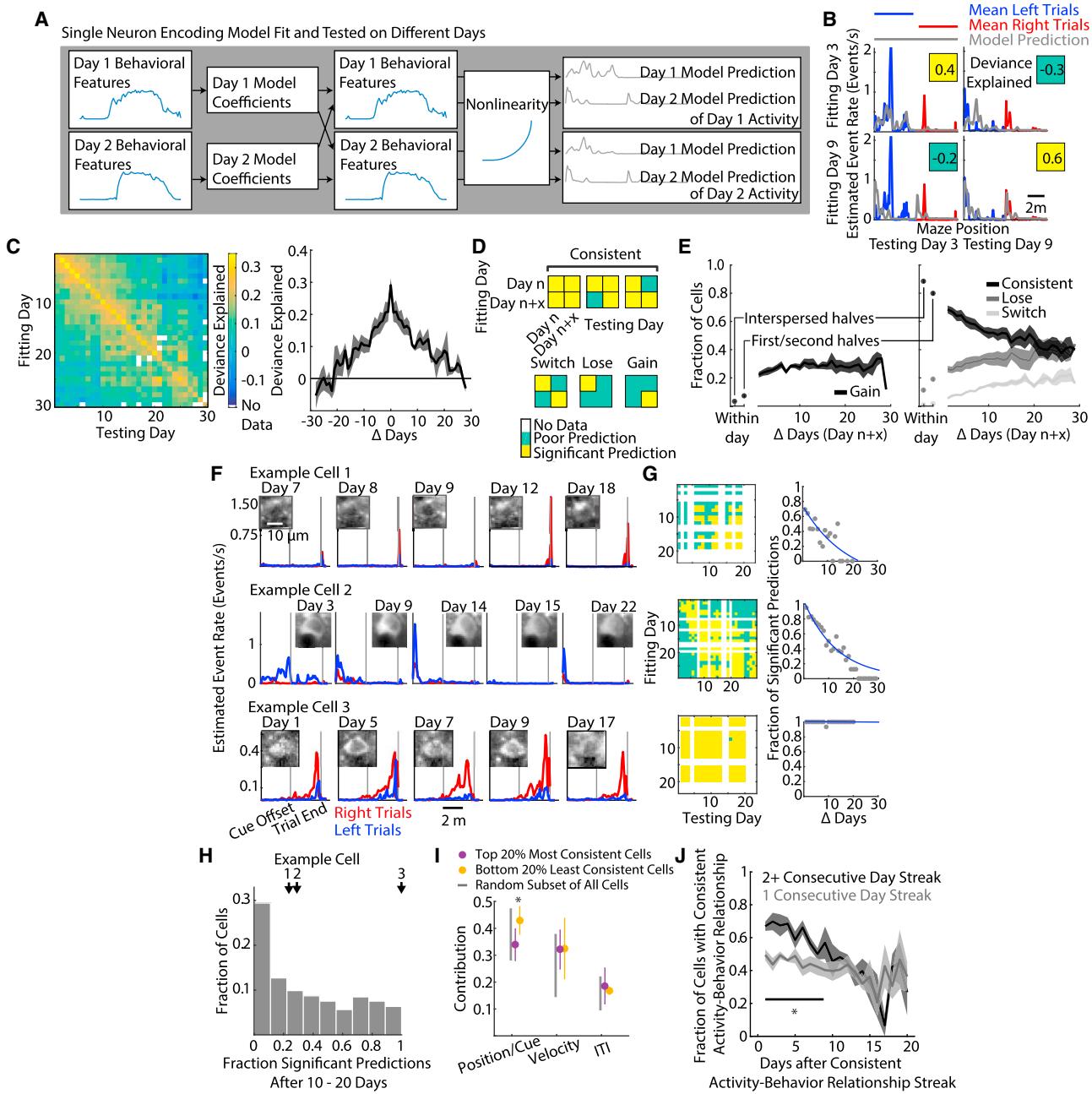


Figure 4. Using a GLM to Track Changes in Neuronal Activity-Behavior Relationships across Days

(A) For each neuron on each day, a GLM was fit to the activity of the neuron based on behavior features. Model coefficients for behavioral features were fit on one day and applied to behavioral data from another day to predict neuronal activity across days. See Figures S3, S4, S5, and S6.

(B) For an example cell, mean activity for white cue-left turn (blue) and black cue-right turn (red) trials on imaging days 3 and 9, with model predictions (gray) for fitting and testing on the same or opposite days.

(C) Left: deviance explained by models fit on one day and tested on the same or a different day, averaged across cells and then mice. See Figure S1C for n values. Right: average deviance explained as a function of time between the fitting and testing days is shown. Shading: mean \pm SEM.

(D) Schematic for model comparisons binarized as significant and poor predictions; threshold of 0.2 deviance explained, chosen based on a bootstrap analysis (STAR Methods).

(E) Left: for cells without a significant model prediction on day n , the fraction of cells with a significant model prediction after a given interval. Right: for cells with a significant model prediction on day n , the fraction of cells with consistent (black), lost (medium gray), or switched (light gray) activity-behavior relationships after a given interval is shown. Shaded area: mean \pm SEM. $n = 5$ mice, except for large intervals (Figure S2K).

(F) Mean activity for example cells with varying consistency in activity-behavior relationships.

(legend continued on next page)

distribution of the entire population, indicating that this subpopulation was not a special set of highly selective neurons across all time points (Figure 3E).

We also examined whether the neurons with selective activity for one trial type switched to having a preference for the other trial type (Figures 3F–3H). The most selective cells often lost their selectivity or gained additional selectivity, at other points in the maze, for the other trial type (Figures 3F and 3G). The trial type preferences of the neurons with the strongest selectivity on a given day approached that of the entire population over days (Figure 3H). Only a small fraction of neurons switched from having significantly higher activity on one trial type to having significantly higher activity on the opposite trial type ($4.7\% \pm 1.7\%$ of cells; lower bound for chance: $1.4\% \pm 0.7\%$ based on switches within a day using a held-out set; upper bound for chance: 45%–50% based on switches in selectivity when cell IDs were shuffled across days). Switches in trial type preferences were thus rare from one day to the next, and gains or losses of selectivity were more common. Together, these results indicate that trial-type-specific information was encoded by different populations of neurons over time.

Using a Generalized Linear Model to Compare Relationships between Neuronal Activity and Behavioral Features across Days

Although these findings provide evidence for major changes in neuronal activity patterns, thus far, we considered two aspects of the task (maze position and trial type) and did not include other task features that could potentially be represented in the neuronal activity, such as the mouse's running or movement of visual stimuli. We wanted to understand whether behavioral variability across days could explain the changes in neuronal activity or whether these changes were due primarily to single neurons having different relationships between their activity and behavior across time.

We developed a generalized linear model (GLM) in which we modeled the activity of an individual neuron based on variables that described the task and mouse's behavior: running patterns of the mouse; virtual maze position (visual scene); trial type; reward events; and whether the mouse was in the inter-trial interval period (Friedman et al., 2010; Park et al., 2014; Figures S3 and S4). We fit the relationship between a cell's activity and these behavior and task features to develop a model of that cell's activity-behavior relationship. We tested the quality of this model by predicting the cell's activity using behavioral and task features in a subset of trials not used for fitting. Across cells, models explained a large fraction of neuronal activity ($57.9\% \pm 2.6\%$ of cells had significant fits measured as the explained deviance in

the neuronal activity compared to a null model; Figure S5). The distribution of model prediction performance across the population was consistent over weeks (Figure S5D).

We used these models to compare the relationship between a cell's activity and behavioral features across days. Using the model of a cell's activity-behavior relationship fit on a single day, we predicted the cell's activity based on behavior features in other days (Figures 4A and 4B). If the model from one day predicted activity on a subsequent day, we concluded that a consistent activity-behavior relationship existed. In contrast, if a model from one day failed to predict the activity on subsequent days, then we concluded that a consistent activity-behavior relationship was absent. Behavioral features were variable across trials but maintained a similar distribution and range on each day, suggesting that models could be transferable across days (Figure S6). We limited effects due to fitting procedures, such as regularization, and due to correlated task variables by fitting and testing bi-directionally for each pair of days (STAR Methods).

Changing Activity-Behavior Relationships in Single Neurons over Days

Models of activity-behavior relationships developed on a given day, on average, predicted activity patterns well both within the same session and on neighboring days. However, these models did a poor job of predicting activity patterns as the time between the compared days increased (Figure 4C). Over long intervals, model predictions reached that of a null model, indicating that single-cell activity-behavior relationships were generally inconsistent over weeks (Figure 4C). We also quantified the similarity of models for a given cell across days using Kendall rank correlations of model parameters and found a comparable decay over time (Figure S5F). The changes in these activity-behavior relationships were made up of cells that lost well-modeled relationships, gained well-modeled relationships, and switched relationships across days. To quantify the prevalence of these events, we used a statistical threshold, based on shuffled data, to binarize model performance into predictions of activity above chance levels and predictions that we could not identify as significant (STAR Methods). We then compared pairs of models fit on separate days for a given cell. If the models developed on one day provided good predictions of the activity patterns on the other day, the cell was considered to have a consistent activity-behavior relationship (Figure 4D). Instead, if one model with a significant prediction of activity could be developed for one day's activity, but not for the other day's activity, the cell was considered to have lost or gained an activity-behavior relationship (Figure 4D). If models with significant predictions could be developed on both days but these models

(G) Left: for example cells from (F), fitting and testing comparisons as in (D). Right: fraction of models with significant predictions as a function of time between the fitting and test days is shown. Exponential fits are shown.

(H) Histogram of the fraction of significant model predictions after 10–20 days between fitting and training days for cells identified on ≥ 15 days. $n = 690$ cells.

(I) Contribution of different categories of behavior features to neuronal activity, estimated as the SD of the linear part of the model (STAR Methods). Comparisons of cells with the 20% most and 20% least consistent models is shown: position/cue, $p = 0.026$; treadmill velocity, $p = 0.98$; whether the mouse was in the inter-trial interval, $p = 0.79$; t test. Error bars: mean \pm SEM. $n = 5$ mice. Gray lines, 95% confidence intervals for randomly selected cells.

(J) For a given day, how many previous days the model of that day's activity provided a good prediction of previous days' activity. Models with significant predictions on ≥ 2 consecutive previous days were more likely to provide a good prediction in future days than models with significant predictions on only 1 previous day. * $p < 0.05$; t test.

See also Figures S3–S6.

provided poor predictions of activity on the other day, the cell was considered to have switched activity-behavior relationships (Figure 4D). Although these categories were not perfect descriptors due to statistical thresholds, they provided an overview of the changes that occurred. The likelihood that a cell lacking an activity-behavior relationship gained such a relationship remained constant over weeks (Figure 4E). Within a session, between the first and second half of trials, changes in activity-behavior relationships were small but statistically significant ($p < 10^{-3}$ versus changes in interspersed trials; Figure 4E). As the interval between the compared days increased, the likelihood that a cell had a consistent relationship decreased and the likelihood that a cell lost or switched a relationship increased (Figure 4E). After ~20 days, a cell with a well-described activity-behavior relationship was more likely to have lost or switched this relationship than to have maintained it (Figure 4E).

The rate of changes in activity-behavior relationships varied greatly across neurons (Figure 4F). For each neuron, we calculated the likelihood that a model developed on one day provided a significant prediction of another day's activity (Figure 4G, left) and fit an exponential to this likelihood over time to define a metric of consistency for each cell (Figure 4G, right). Some neurons had slow decays and thus relatively consistent activity-behavior relationships, whereas others had fast decays indicative of rapid changes. Over a 20-day interval, the large majority of neurons had a low likelihood of consistent models (Figure 4H). Only ~7% of neurons had consistent activity-behavior relationships over the entire interval (defined as >95% significant predictions after 10–20 days).

To understand whether the neurons with the most and least consistent activity-behavior relationships represented different types of behavioral information, we examined the contribution of various parameters to each cell's activity (the extent to which the behavioral parameter of interest modulated a given cell's model prediction). Cells with the most and least consistent relationships had a distribution of contributions for trial type, maze position, running pattern, and inter-trial interval times that overlapped with the distribution in the full population (Figure 4I). However, neurons with the least consistent relationships more often had greater contributions from trial type and maze position than neurons with the most consistent relationships, suggesting that activity related to learned task features may have less consistency across time (Figure 4I).

The changes in activity-behavior relationships were not independent across days. Rather, consistent relationships in the recent past predicted consistent relationships in the near future. Neurons with a consistent activity-behavior relationship for 2 or more consecutive days were more likely to maintain that relationship than neurons with a consistent relationship for only 1 of the immediately preceding days (Figure 4J). Neurons therefore potentially operated with modes of activity that tended to persist for neighboring days (Figure S5G).

The GLM analyses suggest that the changes in activity were likely due to unstable activity-behavior relationships rather than changes in behavioral patterns across days. We supported this finding by comparing the similarity of population activity patterns on trials with the most or least similar behavioral patterns across all days. Population activity was more similar on trials with more

similar behavioral features, measured as correlations between population activity vectors (Figure S6D). However, the difference in activity between the most and least similar behavioral trials was small compared to the population activity changes across time (Figure S6D). In addition, mice did not appear to forget and re-learn the task each day. Performance was near perfect on the first few trials of each day (Figures 1B and S6E).

Consistent Statistical Features of Population Activity on Each Day

Despite changes in the activity of individual neurons, we noticed consistent patterns in the population activity on each day. Neuronal activity that tiled the trial was present on each day and was made up of different neurons across days (Figure 2A). The distribution of population activity across the trial was not uniform, but this distribution was similar across all days in a given population of neurons (Figures 5A and 5B). In addition, on each day, a decoder for trial type based on population activity achieved similar levels of performance and had similar distributions of performance across time points in the trial (Figures 5C and 5D). Interestingly, in each population of neurons from different mice, differences between mice were maintained across days. For example, the population of neurons in one mouse (red) had higher decoding accuracy of trial type than in another mouse (green) across all days (Figure 5D). Other properties of population activity had similar distributions on each day, including for neuron-neuron activity correlations, trial-trial population activity correlations, estimated population firing rates, and decoding accuracy of trial type for individual neurons (Figures 5E–5L). Therefore, the population appeared to have a “set point” of similar activity each day, using different neurons and neurons in different ways.

Decoding of Information from Dynamic Neuronal Representations

The changes in neuronal activity-behavior relationships raise questions about how information could be read out from a dynamic neuronal population. The cells with the most consistent activity-behavior relationships could preferentially carry information for the readout. Alternatively, cells with less consistent activity-behavior relationships could contribute to decoding of information over time. We tested various decoding strategies for reading out the trial type on the basis of population activity.

We first trained and tested a linear decoder on each day separately using all neurons. Trial type information could be decoded throughout the trial, with higher decoding accuracies at the end of the trial, when the mouse executed a turn at the T-intersection (Figure 6A). We compared the decoding performance using the cells with the most or least consistent activity-behavior relationships (from Figure 4G; STAR Methods). The cells with the least consistent relationships had better decoding accuracy throughout the majority of the trial (Figure 6B).

To analyze the stability of information in population-activity patterns, we tested decoding performance across days. We trained a decoder on a given day and tested it on subsequent days (Figure 6C). For a random subset of cells, decoding performance decreased as the interval between compared days increased (Figure 6C). Decoding performance scaled with the

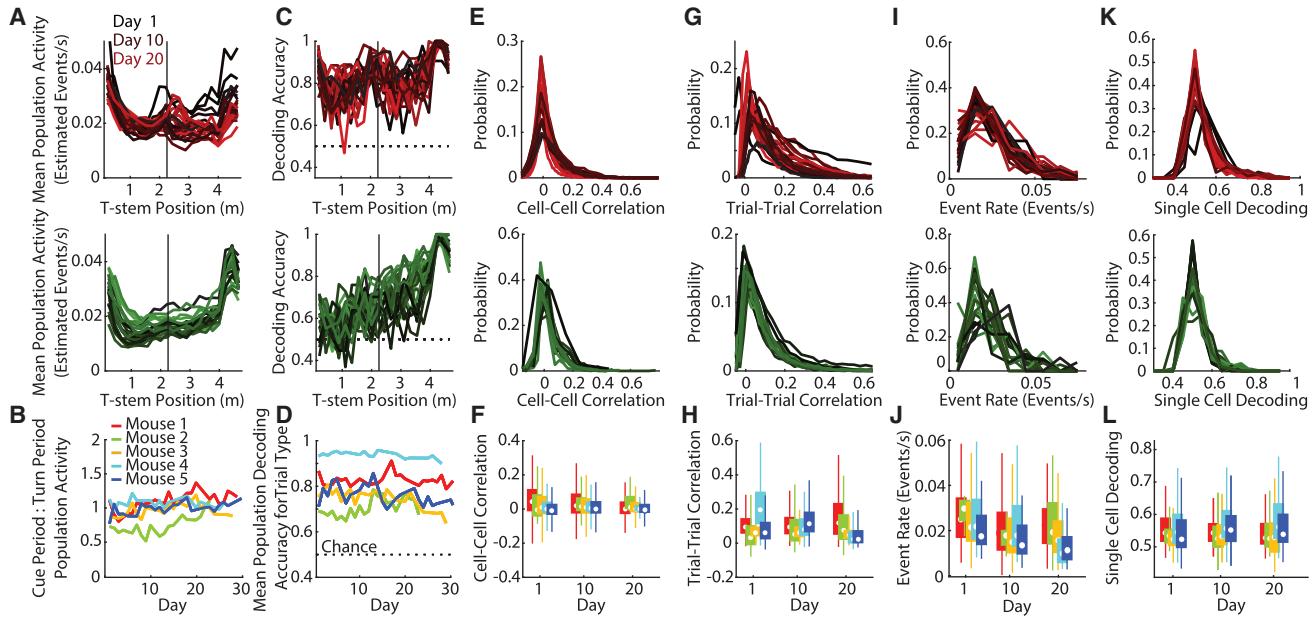


Figure 5. Stable Statistical Features of Population Activity

- (A) For two example mice, mean population activity versus maze position.
- (B) Ratio of activity in the cue period to the delay period.
- (C) For two example mice, population-decoding accuracy of trial type versus maze position. Separate decoders were trained at each spatial bin and on each day.
- (D) Population-decoding accuracy of trial type.
- (E) For two example mice, distributions of cell-cell correlations of deconvolved calcium signals smoothed with a 2-s sliding window.
- (F) Summary of cell-cell correlation distributions. Boxes, 25th and 75th percentiles; white dots, mean; whiskers, 99% range.
- (G–L) Same as in (E)–(F), except for correlations of population activity (cells \times maze position) in trials of the same type (G and H), population activity event rates (I and J), and classification accuracy of trial type based on single-cell activity (K and L).

size of the neuronal population considered (Figure 6D). In the cells with the most consistent activity-behavior relationships, decoding performance was low for the majority of the trial, until the final segment, in which decoding performance was high and consistent over days (Figure 6C). As expected, the performance of a decoder trained on one day and tested on other days decreased with time for the cells with the least consistent activity-behavior relationships (Figure 6C). Interestingly, however, over intervals within 1 week, these cells performed better in the majority of the trial than the cells with the most consistent relationships (Figure 6C, left and middle). Therefore, information in the population was not stable over time, but some information remained for days and weeks, even in populations of neurons with the least consistent activity patterns. For the binary classification of trial type over the intervals examined, a stable readout could suffice for above chance performance, but, to achieve higher performance, a readout that changes dynamically with the encoding network would likely be necessary.

If the relevant information for the task is read out when the mouse executes a turn, then it might be beneficial to weight strongly the cells with the most consistent activity-behavior relationships. In contrast, if information in the T-stem is more relevant for behavior, then weighting the cells with the less consistent relationships might be beneficial. Based on this reasoning, we returned to our optogenetic inactivation experiments and inhibited PPC activity either during the first half or the second half

of the trial. Inhibiting PPC activity in the first half of the trial greatly impaired the mouse's performance, but silencing PPC activity in the second half of the trial had no significant effect (Figure S7). It is thus possible that weighting the information in cells with the least consistent activity-behavior relationships would be useful for PPC's role in this task.

Incorporation of Representations for Newly Learned Associations into Existing Population Activity Patterns

Computational models suggest that an advantage of dynamic neuronal representations could be the flexibility of incorporating new information into the population (Ajemian et al., 2013; Rokni et al., 2007). We tested how existing representations were affected by the learning of a new association. We trained the same mice that had stably performed the task described above to learn a new association. We introduced a third possible cue (X pattern) that required a specific turn at the intersection for a reward (the turn direction was randomly selected for each mouse; Figure 7A). After a mouse learned the novel third trial type, we introduced a fourth cue (triangle pattern) with the opposite turn association (Figure 7A). Mice learned the novel cue-response associations while maintaining high performance for the familiar two-cue-response associations (Figure 7B).

We first asked whether the neuronal activity patterns were different between trials in which the mouse saw the novel cues or the familiar cues. The novel trial types could be distinguished

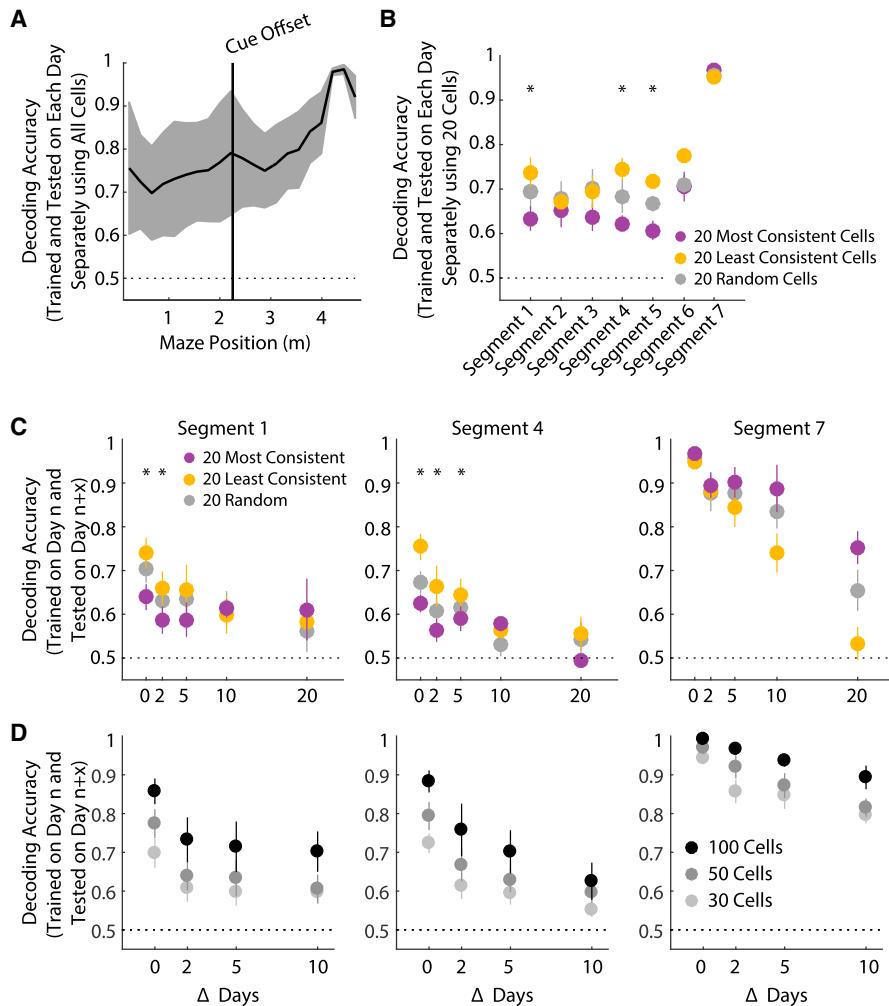


Figure 6. Decoding Task Information across Days

(A) Population-decoding accuracy of trial type on correct trials as a function of position in the maze. Separate decoders were trained at each spatial bin and on each day. Shading: mean \pm SEM. $n = 5$ mice. Dashed line indicates chance.

(B) Decoding accuracy of trial type on correct trials using 20 cells with the least (yellow) or most (purple) consistent activity-behavior relationships or 20 randomly selected cells (gray). Error bars: mean \pm SEM. $n = 5$ mice. * $p < 0.05$; t test; between least and most consistent groups. See Figure S7 for inactivation experiments.

(C) Decoders trained on a given session and tested on a later session using 20 cells. Error bars: mean \pm SEM. $n = 5, 4, 4, 4, 2$ mice for Δ days 0, 2, 5, 10, and 20, respectively. * $p < 0.05$; t test; between least and most consistent groups.

(D) Decoders trained on a given session and tested on a later session using 30, 50, or 100 randomly selected cells. Decoding performance versus number of neurons used for decoding: $p < 10^{-3}$, ANOVA.

See also Figure S7.

with a linear population decoder from the familiar trial types during the cue and delay period, even for trials with the same turn direction (Figure 7C). However, during the turn period, the novel trial type was only distinguishable from trials with opposite behavioral choices (Figure 7C). The relationship between population activity between trial types could be visualized on a single day in a dimensionality-reduced space (Figure 7D). These results indicate that a distinct representation of the new trial types was present, except during the turn period in which activity followed the turn direction.

We tested whether the new learned associations altered the rate at which neuronal activity patterns changed during perfor-

mance of previously learned trial types. We might expect learning to increase the rate of change as new information is incorporated. In agreement with this idea, we found that the rate of change within a session increased with the introduction of a novel trial type (Figure 7E). In contrast, we found that the rate of change across days was comparable between the days with two familiar trial types and during learning of the novel trial types (Figure 7F). New trial types therefore appeared to cause a small increase in the rate of change over short timescales but did not significantly alter the overall rate of drift over days.

The similar rate of change over days before and after the introduction of new trial types could have occurred if the cells

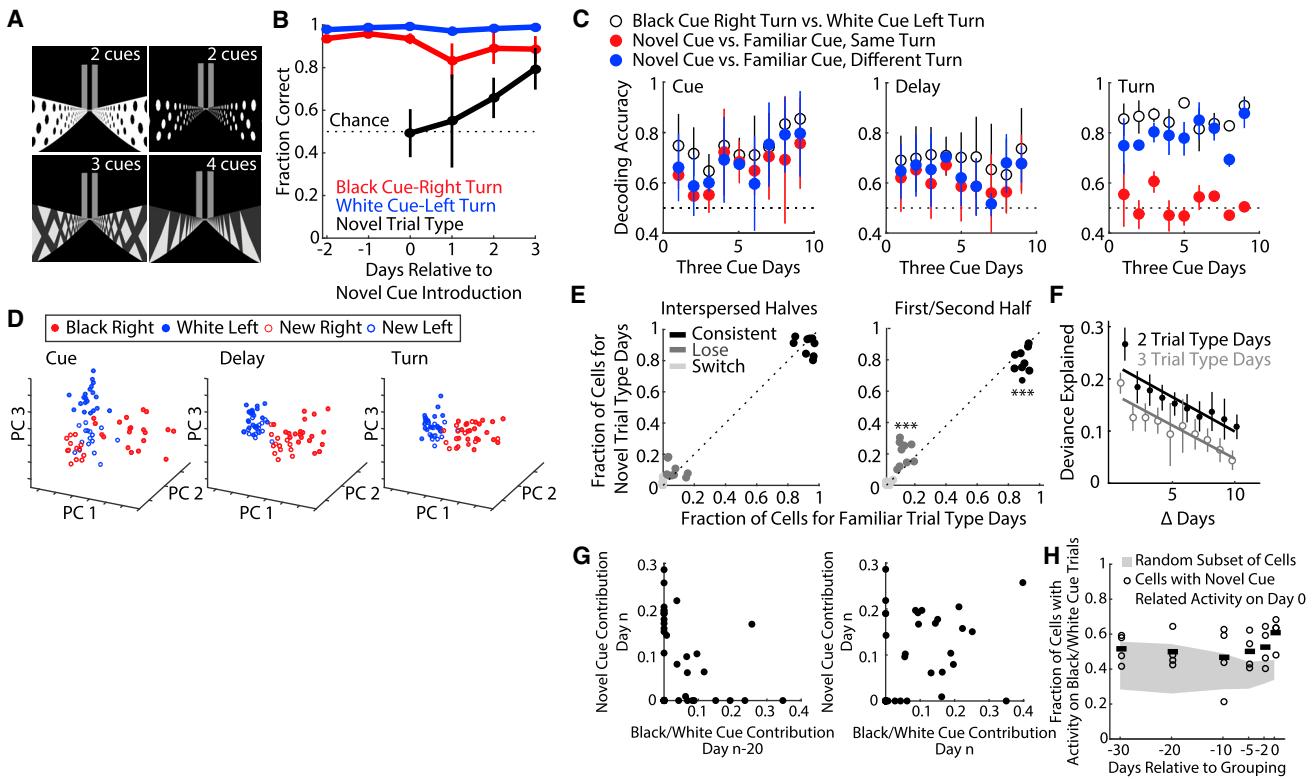


Figure 7. Neuronal Activity during Learning of a Novel Trial Type

- (A) Mouse perspective of virtual maze trial types.
- (B) Behavioral performance on 2 days preceding and 4 days following introduction of a novel trial type. Error bars: mean ± SEM. n = 4 (2 mice with 2 novel trial types each).
- (C) Population decoding performance of different trial types during the cue, delay, and turn periods. Error bars: mean ± SEM; n = 2 mice.
- (D) Example population activity during cue, delay, and turn periods in the same dimensionality reduced space (one mouse, four trial types). Dots, single trials.
- (E) Fraction of cells with consistent (black), lost (medium gray), and switched (light gray) activity-behavior relationships within one session based on our GLM analysis on days with only familiar trials or with novel trial types. Dots, single sessions. First/second half, consistent: p < 10⁻³; lose: p < 10⁻³; t test.
- (F) For models fit to activity-behavior relationships on one day, deviance explained of predictions of activity on other days. p = 0.91; ANOVA; comparing linear regression slopes for two-trial-type and novel-trial-type days.
- (G) Left: for individual cells, novel cue-related activity on day n (measured by >0 contribution for novel cue onset/offset in the GLM) versus familiar cue-related activity on day n-20. Right: for the same cells, novel cue-related activity on day n versus familiar cue-related activity on day n is shown.
- (H) For cells with novel cue-related activity on day n, the fraction of cells with activity related to either white or black cue onset/offset on the same day (day n) and on previous days. n = 4 circles (2 mice with 2 novel trial types each); bars, means; shading, 95% confidence interval from a random subset of neurons.

with activity related to the novel trials were different from the cells with activity related to familiar trial types. Surprisingly, cells with activity related to novel cues were more likely to come from the group of cells that recently (within the past 10 days) had activity related to the familiar cues than from a random sample of neurons (Figures 7G and 7H). The evolving pool of cells involved in representing task features was thus more likely to incorporate new task-relevant information than the group of cells presently without task-relevant information. This finding suggests that new information can be incorporated into the pool of cells with task-relevant activity as this pool continuously shifts over time, without disrupting baseline functionality. We speculate that the network's ongoing changes provide a framework for the addition of new associations using "multitasking" neurons that allow for flexibility during learning.

DISCUSSION

Our results differ from studies in sensory and motor regions that reported no changes or subtle shifts in activity (Margolis et al., 2012; Peron et al., 2015; Peters et al., 2014). We report changes similar to those in hippocampus (Ziv et al., 2013). The rate of change in neuronal representations may vary depending on the function of the population of neurons. In areas closely connected to sensory coding or the generation of motor actions, there may be a greater need for stability. In contrast, association regions require flexibility for learned behaviors as one of their key properties, in which case, malleable activity patterns would be advantageous. Thus, the changes we observed may not be general to cortex but rather a specific property of association areas. Consistent with these ideas, dendritic spines turn over at a higher rate in hippocampus than in primary somatosensory cortex

(Attardo et al., 2015). It will be of interest to directly compare the rates of activity changes in populations of neurons in different brain regions to test whether abstract learned representations drift at faster rates than activity patterns that represent sensory stimuli or generate motor actions.

Dynamic representations were likely used to perform the task. Inactivation of cue period activity centered at PPC resulted in a decrease in behavioral performance. It seems unlikely that PPC was essential for visual perception alone given that our previous work found little evidence of strong responses in PPC during passive viewing and showed PPC was not required for a simple visually guided task (Harvey et al., 2012). Also, because PPC activity was not required during the delay period, it seems unlikely that PPC had an essential short-term memory of the cue or upcoming action. Collectively, the results here and from previous work support the hypothesis that PPC functions in the visual-to-motor transformation (Goard et al., 2016; Harvey et al., 2012; Licata et al., 2017; Raposo et al., 2014). This hypothesis is consistent with PPC's connectivity in which it receives multisensory input, has recurrent connections with frontal regions, and has outputs to motor-related structures (Harvey et al., 2012).

Our work reveals a strategy for PPC neurons to achieve stability of learned associations while allowing flexibility to incorporate new information. We illustrate this idea by focusing on the binary classification of trial type. Over weeks, individual neurons did not maintain different activity between trial types. Many neurons gained or lost selectivity, and a small fraction of neurons switched trial type preferences. Despite these changes, information about trial type could be read out at above chance levels from the population using a stable decoder. The decoder worked because the majority of neurons lost or gained selectivity but did not switch from one trial type to the other. Many changes therefore occurred in a mostly null dimension relative to the dimension important for decoding trial type (Ajemian et al., 2013; Rokni et al., 2007). These changes need not be coordinated in an “intelligent” or “structured” way to occur specifically in a null dimension. Rather, this effect is characteristic of a high-dimensional activity space, in which most dimensions are orthogonal to each other. However, more information would be accessible to the readout if it functioned as a dynamically adaptive decoder. Over long timescales, an ideal decoder could slowly drift in response to the dynamically encoded information but could change at a slow pace or with a lag given that drastic changes in the encoding network did not occur from one day to the next. Learned associations that are not practiced could be lost if the readout is not updated for an extended interval, whereas learned associations that are often practiced could maintain a tight link between drifting activity patterns and the relevant readout. Slow drifts in activity have been shown to provide computational advantages, such as avoiding local minima and providing exploratory information for reinforcement learning (Kappel et al., 2015).

Our work provides evidence that individual PPC neurons do not have specified roles in network activity. Not only did we observe that neurons lost or gained activity-behavior relationships over time but we also found neurons that switched

their activity-behavior relationships. Neuronal activity was best described by combinations of behaviorally relevant features similar to recent work showing multiplexed information in cortical neurons (Cromer et al., 2010; Rigotti et al., 2013). Changes in activity suggest that a neuron might not be confined to a specific class of activity pattern (Raposo et al., 2014). Although our decoding experiments revealed that information could be read out even from cells with the least consistent activity-behavior relationships, it will be important to test experimentally which neurons are most important for behavior. Furthermore, we examined changes in activity relative to the behavioral and task features we measured, but it remains possible that PPC activity could have long-term stability with respect to task, behavior, or internal parameters that we did not monitor or to which we did not have access. For these reasons, it will be of interest to repeat these experiments with different tasks and recording methods.

Together, our work and others suggest that the role of individual neurons could be less important than the overall population activity pattern (Yuste, 2015). Consistent with this idea, PPC population activity had similar statistics on each day, using different neurons or the same neurons in different ways. This result suggests that the population activity reached a set point of activity that was necessary for the PPC's role in the task. This finding is conceptually similar to the homeostatic properties of the stomatogastric ganglion (O'Leary et al., 2014; Prinz et al., 2004). In that system, the same firing patterns of neurons can be achieved through different combinations of ion channels and ion channel expression levels. We speculate that neuronal populations similarly maintain homeostasis. In this case, features of population activity remain constant, but the role for individual neurons is flexible.

STAR★METHODS

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

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Mice
- METHOD DETAILS
 - Virtual reality system
 - Behavioral training
 - Optogenetic inactivation experiments
 - Surgical procedures
 - Field sign maps
 - Two-photon imaging
 - Pre-processing of imaging data
 - Data inclusion criteria
- QUANTIFICATION AND STATISTICAL ANALYSES
 - General analysis procedures
 - Identification of significant peaks of activity
 - GLM encoding model
 - Decoding
 - Dimensionality reduction of neuronal and behavioral data
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2017.07.021>.

AUTHOR CONTRIBUTIONS

L.N.D. and C.D.H. conceived of the project and wrote the manuscript. L.N.D. performed all experiments and analyses, except N.L.P. performed the optogenetics experiments and analyzed associated data. M.M. designed the retinotopy experiments, and S.N.C. and M.M. developed the cell selection algorithm for single sessions.

ACKNOWLEDGMENTS

We thank Tim O’Leary, Ofer Mazor, and the Harvey lab for comments on the manuscript; Paola Patella for contributing to the setup of the optogenetics experiments; and the Research Instrumentation Core at Harvard Medical School. This work was supported by a Burroughs-Wellcome Fund Career Award at the Scientific Interface, the Searle Scholars Program, the New York Stem Cell Foundation, the Alfred P. Sloan Research Foundation, a NARSAD Brain and Behavior Research Young Investigator Award, NIH grants from the NIMH BRAINS program (R01MH107620) and NINDS (R01NS089521), an Armenian-Harvard Foundation Junior Faculty Grant, an Edward R. and Anne G. Leffler Center Predoctoral Fellowship (L.N.D.) and Junior Faculty Award, the Albert J. Ryan Fellowship (L.N.D.), the Stuart H.Q. and Victoria Quan Fellowship (L.N.D. and N.L.P.), a Boehringer Ingelheim Fonds Fellowship (M.M.), a Herchel Smith Graduate Fellowship (M.M.), and a NSF Graduate Research Fellowship (S.N.C.). C.D.H. is a New York Stem Cell Foundation Robertson Neuroscience Investigator.

Received: January 30, 2017

Revised: May 28, 2017

Accepted: July 13, 2017

Published: August 17, 2017

REFERENCES

- Ajemanian, R., D’Ausilio, A., Moorman, H., and Bizzi, E. (2013). A theory for how sensorimotor skills are learned and retained in noisy and nonstationary neural circuits. *Proc. Natl. Acad. Sci. USA* **110**, E5078–E5087.
- Andermann, M.L., Kerlin, A.M., and Reid, R.C. (2010). Chronic cellular imaging of mouse visual cortex during operant behavior and passive viewing. *Front. Cell. Neurosci.* **4**, 3.
- Aronov, D., and Tank, D.W. (2014). Engagement of neural circuits underlying 2D spatial navigation in a rodent virtual reality system. *Neuron* **84**, 442–456.
- Attardo, A., Fitzgerald, J.E., and Schnitzer, M.J. (2015). Impermanence of dendritic spines in live adult CA1 hippocampus. *Nature* **523**, 592–596.
- Chen, T.W., Wardill, T.J., Sun, Y., Pulver, S.R., Renninger, S.L., Baohan, A., Schreiter, E.R., Kerr, R.A., Orger, M.B., Jayaraman, V., et al. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300.
- Chestek, C.A., Batista, A.P., Santhanam, G., Yu, B.M., Afshar, A., Cunningham, J.P., Gilja, V., Ryu, S.I., Churchland, M.M., and Shenoy, K.V. (2007). Single-neuron stability during repeated reaching in macaque premotor cortex. *J. Neurosci.* **27**, 10742–10750.
- Cromer, J.A., Roy, J.E., and Miller, E.K. (2010). Representation of multiple, independent categories in the primate prefrontal cortex. *Neuron* **66**, 796–807.
- Freeman, D.J., and Assad, J.A. (2016). Neuronal mechanisms of visual categorization: an abstract view on decision making. *Annu. Rev. Neurosci.* **39**, 129–147.
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22.
- Ganguly, K., and Carmena, J.M. (2009). Emergence of a stable cortical map for neuroprosthetic control. *PLoS Biol.* **7**, e1000153.
- Goad, M.J., Pho, G.N., Woodson, J., and Sur, M. (2016). Distinct roles of visual, parietal, and frontal motor cortices in memory-guided sensorimotor decisions. *eLife* **5**, 4841–4843.
- Greenberg, D.S., and Kerr, J.N.D. (2009). Automated correction of fast motion artifacts for two-photon imaging of awake animals. *J. Neurosci. Methods* **176**, 1–15.
- Harvey, C.D., Collman, F., Dombeck, D.A., and Tank, D.W. (2009). Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature* **461**, 941–946.
- Harvey, C.D., Coen, P., and Tank, D.W. (2012). Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature* **484**, 62–68.
- Huber, D., Gutnisky, D.A., Peron, S., O’Connor, D.H., Wiegert, J.S., Tian, L., Oertner, T.G., Looger, L.L., and Svoboda, K. (2012). Multiple dynamic representations in the motor cortex during sensorimotor learning. *Nature* **484**, 473–478.
- Kappel, D., Habenschuss, S., Legenstein, R., and Maass, W. (2015). Network plasticity as Bayesian inference. *PLoS Comput. Biol.* **11**, e1004485.
- Kentros, C.G., Agnihotri, N.T., Streater, S., Hawkins, R.D., and Kandel, E.R. (2004). Increased attention to spatial context increases both place field stability and spatial memory. *Neuron* **42**, 283–295.
- Liberti, W.A., 3rd, Markowitz, J.E., Perkins, L.N., Liberti, D.C., Leman, D.P., Guitchounts, G., Velho, T., Kotton, D.N., Lois, C., and Gardner, T.J. (2016). Unstable neurons underlie a stable learned behavior. *Nat. Neurosci.* **19**, 1665–1671.
- Licata, A.M., Kaufman, M.T., Raposo, D., Ryan, M.B., Sheppard, J.P., and Churchland, A.K. (2017). Posterior parietal cortex guides visual decisions in rats. *J. Neurosci.* **37**, 4954–4966.
- Mank, M., Santos, A.F., Direnberger, S., Mrsic-Flogel, T.D., Hofer, S.B., Stein, V., Hendel, T., Reiff, D.F., Levelt, C., Borst, A., et al. (2008). A genetically encoded calcium indicator for chronic *in vivo* two-photon imaging. *Nat. Methods* **5**, 805–811.
- Margolis, D.J., Lütcke, H., Schulz, K., Haiss, F., Weber, B., Kügler, S., Hasan, M.T., and Helmchen, F. (2012). Reorganization of cortical population activity imaged throughout long-term sensory deprivation. *Nat. Neurosci.* **15**, 1539–1546.
- Marshel, J.H., Garrett, M.E., Nauhaus, I., and Callaway, E.M. (2011). Functional specialization of seven mouse visual cortical areas. *Neuron* **72**, 1040–1054.
- McNaughton, B.L., Mizumori, S.J.Y., Barnes, C.A., Leonard, B.J., Marquis, M., and Green, E.J. (1994). Cortical representation of motion during unrestrained spatial navigation in the rat. *Cereb. Cortex* **4**, 27–39.
- Messinger, A., Squire, L.R., Zola, S.M., and Albright, T.D. (2001). Neuronal representations of stimulus associations develop in the temporal lobe during learning. *Proc. Natl. Acad. Sci. USA* **98**, 12239–12244.
- Morcos, A.S., and Harvey, C.D. (2016). History-dependent variability in population dynamics during evidence accumulation in cortex. *Nat. Neurosci.* **19**, 1672–1681.
- Mouse, A., and Coordinate, C. (2016). Allen mouse common coordinate framework. http://help.brain-map.org/download/attachments/2818171/mouse_common_coordinate_framework.pdf?version=2&modificationdate=1477414987120
- Nitz, D.A. (2006). Tracking route progression in the posterior parietal cortex. *Neuron* **49**, 747–756.
- O’Leary, T., Williams, A.H., Franci, A., and Marder, E. (2014). Cell types, network homeostasis, and pathological compensation from a biologically plausible ion channel expression model. *Neuron* **82**, 809–821.
- Otchy, T.M., Wolff, S.B.E., Rhee, J.Y., Pehlevan, C., Kawai, R., Kempf, A., Gobes, S.M.H., and Ölveczky, B.P. (2015). Acute off-target effects of neural circuit manipulations. *Nature* **528**, 358–363.
- Padoa-Schioppa, C., Li, C.-S.R., and Bizzi, E. (2004). Neuronal activity in the supplementary motor area of monkeys adapting to a new dynamic environment. *J. Neurophysiol.* **91**, 449–473.

- Park, I.M., Meister, M.L.R., Huk, A.C., and Pillow, J.W. (2014). Encoding and decoding in parietal cortex during sensorimotor decision-making. *Nat. Neurosci.* **17**, 1395–1403.
- Peron, S.P., Freeman, J., Iyer, V., Guo, C., and Svoboda, K. (2015). A cellular resolution map of barrel cortex activity during tactile behavior. *Neuron* **86**, 783–799.
- Peters, A.J., Chen, S.X., and Komiyama, T. (2014). Emergence of reproducible spatiotemporal activity during motor learning. *Nature* **510**, 263–267.
- Pnevmatikakis, E.A., Soudry, D., Gao, Y., Machado, T.A., Merel, J., Pfau, D., Reardon, T., Mu, Y., Lacefield, C., Yang, W., et al. (2016). Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron* **89**, 285–299.
- Poort, J., Khan, A.G., Pachitariu, M., Nemri, A., Orsolic, I., Krupic, J., Bauza, M., Sahani, M., Keller, G.B., Mrsic-Flogel, T.D., and Hofer, S.B. (2015). Learning enhances sensory and multiple non-sensory representations in primary visual cortex. *Neuron* **86**, 1478–1490.
- Prinz, A.A., Bucher, D., and Marder, E. (2004). Similar network activity from disparate circuit parameters. *Nat. Neurosci.* **7**, 1345–1352.
- Raposo, D., Kaufman, M.T., and Churchland, A.K. (2014). A category-free neural population supports evolving demands during decision-making. *Nat. Neurosci.* **17**, 1784–1792.
- Ratzlaff, E.H., and Grinvald, A. (1991). A tandem-lens epifluorescence macroscope: hundred-fold brightness advantage for wide-field imaging. *J. Neurosci. Methods* **36**, 127–137.
- Rigotti, M., Barak, O., Warden, M.R., Wang, X.-J., Daw, N.D., Miller, E.K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**, 585–590.
- Rokni, U., Richardson, A.G., Bizzi, E., and Seung, H.S. (2007). Motor learning with unstable neural representations. *Neuron* **54**, 653–666.
- Rose, T., Jaepel, J., Hübener, M., and Bonhoeffer, T. (2016). Cell-specific restoration of stimulus preference after monocular deprivation in the visual cortex. *Science* **352**, 1319–1322.
- Sakai, K., and Miyashita, Y. (1991). Neural organization for the long-term memory of paired associates. *Nature* **354**, 152–155.
- Stettler, D.D., Yamahachi, H., Li, W., Denk, W., and Gilbert, C.D. (2006). Axons and synaptic boutons are highly dynamic in adult visual cortex. *Neuron* **49**, 877–887.
- Stevenson, I.H., Cherian, A., London, B.M., Sachs, N.A., Lindberg, E., Reimer, J., Slutsky, M.W., Hatsopoulos, N.G., Miller, L.E., and Kording, K.P. (2011). Statistical assessment of the stability of neural movement representations. *J. Neurophysiol.* **106**, 764–774.
- Tolias, A.S., Ecker, A.S., Siapas, A.G., Hoenselaar, A., Keliris, G.A., and Logothetis, N.K. (2007). Recording chronically from the same neurons in awake, behaving primates. *J. Neurophysiol.* **98**, 3780–3790.
- Tonegawa, S., Liu, X., Ramirez, S., and Redondo, R. (2015). Memory engram cells have come of age. *Neuron* **87**, 918–931.
- Trachtenberg, J.T., Chen, B.E., Knott, G.W., Feng, G., Sanes, J.R., Welker, E., and Svoboda, K. (2002). Long-term *in vivo* imaging of experience-dependent synaptic plasticity in adult cortex. *Nature* **420**, 788–794.
- Wang, Q., and Burkhalter, A. (2007). Area map of mouse visual cortex. *J. Comp. Neurol.* **502**, 339–357.
- Whitlock, J.R., Pfuhl, G., Dagslott, N., Moser, M.-B., and Moser, E.I. (2012). Functional split between parietal and entorhinal cortices in the rat. *Neuron* **73**, 789–802.
- Yuste, R. (2015). From the neuron doctrine to neural networks. *Nat. Rev. Neurosci.* **16**, 487–497.
- Ziv, Y., Burns, L.D., Cocker, E.D., Hamel, E.O., Ghosh, K.K., Kitch, L.J., El Gamal, A., and Schnitzer, M.J. (2013). Long-term dynamics of CA1 hippocampal place codes. *Nat. Neurosci.* **16**, 264–266.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Organisms/Strains		
B6.129P2-Pvalbtm1(cre)Arbr/J	The Jackson Laboratory	Stock No: 017320 Black 6
C57BL/6J mice	The Jackson Laboratory	Stock No: 000664 Black 6
C57BL/6J-Tg(Thy1-GCaMP6s)GP4.12Dkim/J mice	The Jackson Laboratory	Stock No: 025776 Black 6
Bacterial and Virus Strains		
AAV1.EF1.d flox.hChR2(H134R)-mCherry.WPRE.hGH	UPenn Vector Core	Catalog No: AV-1-20297P
AAV1.Syn.Flex.GCaMP6m.WPRE.SV40	UPenn Vector Core	Catalog No: PV2823
AAV1.CAG.tdTomato.WPRE.SV40	UPenn Vector Core	Catalog No: AV-1-PV2126
Software and Algorithms		
MATLAB	MathWorks	http://www.mathworks.com
glmnet	(Friedman et al., 2010)	https://CRAN.R-project.org/package=glmnet
Scanimate	Vidrio Technologies	http://scanimate.vidriotechnologies.com
Calcium imaging deconvolution algorithm	(Pnevmatikakis et al., 2016)	https://github.com/epnev/constrained-foopsi
ViRME (Virtual Reality Mouse Engine)	(Aronov and Tank, 2014)	https://pni.princeton.edu/pni-software-tools/virmen-virtual-reality-matlab-engine
Custom MATLAB code for motion correction, selecting cell regions-of-interest, and extracting fluorescence timeseries	This paper	https://github.com/HarveyLab/Acquisition2P_class

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Christopher Harvey (harvey@hms.harvard.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice

All experimental procedures were approved by the Harvard Medical School Institutional Animal Care and Use Committee and were performed in compliance with the Guide for the Care and Use of Laboratory Animals. Three male C57BL/6J-Tg(Thy1-GCaMP6s)GP4.12Dkim/J mice were used for widefield retinotopic mapping experiments. All other data were obtained from five male C57BL/6J mice (The Jackson Laboratory), which were 8–10 weeks old at the start of behavioral training, and 14–32 weeks old during imaging. A surgery was performed on each mouse before training to affix a titanium headplate to the skull using dental cement (Metabond, Parkell). At least one day after headplate implantation, mice began a water schedule, in which they received 800 µL of water/day. Mouse health was monitored daily. Mice were given additional water if their weight fell below 80% of their pre-schedule weight (mean ± sem 23.2 ± 0.5 g). Mice were housed in pairs of littermates.

METHOD DETAILS

Virtual reality system

Virtual reality environments were constructed and operated using MATLAB-based ViRME software (Virtual Reality Mouse Engine) (Aronov and Tank, 2014; Harvey et al., 2009). A PicoP microprojector (MicroVision Inc.) projected the virtual environment onto the back side of a 24-inch diameter half cylindrical screen. The virtual environment was updated in response to the mouse's manipulations of an open cell Styrofoam spherical treadmill (8-inch diameter, ~135 g). An optical sensor positioned beneath the spherical treadmill measured movements in pitch and roll of the ball (relative to the mouse's body axis). These signals controlled forward/backward and rotational movement in virtual reality, respectively. We recorded the mouse's position in the virtual environment (x/y position), the rotational speed of the spherical treadmill (about the pitch and roll axes), and the mouse's view angle in the environment.

Behavioral training

Mice were on the water schedule for at least five days before behavioral training began. Training sessions were performed daily and lasted 45–60 min at approximately the same time of day each day. Rewards (4 µL of 10% sweetened condensed milk in water) were delivered through a lick spout. Mice were trained to perform the T-maze task using a program of five mazes.

2-Cue Task Training

Maze 1 was a linear track in which the mouse had to run forward to get a reward. After each trial the maze was either lengthened or shortened to maintain an approximate reward rate of 4 rewards/minute. When the mouse completed a trial in less than 15 s, the central corridor would grow by 10 cm on the next trial or shrink by the same amount if the trial was completed in greater than 15 s. The minimum length of the maze was 37.5 cm and the maximum was 3 m (measured as running distance on the treadmill).

In maze 2 the mice had to turn toward a tower above the left or right T-arm in the virtual world. This maze trained the mice to follow a visual cue for reward and improved running skill on the treadmill. Again, after each trial the maze was either lengthened or shortened to maintain an approximate reward rate of 4 rewards/minute with a minimum length of 70 cm and a maximum length of 3.5 m.

In maze 3, mice began to associate colored walls with the cued turn direction. When the tower was on the right, the walls were black. When the tower was on the left, the walls were white. On alternating trials we added a second tower so that there was one on each side, and the mouse had to use the wall color to plan the turn direction. Maze 3 was 4 m in length.

The delay period was gradually incorporated into maze 4, such that the cue offset (delay onset) shifted earlier in the trial. The criteria for advancement to the next maze in the sequence, on mazes 3 and 4, were a trial rate of > 4 trials/minute and > 80% correct for 2–3 consecutive days.

In maze 5, the lengths were fixed with the total length of 4.5 m and a delay period of 2.25 m. The colored walls were either black with white dots or white with black dots followed by a gray striped segment throughout the delay period that was identical across trial types. The entire training program was completed in 4–8 weeks.

3- and 4-Cue Task Training

Training for novel trial type associations was performed during imaging. Mazes were identical to maze 5. On each day, mice were presented with novel trial types after 40 trials of the familiar trial types (black cue-right turn and white cue-left turn). After novel trial types were introduced, familiar trial types and novel trial types were interleaved such that there were equal fractions of left and right turn trials. Mice were first presented with a 3rd cue (crosshatch) and after mice performed all three trial types at above 80% for three consecutive days, we introduced a 4th cue (triangles). For mouse 1, the 3rd cue instructed left turns and the 4th cue instructed right turns. For mouse 2, the cue-turn relationship for novel trial types was reversed. White and black cues maintained consistent cue-turn relationships for both mice. Mice learned novel trial type cue-turn relationships by trial and error while maintaining previously learned relationships for black and white cues.

Bias Correction

Some mice developed biases during training such that left or right turns were favored. During training, we implemented a bias correction. On each trial, the probability that a mouse would be presented with a left turn trial was the fraction of times the mouse turned right on the previous 20 trials. Once mice reached expert levels, biases were rare and bias correction was unnecessary. Bias correction was not used during imaging sessions.

Optogenetic inactivation experiments

AAV1.EF1.dflox.hChR2(H134R)-mCherry.WPRE.hGH was injected into the left and right PPC (2 mm posterior to bregma and 1.7 mm lateral of the midline) of 6–8 week old parvalbumin-cre mice (B6.129P2-Pvalbtm1(cre)Arbr/J, stock number 017320, Jackson Laboratory), and the skull was thinned bilaterally over the injection sites. LEDs of wavelength 465 nm (Optek Technology OVS5MBBCR4) were glued and cemented onto the thinned portion of skull and covered with custom aluminum heat sinks. Training began one week after the surgery. Inactivation experiments began after the mice learned the final version of the task. In a given session, a subset of trials (9%–26%) were randomly selected for inactivation. Stimulation occurred either throughout the first half of the T-stem, the second half of the T-stem, or throughout the entire T-stem. On whole trial inactivations, stimulation occurred from the start of the trial until the mouse either entered a T-arm or until 10 s had elapsed, whichever happened first. Stimulation consisted of 10 ms LED light pulses delivered at 50 Hz (50% duty cycle) with peak power of approximately 4–5 mW/mm². Sessions in which the mouse performed below 65% correct were excluded from the analysis. The difference between performance on control and inactivation trials was compared with the performance-difference generated from 1000 trial-label shuffles.

Surgical procedures

After mice achieved performance greater than 80% correct on the task for five consecutive days, they received ad lib access to water for three days before a cranial window implant surgery. A circular craniotomy with a diameter of 3.1 mm was made over left PPC (stereotaxic coordinates: 2 mm posterior, 1.7 mm lateral of bregma). Three 10 nL injections of a virus mixture containing a 4:1 volumetric ratio of tdTomato (AAV2/1-CAG-tdTomato) to GCaMP6m (AAV2/1-synapsin-1-GCaMP6m) (University of Pennsylvania Vector Core Facility) were made near the center of the craniotomy at a depth of ~275 µm below the dura. Injections were slow (5 min/injection) and continuous (custom air pressure injection system). The pipette (15 µm tip diameter) was advanced using a micromanipulator (Sutter MP285) at a 30-degree angle relative to horizontal to minimize compression of the brain. A glass plug consisting of a single 5 mm diameter coverslip on top of two 3 mm diameter coverslips (#1 thickness; CS-5R and CS-3R, Warner

Instruments) were combined using UV-curable optically transparent adhesive (Norland Optics) and were affixed to the brain with minimal Kwik-Sil (World Precision Instruments) and affixed to the skull using Metabond on the perimeter of the 5 mm coverslip lip. The metabond mixture contained 5% vol/vol India ink, to prevent light contamination from the visual display. Additionally a titanium ring was mounted on top of the headplate. This ring interfaced with the microscope's objective lens through a cylinder of black rubber, to prevent light contamination. Mice resumed training after at least one day of recovery. Imaging began at least three weeks post-injection and was continued for up to 8 weeks. On a given day, we imaged 100 - 300 neurons simultaneously during approximately 200 trials (Figure S2J).

Field sign maps

To visualize Bregma-guided PPC coordinates relative to the retinotopic map in visual areas of the mouse cortex, we performed wide-field imaging in GCaMP6s transgenic mice (C57BL/6J-Tg(Thy1-GCaMP6s)GP4.12Dkim/J) during a stimulus protocol for retinotopic mapping.

Widefield microscope design

Retinotopic mapping was performed with a tandem-lens epifluorescence macroscope (Ratzlaff and Grinvald, 1991). Excitation light (455 nm LED, Thorlabs) was filtered (469 nm with 35 nm bandwidth, Thorlabs) and reflected onto the brain through an inverted camera lens (NIKKOR AI-S FX 50 mm f/1.2, Nikon). Emission light was collected by the same lens, emission-filtered (525 nm with 39 nm bandwidth, Thorlabs), and imaged by a second camera lens (SY85MAE-N 85 mm F1.4, Samyang) onto a CMOS camera (ace acA1920-155um, Basler). Images were collected at 60 Hz, synchronized to the visual stimulus presented on a gamma-corrected 27 inch IPS LCD monitor (MG279Q, Asus).

Visual stimulus design

Stimulus presentation was performed as described in (Marshel et al., 2011). The monitor was placed in front of the right eye at an angle of 30 degrees from the mouse's midline. The stimulus was a black and white checkered moving bar on a gray background, corrected to have constant width (5 degrees) and speed (7 deg/s) in spherical coordinates centered on the mouse's eye (Marshel et al., 2011). The stimulus was presented in blocks containing six repeats of each of four movement directions (up, down, forward, backward). Six blocks were presented per session.

Retinotopy analysis design

Retinotopy was determined by computing the temporal Fourier transform at each pixel and extracting the phase at the frequency of stimulus presentation. The phase images were averaged across all trials of the same orientation and smoothed using a Gaussian kernel with standard deviation 30 μm to obtain horizontal and vertical retinotopic maps. The field sign map was then calculated as the sine between the gradient angles of the horizontal and vertical retinotopic maps. Field sign maps were aligned to Allen Institute field sign maps using control point registration and overlaid with a dorsal map of cortical areas defined by the Allen Mouse Common Coordinate Framework (Mouse and Coordinate, 2016)

Two-photon imaging

Field-of-view location

The field-of-view was selected near the center of the craniotomy, at the center of the viral injection sites. This location was defined based on stereotaxic coordinates but, in separate experiments, showed reproducible positioning relative to areas identified by retinotopic mapping. The imaged fields-of-view are expected to be anterior to visual area AM.

Two-photon microscope design

Data were collected using a custom-built two-photon microscope. A resonant scanning mirror and galvanometric mirror separated by a scan lens-based relay telescope on the scan head allowed fast scanning. An Olympus 25x 1.05 NA objective lens was mounted on a piezo collar (Physik Instrumente) that allowed slower axial scanning. An aluminum box housed collection optics to block light interference from the visual display. Green and red emission light were separated by a dichroic mirror (580 nm long-pass, Semrock) and bandpass filters (525/50 and 641/75 nm, Semrock) and collected by GaAsP photomultiplier tubes (Hamamatsu). A Ti:sapphire laser (Coherent) delivered excitation light at 920 nm with an average power of ~35-70 mW at the sample. The microscope was controlled by ScanImage (version 4; Vidrio Technologies). The spherical treadmill was mounted on an XYZ translation stage (Dover Motion) to position the mouse under the objective.

Image acquisition

Four imaging planes were acquired by volumetric scanning at 5.3 Hz with a resolution of 512 \times 512 pixels (500 $\mu\text{m} \times$ 500 μm) for each plane. Planes were separated by 25 μm axially between 120 and 250 μm below the dura. Imaging was continuous over behavioral sessions lasting 45 min to 1 hr. Bleaching of GCaMP6m was negligible over this time. Approximately every 20 min, slow drifts of the field of view were manually corrected using comparison to a reference image. The imaging frame clock and an iteration counter in VIRMEn were recorded to synchronize imaging and behavioral data.

Chronic imaging

One field-of-view was acquired for each of the five mice over a period of 3 to 8 weeks. The same plane was identified on consecutive days using coarse alignment based on superficial blood vessels followed by careful alignment to reference images at various levels of magnification in the red channel (using tdTomato expression). AAV-mediated expression of GCaMP6m provides high signal-to-noise compared to other methods; however, viral expression is known to increase over months which can lead to compromised signal over

time, which is correlated with nuclear localization of the indicator (Chen et al., 2013). For this reason, imaging was discontinued when fields-of-view contained several cells with GCaMP6 in the nucleus, and all cells with nuclear localization were excluded from analysis (Figure S2I). These methods are in accordance with other long-term imaging studies (Huber et al., 2012). Event rates of all analyzed cells were stable across time along with other properties of the population activity (Figure 5). Moreover, our ability to model and predict neuronal activity using behavioral features remained consistent throughout the duration of this experiment. For these reasons we have no reason to believe cell health was an issue in this work.

Pre-processing of imaging data

Within-session processing

We developed an approach to identify cell bodies in calcium imaging data that combines automated proposals based on image time series statistics with human supervision to provide an efficient but transparent signal extraction procedure. The approach was implemented as part of a custom MATLAB software pipeline for motion correction, definition of putative cell bodies, and extraction of fluorescence traces (https://github.com/HarveyLab/Acquisition2P_class). We chose to develop custom scripts for this pipeline because activity-based methods for ROI segmentation were not available at the time of our experiments.

Following motion correction using the Lucas-Kanade method (Greenberg and Kerr, 2009), candidate locations containing putative cell bodies were selected manually in the mean intensity image of the acquisition (Figure S2A). Fluorescence sources within a square neighborhood (60 μm edge width) around the selected location were then identified automatically based on the correlation structure of the pixel time series. Since the fluorescence time series of pixels belonging to the same source are expected to be highly correlated, sources larger than a single pixel appear in the pairwise pixel time series correlation matrix as clusters with strong within-cluster and weaker across-cluster correlation (Figure S2B). Formally, the correlation matrix was considered to represent a weighted undirected graph with one vertex per pixel, connected by edges with weights given by the correlation between the fluorescence time series of the pixels at either end. The optimal segmentation of the graph was then found using an eigenvector-based approximation of the normalized cuts criterion, followed by k-means clustering to obtain a binary mask for each source (Figures S2C and S2D). This criterion maximizes the fraction of within-segment weights (correlations) over total weights. Using segmentations based on pixel time series correlations prevents the segmentation of inactive cells, even when these cells were visible in the mean image. Fluorescence time series were computed by averaging across all pixels within the binary mask.

These source proposals were manually classified into cell bodies, non-cell sources (excluded from further analysis), and background (neuropil), based on their appearance in the mean intensity image and their fluorescence time series (Figures S2A and S2D). Each putative cell was paired with a background source from the same 60 μm neighborhood. Neuropil contamination was removed from the cell fluorescence time series by subtracting the associated background time series, scaled by a contamination factor. The contamination factor was calculated by regressing the cell fluorescence against the background time series using an iteratively re-weighted least-squares algorithm (*robustfit* in MATLAB) that discounts large deviations from the fitted linear relationship, such as fluorescence transients in the cell (Figures S2E and S2F).

Segmentation and neuropil subtraction were manually verified for each putative cell and adjusted when necessary using a graphical user interface that showed the mean intensity image, current segmentation results, and both raw and background-subtracted fluorescence time series. Manual adjustments of the segmentation were usually made to obtain clean background fluorescence traces absent of distinct sources. Manual adjustments of neuropil subtraction were used to correct for an overestimation bias of the re-weighting procedure when a cell's activity was highly correlated with the neuropil. In such cases, subtraction was adjusted to the highest level that did not result in visually apparent negative-going transients in the neuropil-subtracted trace.

The event rate was estimated using a previously described deconvolution algorithm (Pnevmatikakis et al., 2016) to minimize the impact of indicator kinetics. This method estimates the relative firing rate of each neuron over time but cannot be used to confidently identify single spikes. We therefore refer to deconvolved traces as an estimated event rate.

Across-session processing

Binary masks for all fluorescence sources were identified on each day separately and then aligned across days. The algorithm ranked cells across imaging days with their most likely matches based on proximity after alignment and anatomical image correlation (a 60 μm box around the centroid of the cell). Matches were then verified by eye. This method has advantages over other commonly used approaches. Other approaches often use a single map of ROI masks for all days, such that this map is transformed on each day to best fit that day's imaging alignment. Slight deviations in the axial plane of the image or other sources of in-plane distortion could lead to slight offsets in masks from day-to-day relative to the ideal alignment. Such slight offsets could result in contamination from activity in other cells, dendrites, and axons. Our approach identifies signal sources on each day and thus avoids any potential contamination from other signal sources. We then align the signal sources identified on each day to those from other days. The only error that could result is in incorrectly calling two signal sources as the same across days. However, to prevent such errors we visually compared the anatomical images to make sure the signal sources appeared to correspond to the same cell. If a cell could not be confidently identified on a given day, the data were excluded on that day. As a result, our approach resulted in an incomplete map of all cells across all days. We note that cells had to have some activity (calcium transients) in order to be identified on a given day. This activity requirement for the identification of each cell could potentially result in an underestimation in the extent to which cells gain and lose task related activity. Cells were more likely to have a defined mask on days that were nearby in time due to variable activity and viral expression of the indicator GCaMP6m (Figure S2).

Data inclusion criteria

To include data from a population of neurons in a given mouse, behavioral performance had to exceed 70% correct for greater than 20 consecutive days to ensure that the mouse had learned the task. All imaging sessions were included between the first day with qualitatively bright GCaMP6m expression and the first day in which nuclear localization of GCaMP6m was present in more than a few cells.

No statistical methods were used to predetermine sample size. The experiments were not randomized, and the investigators were not blinded to allocation during any stage of the study. Data groups for randomization and blinding were not present in our study. Replication of results was not performed.

QUANTIFICATION AND STATISTICAL ANALYSES

For analyses of behavior, n refers to the number of mice studied, which is typically n = 5 mice. For comparisons of population activity, n refers to the number of simultaneously imaged neuronal populations studied. One simultaneously imaged neuronal population was studied per mouse. Typically n = 5 neuronal populations for these comparisons. For analyses of single cells, n refers to the number of cells and varied depending on the analysis, such as the interval between days being compared. n values are reported in figure legends and can also be seen in [Figures S2J](#) and [S2K](#).

Statistical analyses are described in the [Results](#), figure legends, and below in this section. In general, 95% confidence intervals were used to define significance. Most statistical analyses used standard parametric approaches, such as t tests and ANOVA. When assumptions of these parametric approaches likely would not be met, such as assumptions about the distribution of the data, we used non-parametric permutation tests. Quantitative approaches were not used to determine if the data met the assumptions of the parametric tests.

General analysis procedures

For visualization of mean and single trial activity, activity was spatially binned into 66 segments (60 in the stem and 6 in the arms) followed by ~2 s (10 imaging frames) during the inter-trial interval. Spatially binned segments typically contained 1 imaging frame each per trial. Data were interpolated to fill in gaps in which a bin contained no imaging frames on a trial. These analyses were only used for fine-resolution measurements of locations of peak activity. For the majority of the analyses (all except [Figure 2](#)), we binned the data into larger segments (21 segments, 23 cm/bin). Neuronal activity and behavioral parameters were averaged in each bin in this case. Each bin typically contained 2-3 frames/trial. Unless otherwise noted, all analyses were performed on correct trials only. Correlation coefficients were calculated from Pearson's correlations unless otherwise noted.

Identification of significant peaks of activity

Here we used 60 spatial bins during the T-stem rather than 21 spatial bins because we were interested in achieving greater resolution of spatial reorganization. Mean activity across 60 spatial bins were calculated for correct white cue-left turn and black cue-right turn trials. To identify statistically significant peaks of activity, behavior data time courses were circle-shifted by a random amount relative to neuronal activity time courses, and new means were calculated for 1000 random shifts. All locations where the mean activity in the unshifted data was greater than activity in 950 shuffles for three consecutive bins were considered to contain a significant peak of activity. For all analyses where peaks were compared across days, we tracked peaks that were labeled with high confidence (unshifted data was greater than activity in 990 shuffles for three consecutive bins). Peaks were labeled as 'gained' or 'lost' if in the absent session, there was below 95% significance for a peak at that location. We found this gap between thresholds for the presence or absence of a peak to be important for limiting measurement noise. By these criteria for change, we found peak consistency across odd and even trials within one session to be $83.2 \pm 2.1\%$.

GLM encoding model

On each day separately, we fit a Poisson Generalized Linear Model ([Friedman et al., 2010](#)) to the estimated event rates of each cell based on measured behavioral and task variables. While our estimated event rates were not necessarily Poisson distributed, mean responses scaled with the variance and rates were larger than zero, suggesting that the Poisson model was appropriate for our data.

Model parameters

All measured behavioral and task-related variables were temporally averaged into bins to match the sampling rate of imaging frames. Variables include x and y position in the virtual world, running speed on the pitch and roll axes of the spherical treadmill, visual cue onset and offset locations, reward delivery events and trial end ([Figure S4](#)). Variables provided input for basis functions that were distributed either in space or in time to produce 144 predictors in total. The maze was divided into 36 spatial boxcar filters and convolved with a Gaussian filter separately for right and left turn trials to make the first 72 filters. The onset of each visual cue contributed 4 basis functions (16 in total for all four cues) that spanned the first 2 s of the trial. For cue offset (delay period onset), 2 basis functions extended for 1 s preceding cue offset, and 4 basis functions extended for 2 s following cue offset (24 in total for all four cues). Running speed signals were extended 1 s forward and backward in time for translational and clockwise and counterclockwise rotational motion (4 filters forward and 4 filters backward for each of 3 speed signals for a total of 24 filters) to model predictive and

responsive signals. Trial-end and reward events each contributed 4 basis functions that extended for 2 s forward in time (8 filters total). All temporal filters spanned 1 s, overlapping and evenly distributed within each set.

Fitting

We used the `glmnet` package in R to fit GLMs. Each day was divided into 10 evenly distributed chunks (first tenth of session, second tenth of session, and so on) and then sub-divided into 11 numbered pieces within each chunk. All pieces with the same number were then combined into groups 1-11. This resulted in 11 groups that contained data that was evenly distributed across the imaging session (Figure S3). Eight of these groups were randomly chosen and used as cross validation folds during fitting, and the other three groups were combined and used to test model predictions (8 groups made up 73% fitting and 3 groups made up 27% testing predictions). For some figures, we compare models that were trained on a subset of data from a single session. First we divided data from a single session into halves, either interspersed chunks throughout the duration of the session or simply the first and second half of the session. We then treated these halves as independent sessions that were divided as we previously describe (73% fitting and 27% testing predictions). These models were only used for comparison to the appropriate opposite half within a single session. Parameters were fitted for each cell separately with elastic net regularization consisting of 10% L2 and 90% L1 methods. Model fitting was conducted on the Orchestra High Performance Compute Cluster at Harvard Medical School. This shared facility is partially supported by NIH through grant NCRR 1S10RR028832-01.

Analysis of model

Deviance explained was used as the metric of model fit. It was calculated by comparing the activity predicted by the model to the actual activity, based on the average real and predicted activity levels on white cue-left and black cue-right turn trials. Deviance explained was calculated based only on data not included in the fitting procedure. It was compared to a null model in which the predicted event rate was 1 (the normalized mean rate of the entire fitting set). We also computed deviance explained using predictions of the full time series of activity for single frames. In this case, the distribution of fits was similar, except lower than for trial-averaged data, as expected. To determine whether a model prediction described test data significantly above chance, we compared the deviance explained to the distribution of deviance explained for models in which, before model fitting, the time series of neuronal activity was circle-shifted relative to the time series of behavior data. We found the upper limit of deviance explained in models that were fit on shuffled data to be 0.2 (for trial-averaged activity). Therefore, we determined models with greater than 0.2 deviance explained on test data to have predictions significantly above chance performance. Cells with activity that could not be significantly predicted by our models could still have behaviorally relevant activity patterns that were simply not described by our set of filters. Therefore, when we refer to cells having gained or lost activity-behavior relationships, we refer specifically to the set of relationships included in our models.

By comparing model fits between any two days or between subsets of data from a single day, we could determine whether a given neuron's activity-behavior relationship was consistent over time, gained or lost a relationship to a behavioral feature, or switched from encoding a particular set of behavioral features to encoding another. If models with significant predictions could be fit on each individual day's data for a given pair of days and if the models also provided significant predictions of the other day's activity, then we called the activity-behavior relationship consistent. If models with significant fits could be fit on each individual day's data for a given pair of days and if the models did not provide significant predictions on the other day, we determined a switch in activity-behavior relationship had occurred. If only one day had a model that could predict the neuron's activity significantly well, then we concluded that a gain or a loss of activity-behavior relationship had occurred. This approach works well when behavioral variables are not highly correlated. However, if behavioral variables are very highly correlated, then it is difficult for the model to determine to which variable weight should be attributed. If variables were highly correlated on one day but not correlated in another day, then incorrect assignments of switches in activity-behavior relationships could emerge. To be conservative and to account for any potential cases in which non-modeled variables were highly correlated with modeled variables (on some days but not others), we did not require symmetry in model performance across days for an activity-behavior relationship to be considered consistent. That is, if a model prediction on day i and day j both had significant predictions of activity and if the model from day i predicted well the activity from day j , but not vice versa, then a consistent activity-behavior relationship would still be considered to be present. In this case, if behavior variables x and y were highly correlated on day j but not on day i , then the model on day j might incorrectly attribute weights on day j leading to a poor prediction on day i . However, because the variables x and y were not highly correlated on day i , then the model would correctly attribute weights on day i , successfully predicting activity on day j despite having correlated variables x and y on day j . This method would still fail to identify a consistent activity-behavior relationship if the 'real' encoded feature (for example, mental state) was correlated with different filters in the model on each day. These subtle behavioral confounds are difficult to remove completely and should be considered in the interpretation of this work and most other behavioral studies. Through the use of virtual reality we could control the sensory environment and track behavioral metrics with great precision, incorporating these features into our models and thus minimizing the aforementioned concerns.

Metric of consistency in activity-behavior relationships

We defined the consistency of activity-behavior relationships for each cell by the likelihood that an encoding model fit on a given day continued to provide significant predictions of the cell's activity for days and weeks. We found the fraction of models that provided a significant prediction of a given cell's activity over each interval between fitting and prediction sets (1-30 days). We then fit an exponential decay, weighting each point by the number of model comparisons available (for example, there were more model comparisons where Δ days = 1 than Δ days = 30). We only included cells that had significant fits on at least half of

days where Δ days = 0 because cells with a low starting point necessarily had slow decays because the fraction of significant model fits was bounded by zero.

Contribution of task and behavioral features to the activity of a cell

To calculate the contribution of each task and behavioral variable to a given model, we computed the standard deviation of the linear part of the model that was related to the behavioral variable of interest (the standard deviation of beta coefficients crossed with relevant behavioral filters of behavioral data). This provided us with the extent to which a behavioral variable modulated the activity of the neuron. For filter groupings in position/cue, velocity and ITI contribution calculation see [Figure S4](#).

Additional notes on GLM

A main benefit of the GLM is that it mitigates effects of behavioral variability on neural changes in our analyses. Because the GLM does not include every possible behavioral variable, it remains possible that behavioral variability that was not modeled could potentially contribute to the neural activity changes we observed. Importantly, however, our analyses using the GLM put an important constraint on any potential additional behavioral features that could contribute to neural activity changes. We modeled a cell's activity based only on the measured behavioral features and then only considered changes in activity in cells that had a good model of their activity based on the behavioral features measured. That means, if a cell was strongly driven by non-modeled behavioral features, then it is unlikely to have a good model fit and thus unlikely to be included in our analysis. The one exception is if an un-modeled behavioral feature is highly correlated with a modeled behavioral feature on some sessions but not on other sessions. Therefore, our GLM analyses are limited not simply because of all un-modeled behavioral features but rather only because of un-modeled behavioral features that are highly correlated with modeled features. The design of the GLM analyses was thus intended to limit as best as possible the potential contributions from un-modeled behavioral features.

The GLM approach has the potential to track stable relationships between neuronal activity and behavior features across days that traditional approaches might miss. For example, if a neuron had activity related to the running patterns of a mouse and if these running patterns changed relative to maze positions across days, the GLM could reveal a stable activity-behavior relationship over time that would be missed if only maze position were analyzed.

Decoding

We used C-Support Vector Classification with a linear kernel for all decoders. For some decoders, we considered the activity of single cells ([Figures 3 and 5K](#)) across the entire trial. In these cases, the decoder was trained using data from all 21 spatial bins in the trial. In other cases, in which we wanted to assess the time course of information, we trained decoders on each spatial bin separately using the activity of all neurons or subsets of neurons ([Figures 5A and 6](#)). When decoders incorporated a smaller number of cells (20 cells each for subgroups in [Figure 6](#)) we divided the maze into larger bins to account for the fact that each cell only had trial type information for a small portion of the trial. For decoding analyses, the data were divided into two-thirds for training/validation and one-third for testing. The regularization weight hyperparameter C was selected using a random search with 10-fold cross validation on a subset of training sets across mice. The specific setting of the hyperparameter did not greatly affect the accuracy of our decoders. The same hyperparameter value (C = 100) was used for all datasets. Significant decoding accuracies in [Figure 3](#) were determined by bootstrap analysis in which behavioral data were circularly rotated relative to each cell's neuronal activity. Cells in which real data performed better than 950 of 1000 shuffles were determined to have significant decoding accuracy. Using this statistical threshold, it is thus expected that 5% of cells would have significant decoding accuracy by chance.

Dimensionality reduction of neuronal and behavioral data

We reduced the dimensionality of single trial data into a space that captured the most variance across trial types and epochs of the trial. We performed principal component analysis (PCA) on averaged population responses across these different conditions. We constructed a data matrix \mathbf{X} of size $N_{\text{neurons}} \times N_{\text{conditions}}$, in which columns corresponded to spatially binned z-scored population response vectors for a given segment of the trial ([Figure 7](#)). The principal components (PCs) of this data matrix were vectors \mathbf{v}_a of length N_{neurons} indexed by the number of PCs. For visualization, we then projected single trial data onto the first 3 PCs.

The motivation for this analysis was to visualize differences in single trial data that corresponded to differences in our conditions of interest. In [Figure 7D](#) we were interested in the evolving population activity throughout the duration of the trial for different trial types on a single day. Our neuronal data matrix \mathbf{X} included 345 neurons identified on a single day and 7 spatial bins evenly spanning the duration of the trial for all 4 trial types (28 conditions). We projected single trial data onto the first 3 PCs of this matrix from individual spatial bins 1, 4 and 7 in the cue period, delay period and turn period. This analysis was solely for visualization. All quantified decoding was performed in the full dimensional space.

DATA AND SOFTWARE AVAILABILITY

Data are available upon request to the Lead Contact. Custom MATLAB code for motion correction, selecting cell regions-of-interest, and extracting fluorescence timeseries is available on Github at https://github.com/HarveyLab/Acquisition2P_class.

Supplemental Figures

Cell

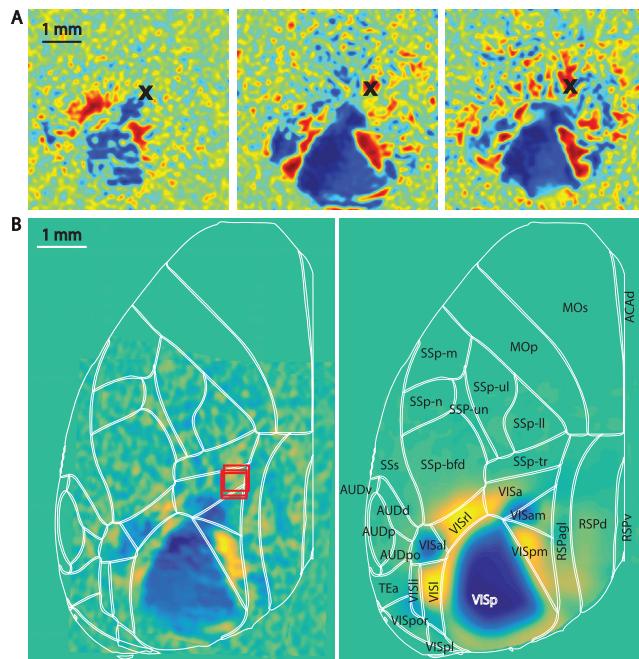
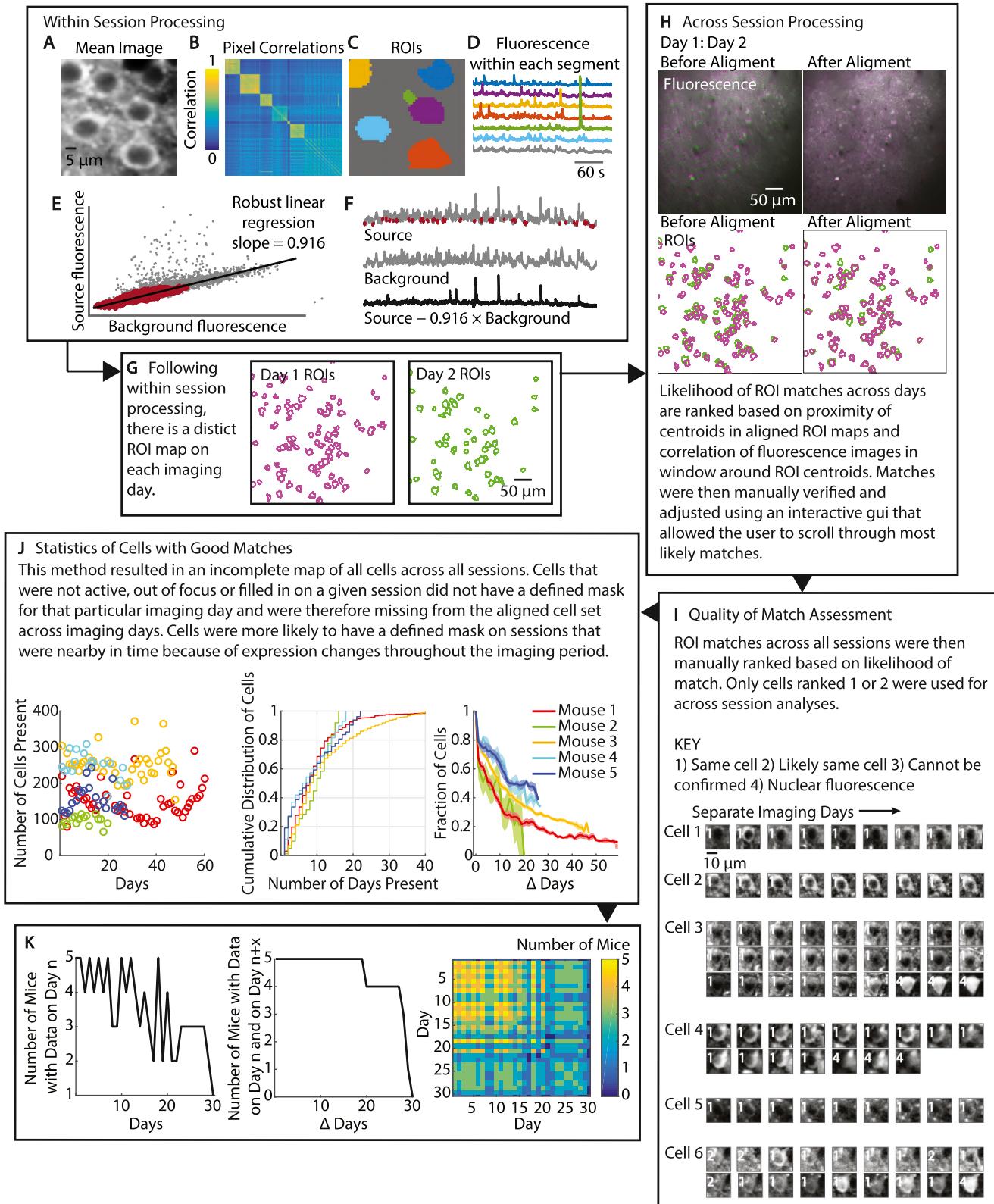


Figure S1. PPC Coordinates Relative to Retinotopic Maps, Related to Figure 1

(A) Field sign maps for three mice. 'X' indicates center of PPC coordinates.

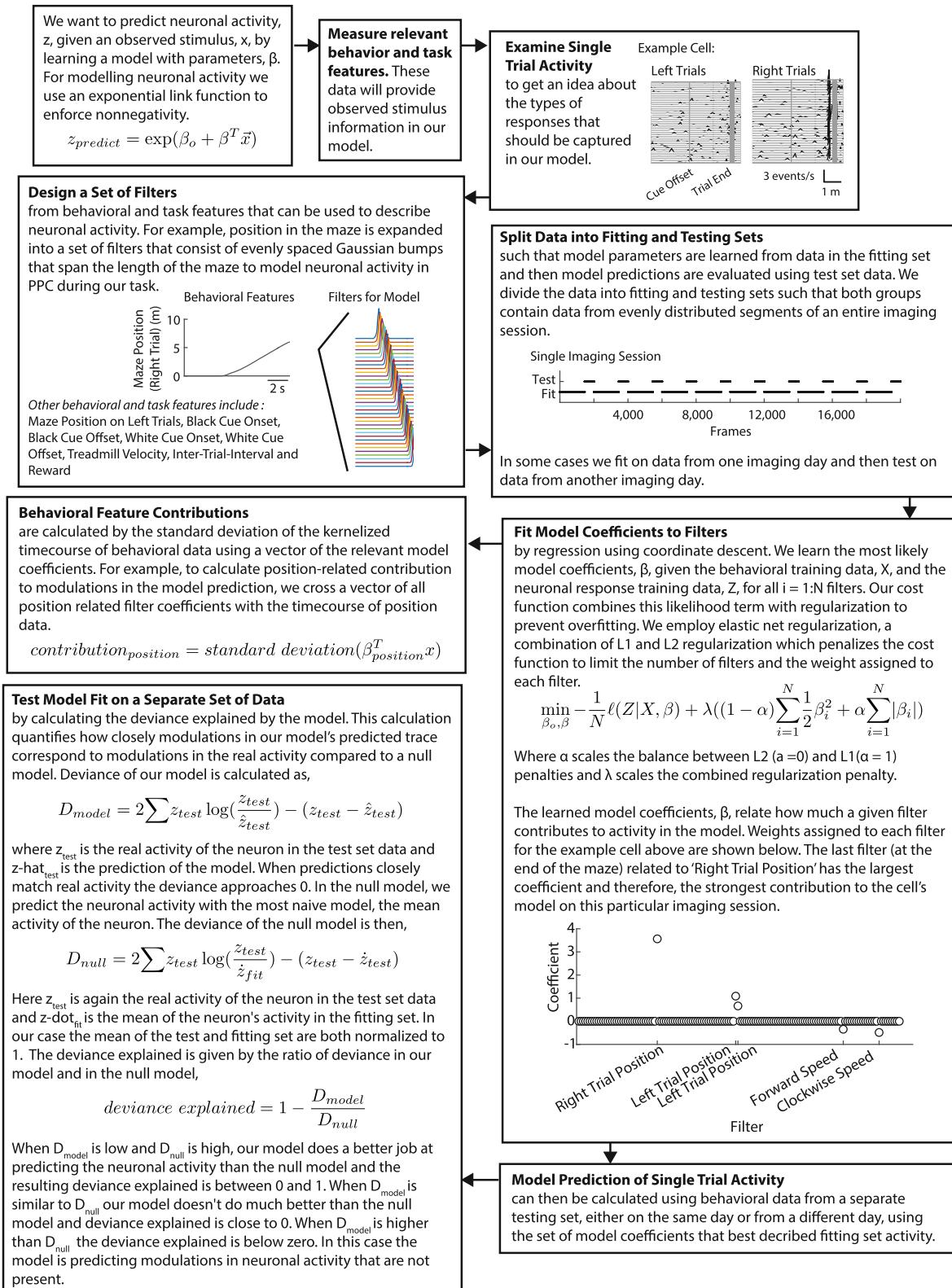
(B) Left: Combined field sign maps overlaid with a dorsal map of cortical areas defined by the Allen Mouse Common Coordinate Framework ([Mouse and Coordinate, 2016](#)). Right: Mean field sign maps from 79 experiments (data from Allen Institute for Brain Science).



(legend on next page)

Figure S2. Cell Identification Protocol across Days and Number of Samples for Various Day Comparisons, Related to Figure 1

- (A) Fluorescence image of a 60 μm neighborhood of cells.
- (B) Correlation matrix between the time series of pixel values within the 60 μm neighborhood.
- (C) Segmentation of pixels within 60 μm neighborhood.
- (D) Mean fluorescence time series of pixels within each segment.
- (E) ROI fluorescence regressed against background fluorescence. The neuropil contamination factor is the slope of the best fit line using bottom 8th percentile of ROI fluorescence (red).
- (F) Top: ROI fluorescence, bottom 8th percentile labeled in red. Middle: Neuropil fluorescence. Bottom: ROI fluorescence with (contamination factor * background) subtracted.
- (G) ROI maps found using protocol in panels A-F from two example imaging days.
- (H) Top: Overlay of average fluorescence signal from the example imaging days before and after registration based on intensity of tdTomato expression. Bottom: ROI outlines found using protocol in panels A-F before and after image registration using the transform from fluorescence image alignment.
- (I) Six example cells across days.
- (J) Left: Number of cells with good matches identified on each imaging day for each mouse. Middle: Cumulative distribution of number of imaging days with good matches for all cells in each mouse. Right: Probability that a cell was identified with a good match on two days separated by various intervals.
- (K) Left: Number of mice with data on each imaging day. Middle: Number of mice with data on two days separated by various intervals. Right: Number of mice with data for each day comparison.

**Figure S3. GLM Fitting Procedure, Related to Figure 4**

See figure for details.

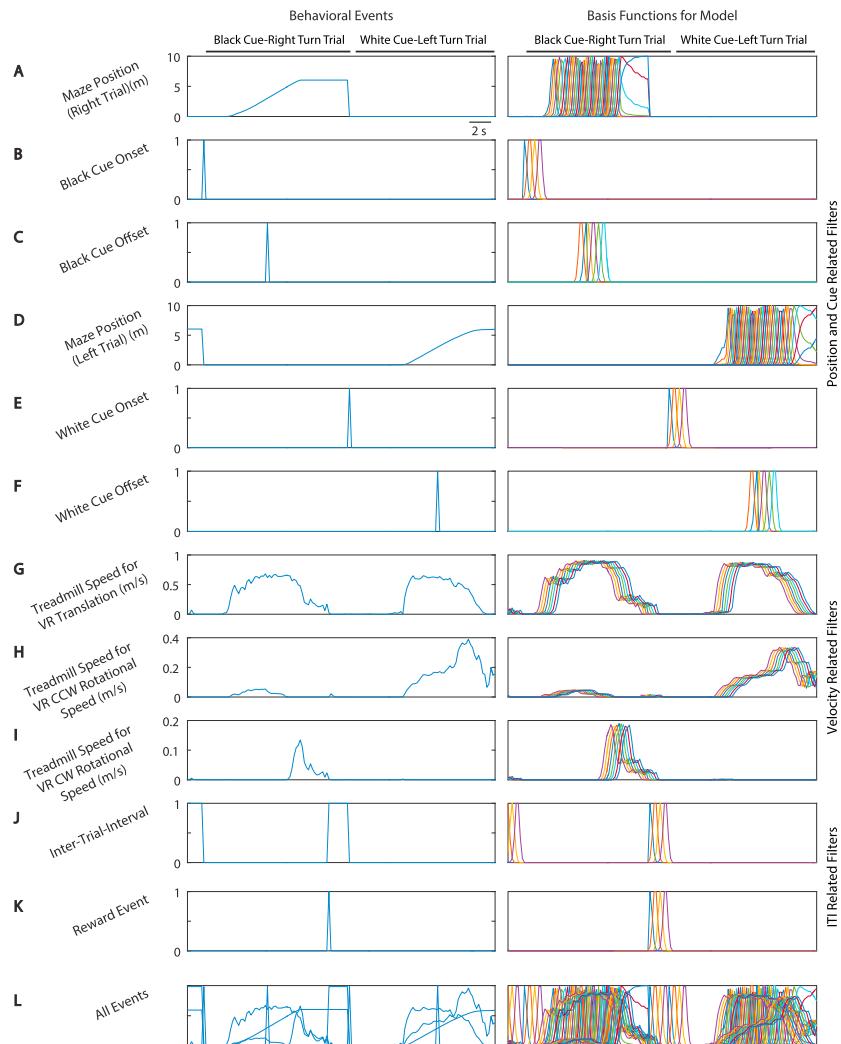


Figure S4. Basis Functions for Encoding Model, Related to Figure 4

Trial and behavioral measurements during each imaging frame (left column) were expanded into a set of basis functions that were incorporated into the GLM (right column). Filter groupings used in contribution calculation for Figure 4I shown in right margin.

(A) Left: Maze position on right turn trials. Right: 36 spatial boxcar filters of position spanning the length of the maze were convolved with a Gaussian filter for right turn trials.

(B) Left: Black cue onset. Right: 4 Gaussian basis functions that span the first 2 s of black cue-right turn trials.

(C) Left: Black cue offset (delay period onset). Right: 6 total basis functions, 2 basis functions extended for 1 s preceding cue offset and 4 basis functions extended for 2 s following cue offset.

(D) Left: Maze position on left turn trials. Right: 36 spatial boxcar filters of position spanning the length of the maze were convolved with a Gaussian filter for left turn trials.

(E) Left: White cue onset. Right: 4 Gaussian basis functions that span the first 2 s of white cue-left turn trials.

(F) Left: White cue offset. (delay period onset). Right: 6 total basis functions, 2 basis functions extended for 1 s preceding cue offset and 4 basis functions extended for 2 s following cue offset.

(G-I) Left: Movement of spherical treadmill. Right: 8 basis functions total for each of 3 running speed signals were extended 1 s forward and backward in time to model predictive and responsive signals.

(J) Left: Inter-trial interval. Right: 4 basis functions that extended for 2 s forward in time following trial end.

(K) Left: Reward times. Right: 4 basis functions that extended for 2 s forward in time following reward.

(L) Left: All trial and behavioral measurements. Right: All basis functions for GLM, excluding novel cue onset and offset (same as black and white cue basis functions).

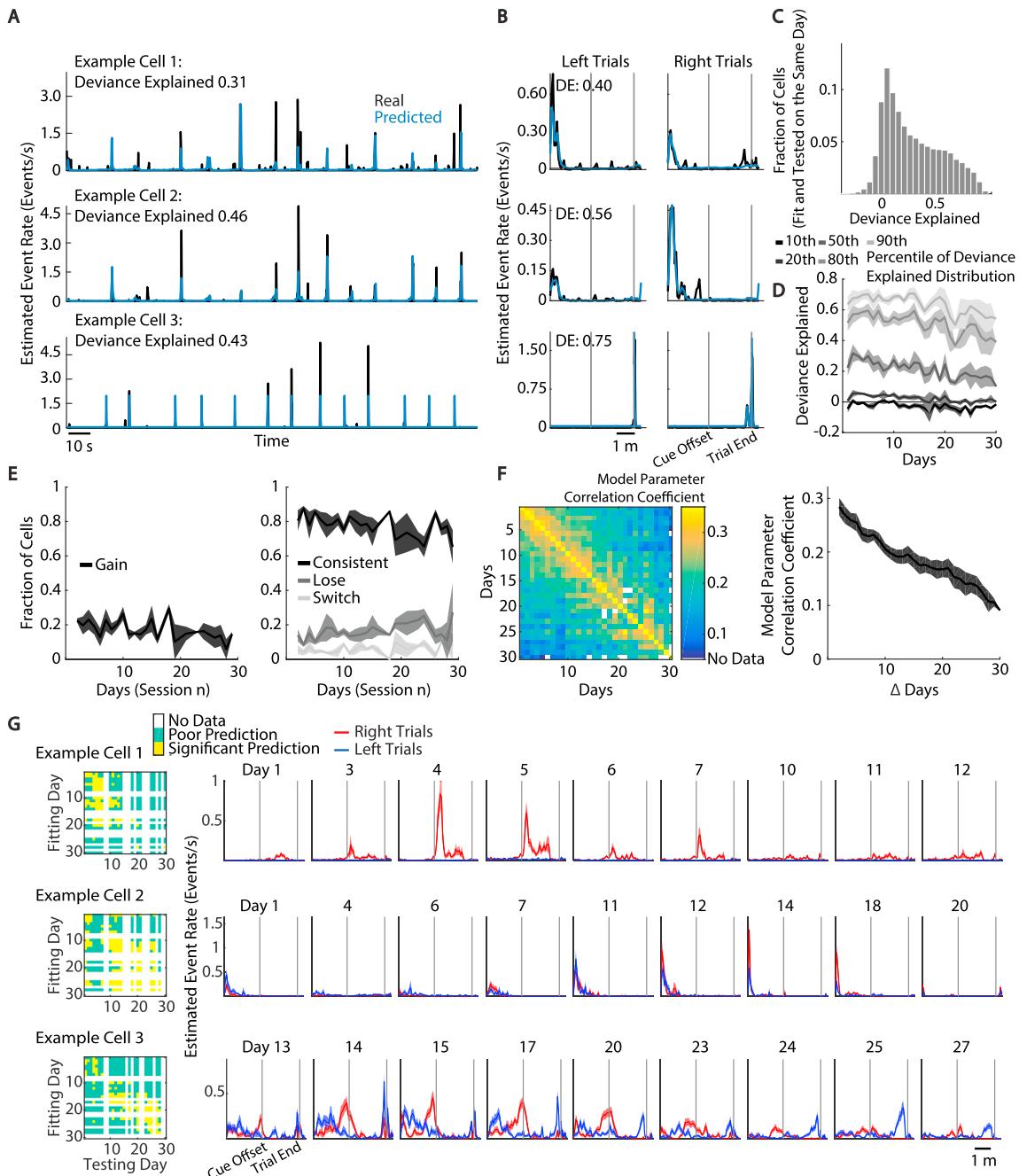


Figure S5. Fitting Activity-Behavior Relationships with a Generalized Linear Model, Related to Figure 4

- (A) For three example cells, a segment of activity is shown (black) along with the activity predicted from the GLM (blue). Deviance explained is calculated directly from single frame predictions.
- (B) Same as in panel A, except for the mean activity on white cue-left turn and black cue-right turn trials. Deviance explained is calculated from trial averaged predictions, concatenated white cue-left turn and black cue-right turn mean activity.
- (C) Distribution of the quality of model fits measured by deviance explained compared to a null model for trial averaged predictions (STAR Methods). The model was fit and tested on data from the same day. $n = 17,353$ model fits across cells and days.
- (D) Distribution of deviance explained for models fit on each day divided into five groups. Similar distributions of fits were apparent on each day. Shaded regions indicate mean \pm sem for $n = 5$ mice (fewer mice on some later days, see Figure S2K for the number of mice at each interval).
- (E) Left: For cells in which a model with a significant prediction of activity could not be developed on day n , the fraction of cells that had a model with a significant prediction of activity that could be developed on day $n+1$. Shaded region indicates mean \pm sem across mice $n = 5$ (fewer mice on some later days). Right: For cells with a significant model prediction on day n , the fraction of cells that had consistent (black), lost (medium gray), and switched (light gray) activity-behavior relationships on day $n+1$. Changes were as defined in Figure 4D. Shaded region indicates mean \pm sem. $n = 5$ (fewer mice on some later days).

(legend continued on next page)

(F) Left: Kendall rank correlation coefficients of model beta coefficients fit on separate days. Diagonal values are 1. The values were averaged across all cells for each mouse and then averaged across mice. Each entry has a variable number of data points (and in some case no data points, white values) due to varying durations of imaging periods for mice and gaps between imaging sessions. See [Figure S2K](#) for the number of mice at each interval. Right: Average correlation coefficient when fit and tested on data separated by n days. Shading indicates mean \pm sem. See [Figure S2K](#) for the number of mice at each interval.

(G) Examples of cells that had ‘modes’ of activity. ‘Modes’ were commonly observed such that neurons had long stretches of similar activity followed by abrupt changes. Left: Matrix of fitting and testing comparisons binarized as significant predictions and poor predictions. Right: Mean activity for example cells.

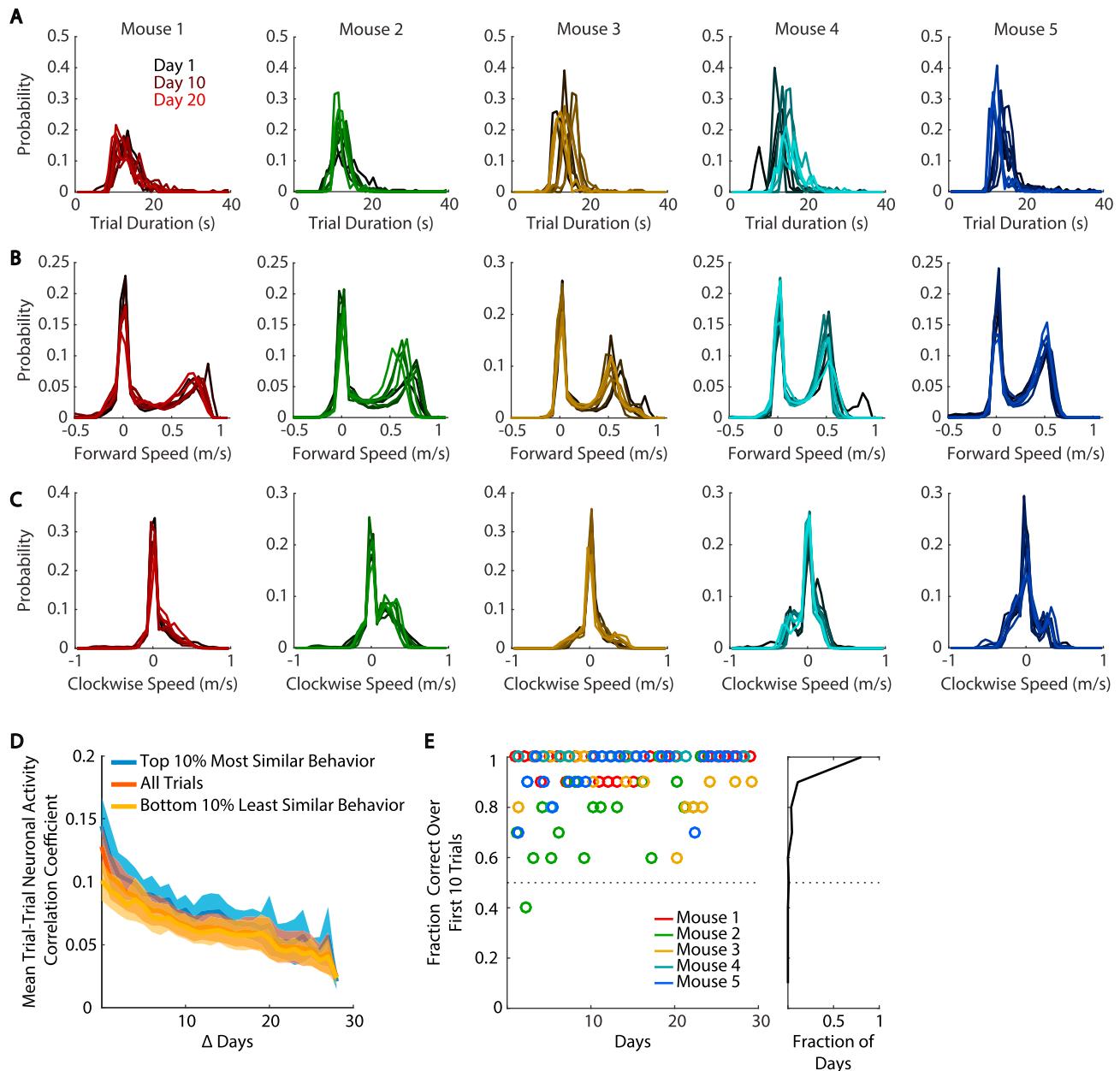


Figure S6. Behavioral Features Were Variable but Had Overlapping Distributions across Days, Related to Figure 4

- (A) For each mouse, the distribution of trial durations from cue onset to trial end.
- (B) For each mouse, the distribution of treadmill speed for forward translation in the virtual environment (about the pitch axis relative to the mouse's body axis).
- (C) For each mouse, the distribution of treadmill speed for rotation in the virtual environment (about the roll axis relative to the mouse's body).
- (D) Similarity in population activity as a function of days separating sessions for the trials with the most or least similar behavioral patterns. Behavioral patterns were compared using an all-session mean-subtracted vector of all measured behavioral parameters (forward speed, clockwise speed, view angle, position along stem axis, and position along arm axis) at each position in the maze (5 parameters x 21 spatial bins). Similarity was determined as the pairwise trial-trial correlation coefficient of this behavioral parameter vector. Population activity patterns were compared using a vector of population activity (neurons x 21 spatial bins). Similarity was determined as the pairwise trial-trial correlation coefficient of this population activity vector. Trials with more similar behavioral patterns were often found on sessions that were separated by a smaller time interval. However, there were some pairs of sessions in the top 10% most similar that spanned the full duration of our experiment, and some pairs of sessions in the 10% least similar that were found in neighboring days. Error bars indicate mean \pm sem ($n = 5$ mice; some large interval data points had fewer than 5 mice, see Figure S2K).
- (E) Left: Behavioral performance over the first ten trials on each imaging day for all mice. Right: Cumulative density plot of behavioral performance on the first ten trials of imaging days.

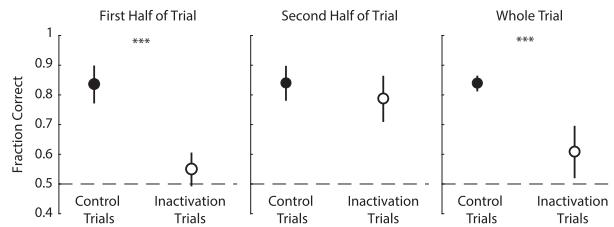


Figure S7. Optogenetic Inactivation during First Half and Second Half of Trials, Related to Figure 6

Optogenetic inactivation during the first half of the T-stem (left), second half of the T-stem (middle), and entire T-stem (right). For each manipulation, trials were pooled across multiple sessions. Points indicate mean \pm sem. $n = 4$ mice. *** indicates $p < 0.001$ based on bootstrap shuffle of control and inactivation trial labels. $p = 0.06$ for the second half of the T-stem (middle).